JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

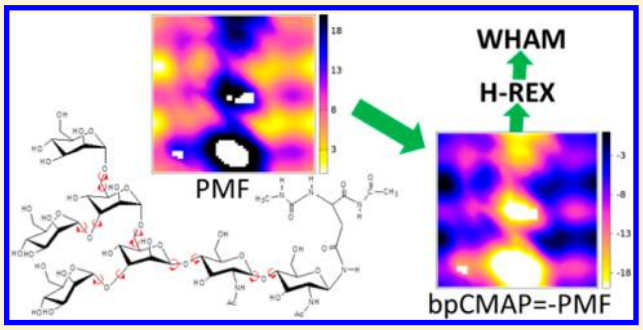
# Conformational Sampling of Oligosaccharides Using Hamiltonian Replica Exchange with Two-Dimensional Dihedral Biasing Potentials and the Weighted Histogram Analysis Method (WHAM)

Mingjun Yang and Alexander D. MacKerell, Jr.*

Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland, Baltimore, Maryland 21201, United States

Supporting Information

**ABSTRACT:** Oligosaccharides and polysaccharides exert numerous functional roles in biology through their structural diversity and conformational properties. To investigate their conformational properties using computational methods, Hamiltonian replica exchange (H-REX) combined with two-dimensional grid-based correction maps as biasing potentials (bpCMAP) significantly improves the sampling efficiency about glycosidic linkages. In the current study, we extend the application of H-REX with bpCMAP to complex saccharides and establish systematic procedures for bpCMAP construction, determination of replica distribution, and data analysis. Our main findings are that (1) the bpCMAP for each type of glycosidic linkage can be constructed from the corresponding disaccharide using gas-phase umbrella sampling simulations, (2) the replica distribution can be conveniently determined following the exact definition of the average acceptance ratio based on the assigned distribution of biasing potentials, and (3) the extracted free energy surface (or potential of mean force (PMF)) can be improved using the weighted histogram analysis method (WHAM) allowing for the inclusion of data from the excited state replicas in the calculated probability distribution. The method is applied to a branched N-glycan found on the HIV gp120 protein, and a linear N-glycan. Considering the general importance of N-glycans and the wide appreciation of the sampling problem, the present method represents an efficient procedure for the conformational sampling of complex oligo- and polysaccharides under explicit solvent conditions. More generally, the use of WHAM is anticipated to be of general utility for the calculation of PMFs from H-REX simulations in a wide range of macromolecular systems.




## INTRODUCTION

Oligo- and polysaccharides, as well as other glycoconjugates, are essential components in all forms of life, where they have roles as fuels, in energy storage, as structural scaffolds, as molecular recognition motifs, and modulatory factors.[1] Besides numerous biological roles, these molecules have applications as biofuels,[2] biocompatible, and biodegradable materials,[3] human vaccines against viruses,[4−6] and excipients in drugs, as well as other pharmaceutical uses.[7,8] These distinct functions strongly depend on their structural diversity, various functional groups and stereoisomers in each monosaccharide, multiple glycosidic linkage types and branching patterns between contiguous sugar units, and the sequence of monosaccharides in the glycopolymers. These variations give rise to significant differences in physical and chemical properties that are responsible for their functional diversity.[9,10] In addition, the variations also impact the 3D conformational properties of these molecules, which has stimulated both experimental and theoretical studies in recent years.[11−17]

In oligo- and polysaccharides, individual monosaccharide units are usually connected together through O-glycosidic linkages that are formed between the hemiacetal or hemiketal group of one sugar and a hydroxyl group of the subsequent sugar. The relative orientation of two contiguous sugars can be characterized by the torsional angles along the linkage.[13,15,17−20] Because of the inherent flexibility of these linkages, most saccharide polymers do not have a single, well-defined structure in solution and available experimental structures are mainly for saccharides conjugated or interacting with other molecules, such as proteins. For conformational analysis of saccharide polymers, various spectroscopic methods, mainly NMR, have been adopted to generate distance or dihedral distribution information,[1] indicative of the range of conformations accessible to these molecules.

To obtain atomistic models of the 3D conformational properties of complex saccharides theoretical simulations based on carefully parametrized molecular mechanical (MM) force fields offer great potential.[21−27] The use of MM models in combination with molecular dynamics (MD) simulation allows for the exploration of conformational properties in aqueous solution.[28−30] For complex molecules such as oligosaccharides, multiple metastable states (i.e., local minima) are separated from each other by energy barriers of various heights yielding

Received: November 5, 2014





DOI: 10.1021/ct500993h
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

complex energy landscapes. However, standard MD simulations are inefficient in crossing the energy barriers making it difficult to sample all the statistically important minima, an issue referred to as the quasi-ergodic problem.[31,32]

To tackle the sampling problem in MD simulations in the context of complex saccharides, several enhanced sampling methods have been employed including metadynamics, umbrella sampling, and replica exchange.[33−36] Metadynamics has been primarily applied to mono- or disaccharides using either quantum mechanical (QM) or MM potentials.[37−40] The application of umbrella sampling and its variants is limited as only 1 or 2 reaction coordinates are generally scanned due to the rapid increase in computational cost with the number of reaction coordinates.[41−44] Replica exchange MD in temperature space (T-REMD) has been employed to study disaccharides or trisaccharides,[45] oligosaccharides with as many as 6 sugar units,[46] a disaccharide N-linked to a peptide through an Asn residue,[47] and some biantennary oligosaccharide molecules in solution.[19,20,48] While temperature dependent properties can be obtained from T-REMD simulations,[46] the number of replicas in T-REMD increases with respect to the square root of the total number of atoms in the system as required to guarantee sufficient acceptance ratio between neighboring replicas and thus it is infeasible for large systems in explicit solvent. An alternative is the Hamiltonian replica exchange (H-REX) approach using different biased potential functions for each replica.[49,50] Our previous studies show that H-REX simulations with 1D and 2D biasing potentials applied to the glycosidic linkage dihedrals can greatly enhance conformational sampling for disaccharides and for disaccharides in a glycopeptide.[15,17] In addition, H-REX has been adopted using a 2D Gaussian-like biasing potential for each linkage in di- and trisaccharides in solution.[51] In the studies in our laboratory the 2D biasing potentials were constructed as the negative of the potential of mean force (PMF) along two linkage dihedrals using the grid-based correction map (CMAP) method, which was originally developed for the protein backbone Φ/Ψ dihedrals in the CHARMM force field.[52,53] This H-REX scheme is promising for use in large systems since it minimizes the need for additional replicas with increasing system size by including only biasing potentials in the calculation of the Metropolis ratio. In the present work, we extend the use of bpCMAPs in H-REX to more complex saccharides developing an automated procedure for bpCMAP construction and for the determination of the number and spacing of the replicas. These approaches are then applied to a branched N-glycan that is present in the HIV envelope protein gp120 and a linear N-glycan selected as a second model system to test the developed methodology.

The remainder of the paper is organized as follows. The H-REX method with bpCMAPs is described followed by the derivation of the replica distribution in the biasing potential space. WHAM is extended to allow for data from all the H-REX replicas to be used in analysis. The application of H-REX and accompanying data analysis is illustrated using a branched N-glycan found on the HIV gp120 protein and a linear model N-glycan. Results indicate that the bpCMAPs can be efficiently constructed from gas-phase disaccharide simulations and used to effectively access the conformational space of complex oligo- and polysaccharides.

## THEORETICAL CONSIDERATIONS

**Hamiltonian−Replica Exchange (H-REX) with CMAP Biasing Potentials (bpCMAPs).** In H-REX simulations, the $m$-th simulation replica is conducted with a biased potential function $U_m(\mathbf{R})$,

$$U_m(\mathbf{R}) = U_0(\mathbf{R}) + \lambda_m V_b(\Omega(\mathbf{R})) \quad (1)$$

where $U_0(\mathbf{R})$ is the potential energy of the system at conformation $\mathbf{R}$ and $V_b(\Omega(\mathbf{R}))$ is the biasing potential contribution, in this study the bpCMAPs applied along the two torsion angles, $w$, for each glycosidic linkage with $\Omega(\mathbf{R}) = (w1(\mathbf{R}), w2(\mathbf{R}))$. $\lambda_m$ is a scaling factor with a zero value for the ground-state, unbiased replica and a positive value for the other, excited-state replicas. The bpCMAPs are constructed within the theoretical framework of CMAP, which inverts the potential of mean force (PMF) along the linkage dihedrals in the disaccharide models (Figure 1a). In the replicas with positive scaling factors, the energy barriers about the dihedrals are lowered by the addition of the biasing potentials, resulting in accelerated barrier transitions between different local minima. The exchange between neighboring replicas reinitializes the lower replica with a new conformation and thus enhanced sampling can be achieved in the ground-state replica. Exchange attempts between two neighboring replicas $i$ and $j$ ($j = i + 1$ or $j = i - 1$) are accepted according to the Metropolis criterion,

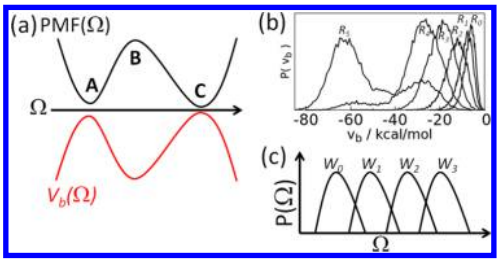



**Figure 1.** (a) 1D schematic representation of the PMF (black) and the corresponding biasing potential (red) along the collective variable Ω constructed as the negative of PMF. (b) Probability distribution of the biasing potential $V_b$ in different replicas ($R_0$−$R_5$) of the H-REX simulation for N-glycan1 molecule using the gas-phase bpCMAP. (c) Schematic plot of the probability distribution along the reaction coordinate Ω in different windows ($W_0$−$W_3$) of the umbrella sampling simulation.

$$\min[1, e^{-\beta(\lambda_i-\lambda_j)(V_b(\Omega(\mathbf{R}_j))-V_b(\Omega(\mathbf{R}_i)))}] \quad (2)$$

with $\beta$ the inverse temperature ($\beta = 1/k_bT$) at which all the replicas are simulated. Many studies have shown that a uniform acceptance ratio between different neighboring replicas can maximize the round trip rate between the ground-state replica and the highest excited replica.[54−58] To this end, the $\lambda$ value of each replica can be derived by equalizing the average acceptance ratio (AR) between two neighboring replicas $i$ and $j$[59]

$$\begin{aligned}\overline{AR_{ij}} &= \int \rho_i(\mathbf{R}_i)\rho_j(\mathbf{R}_j) \\ &\quad \times \min[1, e^{-\beta(\lambda_i-\lambda_j)(V_b(\Omega(\mathbf{R}_j))-V_b(\Omega(\mathbf{R}_i)))}]d\mathbf{R}_i d\mathbf{R}_j \\ &= \overline{AR_1} + \overline{AR_2} = \text{const}\end{aligned} \quad (3)$$

with $\overline{AR_1}$ and $\overline{AR_2}$ expressed as


DOI: 10.1021/ct500993h
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

$$\overline{AR_1} = \int_{-\infty}^{+\infty}\int_{-\infty}^{V_i}\rho_i(V_i)\rho_j(V_j)\mathrm{d}V_i\mathrm{d}V_j$$
$$\overline{AR_2} = \int_{-\infty}^{+\infty}\int_{-\infty}^{V_j}\rho_i(V_i)\rho_j(V_j) \times \exp(-\beta(\lambda_i-\lambda_j)(V_j-V_i)\mathrm{d}V_i\mathrm{d}V_j \quad (4)$$

where $\rho_i(\mathbf{R}_i)$ and $\rho_j(\mathbf{R}_j)$ are conformation distributions in replica $i$ and $j$, respectively; $\rho_i(V_i)$ and $\rho_j(V_j)$ are the biasing potential distributions in replicas $i$ and $j$, respectively. Assuming a $\delta$ function distribution of the biasing potentials, the $\lambda$ parameters can be determined as

$$\overline{AR_{ij}} = \overline{AR_{kl}} = \text{const} = \exp(-\beta(\lambda_i-\lambda_j)(a_j-a_i)) = \exp(-\beta(\lambda_k-\lambda_l)(a_l-a_k)) \quad (5)$$

where $a_i$ and $a_j$ are the average values of the biasing potentials in replicas $i$ and $j$, respectively, and $k$ and $l$ ($l = k + 1$ or $l = k - 1$) are another two neighboring replicas. A similar formula has been derived under the requirement of a uniform exchange probability and average energy approximation or the first-order condition for simulated tempering or temperature-REX simulations.[58,60,61] Here we show that this result can be rigorously obtained following the definition of the average acceptance ratio and the $\delta$ function assumption of the biasing potentials. If a Gaussian distribution is assumed under the central limit theorem when a large number of biasing potentials are used, the $\lambda$ parameters can be computed from

$$\begin{aligned}\overline{AR_{ij}} = \overline{AR_{kl}} = \text{const} &= 0.5\left(1+\text{erf}\left(\frac{a_i-a_j}{\sqrt{2(\sigma_i^2+\sigma_j^2)}}\right)\right)\\ &+ 0.5e^{\beta\Delta\lambda(a_i-a_j)+(\beta\Delta\lambda)^2(\sigma_i^2+\sigma_j^2)/2}\\ &\times\left(1+\text{erf}\left(\frac{a_j-a_i-\beta\Delta\lambda(\sigma_i^2+\sigma_j^2)}{\sqrt{2(\sigma_i^2+\sigma_j^2)}}\right)\right)\\ &= 0.5\left(1+\text{erf}\left(\frac{a_k-a_l}{\sqrt{2(\sigma_k^2+\sigma_l^2)}}\right)\right)\\ &+ 0.5e^{\beta\Delta\lambda(a_k-a_l)+(\beta\Delta\lambda)^2(\sigma_k^2+\sigma_l^2)/2}\\ &\times\left(1+\text{erf}\left(\frac{a_l-a_k-\beta\Delta\lambda(\sigma_k^2+\sigma_l^2)}{\sqrt{2(\sigma_k^2+\sigma_l^2)}}\right)\right)\end{aligned} \quad (6)$$

where $\sigma_i$ and $\sigma_j$ are the standard errors of the sampled biasing potentials in the two replicas and $\Delta\lambda = \lambda_i - \lambda_j$. The const in eq 5 and 6 is a constant factor used to regulate how much the overlap of probability distribution is maintained between neighboring replicas.

**Weighted Histogram Analysis Method (WHAM) to Combine H-REX Trajectories.** In H-REX simulations with bpCMAPs the biasing potentials are locally applied to the linkage dihedrals and can be accurately recomputed in posterior analysis. Traditionally, only the ground-state, unbiased replica is used for data analysis.[15,17,50,51,62,63] In this study, we propose to combine all the replicas with the weighted histogram analysis method (WHAM) to obtain an optimal distribution under the original potential of the system. Following the original WHAM[64] implementation, the probability distribution from H-REX can be expressed as the combination of the distribution from each replica[64,65]

$$\rho(\mathbf{R}) = \sum_{i=1}^{N}\frac{n_i\rho_i(\mathbf{R})}{\sum_{j=1}^{N}n_je^{-\beta(\lambda_jV_b(\Omega(\mathbf{R}))-f_j)}} \quad (7)$$

in which $N$ is the total number of replicas and $n_i$ the number of frames recorded in the $i$-th replica. The factors $\{f_j\}$ can be computed iteratively through

$$e^{-\beta f_k} = \sum_{i=1}^{N}\sum_{l=1}^{n_i}\frac{e^{-\beta\lambda_kV_b(\Omega(\mathbf{R}_{il}))}}{\sum_{j=1}^{N}n_je^{-\beta(\lambda_jV_b(\Omega(\mathbf{R}_{il}))-f_j)}} \quad (8)$$

From $\rho(\mathbf{R})$ the distribution along a set of collective variables $\Theta$ can be computed as

$$\rho(\Theta) = \int\rho(\mathbf{R})\delta(\Theta(\mathbf{R})-\Theta)\mathrm{d}\mathbf{R} = \sum_{i=1}^{N}\sum_{l=1}^{n_i}\frac{\delta(\Theta(\mathbf{R}_{il})-\Theta)}{\sum_{j=1}^{N}n_je^{-\beta(\lambda_jV_b(\Omega(\mathbf{R}_{il}))-f_j)}} \quad (9)$$

with $\delta(\Theta(\mathbf{R}_{il})-\Theta) = 1$ at $\Theta(\mathbf{R}_{il}) = \Theta$ and $\delta(\Theta(\mathbf{R}_{il})-\Theta) = 0$ at $\Theta(\mathbf{R}_{il})!= \Theta$.

And the free energy profile along $\Theta$ can thus be computed from $\rho(\Theta)$ as

$$G(\Theta) = -\frac{1}{\beta}\ln\rho(\Theta) \quad (10)$$

As indicated in eq 9 the frame with a more negative $V_b(\Omega(\mathbf{R}))$ value results in a smaller contribution to the distribution along $\Theta$ due to the larger denominator ($\sum_{j=1}^{N}n_je^{-\beta(\lambda_jV_b(\Omega(\mathbf{R}_{il}))-f_j)}$) in the summation. In H-REX simulations, the snapshots from the higher replicas have more negative biasing potentials than those from lower replicas (Figure 1a and b). Therefore, in comparison to the results from only the ground-state replica, the use of WHAM for H-REX mainly improves the sampling of the high-energy barriers regions that the lower replicas rarely sample. This is a different picture from WHAM for umbrella sampling in which each umbrella window contributes mainly around the center of the specific window along the reaction coordinate. In comparison to WHAM for umbrella sampling, the biasing potential in H-REX is the negative of the PMF and aims to reduce the energy barrier versus the umbrella potential that used to restrain sampling around a given window. The biasing potential can thus be applied to multiple variables instead of only 1 or 2 reaction coordinates typically used in umbrella sampling and all these variables in H-REX can be simultaneously sampled from a given replica. Additionally, the overlap between contiguous windows is required to guarantee the accuracy of computed PMF in umbrella sampling whereas in H-REX the overlap is not necessary in biasing potential space but is naturally maintained by the requirement of a sufficient acceptance ratio (Figure 1b-1c, and Figure S1 in Supporting Information). We note that while the use of WHAM with H-REX based on biasing potentials is applied to complex carbohydrates in the present study, the approach can be used with any system using such a H-REX approach.

## COMPUTATIONAL DETAILS

2D umbrella sampling simulations were used to generate the PMFs required to create the bpCMAPs. The bpCMAPs for 1


DOI: 10.1021/ct500993h
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

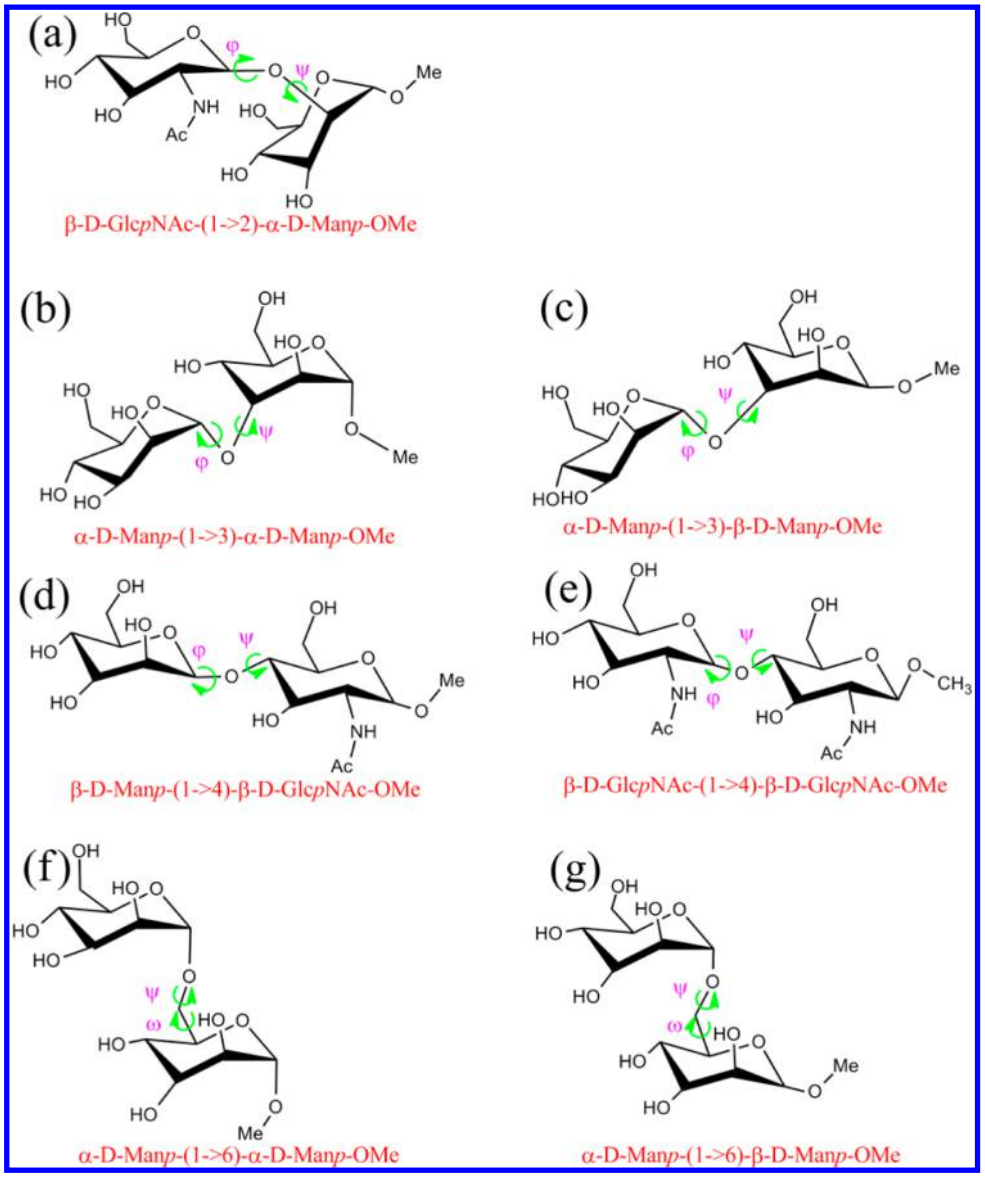



**Figure 2.** Disaccharides used to construct the bpCMAPs along the torsion angles $\varphi(O_5'-C_1'-O_n-C_n)/\psi(C_1'-O_n-C_n-C_{n+1})$ for 1 → 2, 1 → 3, and 1 → 4 linkages and $\psi(C_1'-O_6-C_6-C_5)/\omega(O_6-C_6-C_5-O_5)$ for 1 → 6 linkages.

→ *n* (*n* = 2, 3, 4, and 6) glycosidic linkages were constructed from different disaccharide models, with the hemiacetal group on the reducing end residue methylated (Figure 2). Umbrella sampling simulations were carried out along the torsional angles $\varphi(O_5'-C_1'-O_n-C_n)/\psi(C_1'-O_n-C_n-C_{n+1})$ for 1 → 2, 1 → 3, and 1 → 4 linkages and $\psi(C_1'-O_6-C_6-C_5)/\omega(O_6-C_6-C_5-O_5)$ for 1 → 6 linkages in both the gas phase and aqueous solution.[36] For each system, a total of 18 × 18 umbrella windows were evenly distributed at an interval of 20° along the two linkage dihedrals. A harmonic restraining potential was applied to all sampling windows with a uniform force constant of 45 kcal/mol/rad$^2$. Simulations of 4 ns was carried out for each window with a 2 fs integration step, with SHAKE used to constrain the distances of hydrogen-parent atom covalent bonds.[66] In the gas-phase simulations, a sufficiently long cutoff value was used to take all the nonbonded interactions into account and the temperature was maintained at 298 K with Langevin dynamics using the CHARMM package.[67] For umbrella sampling in solution, the disaccharide was immersed into a 36 Å × 36 Å × 36 Å TIP3P water box.[68] The system was first heated from 100 to 298 K under the constant volume-energy (NVE) ensemble (100 ps) and then equilibrated with 100 ps constant volume-temperature (NVT) and then 100 ps constant pressure−temperature (NPT) MD simulations at 298 K and 1 atm. The short-range nonbonded interactions were computed within a cutoff value of 12 Å. The Lennard-Jones (LJ) interactions were gradually switched off within the range 10−12 Å. The particle mesh Ewald (PME) method was employed to treat long-range electrostatic interactions with a charge grid of 1 Å.[69] All the solvated production simulations were carried out under the NPT ensemble using the GROMACS software.[70] The weak-coupling algorithm was adopted to maintain the system pressure at 1 atm and temperature at 298 K.[71] As the weak-coupling algorithm does not correspond to the correct ensemble,[72] the PMFs of two of the disaccharides were recalculated using the Nose-Hoover and Parrinello−Rahman methods[73−75] as the thermostat and barostat, respectively. Results show the PMFs from the two NPT methods to be similar (see Figure S2 in Supporting Information) and, therefore, adequate for the purpose of


DOI: 10.1021/ct500993h
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

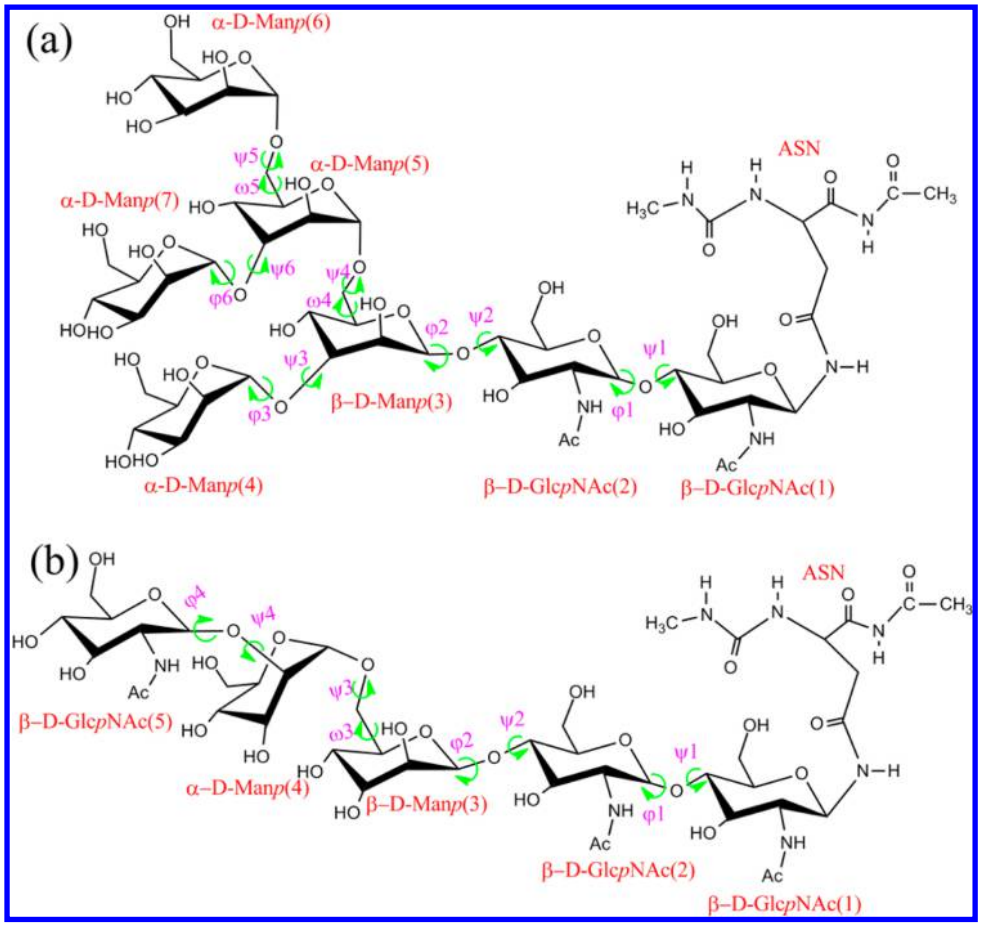



**Figure 3.** Structures of (a) N-glycan1 and (b) N-glycan2. The bpCMAPs along the marked linkage dihedrals are used in the H-REX simulations.

**Table 1. $\lambda$ Distribution and Acceptance Ratios (AR) in the H-REX Simulations of the N-Glycan Molecules**

| | N-glycan1 | | | N-glycan2 | |
|---|---|---|---|---|---|
| replica index | $\lambda_{gas}$[a] (AR[b] (%)) | $\lambda_{solv}$[c] (AR[b] (%)) | $\lambda_{solv2}$[d] (AR[b] (%)) | $\lambda_{gas}$[a] (AR[b] (%)) | $\lambda_{solv}$[b] (AR[b] (%)) |
| 0 | 0.00(61.0) | 0.00(69.1) | 0.00(53.2) | 0.00(57.4) | 0.00(66.3) |
| 1 | 0.17(52.0) | 0.16(64.2) | 0.22(49.2) | 0.20(47.6) | 0.19(62.0) |
| 2 | 0.34(39.7) | 0.31(51.2) | 0.43(44.6) | 0.40(40.4) | 0.38(54.6) |
| 3 | 0.50(37.2) | 0.48(39.9) | 0.58(49.3) | 0.53(50.5) | 0.53(57.0) |
| 4 | 0.61(32.3) | 0.60(23.8) | 0.64(39.9) | 0.61(46.6) | 0.62(51.5) |
| 5 | 0.67(26.6) | 0.67(11.0) | 0.67(25.2) | 0.67(35.0) | 0.67(40.6) |
| AAR[e] (%) | 41.5 | 43.2 | 43.6 | 46.2 | 55.3 |
| ART[f] (ns) | 3.1 | 2.9 | 1.9 | 2.0 | 0.7 |

[a] $\lambda_{gas}$: H-REX simulation with bpCMAPs constructed from gas-phase disaccharides. [b](AR): The acceptance ratio (in parentheses) for each replica in every H-REX simulation. [c] $\lambda_{solv}$: H-REX simulation with bpCMAPs constructed from solvated disaccharides. [d] $\lambda_{solv2}$: H-REX simulation with solvated bpCMAPs and the scaling factor $\lambda$ optimized from the 40 ns H-REX production run of N-glycan1 with $\lambda_{solv}$. [e]AAR: The average acceptance ratio over all 6 replicas in each H-REX simulation. [f]ART: The average round trip time for each H-REX simulation, which was the total simulation time for all 6 replicas divided by the total number of roundtrip in the 6 replicas.

constructing the bpCMAPs. From the umbrella sampling simulations, WHAM was used to compute the free energy profiles along the reaction coordinates.[65,76] The bpCMAPs were constructed as the negative of the free energies along the two linkage dihedrals.

To compare the sampling performance of bpCMAPs derived from gas phase and aqueous solution simulations, two oligosaccharides were employed for H-REX simulations. A branched N-glycan (N-glycan1), which corresponded to an N-glycan found on HIV gp120 protein (PDB ID 4NCO)[77] and a linear N-glycan (N-glycan2) (Figure 3) were selected as model systems. The monosaccharide units in these two glycans are linked via several different (1 → 2, 1 → 3, 1 → 4, and 1 → 6) glycosidic linkages. Two H-REX simulations were performed for each system, one with solvated bpCMAPs and the other with gas-phase bpCMAPs (Table 1). The size of the TIP3P water box used to solvate N-glycan1 was 52 Å × 52 Å × 52 Å and that for N-glycan2 was 40 Å × 40 Å × 40 Å. The short-range nonbonded interaction was computed within a cutoff of 12 Å, with a smoothing switch function over the range 10−12 Å for LJ interactions. The long-range electrostatic interactions were treated by PME with a charge grid of 1 Å and a sixth order


DOI: 10.1021/ct500993h
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

spline function for mesh interpolation. Each N-glycan system was heated from 100 to 298 K under the NVE ensemble (100 ps) and equilibrated with 100 ps NVT followed by a 100 ps NPT MD simulations at 298 K and 1 atm. Prior to the H-REX production run, a set of randomly assigned initial scaling factors, $\{\lambda_i\}$, were used for the trial H-REX simulations under the NPT ensemble for 2 ns. The average biasing potential, $\{a_i\}$, was computed for each replica from the trial simulations. Next the $a(\lambda)$ function was fitted from these $\{\lambda_i - a_i\}$ values using different basis functions, for example, polynomial or Gaussian. Finally, the fitted $a(\lambda)$ function was substituted into eq 5 at a given value of const to get the optimized $\lambda$ parameters, that, with $\lambda_0 = 0$ and const, $\lambda_1$ can be calculated; using the resulting $\lambda_1$ and const to compute $\lambda_2$, and so forth. The const value is adjusted to regulate the overlap between two neighboring replicas, for example, a larger const value represents more overlap of the biasing potential distribution and thus more replicas are needed within a given $\lambda$ range. In the simulations of N-glycan1 and N-glycan2, a total of 6 replicas were adopted for each simulation. The neighboring replicas were exchanged every 2 ps according to the acceptance ratio in eq 2. Each replica was simulated for 40 ns under the NPT ensemble. An additional 20 ns H-REX simulation was performed for N-glycan1 with solvated bpCMAPs and refitted $\lambda$ values from the 40 ns production run to examine the dependence of scaling factor on the average values of the biasing potentials (Table 1). All covalent bonds involving hydrogen atoms were constrained with the SHAKE algorithm and a time step of 2 fs was used.[66] The temperature was maintained at 298 K using the extended system algorithm with a thermal piston mass of 1000 kcal/mol·ps$^2$. A constant pressure of 1 atm was realized using the Langevin piston algorithm with a collision frequency of 20 ps$^{-1}$ and mass of 1630 amu.[78] All H-REX simulations were carried out with the REPDSTR module in the CHARMM package.[67,79] The CHARMM36 additive force field for proteins[80,81] and carbohydrates[25,27,82] was used for all the simulations in this study.

## RESULTS AND DISCUSSIONS

The application of bpCMAPs in H-REX overcomes energy barriers and enables efficient sampling of the local minima in the conformational space defined by the glycosidic linkage dihedrals. However, the bpCMAPs have to be constructed prior to H-REX simulation with the requirement of having a free energy landscape similar to that of the sugar polymer under study. One way to construct bpCMAPs is from high-temperature Langevin dynamics simulations of the target molecule in the gas phase with the assumption that the PMF profiles from these simulations are similar to those under ambient temperature and solvent environment. This scheme is efficient for disaccharides as shown in our previous studies,[15,17] and is allowed due to the biasing potentials not being required to be identical to the fully solvated surface but rather be sufficiently close to facilitate the sampling of the high energy regions. However, with increasing complexity of the system, the construction of bpCMAPs using the molecule itself becomes time-consuming due to inefficient sampling even with high-temperature gas-phase simulations. Alternatively, with oligo- and polysaccharides the bpCMAPs can be constructed from disaccharide models under the assumption that the free energy landscapes of the glycosidic linkages in disaccharides are adequately similar to that of the target saccharide polymer. On the basis of this, in the current study, we aim to develop a systematic procedure for H-REX simulations of complex oligo- and polysaccharides in solution, which includes the construction of bpCMAPs, determination of replica distributions in the biasing potential space, and application of the weighted histogram analysis method (WHAM) to improve the computed PMFs.[65,76]

**bpCMAP Construction from Disaccharide Molecules.** The conformational properties of oligo- and polysaccharides are largely governed by the glycosidic linkages that covalently connect the sugar units. However, the motion along these linkages is easily trapped in local minima in standard MD simulations because of the high-energy barriers between those minima. The application of bpCMAPs in H-REX can accelerate transitions over those barriers and enable efficient sampling of different conformational states. In principle, the bpCMAPs are the negative of the PMFs along these linkages thereby balancing the sampling around different metastable states. However, the PMF is the primary quantity we aim to compute from the H-REX simulations and, thus, cannot be obtained a prior. Alternatively, approximate bpCMAPs may be constructed to bias the sampling out of the local minima in the higher replicas. These new conformations are exchanged with the lower replicas during the simulation such that energy barriers are overcome in the ground-state replica. In this context, the conformations sampled in the ground-state replica should not be influenced by details of the constructed bpCMAPs though they should share similar free energy landscapes with the saccharide polymers under study. For simple oligosaccharides, for example, di- and trisaccharides, high-temperature Langevin dynamics can be performed in the gas phase to construct the bpCMAPs with the assumption that the free energy landscape from these simulations are similar to those in the presence of solvent and at ambient temperature. However, with increasing saccharide complexity, the construction of bpCMAPs using gas-phase simulations of the target molecule becomes time-consuming because of inefficient sampling. In addition, there are potential limitations in bpCMAP constructed from MD simulations alone. If sampling of the high-energy regions in the simulations is not adequate, the bpCMAP in that region will be constructed as a flat surface such that transitions in higher replicas will be inefficient since the biasing force around the transitions is zero. This is important for systems with several possible reaction paths of comparable barriers. As an alternative, the bpCMAPs may be constructed from umbrella sampling simulations using disaccharide models with the same linkages and then used in H-REX simulation of oligosaccharides and polysaccharides. This scheme is efficient and independent of system size of the saccharide molecule under study. However, this approach assumes that the free energy landscape of disaccharides is similar to the oligosaccharides along the same type of glycosidic linkage. To test this we first examined the free energy map of disaccharides in both solvated and gas-phase conditions.

The disaccharide models in Figure 2 were built and subjected to 2D umbrella sampling simulations in both the gas phase and aqueous solution. The linkages in these disaccharides include different 1 → 2, 1 → 3, 1 → 4, and 1 → 6 types that exist in many N-glycan molecules, including the model systems used in the present study. Comparison of the gas phase and aqueous solution PMFs is shown in Figure 4. As is evident in all cases the PMFs from the two environments are quite similar. In addition, the PMFs for the same linkages but different anomers at the reducing end are similar. For example the two 1 → 3


DOI: 10.1021/ct500993h
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

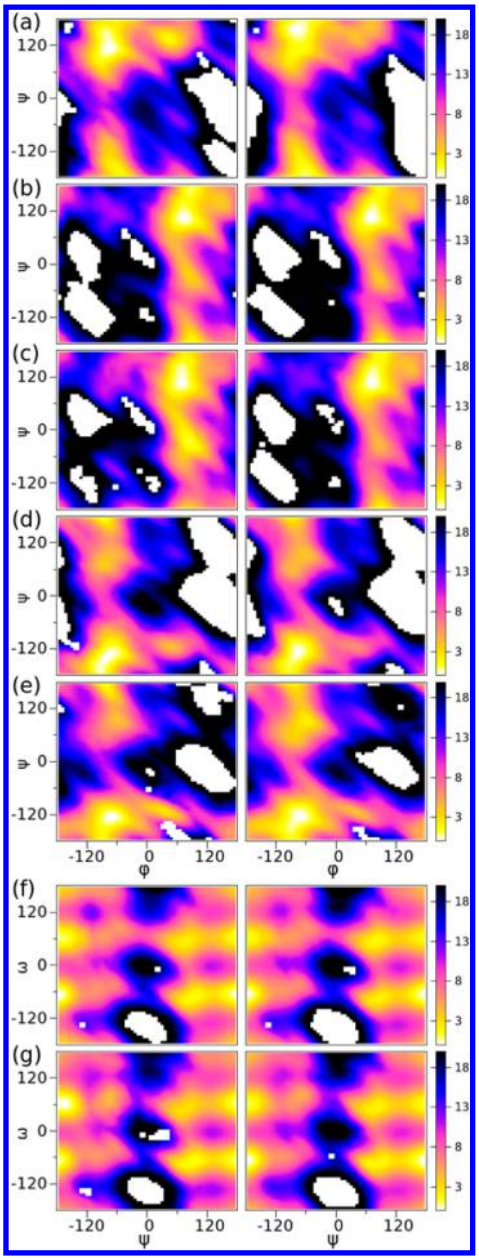



**Figure 4.** 2D $\varphi/\psi$ or $\psi/\omega$ (degrees) PMFs (kcal/mol) used for construction of bpCMAPs for each disaccharide. Gas-phase and solvated PMFs are shown in the left and right columns, respectively. (a) $\beta$-D-Glc*p*NAc-(1 → 2)-$\alpha$-D-Man*p*-OMe, (b) $\alpha$-D-Man*p*-(1 → 3)-$\alpha$-D-Man*p*-OMe, (c) $\alpha$-D-Man*p*-(1 → 3)-$\beta$-D-Man*p*-OMe, (d) $\beta$-D-Glc*p*NAc-(1 → 4)-$\beta$-D-Glc*p*NAc-OMe, (e) $\beta$-D-Man*p*-(1 → 4)-$\beta$-D-Glc*p*NAc-OMe, (f) $\alpha$-D-Man*p*-(1 → 6)-$\alpha$-D-Man*p*-OMe, and (g) $\alpha$-D-Man*p*-(1 → 6)-$\beta$-D-Man*p*-OMe.

linked disaccharides, $\alpha$-D-Man*p*-(1 → 3)-$\alpha$-D-Man*p*-OMe and $\alpha$-D-Man*p*-(1 → 3)-$\beta$-D-Man*p*-OMe in Figure 4b and c, respectively, are similar, though not identical. This is repeated in the 1 → 2 linkage (Figure 4a), the two 1 → 4 linked (Figure 4d, e) and 1 → 6 linked (Figure 4f, g) disaccharides. These results indicate that for H-REX simulation of solvated disaccharides, the bpCMAPs can be efficiently constructed from gas-phase umbrella sampling simulations. Furthermore, the bpCMAP are not sensitive to the nature of the reducing end glycosidic linkage, though given that these linkages are based on an O-methyl group, their similarity was anticipated.

To use these bpCMAPs from disaccharides in H-REX simulations of the oligo- and polysaccharides, two other points need to be examined: (1) the similarity of free energy landscapes between the disaccharides and saccharide polymers calculated with gas-phase versus solvated bpCMAPs and (2) the sampling efficiency using gas-phase and solvated bpCMAPs.

**H-REX Simulations of the N-Glycans Using bpCMAPs from Gas-Phase and Solvated Disaccharides.** The H-REX sampling performance with bpCMAPs constructed from gas-phase and solvated disaccharides was examined in one branched (N-glycan1) and one linear N-glycan (N-glycan2) (Figure 3). N-Glycans consist of an oligo- or polysaccharide linked to the side-chain nitrogen of Asn residues in proteins and are ubiquitous as posttranslational modification of proteins and function in a broad range of biological processes. The conformational properties of N-glycans play a key role in their biological functions and, thus, understanding these properties is motivating both experimental and computational studies. Accordingly, we selected two N-glycans linked to Asn dipeptides to evaluate the sampling performance of H-REX with bpCMAPs constructed from gas-phase and solvated disaccharides. Two 40 ns H-REX simulations were carried out for each N-glycan, one with solvated bpCMAPs and the other with gas-phase bpCMAPs (Table 1). In the remainder of the manuscript, we focus the results for N-glycan1, with analogous plots for N-glcyan2 presented in Supporting Information Figures S6−S9.

To evaluate the performance of scaling factors derived from a short trial run using eq 5, the acceptance ratio was computed for each replica from the H-REX production simulations on N-glycan1. The acceptance is larger than 20% for all replicas except the highest one in the N-glycan1 simulation with solvated bpCMAP (Table 1), although variability of the acceptance ratios between replicas is present. Accordingly, an additional 20 ns H-REX simulation was carried out for this system using the reoptimized $\lambda$ from the first 40 ns H-REX trajectories. The result shows an improved uniformity in the acceptance ratio distribution, suggesting the use of eq 5 depends on the accuracy of the estimated average biasing potential values. Accordingly, for more complex systems with more biasing potentials, we suggest a two-step scheme to optimize the scaling factors, in which the initial values are derived from eq 5 based on short H-REX simulation followed by additional adjustment of the scaling factors with refined average biasing potential energies following a more extended H-REX simulation. This approach avoids manual tuning of the $\lambda$ values, which may require multiple adjustment to achieve acceptable values, and the results from all the H-REX simulations may be combined for the ground-state replica alone or for the ground-state and high-energy replicas using the WHAM procedure (see below).

The sampling efficiency was examined using the average round-trip time between the ground-state replica and the highest excited-state replica (Table 1). One round trip can be completed within a simulation time of less than 3.1 ns, even for the simulation with solvated bpCMAP and $\lambda_{solv}$ that has an acceptance ratio of 11.0% in the highest replica. Importantly, the round trip time is less than 1/10 of the total simulation time and shows an efficient random walk in scaling factor space. These results indicate that multiple transits from the ground state to the highest excited state with both the gas-phase and aqueous-solution biasing potentials, suggesting that sufficient sampling is obtained with both types of bpCMAPs. This is


DOI: 10.1021/ct500993h
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

supported by analysis of the 1D PMFs from the first 20 ns and second 20 ns of the H-REX simulations, as shown in Figure S3 of the Supporting Information for selected linkages.

In the present work, we propose the use of WHAM to allow for the inclusion of conformational sampling from all the replicas to improve the PMF calculations. The utility of this method was first examined by comparing the PMFs computed from the ground-state replica alone and from all 6 replicas in the N-glycan1 simulations. Shown in Figure 5 are 2D free

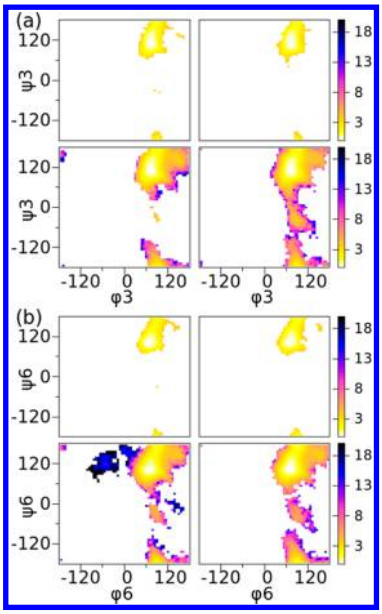



**Figure 5.** 2D Free energy surfaces for the two 1 → 3 linkages in N-glycan1 computed from the ground-state replica with gas-phase bpCMAPs (top left), ground-state replica with solvated bpCMAPs (top right), all 6 replicas with gas-phase bpCMAP (bottom left), and all 6 replicas with solvated bpCMAPs (bottom right) in H-REX simulations.

energy surfaces for the 1 → 3 linkages. The top panels show the sampling based on the ground-state replicas for both gas- and aqueous phase bpCMAPs. Both surfaces are similar, consistent with the ability to sample similar regions of conformational space, but significant sampling of the high-energy regions is lacking. However, when sampling from all the replicas is included using the WHAM approach presented above, additional sampling of the high-energy regions is evident (Figure 5, bottom panels). Notably, the shape of sampling in the low energy regions is similar for the ground-state only versus all-replica analyses, indicating that the WHAM procedure is effective in filtering out possible contamination from the high energy replicas. Similar results were obtained with the 1 → 4 and 1 → 6 linkages, as shown in Figures 6 and 7, respectively.

Additional analysis involved 1D PMF profiles in Figures 8−10 for the 1 → 3, 1 → 4, and 1 → 6 linkages, respectively. Again, significant increases in the sampling of high-energy regions are obtained with the all-replica analysis, while the low energy regions are similar for the ground-state only and all-replica analyses. In addition, the figures include PMFs from the solvated disaccharide umbrella sampling calculations, showing the sampling to generally be similar to that occurring in N-glycan1, as discussed in more detail below. Thus, the inclusion of conformational sampling from all replicas using WHAM

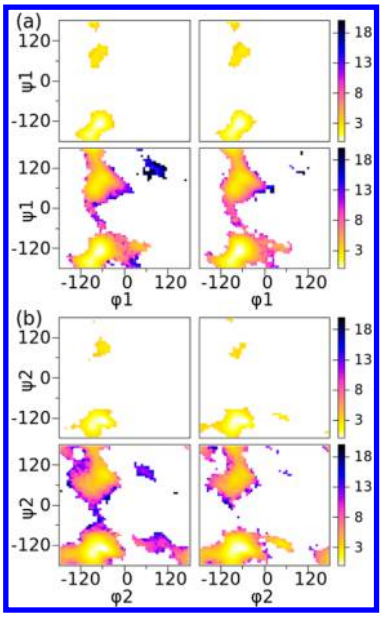



**Figure 6.** 2D free energy surfaces for the two 1 → 4 linkages in N-glycan1 computed from the ground-state replica with gas-phase bpCMAPs (top left), ground-state replica with solvated bpCMAPs (top right), all 6 replicas with gas-phase bpCMAP (bottom left), and all 6 replicas with solvated bpCMAPs (bottom right) in H-REX simulations.

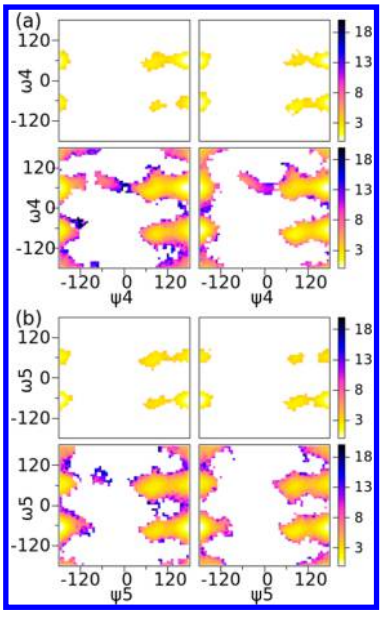



**Figure 7.** 2D free energy surfaces for the two 1 → 6 linkages in N-glycan1 computed from the ground-state replica with gas-phase bpCMAPs (top left), ground-state replica with solvated bpCMAPs (top right), all 6 replicas with gas-phase bpCMAP (bottom left), and all 6 replicas with solvated bpCMAPs (bottom right) in H-REX simulations.

leads to significant improvement in the sampling of high-energy regions.

The improvement in sampling of the high-energy regions of the PMFs using all replicas can be explained by the WHAM for H-REX formula in eq 9. As discussed above, the inclusion of the biasing energy in the denominator leads to the low free energy region being dominated by the lower replicas and the higher replicas mainly contributing to the high-energy regions


DOI: 10.1021/ct500993h
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

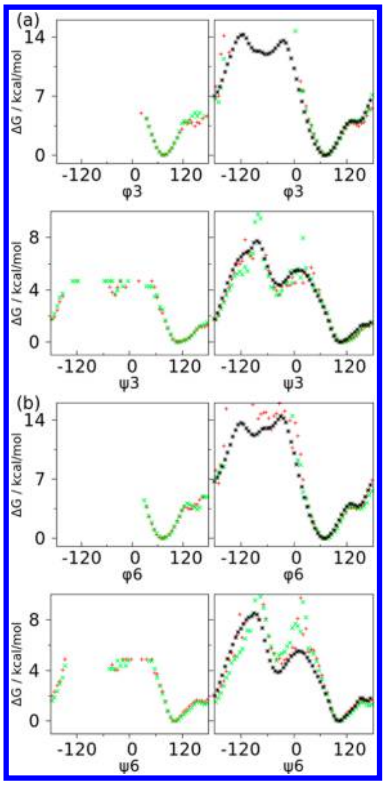



**Figure 8.** 1D free energy surfaces for the two 1 → 3 linkages in N-glycan1 computed from the ground-state replica (left) and all 6 replicas (right) for H-REX simulations using gas-phase (green) and solvated (red) bpCMAPs. The PMF profiles from corresponding solvated (black) disaccharides, *α*-D-Man*p*-(1 → 3)-*β*-D-Man*p*-OMe in (a) and Man*p*-(1 → 3)-*α*-D-Man*p*-OMe in (b), using umbrella sampling were also plotted on the right panels.

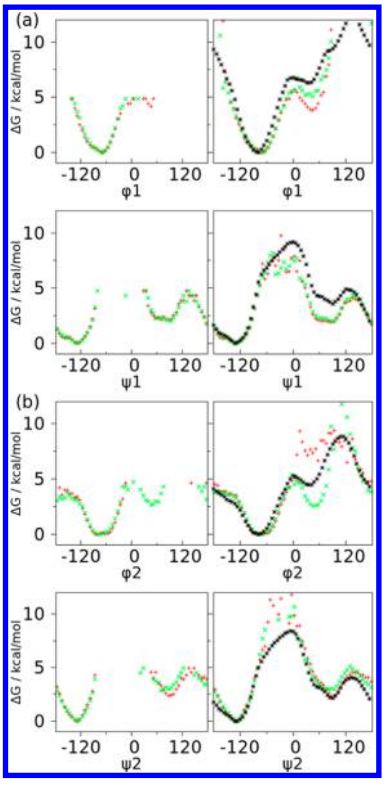



**Figure 9.** 1D free energy surfaces for the two 1 → 4 linkages in N-glycan1 computed from the ground-state replica (left) and all 6 replicas (right) for H-REX simulations using gas-phase (green) and solvated (red) bpCMAPs. The PMF profiles from corresponding solvated (black) disaccharides, *β*-D-Glc*p*NAc-(1 → 4)-*β*-D-Glc*p*NAc-OMe in (a) and *β*-D-Man*p*-(1 → 4)-*β*-D-Glc*p*NAc-OMe in (b), using umbrella sampling were also plotted on the right panels.

rarely sampled from the lower replicas. Clearly, the use of WHAM for H-REX can improve the PMF calculation, facilitating the identification of conformational states around the barrier regions. While all 6 replicas were used in the WHAM analysis in the present study, if larger numbers of replicas are used in more complex systems, the use of all replicas in WHAM analysis may not be necessary since the highest replicas will contribute even less to the probability distributions in accord with eq 9. In such cases, the first several replicas with less negative biasing potential values may be combined in WHAM for data analysis.

It is important to understand the relationship of the free energy landscapes of the linkages in the more complex N-glycans and in disaccharides from which the bpCMAPs were constructed. Notably, the similarity of the surfaces determines the efficiency of H-REX simulation. This is analyzed by inclusion of the solvated disaccharide PMFs from the umbrella sampling simulations in Figures 8−10. For all the linkages studied the free energy surfaces have the same locations of the minima and maxima regions for the disaccharides and N-glycan1. Moreover, the free energy differences in the barrier regions are also comparable. However, some differences in the surfaces are evident as anticipated given the presence of longer-range interactions in the N-glycan that do not occur in the disaccharides. In addition, PMFs calculated from the distances between the centers-of-mass of selected residues in N-glycan1 show that a wide range of conformations of the complex carbohydrate are being sampled and that the gas-phase and solution bpCMAPs yield similar surfaces for these degrees of freedom (see Figure S5 in Supporting Information).

Although the free N-glycan is rather flexible in solution environment, changes in sampling in the full N-glycan1 versus the disaccharides are evident in Figures 8−10. While such behavior is expected given the presence of long-range interactions in the complex N-glycan1, hydrogen bond analysis was undertaken to understand the molecular origins (Figure 11). Hydrogen bonding interactions are frequently formed between the noncontiguous sugars *α*-D-Man*p*(7), *α*-D-Man*p*(5), and *β*-D-Glc*p*NAc(2) (Figure 11a−c). As shown in Figure 11d, these long-range interactions stabilize two conformations with significantly different relative orientations of the flexible sugar branch (*α*-D-Man*p*(5), *α*-D-Man*p*(6), and *α*-D-Man*p*(7)) with respect to the rigid sugar stem (*β*-D-Glc*p*NAc(1), *β*-D-Glc*p*NAc(2), and *β*-D-Man*p*(3)).

In addition to the long-range nonbonded interaction in N-glycans, hydrogen bonding can occur between successive sugars that may impact the conformational properties with such interactions particularly sensitive to the presence or absence of solvent. Indeed this was a concern when considering the use of gas-phase bpCMAPs for H-REX. Accordingly, we analyzed two disaccharide models (*α*-D-Man*p*1 → 3 *β*-D-Man*p* and *β*-D-Man*p*1 → 4 *β*-D-Glc*p*NAc) that are present in N-glycan1 that can form hydrogen bonding interactions between adjacent units (see Figure S4 in the Supporting Information). The results do show significant difference in the hydrogen bonding distribution between the disaccharides in the gas phase and in solution. However, based on the similarity of the 2D free energy maps


DOI: 10.1021/ct500993h
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

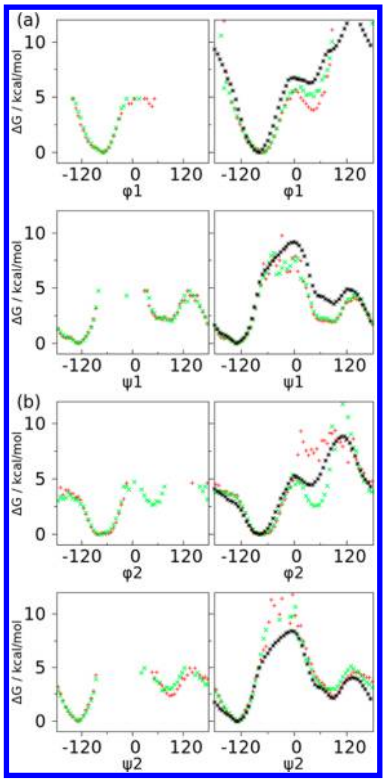



**Figure 10.** 1D free energy surfaces for the two 1 → 6 linkages in N-glycan1 computed from the ground-state replica (left) and all 6 replicas (right) for H-REX simulations using gas-phase (green) and solvated (red) bpCMAPs. The PMF profiles from corresponding solvated (black) disaccharides, α-D-Man*p*-(1 → 6)-β-D-Man*p*-OMe in (a) and α-D-Man*p*-(1 → 6)-α-D-Man*p*-OMe in (b), using umbrella sampling were also plotted on the right panels.

along the two linkages used to construct the bpCMAP (Figure 4) and, more importantly, from the resulting PMFs of the full N-glycan1 from the gas- and solvated phase bpCMAP H-REX simulations (Figure 8−10), it is evident that the differential hydrogen bonding in the gas- versus solution-phase disaccharides does not impact the sampling obtained on N-glycan1 in the solution H-REX simulations.

The above conclusions for bpCMAP construction and data analysis can also be drawn from the linear molecule N-glycan2. Multiple round trips were observed for H-REX simulations using bpCMAPs from gas-phase and solvated disaccharides (Table 1). The PMFs were improved when using WHAM to combine all replicas for data analysis (see Figure S6−S7 for 2D and Figure S8−S9 for 1D PMFs in the Supporting Information). Similar free energy profiles were produced from disaccharides umbrella sampling and solvated N-glycan2 H-REX simulations using both gas-phase and solvated bpCMAPs, further indicating the utility of bpCMAPs from gas-phase disaccharide simulations to improve sampling in more complex N-glycans in aqueous solution.

## CONCLUSION

H-REX simulations with biasing potentials are useful in studying the conformational properties of systems whose motions can be characterized by a set of predefined collective variables. In comparison to the standard REX simulations, this method minimizes the number of degrees of freedom included in the determination of the Metropolis acceptance ratio and thus can be applied to larger systems. Importantly, all the collective variables can be sampled simultaneously in a given replica, in contrast to other reaction coordinate dependent methods, for example, umbrella sampling, which require an exponential increase in computational resources with respect to

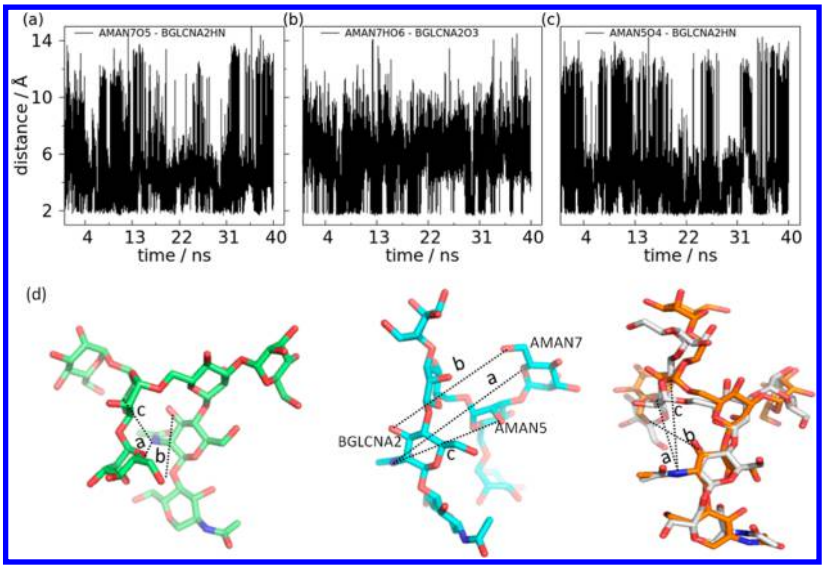



**Figure 11.** Long-range interactions between noncontiguous sugars in the ground-state replica of the H-REX simulation with gas-phase bpCMAPs for N-glycan1. (a−c) Hydrogen-bond distances during the simulations between (a) O5 atom of α-D-Man*p*(7) (AMAN7O5) and HN atom of β-D-Glc*p*NAc(2) (BGLCNA2HN), (b) HO6 atom of α-D-Man*p*(7) (AMAN7HO6) and O3 atom of β-D-Glc*p*NAc(2) (BGLCNA2O3), and (c) O4 atom of α-D-Man*p*(5) (AMAN5O4) and HN atom of β-D-Glc*p*NAc(2) (BGLCNA2HN). The 3 distances are marked on the structure in (d). (d) 3 representative conformations showing the relative orientation of the flexible sugar branch (α-D-Man*p*(5), α-D-Man*p*(6), and α-D-Man*p*(7)) to the rigid sugar stem (β-D-Glc*p*NAc(1), β-D-Glc*p*NAc(2), and β-D-Man*p*(3)). Long-range interactions are formed between the sugar branch and stem in the left and middle conformations. The conformation on the right represents one similar to that observed in the crystal structure of N-glycan1 (shown in gray) bound to an antibody targeting gpHIV120 (PDB ID 4NCO[77]). Oxygen atoms are shown in red and nitrogens in blue. Hydrogen atoms are omitted for clarity.


DOI: 10.1021/ct500993h
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

the number of collective variables predefined for the system. However, the biasing potentials have to be constructed prior to the H-REX simulations and those biasing potentials should share similar free energy landscapes with the system under study.

In the current study, the use of H-REX with bpCMAPs on the glycosidic linkage dihedrals is implemented to allow for conformational sampling of complex oligosaccharides. It is shown that bpCMAPs for each linkage can be efficiently constructed from umbrella sampling simulations of the corresponding disaccharides in the gas phase. The major features of the free energy landscapes in the disaccharides and the oligosaccharides are shown to be similar, as required for the use of the disaccharide bpCMAPs as the biasing potentials. In addition, we show that WHAM can be adopted to combine the data from all replicas to improve the determination of the high-energy regions on the free energy landscapes.

The ability to use a biasing potential derived from a PMF that does not correspond to the "exact" PMF to calculate the PMF itself may be considered in the context of free energy being a state function. The path (or method) by which the system is perturbed to sample conformational space is not important as long as the correct sampling of conformational space is obtained as required to calculate the PMF. The correct sampling of conformational space is inherent in the H-REX approach because of the lack of biasing in the ground-state replica. More generally, this may be thought of in the context of alchemical free energy perturbation calculations where unphysical intermediate states are used to calculate the free energy difference between two physical end states.[31]

On the basis of the present study, the following procedure is suggested for oligosaccharide simulations using H-REX with bpCMAPs: (1) Construction of bpCMAPs using disaccharides corresponding to those in the oligosaccharide calculated using umbrella sampling in the gas phase; these bpCMAPs can be deposited into a repository and used in different saccharide molecules. (2) Determination of the $\lambda$ values for each replica using short trial runs. (3) H-REX production runs with bpCMAPs used in multiple replicas at determined $\lambda$ values. (4) Calculation of free energy profiles using WHAM to combine the sampling from both the ground-state and excited-state replicas.

We anticipate that the presented H-REX with bpCMAP approach combined with WHAM analysis will be a powerful tool for studying the conformational properties of oligo- and polysaccharides. More generally, the WHAM analysis is anticipated to be of utility in combination with biasing-potential H-REX methods to facilitate the calculations of free energy surfaces in a range of complex systems in aqueous solution or other condensed phase environments.

## ■ ASSOCIATED CONTENT

### Supporting Information

Probability distribution of the biasing potential plotted in Figure S1, the PMF profiles using different thermostat and barostat for two disaccharide molecules in Figure S2, the PMFs of N-glycan1 in H-REX simulations in Figure S3, the probability of hydrogen bonding distance in two successive sugars in Figure S4, the PMF along distance between sugar centers in Figure S5, and the results from H-REX simulation of N-glycan2 listed in Figures S6–S9. This material is available free of charge via the Internet at http://pubs.acs.org.

## ■ AUTHOR INFORMATION

### Corresponding Author

*E-mail: alex@outerbanks.umaryland.edu.

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

Financial support from the NIH (GM070855 and GM051501) and computational support from the University of Maryland Computer-Aided Drug Design Center are acknowledged. Appreciation to Dr. Jing Huang for helpful discussions.

## ■ ABBREVIATIONS

1D, one-dimensional; 2D, two-dimensional; 3D, 3-dimensional; bpCMAP, biasing potential using 2D grid-based correction map; H-REX, Hamiltonian replica exchange; MD, molecular dynamics; MM, molecular mechanical; NPT, thermodynamic ensemble of constant particle number, pressure and temperature; NVE, thermodynamic ensemble of constant particle number, volume and energy; NVT, thermodynamic ensemble of constant particle number, volume and temperature; PME, particle mesh Ewald; PMF, potential of mean force; WHAM, weighted histogram analysis method

## ■ REFERENCES

(1) *Essentials of Glycobiology*, 2nd ed.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, New York, 2009.
(2) Alonso, D. M.; Wettstein, S. G.; Dumesic, J. A. *Chem. Soc. Rev.* **2012**, *41* (24), 8075–8098.
(3) Slaney, A. M.; Wright, V. A.; Meloncelli, P. J.; Harris, K. D.; West, L. J.; Lowary, T. L.; Buriak, J. M. *ACS Appl. Mater. Interfaces* **2011**, *3* (5), 1601–1612.
(4) Astronomo, R. D.; Burton, D. R. *Nat. Rev. Drug Discovery* **2010**, *9* (4), 308–324.
(5) Huang, Y.; Wun, C. *Expert Rev. Vaccines* **2010**, *9* (11), 1257–1274.
(6) Burton, D. R.; Poignard, P.; Stanfield, R. L.; Wilson, I. A. *Science* **2012**, *337* (6091), 183–186.
(7) Lang, R.; Winter, G.; Vogt, L.; Zuercher, A.; Dorigo, B.; Schimmele, B. *Drug Dev. Ind. Pharm.* **2009**, *35* (1), 83–97.
(8) Shukla, R. K.; Tiwari, A. *Crit. Rev. Ther. Drug* **2011**, *28* (3), 255–292.
(9) Dwek, R. A. *Biochem. Soc. Trans.* **1995**, *23* (1), 1–25.
(10) DeMarco, M. L.; Woods, R. J. *Glycobiology* **2008**, *18* (6), 426–440.
(11) Andre, S.; Kozar, T.; Schuberth, R.; Unverzagt, C.; Kojima, S.; Gabius, H.-J. *Biochemistry* **2007**, *46* (23), 6984–6995.
(12) Stanca-Kaposta, E. C.; Gamblin, D. P.; Cocinero, E. J.; Frey, J.; Kroemer, R. T.; Fairbanks, A. J.; Davis, B. G.; Simons, J. P. *J. Am. Chem. Soc.* **2008**, *130* (32), 10691–10696.
(13) Hatcher, E.; Sawen, E.; Widmalm, G.; MacKerell, A. D., Jr. *J. Phys. Chem. B* **2011**, *115* (3), 597–608.
(14) He, X.; Hatcher, E.; Eriksson, L.; Widmalm, G.; MacKerell, A. D., Jr. *J. Phys. Chem. B* **2013**, *117* (25), 7546–7553.
(15) Mallajosyula, S. S.; Adams, K. M.; Barchi, J. J.; MacKerell, A. D., Jr. *J. Chem. Inf. Model.* **2013**, *53* (5), 1127–1137.
(16) Krishnan, S.; Liu, F.; Abrol, R.; Hodges, J.; Goddard, W. A.; Prasadarao, N. V. *J. Biol. Chem.* **2014**, *289* (45), 30937–30949.
(17) Patel, D. S.; Pendrill, R.; Mallajosyula, S. S.; Widmalm, G.; MacKerell, A. D., Jr. *J. Phys. Chem. B* **2014**, *118* (11), 2851–2871.
(18) Mallajosyula, S. S.; MacKerell, A. D., Jr. *J. Phys. Chem. B* **2011**, *115* (38), 11215–11229.
(19) Re, S.; Miyashita, N.; Yamaguchi, Y.; Sugita, Y. *Biophys. J.* **2011**, *101* (10), L44–L46.
(20) Re, S.; Nishima, W.; Miyashita, N.; Sugita, Y. *Biophys. Rev.* **2012**, *4* (3), 179–187.


DOI: 10.1021/ct500993h
J. Chem. Theory Comput. XXXX, XXX, XXX–XXX

(21) Patel, D. S.; He, X.; MacKerell, A. D., Jr. *J. Phys. Chem. B* **2015**, *119*, 637−652.
(22) Hansen, H. S.; Huenenberger, P. H. *J. Comput. Chem.* **2011**, *32* (6), 998−1032.
(23) Raman, E. P.; Guvench, O.; MacKerell, A. D., Jr. *J. Phys. Chem. B* **2010**, *114* (40), 12981−12994.
(24) Hatcher, E.; Guvench, O.; MacKerell, A. D., Jr. *J. Phys. Chem. B* **2009**, *113* (37), 12466−12476.
(25) Guvench, O.; Hatcher, E.; Venable, R. M.; Pastor, R. W.; MacKerell, A. D., Jr. *J. Chem. Theory Comput.* **2009**, *5* (9), 2353−2370.
(26) Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; Gonzalez-Outeirino, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J. *J. Comput. Chem.* **2008**, *29* (4), 622−655.
(27) Guvench, O.; Greene, S. N.; Kamath, G.; Brady, J. W.; Venable, R. M.; Pastor, R. W.; Mackerell, A. D., Jr. *J. Comput. Chem.* **2008**, *29* (15), 2543−2564.
(28) Naidoo, K. J.; Denysyk, D.; Brady, J. W. *Protein Eng.* **1997**, *10* (11), 1249−1261.
(29) Salisburg, A. M.; Deline, A. L.; Lexa, K. W.; Shields, G. C.; Kirschner, K. N. *J. Comput. Chem.* **2009**, *30* (6), 910−921.
(30) Wood, N. T.; Fadda, E.; Davis, R.; Grant, O. C.; Martin, J. C.; Woods, R. J.; Travers, S. A. *Plos One* **2013**, *8* (11), No. e80301.
(31) In *Free Energy Calculations: Theory and Applications in Chemistry and Biology*; Chipot, C., Pohorille, A., Eds.; Springer: Berlin, 2007; pp 1−511.
(32) Frantz, D. D.; Freeman, D. L.; Doll, J. D. *J. Chem. Phys.* **1990**, *93* (4), 2769−2784.
(33) Barducci, A.; Bonomi, M.; Parrinello, M. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2011**, *1* (5), 826−843.
(34) Swendsen, R. H.; Wang, J. S. *Phys. Rev. Lett.* **1986**, *57* (21), 2607−2609.
(35) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314* (1−2), 141−151.
(36) Kaestner, J. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2011**, *1* (6), 932−942.
(37) Oborsky, P.; Tvaroska, I.; Kralova, B.; Spiwok, V. *J. Phys. Chem. B* **2013**, *117* (4), 1003−1009.
(38) Spiwok, V.; Kralova, B.; Tvaroska, I. *Carbohydr. Res.* **2010**, *345* (4), 530−537.
(39) Autieri, E.; Sega, M.; Pederiva, F.; Guella, G. *J. Chem. Phys.* **2010**, *133* (9), No. 095104.
(40) Biarnes, X.; Ardevol, A.; Planas, A.; Rovira, C.; Laio, A.; Parrinello, M. *J. Am. Chem. Soc.* **2007**, *129* (35), 10686−10693.
(41) Peric-Hassler, L.; Hansen, H. S.; Baron, R.; Huenenberger, P. H. *Carbohydr. Res.* **2010**, *345* (12), 1781−1801.
(42) Zeng, X.; Hu, H.; Zhou, H.; Marszalek, P. E.; Yang, W. *Biophys. J.* **2010**, *98* (4), 733−740.
(43) Islam, S. M.; Richards, M. R.; Taha, H. A.; Byrns, S. C.; Lowary, T. L.; Roy, P.-N. *J. Chem. Theory Comput.* **2011**, *7* (9), 2989−3000.
(44) Wang, D.; Amundadottir, M. L.; van Gunsteren, W. F.; Huenenberger, P. H. *Eur. Biophys. J. Biophys. Lett.* **2013**, *42* (7), 521−537.
(45) Yongye, A. B.; Gonzalez-Outeirino, J.; Glushka, J.; Schultheis, V.; Woods, R. J. *Biochemistry* **2008**, *47* (47), 12493−12514.
(46) Shen, T.; Langan, P.; French, A. D.; Johnson, G. P.; Gnanakaran, S. *J. Am. Chem. Soc.* **2009**, *131* (41), 14786−14794.
(47) Ellis, C. R.; Maiti, B.; Noid, W. G. *J. Am. Chem. Soc.* **2012**, *134* (19), 8184−8193.
(48) Nishima, W.; Miyashita, N.; Yamaguchi, Y.; Sugita, Y.; Re, S. *J. Phys. Chem. B* **2012**, *116* (29), 8504−8512.
(49) Fukunishi, H.; Watanabe, O.; Takada, S. *J. Chem. Phys.* **2002**, *116* (20), 9058−9067.
(50) Kannan, S.; Zacharias, M. *Proteins: Struct., Funct., Bioinf.* **2007**, *66* (3), 697−706.
(51) Mishra, S. K.; Kara, M.; Zacharias, M.; Koca, J. *Glycobiology* **2014**, *24* (1), 70−84.
(52) MacKerell, A. D., Jr.; Feig, M.; Brooks, C. L. *J. Am. Chem. Soc.* **2004**, *126* (3), 698−699.
(53) MacKerell, A. D., Jr.; Feig, M.; Brooks, C. L. *J. Comput. Chem.* **2004**, *25* (11), 1400−1415.
(54) Earl, D. J.; Deem, M. W. *J. Phys. Chem. B* **2004**, *108* (21), 6844−6849.
(55) Kone, A.; Kofke, D. A. *J. Chem. Phys.* **2005**, *122* (20), No. 206101.
(56) Rathore, N.; Chopra, M.; de Pablo, J. J. *J. Chem. Phys.* 2005, *122* (2), No. 024111.
(57) Denschlag, R.; Lingenheil, M.; Tavan, P. *Chem. Phys. Lett.* **2009**, *473* (1−3), 193−195.
(58) Prakash, M. K.; Barducci, A.; Parrinello, M. *J. Chem. Theory Comput.* **2011**, *7* (7), 2025−2027.
(59) Yang, L.; Shao, Q.; Gao, Y. Q. *J. Chem. Phys.* **2009**, *130* (12), 124111.
(60) Denschlag, R.; Lingenheil, M.; Tavan, P.; Mathias, G. *J. Chem. Theory Comput.* **2009**, *5* (10), 2847−2857.
(61) Sanbonmatsu, K. Y.; Garcia, A. E. *Proteins: Struct., Funct., Bioinf.* **2002**, *46* (2), 225−234.
(62) Kara, M.; Zacharias, M. *Biophys. J.* **2013**, *104* (5), 1089−1097.
(63) Kannan, S.; Zacharias, M. *Proteins: Struct., Funct., Bioinf.* **2010**, *78* (13), 2809−2819.
(64) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. *J. Comput. Chem.* **1992**, *13* (8), 1011−1021.
(65) Roux, B. *Comput. Phys. Commun.* **1995**, *91* (1−3), 275−282.
(66) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23* (3), 327−341.
(67) Brooks, B. R.; Brooks, C. L., III; Mackerell, A. D., Jr.; Nilsson, 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.; Karplus, M. *J. Comput. Chem.* **2009**, *30* (10), 1545−1614.
(68) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79* (2), 926−935.
(69) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98* (12), 10089−10092.
(70) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4* (3), 435−447.
(71) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81* (8), 3684−3690.
(72) Morishita, T. *J. Chem. Phys.* **2000**, *113* (8), 2976−2982.
(73) Hoover, W. G. *Phys. Rev. A* **1985**, *31* (3), 1695−1697.
(74) Nosé, S. *J. Chem. Phys.* **1984**, *81* (1), 511−519.
(75) Parrinello, M.; Rahman, A. *J. Appl. Phys.* **1981**, *52* (12), 7182−7190.
(76) Souaille, M.; Roux, B. *Comput. Phys. Commun.* **2001**, *135* (1), 40−57.
(77) Julien, J.-P.; Cupo, A.; Sok, D.; Stanfield, R. L.; Lyumkis, D.; Deller, M. C.; Klasse, P.-J.; Burton, D. R.; Sanders, R. W.; Moore, J. P.; Ward, A. B.; Wilson, I. A. *Science* **2013**, *342* (6165), 1477−1483.
(78) Feller, S. E.; Zhang, Y.; Pastor, R. W.; Brooks, B. R. *J. Chem. Phys.* **1995**, *103* (11), 4613−4621.
(79) Jiang, W.; Hodoscek, M.; Roux, B. *J. Chem. Theory Comput.* **2009**, *5* (10), 2583−2588.
(80) MacKerell, A. D., Jr.; 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* (18), 3586−3616.
(81) Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E. M.; Mittal, J.; Feig, M.; MacKerell, A. D., Jr. *J. Chem. Theory Comput.* **2012**, *8* (9), 3257−3273.
(82) Guvench, O.; Mallajosyula, S. S.; Raman, E. P.; Hatcher, E.; Vanommeslaeghe, K.; Foster, T. J.; Jamison, F. W.; MacKerell, A. D., Jr. *J. Chem. Theory Comput.* **2011**, *7* (10), 3162−3180.


DOI: 10.1021/ct500993h
*J. Chem. Theory Comput.* XXXX, XXX, XXX−XXX