Title

Dipole Moment and Binding Energy of Water in Proteins from Crystallographic Analysis

Journal

Journal of Chemical Theory and Computation, 10(10)

ISSN

1549-9618

Authors

Morozenko, A
Leontyev, IV
Stuchebrukhov, AA

Publication Date

2014-10-14

DOI

10.1021/ct500358r

Peer reviewed

JCTC Journal of Chemical Theory and Computation

Article

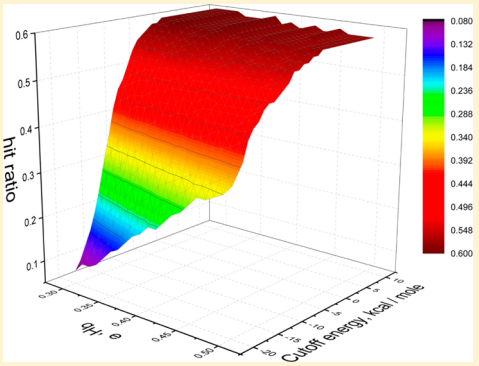
pubs.acs.org/JCTC

Dipole Moment and Binding Energy of Water in Proteins from Crystallographic Analysis

A. Morozenko, I. V. Leontyev, and A. A. Stuchebrukhov*

Department of Chemistry, University of California Davis, One Shields Avenue, Davis, California 95616, United States

ABSTRACT: The energetics of water molecules in proteins is studied using the water placement software Dowser. We compared the water position predictions for 14 high-resolution crystal structures of oligopeptide-binding protein (OppA) containing a large number of resolved internal water molecules. From the analysis of the outputs of Dowser with variable parameters and comparison with experimental X-ray data, we derived an estimate of the average dipole moment of water molecules located in the internal cavities of the protein and their binding energies. The water parameters thus obtained from the experimental data are then analyzed within the framework of charge-scaling theory developed recently by this group; the parameters are shown to be in good agreement with the predictions that the theory makes for the dipole moment in a protein environment. The water dipole in the protein environment is found to be much different from that in the bulk and in such models as SPC or TIPnP. The role of charge scaling due to electronic polarizability of the protein is discussed.



1. INTRODUCTION

Water molecules inside proteins often play important role in proteins' functions.[1−8] For instance, the long-range proton transfer inside of proteins is usually carried out by the Grotthuss mechanism and requires a chain of hydrogen bonds that is composed of internal water molecules and amino acid residues of the protein. In other cases, water molecules can facilitate the enzymes catalytic reactions by becoming a temporary proton donor/acceptor. Also, being a high dielectric, water screens electrostatic interaction of charges, which can lead to stabilization of charge on ionizable residues or can affect their pK_a values. However, the dynamic nature of water in a protein does not allow resolution of all the molecules, and only some of the protein's water molecules are typically seen in the X-ray structure.[9−11] To predict how many molecules are actually present in the protein, the usage of special software is required.

The program Dowser[12−14] is a package for automatically hydrating protein structures, which is widely used in computational biophysics (e.g., it is part of VMD[15]). This program requires a PDB file with atomic coordinates of a protein as an input file and as an output it produces another PDB file with a list of coordinates of placed water molecules together with their energy values. To accomplish this, the program finds internal cavities in the protein and fills the cavities with water molecules depending on their minimized energies. The energy of the waters is evaluated by using a specific model of water−protein interaction, and only molecules with binding energy greater than a threshold value of 10 kcal/mol are retained. (The entropy and the kinetic energy are not counted; only potential energy of interaction between a water molecule and protein is taken into account.) The value of 10 kcal/mol is very close to the total enthalpy of vaporization (10.5 kcal/mol),[16] which is (within a kT) the amount of energy required to bring a single water molecule from bulk to vacuum. However, the threshold value is ill-defined and essentially is an empirical parameter; one can argue that the binding energy of water in the protein should be compared to the amount of energy required to bring a water molecule from the *bulk* to the protein environment, rather than to a vacuum, which is significantly smaller than the vacuum 10 kcal/mol. (It is clear that rigorous criterion of binding should be based on the comparison of chemical potentials of water in the bulk and in the hydration site of the protein. The entropy change between bulk and protein is not counted by Dowser, which may appear to be a too crude approximation; however, the entropy contribution is not expected to be significant (∼1 kcal/mol) on the scale of interaction energies; see the discussion section for details, and the approach of Dowser can be justified as semiquantitative one).

The insertion model of water is also ill-defined; the one that is implemented in Dowser is SPC (simple point charge).[17] This model treats a water molecule as a rigid body with three interaction sites, corresponding to three atoms of the water molecule. Hydrogen atoms are assigned a point charge q_H = +0.41 e and the oxygen atom assigned q_O = −0.82 e. It is one of the simplest nonpolarizable models of water, along with TIP3P[18] and TIP4P[18] (transferable intermolecular potential with 3 or 4 points) and their modifications, which have been in use in computer simulations for over 40 years. These models describe reasonably well the properties of pure *bulk* water, such as density, radial distribution function, diffusion coefficient, static dielectric constant, and the enthalpy of vaporization (for

Received: April 24, 2014
Published: September 15, 2014

   dx.doi.org/10.1021/ct500358r | *J. Chem. Theory Comput.* 2014, 10, 4618−4623

details, see discussion in our recent papers[19,20] and references therein).

However, water molecules in a protein environment are different from those in the bulk phase.[19] Indeed, due to its polarizability, the dipole moment of water strongly depends on the dielectric properties of the surrounding medium. For instance, the dipole moment of water in the gas phase (dielectric constant $\varepsilon_{gas} = 1$) is 1.85 D,[21] while in the high-dielectric environment of liquid water ($\varepsilon_{liquid} = 78$) it is enhanced to about 3.0 D.[22−25] Taking the dielectric constants of a protein $\varepsilon_{protein} = 4$,[26−28] one can argue that the dipole moment of water molecules in proteins should be in the range between 1.85 and 3.0 D. Therefore, the models of bulk water described above may not be suitable for simulations in a protein environment. (Our earlier studies,[19,20] where properties of water in different media such as polar, nonpolar, and charged environment were examined, also support this conclusion.)

Changing the dipole moment of the water model implemented in Dowser affects the total energy of a molecule placed by the program, particularly its electrostatic part. Also, raising the cutoff energy decreases the number of molecules discarded due to the energy criterion. Hence, the total amount of inserted molecules and the number of those of them that match water molecules seen in the X-ray structure depend on the cutoff energy E_{cutoff} and the dipole moment μ of the model of water implemented in the program.

In this paper, we report an attempt to estimate the best empirical value of the dipole moment of water and its binding energy appropriate for simulations in a protein environment using Dowser as a tool for the analysis of experimental data. To this end, we applied Dowser with variable parameters of the cutoff energy E_{cutoff} and the effective dipole moment μ to 14 high-resolution crystal structures of oligopeptide-binding protein (OppA) containing a large number of resolved internal water molecules. From the analysis of the outputs of this program and comparison with experimental X-ray data, we derived the best estimates of an average dipole moment of water molecules located in the internal cavities of the protein and their binding energies. The water parameters are then analyzed within the framework of charge-scaling theory developed recently by this group;[20] the parameters are shown to be in good agreement with the predictions that the theory makes for the dipole moment in a protein environment. The water dipole in the protein environment found to be much different from that in the bulk and in such models as SPC or TIPnP. The role of charge scaling due to electronic polarizability of the protein is discussed.

2. COMPUTATIONAL MODEL

To investigate how varying μ and E_{cutoff} affects the prediction ability of Dowser, we performed a series of computational experiments on 14 structures of oligopeptide-binding protein (OppA) bound to different lysine-X-lysine tripeptides.[29] For the analysis, we chose the same data set of proteins as in refs 30 and 31. To this end, we first removed all the existing water molecules (crystallographic) from the protein structures and saved them as separate PDB files. Then, making the necessary changes in the code of Dowser, we varied either μ or E_{cutoff} and ran the recompiled program on each of the proteins in the data set. After every program's run we compared its output with the experimental data.

To analyze how successful Dowser is in predicting crystallographic waters, we calculated how many molecules from the experimental data sets matched molecules from the Dowser output files. (The criterion of the quality of prediction of hydration water is a nontrivial one, because only a part of the hydrated water molecules are usually seen in the crystal structure; the disordered, or mobile, water molecules are not observed. Thus, an ideal quality program should not only predict those molecules that are observed in the crystal structure but also those that are not seen. The following detailed analysis partially addresses this nontrivial issue). An experimental molecule was considered to match a molecule placed by the program if the displacement between the positions of their oxygens was less than 1.4 A, the radius of a water molecule. In such a case, an experimental molecule was referred to as *predicted*. To estimate the success of a single Dowser run, we introduced a parameter hit ratio χ, the ratio of the number of predicted molecules to the total number of crystallographic waters,

$$\chi = \frac{N_{predicted}}{N_{crystal}} \quad (1)$$

Since the total number of inserted water molecules depends on both μ and E_{cutoff}, the number of predicted molecules and hit ratio χ also depend on these two parameters. We varied parameters μ and E_{cutoff} and ran Dowser on each of the protein structures from the data set. χ was calculated for every protein and every set of μ and E_{cutoff} in question. During this computational experiment, we changed the cutoff energy from −20 to 10 kcal/mol and varied the charge on the hydrogen atoms in the range from 0.30 to 0.50 e (in SPC model $q_H = +0.41e$), while keeping the molecule neutral by assigning the oxygen atom twice the negative charge value of that of hydrogen $q_O = -2q_H$. Then, χ was averaged over all proteins and a three-dimensional surface $\chi_{ave}(E_{cutoff},q)$ was constructed; the result is shown in Figure 1.

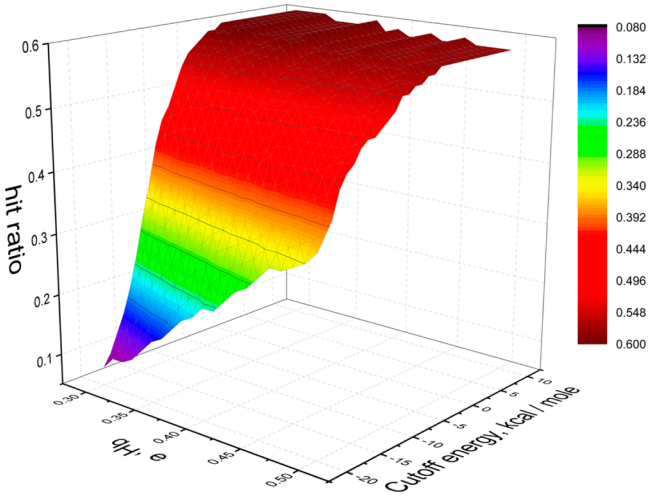


Figure 1. Averaged hit ratio $\chi_{ave}(q,E_{cutoff})$ as a function of cutoff energy and charge on the hydrogen atom of the inserted water.

Figure 1 shows χ as a function of q_H and E_{cutoff}. The charge q_H sets the dipole μ of the water model. In the region of low cutoff energies, χ is very small because all molecules are discarded by Dowser according to its energy criterion. As the cutoff energy grows, χ monotonously increases and eventually reaches a plateau. The plateau is a region where further increase of inserted molecules, which in general occurs with increasing

dx.doi.org/10.1021/ct500358r | J. Chem. Theory Comput. 2014, 10, 4618−4623

the cutoff energy, no longer changes the number of predicted water molecules matching those observed experimentally. At very high cutoff energies, all water accessible sites found by Dowser are occupied by the inserted molecules, and further, the increase of the cutoff energy does not change the number of inserted molecules.

It should be noticed that the maximum hit ratio in Figure 1 is only 0.6, which is 40% lower than a desirable 100% prediction. This result signifies that even upon removal of all energy constrains, Dowser can predict (in our case) only up to 60% of crystallographic molecules. This low percentage indicates that the wrong choice of the parameters μ and E_{cutoff} is only a part of the program's shortcomings (e.g., the program neglects protein structural flexibility).

For a given cutoff energy E_{cutoff}, there is an optimum value of charge q_H^* that maximizes the prediction quality χ. The dependence of the optimum charge q_H^* on the cutoff energy is shown in Figure 2.

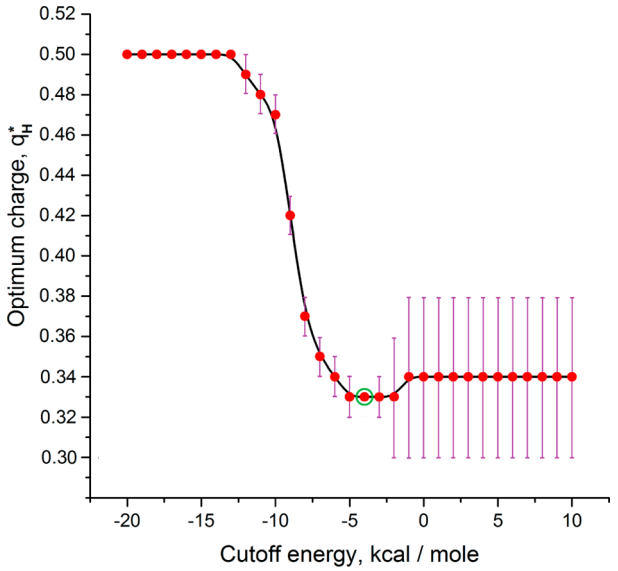


Figure 2. Optimum value of charge q_H^* versus cutoff energy.

At different cutoff energies, the maximum hit ratio is achieved at different values of q_H. The maximum is rather broad, so that at each cutoff energy there is an interval of charge values that maximize the hit ratio. These intervals of optimum values of q_H^* are shown as error bars in Figure 2. The intervals of best values of charge depend on the cutoff energy, as do the average best values themselves. We notice that at cutoff energy −4 kcal/mol, the uncertainty of the optimum charge q_H^* is lowest, and the maximum $\chi_{ave}(E_{cutoff}q_H^*)$ is achieved rigorously at q_H^* = 0.33 e (indicated by a circle). The special meaning of the cutoff energy of −4 kcal/mol and the value of the optimum charge at this energy is discussed later in this section.

The existence of the optimum charge (or the dipole of water molecule) can be qualitatively rationalized by the analysis of interactions of the inserted water molecules with the protein atoms as follows.

The total energy of a water molecule inserted by Dowser E_{water}^{total} is a sum of three interactions: 6−12 Lennard-Jones potential, charge−dipole interaction of molecules with background charges of a protein and an electrostatic dipole−dipole interaction of molecules with each other,

$$E_{water}^{total} = E_{LJ} + U_{protein-water} + U_{water-water} \quad (2)$$

The dipole moment of a model of water is directly proportional to the charges assigned to the atoms $\mu \sim q_O \sim q_H$. The Lennard-Jones potential E_{LJ} is independent of charges, the protein-molecule interaction is linear in charge, $U_{protein-water} \sim q$, and water−water interaction is quadratic, $U_{water-water} \sim q^2$. Overall, the total energy is a quadratic function of charges and has a stationary value at some charge q^*.

When the charges assigned to water atoms are very small, the electrostatic interactions are negligible, and the energy of a molecule is determined by the Lennard-Jones potential E_{LJ} alone, independent of charges. As the charges of water increase, the magnitude of the electrostatic terms increases. First, the favorable protein−water interactions, which are linear in charge, play the major role. The protein−water term $U_{protein-water}$ is negative when interaction is favorable, and hence, this term decreases the total energy of the inserted water molecules E_{water}^{total}. The decrease in energy reduces the number of discarded water molecules and χ goes up. However, further increase of charges changes this trend.

When minimizing energy of electrostatic interactions between the dipoles of water molecules and the background charges, we noticed that often several adjacent molecules of a cluster tend to line up pointing at the same charge of the protein, thereby making up a system of parallel dipoles. The water−water interaction now begins to play an important role. The favorable protein−molecule interaction is linear in charge $U_{protein-water} \sim q$, but the unfavorable water−water interaction is quadratic $U_{water-water} \sim q^2$. At small charges, favorable linear term dominates, but at large charges, the unfavorable quadratic term plays major role. As a result, an optimum value of charge exists at which energy is minimized; the lower the energy of a molecule, the higher the probability that the criterion $E_{water}^{total} < E_{cutoff}$ is satisfied, and hence, the higher the hit ratio χ. Hence, for a given cutoff energy, there is an optimum value of charge q_H^* that maximizes the hit ratio χ.

As an example, Figure 3 shows the dependence of hit ratio on charge for cutoff energy −4 kcal/mol. The maximum is reached at q_H^* = 0.33 e. At all other energies, the optimum charge is shown in Figure 2.

Figure 4 shows the maximum hit ratio χ $(q_H^*(E_{cutoff}),E_{cutoff})$ for different proteins and the average χ_{ave} $(q_H^*(E_{cutoff}),E_{cutoff})$ over all 14 protein structures as a function of the cutoff energy; the value of charge for each curve is q_H^*, which maximizes χ at a given cutoff energy. The dependence $q_H^*(E_{cutoff})$ is shown in Figure 2.

As seen from Figure 4, the maximum hit ratio increases with the increase of the cutoff energy. However, for every protein in the data set there is a "saturation point" at which further increase of the cutoff energy no longer results in an increase of the hit ratio. Since the energies of "saturation" and maximum hit ratios are slightly different for different proteins, we performed averaging of hit ratios over all the proteins in the data set. The full line on Figure 4 represents the averaged hit ratio χ_{ave} $(q_H^*(E_{cutoff}),E_{cutoff})$ as a function of cutoff energy. The average hit ratio grows monotonously until it reaches its maximum value 0.60 at −4 kcal/mol, which indicates that −4 kcal/mol is the optimum value of this parameter.

Setting the cutoff energy *lower* than −4 kcal/mol results in discarding some crystallographic molecules, that is, decreasing the hit ratio χ_{ave}; whereas setting the cutoffs *above* −4 kcal/mol, the hit ratio does not change, but the number of "wrong"

dx.doi.org/10.1021/ct500358r | J. Chem. Theory Comput. 2014, 10, 4618−4623

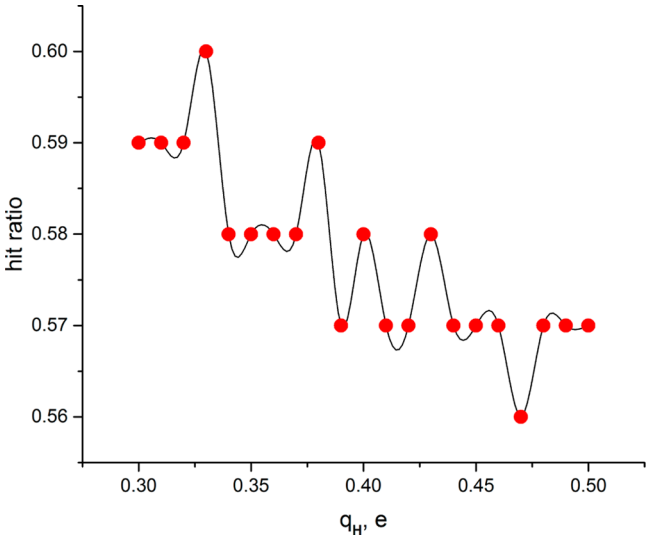


Figure 3. $\chi_{ave}(q_H)$ at E_{cutoff} = −4 kcal/mol. The discrete nature of the dependence is due to the nature of the χ function as a ratio of two integer numbers, eq 1.

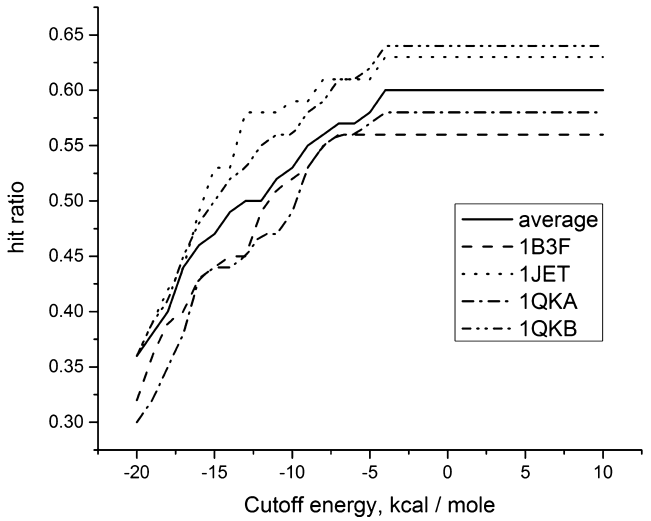


Figure 4. Hit ratio χ $(q_H^*(E_{cutoff}),E_{cutoff})$ versus cutoff energy for different protein structures from the data set, and their average χ_{ave} $(q_H^*(E_{cutoff}),E_{cutoff})$.

molecules, that is, those molecules that should not be in the structure, increases. Thus, the balance between correctly predicted molecules, and the molecules that Dowser places into protein cavities that should not be there sets the optimum value of the cutoff energy around the value where χ_{ave} reaches its maximum. Formally, the criterion could be made more rigorous; however, in practice, it would still involve some uncertainty due to arbitrary definition of the numerical criterion based on the balance of correctly and incorrectly predicted molecules, thus further improvement is not justified for this analysis. (We did try to modify the criterion based on eq 1 and introduced the penalty of false prediction; however, the overall results did not change significantly compared with those described. Actually, the issue is rather subtle, as it involves false negative data, and required a separate study, which is underway in our laboratory.)

Setting the cutoff energy at −4 kcal/mol means that a molecule can occupy a site in the protein's interior only if the total energy of interaction of the molecule with its environment is $E_{binding}^{max}$ = −4 kcal/mol or less.

As seen from Figures 2 and 3, the maximum of $\chi(q_H)$ at optimum value of the cutoff −4 kcal/mol is achieved at q_H^* = 0.33 e. This gives the following estimate of the optimum value of the dipole moment:

$$\mu_{protein}^{eff} = \frac{q_H^*}{q_H^{spc}}\mu_{spc} = \frac{0.33e}{0.41e}2.27D = 1.83D \quad (3)$$

It is clear from the nature of the analysis, this value characterizes only the average dipole moment of multiple hydration sites of the proteins considered; yet, this value can be taken as representing the changes that occur with a water molecule within the protein environment. It was shown earlier,[20,32] the magnitude of the effective dipole of water should be understood as a scaled value $\mu^{eff} = \mu/\sqrt{\varepsilon_{el}}$ (ε_{el} =1.78 is the electronic dielectric constant of water).[20] Hence, the actual (average) dipole moment of water in proteins of our data set is

$$\mu_{protein}^{real} = 1.83D\sqrt{\varepsilon_{el}} = 1.83D\sqrt{1.78} = 2.48D \quad (4)$$

For comparison, SPC and TIP3P effective dipoles are around 2.3 D, and the actual water dipole in bulk water, to which SPC and TIP3P models correspond, is estimated to be around 3 D.[22,23] We see, therefore, that the effective dipole of water in the protein is significantly lower than that of SPC or TIP3P, which raises a question of applicability of such models for protein water simulations. In the next section, we provide theoretical analysis of the obtained result.

3. THEORY

In the spirit of the Born solvation model, the solvation energy of a water molecule in a medium with dielectric constant ε_0 can be estimated according to the Kirkwood−Onsager dielectric cavity model,[33,34] which represents a solvent molecule as a point dipole at the center of a spherical cavity surrounded by dielectric medium. According to this model, see details in ref 20, solvation energy is

$$\Delta F_{solv}(\mu) = -\frac{1}{\frac{\pi(2\varepsilon_0+1)(\varepsilon_{el}+2)}{12(\varepsilon_0-1)(\varepsilon_{el}-1)} - 1}\frac{\mu_0^2}{2\alpha} \quad (5)$$

where μ_0 = 1.85 D is the gas value of the water dipole, α = 1.47A^3 is the polarizability of water and ε_{el} =1.78 is its electronic (high-frequency) dielectric constant. Clearly, the "solvation" energy introduced above does not involve entropy contribution (which amounts to about −5 kcal/mol, as follows from the entropy of the liquid state of water), and only refers to interaction energy of water dipole with the environment (for additional discussion of the relation of enthalpy and solvation energy of water, see ref 20). Therefore, this energy can be directly related to enthalpy changes (Here, we neglect unimportant PV term in enthalpy, which is equal to RT and in gas phase is only 0.6 kcal/mol.) Plugging ε_0 = 78 into this formula gives the free energy of the molecule in bulk $\Delta F_{solv}(\mu_l)$ = −10.3 kcal/mol, which is close to the experimental value of vaporization enthalpy −10.5 kcal/mol. Considering protein as a medium with $\varepsilon_{protein}$ = 4, and using the above formula, one can find that energy per molecule in such environment is $\Delta F_{solv}(\mu_{protein})$ = −6.0 kcal/mol. This is the gain in energy associated with the transition of a water molecule from vacuum

dx.doi.org/10.1021/ct500358r | J. Chem. Theory Comput. 2014, 10, 4618−4623

to the protein interior (again, there is no entropy counting here). In the reverse process, one can also say that the evaporation of a molecule from the protein interior requires 6.0 kcal/mol,

$$\Delta H_{\text{protein}} \approx -\Delta F_{\text{solv}}(\mu_{\text{protein}}) = 6 \text{ kcal/mol} \quad (6)$$

This is enthalpy of "evaporation" from the protein medium, understood exactly as that for bulk water, less the *PV* term.

The amount of energy required to bring a molecule from the surrounding bulk into protein medium is the difference between the enthalpy of vaporization and the solvation energy of a water molecule in protein $\Delta H_{\text{vap}} - \Delta H_{\text{protein}} = 10 - 6 = 4$ kcal/mol. For water molecule to stay inside the protein, this gap in energy must be compensated by the additional interaction energy of the molecule with charges of the protein. (For the discussion of the entropy contribution, see the next section.) The energy should not exceed −4 kcal/mol, which is fully consistent with the value $E^{\text{max}}_{\text{binding}} = -4$ kcal/mol found in our analysis.

In the medium with dielectric constant ε_0, due to reaction field polarization the molecular dipole moment of water is changing from its gas-phase value μ_0 to its medium value μ, which according to ref 20 is given by

$$\mu = \frac{\mu_0}{1 - \frac{6(\varepsilon_{el}-1)2(\varepsilon_0-1)}{\pi(\varepsilon_{el}+2)(2\varepsilon_0+1)}} \quad (7)$$

Thus, the dipole moment of water in liquid ($\varepsilon_{\text{liquid}} = 78$) is

$$\mu_1 = \frac{\mu_0}{1 - \frac{6(\varepsilon_{\text{el}}-1)2(\varepsilon_{liquid}-1)}{\pi(\varepsilon_{\text{el}}+2)(2\varepsilon_{\text{liquid}}+1)}} = 3.0D \quad (8)$$

which is in good agreement with experimental and simulation data.[22−25] In the low-dielectric environment of a protein ($\varepsilon_{\text{protein}} = 4$) the dipole moment is lower:

$$\mu_{\text{protein}} = \frac{\mu_0}{1 - \frac{6(\varepsilon_{\text{el}}-1)2(\varepsilon_{\text{protein}}-1)}{\pi(\varepsilon_{\text{el}}+2)(2\varepsilon_{\text{protein}}+1)}} = 2.51 \text{ D} \quad (3.5)$$

This estimate is very close to the one derived from the analysis of Dowser output (2.5 D).

4. CONCLUSIONS AND DISCUSSION

To conclude, our work can be summarized as follows. From the analysis of a set of protein structures, we determined the optimum values of Dowser parameters μ and E_{cutoff}, which essentially are the effective dipole moment and binding energy of water molecules in protein environment. Once the correct cutoff/binding energy of −4 kcal/mol is implemented in Dowser, the amount of predicted crystallographic waters increases to its maximum value.

Rigorously speaking, the binding energy should be discussed in terms of Gibbs energy per molecule, or chemical potential μ; and the criterion should be $\mu_{\text{prot}} < \mu_{\text{bulk}}$, the chemical potential of the hydration site in the protein interior should be lower than that of the bulk water. The difference with the above analysis is the entropy change between the bulk and the protein sites. Dowser neglects this contribution. This is certainly an approximation that perhaps is partially responsible for the low predictability of the program; however, on average, the entropy contributions are expected to be rather small compared with the binding energies that are on the scale of 10 kcal/mol. To see that, one can argue as follows.

For immobile water, the maximum difference (decrease) in entropy between *bulk* (liquid) state water and protein can be estimated as that between liquid and ice. In energy units $RT\Delta S$, this is only 1.4 kcal/mol (from the enthalpy of ice melting). Realistically, internal water *on average* is less immobile than in ice, hence the difference is even smaller, perhaps lower than 1 kcal/mol. Earlier, Dunitz[35] reached the same conclusion using a different argument. Given the crude nature of Dowser approximations, this difference is negligible, compared with enthalpy/energy changes.

For mobile protein water, that is, more mobile than in the bulk, free rotations could give rise to entropy gain (in energy units) of up to 3 kcal/mol, in our estimate; but then, some hydrogen bonds are lost, 5 kcal/mol each; therefore, entropy by itself cannot be the driving force of hydration (unlike in vaporization). Obviously, a subtle balance between energy and entropy in the range 1−2 kcal/mol related to hindered rotation and translation can play a role. However, given the crude nature of Dowser and the averaged characteristic of the determined values, such subtle effects cannot be captured within this approach. Instead, a crude approximation is adopted, and an empirical value of the cutoff energy, attributed to change of interaction energy (i.e., the difference between bulk and protein), is determined from the experimental data. As we showed in the Theory section, the determined empirical value (4 kcal/mol) can be rationalized quantitatively within the Kirkwood−Onsager model as the difference in interaction energy between the liquid bulk and protein environments.

The value of the *effective* dipole moment (averaged over many hydration sites of the protein) maximizing the number of predicted molecules in the protein interior turned out to be 1.83 D, which is to be compared with 2.3 D of approximate SPC or TIP3P models of bulk water. Qualitatively, this change is not unexpected; however, it is interesting that the obtained value can be rationalized quantitatively within a simple charge scaling model.

The effective dipole allows one to estimate the *real* dipole moment in a protein environment by taking into account rescaling of electrostatic charges. According to charge scaling theory,[32] the magnitude of the effective dipole moment should be understood as a scaled value $\mu^{\text{eff}} = \mu/\sqrt{\varepsilon_{\text{el}}}$, which is due to the screening of electrostatic interactions by electronic continuum. This theory explains the difference between the value of water dipole $\mu^{\text{eff}} \approx 2.3$ D used in conventional models of liquid water and the value $\mu \approx 3$ D reported in experimental studies[22] and ab initio simulations.[24,25] In order to find out the real value of the dipole moment, one needs to multiply the effective dipole moment by a factor of $\sqrt{\varepsilon_{\text{el}}}$. Employing the charge rescaling procedure, we find that the average dipole moment of water molecules inside a protein of our data set is 2.5 D, which should be compared with 3.0 D in bulk water.

The value of 2.5 D is fully consistent with a theoretical estimate based on a Kirkwood−Onsager model of a point dipole in dielectric medium. The significant enhancement of the dipole moment from the gas phase value of 1.85 to 2.5 D is caused by the polarization of water in a protein environment due to both protein charges/(hydrogen bonding) and the electronic continuum of the medium. In comparison with the bulk dipole of 3.0 D, the decreased value of 2.5 D indicates that in the protein environment there are on average less than four hydrogen bonds typical for bulk water.

The effective dipole that best described water in the protein environment is 1.83 D, significantly smaller than 2.3 D of SPC or TIP3P models. This indicates that in nonpolarizable MD simulations, the SPC or TIP3P water models may not be suitable for protein interior. These results underscore the importance of the effects of electronic polarizability and charge rescaling in biological and other condensed matter simulations.

AUTHOR INFORMATION

Corresponding Author

*Email: stuchebr@chem.ucdavis.edu.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work has been supported in part by the National Science Foundation (NSF), grant PHY 0646273, and the National Institutes of Health (NIH), grant GM054052.

REFERENCES

(1) Clarke, C.; Woods, R. J.; Gluska, J.; Cooper, A.; Nutley, M. A.; Boons, G.-J. Involvement of Water in Carbohydrate−Protein Binding. *J. Am. Chem. Soc.* **2001**, *123* (49), 12238−12247.

(2) García-Sosa, A. T.; Firth-Clark, S.; Mancera, R. L. Including Tightly-Bound Water Molecules in De Novo Drug Design. Exemplification through the in Silico Generation of Poly(ADP-ribose)polymerase Ligands. *J. Chem. Inf. Model.* **2005**, *45* (3), 624−633.

(3) García-Sosa, A. T.; Mancera, R. L. Free Energy Calculations of Mutations Involving a Tightly Bound Water Molecule and Ligand Substitutions in a Ligand−Protein Complex. *Mol. Inf.* **2010**, *29* (8−9), 589−600.

(4) Kadirvelraj, R.; Foley, B. L.; Dyekjær, J. D.; Woods, R. J. Involvement of Water in Carbohydrate−Protein Binding: Concanavalin A Revisited. *J. Am. Chem. Soc.* **2008**, *130* (50), 16933−16942.

(5) Mancera, R. De Novo Ligand Design with Explicit Water Molecules: An Application to Bacterial Neuraminidase. *J. Comput.-Aided Mol. Des.* **2002**, *16* (7), 479−499.

(6) Michel, J.; Tirado-Rives, J.; Jorgensen, W. L. Energetics of Displacing Water Molecules from Protein Binding Sites: Consequences for Ligand Optimization. *J. Am. Chem. Soc.* **2009**, *131* (42), 15403−15411.

(7) Mikol, V.; Papageorgiou, C.; Borer, X. The Role of Water Molecules in the Structure-Based Design of (5-Hydroxynorvaline)-2-Cyclosporin: Synthesis, Biological Activity, and Crystallographic Analysis with Cyclophilin A. *J. Med. Chem.* **1995**, *38* (17), 3361−3367.

(8) Wong, S. E.; Lightstone, F. C. Accounting for Water Molecules in Drug Design. *Expert Opin. Drug Discovery* **2010**, *6* (1), 65−74.

(9) Carugo, O.; Bordo, D. How Many Water Molecules Can Be Detected by Protein Crystallography? *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1999**, *55* (2), 479−483.

(10) Davis, A. M.; Teague, S. J.; Kleywegt, G. J. Application and Limitations of X-ray Crystallographic Data in Structure-Based Ligand and Drug Design. *Angew. Chem., Int. Ed.* **2003**, *42* (24), 2718−2736.

(11) Ernst, J.; Clubb, R.; Zhou, H.; Gronenborn, A.; Clore, G. Demonstration of Positionally Disordered Water within a Protein Hydrophobic Cavity by NMR. *Science* **1995**, *267* (5205), 1813−1817.

(12) Dowser. http://danger.med.unc.edu/hermans/dowser/dowser.htm (accessed Aug. 25, 2011).

(13) Dowser Plugin. http://www.ks.uiuc.edu/Research/vmd/plugins/dowser/ (accessed March 5, 2013).

(14) Dowser—SBGrid Consortium. http://www.sbgrid.org/software/title/Dowser (accessed February 3, 2014).

(15) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graph.* **1996**, *14* (1), 33-8−27-8.

(16) *CRC Handbook of Chemistry and Physics (Internet Version 2010)*, 90 ed.; Lide, D. R., Ed.; CRC Press/Taylor and Francis: Boca Raton, FL, 2010.

(17) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. Interaction Models for Water in Relation to Protein Hydration. In *Intermolecular Forces*; Pullmann, B., Ed.; D. Reidel Publishing Company: Dordrecht, 1981; pp 331−342.

(18) 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.

(19) Leontyev, I. V.; Stuchebrukhov, A. A. Polarizable Mean-Field Model of Water for Biological Simulations with AMBER and CHARMM Force Fields. *J. Chem. Theory Comput.* **2012**, *8*, 3207−3216.

(20) Leontyev, I. V.; Stuchebrukhov, A. A. Electronic Polarizability and the Effective Pair Potentials of Water. *J. Chem. Theory Comput.* **2010**, *6* (10), 3153−3161.

(21) Lovas, F. J. Microwave Spectral Tables. II. Triatomic Molecules. *J. Phys. Chem. Ref. Data* **1978**, *7* (4), 1445−1750.

(22) Badyal, Y. S.; Saboungi, M. L.; Price, D. L.; Shastri, S. D.; Haeffner, D. R.; Soper, A. K. Electron Distribution in Water. *J. Chem. Phys.* **2000**, *112*, 9206−9208.

(23) Gregory, J. K.; Clary, D. C.; Liu, K.; Brown, M. G.; Saykally, R. J. The Water Dipole Moment in Water Clusters. *Science* **1997**, *275*, 814−817.

(24) Silvestrelli, P. L.; Parrinello, M. Structural, Electronic, and Bonding Properties of Liquid Water from First Principles. *J. Chem. Phys.* **1999**, *111* (8), 3572−3580.

(25) Sharma, M.; Resta, R.; Car, R. Dipolar Correlations and the Dielectric Permittivity of Water. *Phys. Rev. Lett.* **2007**, *98* (24), 247401.

(26) Nakamura, H.; Sakamoto, T.; Wada, A. A Theoretical Study of the Dielectric Constant of Protein. *Protein Eng.* **1988**, *2* (3), 177−183.

(27) Kukic, P.; Farrell, D.; McIntosh, L. P.; Garcia-Moreno, E. B.; Jensen, K. S.; Toleikis, Z.; Teilum, K.; Nielsen, J. E. Protein Dielectric Constants Determined from NMR Chemical Shift Perturbations. *J. Am. Chem. Soc.* **2013**, *135* (45), 16968−76.

(28) Pitera, J. W.; Falta, M.; van Gunsteren, W. F. Dielectric Properties of Proteins from Simulation: The Effects of Solvent, Ligands, pH, and Temperature. *Biophys. J.* **2001**, *80* (6), 2546−2555.

(29) Sleigh, S. H.; Seavers, P. R.; Wilkinson, A. J.; Ladbury, J. E.; Tame, J. R. H. Crystallographic and Calorimetric Analysis of Peptide Binding to OppA Protein. *J. Mol. Biol.* **1999**, *291* (2), 393−415.

(30) Ross, G. A.; Morris, G. M.; Biggin, P. C. Rapid and Accurate Prediction and Scoring of Water Molecules in Protein Binding Sites. *PloS One* **2012**, *7* (3), e32036.

(31) Rossato, G.; Ernst, B.; Vedani, A.; Smieško, M. AcquaAlta: A Directional Approach to the Solvation of Ligand−Protein Complexes. *J. Chem. Inf. Model.* **2011**, *51* (8), 1867−1881.

(32) Leontyev, I. V.; Stuchebrukhov, A. A. Accounting for Electronic Polarization in Non-Polarizable Force Fields. *Phys. Chem. Chem. Phys.* **2011**, *13* (7), 2613−2626.

(33) Kirkwood, J. G. The Dielectric Polarization of Polar Liquids. *J. Chem. Phys.* **1939**, *7* (10), 911−919.

(34) Onsager, L. Electric Moments of Molecules in Liquids. *J. Am. Chem. Soc.* **1936**, *58*, 1486−1493.

(35) Dunitz, J. D. The Entropic Cost of Bound Water in Crystals and Biomolecules. *Science* **1994**, *264* (5159), 670.