CONF-961202--35

Title: HYDRATION OF BIOLOGICAL MACROMOLECULES: FROM
SMALL SOLUTES TO PROTEINS AND NUCLEIC ACIDS

Author(s): Shekhar Garde, T-10, Los Alamos National Laboratory
Gerhard Hummer, T-10, Los Alamos National Laboratory
Michael E. Paulaitis, University of Delaware, DE
Angel E. Garcia, T-10, Los Alamos National Laboratory

RECEIVED

JAN 21 1997

OSTI

Submitted: Material Research Society
Proceedings-Symposium on Statistical Mechanics on Biological
System

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

MASTER

Theoretical Biology and Biophysics Group (T-10)
MS K710, T-10, Los Alamos, NM 87545

Los Alamos
NATIONAL LABORATORY

Los Alamos National Laboratory, an affirmative action/equal opportunity employer, is operated by the University of California for the U.S. Department of Energy under contract W-7405-ENG-36. By acceptance of this article, the publisher recognizes that the U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or to allow others to do so, for U.S. Government purposes. Los Alamos National Laboratory requests that the publisher identify this article as work performed under the auspices of the U.S. Department of Energy. The Los Alamos National Laboratory strongly supports academic freedom and a researcher's right to publish; as an institution, however, the Laboratory does not endorse the viewpoint of a publication or guarantee its technical correctness.

DISCLAIMER

**Portions of this document may be illegible
in electronic image products. Images are
produced from the best available original
document.**

Hydration of Biological Macromolecules: From Small Solutes to Proteins and Nucleic Acids.

Shekhar Garde,^{1,2} Gerhard Hummer,¹ Michael E. Paulaitis,^{2,3} and Angel E. García^{1,*}

¹*Theoretical Biology and Biophysics T-10, MS K710, Los Alamos National Laboratory, Los Alamos, NM 87545*

²*Center for Molecular and Engineering Thermodynamics, Department of Chemical Engineering, University of Delaware, Newark, DE 19716*

³*Department of Chemical Engineering, Johns Hopkins University, Baltimore, MD 21218.*

We present a method that uses two- and three-particle correlation functions between solute atoms and water molecules to approximate the density profile of water surrounding biomolecules. The method is based on a potential of mean force expansion and uses X-ray crystallography, NMR, or modeling structural input information on the biomolecule. For small hydrophobic solutes, we have calculated entropies of hydration using the predicted water densities that are in good agreement with experimental results. We have also predicted the hydration of the *catabolite activator protein*-DNA complex. The method is extremely efficient and makes possible the study of hydration of large biomolecules within CPU minutes.

1 Introduction

The unique balance of forces underlying biological processes such as protein folding, molecular recognition, and formation of biological membranes owes its origin in part to the surrounding aqueous medium.[1] Hydrophobic interactions, in particular, are believed to contribute to the overall stability of folded proteins, protein-ligand complexes, and biological membranes. The hydrophobic driving force for self-assembly is mainly entropic in origin, and is directly connected to the inhomogeneous structure of water in the vicinity of hydrophobic solutes.[2] Thus, quantifying structural hydration of biomolecules and relating water structure to the key thermodynamic properties of hydration constitute important steps towards understanding of the role of water in biomolecular structure-function relationship.

Experimental information about the hydration of biomolecules at atomic resolution is obtained mainly from X-ray and neutron diffraction,[3] and NMR spectroscopic [4] techniques. Even though these techniques are widely used, unambiguous assignment of water molecules on the surface and in the interior of biomolecules is complicated due to methodological difficulties. Further, the disorder of water molecules hydrating biomolecules in different unit cells of crystals presents difficulties in the interpretation of experimental results.[5,6] This disorder is quantified best by an inhomogeneous density of water molecules in space. The experimental analysis is however limited to identifying the regions of high localization of water molecules.

Theoretical methods for the detailed study of the structural hydration of biomolecules rely mostly on the molecular dynamics (MD) and Monte Carlo (MC) computer simulation techniques.[7] Computer simulations of hydration of small molecules have advanced our fundamental understanding of solvation phenomena. However, the computational effort needed to obtain satisfactory statistical precision increases significantly when applied to the study of biomolecular hydration. Further, analysis of systems like protein-DNA complexes would need simulations of the individual protein, the DNA, and the complex to understand the role of hydration in complexation.

We present here an alternative approach to predicting position-dependent water densities near biomolecules. Our method is based on the potential-of-mean-force (PMF) expansion [8] applied previously to hydration studies of B-DNA,[9] an α -helix,[10] t-RNA, the photosynthetic reaction center, an antigen-antibody complex,[11,12] and a host of other systems. We will show that the water densities surrounding biomolecules can be approximated by a limited number of well-defined, lower-order correlation functions accurately calculated from simulations of small molecules in water. The efficiency of the method allows analysis of hydration of biomolecules in CPU minutes rather than weeks. For purely hydrophobic molecular solutes we relate the calculated water densities to the entropies of hydration *via* an entropy expansion to show the feasibility of accurately estimating macroscopic thermodynamic properties from water density profiles. Finally, we analyse of the hydration of the interface of the *catabolite activator protein* (CAP)-DNA complex.

2 Theory and Computer Simulations

The hydration of biomolecules can be quantified by the equilibrium position-dependent density of water molecules surrounding them. The biomolecule is characterized by a set of atomic coordinates $\{\mathbf{s}_{i\alpha}\}$, where α indicates the type of atom, $\alpha = 1, \dots, M$, and $i = 1, \dots, N_\alpha$. The conditional density of water molecules at a position $\mathbf{r}$ (given that biomolecule is fixed in a given configuration $\{\mathbf{s}_{i\alpha}\}$) is then given by,[11]

$$\rho(\mathbf{r}|\{\mathbf{s}_{i\alpha}\}) = \rho_0 \frac{g^{(1;\{N_\alpha\})}(\mathbf{r}, \{\mathbf{s}_{i\alpha}\})}{g^{(\{N_\alpha\})}(\{\mathbf{s}_{i\alpha}\})}. \quad (1)$$

The n -particle distribution function, $g^{(n)}$, is related to the n -particle potential of mean force, $W^{(n)}$, by $g^{(n)} = \exp(-W^{(n)}/kT)$, and ρ_0 is the bulk density of water. Successful prediction of density thus depends on how accurately one can approximate the n -particle PMF. For strictly nonpolar molecular solutes, water structure is found to be only locally sensitive to the structural details of the solute. Thus, the water density around nonpolar solutes can be approximated simply by its correlations with one or two nearest solute sites.[13,14] Namely,

$$\rho^{(n,1)}(\mathbf{r}|\mathbf{r}_1, \dots, \mathbf{r}_n) \approx \rho_0 g(\mathbf{r}|\mathbf{r}_j), \quad (2)$$

$$\rho^{(n,2)}(\mathbf{r}|\mathbf{r}_1, \dots, \mathbf{r}_n) \approx \rho_0 g(\mathbf{r}|\mathbf{r}_j, \mathbf{r}_k), \quad (3)$$

with j and k such that $|\mathbf{r} - \mathbf{r}_j| = \min_{l=1, \dots, n} |\mathbf{r} - \mathbf{r}_l|$ and $|\mathbf{r} - \mathbf{r}_k| = \min_{m=1, \dots, n; m \neq j} |\mathbf{r} - \mathbf{r}_m|$. Eqs. (2) and (3) are defined, respectively, as the one- and two-site proximity approximations.

In contrast to nonpolar molecules, biomolecules are a complex mix of a variety of sites. In this case, we formally expand $W^{(n)}$ in terms of multiparticle potentials of mean force. Difficulties in calculating correlation functions of order four and higher necessitates truncation of this expansion at the three-particle level. Further, geometric and energetic considerations of interactions of nitrogen, oxygen, and sulphur atoms with water molecules allow us to effectively group these atoms into one class of solute sites. That is, their effect on ordering nearby water molecules is assumed to be similar to that of water oxygens. Also, the effect on water structure of nonpolar carbon atoms, represented as methyl [Me] sites, is treated using the one-site proximity approximation. That these approximations are reasonable for applications to biomolecules has been shown elsewhere.[9-12] These approximations render the complex problem of biomolecular hydration tractable. The following approximate expression

for the conditional water density is obtained.

$$\rho(\mathbf{r}|\mathbf{s}_1, \dots, \mathbf{s}_n) \approx \rho_0 \prod_{j=1}^n g^{(2)}(\mathbf{r}, \mathbf{s}_j) \prod_{j=1}^{n-1} \prod_{k=j+1}^n \frac{g^{(3)}(\mathbf{r}, \mathbf{s}_j, \mathbf{s}_k)}{g^{(2)}(\mathbf{r}, \mathbf{s}_j)g^{(2)}(\mathbf{s}_j, \mathbf{s}_k)g^{(2)}(\mathbf{s}_k, \mathbf{r})} \times g_{Me-O}(\mathbf{r}, \mathbf{s}_{C-prox}). \quad (4)$$

Additional effects, such as polar hydrogens forming directional hydrogen bonds, can also be incorporated into this expression.[11]

Application of eqn. (4) to biomolecules requires the calculation of the O-O-O triplet correlation function, $g^{(3)}$, as well as Me-O pair correlation function g_{Me-O} . The $g^{(3)}$ is calculated from simulations of SPC water.[9,18] Simulations were also performed for a single Me and the Me-Me dimer in water where Me-Me dimer was held fixed at separations between 0.12 to 0.66 nm. [14] These simulations were used to calculate Me-O pair, and Me-Me-O triplet correlations required in the proximity approximations, as well as eqn. (4).

3 Results and Discussion

3.1 Hydration of nonpolar molecules

Figure 1: Spherically averaged water-oxygen density in units of bulk density as a function of distance r from the center of benzene calculated from simulation (symbols) and the two-site proximity approximation (dashed line). Inset shows the C_{20} and C_{40} coefficients in the spherical harmonic expansion of the density.[14]

Figure 1 shows a comparison of water-oxygen densities calculated from MC simulation of benzene (represented by six Me sites in a planar configuration) with densities calculated from the two-site proximity approximation. [14] The positions and heights of both peaks corresponding to the first and the second hydration shells, respectively, are well reproduced. The symmetry of the density profile about the plane of benzene, captured by the higher order coefficients in the spherical harmonic expansion, is also very well reproduced by the two-site proximity approximation. Application to other model solutes also gives similar results.[14]

This suggests a key feature of the hydration of nonpolar solutes is that water structure in their vicinity is only locally sensitive to the structural details of the nonpolar solute.

Table 1. Standard entropies of hydration calculated using an entropy expansion.

solute	$\Delta S_{\text{hydration}}^*/(\text{cal/mol/K})$	
	calculated	expt
ethane	-18.0	-20.1
propane	-21.7	-22.8
<i>n</i> -butane ^a	-23.3	-26.0
benzene	-28.8	-21.2

^a56/44 % *trans/gauche* mixture.

The calculated entropies include two terms corresponding to positional and orientational correlations of water molecules with nonpolar solute approximated using two- and one-site proximity approximations, respectively. The latter amounts to an orientational entropy of 0.35 cal/mol/K per water molecule in the first hydration shell.[15]

We have used the water density profiles obtained from proximity approximations to calculate standard entropies of hydration for various nonpolar solutes.[13,16,17] The results are given in Table 1. The entropies of hydration are large and negative and are in good agreement with experimental results for alkanes. A value significantly more negative than the experimental value is obtained for benzene. Note, however, that the description of benzene as six overlapping Me sites is a severe approximation for this molecule which neglects its complex electronic structure.[19]

The PMF between two methane molecules in water and water contribution to the conformational equilibria of *n*-butane and *n*-pentane have also been predicted from water densities obtained using the proximity approximations to show that hydration favors the more compact conformations, as expected. [15]

3.2 Ice-water interface

Eq. (4) was applied to the ice/water interface. The ice phase is described by fixing SPC water oxygens at ideal hexagonal-ice I_h lattice positions while allowing for molecular reorientations representing thermal disorder in the ice phase. The basal plane of the ice layer is covered by liquid water. Figure 2 shows a comparison of the oxygen density profiles predicted by the PMF expansion truncated at the level of pair and triplet correlations with calculations from explicit MC computer simulations.[9] It is seen that truncation of the PMF expansion at the level of pair correlations results in water densities much higher than those observed in MC simulations in the region close to the basal plane ($Z < 0.2\text{nm}$). Including triplet correlations improves the agreement. The position as well as the height of the first peak are well reproduced. The agreement at longer distances is qualitative. Clearly, information on triplet correlations is needed for identifying high density regions accurately. We thus include the terms up to triplet correlations in the PMF expansion eqn. (4) for the applications to the biomolecular hydration.

Figure 2: Water-oxygen density at the ice/water interface in units of the bulk- water density, ρ_0 . Z is the distance from the closest plane of oxygen atoms in the ice phase. (+) PMF expansion including only pair correlations; ($\diamond$) PMF expansion including both pair and triplet correlations. Solid lines: explicit MC simulations using different system sizes and methods for Coulombic interactions.[9]

3.3 CAP-DNA complex

The catabolite activator protein (CAP), also known as cyclic adenosine monophosphate receptor protein (CRP), plays an important role in mediating transcription activation of several genes in *E. coli*. The CAP-cAMP complex binds, as a dimer, to the specific sequences in several *E. coli* promoter regions.[20] The activation process is complex, involving not only the binding of cAMP, but also the interactions of CAP with the DNA [21, 22] and with the RNA polymerase.[23] Atomic resolution structure of CAP bound to the consensus DNA sequence has been determined by X-ray crystallographic studies.[21, 22] Upon binding, the DNA molecule is bent at approximately 90° near the position where helix-turn-helix motif of each CAP monomer binds. Previous crystallographic studies of protein-DNA complexes have shown that the sequence recognition can also be mediated by water molecules bridging between hydrogen-bond acceptor groups in the protein and the DNA.[24] A recent moderately high resolution (2.7 Å) crystal structure of the CAP-DNA complex also shows water mediated interactions. To elucidate the extent of water mediated recognition, we have studied the hydration of the regions in the vicinity of the recognition sites by means of a local water density calculation using the PMF expansion. Here we present the results for one of the recognition motifs.

Points corresponding to a high local density of water molecules near the binding region in the complex are indicated in the panels A and B. The binding region is seen to be considerably hydrated. We have also calculated water densities near the isolated protein and DNA molecule. Both the amino acids (panel D) as well as the DNA bases (panel C) comprising the recognition sites are considerably hydrated. Thus, water molecules corresponding to high

Figure 3: Hydration of the CAP-DNA complex. Regions of water density $\rho(r)/\rho_0 > 3$ and $\rho(r)/\rho_0 > 5$ are identified in each panel by small and medium sized dark gray spheres. Panel (A) and (B) show two different views of the CAP (white) and DNA (gray) complex at the recognition motif (5'-TGTGA-3'/3'-ACACT-5'). Panel (C) shows the hydration of the DNA recognition site in the absence of CAP. (white: DNA bases comprising the recognition site, gray: other bases). Panel (D) shows the hydration of the CAP recognition site in absence of the DNA. (white: amino acids comprising recognition site, gray: other amino acids.) The crystal structure from pdb file 1ber [22] was used for the calculations.

local densities near the isolated recognition sites have to be removed upon binding. The release of bound water represents an entropic gain upon binding while the water molecules that hydrate the interface of the complex (as shown in panel A and B) give a favorable enthalpic contribution, *e.g.*, by mediating hydrogen bonds.

Our results demonstrate that we can efficiently investigate medium-to-low resolution structures (> 0.2 nm crystallographic resolution) where water molecules are difficult to assign. Further, we get the density profiles of water hydrating biomolecules which represents a more complete picture compared to assignment of only individual water molecules. We

have studied the hydration of the recognition sites of the CAP and the DNA for which the binding free energies have been measured previously.[25] Based on our calculations, it is our hypothesis that water-mediated interaction in this complex can partly account for the observed specificity.

4 Conclusions

We have presented a method for studying the hydration of biological macromolecules. Detailed simulations of the hydration of purely hydrophobic solutes and the ice-water interface indicate that the water structure is only locally sensitive to the structure of hydrophobic solutes and that information at the level of three-particle correlations is essential for reproducing water density profiles observed in simulations of polar interface. These results have been incorporated into a truncated PMF expansion for describing biomolecular hydration.

An advantage of this approach is that the database of well-defined, lower order correlation functions required for the determination of water densities around a macromolecule must be obtained only once. This makes the application of the PMF expansion extremely efficient. The direct connection to the statistical mechanics also allows us to include any additional information on correlations (*e.g.* directional hydrogen bonding interactions) in an unambiguous, systematic manner.

The PMF expansion formalism can be applied to hydration studies in both crystals or solutions to give average density profiles rather than assigning individual water molecules. The method can also be applied to medium-to-low resolution X-ray structures for the assignment of water molecules in the process of refinement.

The water density profiles surrounding hydrophobic molecules have also been used to calculate entropies of hydration. Similar application to an antigen-antibody complex was performed recently using water densities calculated from the PMF expansion. Qualitative correlation with experimental entropies of binding was observed.[11,26] Accurate calculations of free energies of binding are complicated by the orientational component in the entropy and the electrostatic part in the enthalpy. Thus, more detailed calculations on biological systems are needed before quantitative agreement with experiments and possibly predictions of thermodynamic quantities can be achieved. However, our results for the CAP-DNA complex presented here underscore the importance of water in biomolecular interactions.

Acknowledgements

G.H. would like to thank Dr. D. M. Soumpasis for many discussions on the role of water and ions in the biomolecular systems.

References

- [1] J. A. Rupley and G. Careri, *Adv. Protein. Chem.*, **41**, 37, (1991).
- [2] M. E. Paulaitis, S. Garde, and H. S. Ashbaugh, *Curr. Opin. Coll. Int. Sci.*, **1**, 376, (1996).
- [3] H. Savage and A. Wlodawer, *Meth. Enzymol.*, **127**, 162, (1986).
- [4] G. Otting, E. Liepinsh, and K. Wüthrich, *Science*, **254**, 974, (1991).

- [5] E. Westhf, *Annu. Rev. Biophys. Biophys. Chem.*, **17**, 125, (1988).
- [6] X-J. Zhang and B. W. Matthews, *Protein Sci.* **3**, 1031, 1994.
- [7] C. L Brooks III and M. Karplus, *Meth. Enzymol.*, **127**, 369, (1986).
- [8] D. M. Soumpasis, in Computation of Biomolecular Structures, edited by D. M. Soumpasis and T. M. Jovin (Springer, Berlin, 1993) p. 223.
- [9] G. Hummer and D. M. Soumpasis, *Phys. Rev. E.*, **49**, 591, (1994); *ibid.* **50**, 5085, (1994).
- [10] G. Hummer and A. E. Garca, *Prot. Struct. Funct. Gen.*, (1996), in press.
- [11] G. Hummer, A. E. Garca, and D. M. Soumpasis, *Faraday Discuss.*, **103**, (1996), in press.
- [12] G. Hummer, A. E. Garca, and D. M. Soumpasis, *Biophys. J.*, **68**, 1639, (1995).
- [13] H. S. Ashbaugh and M. E. Paulaitis, *J. Phys. Chem.*, **100**, (1996).
- [14] S. Garde, G. Hummer, A. E. Garca, L. R. Pratt, and M. E. Paulaitis, *Phys. Rev. E.*, **53**, R4310, (1996).
- [15] S. Garde, G. Hummer, and M. E. Paulaitis, *Faraday Discuss.*, **103**, (1996), in press.
- [16] T. Lazaridis and M. E. Paulaitis, *J. Phys. Chem.*, **96**, 3847, (1992); *ibid.* **98**, 635, (1994).
- [17] M. E. Paulaitis, H. S. Ashbaugh, and S. Garde, *Biophys. Chem.*, **51**, 349, (1994).
- [18] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, and J. Hermans, in Intermolecular Forces: Proceedings of the 14th Jerusalem Symposium on Quantum Chemistry and Biochemistry, edited by B. Pullman, (Reidel Dordrecht, The Netherlands), p. 331, (1981).
- [19] W. L. Jorgensen and D. L. Severance, *J. Am. Chem. Soc.*, **112**, 4768, (1990).
- [20] J. L. Boshford and J. G. Harman, *Microbiol. Review*, **56**, 100, (1992).
- [21] S. Schultz, G. Shields, and T. Steitz, *Science*, **253**, 1001, (1991).
- [22] G. Parkinson, C. Wilson, A. Gunasekera, Y. W. Ebright, R. H. Ebright, and H. M. Berman, *J. Mol. Biol.* **260**, 395, (1996).
- [23] R. H. Ebright, *Mol. Microbiol.*, **8**, 797, (1993).
- [24] B. F. Luisi, W. X. Xu, Z. Otinowski, L. P. Freedman, K. R. Yamamoto, and P. B. Sigler, *Nature*, **352**, 495, (1991).
- [25] A. Gunasekera, Y. W. Ebright, and R. H. Ebright, *J. Biol. Chem.*, **267**, 14713, (1992).
- [26] T. N. Bhat, G. A. Bentley, G. Boulot, M. I. Green, D. Tello, W. Dall'Acqua, H. Souchon, F. P. Schwartz, R. A. Mariuzza, and R. A. Poljak, *Proc. Nat'l. Acad. Sci. USA.*, **91**, 1089, (1994).