Article

Comparative Assessment of Water Models in Protein−Glycan Interaction: Insights from Alchemical Free Energy Calculations and Molecular Dynamics Simulations

Deng Li and Mona S. Minkara*

Cite This: https://doi.org/10.1021/acs.jcim.4c01361

Read Online

ACCESS | Metrics & More | Article Recommendations | SI Supporting Information

ABSTRACT: Accurate computational simulations of protein−glycan dynamics are crucial for a comprehensive understanding of critical biological mechanisms, including host−pathogen interactions, immune system defenses, and intercellular communication. The accuracy of these simulations, including molecular dynamics (MD) simulation and alchemical free energy calculations, critically relies on the appropriate parameters, including the water model, because of the extensive hydrogen bonding with glycan hydroxyl groups. However, a systematic evaluation of water models' accuracy in simulating protein−glycan interaction at the molecular level is still lacking. In this study, we used full atomistic MD simulations and alchemical absolute binding free energy (ABFE) calculations to investigate the performance of five distinct water models in six protein−glycan complex systems. We evaluated water models' impact on structural dynamics and binding affinity through over 5.8 μs of simulation time per system. Our results reveal that most protein−glycan complexes are stable in the overall structural dynamics regardless of the water model used, while some show obvious fluctuations with specific water models. More importantly, we discover that the stability of the binding motif's conformation is dependent on the water model chosen when its residues form weak hydrogen bonds with the glycan. The water model also influences the conformational stability of the glycan in its bound state according to density functional theory (DFT) calculations. Using alchemical ABFE calculations, we find that the OPC water model exhibits exceptional consistency with experimental binding affinity data, whereas commonly used models such as TIP3P are less accurate. The findings demonstrate how different water models affect protein−glycan interactions and the accuracy of binding affinity calculations, which is crucial in developing therapeutic strategies targeting these interactions.

Water models: SPC/E, TIP3P, TIP3P-FB, OPC, TIP4P-Ewald → **Protein-Glycan complex**
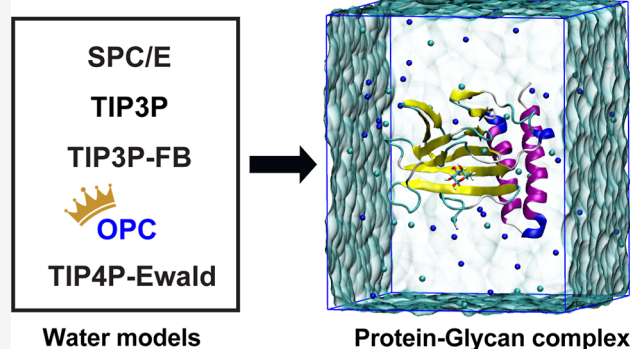
INTRODUCTION

Protein−glycan interactions play a crucial role in various biological processes, such as cell adhesion, immune responses, and pathogen recognition.[1−3] These interactions are integral to how cells communicate and respond to their external environment, particularly in the context of disease and defense mechanisms. For example, in infectious diseases, the way pathogens interact with host cells is often mediated by these complex protein−glycan interactions. Highlighting this, studies by Craig et al. and Smedley have shown that the pilin glycan on the *Pseudomonas aeruginosa*'s surface can enhance its invasion ability to the host cell by increasing *P. aeruginosa*'s adhesion to the host tissue.[4,5] Furthermore, C-type lectins, such as surfactant protein D (SP-D), surfactant protein A, and mannose-binding lectin, are the immunoproteins that specifically target pathogenic surface glycans related to the innate immune function in human.[6−10] They can recognize various pathogens by binding the glycans on their surface, including Influenza, SARS-CoV-2, HIV, and bacteria like *P. aeruginosa*.[11−15] These interactions are highly specific and play a critical role in molecular signaling

pathways. Despite the significance of protein−glycan interactions, the detailed mechanisms governing protein−glycan interactions remain only partially understood, primarily due to their complex and dynamic nature. This complexity underscores the significance of utilizing full atomistic modeling to provide a deeper insight into these interactions at the molecular level.
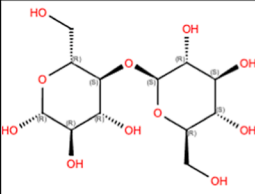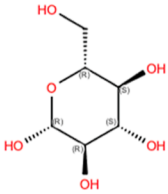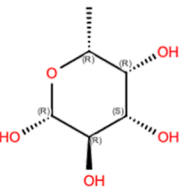
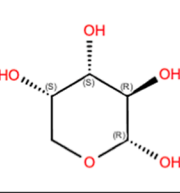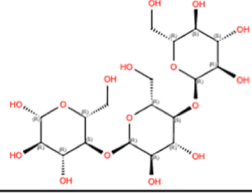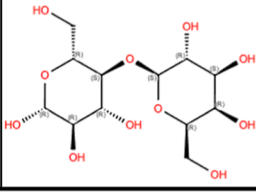Full atomistic modeling, specifically molecular dynamics (MD) simulations and alchemical binding free energy (ABFE) calculations, has proven pivotal in elucidating the dynamic behavior and glycan-binding specificity of protein−glycan complexes, providing comprehensive insights into glycan binding specificity, atomic-scale interactions, and conformational dynamics integral to pathogen recognition mecha-

A

Table 1. Protein−Glycan Complex Model Information with Corresponding Experimental ΔG

PDB ID	Protein	Glycan	2D Glycan Structure	Experimental ΔG (kcal/mol)
1I82	CBM9-2	DGlcpb1-4DGlcpb1-OH		-8.65
1I8A	CBM9-2	DGlcpb1-OH		-5.64
2AAC	AraC	DFucpb1-OH		-3.03
2ARC	AraC	LArapa1-OH		-4.09
2GGU	SP-D	DGlcpa1-4DGlcpa1-4DGlcpb1-OH		-4.13
2NN8	Gal-3 CRD	DGalpb1-4DGlcpb1-OH		-4.89

nisms.[16−23] The accuracy of these simulations is critically dependent on the appropriate choice of modeling parameters, especially the water models used to simulate the aqueous environment for the protein−glycan system.[24] The selection of an appropriate water model is crucial in simulating protein−glycan systems, particularly due to the extensive existence of hydroxyl groups in glycans. These hydroxyl groups engage in extensive hydrogen bonding, both within the glycan structure and with surrounding water molecules, significantly influencing the conformation and dynamics of the glycan.[25] Accurate representation of water is, therefore, essential for modeling these interactions realistically. Previous studies have demonstrated that different water models can yield varied results in terms of solvation properties and hydrogen bonding patterns.[26,27] Despite numerous studies employing various water models, a systematic evaluation of water models' accuracy in simulating

protein−glycan interaction at the molecular level is still lacking.[16,24,28−30]

Although one study assesses the effectiveness of various carbohydrate force fields in simulating protein−glycan interactions, various factors, including glycan force fields, water models, and MD parameters like cutoff distance for nonbonded interaction, protein force fields, and water models, were not well controlled.[24] The presence of multiple variables complicates the ability to draw definitive conclusions about the performance of any individual component within the simulation framework. Furthermore, the correlation between computational predictions, especially regarding binding free energies using different water models, and experimental data in protein−glycan complexes has not been investigated. Understanding the correlation can reveal how different water models impact binding free energy calculation accuracy, guiding the optimization of simulations for more reliable molecular predictions.

We aim to fill this gap by evaluating the accuracy of five water models (TIP3P, OPC, SPC/E, TTIP3P-FB, and TIP4P-Ewald) through extensive MD simulations of six protein−glycan complexes. Microsecond-scale, full-atomistic MD simulation trajectories for each model under different water models were analyzed, focusing on elucidating the dynamical behaviors and interaction differences in different solvation models. Additionally, we utilize alchemical absolute binding free energy (ABFE) calculations to evaluate the binding affinities between proteins and glycans in different solvation conditions and compare them with the experimental results to quantify each water model's accuracy. This study aims to establish a robust benchmark for selecting the most appropriate water model for accurately simulating protein−glycan interactions.

■ METHODS

Protein−Glycan Complex Models. In this study, we assessed the performance of five distinct water models (TIP3P, OPC, SPC/E, TIP3P-FB, and TIP4P-Ewald) in MD simulations and alchemical ABFE calculations. We selected six protein−glycan complexes from the protein data bank (PDB) for this evaluation. The complexes and their characteristics are as follows:

1. The C-terminal module of thermostable *Thermotoga maritima* xylanase 10A (CBM9-2), comprising residues 1−189 and cataloged under PDB ID: 1I82, specifically binds the disaccharide DGlcpb1−4DGlcpb1-OH, which consists of two D-glucoses linked by a 1−4 glycosidic bond.[31] The crystal structure's resolution is 1.90 Å, and its binding affinity is measured by isothermal titration calorimetry.[31,32] The CBM9-2 is a carbohydrate-binding module that facilitates the interaction between the enzyme and its substrate.[31]

2. Additionally, CBM9-2 interacts with the monosaccharide DGlcpb1-OH (PDB ID: 1I8A).[31] This crystal structure also has a resolution of 1.90 Å and uses the same method to measure the binding affinity as that in 1I82.[31,32]

3. The *Escherichia coli* (*E. coli*) gene regulatory protein AraC (PDB ID: 2AAC), containing residues 6−168, forms a complex with DFucpb1-OH, a D-fucose monosaccharide.[33] This crystal structure has a resolution of 1.6 Å, and its binding affinity was measured by tryptophan fluorescence quenching (TFQ).[33,34] AraC is a transcriptional regulator that controls the expression of genes involved in the metabolism of arabinose and other sugars in *E. coli*.[33]

4. Additionally, AraC, described in the PDB with PDB ID 2ARC for residues 7−167, binds with LArapa1-OH. LArapa1-OH is an L-arabinose monosaccharide.[35] This structure is resolved at 1.5 Å, and binding affinity was measured by TFQ.[34,35]

5. The human SP-D, with its structure in the Protein Data Bank (PDB ID: 2GGU) consisting of residues 205−355, is complexed with a glycan DGlcpa1−4DGlcpa1−4DGlcpb1-OH.[36] DGlcpa1−4DGlcpa1−4DGlcpb1-OH is a trisaccharide that contains three D-glucoses (Glc) connected through two 1−4 glycosidic linkages. This crystal structure has a resolution of 1.9 Å, and its binding affinity was measured by enzyme-linked immunosorbent assays (ELISAs).[36−38] SP-D acts as a collectin in the lung, recognizing and opsonizing pathogens for phagocytosis while modulating inflammation.[36]

6. The human galectin-3 (Gal-3) carbohydrate-recognition domain, with its structure in the protein data bank (PDB ID: 2NN8) containing residues 113−250, interacts with a DGalpb1−4DGlcpb1-OH glycan.[39] DGalpb1−4DGlcpb1-OH is a disaccharide with a D-galactose (Gal) and a D-glucose (Glc) through a 1−4 glycosidic linkage. The crystal structure has a resolution of 1.35 Å and its binding affinity was measured by frontal affinity chromatography.[38−40] Gal-3 is a multifunctional protein that binds to sugars on various molecules, influencing cell adhesion, proliferation, and immune response.[39]

Detailed descriptions of these six protein−glycan complexes are provided in Table 1, and the corresponding 3D complex structures are provided in Figure S1.

Structural Preparation, Solvation, and Neutralization. For the preparation of the molecular structures, we utilized the tleap module from AmberTools23 to add hydrogen atoms.[41] The protein structures were modeled using the Amber ff19SB force field, while the glycan structures were represented using the latest GLYCAM06j force field in Amber.[42,43] To evaluate the performance of different water models in simulating protein−glycan interactions, we used five distinct water models: TIP3P, OPC, SPC/E, TIP3P-FB, and TIP4P-Ewald.[27,44−47] We did not consider five points or more complex water models as they are usually used in more complex liquid environments such as ice and liquid-ice coexistence conditions.[48−51] Each protein−glycan complex was solvated in a rectangular water box, ensuring a minimum distance of 10 Å from any box edge to the nearest atom of the complex. This was achieved by creating a water shell around the complex. The system's neutrality was maintained by appropriately placing sodium (Na^+) and chloride (Cl^-) ions. Furthermore, the ionic strength of the solution was adjusted to mirror physiological conditions by setting the salt (NaCl) concentration to 0.15 M.

Molecular Dynamics Simulations. All simulations commenced with a two-stage minimization process. Initially, the entire protein−glycan complex structure was fixed by applying harmonic restraints of 50 kcal/(mol·Å^2) on all atoms. This was followed by an unrestrained minimization of the entire system to remove any residual structural strains. After minimization, the system underwent a two-stage equilibration. In the first stage, the system's temperature was incrementally increased from 0 to 298 K over 2 ns of MD with a time step of 2 fs in the *NVT* ensemble. During the heating stage, a harmonic restraint of 50 kcal/(mol·Å^2) was maintained on the entire complex structure. The final equilibration stage involved maintaining the system at a constant temperature (298 K) and pressure (1 bar) in the *NPT* ensemble for a duration of 50 ns. Temperature control was achieved using the Langevin dynamics algorithm with a collision frequency of 1.0 ps^{-1}, while pressure was regulated using the Monte Carlo barostat.[52−54] Throughout the equilibration phase, all hydrogen-containing bonds were constrained using the SHAKE algorithm.[55] Nonbonded interactions were calculated using the particle mesh ewald method with a cutoff distance of 10 Å. All MD simulations were performed using GPU-accelerated implementation of Amber22.[56] The production MD run was conducted for a duration of 1000 ns in two replicates, employing the isobaric−isothermal (*NPT*) ensemble settings as detailed in the equilibration section.

Alchemical Absolute Binding Free Energy (ABFE) Calculation. In this study, we applied alchemical ABFE

calculations using free energy perturbation to assess the accuracy of binding affinity predictions across various water models. We adapted a protocol originally proposed by Heinzelmann and Gilson, modifying the BAT.py v2.2 source code specifically for ABFE calculation in protein−glycan systems with specific force fields.[57] We evaluated how accurate ABFE calculations within the protein−glycan complex are impacted by the five different water models.

During the ABFE calculation, proper ligand restraints are crucial for accurate binding affinity prediction in these calculations. To prevent ligand displacement from the receptor binding pocket, harmonic translational and rotational (TR) restraints were applied to both protein and ligand atoms during the perturbation process. These TR restraints comprise one distance restraint, two angle restraints, and three dihedral restraints.[57−59] Additionally, conformational restraints were imposed on both the receptor and ligand, following the methodology described by Heinzelmann and Gilson.[57]

Prior to the free energy calculation simulations, the systems were equilibrated using similar procedures to those outlined in the MD Simulations section. The equilibrated complex structures were then utilized in simulations for ABFE calculations. This involved 12 λ windows (0.00922, 0.04794, 0.11505, 0.20634, 0.31608, 0.43738, 0.56262, 0.68392, 0.79366, 0.88495, 0.95206, 0.99078) based on a Gaussian quadrature distribution for decoupling/recoupling ligand Lennard-Jones (ΔG_{LJ}) and charge interactions (ΔG_{elec}).[57,60] The contributions of various restraints to the binding free energy, including the attachment of conformational restraints to protein ($\Delta G_{p,att}$), ligand ($\Delta G_{l,att}$), and the TR restraints ($\Delta G_{l,TR,att}$), and the release of conformational restraints for the protein ($\Delta G_{p,rel}$), ligand ($\Delta G_{l,rel}$), and the TR restraints ($\Delta G_{l,TR,rel}$), were computed across 16 windows based on the protocol by Heinzelmann and Gilson.[57] The ABFE is thus represented as below

$$-\Delta G_{bind}^0 = \Delta G_{elec} + \Delta G_{LJ} + \Delta G_{p,att} + \Delta G_{l,att}$$
$$+ \Delta G_{l,TR,att} + \Delta G_{l,rel} + \Delta G_{p,rel} + \Delta G_{l,TR,rel}$$

For each perturbation window, the system was heated from 50 to 298 K over 50 ps and subsequently equilibrated for 70 ps. Data for the binding free energy calculation was gathered over a 10 ns production run. In total, 1280 ns of simulation time was collected for the calculation in each system. The free energy changes were estimated using the multistate Bennett-acceptance ratio.[61] All alchemical ABFE calculations were carried out using GPU-accelerated implementation of Amber22.[56] The ABFE values are corrected to the standard state.[57] To ensure the robustness of the results, three replicates were conducted for each system to determine the final binding free energy.

Statistical metrics, including mean unsigned error (MUE) and root-mean-square-error (RMSE), are calculated to measure the difference between simulated ABFE values and experimental binding affinities and to further evaluate the performance of ABFE in different water models.

Structural Analysis Methods. Root mean square deviation (RMSD) was calculated based on the coordinates of the protein−glycan's non-hydrogen atoms for each model across the entire simulation trajectory, using the pre-equilibrated, experimentally solved crystal complex structure as the reference. The rmsd's coefficient of variation (CV) is calculated using the equation below

$$CV = \frac{\sigma}{\mu}$$

where σ is the standard deviation of RMSD and μ is the mean value of RMSD. Additionally, root mean square fluctuation (RMSF) was determined based on the coordinates of the protein−glycan's non-hydrogen atoms. Hydrogen bond analysis was conducted using the CPPTRAJ program, using a default angle cutoff of 135° and a distance cutoff of 3 Å.[62] Regarding the glycosidic torsion angles, we specifically focused on the Phi and Psi angles, defined respectively as O5−C1−O4−C4 and C1−O4−C4−C5. These torsion angles were also computed using the CPPTRAJ program.[62] Visual molecular dynamics and Schrodinger-Maestro (2023−2) were used to visualize the snapshots of the simulation.[63,64] All analyses are the average over the two replicates of the microsecond MD simulations.

Density Functional Theory Calculations. We analyzed the conformations of the most populated glycosidic torsion angles in the glycan components of systems 1I82, 2GGU, and 2NN8. These conformations were then subjected to single-point energy calculations to quantify their stability in the bound state. Tables 2, 3, and S7 detail the most populated glycosidic torsion

Table 2. Single Point Energy of DGlcpb1-4DGlcpb1-OH at Bound State in 1I82 Complex at the Phi Mode and Psi Mode Conformation in Different Water Models

Water model	Phi mode (deg)	Psi mode (deg)	Single point energy (hartrees)	Relative energy (kcal/mol)
TIP3P	−82.34	−155.14	−1298.375078	0.000
OPC	−82.70	−148.29	−1298.379260	−2.624
SPC/E	−84.14	−151.89	−1298.378725	−2.289
TIP3P-FB	−82.70	−151.89	−1298.377308	−1.399
TIP4P-Ewald	−82.34	−150.45	−1298.377798	−1.707

angles for these systems, as observed across different water models. For the single-point energy calculations, we used the B3LYP-D3/6-311++G** level of theory, which considers additional dispersion correction terms. All density functional theory (DFT) calculations were performed using the Jaguar.[65]

■ RESULTS AND DISCUSSION

Structural Dynamics of the Systems. To assess the structural fluctuation of the six distinct protein−glycan complexes, as shown in Table 1, we conducted RMSD analysis and the corresponding CV under five different water models (Figure 1). Among these, the complexes 1I82, 1I8A, and 2NN8 are stable across all water models, as shown in Figure 1a,b,f. The stability is evidenced by minimal fluctuations in their RMSD curves and small CVs in different water models.

In contrast, certain complexes exhibited more significant fluctuations under specific water models. For example, the 2AAC complex showed larger variability when simulated with the TIP4P-Ewald water model, yielding an average RMSD and corresponding CV of 2.10 Å and 9.39%, as shown in Figure 1c. For the rest of the water models, the average RMSDs are 1.94 Å for TIP3P, 2.19 Å for OPC, 1.89 Å for SPC/E, and 1.89 Å for TIP3P-FB. Their corresponding CVs are 6.18%, 6.49%, 5.74%, and 5.68%. Similarly, the 2ARC complex experienced greater fluctuations under the TIP3P water model, with an average RMSD of 3.13 Å and CV of 17.85%, as shown in Figure 1d. For the remaining water models, the average RMSDs are 2.12, 2.03, 1.98, and 2.00 Å for OPC, SPC/E, TIP3P-FB, and TIP4P-

Table 3. Single Point Energy of DGlcpa1-4DGlcpa1-4DGlcpb1-OH at Bound State in 2GGU Complex at the Phi Mode and Psi Mode Conformation in Different Water Models

Water model	Group 1		Group 2		Single point energy (hartrees)	Relative energy (kcal/mol)
	Phi mode (deg)	Psi mode (deg)	Phi mode (deg)	Psi mode (deg)		
TIP3P	95.68	−133.51	90.63	−142.16	−1909.253062	0.000
OPC	92.79	−133.15	91.71	−142.16	−1909.254107	−0.656
SPC/E	100.36	−132.07	94.23	−139.28	−1909.253970	−0.570
TIP3P-FB	99.64	−130.63	94.59	−138.92	−1909.253550	−0.306
TIP4P-Ewald	96.04	−132.07	95.68	−138.92	−1909.260350	−4.573

Ewald, respectively, and their corresponding CVs are 6.20%, 8.12%, 6.68%, and 8.74%. The significant RMSD fluctuation observed with TIP3P is attributed to glycan unbinding in one of the replicate MD simulations for the 2ARC complex. The 2GGU complex under the TIP3P water model achieved large fluctuation, with an average RMSD of 2.56 Å and CV of 18.28%, as shown in Figure 1e. In the other water models, the average RMSDs are 2.05, 2.22, 2.21, and 2.05 Å for OPC, SPC/E, TIP3P-FB, and TIP4P-Ewald. Their corresponding CVs are 11.72%, 16.45%, 12.10%, and 12.94%. These results suggest that the choice of water model can impact the dynamic behavior of protein−glycan complexes.

Root-mean-square fluctuation (RMSF) analysis was performed to evaluate each residue's fluctuation on six protein−glycan complex models across five water models (Figure S2). Complexes 1I82, 1I8A, 2AAC, and 2NN8 demonstrated consistent RMSF curves in all water models, indicating stable residue fluctuations. In contrast, the 2ARC complex showed significant fluctuations in the terminal residues, specifically in the TIP3P water model. The residues from the terminal region (residues 7 to 18) are key components of the binding pockets, suggesting their fluctuations could influence glycan binding. Notably, when the glycan unbound from the binding pocket in the TIP3P water model, these residues experienced increased fluctuation. However, the same protein, AraC, complexes with a different monosaccharide in 2AAC, D-fucose, and exhibits stable fluctuation in the terminal motif. We further investigated the interaction between the identical protein, AraC, and two monosaccharides in detail through hydrogen bond analysis.

Protein−Glycan Hydrogen Bond Information. To better understand how the water models influence the interaction between the protein and the glycan, we conducted a comprehensive comparison of hydrogen bond interactions within each of the six protein−glycan complexes across five water models. Details of hydrogen bonds between protein and glycan in various water models are provided in Tables S1−S6.

In the complexes 2AAC and 2ARC, the identical protein, AraC, interacts with two distinct monosaccharides, D-fucose and L-arabinose, respectively. We found that the hydrogen bonds involving THR24, ARG38, TYR82, and HIS93 are consistent in both monosaccharides. Notably, significant variations in the hydrogen bonding interactions between PRO8 and the two monosaccharides were observed. Specifically, the average PRO8-Fucose hydrogen bond fraction in the 2AAC complex is 0.792 ± 0.055 across all water models, exhibiting minimal variation. In contrast, within the 2ARC complex, the PRO8-Arabinose average hydrogen bond fraction significantly decreases to 0.105 ± 0.146 in all water models. PRO8 is situated in the terminal region of the protein and involved in its binding pocket, which explains the high RMSF region (residues 7 to 18) in the 2ARC complex model, as shown in Figure 2d.

In contrast, the remaining complexes (1I82, 1I8A, 2GGU, and 2NN8) exhibited similar hydrogen bond details across different water models. The 1I82 complex maintained an average hydrogen bond number of 3.50 ± 0.075, with consistent bond types. The 1I8A complex showed a similar pattern with an average of 3.158 ± 0.156 hydrogen bonds. The 2GGU complex's average was around 2.948 ± 0.247, and the 2NN8 complex presented an average of 4.79 ± 0.0430 hydrogen bonds under different water models.

Binding Motif Stability. As indicated by hydrogen bonding information and RMSF curves in the previous sections, significant fluctuations were observed in the binding motif (residues 7 to 18) of the 2ARC complex models across different water models. We calculated the RMSD of the motif region in each water model to further understand how water models influence the binding motif's conformation.

In the 2ARC complex, the binding motif experienced different structural stability depending on the water model. The OPC, SPC/E, TIP3P-FB, and TIP4P-Ewald models exhibited high structural stability in the region, as evidenced by the average RMSD values: 2.94 ± 0.59, 2.67 ± 0.49, 2.46 ± 0.44, and 2.83 ± 0.53 Å respectively, shown in Figure 2a. Conversely, the TIP3P models presented obvious structural fluctuations within the binding motif, with average RMSD values recorded at 4.75 ± 1.05 Å. This observation of increased structural variability is further corroborated by the spatial dispersion of motif superpositions, as illustrated in Figure 2b. These findings collectively underscore the pivotal role of water model selection in modulating the conformational stability of terminal residues, particularly within the vicinity of binding pockets, thereby influencing the overall MD and interaction patterns within the complex.

By contrast, for the 2AAC complex, the RMSD analysis revealed no significant fluctuation within the binding motif when simulated in the five water models, yielding an average RMSD of 1.98 ± 0.28, 2.05 ± 0.27, 1.76 ± 0.30, 1.76 ± 0.27, and 1.90 ± 0.28 Å, respectively, as shown in Figure 3a. The spatial superpositions of the binding motif maintain a state of relative stability, with minimal fluctuations observed in the five water models, as illustrated in Figure 3.

From the comparative analysis of these two highly similar complexes, it seems that the sensitivity to the choice of water model may depend on the strength of protein−glycan hydrogen bonds. The 2ARC complex, where PRO8 of the binding motif forms a weaker hydrogen bond than the 2AAC complex, exhibited water model-dependent dynamics. We hence speculate that the water model has more impact on protein−glycan interactions when the binding motif forms fewer hydrogen bonds with the glycan.

Glycan Conformation Stability in the Bound State. In addition to proteins' conformation analysis, we also performed the DFT calculations to evaluate the conformational stability of
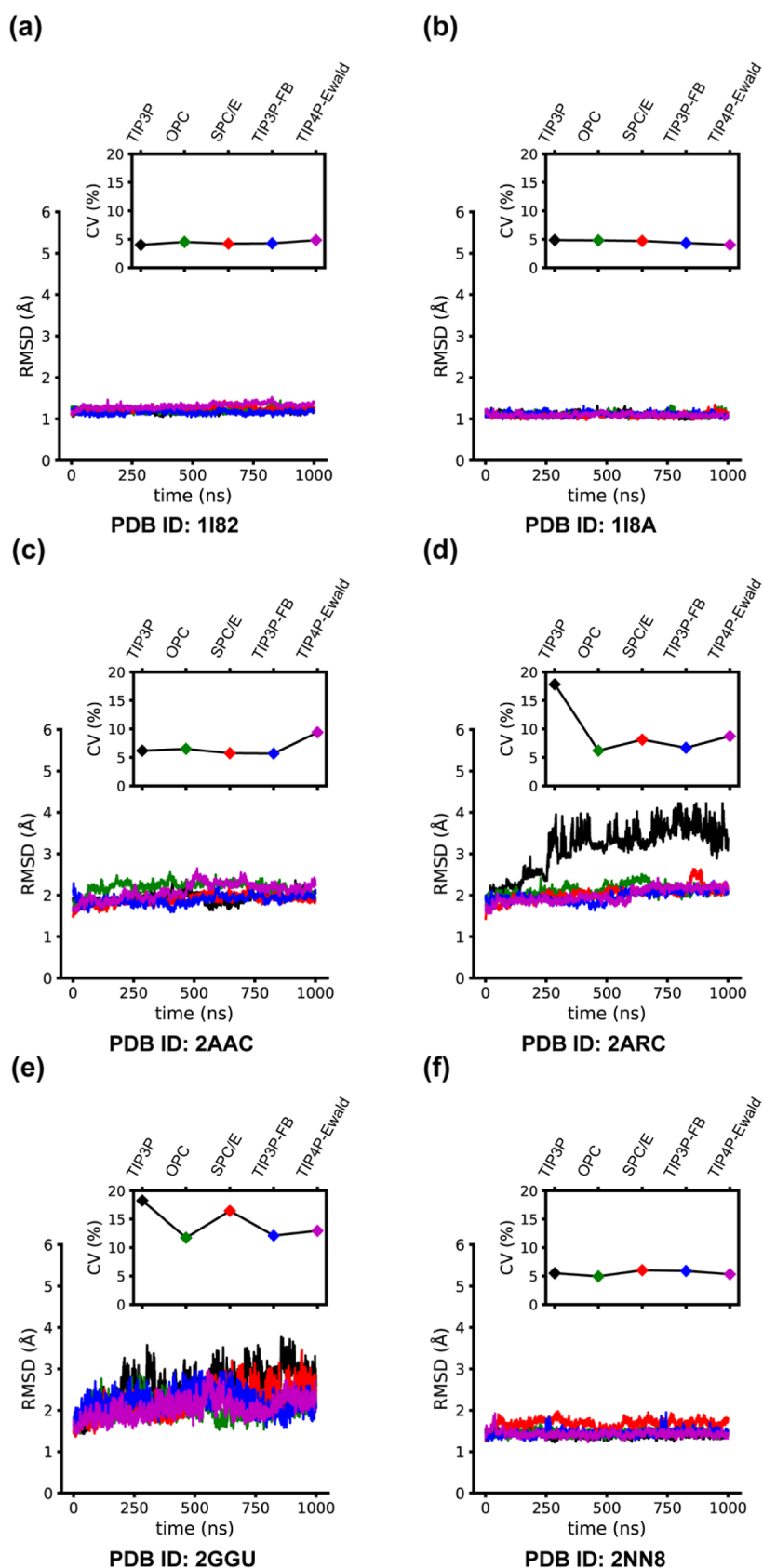
Figure 1. RMSD and corresponding CV of (a) 1I82, (b) 1I8A, (c) 2AAC, (d) 2ARC, (e) 2GGU, and (f) 2NN8 with different water models (black for TIP3P, green for OPC, red for SPC/E, blue for TIP3P-FB, and magenta for TIP4P-Ewald).

glycans in the bound state. We previously established a quantified correlation between the bound-state conformation

of glycans and the conformational stability.[23] Here, we investigate glycan conformation stability in complex models,
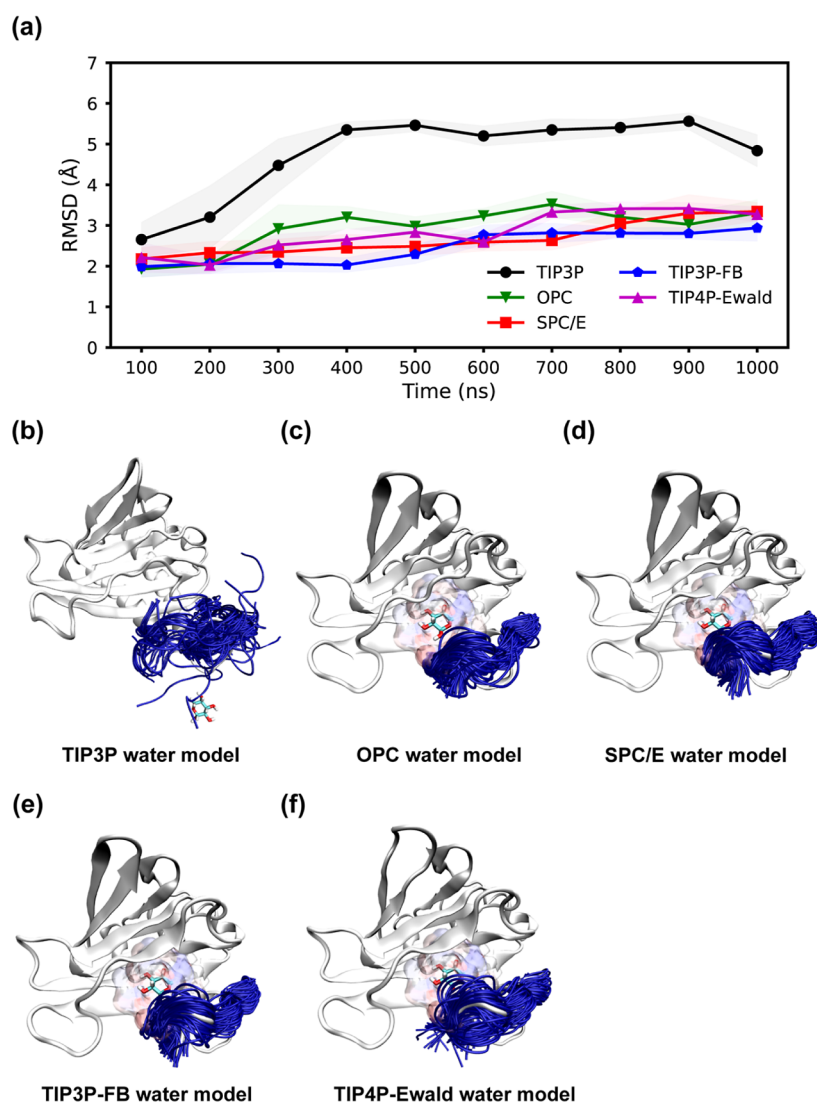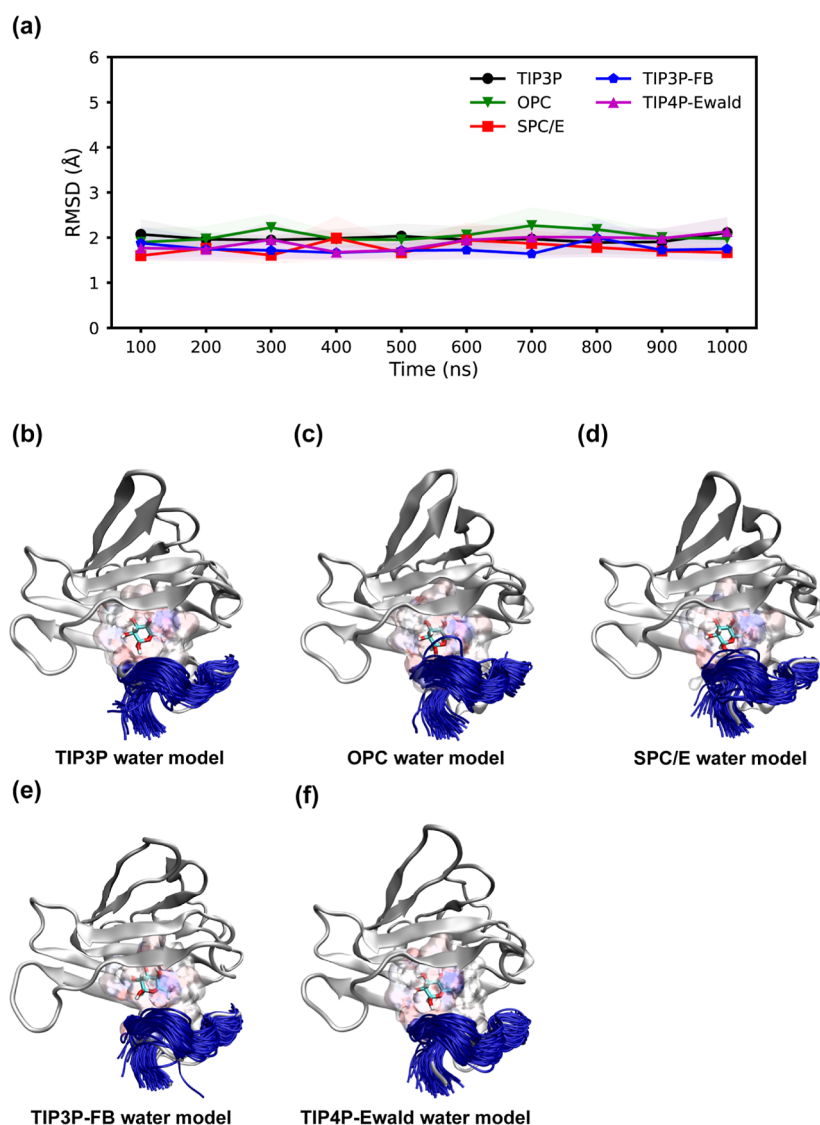
Figure 2. (a) Average RMSD for every 100 ns simulations of the binding motif (residues 7 to 18) in 2ARC in different water models. The shaded regions represent the standard deviation of the mean RMSD values. Superposition of the binding motif (highlighted in blue) in MD conformations of 2ARC in (b) TIP3P water model, (c) OPC water model, (d) SPC/E water model, (e) TIP3P-FB, and (f) TIP4P-Ewald water model, respectively.

including DGlcpb1−4DGlcpb1-OH in 1I82, DGlcpa1−4DGlcpa1−4DGlcpb1-OH in 2GGU, and DGalpb1−4DGlcpb1-OH in 2NN8. These analyses were conducted using the most populated glycosidic torsion angles to represent glycan conformation as in our previous study. The glycans in the rest of the complexes (1I8A, 2AAC, and 2ARC) are simple monosaccharides and are excluded from the glycan conformation stability analysis.

For the 1I82 complex, the glycosidic torsion angle distributions for DGlcpb1−4DGlcpb1-OH are shown in Figure 4. Here, the Phi torsion angles across all five water models exhibit minimal variance, generally within 1 degree of each other. However, there is a significant difference in the Psi angle within the OPC water model ($-148.29°$) compared to other water models ($\sim-153°$). To further assess the stability of these conformations, DFT was utilized to calculate the single-point energies of the predominant glycan conformations for each water model, as outlined in Table 2. The more negative energy value associated with the OPC model indicates higher stability for the DGlcpb1−4DGlcpb1-OH conformation.

For the 2GGU complex, the trisaccharide DGlcpa1−4DGlcpa1−4DGlcpb1-OH exhibited two sets of glycosidic torsion angle distributions as shown in Figure 5. The first set displayed minimal variation in Phi mode angles, with the smallest value being $92.79°$, as recorded in the OPC water model (Figure 5b). The Phi modal angles and Psi angles of the second set demonstrated proximity across the five water models (Figure 5c). Furthermore, the similar single-point energy at the bound conformations of DGlcpa1−4DGlcpa1−4DGlcpb1-OH in TIP3P, SPC/E, and TIP3P-FB water models indicated a comparable level of stability, as shown in Table 3. In contrast, the trisaccharide bound conformation stability in OPC and TIP4P-Ewald water models exhibited more negative single-point energy, suggesting more favorable bound conformations in OPC and TIP4P-Ewald water models.

In the context of the 2NN8 complex, the glycosidic torsion angle distributions corresponding to DGalpb1−4DGlcpb1-OH maintained a high degree of congruence, with Phi and Psi modal angles differing by less than $1°$, as illustrated in Figure S3. This alignment was further corroborated by the similar single-point energies in the corresponding bound conformations, which

(a)

(b)   **(c)**   **(d)**

TIP3P water model   OPC water model   SPC/E water model

(e)   **(f)**

TIP3P-FB water model   TIP4P-Ewald water model

Figure 3. (a) Average RMSD for every 100 ns simulations of the binding motif (residues 7 to 18) in 2AAC in different water models. The shaded regions represent the standard deviation of the mean RMSD values. Superposition of the binding motif (highlighted in blue) in MD conformations of 2AAC in (b) TIP3P water model, (c) OPC water model, (d) SPC/E water model, (e) TIP3P-FB, and (f) TIP4P-Ewald water model, respectively.

indicated uniform conformational stability, as shown in Table S7.

Alchemical ABFE. The findings in previous sections have dissected individual factors that influence protein−glycan binding affinity. While this dissection is critical, it is the holistic integration of these factors that provides a comprehensive understanding of the interaction dynamics. Alchemical ABFE calculations are pivotal in this integration, offering a quantitatively precise method that considers all contributing elements, including the often critical roles of solvent dynamics and protein conformational flexibility.[60,66,67] The alchemical ABFE is crucial for revealing glycan binding specificity, which is essential in understanding the molecular mechanisms of host−pathogen interaction.[23,68,69] We hypothesize that the precision of ABFE estimations is influenced by the choice of water model, especially in the protein−glycan complex system. Hence, we evaluated the accuracy of the five different water models in alchemical ABFE calculation in six protein−glycan complexes.

Table 4 shows the metrics of ABFE results for different water models. Alchemical binding free energy calculations are

categorized into relative and absolute types.[60,66] Relative binding free energy focuses on the impact of functional group substitutions on a ligand's binding affinity, emphasizing comparative efficacy. It is more commonly used in comparing the binding affinities between two similar molecules. However, for ABFE calculations, like those in this study, accuracy against experimental values is crucial. Here, the MUE and RMSE are superior metrics compared to the correlation coefficient, as they directly reflect the precision of computational predictions against experimental data.

In the evaluation of five water models, the OPC water model demonstrates the highest accuracy, achieving both the lowest MUE of 1.57 kcal/mol and the lowest RMSE of 1.85 kcal/mol. Figure 6 illustrates this comparison between experimental and computationally calculated binding free energies, and the corresponding values are listed in Table S8. The ABFE results obtained using the OPC water model exhibit remarkable precision, with almost all deviations falling within the margin of error of 1 kcal/mol. In contrast, the widely used water model, TIP3P, demonstrates comparatively lower precision, with an
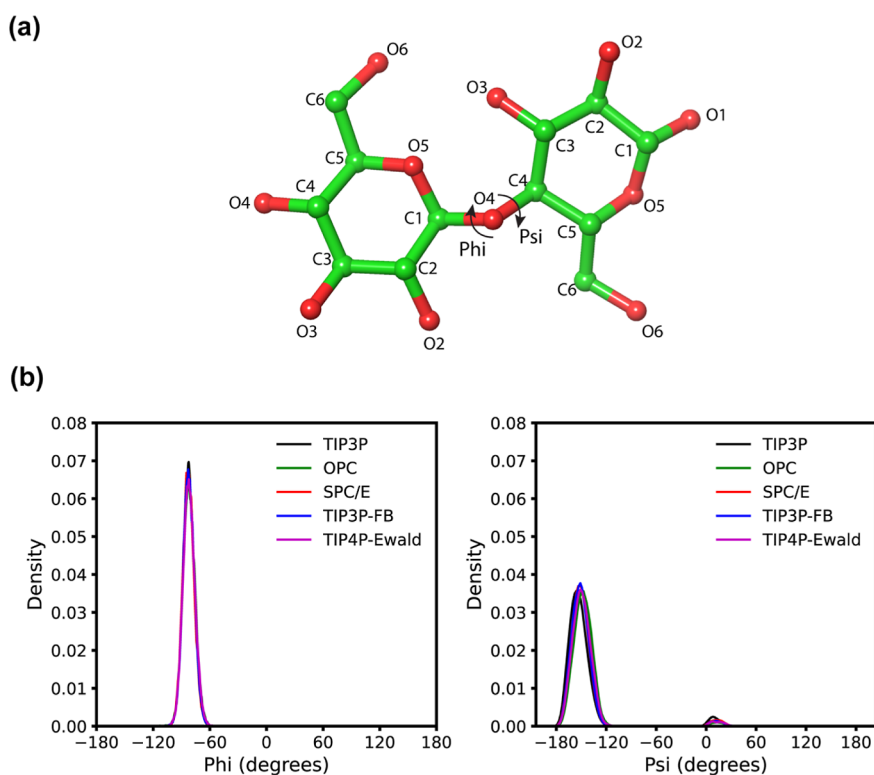
Figure 4. Glycosidic torsion angles (Phi and Psi) distribution of DGlcpb1−4DGlcpb1-OH in 1I82 complex. (a) Illustration of structure of DGlcpb1−4DGlcpb1-OH. Glycosidic torsion angles are defined as Phi = O5−C1−O4−C4 and Psi = C1−O4−C4−C5. (b) The probability density distribution curves of the glycosidic torsion angles in different water models (black for TIP3P, green for OPC, red for SPC/E, blue for TIP3P-FB, and magenta for TIP4P-Ewald).

MUE of 3.52 kcal/mol and the highest RMSE among the evaluated water models, recorded at 4.17 kcal/mol. This elevated RMSE is largely influenced by a significant outlier observed for the 1I82 protein−glycan system. An analysis of single-point energy calculations for the glycan in its bound state across different water models (see Table 2) reveals that TIP3P yields the highest energy (−1298.375078 hartree), indicating that the glycan is least stable in this model. In comparison, the OPC model provides a more stable energy (−1298.379260 hartree), making the glycan approximately 2.624 kcal/mol more stable than in TIP3P. These small yet significant energy differences at the molecular scale likely contribute to the anomalously high ABFE observed with TIP3P for 1I82. Additionally, the backbone dihedral angles Phi and Psi for TIP3P (−82.34 and −155.14°, respectively) differ notably from those in other models, particularly in the Psi angle, which may further influence the binding free energy calculations.

Regarding the other water models, the performance metrics are as follows: the SPC/E model exhibits a MUE of 3.02 kcal/mol and an RMSE of 3.53 kcal/mol, the TIP3P-FB model shows a MUE of 3.41 kcal/mol and an RMSE of 3.83 kcal/mol, and the TIP4P-Ewald model presents a MUE of 2.42 kcal/mol and an RMSE of 3.05 kcal/mol. When considering the RMSE metric as a standard for evaluating the accuracy of water models in ABFE calculations, the ranking from most to least accurate is as follows: OPC, TIP4P-Ewald, SPC/E, TIP3P-FB, and TIP3P. It is worth noting that this study focuses on the protein−glycan system, concluding that TIP3P has lower accuracy for ABFE predictions in this context. However, TIP3P may perform well in other systems, such as small molecules without standard force

fields,[70] so our conclusions are specific to protein−glycan systems.

We conducted additional analysis focusing on a crucial component of the ABFE cycle, the ligand decoupling in bulk in different water models. This analysis provided a more direct evaluation of the performance of different water models. As illustrated in Figure S4, the OPC water model consistently produces less negative free energy values compared to TIP3P, SPC/E, TIP3P-FB, and TIP4P-Ewald. These findings are consistent with previous research, which has demonstrated that the OPC model predicts bulk water properties more accurately than other water models.[27] Accurate prediction of bulk water properties is critical as it affects the accuracy of the decoupling free energy of glycans in bulk water. The more negative decoupling free energies in bulk water for the other water models lead to overestimated ABFE values. Additionally, the OPC model has been shown to provide a more accurate prediction of solvation free energy for small molecules, reflecting its superior capability in representing solute−water interactions.[27,71]

Which Water Model to Use? Previous studies, especially those focused on solvent benchmarking, have greatly enhanced our understanding of how water models influence the energetics and structural properties of biomolecular systems such as glycosaminoglycans and other polymers. It was shown that the choice of water model can profoundly impact the structural and energetic properties of glycosaminoglycans.[72,73] Explicit water models like TIP3P, TIP4P, and OPC have shown superior ability to capture detailed interactions of protein−glycosaminoglycan complexes. Conversely, implicit water models often produce less fluctuation and may overestimate binding energies
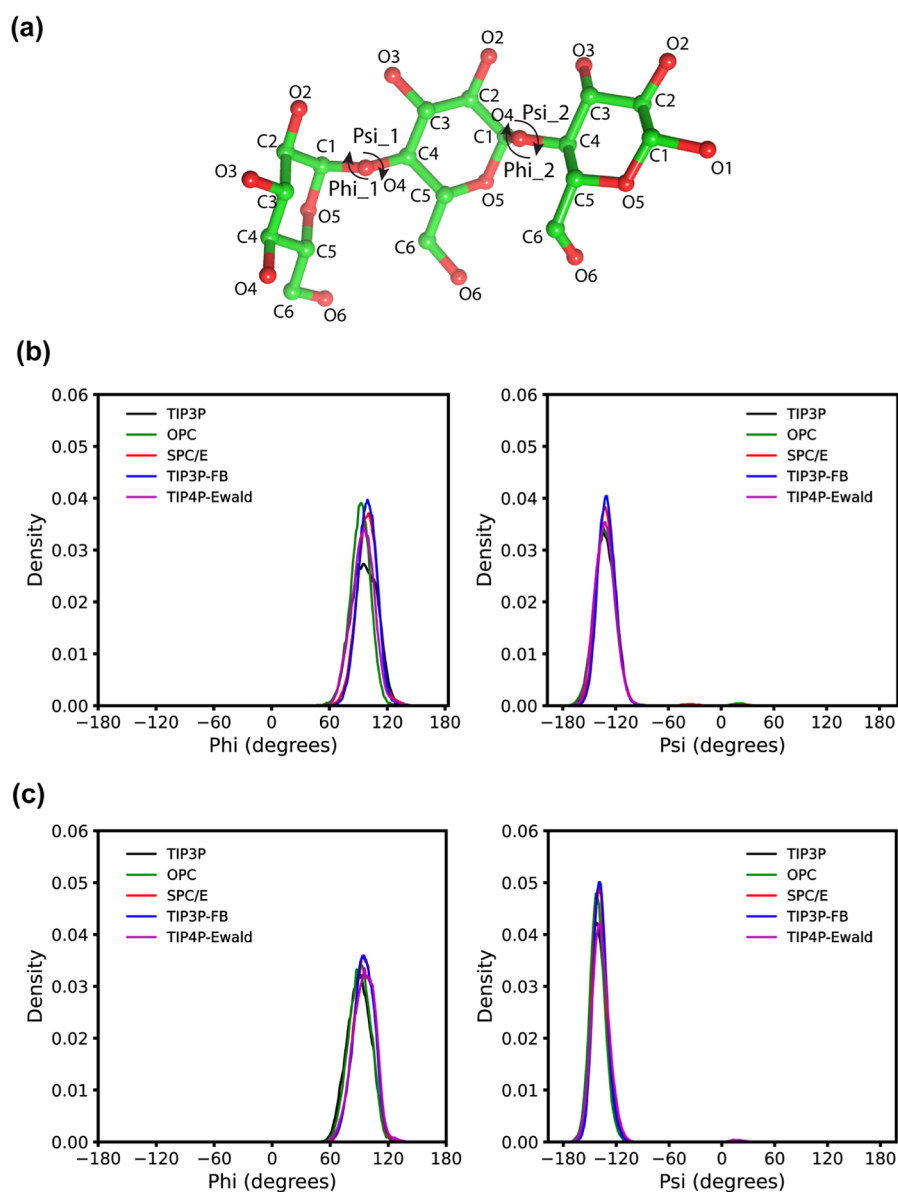
Figure 5. Glycosidic torsion angles (Phi and Psi) distribution of DGlcpa1−4DGlcpa1−4DGlcpb1-OH in 2GGU complex. (a) Illustration of structure of DGlcpa1−4DGlcpa1−4DGlcpb1-OH. Glycosidic torsion angles are defined as Phi = O5−C1−O4−C4 and Psi = C1−O4−C4−C5. The probability density distribution curves of the glycosidic torsion angles (b) group 1 and (c) group 2 in different water models (black for TIP3P, green for OPC, red for SPC/E, blue for TIP3P-FB, and magenta for TIP4P-Ewald).

Table 4. Statistics of Alchemical ABFE Results for Different Water Models

Water model	MUE (kcal/mol)	RMSE (kcal/mol)
TIP3P	3.52	4.17
OPC	1.57	1.85
SPC/E	3.02	3.53
TIP3P-FB	3.41	3.83
TIP4P-Ewald	2.42	3.05

and non-native contacts, underscoring their limitations in accurately representing solvent effects. Notably, the commonly used TIP3P water model failed to produce proper structural conformations of the specific glycosaminoglycan, exhibiting unique behavior in longer simulations. Another study on the thermosensitivity of poly(*N*-isopropylacrylamide) (PNIPAM), reveals that water models can induce substantial shifts in

transition temperatures and alter polymer−solvent interactions.[74] These factors are crucial for understanding phase behavior and structural transitions. They found that SPC and TIP3P water models can lead to significant deviations due to their less accurate bulk solvent descriptions. These insights emphasize the necessity of careful selection of water models in MD simulations to ensure accurate representation of molecular behavior.

The OPC water model represents one of the latest advancements in the simulation of aqueous environments.[27] First integrated into the AMBER environment, the OPC water model has demonstrated significant efficacy, particularly when combined with the latest AMBER protein force field, ff19SB.[27,42] Our study corroborates these findings, illustrating that the combination of OPC, ff19SB, and GLYCAM06j can precisely predict protein−glycan interactions, both in dynamic behaviors and alchemical binding free energy calculations.
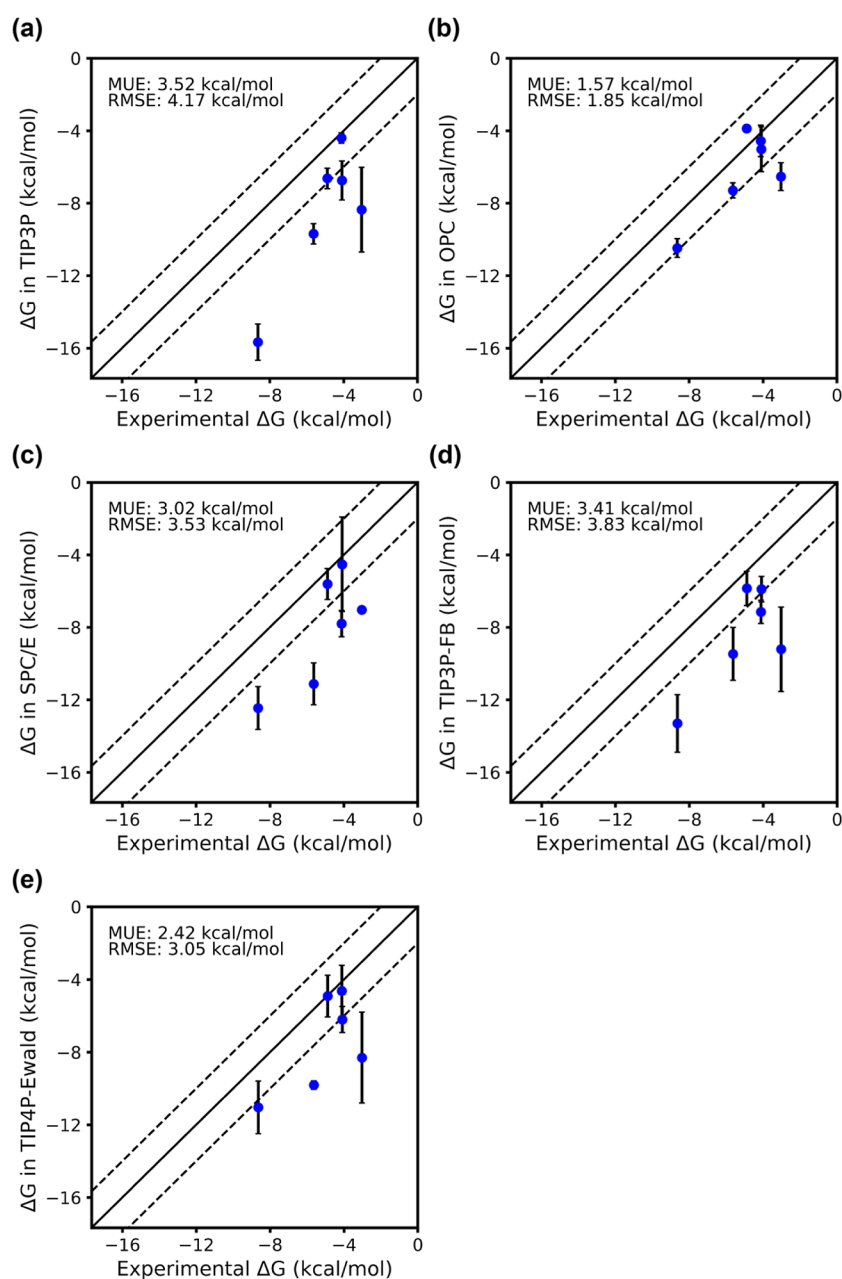
Figure 6. Comparison of calculated alchemical absolute binding free energies versus experimentally measured values in different water models as shown in panels (a−e) for TIP3P, OPC, SPC/E, TIP3P-FB, and TIP4P-Ewald, respectively. The calculations were performed for six protein−glycan complexes with PDB IDs: 1I82, 1I8A, 2AAC, 2ARC, 2GGU, and 2NN8. The corresponding glycans are DGlcpβ1−4DGlcpβ1-OH, DGlcpβ1-OH, DFucpβ1-OH, LArapα1-OH, DGlcpα1−4DGlcpα1−4DGlcpβ1-OH, and DGalpβ1−4DGlcpβ1-OH. The detailed ABFE values for each data point are listed in Table S8. The solid line indicates $y = x$, while the dashed lines correspond to $y = x \pm 1$ kcal/mol.

The widespread utilization of the TIP3P water model in biomolecular simulations is largely due to its compatibility across a variety of force fields and its computational efficiency. While the TIP3P water model can adequately simulate dynamic behaviors of common protein systems at physiological temperatures (approximately 310 K), it struggles to accurately represent the properties of more complex systems, such as lipid monolayers.[27,75] Previous studies indicate that, in comparison, the OPC water model provides a more precise simulation of lipid monomer behaviors.[75] This distinction underscores the critical need for careful selection of water models in simulations for specific systems, where the choice of model can significantly impact the accuracy of the results.

Further, in protein−ligand complexes featuring flexible binding motifs, the TIP3P model may induce anomalous fluctuations and potentially affect ligand binding. Other water models, when combined with suitable force fields, demonstrate commendable performance. For example, the combination of the SPC/E water model and the OPLS force field yields accurate results, especially in simulations involving small drug molecules.[76,77]

■ **CONCLUSION**

In this study, we thoroughly investigated protein−glycan interactions in various water models, yielding critical insights into the effectiveness of these models for simulating such

interactions. Our findings, focusing on overall structural stability, reveal that certain protein−glycan complexes maintain remarkable stability across all water models while others display notable fluctuations under specific conditions. Furthermore, our study demonstrates that when the residues from the binding motif form weak hydrogen bonds with the glycan, the choice of water model can significantly affect the conformation stability to further influence the binding with the glycan. The investigation into hydrogen bond dynamics within these complexes further elucidates how different water models can alter hydrogen bond patterns, thereby affecting the stability and behavior of these complexes. Additionally, our study delves into the conformational stability of binding motifs and the conformational stability of glycans in the bound state. These analyses reveal that the choice of water model profoundly impacts the stability of these flexible binding motifs and the glycan conformational stability. These findings are vital for understanding the structural dynamics of protein−glycan complexes.

The investigation of alchemical ABFE calculations revealed that the OPC water model outperformed others, demonstrating the highest accuracy, making it a reliable choice for such calculations. In contrast, the widely used TIP3P water model exhibited a higher RMSE, signifying its inadequacy in alchemical binding free energy calculations. These findings emphasize the significant disparities in the predictive capabilities of water models regarding binding free energies, with the hierarchy of accuracy ranking them from most to least effective as OPC, TIP4P-Ewald, SPC/E, TIP3P-FB, and TIP3P. We thus emphasize the importance of water models in alchemical binding free energy calculation, especially for the protein−glycan system.

In summary, this study not only enhances our understanding of protein−glycan interactions but also establishes a comprehensive guideline for selecting the most suitable water models in computational simulations. The insights gained hold significant potential for advancing our knowledge of protein−glycan interaction mechanisms and binding specificity through full atomistic modeling. These findings could significantly contribute to the development of new therapeutic strategies targeting protein−glycan interactions, paving the way for more effective and targeted treatments in drug design.

■ ASSOCIATED CONTENT

Data Availability Statement

PDB files of the protein−glycan complexes are available in the Supporting Information, along with topology and parameter files for the proteins and ligands. All model constructions are performed using tleap in AmberTools23 (https://ambermd.org/AmberTools.php) and VMD 1.9.3 (https://www.ks.uiuc.edu/Research/vmd/vmd-1.9.3/). All molecular dynamics simulations and alchemical absolute binding free energy (ABFE) calculations are conducted using GPU-accelerated implementation of Amber22 (https://ambermd.org/AmberMD.php). The script for setting up ABFE calculation can be accessed at https://github.com/GHeinzelmann/BAT.py. The trajectory analyses were performed using cpptraj in AmberTools23 and VMD 1.9.3. Jaguar 12.1 from the Schrödinger Suite of software tools was used to conduct the single-point energy calculation https://newsite.schrodinger.com/platform/products/jaguar/.

Ⓢ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jcim.4c01361.

3D conformations of each complex structure are shown in Figure S1. RMSF curves of six protein−glycan complex models across five water models are shown in Figure S2. Detailed Hydrogen bond information for the protein−glycan complexes is in Tables S1−S6. Glycosidic torsion angle distributions of DGalpb1−4DGlcpb1-OH in 2NN8 are in Table S7 and Figure S3. Detailed experimentally measured ΔG and computationally predicted ΔG are listed in Table S8. Free energy contributions in ABFE for ligand decoupling in bulk solvent in different water models are shown in Figure S4 (ZIP)

■ AUTHOR INFORMATION

Corresponding Author

Mona S. Minkara − *Department of Bioengineering, Northeastern University, Boston, Massachusetts 02120, United States;* ⊙ orcid.org/0000-0003-1821-2725; Email: m.minkara@northeastern.edu

Author

Deng Li − *Department of Bioengineering, Northeastern University, Boston, Massachusetts 02120, United States;* ⊙ orcid.org/0009-0002-3030-8009

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jcim.4c01361

Author Contributions

M.S.M. provided funding to support this work, supervised the work, and revised the manuscript. D.L. designed the research and carried out all simulations. D.L. collected and analyzed the data and wrote the manuscript.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

■ REFERENCES

(1) Collins, B. E.; Paulson, J. C. Cell surface biology mediated by low affinity multivalent protein-glycan interactions. *Curr. Opin. Chem. Biol.* **2004**, *8* (6), 617−625.

(2) van Kooyk, Y.; Rabinovich, G. A. Protein-glycan interactions in the control of innate and adaptive immune responses. *Nat. Immunol.* **2008**, *9* (6), 593−601.

(3) Raman, R.; Tharakaraman, K.; Sasisekharan, V.; Sasisekharan, R. Glycan-protein interactions in viral pathogenesis. *Curr. Opin Struc Biol.* **2016**, *40*, 153−162.

(4) Craig, L.; Pique, M. E.; Tainer, J. A. Type IV pilus structure and bacterial pathogenicity. *Nat. Rev. Microbiol.* **2004**, *2* (5), 363−378.

(5) Smedley, J. G., 3rd; Jewell, E.; Roguskie, J.; Horzempa, J.; Syboldt, A.; Stolz, D. B.; Castric, P. Influence of pilin glycosylation on

Pseudomonas aeruginosa 1244 pilus function. *Infect. Immun.* **2005**, *73* (12), 7922−7931.

(6) Holmskov, U.; Malhotra, R.; Sim, R. B.; Jensenius, J. C. Collectins: collagenous C-type lectins of the innate immune defense system. *Immunol. Today* **1994**, *15* (2), 67−74.

(7) Cambi, A.; Koopman, M.; Figdor, C. G. How C-type lectins detect pathogens. *Cell. Microbiol.* **2005**, *7* (4), 481−488.

(8) Sano, H.; Kuroki, Y. The lung collectins, SP-A and SP-D, modulate pulmonary innate immunity. *Mol. Immunol.* **2005**, *42* (3), 279−287.

(9) Reid, K. B. Functional roles of the lung surfactant proteins SP-A and SP-D in innate immunity. *Immunobiology* **1998**, *199* (2), 200−207.

(10) Ip, W. K.; Takahashi, K.; Ezekowitz, R. A.; Stuart, L. M. Mannose-binding lectin and innate immunity. *Immunol. Rev.* **2009**, *230* (1), 9−21.

(11) Hartshorn, K. L.; Crouch, E. C.; White, M. R.; Eggleton, P.; Tauber, A. I.; Chang, D.; Sastry, K. Evidence for a Protective Role of Pulmonary Surfactant Protein-D (Sp-D) against Influenza-a Viruses. *J. Clin. Invest.* **1994**, *94* (1), 311−319.

(12) Hsieh, M.-H.; Beirag, N.; Murugaiah, V.; Chou, Y.-C.; Kuo, W.-S.; Kao, H.-F.; Madan, T.; Kishore, U.; Wang, J.-Y. Human Surfactant Protein D Binds Spike Protein and Acts as an Entry Inhibitor of SARS-CoV-2 Pseudotyped Viral Particles. *Original Research* **2021**, *12*, 641360.

(13) Meschi, J.; Crouch, E. C.; Skolnik, P.; Yahya, K.; Holmskov, U.; Leth-Larsen, R.; Tornoe, I.; Tecle, T.; White, M. R.; Hartshorn, K. L. Surfactant protein D binds to human immunodeficiency virus (HIV) envelope protein gp120 and inhibits HIV replication. *J. Gen. Virol.* **2005**, *86*, 3097−3107.

(14) Ji, X.; Gewurz, H.; Spear, G. T. Mannose binding lectin (MBL) and HIV. *Mol. Immunol.* **2005**, *42* (2), 145−152.

(15) Mnich, M. E.; van Dalen, R.; van Sorge, N. M. C-Type Lectin Receptors in Host Defense Against Bacterial Pathogens. *Front. Cell. Infect. Microbiol.* **2020**, *10*, 309.

(16) Goh, B. C.; Rynkiewicz, M. J.; Cafarella, T. R.; White, M. R.; Hartshorn, K. L.; Allen, K.; Crouch, E. C.; Calin, O.; Seeberger, P. H.; Schulten, K.; et al. Molecular Mechanisms of Inhibition of Influenza by Surfactant Protein D Revealed by Large-Scale Molecular Dynamics Simulation. *Biochemistry* **2013**, *52* (47), 8527−8538.

(17) Goh, B. C.; Wu, H. X.; Rynkiewicz, M. J.; Schulten, K.; Seaton, B. A.; McCormack, F. X. Elucidation of Lipid Binding Sites on Lung Surfactant Protein A Using X-ray Crystallography, Mutagenesis, and Molecular Dynamics Simulations. *Biochemistry* **2016**, *55* (26), 3692−3701.

(18) Newhouse, E. I.; Xu, D.; Markwick, P. R. L.; Amaro, R. E.; Pao, H. C.; Wu, K. J.; Alam, M.; McCammon, J. A.; Li, W. W. Mechanism of Glycan Receptor Recognition and Specificity Switch for Avian, Swine, and Human Adapted Influenza Virus Hemagglutinins: A Molecular Dynamics Perspective. *J. Am. Chem. Soc.* **2009**, *131* (47), 17430−17442.

(19) 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.

(20) Casalino, L.; Gaieb, Z.; Goldsmith, J. A.; Hjorth, C. K.; Dommer, A. C.; Harbison, A. M.; Fogarty, C. A.; Barros, E. P.; Taylor, B. C.; McLellan, J. S.; et al. Beyond Shielding: The Roles of Glycans in the SARS-CoV-2 Spike Protein. *ACS Cent. Sci.* **2020**, *6* (10), 1722−1734.

(21) Stewart-Jones, G. B. E.; Soto, C.; Lemmin, T.; Chuang, G. Y.; Druz, A.; Kong, R.; Thomas, P. V.; Wagh, K.; Zhou, T. Q.; Behrens, A. J.; et al. Trimeric HIV-1-Env Structures Define Glycan Shields from Clades A, B, and G. *Cell* **2016**, *165* (4), 813−826.

(22) Zhao, P.; Praissman, J. L.; Grant, O. C.; Cai, Y. F.; Xiao, T. S.; Rosenbalm, K. E.; Aoki, K.; Kellman, B. P.; Bridger, R.; Barouch, D. H.; et al. Virus-Receptor Interactions of Glycosylated SARS-CoV-2 Spike and Human ACE2 Receptor. *Cell Host Microbe* **2020**, *28* (4), 586−601.e6.

(23) Li, D.; Minkara, M. S. Elucidating the enhanced binding affinity of a double mutant SP-D with trimannose on the influenza A virus using molecular dynamics. *Comput. Struct. Biotechnol. J.* **2022**, *20*, 4984−5000.

(24) Plazinska, A.; Plazinski, W. Comparison of Carbohydrate Force Fields in Molecular Dynamics Simulations of Protein-Carbohydrate Complexes. *J. Chem. Theory Comput.* **2021**, *17* (4), 2575−2585.

(25) Battistel, M. D.; Pendrill, R.; Widmalm, G.; Freedberg, D. I. Direct Evidence for Hydrogen Bonding in Glycans: A Combined NMR and Molecular Dynamics Study. *J. Phys. Chem. B* **2013**, *117* (17), 4860−4869.

(26) Bandyopadhyay, D.; Mohan, S.; Ghosh, S. K.; Choudhury, N. Correlation of Structural Order, Anomalous Density, and Hydrogen Bonding Network of Liquid Water. *J. Phys. Chem. B* **2013**, *117* (29), 8831−8843.

(27) Izadi, S.; Anandakrishnan, R.; Onufriev, A. V. Building Water Models: A Different Approach. *J. Phys. Chem. Lett.* **2014**, *5* (21), 3863−3871.

(28) Kumar, S.; Frank, M.; Schwartz-Albiez, R. Understanding the Specificity of Human Galectin-8C Domain Interactions with Its Glycan Ligands Based on Molecular Dynamics Simulations. *PLoS One* **2013**, *8* (3), No. e59761.

(29) Zhang, J. L.; Zheng, Q. C.; Zhang, H. X. Insight into the Dynamic Interaction of Different Carbohydrates with Human Surfactant Protein D: Molecular Dynamics Simulations. *J. Phys. Chem. B* **2010**, *114* (21), 7383−7390.

(30) Lepsik, M.; Sommer, R.; Kuhaudomlarp, S.; Lelimousin, M.; Paci, E.; Varrot, A.; Titz, A.; Imberty, A. Induction of rare conformation of oligosaccharide by binding to calcium-dependent bacterial lectin: X-ray crystallography and modelling study. *Eur. J. Med. Chem.* **2019**, *177*, 212−220.

(31) Notenboom, V.; Boraston, A. B.; Kilburn, D. G.; Rose, D. R. Crystal Structures of the Family 9 Carbohydrate-Binding Module from Thermotoga maritima Xylanase 10A in Native and Ligand-Bound Forms'. *Biochemistry* **2001**, *40* (21), 6248−6256.

(32) Boraston, A. B.; Creagh, A. L.; Alam, M. M.; Kormos, J. M.; Tomme, P.; Haynes, C. A.; Warren, R. A. J.; Kilburn, D. G. Binding Specificity and Thermodynamics of a Family 9 Carbohydrate-Binding Module from Thermotoga maritima Xylanase 10A. *Biochemistry* **2001**, *40* (21), 6240−6247.

(33) Soisson, S. M.; MacDougall-Shackleton, B.; Schleif, R.; Wolberger, C. The 1.6 A crystal structure of the AraC sugar-binding and dimerization domain complexed with D-fucose. *J. Mol. Biol.* **1997**, *273* (1), 226−237.

(34) Wilcox, G. The interaction of L-arabinose and D-fucose with AraC protein. *J. Biol. Chem.* **1974**, *249* (21), 6892−6894.

(35) Soisson, S. M.; MacDougallShackleton, B.; Schleif, R.; Wolberger, C. Structural basis for ligand-regulated oligomerization of AraC. *Science* **1997**, *276* (5311), 421−425.

(36) Crouch, E.; McDonald, B.; Smith, K.; Cafarella, T.; Seaton, B.; Head, J. Contributions of Phenylalanine 335 to Ligand Recognition by Human Surfactant Protein D. *J. Biol. Chem.* **2006**, *281* (26), 18008−18014.

(37) Lim, B. L.; Wang, J. Y.; Holmskov, U.; Hoppe, H. J.; Reid, K. B. M. Expression of the Carbohydrate-Recognition Domain of Lung Surfactant Protein-D and Demonstration of Its Binding to Lipopolysaccharides of Gram-Negative Bacteria. *Biochem. Biophys. Res. Commun.* **1994**, *202* (3), 1674−1680.

(38) Wang, R. X.; Fang, X. L.; Lu, Y. P.; Wang, S. M. The PDBbind database: Collection of binding affinities for protein-ligand complexes with known three-dimensional structures. *J. Med. Chem.* **2004**, *47* (12), 2977−2980.

(39) Collins, P. M.; Hidari, K. I. P. J.; Blanchard, H. Slow diffusion of lactose out of galectin-3 crystals monitored by X-ray crystallography: possible implications for ligand-exchange protocols. *Acta Crystallogr. D* **2007**, *63*, 415−419.

(40) Hirabayashi, J.; Hashidate, T.; Arata, Y.; Nishi, N.; Nakamura, T.; Hirashima, M.; Urashima, T.; Oka, T.; Futai, M.; Muller, W. E. G.; et al. Oligosaccharide specificity of galectins: a search by frontal affinity chromatography. *Bba-Gen Subjects* **2002**, *1572* (2−3), 232−254.

(41) Case, D. A.; Aktulga, H. M.; Belfon, K.; Cerutti, D. S.; Cisneros, G. A.; Cruzeiro, V. W. D.; Forouzesh, N.; Giese, T. J.; Götz, A. W.;

Gohlke, H.; et al. AmberTools. *J. Chem. Inf. Model.* **2023**, *63* (20), 6183−6191.

(42) Tian, C.; Kasavajhala, K.; Belfon, K. A. A.; Raguette, L.; Huang, H.; Migues, A. N.; Bickel, J.; Wang, Y. Z.; Pincay, J.; Wu, Q.; et al. ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. *J. Chem. Theory Comput.* **2020**, *16* (1), 528−552.

(43) 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.

(44) 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.

(45) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. The Missing Term in Effective Pair Potentials. *J. Phys. Chem-Us* **1987**, *91* (24), 6269−6271.

(46) Wang, L. P.; Martinez, T. J.; Pande, V. S. Building Force Fields: An Automatic, Systematic, and Reproducible Approach. *J. Phys. Chem. Lett.* **2014**, *5* (11), 1885−1891.

(47) Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T. Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *J. Chem. Phys.* **2004**, *120* (20), 9665−9678.

(48) Vega, C.; Abascal, J. L. F.; Conde, M. M.; Aragones, J. L. What ice can teach us about water interactions: a critical comparison of the performance of different water models. *Faraday Discuss.* **2009**, *141*, 251−276.

(49) Nada, H.; van der Eerden, J. P. J. M. An intermolecular potential model for the simulation of ice and water near the melting point: A six-site model of H_2O. *J. Chem. Phys.* **2003**, *118* (16), 7401−7413.

(50) Zhao, C. L.; Zhao, D. X.; Bei, C. C.; Meng, X. N.; Li, S. M.; Yang, Z. Z. Seven-Site Effective Pair Potential for Simulating Liquid Water. *J. Phys. Chem. B* **2019**, *123* (21), 4594−4603.

(51) Zhao, C. L.; Zhao, D. X.; Jiang, Q. Y.; Zhang, H. X.; Li, S. M.; Yang, Z. Z. Polarizable TIP7P Water Model with Perturbation Charges Evaluated from ABEEM. *J. Phys. Chem. B* **2020**, *124* (12), 2450−2464.

(52) Uberuaga, B. P.; Anghel, M.; Voter, A. F. Synchronization of trajectories in canonical molecular-dynamics simulations: Observation, explanation, and exploitation. *J. Chem. Phys.* **2004**, *120* (14), 6363−6374.

(53) Sindhikara, D. J.; Kim, S.; Voter, A. F.; Roitberg, A. E. Bad Seeds Sprout Perilous Dynamics: Stochastic Thermostat Induced Trajectory Synchronization in Biomolecules. *J. Chem. Theory Comput.* **2009**, *5* (6), 1624−1631.

(54) Åqvist, J.; Wennerström, P.; Nervall, M.; Bjelic, S.; Brandsdal, B. O. Molecular dynamics simulations of water and biomolecules with a Monte Carlo constant pressure algorithm. *Chem. Phys. Lett.* **2004**, *384* (4−6), 288−294.

(55) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977**, *23* (3), 327−341.

(56) Case, D. A.; A, H. M.; Belfon, K.; Ben-Shalom, I. Y.; Berryman, J. T.; Brozell, S. R.; Cerutti, D. S.; Cheatham, T. E., III; Cisneros, G. A.; Cruzeiro, V. W. D.; Darden, T. A.; Forouzesh, N.; Giambaşu, G.; Giese, T.; Gilson, M. K.; Gohlke, H.; Goetz, A. W.; Harris, J.; Izadi, S.; Izmailov, S. A.; Kasavajhala, K.; Kaymak, M. C.; King, E.; Kovalenko, A.; Kurtzman, T.; Lee, T. S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Machado, M.; Man, V.; Manathunga, M.; Merz, K. M.; Miao, Y.; Mikhailovskii, O.; Monard, G.; Nguyen, H.; O'Hearn, K. A.; Onufriev, A.; Pan, F.; Pantano, S.; Qi, R.; Rahnamoun, A.; Roe, D. R.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shajan, A.; Shen, J.; Simmerling, C. L.; Skrynnikov, N. R.; Smith, J.; Swails, J.; Walker, R. C.; Wang, J.; Wang, J.; Wei, H.; Wu, X.; Wu, Y.; Xiong, Y.; Xue, Y.; York, D. M.; Zhao, S.; Zhu, Q.; Kollman, P. A. *Amber 2023*; University of California: San Francisco, 2023.

(57) Heinzelmann, G.; Gilson, M. K. Automation of absolute protein-ligand binding free energy calculations for docking refinement and compound evaluation. *Sci. Rep.* **2021**, *11* (1), 1116.

(58) Irwin, B. W. J.; Huggins, D. J. Estimating Atomic Contributions to Hydration and Binding Using Free Energy Perturbation. *J. Chem. Theory Comput.* **2018**, *14* (6), 3218−3227.

(59) Boresch, S.; Tettinger, F.; Leitgeb, M.; Karplus, M. Absolute binding free energies: A quantitative approach for their calculation. *J. Phys. Chem. B* **2003**, *107* (35), 9535−9551.

(60) Lee, T. S.; Allen, B. K.; Giese, T. J.; Guo, Z. Y.; Li, P. F.; Lin, C.; McGee, T. D.; Pearlman, D. A.; Radak, B. K.; Tao, Y. J.; et al. Alchemical Binding Free Energy Calculations in AMBER20: Advances and Best Practices for Drug Discovery. *J. Chem. Inf. Model.* **2020**, *60* (11), 5595−5623.

(61) Shirts, M. R.; Chodera, J. D. Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys.* **2008**, *129* (12), 124105.

(62) Roe, D. R.; Cheatham, T. E., 3rd PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, *9* (7), 3084−3095.

(63) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graphics Modell.* **1996**, *14* (1), 33−38.

(64) Schrödinger, L. *New York, NY. Schrödinger Release 2023-2: Maestro*, 2023.

(65) Bochevarov, A. D.; Harder, E.; Hughes, T. F.; Greenwood, J. R.; Braden, D. A.; Philipp, D. M.; Rinaldo, D.; Halls, M. D.; Zhang, J.; Friesner, R. A. Jaguar A high-performance quantum chemistry software program with strengths in life and materials sciences. *Int. J. Quantum Chem.* **2013**, *113* (18), 2110−2142.

(66) Chodera, J. D.; Mobley, D. L.; Shirts, M. R.; Dixon, R. W.; Branson, K.; Pande, V. S. Alchemical free energy methods for drug discovery: progress and challenges. *Curr. Opin Struc Biol.* **2011**, *21* (2), 150−160.

(67) Aleksandrov, A.; Thompson, D.; Simonson, T. Alchemical free energy simulations for biological complexes: powerful but temperamental. *J. Mol. Recognit.* **2010**, *23* (2), 117−127.

(68) Perez, S.; Makshakova, O. Multifaceted Computational Modeling in Glycoscience. *Chem. Rev.* **2022**, *122* (20), 15914−15970.

(69) Mishra, S. K.; Koca, J. Assessing the Performance of MM/PBSA, MM/GBSA, and QM-MM/GBSA Approaches on Protein/Carbohydrate Complexes: Effect of Implicit Solvent Models, QM Methods, and Entropic Contributions. *J. Phys. Chem. B* **2018**, *122* (34), 8113−8121.

(70) Huggins, D. J. Comparing the Performance of Different AMBER Protein Forcefields, Partial Charge Assignments, and Water Models for Absolute Binding Free Energy Calculations. *J. Chem. Theory Comput.* **2022**, *18* (4), 2616−2630.

(71) Jorgensen, W. L.; Tirado-Rives, J. Potential energy functions for atomic-level simulations of water and organic and biomolecular systems. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102* (19), 6665−6670.

(72) Marcisz, M.; Samsonov, S. A. Solvent Model Benchmark for Molecular Dynamics of Glycosaminoglycans. *J. Chem. Inf. Model.* **2023**, *63* (7), 2147−2157.

(73) Anila, S.; Samsonov, S. A. Benchmarking Water Models in Molecular Dynamics of Protein-Glycosaminoglycan Complexes. *J. Chem. Inf. Model.* **2024**, *64* (5), 1691−1703.

(74) Quoika, P. K.; Kamenik, A. S.; Fernandez-Quintero, M. L.; Zacharias, M.; Liedl, K. R. Water model determines thermosensitive and physicochemical properties of poly(N-isopropylacrylamide) in molecular simulations. *Front. Mater.* **2023**, *10*, 1005781.

(75) Tempra, C.; Ollila, O. H. S.; Javanainen, M. Accurate Simulations of Lipid Monolayers Require a Water Model with Correct Surface Tension. *J. Chem. Theory Comput.* **2022**, *18* (3), 1862−1869.

(76) Harder, E.; Damm, W.; Maple, J.; Wu, C. J.; Reboul, M.; Xiang, J. Y.; Wang, L. L.; Lupyan, D.; Dahlgren, M. K.; Knight, J. L.; et al. OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. *J. Chem. Theory Comput.* **2016**, *12* (1), 281−296.

(77) Coskun, D.; Chen, W.; Clark, A. J.; Lu, C.; Harder, E. D.; Wang, L. L.; Friesner, R. A.; Miller, E. B. Reliable and Accurate Prediction of

Single-Residue pKa Values through Free Energy Perturbation Calculations. *J. Chem. Theory Comput.* **2022**, *18* (12), 7193−7204.

(78) Boerner, T. J.; Deems, S.; Furlani, T. R.; Knuth, S. L.; Towns, J. ACCESS: Advancing Innovation: NSF's Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support. *Practice and Experience in Advanced Research Computing*: Portland, OR, USA, 2023.