THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

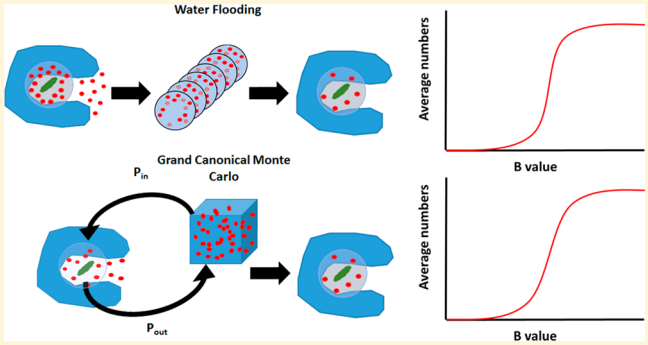
Validating the Water Flooding Approach by Comparing It to Grand Canonical Monte Carlo Simulations

Hanwool Yoon, Vesselin Kolev, and Arieh Warshel*

Department of Chemistry, University of Southern California, 418 SGM Building, 3620 McClintock Avenue, Los Angeles, California 90089-1062, United States

Supporting Information

ABSTRACT: The study of the function of proteins on a quantitative level requires consideration of the water molecules in and around the protein. This requirement presents a major computational challenge due to the fact that the insertion of water molecules can have a very high activation barrier and would require a long simulation time. Recently, we developed a water flooding (WF) approach which is based on a postprocessing Monte Carlo ranking of possible water configurations. This approach appears to provide a very effective way for assessing the insertion free energies and determining the most likely configurations of the internal water molecules. Although the WF approach was used effectively in modeling challenging systems that have not been addressed reliably by other microscopic approaches, it was not validated by a comparison to the more rigorous grand canonical Monte Carlo (GCMC) method. Here we validate the WF approach by comparing its performance to that of the GCMC method. It is found that the WF approach reproduces the GCMC results in well-defined test cases but does so much faster. This established the WF approach as a useful strategy for finding correct water configurations in proteins and thus to provide a powerful way for studies of the functions of proteins.



I. INTRODUCTION

It is now widely recognized[1] that water molecules inside and around biological molecules play a major role in determining key functions. It is also appreciated that the effect of internal water molecules must be considered in evaluation of the energetics of biological processes, ranging from enzymatic reactions,[2] ligand binding,[3−6] redox reactions,[7] ion conduction,[8] proton transport,[9,10] and ionization of deeply buried protein residues.[11,12] The protein dipole Langevin dipole (PDLD) model has been arguably the first attempt to represent the water in and around the protein in an explicit yet simplified way in studies of the energetics of biological processes.[13−15] This model allowed one to consider water penetration but focused on the overall effective energetics rather than on reproducing the possible exact hydrogen bonding pattern. The earliest attempts to incorporate water molecules in atomistic free energy calculations of ionized groups in proteins were reported in ref 16, using the surface constrained all-atom solvent (SCAAS) model.[17,18] Different adaptations of the SCAAS and its earlier SCSSD version[19] boundary conditions ideas have emerged and been applied by other groups (e.g., ref 20) and eventually have used proper electrostatic boundaries.[21,22] Continuum models were also developed as a powerful tool of representing the effect of the external water molecules, but attempts to consider in such models a few water molecules implicitly around the protein have not been so successful (see discussion in ref 23).

Despite the effectiveness of the PDLD, the advances in computer power led to subsequent different atomistic models and to detailed studies of the energetics of many biological processes (e.g., ref 15). However, notable problems emerged in some studies of relevant systems. One of the most serious problems has been the overestimate of the solvation penalty for moving charges to nonpolar sites in proteins where, for example, if we consider the important benchmark of ionizable residues provided by Garcia-Moreno and co-workers,[11,12,24] we can obtain major overestimates of the penalty of moving the given charges from bulk water to the protein sites by simple free energy perturbation (FEP) microscopic calculations (see refs 9 and 25). This problem could be reduced by using a semimacroscopic model such as the PDLD/S-LRA, with a relatively high dielectric for the charging energy in problematic sites (around 6−8).[25] Furthermore a way forward has been offered by the overcharging approach that induced a partial unfolding of the protein by artificially increasing the solute charge and forcing accelerated water penetration,[25] but this method requires major computer time.

Received: August 3, 2017
Revised: September 10, 2017
Published: September 14, 2017

DOI: 10.1021/acs.jpcb.7b07726
J. Phys. Chem. B 2017, 121, 9358−9365

Another area where the water positioning issue has become a critical problem is the study of ligand binding. Here it has been noted (e.g., refs 3 and 26−28) that performing free energy perturbation (FEP) or linear response approximation (LRA) may not allow for proper water equilibration. Obviously, with an infinite amount of simulation time, it is possible to obtain the proper water configurations, but this may not be practical. Attempts to reach a formal rigor in water insertion in protein cavities have focused on grand canonical Monte Carlo (GCMC) methods,[29,30] as reported in several studies.[31−33,27,28] However, while such methods seem attractive, they require major computational resources, as the acceptance of the insertion/deletion attempts is very low. The search for a faster approximated approach led ref 4 to develop a method termed "just add water molecules" (JAWS) where water molecules can appear/disappear from the system based on the occupancy of different sites. This method also involves a refinement using explicit MC moves. However, the above method has been focused on reproducing the water sites observed in high resolution crystal structures, rather than the energetics of the penetration of clusters of water molecules.

A clustering approach (WaterMap) was introduced in refs 5 and 34, and the introduction of empirical scaling seems to improve the results of binding calculations. We would like to point out, however, that the implication that this approach leads to enormous improvement in calculations of ligand binding (e.g., ref 35) are unjustified. That is, a proper validation of the pure effect of better placement of water molecules must include comparison of the use of the given free energy sampling approach with water positions evaluated by a standard water placing method and the WaterMap approach. Such a comparison[36] that was obtained by the JAWS approach produced a moderate improvement.

Recent interesting studies were reported by Essex and co-workers[27,28] who compared several methods and in particular advanced actual GCMC, trying to determine the energetics of water insertion in proteins. These studies gave useful insight and demonstrated that GCMC can be used with sufficient investment of computer time, although special care is needed in attempts to reach convergence.

Despite the potential of the above approaches, they were not validated by comparing the calculated energetics of electrostatic calculations (e.g., pK_a calculations) with and without the corresponding water insertion approach. This is important, since in cases of internal charges the effect of water molecules can be very large and this should be useful in assessing the error range.

Considering the challenges of developing both effective and reliable methods for water insertion, we introduced recently the water flooding (WF) approach.[26] In this approach, we determine the energy of rationally inserted water molecules using the LRA and the linear interaction energy (LIE) approaches and then sort the energetics by a postprocessing MC approach. The postprocessing strategy makes the method much faster than any GCMC, since it avoids the need to perform any explicit MC simulation during the energy evaluation.

Despite the impressive results obtained with the WF approach,[26,37] we feel that this strategy should be further validated by comparing it to the more rigorous GCMC approach. Thus, we conduct in this work a systematic comparison between the results of the WF and GCMC approaches in the evaluation of water penetration to proteins.

II. METHODS

II.a. Grand Canonical Monte Carlo. Our GCMC model generated the usual SCAAS model for the protein and the surrounding solvent and selected a central point in the protein region of the study, using a radial boundary that prevented movement of water in or out of a specific radius. The subsequent step started by inserting and deleting water from the inner region in the following way. Any insertion attempt was followed by finding 20 orientations of the inserted water and then choosing the one with the lowest potential ($U_{in,min}$) and then accepting or rejecting the move by

$$P_{in} = \min[1, (1/(N+1)) \exp(-(U_{in,min}\beta - B))] \quad (1)$$

where P_{in} is the probability of acceptance of an insertion move and $B = (\Delta G_{bulk}\beta + \ln\langle N\rangle)$, N is the total number of water molecules in the system, $\langle N\rangle$ is the average N for the given B, and ΔG_{bulk} is the assumed free energy (more precisely chemical potential) of a water molecule in the bulk. The selection of a single orientation was a simplification of previous approaches that used the Boltzmann average.[21,38] The condition of eq 1 was satisfied by a standard Metropolis approach.[39] Our GCMC method was implemented in the MOLARIS-XG program package.[40]

The deletion moves were accepted by evaluating the potential U_{out} of a random water molecule in the inner region, and then applying the criterion

$$P_{out} = \min[1, N \exp(+(U_{out}\beta - B))] \quad (2)$$

where U_{out} is the potential energy of a water molecule that is a candidate for removal.

Each insertion or deletion attempt was followed by 10 MD steps of the entire system that establishes the equilibration. Following ref 27, we performed the GCMC steps in a subregion of the complete system (the equivalent of the box in ref 27), surrounded by a spherical wall of a specified radius.

II.b. The Water Flooding Model. As stated in the Introduction, we developed in our previous report[26] a WF approach which drastically accelerates the insertion process by not performing any explicit MC water insertion but rather using a postprocessing MC strategy. While the details of the method are described in ref 41, we give below the key details.

The WF method starts by generating different configurations with an excess number of internal water molecules, deleting those that collide with the protein, and then evaluating the electrostatic energies of these molecules using the linear response approximation (LRA). This approximation estimates the free energy of each configuration of the internal water molecules by

$$\Delta G_{(\mathbf{m})} = \sum_{i(\mathbf{m}(p))} (\Delta G_i^p)\delta_i(\mathbf{m}) + \sum_{i(\mathbf{m}(p))} \sum_{j(\mathbf{m}(p))<i(\mathbf{m}(p))} \Delta G_{ij}^p \delta_i(\mathbf{m})\delta_j(\mathbf{m}) \quad (3)$$

where $\mathbf{m}$ is a configuration vector that runs over all of the water sites and $\delta_i(\mathbf{m})$ is a function that represents the occupancy of the water sites in the current $\mathbf{m}$th configuration. $\delta_i(\mathbf{m})$ is 1 when the ith site is occupied and zero otherwise. p represents protein sites. The terms in eq 3 are given by

DOI: 10.1021/acs.jpcb.7b07726
J. Phys. Chem. B 2017, 121, 9358−9365

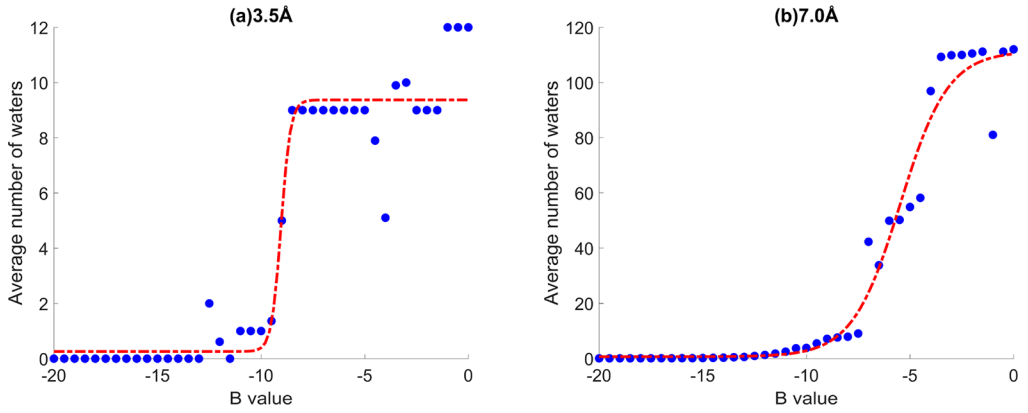


Figure 1. GCMC titration for water molecules in a bulk water sphere with different radii for the GCMC insertion/deletion.

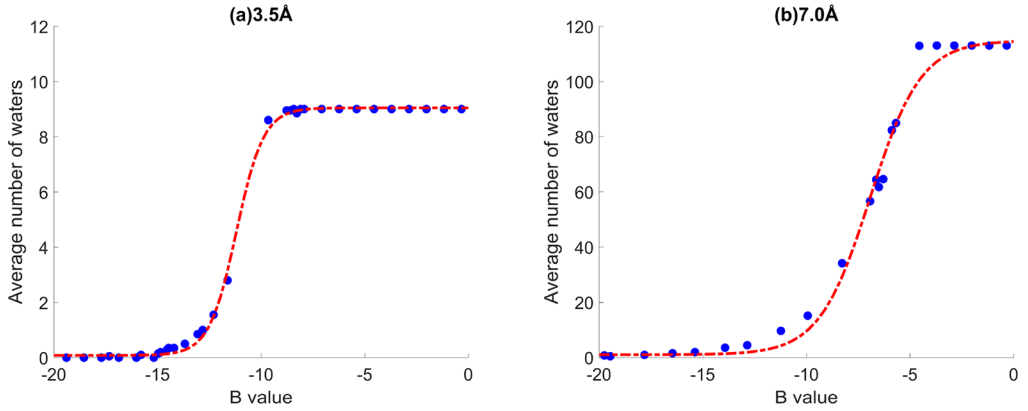


Figure 2. WF titration for water molecules in a bulk water sphere with different radii for the WF region.

$$\Delta G_i = \frac{1}{2}\langle U^i{}_Q - U^i{}_0\rangle_{U,Q} + \beta\langle U^i{}_{vdw}\rangle_{U,Q}$$
$$= \frac{1}{2}\sum_{k\notin WAT}\langle U^{ik}{}_Q - U^{ik}{}_0\rangle_{U,Q} + \beta\langle U^i{}_{vdw}\rangle_{UQ},$$
where $i \in$ WAT (4)

$$\Delta G_{ij} = \frac{1}{2}\langle U^{ik}{}_Q - U^{ik}{}_0\rangle_{U,Q} + \Delta G^{ij}{}_{ins,0ij},$$
where $i, j \in$ WAT (5)

$U^i{}_Q$ and $U^i{}_0$ denote the total "solvation" energy (self-energy) of the ith water molecule, and the subscript denotes the charge distribution of the water molecule. The angular brackets denote the ensemble average obtained by propagating trajectories over the polar state of the water. It is important to mention here that eq 4 does not include contribution from the water molecules, which have already been inserted in the protein cavity. The second part of eq 4 is calculated using the LIE[42−44] which provides a good approximation to the free energy of creating a cavity for the ith water molecule in the protein. Note in this respect that most of the free energy of the water molecules is electrostatic and thus captured quite reliably by the LRA.[23] Equation 5 denotes the total pairwise interaction between all ith and jth pairs of water molecules, where $U^{ij}{}_Q$ is the pairwise interaction of the i−j pair and Q represents the charge on the atoms of the water molecules. The last term in eq 5 denotes the pairwise term of the nonpolar interaction energy.

The calculated free energies of the internal water molecules are used to select different water configurations by a MC procedure and calculate the corresponding free energy using eq 3. The postprocessing WF MC approach was modified relative to our previous treatment[26] and follows the formulation of eq 1. This modification arranged the $\Delta G_{(m)}$ according to the number of water molecules in each configuration and then performed MC with moves that only add or subtract one water molecule.

For adding a water molecule, we use

$$P_{m+1} = \min\{1, (1/(N+1))$$
$$\exp[-((\Delta G_{m+1} - \Delta G_m)\beta - B'_{m+1})]\} \quad (6)$$

Similarly, we used the equivalent of eq 2 for deleting a water molecule.

The parameter B' is given by $B'_{m+1} = \Delta G_{bulk}\beta + \ln N_{m+1}$. This parameter can be evaluated by determining the free energy of inserting a single water molecule. Alternatively, the bulk energy that corresponds to the insertion free energy at the half point between no water and one full water molecule can be taken as B'. This postprocessing MC approach allowed us to select the minimum free energy configurations.

The radius and center of the simulating system for each cavity studied were set as stated in Table SI1. The simulations used the SCAAS surface constraints and the local reaction field (LRF) long-range treatment (see ref 45). The protein structures were subject to 3000 steps of minimization using steepest decent, followed by 300p MD relaxation, using the polarizable ENZYMIX force field[45] with time steps of 1.0 fs and the solute parameters given in refs 46 and 47. These structures were then used for GCMC and WF simulations. During both the GCMC and WF simulations, a spherical hard wall was

DOI: 10.1021/acs.jpcb.7b07726
J. Phys. Chem. B 2017, 121, 9358−9365

placed so that the inside water molecules could not escape and the outside water molecules could not enter. In GCMC, the insertion and deletion were attempted only inside this hard wall.

III. RESULTS AND DISCUSSION

We started by exploring the relationship between the performance of the WF and GCMC in a bulk water system, modeled as a water sphere. The dependence of the number of inserted water molecules on the *B* value for the GCMC simulations is summarized in Figure 1, and the corresponding WF results are depicted in Figure 2. As is clear from the figures, both models give very similar results.

Next, we moved to the much more challenging case of water molecules inside bovine pancreatic trypsin inhibitor (BPTI) (PDB: 5PTI[48]), which has been previously tested by Ross et al.[27]

We started with the site that is known to contain a single water molecule. The corresponding *B* dependence by the GCMC and WF models is described in Figures 3 and 4. As seen

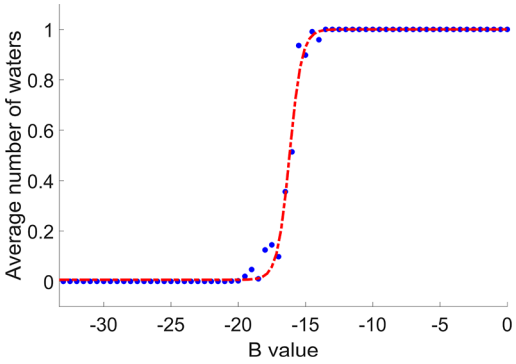


Figure 3. GCMC titration for water molecules in the single water cavity in BPTI.

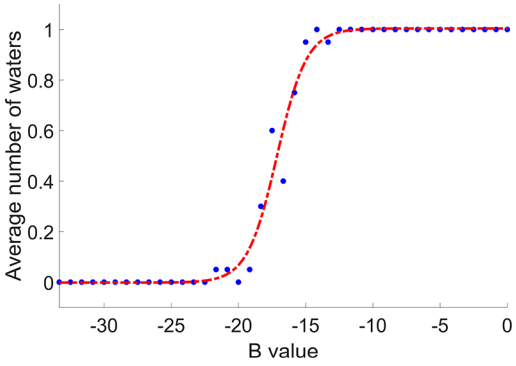


Figure 4. WF titration for water molecules in the single water cavity in BPTI.

from the figures, the trend in both figures is similar. An additional validation was performed by considering the three-water site in BPTI.[27] Here again, we compared the *B* dependence of the GCMC (Figure 5) and WF (Figure 6). The studies presented were done with different constraints on the protein, and the results seem to follow a similar trend. Next, we considered the simulated position of the internal water molecules in the three-water cavity of BPTI. As seen from Figure 7, both methods produce similar structures and *B* dependence profile.

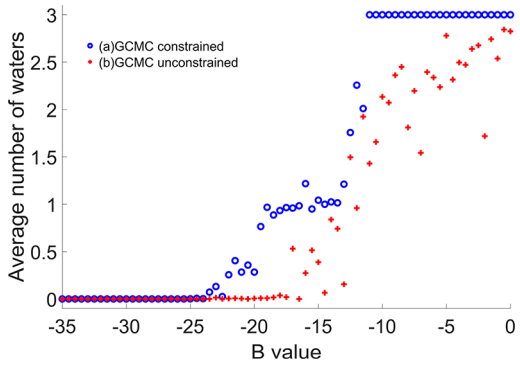


Figure 5. GCMC titration plot for the three-water cavity in BPTI, when the protein is constrained (blue) and unconstrained (orange).

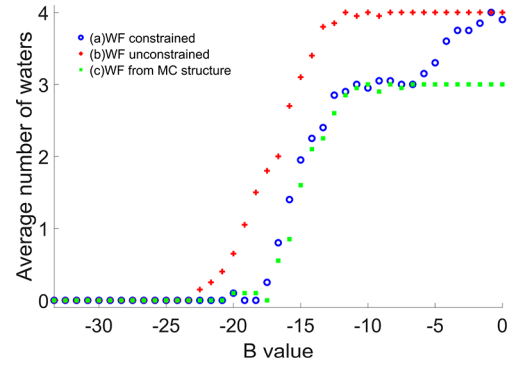


Figure 6. WF titration plot for water molecules in the three-water cavity in BPTI, when the protein is constrained (blue) and not constrained (orange). The WF calculation constrained on the structure generated by GCMC at *B* value = 0 is given in a green color plot.

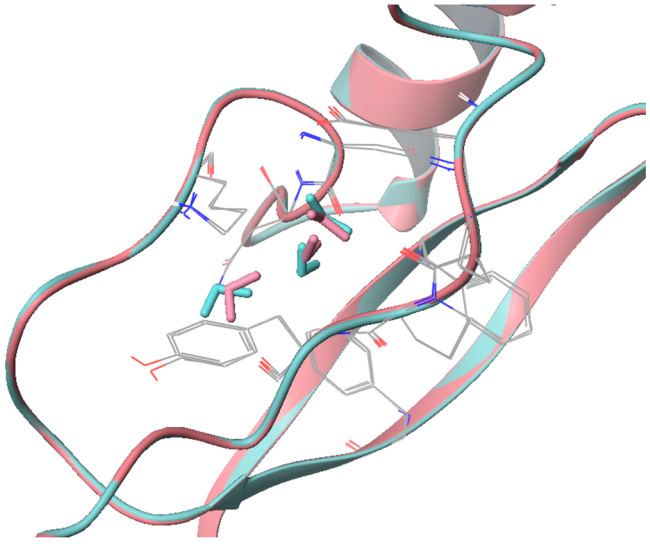


Figure 7. Comparing the three-water cavity sites of BPTI obtained by the WF (blue) and GCMC (pink) approaches, at a *B* value of −8.0. The cavity water molecules are colored the same as the protein backbone.

A crucial aspect that has to be addressed by both the WF and GCMC approaches is the dependence of the number of inserted water molecules on the protein structure. To establish this problem, we considered the three-water site of BPTI and generated open configurations by running short GCMC with

DOI: 10.1021/acs.jpcb.7b07726
J. Phys. Chem. B 2017, 121, 9358−9365

large positive B values. The corresponding configurations were used in WF calculations (Figure 8), generating an increase in

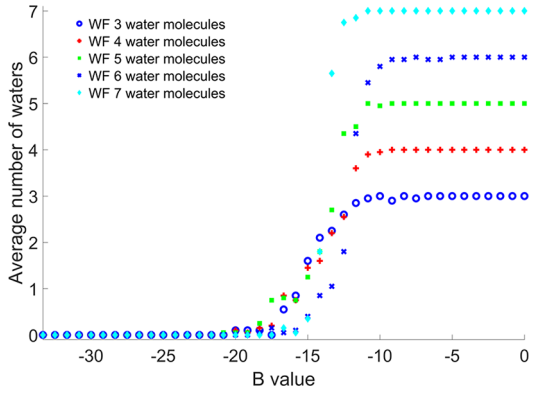


Figure 8. WF titration plot with different initial numbers of water molecules in the three-water cavity in BPTI.

the number of inserted water molecules. A similar trend occurs in the GCMC calculations (Figure 9). The problem is, of

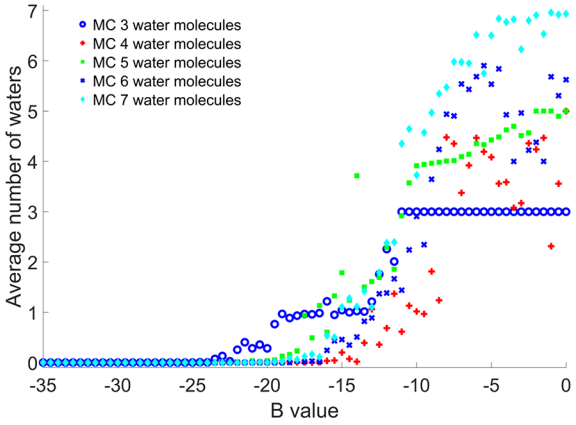


Figure 9. GCMC titration plot with different initial numbers of water molecules in the three-water cavity in BPTI.

course, to determine which configuration has the lowest free energy. This can be done by our specialized LRA type treatment, e.g. (see ref 49),

$$\Delta G(r_1 \rightarrow r_2) = \frac{1}{2}[\langle \Delta\varepsilon(q_m \rightarrow q_{m'})_{r_1,m}\rangle - \langle \Delta\varepsilon(q_m \rightarrow q_{m'})_{r_2,m'}\rangle] \quad (7)$$

where $\Delta\varepsilon$ designates the difference in energy of the two states in the water charging process and $\langle\ \rangle_{r_i,m}$ indicates an average over the charge distribution, q_m, where the system is held by a weak constraint near the position vector of the indicated states ($\mathbf{r}_i$). This expression tells us how much it costs to move the protein from one configuration to another. In order to evaluate the energetics of the protein deformation from the geometry obtained with three water molecules, r_1, to that obtained with seven water molecules, r_2, we performed the calculations while keeping the number of water molecules the same in both states (seven water molecules) but setting to zero the fractional charges on the last four inserted water molecules, when considering state 1. This calculation gave a 5.56 kcal/mol penalty for moving to the conformation of the seven water molecules. This penalty compensates for the gain in the water energy upon moving from three- to seven-water configurations at $B = -8$ (Figure 10). Of course, in the present case with the

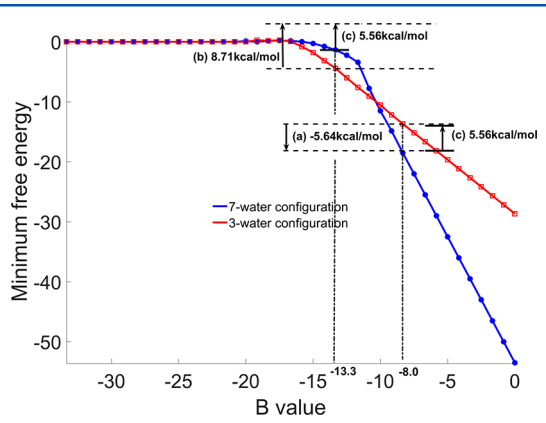


Figure 10. Plot of minimum free energies obtained by the WF's MC postprocessing, for configurations of three and seven water molecules, at the three-water-molecule cavity in BPTI. The difference of the minimum LRA energy between the two configurations is −5.64 kcal/mol at a B value of 8.0 (a) and 3.15 kcal/mol at a B value of 13.3 (b), and the free energy of protein deformation from a three-water-molecule to seven-water-molecule configuration calculated according to eq 6 is 5.56 kcal/mol (c).

correct B (−13.8), we already have negative energy in going from seven to three water molecules and the deformation energy makes this difference even larger. At any rate, our approach can be used as a general tool in cases of significant conformational change due to additional water molecules. It is also possible to evaluate the protein deformation energy by the far more expensive overcharging approach.[25]

At this stage, it may be useful to compare the computational time of the WF and GCMC methods. This is done in Figures 11 and 12 with the GCMC and WF, respectively, for the three-water site of BPTI. As seen from the figure, the WF approach easily converges in less than 10 ps, while the GCMC needs at least 5 ns (probably much longer) to converge.

We also studied the water insertion in V66D Staphylococcal nuclease (SNase, PDB: 2OXP[24]), which was also studied in our previous study[25] as a benchmark of the WF approach. In this case, residue V66 is located in a hydrophobic site and the mutation to ASP creates a case where an internal acid cannot be ionized until it is stabilized by internal water molecules and/or reorganization of the protein. As seen from Figure 13, the WF approach leads to a fast convergence of the internal water molecule to a configuration that was found to reproduce the observed pK_a.[25] On the other hand, we see (Figure 14) that the GCMC has significant convergence difficulties.

IV. CONCLUSIONS

Water molecules play a crucial role in determining the energetics of biological processes. Thus, it is crucial to capture, in computational studies of protein functions, the correct effect of water inside and around proteins. The problem is particularly challenging in cases of highly polar or charged environments in protein interiors. Addressing this challenge by regular MD studies is far from simple, since the equilibration time of water penetration processes might be extremely long. One may use specialized approximated approaches such as our WF method[26] or other strategies (e.g., refs 4 and 33). However, it is important

DOI: 10.1021/acs.jpcb.7b07726
J. Phys. Chem. B 2017, 121, 9358−9365

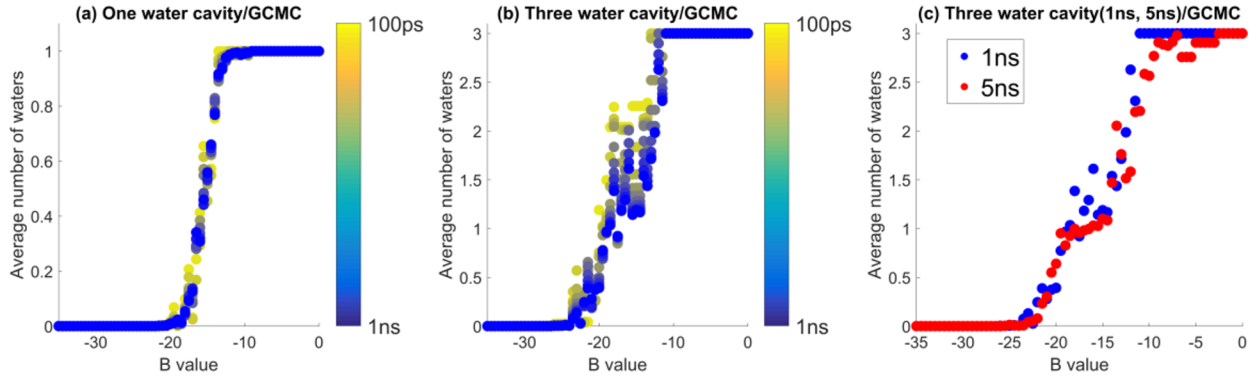


Figure 11. GCMC titration plot for runs from 100 ps to 1 (and 5) ns for the one-water and three-water cavities in BPTI.

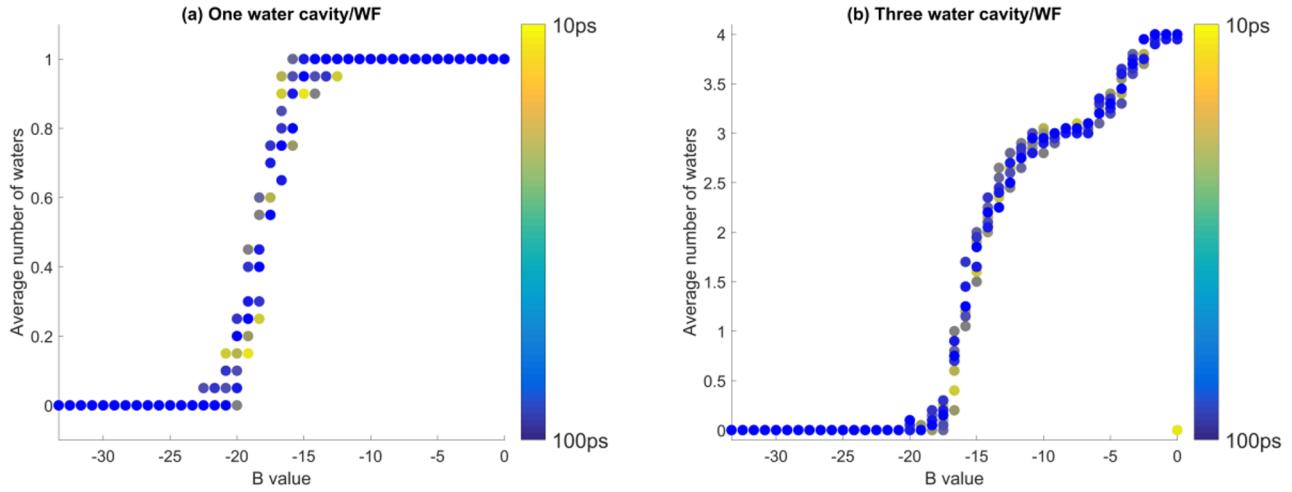


Figure 12. WF titration plot for runs from 10 to 100 ps for the one-water and three-water cavities in BPTI.

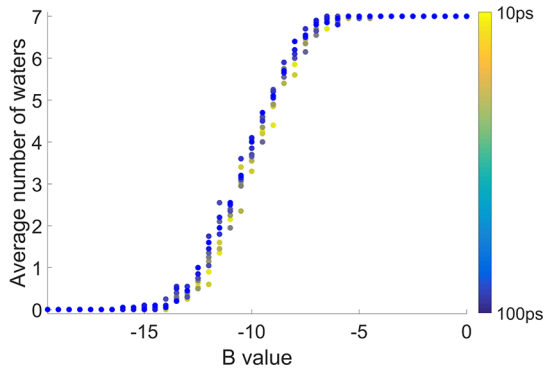


Figure 13. WF titration plot from 10 to 100 ps simulation for SNase.

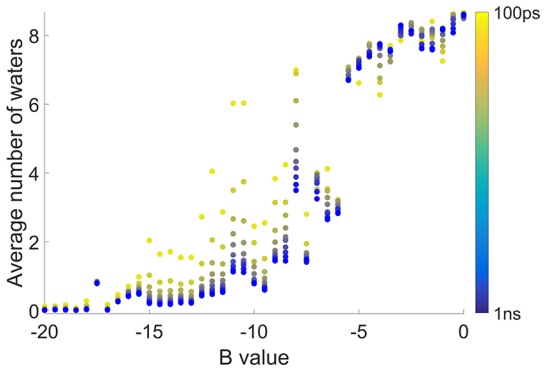


Figure 14. GCMC titration plot from 100 ps to 1 ns simulation for SNase.

to assess the validity of the approximation used. An attempt to validate the just add water molecules (JAWS) approach of ref 4 was reported in ref 27, but this validation concluded that the MC part of the JAWS approach has difficulties in convergence and requires a long simulation time and that converting the water densities to unique positions is far from simple. Furthermore, no validation in terms of the dependence on the *B* value was reported. As for the WaterMap[5,34] approach, one of the arguments has been that this approach can provide a way of capturing the water entropic effect. Now, while the entropic contributions of ordering of the water molecules were expressed by elegant expansion terms (e.g., ref 5), not only does this treatment have possible formal problems (e.g., ref 50), but also, to the best of our knowledge, there are no reported careful quantitative validations of such entropy estimates, as reported in studies that used our restraint release (RR) treatment.[51] It seems to us that at present there is no fast approach that can yield a reliable estimate of the water entropic effect.

Our WF approach has been used successfully in challenging cases but again has not been validated in terms of its rigorous foundation. In the present work, we tried to validate the WF approach by comparing its results to those of the corresponding GCMC simulations. The comparison yielded very encouraging results, basically recovering the GCMC trend. In doing so, the

DOI: 10.1021/acs.jpcb.7b07726
J. Phys. Chem. B 2017, 121, 9358−9365

WF reproduced the GCMC for both configurations and water insertion energy (as is apparent from having a similar B value).

Apparently, as is clear from the present paper and from ref 27, one can use GCMC in studies of water penetration to proteins. Unfortunately, the computer time needed for converging GCMC results is very extensive where the results of Figure 11 show that we need more than 5 ns for the three-water-molecule case. In fact, an estimate of ref 27 seems to suggest that GCMC should use about 200 million MC steps per each B value simulation, which correspond to 2 μs in our MD/GCMC implementation. On the other hand, the WF simulations take about 10 ps for both one and three water molecules. Thus, this is the method of choice, when one is interested in practical yet reliable calculations, for example, in the case of screening for ligand binding. Moreover, if one likes to generate a titration plot of the average number of water molecules at each B value, many independent GCMC simulations are needed to be run for the entire range of B values, while WF only needs one short MD run, since the very fast postprocessing MC can generate a titration plot within a few CPU seconds.

An interesting issue that was addressed in the present work is the dependence of the configuration of the water molecules on the protein conformations. This fundamental issue is far from trivial and in some cases should be treated by eq 7. In fact, the same problem exists with GCMC and other approaches. Perhaps, the dependence on the protein configurations should be handled in the same way they are handled in the stage where we start with different initial protein configurations to obtain the average results for FEP calculations (or in the related replica exchange steps).

In our view, the issue of water insertion is the most critical in studies of the effect of internal water molecules near charged residues, as is the case in cytochrome c oxidase[26] and the V66D in SNase considered as above. Such calculations can be very expensive in particular when we consider pathways of charge transfer, and thus, the WF approach is ideal for such type of problems.

Overall, we believe that the WF approach is very efficient and powerful and is arguably the method of choice in cases of calculations that are limited by the computer time (e.g., fast screening of drugs). However, if one is interested in very expensive, and in principle more rigorous, GCMC, it can be accelerated by starting from the WF configurations.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpcb.7b07726.

The center coordinates and the sizes of the simulating systems and cavities used for the water insertion (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: warshel@usc.edu.

ORCID

Hanwool Yoon: 0000-0002-7761-4321

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We would like to thank NIH grant GM40283 and NIH Grant 5U19CA105010. All computational work was performed on the University of Southern California High Performance Computing and Communication Center (HPCC), whom we would like to thank for having provided us with access to computer time on the HPCC cluster.

■ REFERENCES

(1) Ball, P. Water as an Active Constituent in Cell Biology. *Chem. Rev.* **2008**, *108* (1), 74−108.

(2) Warshel, A. *Computer Modeling of Chemical Reactions in Enzymes and Solutions*; John Wiley & Sons: New York, 1991.

(3) Kato, M.; Braun-Sand, S.; Warshel, A. Chapter 15 Challenges and Progresses in Calculations of Binding Free Energies - What Does It Take to Quantify Electrostatic Contributions to Protein-Ligand Interactions? *Computational and Structural Approaches to Drug Discovery: Ligand-Protein Interactions*; The Royal Society of Chemistry: London, 2008.

(4) Michel, J.; Tirado-Rives, J.; Jorgensen, W. L. Prediction of the Water Content in Protein Binding Sites. *J. Phys. Chem. B* **2009**, *113* (40), 13337−13346.

(5) Young, T.; Abel, R.; Kim, B.; Berne, B. J.; Friesner, R. A. Motifs for Molecular Recognition Exploiting Hydrophobic Enclosure in Protein−Ligand Binding. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104* (3), 808−813.

(6) Helms, V.; Wade, R. C. Thermodynamics of Water Mediating Protein-Ligand Interactions in Cytochrome P450cam: A Molecular Dynamics Study. *Biophys. J.* **1995**, *69* (3), 810−824.

(7) Stephens, P. J.; Jollie, D. R.; Warshel, A. Protein Control of Redox Potentials of Iron-Sulfur Proteins. *Chem. Rev.* **1996**, *96*, 2491.

(8) Morais-Cabral, J. H.; Zhou, Y.; MacKinnon, R. Energetic Optimization of Ion Conduction Rate by the K+ Selectivity Filter. *Nature* **2001**, *414* (6859), 37−42.

(9) Pisliakov, A. V.; Sharma, P. K.; Chu, Z. T.; Haranczyk, M.; Warshel, A. Electrostatic Basis for the Unidirectionality of the Primary Proton Transfer in Cytochrome C Oxidase. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105* (22), 7726−31.

(10) Lee, H. J.; Svahn, E.; Swanson, J. M. J.; Lepp, H.; Voth, G. A.; Brzezinski, P.; Gennis, R. B. Intricate Role of Water in Proton Transport through Cytochrome C Oxidase. *J. Am. Chem. Soc.* **2010**, *132* (45), 16225−16239.

(11) Dwyer, J. J.; Gittis, A. G.; Karp, D. A.; Lattman, E. E.; Spencer, D. S.; Stites, W. E.; García-Moreno, E. B. High Apparent Dielectric Constants in the Interior of a Protein Reflect Water Penetration. *Biophys. J.* **2000**, *79* (3), 1610−1620.

(12) Fitch, C. A.; Karp, D. A.; Lee, K. K.; Stites, W. E.; Lattman, E. E.; García-Moreno, E. B. Experimental Pka Values of Buried Residues: Analysis with Continuum Methods and Role of Water Penetration. *Biophys. J.* **2002**, *82* (6), 3289−3304.

(13) Warshel, A.; Levitt, M. Theoretical Studies of Enzymic Reactions: Dielectric, Electrostatic and Steric Stabilization of the Carbonium Ion in the Reaction of Lysozyme. *J. Mol. Biol.* **1976**, *103*, 227−249.

(14) Warshel, A.; Russel, S. T. Calculations of Electrostatic Interactions in Biological Systems and in Solutions. *Q. Rev. Biophys.* **1984**, *17*, 283−422.

(15) Warshel, A.; Sharma, P. K.; Kato, M.; Parson, W. W. Modeling Electrostatic Effects in Proteins. *Biochim. Biophys. Acta, Proteins Proteomics* **2006**, *1764* (11), 1647−1676.

(16) Warshel, A.; Sussman, F.; King, G. Free Energy of Charges in Solvated Proteins: Microscopic Calculations Using a Reversible Charging Process. *Biochemistry* **1986**, *25*, 8368−8372.

(17) Warshel, A.; King, G. Polarization Constraints in Molecular Dynamics Simulation of Aqueous Solutions: The Surface Constraint All Atom Solvent (Scaas) Model. *Chem. Phys. Lett.* **1985**, *121*, 124−129.

DOI: 10.1021/acs.jpcb.7b07726
J. Phys. Chem. B 2017, 121, 9358−9365

(18) King, G.; Warshel, A. A Surface Constrained All-Atom Solvent Model for Effective Simulations of Polar Solutions. *J. Chem. Phys.* **1989**, *91* (6), 3647−3661.
(19) Warshel, A. Calculations of Chemical Processes in Solutions. *J. Phys. Chem.* **1979**, *83*, 1640−1650.
(20) Brooks, C. L., III; Karplus, M. Deformable Stochastic Boundaries in Molecular Dynamics. *J. Chem. Phys.* **1983**, *79*, 6312−6325.
(21) Im, W.; Berneche, S.; Roux, B. Generalized Solvent Boundary Potential for Computer Simulations. *J. Chem. Phys.* **2001**, *114* (7), 2924−2937.
(22) Ghosh, N.; Prat-Resina, X.; Gunner, M. R.; Cui, Q. Microscopic Pka Analysis of Glu286 in Cytochrome C Oxidase (Rhodobacter Sphaeroides): Toward a Calibrated Molecular Model. *Biochemistry* **2009**, *48* (11), 2468−2485.
(23) Warshel, A.; Sharma, P. K.; Kato, M.; Parson, W. W. Modeling Electrostatic Effects in Proteins. *Biochim. Biophys. Acta, Proteins Proteomics* **2006**, *1764* (11), 1647−1676.
(24) Karp, D. A.; Gittis, A. G.; Stahley, M. R.; Fitch, C. A.; Stites, W. E.; García-Moreno, E. B. High Apparent Dielectric Constant inside a Protein Reflects Structural Reorganization Coupled to the Ionization of an Internal Asp. *Biophys. J.* **2007**, *92* (6), 2041−2053.
(25) Kato, M.; Warshel, A. Using a Charging Coordinate in Studies of Ionization Induced Partial Unfolding. *J. Phys. Chem. B* **2006**, *110* (23), 11566−11570.
(26) Chakrabarty, S.; Warshel, A. Capturing the Energetics of Water Insertion in Biological Systems: The Water Flooding Approach. *Proteins: Struct., Funct., Genet.* **2013**, *81* (1), 93−106.
(27) Ross, G. A.; Bodnarchuk, M. S.; Essex, J. W. Water Sites, Networks, and Free Energies with Grand Canonical Monte Carlo. *J. Am. Chem. Soc.* **2015**, *137* (47), 14930−14943.
(28) Bodnarchuk, M. S.; Viner, R.; Michel, J.; Essex, J. W. Strategies to Calculate Water Binding Free Energies in Protein-Ligand Complexes. *J. Chem. Inf. Model.* **2014**, *54* (6), 1623−1633.
(29) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Oxford University Press: Oxford, U.K., 1989.
(30) Frenkel, D.; Smit, B. *Understanding Molecular Simulation: From Algorithms to Applications*; Academic Press: Massachusetts, 2002.
(31) Resat, H.; Mezei, M. Grand Canonical Monte Carlo Simulation of Water Positions in Crystal Hydrates. *J. Am. Chem. Soc.* **1994**, *116* (16), 7451−7452.
(32) Lynch, G. C.; Pettitt, B. M. Grand Canonical Ensemble Molecular Dynamics Simulations: Reformulation of Extended System Dynamics Approaches. *J. Chem. Phys.* **1997**, *107* (20), 8594−8610.
(33) Woo, H.-J.; Dinner, A. R.; Roux, B. Grand Canonical Monte Carlo Simulations of Water in Protein Environments. *J. Chem. Phys.* **2004**, *121* (13), 6392−6400.
(34) Abel, R.; Young, T.; Farid, R.; Berne, B. J.; Friesner, R. A. Role of the Active-Site Solvent in the Thermodynamics of Factor Xa Ligand Binding. *J. Am. Chem. Soc.* **2008**, *130* (9), 2817−2831.
(35) Bodnarchuk, M. S. Water, Water, Everywhere··· It's Time to Stop and Think. *Drug Discovery Today* **2016**, *21* (7), 1139−1146.
(36) Luccarelli, J.; Michel, J.; Tirado-Rives, J.; Jorgensen, W. L. Effects of Water Placement on Predictions of Binding Affinities for P38 Alpha Map Kinase Inhibitors. *J. Chem. Theory Comput.* **2010**, *6* (12), 3850−3856.
(37) Yoon, H.; Warshel, A. The Control of the Discrimination between Dntp and Rntp in DNA and Rna Polymerase. *Proteins: Struct., Funct., Genet.* **2016**, *84* (11), 1616−1624.
(38) Shelley, J. C.; Patey, G. N. A Configuration Bias Monte-Carlo Method for Water. *J. Chem. Phys.* **1995**, *102* (19), 7656−7663.
(39) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. Equation of State Calculations by Fast Computing Machines. *J. Chem. Phys.* **1953**, *21* (6), 1087−1092.
(40) Warshel, A.; Chu, Z. T.; Villa, J.; Strajbl, M.; Schutz, C. N.; Shurki, A.; Vicatos, S.; Plotnikov, N. V.; Schopf, P. *Molaris-XG, v 9.15*; University of Southern California: Los Angeles, 2012.
(41) Chakrabarty, S.; Warshel, A. Capturing the Energetics of Water Insertion in Biological Systems: The Water Flooding Approach. *Proteins: Struct., Funct., Genet.* **2013**, *81* (1), 93−106.
(42) Lamarre, D.; Anderson, P. C.; Bailey, M.; Beaulieu, P.; Bolger, G.; Bonneau, P.; Bös, M.; Cameron, D. R.; Cartier, M.; Cordingley, M. G. An Ns3 Protease Inhibitor with Antiviral Effects in Humans Infected with Hepatitis C Virus. *Nature* **2003**, *426* (6963), 186−189.
(43) Åqvist, J.; Medina, C.; Samuelsson, J.-E. A New Method for Predicting Binding Affinity in Computer-Aided Drug Design. *Protein Eng., Des. Sel.* **1994**, *7* (3), 385−391.
(44) Sham, Y. Y.; Chu, Z. T.; Tao, H.; Warshel, A. Examining Methods for Calculations of Binding Free Energies: Lra, Lie, Pdld-Lra, and Pdld/S-Lra Calculations of Ligands Binding to an Hiv Protease. *Proteins: Struct., Funct., Genet.* **2000**, *39* (4), 393−407.
(45) Lee, F. S.; Chu, Z. T.; Warshel, A. Microscopic and Semimicroscopic Calculations of Electrostatic Energies in Proteins by the Polaris and Enzymix Programs. *J. Comput. Chem.* **1993**, *14*, 161−185.
(46) Olsson, M. H. M.; Sharma, P. K.; Warshel, A. Simulating Redox Coupled Proton Transfer in Cytochrome C Oxidase: Looking for the Proton Bottleneck. *FEBS Lett.* **2005**, *579* (10), 2026−2034.
(47) Kato, M.; Pisliakov, A. V.; Warshel, A. The Barrier for Proton Transport in Aquaporins as a Challenge for Electrostatic Models: The Role of Protein Relaxation in Mutational Calculations. *Proteins: Struct., Funct., Genet.* **2006**, *64* (4), 829−844.
(48) Wlodawer, A.; Walter, J.; Huber, R.; Sjolin, L. Structure of Bovine Pancreatic Trypsin-Inhibitor - Results of Joint Neutron and X-Ray Refinement of Crystal Form-Ii. *J. Mol. Biol.* **1984**, *180* (2), 301−329.
(49) Braun-Sand, S.; Sharma, P. K.; Chu, Z. T.; Pisliakov, A. V.; Warshel, A. The Energetics of the Primary Proton Transfer in Bacteriorhodopsin Revisited: It Is a Sequential Light-Induced Charge Separation after All. *Biochim. Biophys. Acta, Bioenerg.* **2008**, *1777* (5), 441−452.
(50) Nguyen, C. N.; Kurtzman, T.; Gilson, M. K. Spatial Decomposition of Translational Water-Water Correlation Entropy in Binding Pockets. *J. Chem. Theory Comput.* **2016**, *12* (1), 414−429.
(51) Singh, N.; Warshel, A. A Comprehensive Examination of the Contributions to the Binding Entropy of Protein−Ligand Complexes. *Proteins: Struct., Funct., Genet.* **2010**, *78* (7), 1724−1735.

DOI: 10.1021/acs.jpcb.7b07726
J. Phys. Chem. B 2017, 121, 9358−9365