STRUCTURE AND THERMODYNAMIC PROPERTIES OF AQUEOUS SOLUTIONS OF SMALL MOLECULES AND PROTEINS

Harold A. Scheraga
Baker Laboratory of Chemistry, Cornell University, Ithaca, NY 14853, USA

Abstract — The structures and reactivities of proteins are markedly influenced by water. In conformational energy calculations, the hydration of proteins is treated by a hydration-shell model. Originally, the shell model was parameterized with experimental data, but more recently empirical potentials were introduced to obtain these parameters by Monte Carlo and molecular dynamics techniques. These simulation methods are being used to obtain structural, dynamic and thermodynamic properties of liquid water and aqueous solutions of nonpolar and polar solutes. Calculations have also been carried out on aqueous solutions containing more than one solute molecule, e.g., several nonpolar molecules to investigate the hydrophobic interaction. An array processor, supported by a minicomputer host, has made it possible to treat large systems and to make long runs in Monte Carlo and molecular dynamics simulations. Nevertheless, problems still remain in obtaining convergence to the proper thermodynamic limit and in computing free energies and entropies of hydration. This paper is concerned with the history of these developments and with the problems that remain to be solved in order to provide an accurate treatment of the effect of water on protein structure and reactivity.

INTRODUCTION

Our interest in the properties of liquid water and aqueous solutions stems largely from the need to understand the role that water plays in the stability and interactions of proteins. Among other things, water leads to a tendency for the nonpolar and polar groups of proteins to lie preferentially on the inside and outside, respectively, of the molecule.

In computations on protein conformation, we have found it most convenient to include the effect of water by means of a hydration-shell model, the development of which has been described by Paterson et al. (1). A hydration shell (having the thickness of one water molecule) and an associated free energy of hydration are assigned to each atom or group of atoms of a polypeptide chain. As the conformation of the protein is altered (to minimize its conformational energy), water must be eliminated from the hydration shells whenever they overlap in the manner indicated in Figure 1. The free energy change accompanying this "dehydration" depends on the degree of overlap (which determines the amount of water eliminated from the hydration shells) and on the free energy of hydration of the various groups. It is, therefore, essential to obtain information about the structural aspects of the hydration shells and about the free energies of hydration. The sizes of the hydration shells and the free energies of hydration of the various functional groups were estimated (2-5) from a variety of physical chemical data, but experience with these computations has demonstrated a need to obtain more accurate values of these quantities. This need has stimulated considerable activity in the fields of Monte Carlo and molecular dynamics simulations of liquid water and aqueous solutions in order to acquire more information about protein hydration (6). This paper is concerned with these problems.

BRIEF HISTORY

An indication of the progress that has been made in studies of water, aqueous solutions and protein hydration can be obtained by tracing the evolution of our knowledge in a series of reviews of the subject (1, 6, 7-14). Early
Figure 1. Schematic diagram illustrating three stages of the approach of two groups 1 and 2. (A) No overlaps. (B) Overlap of the hydration spheres of groups 1 and 2; the free energy of hydration of both groups is unaltered. (C) Overlap of the van der Waals sphere of group 2 with the hydration sphere of group 1; the free energy of hydration of group 2 is unaltered, but that of group 1 is changed because of removal of water in the shaded region of the hydration shell of group 1. When groups 1 and 2 approach even more closely, the free energy of hydration of group 2 would change as well. From reference 4.

studies of hydrophobic interactions were based on statistical mechanical treatments of models of water and aqueous solutions of hydrocarbons (15-21). A summary of the physical picture of hydrophobic interactions that emerged from these studies is given in reference 13. These ideas were applied in investigations of the conformations of synthetic polyamino acids and proteins, as will be discussed in the last section of this article. They were also incorporated into models of non-ionic (22) and ionic (23) micelles.

The theoretical parameters for hydrophobic interactions have been verified by various types of experiments (1,7,12). These include adsorption of compounds containing nonpolar groups by nonpolar resins (24,25), studies of equilibria of binary complexes of molecules containing nonpolar group (26-28), transfer of compounds containing nonpolar groups from H2O to D2O (29), and volume changes upon mixing of water with compounds containing nonpolar groups (30). These parameters, in turn, were used (together with empirical data for polar groups) in preliminary applications of the hydration shell model (1-5,31,32), as mentioned in the Introduction.

SIMULATION OF LIQUID WATER

More recently, the reliance on models has been abandoned, and attempts have been made to deduce the properties of solutions by use of empirical potential functions, primarily in computer simulations. A variety of empirical potential functions (reviewed in reference 6) have been used in Monte Carlo and molecular dynamics simulations of liquid water and dilute aqueous solutions. Those most frequently used are the ST2 potential of Stillinger and Rahman (33), the MCY potential of Matsuoka, Clementi and Yoshimine (34), and the EPEN potential (Empirical Potential based on the interactions of Electrons and Nuclei) (35,36). Long simulation runs are required to obtain statistically meaningful results, and the recent introduction of array processors (37) has facilitated such computations.
The MCY potential, for example, has been used in Monte Carlo (38-42) and molecular dynamics (43) simulations of liquid water, and the radial distribution functions from the molecular dynamics simulations are compared with experimental results in Figure 2. Both types of simulations (38-44) led to good agreement between computed and experimental thermodynamic and transport properties, and this agreement appears to hold at low temperatures, i.e. for supercooled water (W.J. Peer and H.A. Scheraga, work in progress). Recent neutron scattering experiments on water at various temperatures (45,46) provide additional data with which to test the results of Monte Carlo and molecular dynamics simulations.

Figure 2. Molecular-dynamics simulated and experimental oxygen-oxygen radial distribution functions. From reference 43.

The "agreement" described above hides a fundamental difficulty that has been encountered in Monte Carlo runs, but has not yet been explored in molecular dynamics simulations. This problem is one of convergence. There are two aspects to the convergence problem. First, in a Monte Carlo simulation of pure water, there was a small periodic change in the energy of the system which occurred about once in every 1.5 million attempted moves (in a very long run of 4 million attempted moves) (40). This has the largest effect on a fluctuation quantity such as the heat capacity. The second aspect of the convergence problem, which arises not so much in computations on a pure solvent but rather in simulations of dilute solutions, involves the radial distribution function. The problem can be demonstrated in Monte Carlo calculations on a pure solvent by treating it as a solution, with one water molecule serving as the "solute" and the remaining water molecules as the solvent. In calculating the radial distribution function for a pure solvent containing N molecules, one samples N(N-1)/2 distances per configuration, in averaging over N molecules, whereas the computation of a solute-solvent radial distribution function involves a sampling of only (N-1) solute-solvent distances per configuration; hence, one would expect to obtain a larger
statistical error in the computation of solution properties (47). Mehrotra et al. (48,49) made a detailed study of this problem and found that the computed radial distribution function varied by an amount greater than the statistical uncertainty as different water molecules were designated as the "solute." Only when they averaged over all "solute" water molecules did they obtain the correct radial distribution function. They circumvented this convergence problem, in anticipation of intended calculations on solutions, by introducing special sampling techniques [force bias method (50) and preferential sampling (51,52)]; this greatly improved the agreement between the radial distribution function calculated by regarding one water molecule as the "solute" and that calculated by averaging over the N molecules, i.e., by taking account of N(N-1)/2 distances per configuration. This problem has also been treated (53) by accelerating the convergence in the Metropolis (54) Monte Carlo procedure by optimizing the maximum step sizes used by the algorithm, as well as by using the technique of preferential sampling (51,52). While these devices improve the rate of convergence in a Monte Carlo run, there is still considerable need for improvement when treating dilute solutions. An important consideration in assessing convergence is the criterion used to determine how well the entire phase space has been sampled during the simulation (W.J. Peer and H.A. Scheraga, work in progress). It is not yet clear to what extent the results of simulations of aqueous solutions, described below, are affected by this convergence problem.

The effects of other aspects of the calculations, such as boundary conditions, truncation of potential, etc., on the results of the computer simulations have been carefully evaluated (55-62). This adds confidence that the simulations, using only a few hundred molecules, faithfully reproduce the macroscopic behavior of a liquid. Furthermore, intercomparison among the results of computer simulations using different potential functions for water aids in recognizing those general features of the simulation models that are applicable to real water.

SIMULATIONS OF AQUEOUS SOLUTIONS

Early treatments of aqueous hydrocarbon solutions were based on a clathrate model for the hydrophobic hydration of a single molecule (16), and the hydrophobic interaction was considered to be a partial reversal of the solution process as the hydrated nonpolar groups came into contact (17). More recently, the clathrate model of hydrophobic hydration has been amply demonstrated by both Monte Carlo and molecular dynamics simulations of a nonpolar solute molecule surrounded by water molecules (47,53,63—70). In order to isolate the physical origin of hydrophobic hydration, which appears to be due to a volume-exclusion effect, very long molecular dynamics runs (~70 psec) were carried out (70) on a system consisting of four nearly-hard-sphere solute molecules (obeying a truncated shifted Lennard-Jones potential), and 339 water molecules, the latter being modeled with the MCY potential. Comparison of the radial distribution functions in Figures 3 and 4 demonstrates the existence of a clathrate-like structure around an isolated solute molecule. Analysis of the orientational preferences of the water molecules around the solute (70) indicated a clear tendency for the water dipoles of the innermost hydration shell to point inward toward the solute, an arrangement consistent with a clathrate-like structure which, of course, is only a time-averaged one.

Figure 3. Comparison of solute-oxygen radial distribution function (gAO) with that for oxygen-oxygen in pure water. From reference 70.
Figure 4. Comparison of solute-hydrogen radial distribution function ($g_{AH}$) with that for oxygen-hydrogen in pure water. From reference 70.

Monte Carlo calculations have also been carried out for polar non-electrolyte solute molecules in water, viz. formaldehyde (71), glycine (72), methanol (73, 74), and ethanol (75). These calculations provided information about the distribution of water around these solutes and, in the case of methanol (73), that the most stable conformations of the solute in the gas phase and in aqueous solution differ.

Extension of these simulations to ions in water faces the additional problem of the long-range nature of the Coulombic potential and the strong electronic polarization induced in solvent molecules by ions. Several simulations of aqueous ionic solutions have been carried out (76-86); however, better simulation techniques will be required before sufficiently realistic results can be obtained for ionic solutions (59). Until then, empirical models will have to be used to treat ionic hydration and ionic interactions in aqueous solution (5, 87), as e.g. in the hydration shell model. The available evidence, summarized in reference 5, seems to support the model proposed by Frank and Wen (88), shown in Figure 5, to describe ionic hydration. Despite the presence of more than one hydration shell around the ion in this model, Paterson et al. (5) have argued that ion-water interactions can be adequately modelled empirically by a single hydration shell.

Figure 5. The model for a hydrated ion in solution as proposed by Frank and Wen (88). The first hydration shell (heavily shaded) consists of immobilized water; the second shell (lightly shaded) contains less ordered water than the bulk medium (unshaded) that surrounds it.

Two nonpolar solutes can participate in a hydrophobic interaction. In an early treatment of this problem (17), the nonpolar molecules were thought to make contact, with an accompanying partial destruction of the surrounding clathrate-type water structure. More recently, Pratt and Chandler (88-91) have calculated the potential of mean force between two nonpolar solute
molecules and found that there are two stable configurations in such a system, one in which the two solutes do indeed make contact and one in which they are separated by one water molecule. Some computer simulations (64, 66, 92, 93), including one in which the potential of mean force between two Lennard-Jones solutes in water was calculated explicitly (66, 92), support these results of Pratt and Chandler. On the other hand, in a very long molecular dynamics run with four nearly-hard-sphere solutes among 339 water molecules [70 psec, which corresponds to about 100 million attempted moves of a normal Monte Carlo simulation (70)], there was no tendency for the solute molecules to associate (see Figure 6). Even when pairs of solutes were placed in contact initially, they showed a tendency to drift apart (70). These results, which differ from earlier ones, suggest that, whereas the picture of closely associated nonpolar molecules in aqueous solution is a reasonable one (and the basis of current discussions of hydrophobic bonding), the problem requires further investigation for clarification; some discussion of this problem has been presented in reference 70.

The potential of mean force is the basic quantity describing solute-solute interactions in liquids, and has been employed in theories of hydrophobic interactions (89, 94). Ideally, these are the type of potentials which should be used to model hydrated proteins. The problem is that these potentials are not pairwise additive and the extent of their non-additivity is not known, so that much basic research is required before they can be employed in conformational energy calculations. Use of the shell model of hydration, however, allows the most important influences of solute-solute interactions to be taken into account. It would be desirable to use simulations of aqueous solutions to obtain free energies of hydration with which to parameterize a shell model, but there have been serious practical difficulties that have prevented the realization of this aim (39, 47, 95-98). This is the reason that empirical free energies are being used temporarily.

The problem of calculating free energies in computer simulations reduces to the following. Whereas it is possible to compute the average potential energy, $U$, efficiently in a Monte Carlo calculation, it is very difficult to obtain the average value of $\exp(U/kT)$, which is essentially the free energy. This difficulty arises because the probability distribution function for $U$ has its maximum in the region where $U$ appears, but it is close to zero in the region where large values of $\exp(U/kT)$ occur; thus, the Monte Carlo technique rarely samples the region of $U$ where the value of $\exp(U/kT)$ is large. Special sampling techniques have been developed (95-97, 99-101) to circumvent this problem; however, these are applicable at best only to monotomic fluids at medium densities. Hopefully, in the future, refinement of such techniques would also make them applicable to simulations of water and aqueous solutions. Recently, Mezei et al. (98), using a concept developed by Onsager (102) and Kirkwood (103), carried out a series of Monte Carlo calculations of the free energy of liquid water by integrating with respect to a coupling parameter (see discussion in reference 97); making use of the MCY potential function, they obtained reasonable results for the free energy and entropy of liquid water.
SOME APPLICATIONS TO POLYPEPTIDES AND PROTEINS

This subject has been treated extensively in a recent review (6), and we present some examples here.

Because of the difficulty of computer simulations of aqueous solutions of small solutes, only a few simulations have been performed for larger solutes. From a molecular dynamics simulation of a dipeptide in water, Rosaky et al. (104,105) found that (a) many results of simulations on small solutes are also applicable to the larger, multi-functional dipeptide, thereby validating the applicability of simulations on small, representative solutes to larger solutes, (b) only one hydration layer seemed to be important in solvating the dipeptide, supporting the basis of the hydration shell model, and (c) hydration did not greatly affect the local structure and dynamics of the high frequency modes of the dipeptide. In a Monte Carlo simulation, Hagler et al. (106), in addition, found that the relative stabilities of two conformations of this dipeptide were very much affected by the presence of the solvent. For larger molecules, simulations can be carried out only for hydrated crystals because of computer limitations on the number of water molecules that can be treated (107,108), but it is not yet clear how relevant this computation is to the question of solvent effects in dilute solutions.

Quantitative treatments have been given for the stabilization of organized structures (e.g. α-helices) in polypeptides and proteins by hydrophobic bonds (109-116), and various experimental studies (117-124) have supported the results of these computations. For example, Figure 7 shows a comparison of experimental and calculated values of the temperature dependence of the Zimm-Bragg helix-growth parameter s (125) for poly(L-valine) in water. It can be seen that the calculations match the observed increase in s with increasing temperature reflecting primarily the characteristic increase (17) in hydrophobic bond strength (involving the valine side chain) in this temperature range.

Figure 7. Comparison of s vs. T curves for poly(L-valine) in water. The line is a calculated one (116), and the squares are the experimental results (121).

The hydration-shell model, thus far parameterized with empirical free energies, has been used in calculations of a variety of polypeptide and protein conformational problems. For example, as shown in Figure 8, the conformational energy contour map of an amino acid residue is influenced by the presence of water (4). The shell model has also been used to include the influence of water in calculations of the conformations of various peptides (2,115,116,126,127) and of oxytocin and vasopressin (128).

In a globular protein, water tends to segregate the polar and nonpolar residues so that the former have a tendency to lie on the surface of the protein whereas the latter tend to lie in the interior. These effects are manifested in a number of empirical observations (reviewed in reference 6) on the distribution of polar and nonpolar residues in the known structures of globular proteins, determined by X-ray analysis.

Finally, in theoretical studies of protein folding, the nucleation sites for initiation of the folding process have been predicted with a model based on optimization of hydrophobic bonding (129).
Figure 8. Part of the conformational energy contour maps of unhydrated (A) and hydrated (B) N-acetyl-N'-methylalanineamide as a function of $\phi$ and $\psi$ for $\chi^1 = 60^\circ$. From reference 4.

Clearly, the solvent plays an important role in determining protein conformation, and further improvements in the simulation techniques discussed here will be required for a proper treatment of protein hydration.

Acknowledgement -- I am indebted to Dr. William J. Peer for helpful comments on this manuscript.

The research described herein that was carried out in the author's laboratory has been supported primarily by grants from the National Science Foundation (PCM79-20279 and PCM79-18336) and from the National Institute of General Medical Sciences of the National Institutes of Health, U.S. Public Health Service (GM-14312 and GM-25138).

REFERENCES

Protein structure and reactivity in aqueous solutions