Differences in Hydration Structure Near Hydrophobic and Hydrophilic Amino Acids

Teresa Head-Gordon,* Jon M. Sorenson,# Alexander Pertsemlidis,§ and Robert M. Glaeser*§
*Life Sciences Division, Lawrence Berkeley National Laboratory; #Department of Chemistry, University of California; and §Department of Molecular and Cell Biology, University of California, Berkeley, California 94720 USA

ABSTRACT We use molecular dynamics to simulate recent neutron scattering experiments on aqueous solutions of *N*-acetyl-leucine-amide and *N*-acetyl-glutamine-amide, and break down the total scattering function into contributions from solute-solute, solute-water, water-water, and intramolecular correlations. We show that the shift of the main diffraction peak to smaller angle that is observed for leucine, but not for glutamine, is attributable primarily to alterations in water-water correlations relative to bulk. The perturbation of the water hydrogen-bonded network extends roughly two solvation layers from the hydrophobic side chain surface, and is characterized by a distribution of hydrogen bonded ring sizes that are more planar and are dominated by pentagons in particular than those near the hydrophilic side chain. The different structural organization of water near the hydrophobic solute that gives rise to the inward shift in the main neutron diffraction peak under ambient conditions may also provide insight into the same directional shift for pure liquid water as it is cooled and supercooled.

INTRODUCTION

Recently reported neutron scattering studies (Pertsemlidis et al., 1996) on hydrophobic and hydrophilic amino acids have provided new experimental evidence that water in the hydration shell of a nonpolar side chain is organized differently than in pure liquid water. The scattering experiment observes a shift in the main neutron diffraction peak of water to a smaller angle for solutions of hydrophobic *N*-acetyl-leucine-amide (NALA), and its side-chain analog isobutanol, whereas the position of the water peak remained unperturbed for solutions of hydrophilic *N*-acetyl-glutamine-amide (NAQA) and the model backbone *N*-acetyl-glycine-amide.

Scattering from aqueous solution can be analyzed in terms of four spherically averaged pair contributions:

$$I_{\text{solution}}(Q) = I_{\text{solute-solute}}(Q) + I_{\text{solute-water}}(Q) + I_{\text{water-water}}(Q) + I_{\text{intra}}(Q) \quad (1)$$

where the last term refers to scattering interference between atoms on the same molecule, and the remaining terms are contributions due to intermolecular correlations. The purpose of the present paper is to use molecular dynamics simulations to 1) evaluate the various contributions to the total scattering, 2) to show that the shift in the main diffraction peak observed for aqueous hydrophobic solutions is indeed due primarily to changes in water-water correlations, and 3) to provide insight into the molecular origin of the observed shift by using polygon distribution functions and the degree of polygon planarity. We show in regard to 3) that water around the leucine side chain is more often organized into hydrogen-bonded rings dominated by pentagons with better hydrogen bonds, when compared to the glutamine side chain, and that little difference in structure is observed near their respective backbones. Furthermore, the dihedral angle distributions generated as a function of hydrogen-bonded ring size (pentagons and hexagons) show less ring puckering near the hydrophobic side chain, consistent with the shift in the measured diffraction peak to a larger effective Bragg spacing. We show that the measured effect by neutron scattering should also be observable by x-ray scattering.

It is worthwhile to explain the term "effective Bragg spacing," which is central to our analysis of the solution scattering experiments. In the case of crystalline materials, x-rays and neutrons are scattered at discrete angles by sinusoidal (Fourier) components of the electron density and nuclear scattering potential of the specimen, respectively. The scattering angle is determined by the spatial period of the Fourier component that is responsible for the scattering; thus, for each scattering angle there is a corresponding Bragg spacing, which is given by

$$d = \frac{\lambda}{2\sin(\theta/2)} = \frac{2\pi}{Q} \quad (2)$$

where θ is the scattering angle and Q is the momentum transfer associated with scattering. Bragg's law was first used to describe diffraction from crystalline materials, where the spatial period of each Fourier component of the electron density is determined by the lengths of the unit cell vectors of the crystal. Representation of the electron density as a sum of Fourier components is equally applicable to noncrystalline materials, however. As a result it is still true that the spatial period of the Fourier component, which we

Received for publication 5 December 1996 and in final form 26 June 1997.
Address reprint requests to Dr. Teresa Head-Gordon, Lawrence Berkeley National Laboratory, 314 Donner, Life Sciences Division, Berkeley, CA 94720. Tel.: 510-486-7365; Fax: 510-486-6488; E-mail: thg@ water.lbl.gov.

Dr. Pertsemlidis's present address is Department of Biochemistry, University of Texas Southwestern Medical Center, Dallas, TX 75235.

0006-3495/97/10/2106/10 $2.00

can call an effective Bragg spacing, can be calculated from the measured scattering angle according to the above equation. As with crystalline materials, the amplitude of each Fourier component of the electron density is given by the square root of the scattered intensity. Information about the vector direction of the Fourier component is lost in scattering from liquids, however, unlike in the case of crystals.

In the case of normal ice, the solid is made up of hydrogen-bonded hexagons arranged such that the largest Bragg spacing observed with x-rays is ~0.39 nm. Cubic ice is also composed exclusively of hydrogen-bonded hexagons, but in this case the largest Bragg spacing is 0.37 nm (Bosio et al., 1983; Dubochet et al., 1982). The strict adherence to hydrogen-bonded hexagons of the ice crystal gives way in the liquid to a broader distribution of hydrogen-bonded configurations, including a variety of polygons of varying sizes and degrees of puckering or distortion, all of which result in a more compact arrangement of water molecules. The electron density of the liquid is now characterized by a continuous water ring rather than a discrete distribution of Fourier components, and the effective Bragg spacing associated with the main water peak is found to be ~0.31 nm.

Furthermore, the distribution is peaked at a distance that remains larger than the center-to-center distance between individual water molecules. Thus it is clear that the most prominent Fourier components of the scattering density of pure liquid water have a repeat distance that is larger than the oxygen-oxygen nearest-neighbor distance, which is ~0.28 nm. The larger Bragg spacing tells us that the fundamental scattering unit in liquid water must be something bigger than a single water molecule. One important aspect of the present paper is to show that the clustering of water molecules into more planar pentagons in the hydration shell of a hydrophobic amino acid is likely an important contributor to the fundamental scattering unit for water in aqueous solutions containing hydrophobic groups.

The next section provides background on solution scattering, describes our simulation protocol and the empirical force fields used, outlines the procedure for obtaining simulated excess scattering curves, and briefly describes the polygon distribution function analysis used to explain the measured differences between leucine and glutamine solutions. The third section contains the results of our structural analysis that pertain to the molecular origin of the shift in the main diffraction peak of water to a smaller angle for a hydrophobic amino acid. We conclude with a summary, discussion, and plans for future work.

METHODS AND MODELS

Simulation of excess solution scattering intensity

Apart from an isotropic scattering term, the intensity measured for a solution is expressed as

$$I(Q) = \sum_A \sum_B c_A c_B b_A b_B H_{AB}(Q) \qquad (3)$$

where b_A is the neutron scattering length for type A atom, and c_A is the atomic fraction of atom A (Egelstaff, 1992; Corongiu and Clementi, 1992; Finney and Soper, 1994). The partial structure factor,

$$H_{AB}(Q) = 4\pi\rho \int_0^\infty r^2[g_{AB}(r) - 1] \frac{\sin(Qr)}{Qr} \, dr \qquad (4)$$

is the Fourier transform of the direct correlation function, $h_{AB}(r) = g_{AB}(r) - 1$, where $g(r)$ is the radial distribution function (rdf), ρ is the total number of atoms per unit volume, and Q is the momentum transfer, related to the scattering angle as $Q = 4\pi\sin(\theta/2)/\lambda$. All water-water rdfs ($g_{OO}(r)$, $g_{OH}(r)$, and $g_{HH}(r)$) and all water-solute rdfs ($g_{OX}(r)$ and $g_{HX}(r)$, where X is an atom of the leucine or glutamine solute) were determined from molecular dynamics simulations and were used to evaluate Eq. 4 (Pertsemlidis et al., 1996). The scattering length appropriate to deuterium (0.663×10^{-12} cm) is used to describe heavy water and all exchangeable hydrogens on the solute to match the experimental conditions.

We define excess scattering as the difference between the intensity measured from solution and the scattering measured for pure water scaled by the factor k, the estimated number of water molecules per unit volume of a given solution divided by the number of water molecules per unit volume of pure water (Pertsemlidis et al., 1996; Pertsemlidis, 1995):

$$I_{excess}(Q) = I_{solution}(Q) - kI_{pure\ water}(Q) \qquad (5)$$

Our definition of excess scattering is operationally equivalent to the definition of a "first-order difference" between scattering of pure water at two different temperatures, where k is analogous to differences in density at the two temperatures. In both cases, the main water diffraction peak at 2.0 $Å^{-1}$ might shift in some direction so that the excess scattering (or first-order difference) shows a ripple in the region of the water ring (a term commonly used in protein x-ray crystallography to describe what is, in that context, the undesired background scattering of water). We analyze the scattering data in the next section by combining Eqs. 1 and 5:

$$I_{excess}(Q) = I_{solute-solute}(Q) + I_{solute-water}(Q) + \Delta I_{water-water}(Q) + I_{intra}(Q) \qquad (6)$$

to isolate changes in solution water structure, if any, compared to that of pure water. The evaluation of each contribution to Eq. 6 is described in the next subsections.

Molecular dynamics simulation of $I_{solute\text{-}water}(Q)$ and $I_{water\text{-}water}(Q)$

The simulation of a single solute in water allows us to evaluate contributions from solute-water and water-water

correlations to $I_{excess}(Q)$. We use AMBER 4.0 parameters (Cornell et al., 1995) to describe a single *N*-acetyl-leucine-amide (NALA) (Fig. 1, *top*) or *N*-acetyl-glutamine-amide (NAQA) (Fig. 1, *bottom*) in solution with 502 and 505 SPC water molecules, respectively (Berendsen et al., 1981). Contributions from $I_{solute\text{-}water}(Q)$ and $I_{water\text{-}water}(Q)$ were multiplied by a factor of 4.56 and 2.60 for leucine and glutamine, respectively, to account for differences in the concentration of solute molecules between the simulated and experimental conditions (Pertsemlidis et al., 1996). An independent simulation of pure water using 512 SPC water molecules was performed to generate the excess scattering differences.

The velocity Verlet algorithm (Swope et al., 1980) was used to integrate the equations of motion at 298 K and ~1 kg/liter in the NVT ensemble. A time step of 1.5 fs was used for all reported calculations. The simulation length comprised 150 ps of equilibration and 300 ps of dynamics, during which statistics were accumulated. Constant temperature was maintained by periodically rescaling velocities (every 0.75 ps). RATTLE was used to maintain the rigidity of the SPC internal degrees of freedom (Andersen, 1983). We used Ewald sums for the aqueous solution simulations to evaluate the long-range electrostatic interactions, where a Gaussian form for the modified charge distribution is assumed (Allen and Tildesley, 1987). We used a convergence parameter $\alpha = 6.0/L$ and truncated the k-space sum at $k_m = 3$. The evaluation of the r-space sum was truncated at $n = 0$, and the minimum image convention was used to evaluate all real-space interactions. The minimum image convention was also used for evaluating the solute-solvent interactions.

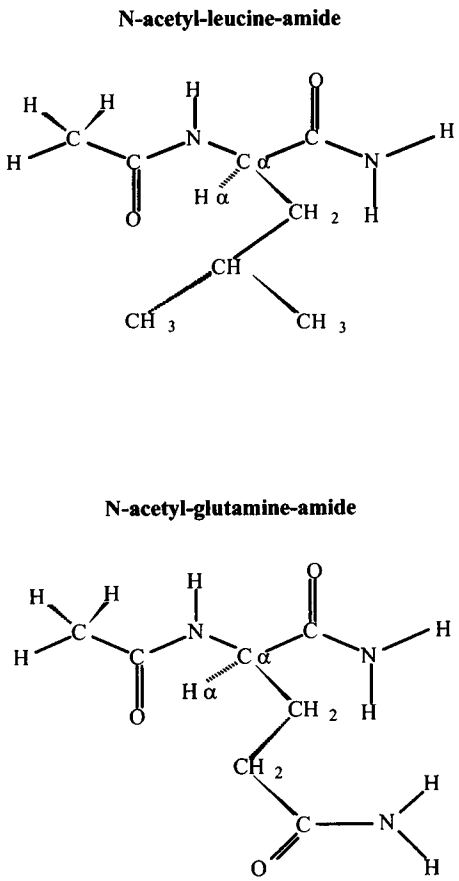

FIGURE 1 Chemical structures of the two solutes used in the neutron scattering experiments (Pertsemlidis et al., 1996) and simulated in this study. (*Top*) *N*-Acetyl-leucine-amide. (*Bottom*) *N*-Acetyl-glutamine-amide.

Molecular dynamics simulation and integral equation estimates for $I_{solute\text{-}solute}(Q)$

Because we are primarily interested in reproducing the experiment and determining the molecular origin of the shift in the diffraction peak at $Q = 2.0$ Å^{-1}, we will not attempt to quantitatively describe the solute-solute correlations in water throughout the full Q range. In particular, the small-angle region ($Q < 1.0$ Å^{-1}) is difficult to simulate because it is dominated by large molecular length scales corresponding to multiples of the diameter of the solute molecules themselves. Simulation of the small-angle region is therefore limited by both our simulation box size (which is valid for $Q > 0.25$ Å^{-1}) and adequate sampling over the full radial separation between solute molecular centers in water.

By contrast, the region corresponding to the water ring, $Q \approx 2.0$ Å^{-1}, probes correlations over smaller length scales with an effective Bragg spacing of ~3.1 Å. Simply based on the effective size of the solutes, which are ~5.0–7.0 Å in diameter, we can rule out their contribution in the region of the water ring. We use Pratt-Chandler theory (Pratt and Chandler, 1977) above to crudely estimate $g_{solute\text{-}solute}(r)$ for hydrophobic spheres of this size in water to show that solute-solute correlations do not contribute in the region of the water ring. Pratt-Chandler theory solves Ornstein-Zernike-like equations for solute-water and solute-solute pair correlation functions, assuming that repulsive interactions dominate those interactions in dilute solution, and therefore can be approximated by treating the solute particles as hard spheres (Pratt and Chandler, 1977). Clearly the attractive forces in a highly associative fluid such as water would be poorly treated by the above approximation, and the theory is rendered semiempirical through the use of the experimental oxygen-oxygen pair correlation function for liquid water (Pratt and Chandler, 1977).

However, there is some possible effect due to smaller length-scale, intersolute interactions, such as the formation of a solute-solute hydrogen bond, that might contribute to scattering in the water ring region. In an attempt to quantify these effects, we have simulated 27 leucines and glutamines confined to a box 30 Å on edge, without water, to evaluate all radial distributions functions between all solute atomic pairs. Although this gas-phase simulation is obviously imperfect for describing the nature of solute-solute correlations in water, it does demonstrate that solute-solute correlations do not significantly contribute to scattering intensity in the water ring region at the concentrations used in the neutron scattering experiments.

In the case of Pratt-Chandler theory, we estimate $I_{\text{solute-solute}}(Q)$ according to

$$I_{\text{solute-solute}}(Q) = \langle F(Q)\rangle^2[S_{\text{solute-solute}}(Q) - 1] \quad (7)$$

where

$$S_{\text{solute-solute}}(Q) - 1 = 4\pi\rho \int_0^\infty r^2 g_{\text{solute-solute}}(r) \frac{\sin(Qr)}{Qr}\, dr \quad (8)$$

We approximate leucine (or glutamine) as a sphere of radius r_a = 2.5 Å, and using an average scattering length, b_{ave}, of all atomic scattering lengths in the solute molecule, we evaluate $\langle F(Q)\rangle^2$ as

$$\langle F(Q)\rangle^2 = b_{ave}^2 \left[\frac{\sin(Qr_a)}{Qr_a}\right]^2 \quad (9)$$

In the case of atomic simulation, we evaluate $I_{\text{solute-solute}}(Q)$ according to Eq. 3.

Calculation of $I_{\text{intra}}(Q)$

To obtain the contribution from intramolecular correlations to the scattered intensity, we have averaged the molecular structure factor (Eq. 10) over many possible molecular conformations, weighting each one by its probability of occurrence, using the library of conformations of Dunbrack and Karplus (1994). The library provides the atomic coordinates for a large set of side-chain and backbone conformations and the number of times a particular conformation appears among proteins of known structure. For each such conformation, the spherically averaged square of the molecular structure factor was calculated as

$$\langle F^2(Q)\rangle = \sum_a \sum_b b_a b_b \frac{\sin Qr_{ab}}{Qr_{ab}} \quad (10)$$

where r_{ab} is the distance between two atoms within one solute molecule. A total of 589 and 661 different conformers taken from 10,491 and 4,857 occurrences in the data base were used to calculate the molecular structure factor for NALA and NAQA, respectively.

Polygon and dihedral angle distribution analysis

One of us has recently used polygon distribution functions (the enumeration of non-short-circuited hydrogen bond pathways) to analyze water structure near two methane solutes studied by molecular dynamics simulation (Head-Gordon, 1995). We have used the polygon definition of Rahman and Stillinger (1973), but with the difference that for multicyclo rings, only the minimum number of the smallest-size polygons that capture all vertices is counted. We note that primitive polygons provide an alternative polygon definition (Speedy et al., 1987). In pure (ambient temperature) liquid water there is nearly a 1:1 ratio of pentagons to hexagons (Rahman and Stillinger, 1973), whereas pentagons overwhelm all other polygon sizes in clathrate hydrates (Davidson, 1973). Polygon distributions calculated in the hydration shells near the two methanes provided direct evidence that clathrate analogies are qualitatively valid, because there are nearly twice as many pentagons as hexagons, but it was shown that larger polygons also contributed to greater structural organization along the surface of the nonpolar groups, which is inconsistent with simple clathrate-like views (Head-Gordon, 1995).

Coordinates were written out every 100 steps, and the polygon distribution functions were evaluated as averages over the resulting 2000 snapshots. We have constructed polygon distributions in two different regions of the solute molecule for our comparison of hydration structure between hydrophobic and hydrophilic residues. The first region considers all water molecules within a specified distance of any solute side-chain atom (CH_2-CH-$(CH3)_2$ for leucine and CH_2-CH_2-CO-NH_2 for glutamine), and the second region is for all water molecules within a specified distance of any backbone atom (CH_3-CO-NH-$C_\alpha H_\alpha$-CO-NH_2).

We construct the nearest-neighbor lists of all water molecules in either of these regions from all water molecules whose oxygens are also present in the same region and whose "hydrogen-bonded" interaction energy falls below V_c. This definition allows for the discovery of a unique hydration structure that might occur along the solute side-chain (backbone) surface. In the next section we will compare how the polygon distributions generated in these regions differ between leucine and glutamine, and how the differences fall off with distance from the solutes.

The shift in effective Bragg spacing to larger values (corresponding to a shift to a smaller angle of the main water diffraction peak) found for hydrophobic solutions suggests that some of the hydrogen-bonded rings are becoming more planar. Therefore we evaluate dihedral angle distributions as a function of polygon size, focusing primarily on pentagons and hexagons, to evaluate whether the rings are indeed more planar near the leucine side chain.

RESULTS

Individual contributions to the total scattering intensity

Fig. 2, *a* and *b*, portrays the high flux beam reactor (HFBR) experimental excess scattering curves (Pertsemlidis, 1995; Pertsemlidis et al., 1996), the simulated excess scattering curve, as well as the integral equation theory results for hard spheres in water, for leucine and glutamine solutes, respectively. The simulated excess scattering curve represents the sum of all terms in Eq. 6. We also display the SANDALS experimental curve for leucine that places the leucine data on an absolute scale, and provides important confirmation that the effect seen for leucine is independent of experimental setup. The HFBR experimental data, which are not on an absolute scale and extend out to only 2.5 $Å^{-1}$, were shifted

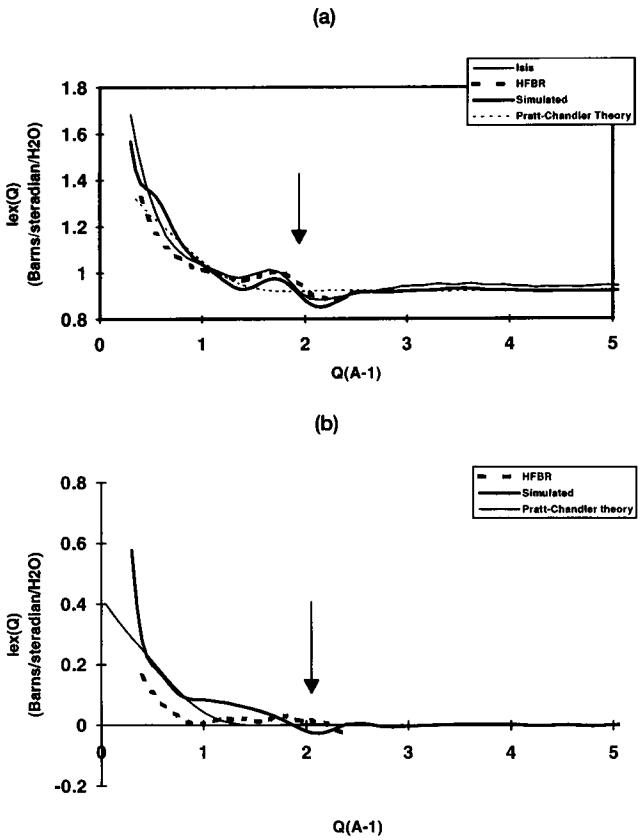

FIGURE 2 A comparison of simulated and experimental excess scattering curves, $I_{excess}(Q)$, in the region of the main water diffraction peak for (*a*) leucine and (*b*) glutamine. Excess scattering represents the difference between scattering from solution (0.5 M) and scattering measured for pure water. The comparison shows that there is reasonable quantitative agreement between simulation and the HFBR and SANDALS experimental curves. The flatness of the excess scattering for glutamine indicates that there is little change for the hydrophilic solution, which can be detected by a scattering measurement. The excess scattering for leucine, on the other hand, shows that the water structure in the aqueous solution of the hydrophobic solute is different from that found for pure water, so that subtraction gives a strong ripple in the water ring. We also plot $I_{solute\text{-}solute}(Q)$, as determined by Pratt-Chandler theory, to show that, based on solute size considerations alone, solute-solute correlations should not be important at this effective Bragg spacing.

and scaled by a two best-parameter fit to the SANDALS data (Pertsemlidis et al., 1996). To match the SANDALS experimental data, the simulated and integral equation results were offset by a factor equal to the sum of the square of the scattering lengths for all leucine atoms (normalized by the number of water molecules per solute), which is the theoretical limit for scattering at high Q.

It is evident that the simulation data are in quite reasonable quantitative agreement with the neutron data, and exhibit the correct qualitative trends between the two solute types, in the region of the water ring. The solute-solute correlations as determined by Pratt-Chandler theory demonstrate that the intermolecular contribution to the total scattering is neglible in the region of the water ring, for hard spheres in water that are roughly the size of the experimental solutes.

Fig. 3, *a* and *b*, shows the individual contributions to $I_{excess}(Q)$ resulting from atomic simulations of solute-solute, solute-water, changes in water-water correlations between solution and pure water, and calculated intramolecular scattering. At the experimental concentrations (roughly one leucine for every 100 waters), there are no significant effects due to either solute-water correlations or, even for hydrogen bonding configurations between solutes, solute-solute correlations. It is evident that the water-water correlations are the dominant contributor to the water ring signature for leucine at ~2.0 Å $^{-1}$. Leucine displays a structural reorganization of water that shifts the main water diffraction peak to smaller Q (Pertsemlidis et al., 1996). As a result, the difference in scattering between the solution and pure water is a positive-negative pair of peaks rather than a flat baseline. Relative to leucine, the curve for glutamine appears flat in this region, indicating that its hydration shell is largely equivalent in structure to that of the pure water background that has been subtracted off, at least for two body correlations. The next section describes an anal-

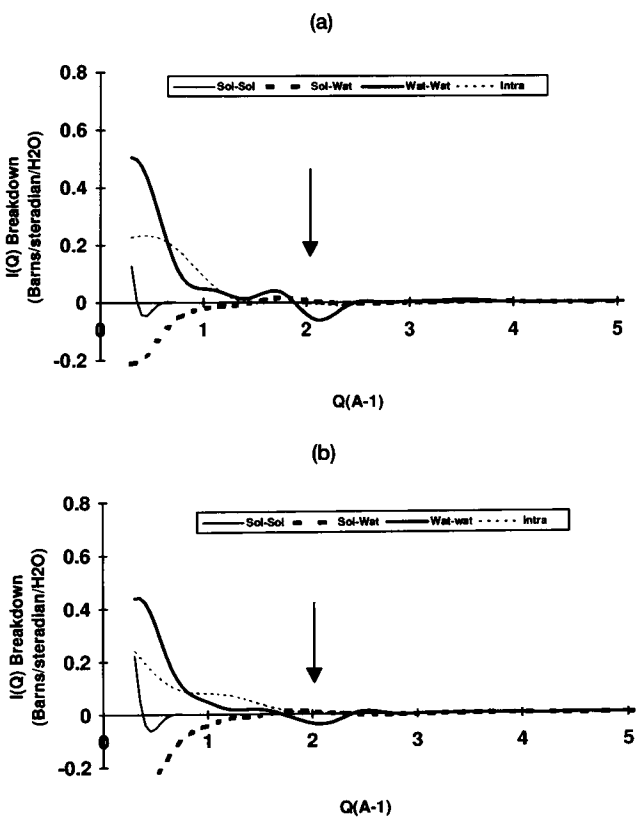

FIGURE 3 The individual contributions to $I_{excess}(Q)$, described in Eq. 6, for (*a*) leucine and (*b*) glutamine. Because we are not evaluating a true solute-solute correlation in water, the small angle region cannot be compared to the experimental data. The atomic simulations of solute-solute and solute-water correlations show that they are not significant contributors to the region 1.5 Å $^{-1} < Q < 2.5$ Å $^{-1}$ at the experimental concentrations. Intramolecular effects are also flat in this region. Essentially the water-water correlations dominate the perturbation of the water ring, and this perturbation is attributable to alterations of the hydration shell around the hydrophobic amino acid.

ysis of the structural organization of water near the leucine side chain, which may explain the shift in the main diffraction peak of water to a smaller angle.

Structural analysis of hydration shells near leucine and glutamine

A measure of hydration structure in terms of many-body functions is the enumeration of polygon sizes (Rahman and Stillinger, 1973; Speedy et al., 1987; Head-Gordon, 1995), which we use to quantify differences in the organization of water structure around the leucine and glutamine side chains and backbones. Polygon distributions in different regions near the two solutes are displayed in Figs. 4-7. P_N in these figures refers to the absolute number of polygons of each size: triangles (P_3), quadrilaterals (P_4), pentagons (P_5), etc., counted over all 2000 snapshots, except for P_0, which corresponds to the number of snapshots when no polygon of any size is found. P_1 and P_2 are not listed, because the smallest possible polygon is three-sided. No triangles were observed in the simulation with the SPC model for the values of V_c used. The numbers next to the bar labels for leucine and glutamine indicate how many waters of hydration are present in any given region.

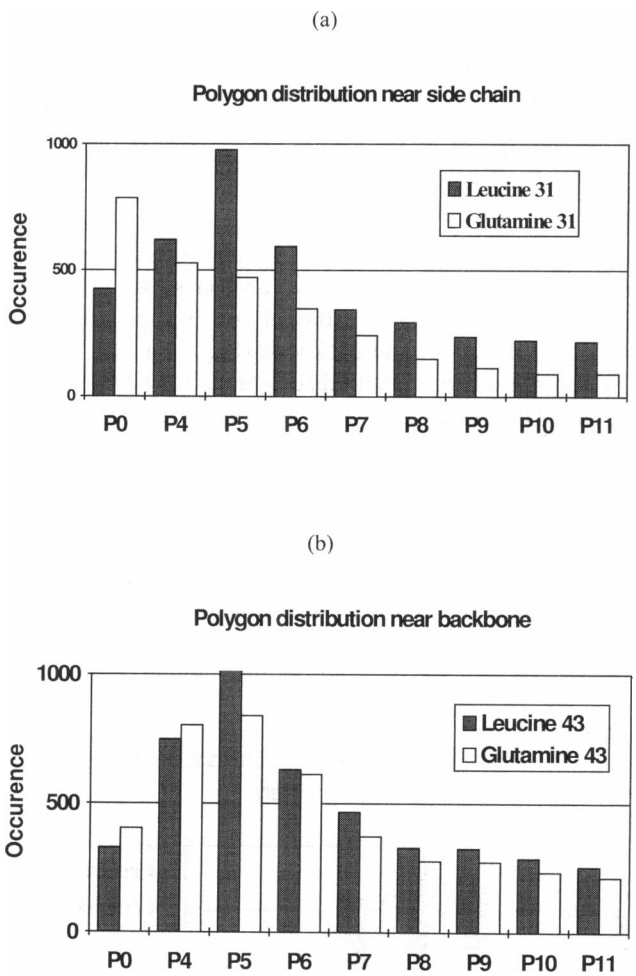

FIGURE 4 Polygon distribution functions of water generated near (*a*) the side chain and (*b*) the backbone of leucine and glutamine for $V_{cut} = -3.0$ kcal/mol and $R_{cut} = 5.0$ Å at room temperature. P_0 corresponds to the number of snapshots in which no polygon is found. P_5 is the number of pentagons, P_6 is the number of hexagons, etc. The number in the legend corresponds to the average number of vertices used to generate the polygon distributions.

Fig. 4 compares polygon distributions for leucine and glutamine side chains and backbones. In this figure a water molecule is counted if it lies 5.0 Å from any solute side-chain atom (Fig. 4 *a*) or backbone atom (Fig. 4 *b*), and V_c is below -3.0 kcal/mol. A comparison of Fig. 4, *a* and *b*, shows that structural differences between hydration shells around leucine and glutamine reside largely near the side chains and not their peptide backbones, consistent with the results obtained by Pertsemlidis et al. (1996) in the neutron scattering experiments using the isobutanol and *N*-acetylglycine-amide solutes. In Fig. 4 *a*, the leucine side chain is seen to have a more ordered hydration shell, because the number of snapshots with no polygons present (P_0) is nearly half that of glutamine. Another observation about the polygon data is that the polygon size making the largest contribution is quadrilaterals for glutamine, whereas pentagons contribute the largest number in the case of leucine. The ratio of quadrilaterals to pentagons or hexagons to pentagons is roughly 1:1 for glutamine, whereas for leucine they are 1:2. Finally, the fraction of the total number of rings that are pentagons is 1/3 for leucine and 1/4 for glutamine. Clearly pentagons are an important feature of the polygon distribution around the leucine side chain, providing evidence for clathrate-like behavior of water molecules, which form the hydration shell around large and complex-shaped hydrophobic solutes like leucine. It is also important not to take the idea of a persistent clathrate cage too literally, because the absolute number of any given polygon size is less than the number of coordinate sets used to generate the distributions, indicating that the hydration structure is continuously formed and unformed. However, the idea that pentagons play a special role in enclosing the solute, similar to the role that pentagons play in the fullerenes when compared to carbon sheets, gives topological justification for the importance and qualitative validity of clathrates (Head-Gordon, 1995).

In Fig. 4 *b* the polygon distributions near the backbone atoms are largely the same for the two amino acids, which is especially manifested by the near-equality in the number of snapshots where no polygons are found (P_0). Two further points are worth making regarding the data in Fig. 4. First the extra number of hydrogen-bonded rings around leucine as compared to glutamine near the backbone is due to the (imperfect) definition of whether a water molecule resides near the side chain or backbone; some water molecules will serve as a vertex in both regions. Second, the absolute number of some polygons near leucine's side chain is as great as the number surrounding the backbone. This does not mean that the hydration structure is similar in the two cases, but rather that there are just more water molecules

near the backbone (43) than near the side chain (31), based on the radial envelope of 5.0 Å, so that there are more vertices and therefore more possibilities for forming hydrogen-bonded rings. The fact that the side chain forms as many rings as the backbone for a given polygon size, with fewer vertices, actually emphasizes the greater degree of networking within the hydration shell near the hydrophobic group.

Fig. 5 makes the same comparison as Fig. 4, except for a stricter energy criterion of $V_c = -3.25$ kcal/mol. The point here is that the observed ordered water structure around the leucine side chain is not sensitive to the definition of V_c. In fact, the trends evident when progressing from Fig. 4 to Fig. 5, which differ only in the strength of the water-water interaction (which is correlated with the degree of linearity of the hydrogen bond; Rahman and Stillinger, 1973), suggest that although more ordered and entropically unfavorable water is present around leucine, compensation is found through better energetic interactions between hydration shell waters, as has been noted in the past (see, for example, Franks, 1975). This is consistent with the idea that "better" (i.e., more linear) hydrogen bonds are formed around the hydrophobic solute.

How far does the altered water structure extend from the amino acid surfaces of NALA and NAQA? The polygon distributions of Figs. 4 and 5 essentially characterize the first and part of the second solvation shell. Fig. 6 displays the polygon distribution generated from all water molecules within 6.0 Å of the side chain, after subtracting off the polygon counts for each size generated from distances out to 5.0 Å, for NALA and NAQA. There are clearly some remaining structural differences between the hydrophobic and hydrophilic peptides, although the structural differences between NALA and NAQA are absent in shells at larger radius (data not shown). For single hydrophobic amino acid solutes in water, the results seem to suggest that the first solvation shell is significantly compromised to accommodate the solute, whereas the second solvation shell is altered to interface between the first solvation shell and the surrounding bulk.

It is evident that water structure in the vicinity of the leucine side chain shows greater numbers of pentagons and more highly connected vertices in general in the first two solvation shells (Figs. 4–6). This is only part of the structural picture that underlies the shift of the diffraction peak to smaller angle in reciprocal space. Figs. 7 and 8 display dihedral angle distribution functions generated from all four sequential vertices available for pentagons near the NALA and NAQA side chains (Fig. 7 *a*) and backbone (Fig. 7 *b*), and the corresponding distributions for hexagons near the side chain (Fig. 8 *a*) and backbone (Fig. 8 *b*).

When the dihedral angle distributions for pentagons are compared between NALA and NAQA near their side chains (Fig. 7 *a*), it is evident that the pentagonal rings are much more planar near the leucine side chain than those near the glutamine side chain. Fig. 7 *b* shows that the dihedral distributions for pentagons are very similar near their backbones. The dihedral angles generated from hexagons indicate that the hexagonal rings are again more planar near the side chains of leucine than those near glutamine (Fig. 8 *a*), while remaining similar near their backbones (Fig. 8 *b*), but

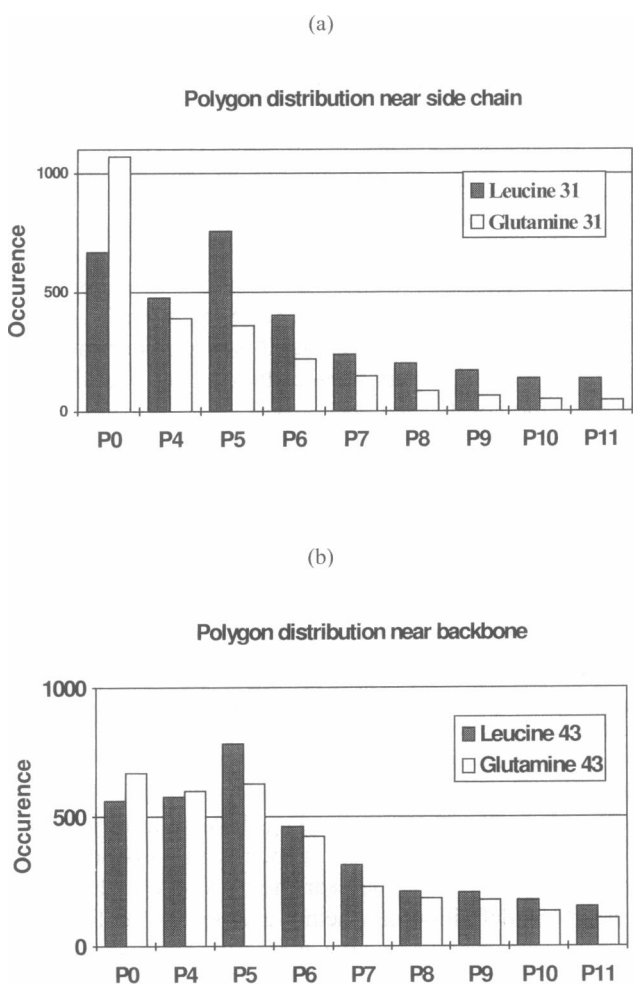

FIGURE 5 Polygon distribution functions of water generated near (*a*) the side chain and (*b*) the backbone of leucine and glutamine for $V_{cut} = -3.25$ kcal/mol and $R_{cut} = 5.0$ Å at room temperature. See Fig. 4 for more details.

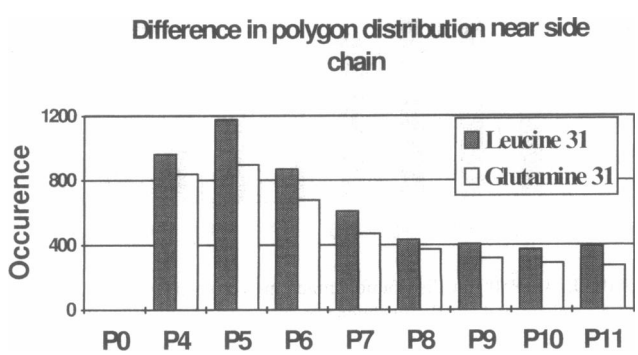

FIGURE 6 Differences of polygon distribution functions generated for $R_{cut} = 6.0$ Å and $R_{cut} = 5.0$ Å of water generated near the side chain of leucine and glutamine, using $V_{cut} = -3.75$ kcal/mol.

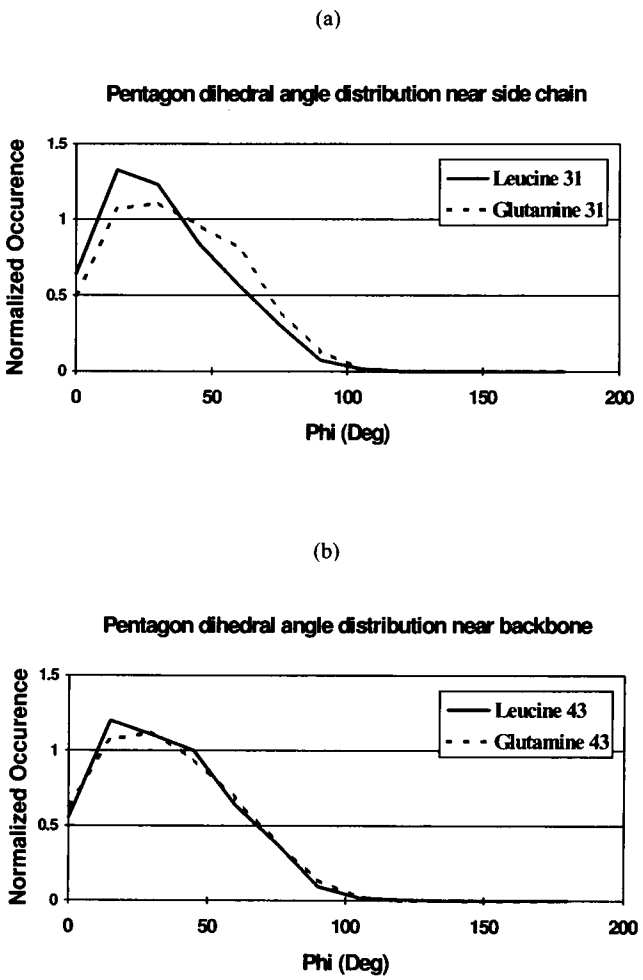

FIGURE 7 Dihedral angle distribution function for pentagons near (*a*) the side chain and (*b*) the backbone of leucine and glutamine for V_{cut} = −3.0 kcal/mol and R_{cut} = 5.0 Å.

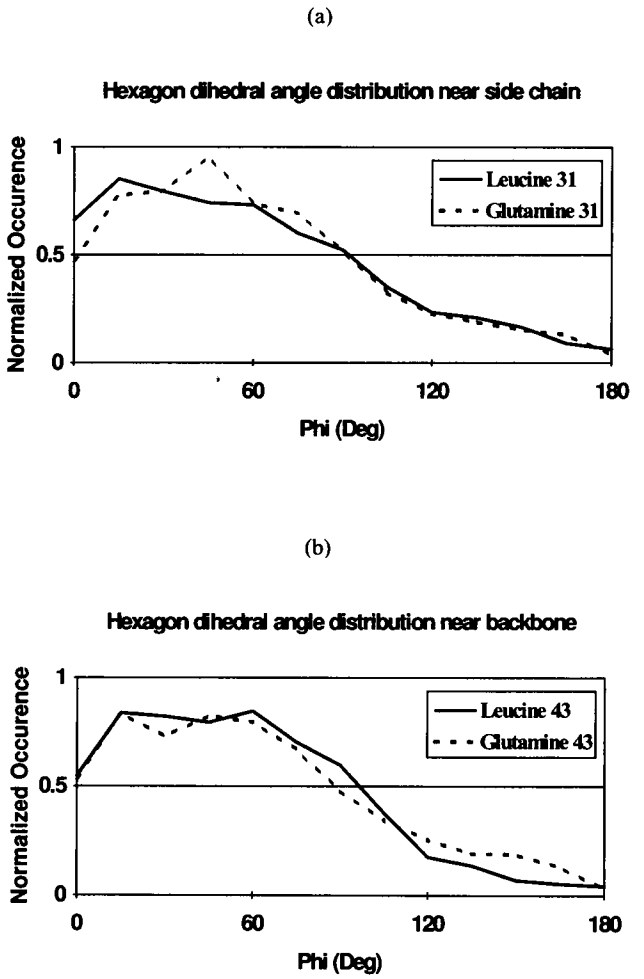

FIGURE 8 Dihedral angle distribution function for hexagons near (*a*) the side chain and (*b*) the backbone of leucine and glutamine for V_{cut} = −3.0 kcal/mol and R_{cut} = 5.0 Å .

the contrast for hexagons is less dramatic than that found for pentagons.

The shift seen in the hydration shell of the leucine side chain is qualitatively analogous to the change in effective Bragg spacing observed in x-ray scattering when going from liquid water (~3.2 Å) to hexagonal ice (~3.9 Å). When going from ice to liquid, the expanded hydrogen-bonded network due to hexagonal rings gives way to contributions made by smaller ring sizes and the greater distortion of hexagonal (and larger) rings, so that the effective Bragg spacing decreases. The greater degree of ring planarity for pentagons near the hydrophobic side chain is likely to be the most direct evidence for why the water peak shifts to a smaller angle and hence for a larger effective Bragg spacing. Figs. 4–8 emphasize the special role played by pentagons in adapting the pure water network to accommodate relatively large hydrophobic solutes such as NALA. Therefore, the greater number of more planar pentagons probably contributes the most to the shift in equivalent Bragg spacing to larger values (liquid water to ice), as is observed experimentally.

Finally, in Fig. 9 we show the differences between the individual partial structure factors, $H_{OO}(Q)$, $H_{OH}(Q)$, and $H_{HH}(Q)$, measured for leucine solutions and pure water. This breakdown into partial structure factors makes it clear

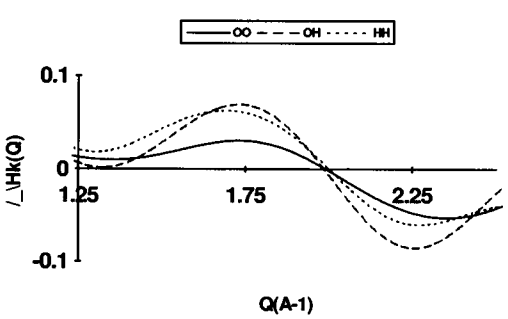

FIGURE 9 Partial structure factor differences between aqueous leucine solution and pure water, weighted by concentration of atomic species, but not atomic scattering lengths. The observed ripple for the simulated oxygen correlations suggests that a shift in the water peak should be discernible by x-ray scattering.

that perturbation of the water ring for hydrophobic solutions should be detectable by x-ray diffraction and could therefore be used as a complementary experimental source for measuring altered hydration structure.

DISCUSSION AND CONCLUSION

In this study we have simulated and analyzed neutron scattering solution studies of hydrophobic and hydrophilic blocked peptides in water. Our simulations and calculations provide fair quantitative agreement between simulation and experiment (except at small angles), especially in the region that brackets the main neutron diffraction peak for water. A breakdown of the simulated total scattering function into contributions from solute-solute, solute-water, changes in water-water correlations between waters in solution and in neat water, and intramolecular correlations, attributes the shift in the peak, observed only for the hydrophobic solute, to structural perturbations of water within the hydration shell.

Our analysis indicates that hydration structure around the leucine side chain is more ordered than water near the glutamine side chain, whereas the hydration water near their backbones is less ordered and largely equivalent between the two amino acids. The special role played by pentagons near the leucine side chain provides qualitative support for clathrate analogies, especially their topological role in enclosing hydrophobic solutes. The complete analogy breaks down because of the still significant presence of polygon sizes beyond octagons and the observation that the ordered structure is not rigid, but breaks up and reforms throughout the simulation.

The altered hydration structure near the leucine side chain, characterized by highly connected water vertices forming rings that are dominated by planar pentagons in particular, extends about two solvation shells from the surface. This structural persistence length suggests that two leucine peptides in water would likely sense each other's presence at a minimum distance of ~10 Å. The distance could be even larger, because the two hydration shells comprising a large number of planar pentagonal rings confine water between them, further ordering the intervening bulk, and entropically driving hydrophobic association over somewhat larger distances. This possibility will be an interesting point to investigate in future work.

The room temperature scattering intensity difference between aqueous solutions of hydrophobic amino acids and pure water is similar to reported measured differences between ambient and supercooled water (Bosio et al., 1983; Bellissent-Funel et al., 1986). The structural interpretation of the shift in the main diffraction peak as water is supercooled is a subject of debate (Stanley and Teixeira, 1980; Stillinger, 1981; Speedy, 1984; Dore, 1990; Walrafen and Chu, 1995). Stillinger has suggested that when water forms bulky polyhedra around hydrophobic groups, a mean attraction between the solutes is induced, because the (more) planar faces of hydrogen-bonded water rings near one hydrophobic group have hydrogen-bonded sites that naturally interface with the more planar water ring faces near the other hydrophobic group (Stillinger, 1981). This mean attraction between bulky closed ring structures is also thought to be dominated by pentagons in particular (Speedy, 1984). In the case of mean attraction between general polyhedra (Stillinger, 1981), clathrates in particular (Walrafen and Chu, 1995), or domains of pentagons (Speedy, 1984), the effect is thought to be highly cooperative as the temperature decreases into the supercooled regime, culminating in a divergence in the temperature dependence of the concentration of these open structures at −46°C (Stillinger, 1981; Speedy, 1984).

The measured shift of the main diffraction peak for the hydrophobic solutions is ~0.05 $Å^{-1}$, which would correspond roughly to a decrease of temperature to ~0–10°C for the pure liquid. The structural origin of the shift in the main diffraction peak to a lower angle for aqueous hydrophobic solutions is found by our analysis to be due to an increase in the number of pentagon rings as well as larger ring structures, and a greater degree of planarity of pentagons in particular. This complex structural picture shows the versatility that the water network has to adapt to perturbations such as the introduction of a foreign solute, and might provide insight into structural changes of the water network as it is cooled and supercooled. Although we have made some definitive connection between the important role of pentagons in hydrophobic phenomena, the water structure connection between hydrophobic association and water as it is supercooled remains controversial and unresolved.

There are several directions to pursue in the immediate future. The first goal of the simulations and neutron scattering experiments is to derive the solute-solute potential of mean force (pmf) profiles to characterize both the magnitude and spatial range of side chain–side chain interaction. The simulated scattering curves using empirical force fields have provided reasonable quantitative agreement with the experimental results. Therefore we might be able to isolate the solute-solute pmf profiles from the experimental data by subtracting off the simulated water-water and water-solute correlations. We hope to extend these studies to derive spherically averaged potentials of mean force for several types of amino acid side-chain pairs.

The second goal is to continue to benchmark the simulation results against x-ray and neutron scattering experiments on other types of amino acids. Ultimately, molecular dynamics simulations will be required to characterize hydration forces between heterologous pairs of amino acids, as it is unlikely that neutron scattering experiments could be used for this purpose. Close interaction with experiment is especially important at present to ensure that the peptide and water parameters are quantitatively correct. Parameterization of solute-water potentials, if necessary, will also be beneficial for those who simulate proteins in water using explicit water models.

Finally, we envision that the derived potentials of mean force would represent "implicit" hydration potentials that would be interfaced with empirical protein force fields (or even ab initio calculations) to be broadly used in computational studies of protein structure prediction and folding (Daggett and Levitt, 1992, 1993; Caflisch and Karplus, 1994; Li and Daggett, 1996). Using the hydration potentials of mean force alone will clearly make exhaustive searches more feasible than would fully explicit models, and their greater complexity in comparison to lattice models might make it possible to address important questions regarding the requirements for folding. During later stages of these simulated pathways, we could usefully introduce the more detailed protein force fields to provide a tertiary structure prediction with atomic resolution.

THG gratefully acknowledges support from the Air Force Office of Sponsored Research (grant no. FQ8671-9601129), the U.S. Department of Energy (contract number DE-AC-03-76SF00098), and the National Energy Research Supercomputer Center for computer time. JMS is supported by a National Science Foundation Graduate Research fellowship. AP and RMG acknowledge support from the W. M. Keck Program in Biomedical Research, National Institutes of Health training grants in molecular biophysics and biotechnology, and the support of the Committee on Research of the University of California at Berkeley.

REFERENCES

Allen, M. P., and D. J. Tildesley. 1987. Computer Simulation of Liquids. Clarendon Press, Oxford.

Andersen, H. C. 1983. Rattle: a "velocity" version of the shake algorithm for molecular dynamics calculations. *J. Comput. Phys.* 52:24–34.

Bellissent-Funel, M.-C., J. Teixeira, L. Bosio, J. Dore, and P. Chieux. 1986. Spatial correlations in deeply supercooled water. *Europhys. Lett.* 2:241–245.

Berendsen, H. J. C., J. P. M. Postma, W. F. van Gunsteren, and J. Hermanns. 1981. Interaction models for water in relation to protein hydration. *In* Intermolecular Forces. B. Pullman, editor. Reidel, Dordrecht, the Netherlands. 331–342.

Bosio, L., J. Teixeira, J. C. Dore, D. Steytler, and P. Chieux. 1983. Neutron diffraction studies of water. IV. The supercooled liquid region (≥ -14.5°C). *Mol. Phys.* 50:733–740.

Caflisch, A., and M. Karplus. 1994. Molecular dynamics simulation of protein denaturation: solvation of the hydrophobic cores and secondary structure of barnase. *Proc. Natl. Acad. Sci. USA.* 91:1746–1750.

Cornell, W. D., P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman. 1995. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* 117: 5179–5197.

Corongiu, G., and E. Clementi. 1992. Liquid water with an ab initio potential: x-ray and neutron scattering from 238 to 368K. *J. Chem. Phys.* 97:2030–2038.

Daggett, V., and M. Levitt. 1992. A model of the molten globule state from molecular dynamics simulation. *Proc. Natl. Acad. Sci. USA.* 89: 5142–5146.

Daggett, V., and M. Levitt. 1993. Protein unfolding pathways explored through molecular dynamics simulation. *J. Mol. Biol.* 232:600–619.

Davidson, D. W. 1973. *In* Water, A Comprehensive Treatise, Vol. 2. F. Franks, editor. Plenum, New York. 115–145.

Dill, K. A. 1990. Dominant forces in protein folding. *Biochemistry.* 29: 7133–7154.

Dore, J. C. 1990. Hydrogen-bond networks in supercooled liquid water and amorphous/vitreous ices. *J. Mol. Struct.* 237:221–232.

Dubochet, J., J. Lepault, R. Freeman, J. A. Berriman, et al. 1982. Electron microscopy of frozen water and aqueous solutions. *J. Microsc.* 128: 219–237.

Dunbrack, R. L., Jr., and M. Karplus. 1994. Conformational analysis of the backbone-dependent rotamer preferences of protein sidechains. *Nature Struct. Biol.* 1:334–340.

Egelstaff, P. A. 1992. An Introduction to the Liquid State. Oxford Series on Neutron Scattering in Condensed Matter. Oxford University Press, Oxford.

Finney, J. L., and A. K. Soper. 1994. Solvent structure and perturbations in solutions of chemical and biological importance. *Chem. Soc. Rev.* 23: 1–10.

Franks, F. 1975. *In* Water, A Comprehensive Treatise, Vols. 2 and 4. F. Franks, editor. Plenum, New York.

Head-Gordon, T. 1995. Is water structure around hydrophobic groups clathrate-like? *Proc. Natl. Acad. Sci. USA.* 92:8308–8312.

Kauzmann, W. 1959. Some factors in the interpretation of protein denaturation. *Adv. Protein Chem.* 14:1–63.

Li, A., and V. Daggett. 1996. Identification and characterization of the unfolding transition state of chymotrypsin inhibitor 2 by molecular dynamics simulation. *J. Mol. Biol.* 257:412–429.

Mettananda, I. H. I. U., and R. J. Speedy. 1984. Supercooled aqueous tetra-*n*-butylammonium bromide solutions. Conductivity and differential thermal analysis studies. *J. Phys. Chem.* 88:4163–4166.

Pertsemlidis, A. 1995. A solution scattering approach to defining the importance of hydration forces in protein folding. Ph.D. thesis. University of California, Berkeley, CA.

Pertsemlidis, A., A. Saxena, A. K. Soper, T. Head-Gordon, and R. M. Glaeser. 1996. Direct structural evidence for modified solvent-structure within the hydration shell of a hydrophobic amino acid. *Proc. Natl. Acad. Sci. USA.* 93:10769–10774.

Pratt, L. R., and D. Chandler. 1977. Theory of the hydrophobic effect. *J. Chem. Phys.* 67:3683–3704.

Rahman, A., and F. H. Stillinger. 1973. Hydrogen-bond patterns in liquid water. *J. Am. Chem. Soc.* 95:7943–7948.

Speedy, R. J. 1984. Self-replicating structures in water. *J. Phys. Chem.* 88:3364–3373.

Speedy, R. J., J. D. Madura, W. L. Jorgensen. 1987. Network topology in liquid water. *J. Phys. Chem.* 91:909–913.

Speedy, R. J., and M. Mezei. 1985. Pentagon-pentagon correlations in water. *J. Phys. Chem.* 89:171–175.

Stanley, H. E., and J. Teixeira. 1980. Interpretation of the unusual behavior of H_2O and D_2O at low temperatures: tests of a percolation model. *J. Chem. Phys.* 73:3404–3422.

Stillinger, F. H. 1980. Water revisited. *Science.* 209:451–457.

Stillinger, F. H. 1981. *In* Waters in Polymers, ACS Symposium Series, Vol. 127. S. P. Rowland, editor. American Chemical Society, Washington, DC. 11–22.

Stillinger, F. H., and T. A. Weber. 1983. Inherent structure in water. *J. Phys. Chem.* 87:2833–2840.

Stillinger, F. H., and T. A. Weber. 1984. Packing structures and transitions in liquids and solids. *Science.* 225:983–989.

Swope, W. C., H. C. Anderson, P. H. Berens, and K. R. Wilson. 1980. A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: application to small water clusters. *J. Chem. Phys.* 76:637–649.

Walrafen, G. E., and Y. C. Chu. 1995. Shear viscosity and self-diffusion evidence for high concentrations of hydrogen-bonded clathrate-like structures in very supercooled liquid water. *J. Phys. Chem.* 99: 10635–10643.