

# On the Role of Water Models in Quantifying the Binding Free Energy of Highly Conserved Water Molecules in Proteins: The Case of Concanavalin A

Elisa Fadda<sup>†</sup> and Robert J. Woods<sup>\*,†,‡</sup>

<sup>†</sup>School of Chemistry, National University of Ireland Galway, University Rd, Galway, Ireland

<sup>‡</sup>University of Georgia, Complex Carbohydrate Research Center, 315 Riverbend Rd, Athens, Georgia 30602, United States

 Supporting Information

**ABSTRACT:** The ability of ligands to displace conserved water molecules in protein binding sites is of significant interest in drug design and is particularly pertinent in the case of glycomimetic drugs. This concept was explored in previous work [Clarke et al. *J. Am. Chem. Soc.* **2001**, *123*, 12238–12247 and Kadirvelraj et al. *J. Am. Chem. Soc.* **2008**, *130*, 16933–16942] for a highly conserved water molecule located in the binding site of the prototypic carbohydrate-binding protein Concanavalin A (Con A). A synthetic ligand was designed with the aim of displacing such water. While the synthetic ligand bound to Con A in an analogous manner to that of the natural ligand, crystallographic analysis demonstrated that it did not displace the conserved water. In order to quantify the affinity of this particular water for the Con A surface, we report here the calculated standard binding free energy for this water in both ligand-bound and free Con A, employing three popular water models: TIP3P, TIP4P, and TIP5P. Although each model was developed to perform well in simulations of bulk-phase water, the computed binding energies for the isolated water molecule displayed a high sensitivity to the model. Both molecular dynamics simulation and free energy results indicate that the choice of water model may greatly influence the characterization of surface water molecules as conserved (TIP5P) or not (TIP3P) in protein binding sites, an observation of considerable significance to rational drug design. Structural and theoretical aspects at the basis of the different behaviors are identified and discussed.

## INTRODUCTION

Protein–carbohydrate recognition is a fundamental step in numerous essential biological processes, such as cell–cell and cell–matrix interactions, and conversely can be exploited in viral and bacterial adhesion and infection.<sup>1</sup> Detailed knowledge of the precise relationships between glycan structure and protein recognition requires the ability to relate atomic structures to thermodynamic properties. Such insight is paramount in the rational design of therapeutic agents that specifically target these interactions. The earliest carbohydrate-binding proteins to have received extensive analyses are the plant lectins.<sup>2,3</sup> Because of the large amount of structural, thermodynamic, and mutational data that have been accumulated for lectins, they serve as excellent models for probing the origins of the specificity of carbohydrate–protein binding.<sup>1,3</sup> A common feature in carbohydrate–protein complexes is the presence of discrete water molecules that appear to mediate the interaction through hydrogen bonding.<sup>2</sup> Such water molecules can also be found to occupy the same hydration sites in both the free and bound forms of the protein, suggesting that they play functional roles in recognition and binding. The presence of highly conserved water molecules in protein binding sites is of significant interest in drug design because there is the potential for a therapeutic agent to gain binding free energy through entropic gains relative to the native ligand upon their displacement.<sup>4–10</sup> The rational design of ligands that can displace waters from specific protein binding sites requires a knowledge of accurate values of the binding free energies of such water molecules; this information can only be obtained via computer

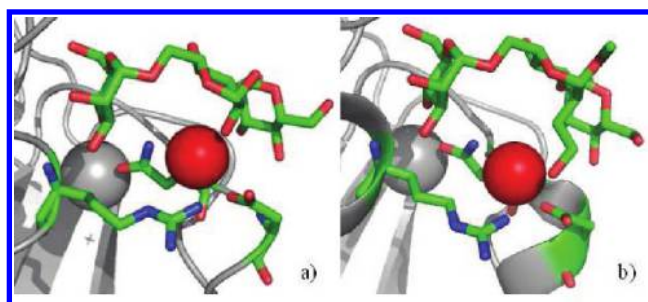
simulation techniques. Despite their apparent significance, the affinities of these conserved waters in carbohydrate–protein complexes remain unknown.

The ability to define a relationship between binding free energies of conserved water molecules and ligand affinity is also of great interest. In their work, Barillari et al.<sup>11</sup> analyzed the binding free energies of several conserved and displaced water molecules in different proteins complexed with a variety of ligands. Nonetheless, they were not able to establish any direct correlation between water binding free energies and ligand affinity. If such a correlation exists, it may well not be generalizable to all proteins and all ligands. It is likely that the gain or loss of binding affinity due to water displacement is system-specific, as well as the balance between entropic and enthalpic binding contributions for the conserved water molecules.<sup>8,9</sup> It is also important to recall that some water molecules are highly conserved because they play a structural role in the scaffold to which the ligand binds. Therefore, as in the mutation of key protein residues in the binding site, it is foreseeable that the displacement of these water molecules could also destabilize the ligand.

Con A, isolated from the jack bean (*Canavalia ensiformis*), is not only one of the best-characterized lectins but also shows a well-defined example of water-mediated carbohydrate binding. Isothermal titration microcalorimetry (ITC) experiments have

Received: June 14, 2011

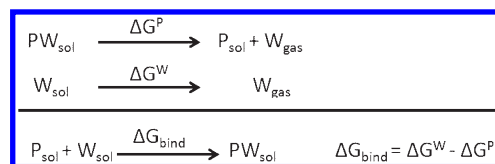
Published: August 23, 2011



**Figure 1.** (a) Detail of the saccharide binding site of Concanavalin A (Con A) bound to 3MAN (PDBID 1CVN, ref 14). (b) Con A bound to 3HET (PDBID 3D4K, ref 25). Structural water is shown as a red sphere, while ligands and key residues, i.e., ASN 14, ASP 16, and ARG 228, are shown as sticks. Oxygen atoms are shown in red, carbon in green, and nitrogen in blue. The remainder of the protein is shown in gray, with the structural metal ions shown as spheres.

established that Con A binds the trisaccharide  $\alpha$ -D-Manp(1–6)-[ $\alpha$ -D-Manp(1–3)] $\alpha$ -D-Manp (3MAN), present in all *N*-linked glycans, with a 60-fold higher affinity than the monomer methyl- $\alpha$ -D-mannopyranose.<sup>12</sup> The high affinity of Con A for the trimannoside core has been the subject of numerous studies.<sup>12–20</sup> X-ray crystallographic data reveal the presence of specificity-defining hydrogen bonds, as well as nonspecific hydrophobic interactions, which involve all three carbohydrate residues and extend along the length of the binding site.<sup>14</sup> The conserved water participates directly in the hydrogen bond network (Figure 1). This water molecule is tetrahedrally coordinated to the O2 of the central (reducing) sugar in 3MAN, and to the protein residues ASN 14, ASP 16, and ARG 228.<sup>14</sup> A water molecule in the same location is also observed in crystal structures of the unbound protein.<sup>21–23</sup> Earlier thermodynamic,<sup>24</sup> crystallographic,<sup>25</sup> and theoretical studies<sup>17,24–26</sup> explored the possibility of displacing this water by means of a synthetic analog of 3MAN, namely 3HET (Figure 1, panel b). In 3HET, a hydroxyethyl group replaces the hydroxyl group at the C2 position of the central mannosyl residue. ITC data<sup>24</sup> showed a more favorable entropy contribution for the binding of 3HET to Con A, relative to 3MAN, which was initially interpreted as indicating the displacement of the conserved water. A subsequent crystallographic and computational study<sup>25</sup> showed that the water was not displaced, although its position was slightly distorted relative to that in the Con A–3MAN complex (Figure 1). In the Con A–3HET complex, the conserved water forms a similar hydrogen bond network as present in the Con A–3MAN complex. In order to determine the molecular basis for the inability of the synthetic ligand to displace the water and to determine whether that particular water could reasonably be targeted for displacement in the first place, we have computed binding free energies for the water in the unbound protein, as well as in the protein bound to the natural and synthetic ligands. Additionally, extended (100 ns) unrestrained molecular dynamics (MD) simulations were run on all systems to investigate this water's residence times, and water–protein–ligand dynamics.

The interpretation of subtle changes in predicted free energy values places strenuous demands on the precision of the calculations. In this respect, equilibrium free energy calculation methods, such as thermodynamic integration (TI) used in this work, in combination with extensive sampling, are among the most reliable. In order to gauge the influence of different water models



**Figure 2.** Thermodynamic cycle used to derive standard binding free energies for the conserved water in the Con A saccharide binding site.  $P_{\text{sol}}$  indicates the protein (Con A) in solution without the conserved water molecule, and  $PW_{\text{sol}}$  indicates Con A with the bound water molecule in solution.  $W_{\text{sol}}$  indicates a water molecule in bulk water, and  $W_{\text{gas}}$  indicates a water molecule in the gas phase.  $\Delta G^P$  is the free energy required for removing a water from the protein binding site, and  $\Delta G^W$  is the hydration free energy.

on the binding free energy calculations, we focused on three popular nonpolarizable models: TIP3P,<sup>27</sup> TIP4P,<sup>27</sup> and TIP5P,<sup>28</sup> which employ three, four, or five sites to describe the electrostatic properties of water, respectively.

Our results demonstrate an unexpected high sensitivity to the water model used in free energy calculations, with the TIP5P water model being the only one that characterizes the water as bound both in the free protein and in all its complexes. Additional unrestrained MD simulations (100 ns) show that in the free protein TIP3P and TIP4P waters exchange with bulk water far more rapidly than TIP5P. Similar conclusions have been recently reported based on peptides simulations.<sup>29</sup> In addition, the TIP3P and TIP4P waters continue to exchange, albeit more slowly, even in the presence of the bound ligands, while for the duration of the simulations the same TIP5P water remains bound in both complexes. Notably, and in contradiction with the structural data,<sup>25</sup> the MD simulations with TIP3P and TIP4P show that the hydroxyethyl moiety in the 3HET ligand spontaneously displaces both bound waters. In contrast, the TIP5P water resists displacement by 3HET.

## MATERIALS AND METHODS

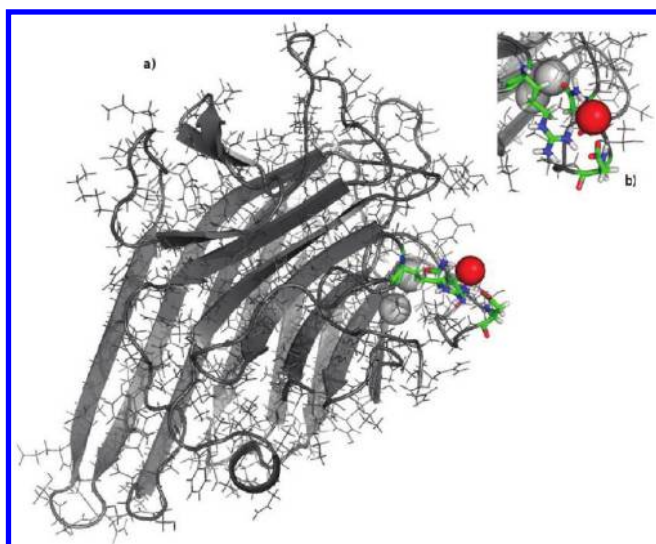
Standard binding free energies were determined by double decoupling (DD),<sup>30</sup> based on the diagram shown in Figure 2. In the DD method, the free energy for ligand binding  $\Delta G_{\text{bind}}^{\circ}$  is computed as the difference between the hydration free energies for the ligand in bulk water  $\Delta G_{\text{TI}}^W$  and the decoupling free energy of the ligand in the protein  $\Delta G_{\text{TI}}^P$ , where the term “decoupling” refers to the stepwise disappearance of the ligand, computed here using thermodynamic integration (TI):

$$\Delta G_{\text{bind}}^{\circ} = \Delta G_{\text{TI}}^W (\Delta G_{\text{TI}}^P + \Delta G_{\text{Corr}}) \quad (1)$$

$\Delta G_{\text{Corr}}$  is the standard state correction term required to correct for a standard concentration ( $C^{\circ}$ ) of 1 M, which in molecular simulations is more conveniently expressed as 1 molecule/1660  $\text{\AA}^3$ . In order to prevent the diffusion of the water out of the binding site during its annihilation, the oxygen of the water molecule is harmonically restrained throughout the calculations. The choice of the restraining potential follows the directives suggested in work by Hamelberg and McCammon.<sup>10</sup> This restraint restricts the volume sampled by the water, and it is accounted for in the correction term:<sup>30</sup>

$$\Delta G_{\text{Corr}} = -RT \frac{\sigma_{\text{PW}}}{\sigma_{\text{P}} \sigma_{\text{W}}} + RT [C^{\circ} (2\pi RT/k)^{3/2}] \quad (2)$$

where  $\sigma_{\text{PW}}$ ,  $\sigma_{\text{P}}$ , and  $\sigma_{\text{W}}$  are the symmetry numbers for the protein–water complex (PW), for the protein (P), and for the



**Figure 3.** (a) Con A (monomer; PDBid 1GKB, ref 40). (b) Detail of the saccharide binding site with the key residues, ASN 14, ASP 16, and ARG 228 highlighted. The structurally conserved water is represented by a red van der Waals sphere, the metal ions by two gray spheres.

water molecule (W), respectively, and  $k$  is the harmonic force constant ( $k = 2.4 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ ). The standard state correction term  $\Delta G_{\text{corr}}$  in our calculations amounts to  $-3.0 \text{ kcal mol}^{-1}$ . The same harmonic restraints were applied in all DD calculations.

The following crystal structures, indicated here by their PDBid, were chosen as starting conformations: 1CVN<sup>14</sup> for Con A bound to the trisaccharide 3MAN, 3D4K<sup>25</sup> for Con A bound to the synthetic analog 3HET, and 1GKB<sup>21</sup> for free Con A. At pH = 7.0, Con A exists both as a dimer and as a tetramer.<sup>31</sup> However, the ligand binding sites are independent of multimeric structure, and only one of the monomers was used in all simulations (see Figure 3). Two metal ions, a Mn(II) and a Ca(II), are required for Con A's structural stability and for ligand binding. Due to the lack of force field parameters for Mn(II), both metals were described as Ca(II). In addition to TIP3P, TIP4P, and TIP5P, the performance of a rarely used modified version of the TIP3P model, namely TIP3P-MOD,<sup>32</sup> was also assessed. The latter differs from TIP3P exclusively in the values of the Lennard-Jones (LJ) parameters. The results obtained are very similar to the ones obtained with the TIP3P model and are included as Supporting Information.

The following simulation protocol was followed in all cases. Crystal structures of the protein were inserted in an equilibrated cubic water box (80 Å cell length). The *genion* routine, included in the GROMACS software package, was used to add the three Na<sup>+</sup> counterions necessary to neutralize the system. The positions of the added water molecules, counterions, and hydrogen atoms were minimized with the steepest descent algorithm in 15 000 steps. An initial 500 ps equilibration at constant volume and temperature (the NVT at 300 K) was performed for the added water molecules, counterions, and hydrogen atoms. This step was followed by another 500 ps equilibration step at atmospheric constant pressure (NPT at 300 K). Protein, ligand, and conserved water were then equilibrated for 1 ns in the NPT ensemble, with the position of the protein C-alpha (C $\alpha$ ) atoms, heavy atoms of the ligand, and the oxygen (OW) of the conserved water restrained

harmonically with a force constant of  $2.4 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ . This equilibration step was followed by a final production step of 10 ns. The conformations obtained after the 10 ns MD simulation were used as the starting point for the free energy calculations. Thermodynamic integration (TI) in bulk water was performed with 20 lambda ( $\lambda$ ) points, while decoupling in the protein binding site involved integration over 37  $\lambda$  points. These steps were separated by 0.01 between 0.00 and 0.10 and between 0.90 and 1.00 and by 0.05 between 0.10 and 0.90. A finer separation around  $\lambda = 0$  and  $\lambda = 1$  was necessary to reduce the noise experienced at the beginning and end of the decoupling process and therefore to improve precision. In all cases, the Coulomb interactions were decoupled first, followed by the vdW interactions. At each  $\lambda$  value, an energy minimization was performed for 15 000 steps of steepest descent, followed by equilibration for 1 ns (NPT). Production was run for 1 ns. Tests with longer production runs, i.e., 5 ns per  $\lambda$  and even 10 ns per  $\lambda$ , did not show significant differences in the results. The final free energy values are based on a total simulation time of 72 ns for each system. During TI, harmonic restraints were applied to the positions of all heavy atoms in the ligand; the OW of the water; and heavy atoms of residues ASN 14, ASP 16, and ARG 228. In particular, ASP 16 and ARG 228 are not only directly hydrogen-bonded to the conserved water but also form a salt bridge. Restraining the salt bridge in a closed position, which corresponds to the crystallographic conformation in all structures available in the literature, was found to be necessary because its opening and closing greatly affects free energy values. Water molecules from the bulk can access the protein binding site as the conserved water is completely decoupled. In similar studies, a square-well potential was applied in order to prevent this from happening.<sup>5,11</sup> When Con A is bound to a ligand, either 3MAN or 3HET, the binding site is not accessible to bulk water, even when the conserved water molecule is annihilated. As the option of applying an external potential is not available in the software package we used (see below for details), in the case of free Con A, we confirmed that no bulk water diffused into the binding site during decoupling by analyzing the trajectories. In order to avoid end-point singularities, the vdW interactions were treated via soft-core interactions with a soft-core parameter  $\alpha = 0.5$ , an exponent of 1, and  $\sigma = 0.3 \text{ nm}$ . Integration was performed with the *scipy* module of python 2.6 (<http://www.python.org/>). For all MD simulations performed in the NPT ensemble, the temperature was held constant at 300 K by a Langevin thermostat with a coupling time constant of 0.1 ps. Pressure was held at 1 bar by exponential relaxation pressure coupling with a time constant of 0.5 ps. All MD simulations were performed with versions 4.0.3 up to 4.0.7 of the GROMACS software package.<sup>33</sup> The equations of motion were integrated using a leapfrog stochastic dynamics integrator with a 2 fs time step. Long-range electrostatics were treated with the particle mesh Ewald (PME) method.<sup>34,35</sup> The maximum spacing for the fast fourier transform grid was chosen as 1 Å. Cutoff values for Coulomb and vdW interactions were set to 0.9 nm. Tests with longer cutoffs values (12 Å) were performed, but no significant differences were observed. Water hydration free energies were calculated in a cubic water box of 25 Å per side. The calculated water densities at 300 K are  $0.985 \text{ g cm}^{-3}$  for TIP3P,  $0.993 \text{ g cm}^{-3}$  for TIP4P, and  $0.983 \text{ g cm}^{-3}$  for TIP5P. A separate set of MD simulations with restraints only on the C $\alpha$  and on the Ca<sup>2+</sup> ions bound to the protein were performed for



100 ns for all systems with all three water models for a total of 900 ns simulation time.

The AMBER99SB force field<sup>36</sup> was chosen to represent the protein, while both carbohydrate ligands were represented with the 06g release of the GLYCAM06 force field.<sup>37</sup>

## RESULTS AND DISCUSSION

**1. Water Hydration Free Energies.** The hydration free energies for each water model are shown in Table 1. The TIP5P hydration free energy is slightly lower compared not only to TIP3P and TIP4P but also to other popular water models.<sup>38</sup> The values reported in the literature range between an upper bound of  $-7.0$  kcal mol<sup>-1</sup> for the SPC/E model to a lower bound of  $-6.1$  kcal/mol<sup>38,39</sup> for TIP3P and TIP4P. The agreement between the TIP3P and TIP4P hydration free energy values here and the values reported in the literature confirm the accuracy of the TI protocol employed in this study. Variations in the electrostatics contribute most significantly to the differences in the hydration free energies between TIP3P, TIP4P, and TIP5P. As shown in Table 1, the free energy required to annihilate the electrostatic interactions is the same for TIP3P and TIP4P, i.e., 8.5 and 8.4 kcal mol<sup>-1</sup>, respectively. For TIP5P, the Coulomb contribution is weaker, i.e., 7.7 kcal mol<sup>-1</sup>. The van der Waals (vdW) contributions are similar and unfavorable for all water models, ranging from 2.0 to 2.3 kcal/mol. These contributions reflect the steric compression resulting from the strongly attractive electrostatic interactions in liquid water.

Electrostatic and Lennard-Jones parameters characterizing TIP3P, TIP4P, and TIP5P are shown in Table S.1 (Supporting Information).

**2. Water Binding Free Energy in Free Con A.** In free Con A, the conserved water is coordinated to the binding site's key residues, ASN 14, ASP 16, and ARG 228, through three hydrogen bonds. The fourth hydrogen bond is directed to a water molecule in the bulk (see Figures 3 and 4).<sup>40</sup> ASP 16 and ARG 228 form a salt bridge. Table S.2 (Supporting Information) shows the hydrogen bond distances in the crystal structure (1GKB) and in the free Con A snapshots used to start the free energy calculations. These snapshots were obtained from 10 ns MD

simulations, where the oxygen of the conserved water was harmonically restrained to its crystallographic position. Such restraint was necessary and sufficient to ensure that water's coordination was preserved, as well as the closed conformation of the salt bridge between ASP 16 and ARG 228. The coordination of TIP3P, TIP4P, and TIP5P in free Con A is shown in Figure 4. All water models are able to reproduce correctly the water coordination, with the only notable difference being that the salt bridge between the carboxylate oxygen in ASP 16 and the amino nitrogen in ARG 228 is shortened in the MD snapshots, and especially in the structures with TIP3P and TIP5P water, relative to the crystal structure.

The binding free energies for the conserved water in free Con A are shown in Table 2. The free energy cost for displacing the conserved water molecule from Con A's binding site is  $-2.3$  kcal mol<sup>-1</sup> for both TIP5P and TIP4P. This value is consistent with crystallographic evidence<sup>22,23,40</sup> indicating the presence of a water molecule in the binding site of the free protein. Conversely, the value of  $+0.1$  kcal mol<sup>-1</sup> obtained for TIP3P does not suggest occupancy of the binding site. The modification of the TIP3P Lennard-Jones parameters characterizing the TIP3P-MOD model does not correct the TIP3P behavior, i.e.,  $-0.3$  kcal mol<sup>-1</sup> (see Table S.4 in Supporting Information).

### 3. Water Binding Free Energy in the Con A–3MAN Complex.

As in free Con A, the conserved water molecule forms four hydrogen bonds also in the binding site of the Con A–3MAN complex. The crystal structure 1CVN indicates that three hydrogen bonds connect the water to the key residues, ASN 14, ASP 16, and ARG 228, while the fourth hydrogen bond is directed to the O2 of the central (reducing) mannosyl residue of the ligand. This water coordination is generally well preserved in all of the structures used as a starting point for the free energy calculations, with only small differences in the hydrogen bonding coordination pattern (see Figure 5). Specifically, while one lone pair in TIP5P forms a hydrogen bond with the ARG 228 imino group ( $-NH-$ ), both TIP3P and TIP4P interact with ARG 228 through one of the two amino group protons ( $-NH_2$ ; see Figure 5 and Table S.2, Supporting Information, for hydrogen bond distances). The presence of the two lone pairs in TIP5P clearly facilitates the formation of linear hydrogen bonds with

**Table 1. Hydration Free Energies<sup>a</sup> ( $\Delta G^\circ$ ) and Individual Contributions Obtained for the TIP3P, TIP4P, and TIP5P Water Models**

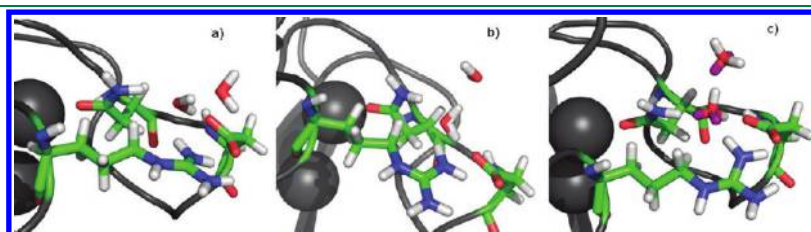
water model	Coulomb	vdW	$\Delta G^\circ$	literature
TIP3P	-8.5	2.2	-6.3 (0.1) <sup>b</sup>	-6.5, <sup>39</sup> -6.1 <sup>38</sup>
TIP4P	-8.4	2.3	-6.1 (0.2)	-6.1 <sup>38</sup>
TIP5P	-7.7	2.0	-5.7 (0.1)	

<sup>a</sup> All energies are in kcal mol<sup>-1</sup>. <sup>b</sup> Errors, calculated by block averaging, are shown in parentheses.

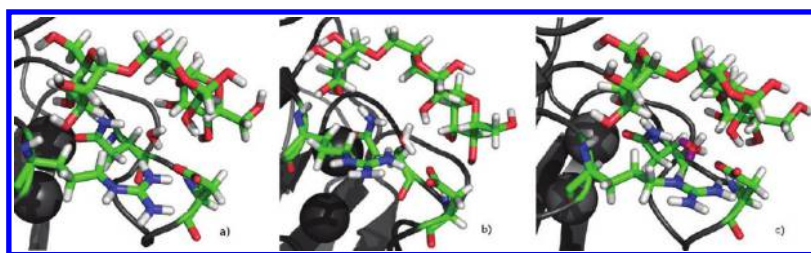
**Table 2. Standard Binding Free Energies<sup>a</sup> ( $\Delta G^\circ$ ) and Individual Contributions for the Conserved Water in the Binding Site of Free Con A**

free Con A	Coulomb	vdW	$\Delta G_p$	$\Delta G^\circ$
TIP3P	-14.9	5.7	-6.2	0.1 (0.1) <sup>b</sup>
TIP4P	-13.7	2.3	-8.4	-2.3 (0.1)
TIP5P	-15.5	4.5	-8.0	-2.3 (0.2)

<sup>a</sup> All energies are in kcal mol<sup>-1</sup>. <sup>b</sup> Errors, calculated by block averaging, are shown in parentheses.  $\Delta G_p$  includes a standard state correction term of 3.0 kcal mol<sup>-1</sup>.



**Figure 4.** Coordination of TIP3P, TIP4P, and TIP5P in free Con A



**Figure 5.** (a) Detail of the binding site in free Con A from the MD simulation with a TIP3P water representing the structurally conserved water. (b) Detail of the binding site in free Con A from the MD simulation with a TIP5P water representing the structurally conserved water. The TIP5P lone pairs are highlighted in magenta.

**Table 3. Standard Binding Free Energies<sup>a</sup> ( $\Delta G^\circ$ ) and Individual Contributions for the Conserved Water in the Binding Site of Con A in Complex with 3MAN**

Con A–3MAN	Coulomb	vdW	$\Delta G_p$	$\Delta G^\circ$
TIP3P	−21.7	11.4	−7.3	−1.0 (0.2) <sup>b</sup>
TIP4P	−21.5	10.1	−8.4	−2.3 (0.3)
TIP5P	−21.1	5.3	−12.8	−7.1 (0.1)

<sup>a</sup>All energies are in kcal mol<sup>−1</sup>. <sup>b</sup>Errors, calculated by block averaging, are shown in parentheses.  $\Delta G_p$  includes a standard state correction term of 3.0 kcal mol<sup>−1</sup>.

all residues. Conversely, the hydrogen bond between the TIP3P and TIP4P oxygen and the  $-\text{NH}_2$  of ARG 228 is not as linear.

As shown in Table 3, in the Con A–3MAN complex, the TIP5P water has a much higher binding free energy value of  $-7.1$  kcal/mol relative to free Con A. High standard binding free energy values for conserved water molecules are not unusual,<sup>11</sup> and they suggest a structural role for water molecules in ligand binding. The values obtained for TIP3P and TIP4P are much lower relative to TIP5P, i.e.  $-1.0$  kcal/mol and  $-2.3$  kcal/mol, respectively. As shown in Table 3, the steric compression, accounted for by the vdW decoupling free energy contributions, is much higher for TIP3P and TIP4P waters relative to TIP5P. Interestingly, the binding free energy value obtained for TIP4P is equal to the value calculated for the TIP4P water in the free protein, suggesting that interchanging bulk water molecules are as effective in coordinating the conserved water as is the O2 in 3MAN, that is, suggesting that the ligand does not enhance the stability of the bound water.

#### 4. Water Binding Free Energy in the Con A–3HET Complex.

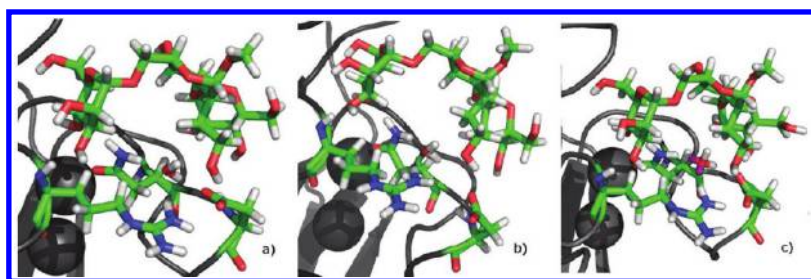
Crystallographic data show that the position of the water in the Con A–3HET complex is only slightly distorted relative to the one in the Con A–3MAN complex.<sup>25</sup> In the crystal structure of the Con A–3HET complex (3D4K), the conserved water is hydrogen-bonded to the same protein residues as it is in the Con A–3MAN complex, but it interacts with the O8 of the hydroxyethyl group instead of the O2 of the central mannosyl residue of 3MAN (see Figures 1 and 6). The conserved water coordination is reproduced correctly by all water models with a few slight differences. As shown in Figure 6 a and b, the coordinations of TIP3P and TIP4P are very similar with direct hydrogen bonds with ASN 14, ASP 16, ARG 228, and the O8 of the 3HET hydroxyethyl. In all structures of the Con A–3HET complex including the crystal structure, the hydrogen bond to the O8 of 3HET is slightly elongated relative to the hydrogen bond to O2 of 3MAN, suggesting a weaker interaction between

the ligand and the binding site (see Table S.2, Supporting Information).

Binding free energies for the water in the Con A–3HET complex are shown in Table 4. The binding free energy for a TIP3P water in the Con A–3HET complex is  $-4.8$  kcal/mol, making TIP3P more tightly bound when Con A is in complex with the synthetic ligand relative to the complex with the natural ligand. The binding free energy obtained for the TIP4P water is  $+0.2$  kcal mol<sup>−1</sup>, which indicates no propensity for the binding site to be hydrated. The binding free energy of a TIP5P water in the Con A–3HET complex is  $1.4$  kcal mol<sup>−1</sup> lower than the binding free energy in the Con A–3MAN complex, i.e.,  $-5.7$  kcal mol<sup>−1</sup>. This value indicates that the water is still firmly bound within the binding site but in a less favorable location relative to the Con A–3MAN complex. Indeed, while all water models make stronger electrostatic contacts, in the Con A–3MAN complex, the steric compression of TIP3P and TIP4P is much higher than for TIP5P (see Tables 3 and 4).

**5. Molecular Dynamics Simulations.** The stability of the conserved water molecule and of the hydrogen bond network connecting it to the ligands and to the protein binding site was monitored for each water model throughout 100 ns of unrestrained MD simulations.

As seen in Figure 4, in the free protein, the conserved water is hydrogen-bonded to three key residues (ASN 14, ASP 16, and ARG 228) and to a water molecule from the bulk. The average hydrogen bond distances calculated over the 100 ns trajectories are shown in Table 5. The dynamics of the residues in the binding site and especially of the salt bridge between ASP 16 and ARG 228 promote the exchange between water molecules from the bulk and the water occupying the binding site. This exchange always occurs between the conserved water and the water hydrogen-bonded to it in a concerted manner so that the binding site never remains vacant. The exchange rate of the water molecules depends both on the water coordination and on the water model chosen. More specifically, as the salt bridge opens, the water loses one hydrogen bond either to ASP 16 or to ARG 228, and it is much more likely to exit the binding site. Most water exchanges are triggered by the opening of the salt bridge. As a complete kinetic analysis of the water exchange rates is beyond the scope of this work, we determined water molecule residence times on the basis of 100 ns of the lifetime of the hydrogen bond network between a specific water and the three key residues, ASN 14, ASP 16, and ARG 228. Residence times were calculated on the basis of 41 exchange events for TIP3P, 33 events for TIP4P, and 26 events for TIP5P. In order to calculate these values, the trajectories were extended from 100 ns up to 140 ns. In the binding site of free Con A, TIP5P has the longest residence time of 2.3 ns, while TIP3P and TIP4P residence times are 1.3 and 1.6 ns, respectively.



**Figure 6.** (a) Detail of the binding site in the Con A–3MAN complex from the MD simulation with a TIP3P water representing the structurally conserved water. (b) Detail of the binding site in the Con A–3MAN complex from the MD simulation with a TIP4P water representing the structurally conserved water. (c) Detail of the binding site in the Con A–3MAN complex from the MD simulation with a TIP5P water representing the structurally conserved water. The TIP5P lone pairs are highlighted in magenta.

**Table 4. Standard Binding Free Energies<sup>a</sup> ( $\Delta G^\circ$ ) and Individual Contributions for the Conserved Water in the Binding Site of Con A in Complex with 3HET**

Con A–3HET	Coulomb	vdW	$\Delta G_p$	$\Delta G^\circ$
TIP3P	−18.7	4.6	−11.1	−4.8 (0.1) <sup>b</sup>
TIP4P	−17.0	8.1	−5.9	+0.2 (0.4)
TIP5P	−19.0	4.6	−11.4	−5.7 (0.2)

<sup>a</sup>All energies are in kcal mol<sup>−1</sup>. <sup>b</sup>Errors, calculated by block averaging, are shown in parentheses.  $\Delta G_p$  includes a standard state correction term of 3.0 kcal mol<sup>−1</sup>.

**Table 5. Hydrogen Bond Distances<sup>b</sup> between the Conserved Water and the Key Residues in the Binding Site of Free Con A<sup>a</sup>**

H-bond distances	1GKB	TIP3P	TIP4P	TIP5P
O <sub>WAT</sub> –O <sub>ASP16</sub>	2.6	2.7 (0.3) <sup>c</sup>	2.7 (0.2)	2.8 (0.3)
O <sub>WAT</sub> –N <sub>ASN14</sub>	2.9	3.1 (0.3)	3.0 (0.2)	2.9 (0.2)
O <sub>WAT</sub> –N <sub>ARG228</sub>	3.0	3.0 (0.2)	2.9 (0.2)	2.8 (0.1)

<sup>a</sup>Data from the crystal structure (1GKB) are compared to average distances calculated during the unrestrained MD simulations. <sup>b</sup>In Ångstroms. <sup>c</sup>Standard deviations are indicated in parentheses.

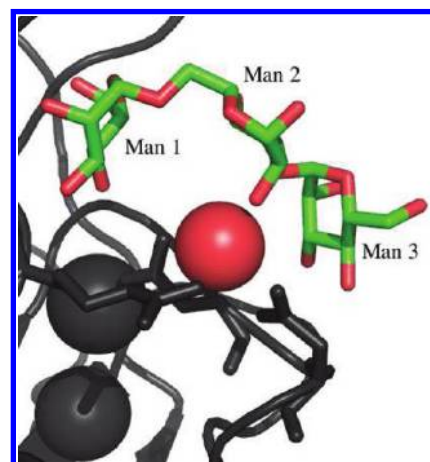
The MD simulation of the Con A–3MAN complex shows that the presence of the trisaccharide limits the mobility of the water, reducing drastically the number of exchange events between the conserved water and water in the bulk. During the 100 ns simulations, a few exchange events are observed for both TIP3P and TIP4P water, but not for TIP5P, where the same water molecule occupied the binding site for the whole 100 ns. Average hydrogen bond distances are shown in Table 6. 3MAN is tightly bound to the protein surface throughout all simulations, and it shows a similar degree of mobility for all water models (RMSD values in Table S.5 and Figure S.1, Supporting Information). The terminal sugar residue more exposed to the bulk, namely, Man 3 (see Figure 7), shows a slightly higher degree of mobility relative to the others, especially in TIP4P. The slightly higher RMSD calculated for 3MAN in TIP4P is due to a partial detachment of the Man 3 residue at around 10 ns triggered by water exchange.

A partial detachment of the ligand from the protein was observed during the MD simulation of the Con A–3HET complex in TIP4P. As shown by the RMSD plot in Figure 8, between 15 and 21 ns, 3HET is highly mobile. During this time interval, 3HET remains bound to the protein only via Man 1. The partial detachment is caused by the conserved water

**Table 6. Hydrogen Bond Distances<sup>a</sup> between the Conserved Water and the Key Residues in the Binding Site of the Con A–3MAN Complex**

H-bond distances	1CVN	TIP3P	TIP4P	TIP5P <sup>b</sup>
O <sub>WAT</sub> –O <sub>ASP16</sub>	2.8	2.6 (0.1) <sup>c</sup>	2.6 (0.1)	2.8 (0.2)
O <sub>WAT</sub> –N <sub>ASN14</sub>	2.7	2.9 (0.1)	3.0 (0.2)	2.8 (0.1)
O <sub>WAT</sub> –N <sub>ARG228</sub>	3.1	3.0 (0.2)	3.0 (0.2)	3.0 (0.3); 3.5(0.3)
O <sub>WAT</sub> –O <sub>23MAN</sub>	2.4	2.8 (0.2)	2.9 (0.2)	2.9 (0.2)

<sup>a</sup>In Ångstroms. <sup>b</sup>For TIP5P, the first O<sub>WAT</sub>–N<sub>ARG 228</sub> value corresponds to the distance to the imino nitrogen of ARG 228, while the second one corresponds to the amino nitrogen. <sup>c</sup>Standard deviations are indicated in parentheses.



**Figure 7.** Residues nomenclature used for 3MAN. The conserved water is shown in red. The structure shown has PDBid 1CVN.<sup>14</sup>

leaving the binding site and not being immediately replaced. As a new water molecules enters the binding site at 21 ns, the 3HET returns to its bound conformation, where it stays until the end of the simulation. RMSD values for the ligand calculated over the 100 ns simulation are shown in Table S.6 (Supporting Information).

An interesting result obtained from the MD simulations on the Con A–3HET complex is that the hydroxyethyl group is able to displace both the TIP3P and the TIP4P water within 100 ns. In the MD simulation with TIP3P water, the water displacement occurs at 53 ns, while in the simulation with TIP4P it occurs at 50 ns. The OH group of the hydroxyethyl forms stable hydrogen bonds to the three binding site key residues



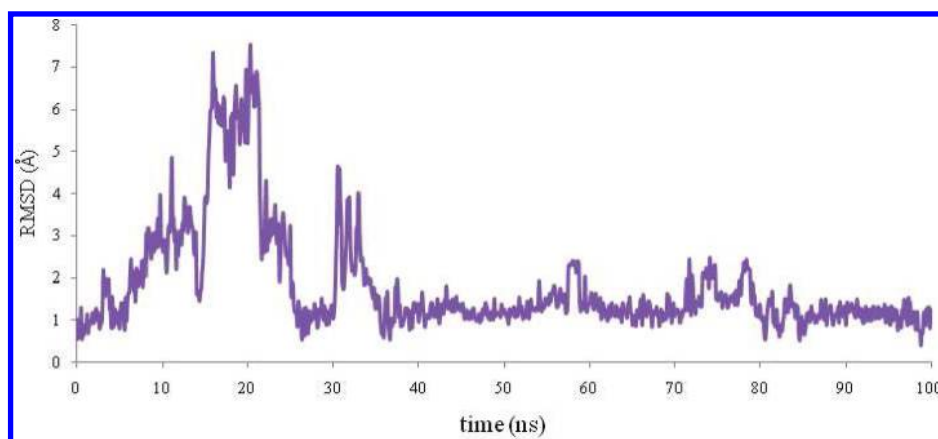


Figure 8. RMSD values for the water oxygen calculated over 100 ns of unrestrained MD simulations of the Con A–3HET complex in TIP4P water.

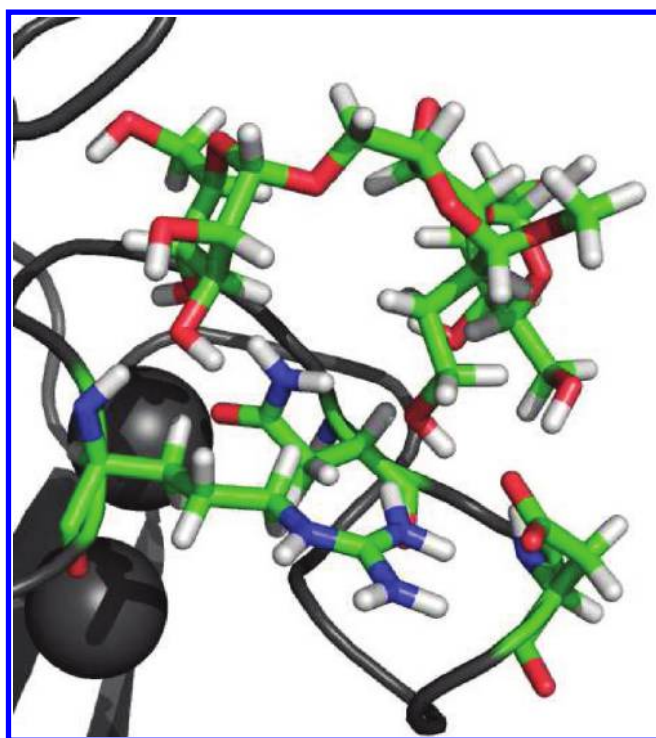


Figure 9. Hydroxyethyl of 3HET occupying Con A's binding site after the displacement of the conserved water. The conformation was obtained from the 100 ns of unrestrained MD simulation of the Con A–3HET complex in TIP3P water.

(see Figure 9), and it occupies the location of the conserved water for the remainder of the simulation. The average hydrogen bond distances are shown in Table 8. A comparison between the data in Tables 7 and 8 shows that the hydroxyethyl OH is coordinated exactly like the conserved water in both the simulation with TIP3P and that with TIP4P.

As seen in the simulation of the Con A–3MAN complex with TIP5P, in the simulation of the Con A–3HET complex with TIP5P, no exchanges between conserved water and the bulk were observed. Additionally, within the 100 ns of unrestrained simulation, the hydroxyethyl group was unable to displace the conserved water (see Table 7 for hydrogen bond distances).

Table 7. Hydrogen Bond Distances<sup>a</sup> between the Conserved Water and the Key Residues in the Binding Site of the Con A–3HET Complex

H-bond distances	3D4K	TIP3P	TIP4P	TIP5P
O <sub>WAT</sub> –O <sub>ASP16</sub>	2.5	2.7 (0.3) <sup>b</sup>	3.0 (0.6)	2.7 (0.1)
O <sub>WAT</sub> –N <sub>ASN14</sub>	2.7	3.0 (0.2)	3.0 (0.2)	2.8 (0.1)
O <sub>WAT</sub> –N <sub>ARG228</sub>	3.0	3.0 (0.3)	3.0 (0.4)	2.9 (0.2)
O <sub>WAT</sub> –O <sub>2HET</sub>	3.0	3.1 (0.5)	3.3 (0.9)	4.3 (0.9)

<sup>a</sup> In Ångstroms. <sup>b</sup> Standard deviations are indicated in parentheses.

Table 8. Hydrogen Bond Distances between<sup>a</sup> the Terminal OH of the Hydroxyethyl Group of 3HET and the Key Residues in the Binding Site of the Con A–3HET Complex<sup>c</sup>

H-bond distances	TIP3P	TIP4P
OH <sub>3HET</sub> –O <sub>ASP16</sub>	2.7 (0.2) <sup>b</sup>	2.8 (0.3)
OH <sub>3HET</sub> –N <sub>ASN14</sub>	3.0 (0.2)	3.0 (0.2)
OH <sub>3HET</sub> –N <sub>ARG228</sub>	3.0 (0.2)	3.0 (0.2)

<sup>a</sup> In Ångstroms. <sup>b</sup> Standard deviations are indicated in parentheses.

<sup>c</sup> Average distances are calculated during the unrestrained MD simulations in TIP3P and TIP4P water only after the conserved water is displaced. TIP5P water was not displaced within the 100 ns trajectory.

## CONCLUSIONS

In this work, we analyzed the function and stability of a structurally conserved water molecule in the binding site of the carbohydrate binding protein Con A by means of binding free energy calculations and MD simulations. The calculations were performed for three systems: the free protein, the protein bound to its natural ligand (3MAN), and the protein bound to a synthetic analog (3HET) that was designed and synthesized in previous work to displace the conserved water.<sup>24,25</sup> Subsequent structural data<sup>25</sup> showed that 3HET was not in fact successful in displacing the key water. To provide insight into the unexpected resistance of this water to displacement, its binding affinity was estimated employing three rigid nonpolarizable water models, namely, TIP3P, TIP4P, and TIP5P. The water binding free energy values, the dynamics, and the stability of the water in the binding site were all found to depend significantly on the choice of water model. On the basis of the free energy values and the dynamics, in the free protein, TIP3P clearly underestimates the

level of order and occupancy of the binding site by the conserved water, while both TIP3P and TIP4P correctly predict that the binding site is specifically hydrated.

In the complex with 3MAN, the TIP3P water is tightly bound and optimally coordinated with the three key amino acid residues in the Con A binding site and the hydroxyl group of the central mannose. In contrast, both TIP3P and TIP4P are less strongly bound and more labile during the simulations.

In the Con A–3HET complex, both TIP3P and TIP4P escape from the binding site to be replaced by the hydroxyl group of the hydroxyethyl moiety. This direct interaction between the synthetic ligand and the Con A binding site was predicted from earlier studies with TIP3P<sup>23</sup> but was refuted by subsequent crystallographic data.<sup>25</sup> Only the TIP3P model maintains its position in the 3HET complex throughout the 100 ns of simulation, although it is less strongly bound than in the complex with the natural ligand.

Although the MD simulations with TIP3P and TIP4P suggest that the displacement of the conserved water by a synthetic ligand is an easier task than it apparently is, the length of the hydroxyethyl group does indeed fit this requirement. However, the flexibility of the hydroxyethyl moiety in 3HET permits it to adopt a conformation that enables 3HET to bind without displacing the water. A modification of the hydroxyethyl group that enhanced the stability of the C2–C7 torsion in a  $g^+$  conformation could conceivably displace this water, provided that the net ligand binding free energy exceeded  $-2.4$  kcal mol<sup>-1</sup>, the binding free energy computed with TIP3P for the conserved water in free Con A.

## ■ ASSOCIATED CONTENT

**S Supporting Information.** The results obtained with the TIP3P-MOD water model together with additional figures and tables have been included. This information is available free of charge via the Internet at <http://pubs.acs.org>.

## ■ AUTHOR INFORMATION

### Corresponding Author

\*Tel.: +1 (706) 542 4454. Fax: +1 (706) 542 4412. E-mail: [rwoods@ccrc.uga.edu](mailto:rwoods@ccrc.uga.edu)

## ■ ACKNOWLEDGMENT

E.F. thanks Chris Neale and Régis Pomès for insightful discussions. The Science Foundation of Ireland (SFI operating grant 08/IN.1/B2070) is gratefully acknowledged for funding.

## ■ REFERENCES

- (1) Dwek, R. A. *Chem. Rev.* **1996**, *96*, 683–720.
- (2) Rini, J. M. *Annu. Rev. Biophys. Biomol. Struct.* **1995**, *24*, 551–577.
- (3) Lis, H.; Sharon, N. *Chem. Rev.* **1998**, *98*, 637–674.
- (4) de Beer, S. B. A.; Vermeulen, N. P. E.; Oostenbrink, C. *Curr. Top. Med. Chem.* **2009**, *10*, 55–66.
- (5) Michel, J.; Tirado-Rives, J.; Jorgensen, W. L. *J. Am. Chem. Soc.* **2009**, *131*, 15403–15411.
- (6) Poornima, C. S.; Dean, P. M. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 500–512.
- (7) Homans, S. W. *Drug Discovery Today* **2007**, *12*, 534–539.
- (8) Baron, R.; Setny, P.; McCammon, J. A. *J. Am. Chem. Soc.* **2010**, *132*, 12091–12097.

- (9) Setny, P.; Baron, R.; McCammon, J. A. *J. Chem. Theory Comput.* **2010**, *6*, 2866–2871.
- (10) Hamelberg, D.; McCammon, J. A. *J. Am. Chem. Soc.* **2004**, *126*, 7683–7689.
- (11) Barillari, C.; Taylor, J.; Viner, R.; Essex, J. W. *J. Am. Chem. Soc.* **2007**, *129*, 2577–2587.
- (12) Gupta, D.; Dam, T. K.; Oscarson, S.; Brewer, C. F. *J. Biol. Chem.* **1997**, *272*, 6388–6392.
- (13) Brewer, C. F.; Bhattacharyya, L. *J. Biol. Chem.* **1986**, *261*, 7306–7310.
- (14) Naismith, J. H.; Field, R. A. *J. Biol. Chem.* **1996**, *271*, 972–976.
- (15) Bhattacharyya, L.; Brewer, C. F. *Eur. J. Biochem.* **1989**, *178*, 721–726.
- (16) Chervenak, M.; Toone, E. J. *Biochemistry* **1995**, *34*, 5685–5695.
- (17) Bryce, R. A.; Hillier, I. H.; Naismith, J. H. *Biophys. J.* **2001**, *81*, 1373–1388.
- (18) Bouckaert, J.; Hamelryck, T. W.; Wyns, L.; Loris, R. *J. Biol. Chem.* **1999**, *274*, 29188–29195.
- (19) Dam, T. K.; Oscarson, S.; Sacchetti, J. C.; Brewer, C. F. *J. Biol. Chem.* **1998**, *273*, 32826–32.
- (20) Mandal, D. K.; Kishore, N.; Brewer, C. F. *Biochemistry* **1994**, *33*, 1149–1156.
- (21) Katardjieff, K.; Höchtl, P.; Segelke, B. W.; Tao, F.-M.; Rupp, B. *Acta Crystallogr.* **2002**, *58*, 735–743.
- (22) Deacon, A.; Gleichmann, T.; Kalb (Gilboa), A. J.; Price, H.; Raftery, J.; Bradbrook, G.; Yariv, J.; Helliwell, J. R. *J. Chem. Soc., Faraday Trans.* **1997**, *93*, 4305–4312.
- (23) Parkin, S.; Rupp, B.; Hope, H. *Acta Crystallogr.* **1996**, *DS2*, 1161–1168.
- (24) Clarke, C.; Woods, R. J.; Gluska, J.; Cooper, A.; Nutley, M. A.; Boons, G. J. *J. Am. Chem. Soc.* **2001**, *123*, 12238–12247.
- (25) Kadirvelraj, R.; Foley, B. L.; Dyekjaer, J. D.; Woods, R. J. *J. Am. Chem. Soc.* **2008**, *130*, 16933–16942.
- (26) Li, Z.; Lazaridis, T. *J. Phys. Chem. B* **2005**, *109*, 662–670.
- (27) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (28) Mahoney, M. W.; Jorgensen, W. L. *J. Chem. Phys.* **2000**, *112*, 8910–8922.
- (29) Florova, P.; Sklenovsky, P.; Banas, P.; Otyepka, M. *J. Chem. Theory Comput.* **2010**, *6*, 3569–3579.
- (30) Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. *Biophys. J.* **1997**, *72*, 1047–1069.
- (31) McKenzie, G. H.; Sawyer, W. H. *J. Biol. Chem.* **1973**, *248*, 549–556.
- (32) Sun, Y. X.; Kollman, P. A. *J. Comput. Chem.* **1995**, *16*, 1164–1169.
- (33) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.
- (34) York, D. M.; Darden, T. A.; Pedersen, L. G. *J. Chem. Phys.* **1993**, *99*, 8345–8348.
- (35) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577–8593.
- (36) Simmerling, C.; Strockbine, B.; Roitberg, A. E. *J. Am. Chem. Soc.* **2002**, *124*, 11258–11259.
- (37) 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*, 622–655.
- (38) Shirts, M. R.; Pande, V. S. *J. Chem. Phys.* **2005**, *122*, 13.
- (39) Olano, L. R.; Rick, S. W. *J. Am. Chem. Soc.* **2004**, *126*, 7991–8000.
- (40) Kantardjieff, K. A.; Hochtl, P.; Segelke, B. W.; Tao, F. M.; Rupp, B. *Acta Crystallogr.* **2002**, *58*, 735–743.