

## Simulation of Water around a Model Protein Helix. 2. The Relative Contributions of Packing, Hydrophobicity, and Hydrogen Bonding

Mark Gerstein<sup>†‡</sup> and R. M. Lynden-Bell<sup>\*‡</sup>

MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, U.K., and University Chemical Laboratory, Lensfield Road, Cambridge CB2 1EW, U.K.

Received: August 13, 1992; In Final Form: November 20, 1992

In the preceding paper, the structure of water around a model protein  $\alpha$ -helix (made from polyalanine) was investigated using two-dimensional projections of the molecular distribution function. Here an attempt is made to assess the relative importance of packing, protein–water hydrogen bonding, and water–water hydrogen bonding in creating this water structure. To isolate the effect of protein–water hydrogen bonding, simulations with the helix charges “switched” on and off were compared. Likewise, these “normal” water simulations were compared to ones done with the water charges switched off to assess the relative contributions of packing and hydrogen bonding. The energy of water molecules around the helix was also investigated. The results show that water–water hydrogen bonding, which underlies hydrophobicity, is the dominant interaction. On average it moves water molecules back from hydrophobic parts of the helix surface as compared to water molecules around hydrophilic parts. Furthermore, completely disrupting this interaction by switching off the water charges moves the solvent molecules in toward the helix, making narrow crevices more accessible to solvent. This result has important implications for the interpretation of Richards–Connelly molecular surfaces.

### I. Introduction

The generally accepted view of the structure of proteins is that three types of nonbonded interactions predominate.<sup>2,3</sup> Hydrogen bonding, which is primarily electrostatic in nature,<sup>4</sup> plays a major role in the stability of secondary structures, i.e.,  $\alpha$ -helices and  $\beta$ -sheets. Hydrophobic interactions, which cause apolar side chains to face inward and polar ones to face outward, are responsible for the overall globular shape of protein tertiary structures. Tight packing and intercalation of side chains determines the details of tertiary structure. Despite the complex and varied shape of the 20 side chains, there are remarkably few gaps and cavities in the protein interior. A folded protein in a sense resembles a solved “jigsaw puzzle”.<sup>5</sup> Research has focused on this close packing as the key to understanding protein structure.<sup>6–8</sup> Protein–protein recognition, moreover, again seems to involve tight packing of the contacting interfaces,<sup>9</sup> so when antibodies bind to antigens or serine proteases bind to their inhibitors, the interfaces pack closely and fit remarkably tightly.

Similarly, the structure of simple liquids, such as liquid argon or liquid carbon dioxide, can be well accounted for in terms of the close packing of hard spheres as the forces between molecules are dominated by harsh, short-ranged repulsions from interaction sites.<sup>10–14</sup> Attractive forces play only a minor role and can often be approximated by a spatially uniform background potential—i.e., a mean field. Liquid water, however, is a different story. The hydrogen bonding between molecules is an attractive force strong enough to be competitive with the hard-core repulsions. The hydrogen bonds are highly directional and give water a more open structure (coordination number 4, tetrahedral) than that for close packed spheres (coordination number 12). Although still not completely understood, the structure of liquid water can be conceptualized as a random network of hydrogen bonds punctuated by mobile defects.<sup>15–18</sup>

Thus, the structure of a folded protein or protein–protein complex is primarily determined by close packing while the structure of liquid water, by hydrogen-bonding. What happens when these two dissimilar substances meet? What is responsible for structuring the interface? If tight packing were primarily responsible for structuring the interface, one would expect the

water structure to conform to all the bumps and grooves in the van der Waals envelope of the protein. If hydrophobicity were responsible for structuring the interface, one might expect water–water hydrogen bonding to be crucial and the water molecules to be arranged in clathrate-like cages. Alternatively, if protein–water hydrogen bonding were responsible for structuring the interface, one would expect hydrogen-bond acceptors and donors on the protein to be associated with notable features in the water structure.

In the previous paper,<sup>1</sup> the distribution and orientation of water molecules around a model protein helix were calculated in a Monte Carlo simulation. The model system contained 14 residues (4 turns) of polyalanine  $\alpha$ -helix surrounded by 321 waters at 300 K in a periodic cell of dimensions  $22.219 \times 22.219 \times 20.93$  Å. The internal coordinates of the helix and water were kept rigid, and the interaction between molecules was parameterized by TIP3P<sup>19</sup> nonbonded parameters on the water and CHARMM<sup>20</sup> parameters on the helix. Helical and straight projections of the molecular distribution functions were calculated.

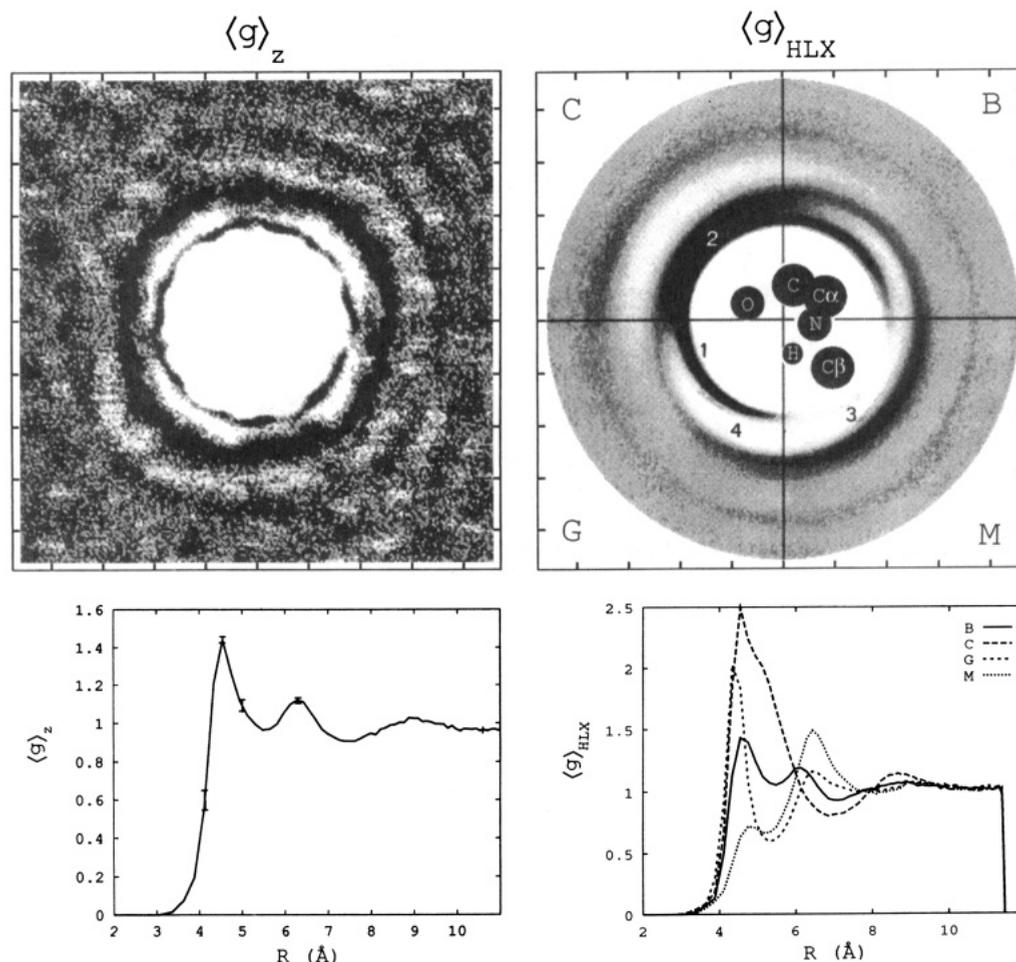
Here an attempt is made to assess the relative contributions of packing, hydrophobicity, and protein–water hydrogen bonding in structuring the solvent around a polyalanine helix. Three additional simulations were done: one with the charges on the helix switched off and two with the charges on the water switched off. Switching off the charges on the helix removes and hence highlights the effect of protein–water hydrogen bonding. Likewise, turning off the water charges removes the effect of water–water hydrogen bonding, leaving only van der Waals (packing) forces. These new simulations are compared with the original simulation of “normal” water around a charged helix. In addition, the energetics of the water around the helix are analyzed in all four simulations.

### II. Comparison of Four Simulations

**A. “Normal” Water around a Charged and Uncharged Helix.** Previously,<sup>1</sup>  $8.76 \times 10^5$  cycles of Monte Carlo simulation were done on 321 water molecules around a polyalanine helix. This will be referred to as the *standard* simulation since the water and the helix were normally charged. Here a second simulation was done with the helix charges set to zero. This will be referred to as the *uncharged-helix* simulation. This simulation used the same

<sup>†</sup> MRC Laboratory of Molecular Biology.

<sup>‡</sup> University Chemical Laboratory.



**Figure 1.** The oxygen distribution in the standard simulation (normal water around a normally charged helix). (a, left) Straight projection of oxygen distribution,  $\langle g(r) \rangle_z$  (in the notation of the previous paper<sup>1</sup>). The helix is centered at  $R = 0$  and in this projection has approximate 7-fold symmetry in  $\phi$ . The top half of figure shows the projection in "gray-level", where black is high. Tick marks are spaced 2.5 Å apart. The innermost ring of density (at  $R = 4.6$  Å) has maxima of  $\sim 1.7$  and minima of  $\sim 0.4$ . The bottom half of the figure shows the radial profile of the projection (i.e., after averaging over  $\phi$ ). Error bars are estimated by deviations from 7-fold symmetry. (b, right) Helical projection of oxygen distribution:  $\langle g(r) \rangle_{HLX}$ . The top half of the figure shows the projection in "gray-level", where black is high. Tick marks are spaced 2.5 Å apart. Positions of the helically projected protein atoms are indicated by circles drawn with a radius one-third of the van der Waals radius. The values of selected peaks and troughs, indicated by footnote numbers, are as follows: maximum near 1, 3.5; maximum near 2, 2.5; minimum near 3, 0.41; minimum near 4, 0.53. The bottom half of the figure shows averages of the projection in each of the four quadrants as a function of the distance from the helix axis ( $R$ ). These quadrants are indicated here by B (backbone), C (carbonyl), G (gap), and M (methyl).

heat-cool-run protocol as the standard simulation and consisted of a total of  $6.75 \times 10^5$  Monte Carlo cycles.

The straight projection of the oxygen distribution illustrates the overall effect of turning off the helix charges. It is shown for the standard simulation in Figure 1a and for the uncharged helix simulation in Figure 2a. In the standard simulation, there are three peaks at 4.6, 6.3, and 8.9 Å from the helix axis. These peaks can be associated with three shells of water molecules around the helix. In the uncharged-helix simulation, the height of the first peak decreases from 1.5 to 1.1, and much of the density moves to the second peak, which, in turn, increases from 1.1 to 1.2 and widens. That is, *on average* the water molecules "move back", away from the helix.

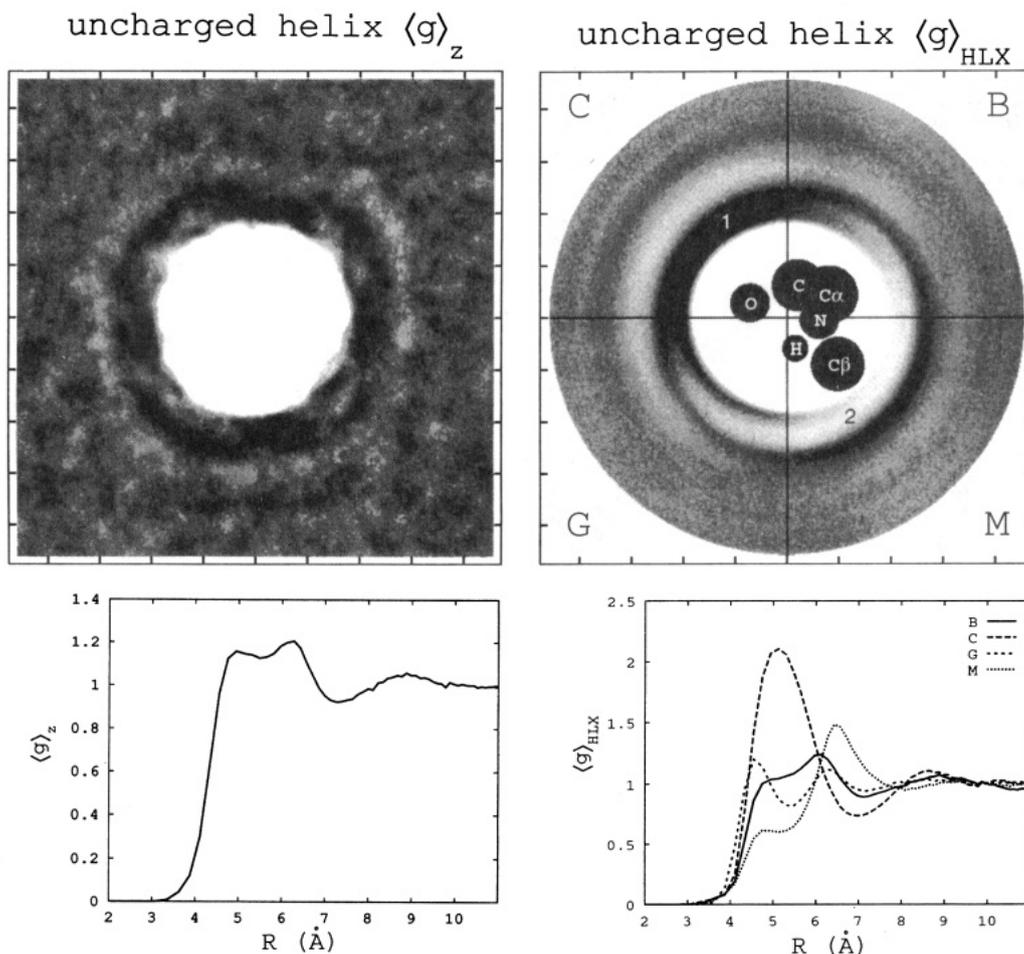
In both simulations, the second and third shells form a continuous band around the helix. In contrast, the first shell has some axial structure. The helical projection of the oxygen distribution, shown in Figures 1b and 2b, reveals the source of this axial structure. In the standard simulation the helical projection shows three well-defined peaks in the backbone and methyl quadrants. In contrast, in the carbonyl quadrant two of the inner peaks have merged into a single large peak. This merged peak is more than twice as high as the first peak in the methyl quadrant. In the uncharged-helix simulation, the helical projection shows that the water "moves back" away from the helix in all quadrants except the methyl quadrant, which is already

uncharged and hydrophobic. There is still a significant density maximum near the carbonyl oxygen, which suggests that packing geometry, in addition to hydrogen bonding, is responsible for the high probability of water molecules surrounding it.

In addition to "moving back" away from the helix, the first-shell water molecules in the uncharged-helix simulation rotate their hydrogens away from the helix so that they can hydrogen bond to other solvent molecules. This change in water orientation shows up clearly in Figure 3, which shows the hydrogen distribution for the standard and uncharged-helix simulations. When the helix charges are switched off, the seven peaks at 3.6 Å, which result from hydrogen bonding to the carbonyl oxygen, completely disappear. However, the peak at  $\sim 5.5$  Å, which comes from hydrogens sandwiched between the first and second water layers, is essentially unchanged.

**B. Helix in Liquid and Fluid "Argon".** Thus, turning off the charges on the helix highlights the effect of protein-water hydrogen bonding. Likewise, switching off the charges on the water highlights the effect of water-water hydrogen bonding.

However, switching off the charges on the water was not as straightforward as switching them off on the helix, so two simulations were done. In the first simulation, no changes were made besides setting all the water charges to zero. As the original TIP3P model has no Lennard-Jones parameters for the hydrogens, this simulation was effectively of a Lennard-Jones liquid with



**Figure 2.** The oxygen distribution in the uncharged-helix simulation. The same conventions as in Figure 1 are used. (a, left) Straight projection,  $\langle g(r) \rangle_z$ . Compared to the oxygen distribution in the standard simulation water molecules have on average "moved back" from the helix. (b, right) Helical projection,  $\langle g(r) \rangle_{HLX}$ . Maximum 1 near the carbonyl is 2.5, which is much lower than the corresponding point in the standard simulation, and minimum 2 near the  $\beta$ -carbon is 0.5.

oxygen TIP3P parameters. Consequently, it is worthwhile to convert the conditions of the simulation to reduced units for comparison with other Lennard-Jones simulations, such as those of liquid argon. The definitions for reduced units are

$$\rho^* = \rho \sigma^3 \quad T^* = kT/\epsilon \quad (1)$$

where  $\sigma$  is the Lennard-Jones radius and  $\epsilon$  is the Lennard-Jones well depth. To calculate the density  $\rho$ , it is necessary to know the volume to the simulation box accessible to water. The box accommodates 375 waters without the helix and 321 with it, so the accessible volume  $V_A$  was estimated at  $(321/375)a^2c$ , where  $a^2c$  is the volume of the cell. Note, there are other ways to estimate this accessible volume, particularly involving the use of tabulated residue volumes<sup>21,22</sup> and Voronoi polyhedra.<sup>7</sup>

In reduced units, the first simulation has  $\rho^* = 1.04$  and  $T^* = 3.92$ . These values do not correspond to those for a Lennard-Jones liquid. For liquid argon, such a reduced temperature corresponds to a real temperature of 470 K, which is far above the critical temperature of 151 K.<sup>14</sup> Consequently, the first simulation was of a supercritical fluid at high pressure. It does not correspond to simulating a liquid.

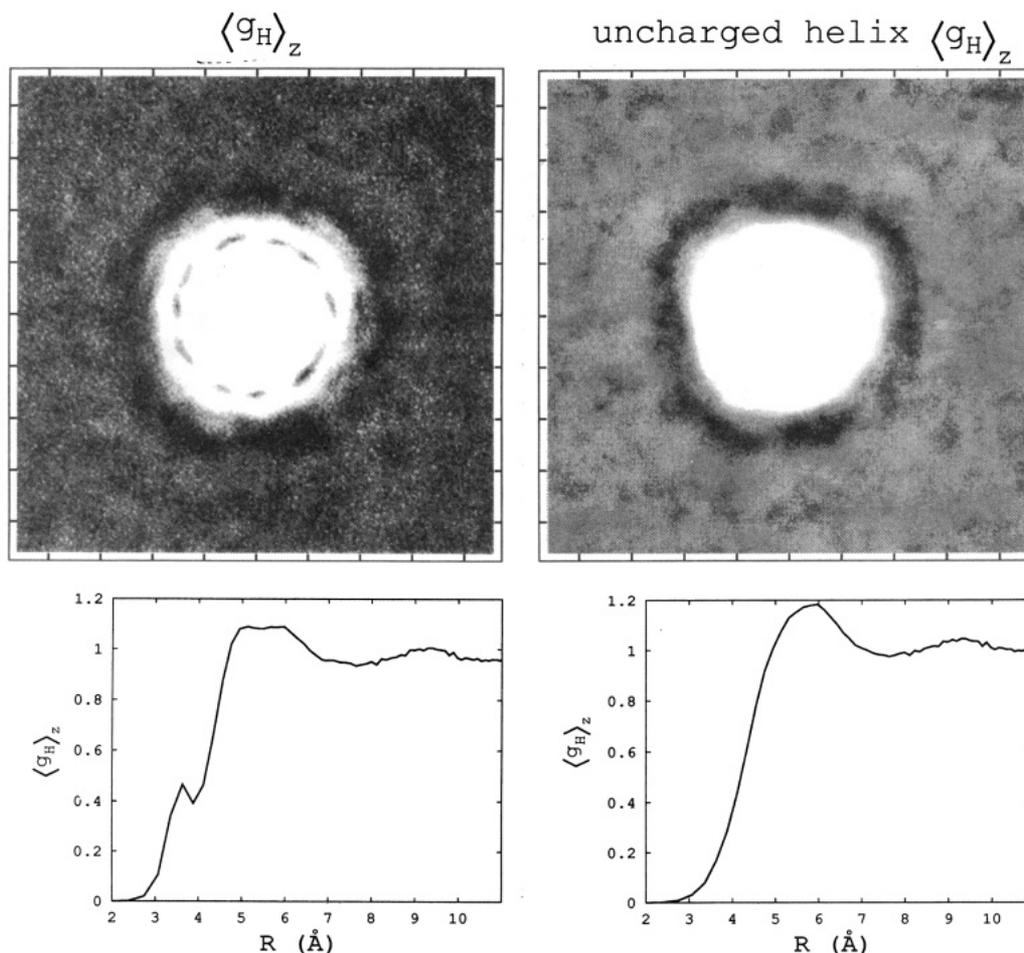
In the second simulation, the number of solvent molecules and the temperature were adjusted to give values appropriate for liquid argon. The conditions  $T^* = 0.94$  and  $\rho^* = 0.88$  were chosen from past simulations of liquid argon.<sup>23</sup> For the TIP3P Lennard-Jones parameters, these conditions correspond to a simulation with 272 solvent molecules in the same size cell at a temperature of 72 K. Because of the similarity of the two simulations of "uncharged water" to the fluid and liquid states of argon, they will be referred to as the *fluid* and *liquid* "argon" simulations,

respectively. They will be collectively contrasted with the two simulations of "normal" water around a charged and uncharged helix.

Liquid argon equilibrates more rapidly than water, so it was only necessary to run the fluid and liquid simulations for  $2.06 \times 10^5$  and  $2.66 \times 10^5$  Monte Carlo cycles, respectively. Both simulations used the same nonbonded cutoffs as the normal water simulations: 7.5 Å, which corresponds to  $2.4\sigma$ . As was the case with the normal water simulations, longer cutoffs (up to 10 Å) were tried and not found to affect the results significantly.

In both "argon" simulations the Lennard-Jones parameters ( $\epsilon$  and  $\sigma$ ) for the helix atoms were unchanged. This parametrization kept the relative sizes of the helix and the solvent the same. It also necessarily implies that "argon" atoms will not be able to approach the helix any more closely than water molecules since the distance of closest approach is determined by the harsh Lennard-Jones repulsion. However, a scaling similar to that achieved in the liquid "argon" simulation could also have been accomplished by multiplying all the Lennard-Jones parameters by an appropriate factor. Since such a scaling would not change the coordinate positions of the atoms, it would implicitly distort the geometry of the system and shift the closest approach of an "argon" atom to the helix.

Highlighting the importance of water-water hydrogen bonding relative to packing, switching off the charges on the water produces much larger changes than switching off the charges on the helix. Figures 4 and 5a show the straight projection of the "argon" distribution in the both fluid and liquid simulations. The results from both are similar, which is encouraging since the simulations suffer from different distortions. In both simulations, the



**Figure 3.** Hydrogen distribution in (a) the standard simulation and (b) the uncharged-helix simulation. Same conventions as in Figure 1a are used. The inner peaks, which result from water molecules hydrogen bonding to the carbonyl oxygens, completely disappear in the uncharged-helix simulation.

Lennard-Jones well-depth parameters  $\epsilon$  for the helix were unchanged. These parameters were developed for room temperature (300 K) simulation and may be too large (and attractive) for simulations at 72 K. Consequently, some of the structure in the liquid "argon" simulation may reflect local solidification near the helix. The fluid "argon" simulation, in contrast, may have an overly gaseous structure—particularly far from the helix.

In both simulations, the "argon" molecules *on average* "move forward" towards the helix in contrast to the water molecules in the standard simulation. This movement is in the opposite direction to that found in the uncharged-helix simulation and is a manifestation of the greater surface "wetting" expected with a Lennard-Jones fluid. The liquid "argon" simulation manifests this "wetting" more dramatically than the fluid one and clearly exhibits many well-defined solvent shells around the helix. The radial distribution has five peaks within 8 Å of the helix axis in contrast to the two found in the standard simulation. The peaks close to the helix are much higher (with magnitudes up to 8) than in the standard simulation. Consequently, an "argon" molecule has a greater probability of being close to the helix than a normal water molecule does. However, as discussed above it does not approach the helix any more closely in an absolute sense.

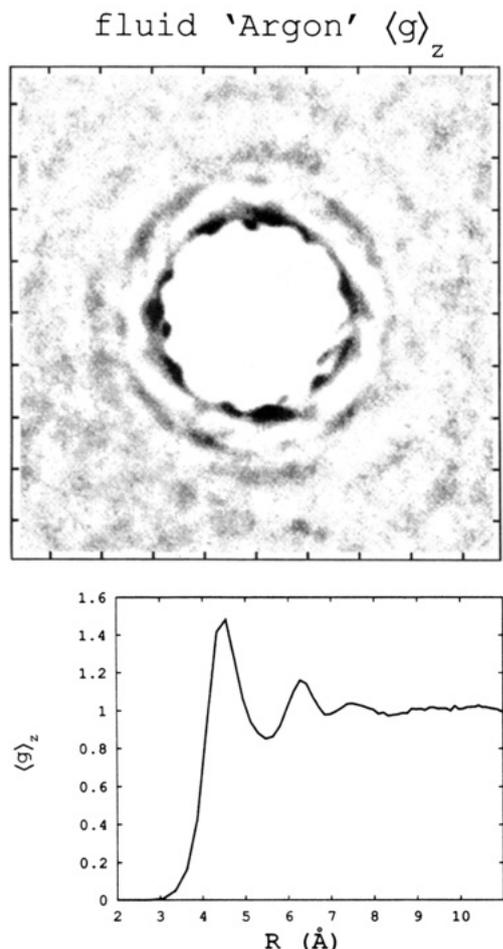
The multiplicity of high peaks in the "argon" simulations results from solvent molecules packing into narrow crevices on the helix surface. These crevices are evident in the helical projection of the "argon" distribution, which is shown for the liquid simulation in Figure 5b. There are peaks (marked by 1, 2, and 3) on either side of the carbonyl oxygen, where inspection of the van der Waals envelope of the helix reveals narrow grooves. The peak in the gap quadrant (marked with a 1) extends much closer to the  $\beta$ -carbon than the analogous peak in the normal water simulations. In fact, in the normal water simulations there is a clear minimum

around the  $\beta$ -carbon (marked by a 3 in Figure 1b). The implication is that there is a narrow crevice around the  $\beta$ -carbon that is geometrically accessible to a water molecule but not normally filled because of hydrogen-bonding constraints. The sharp minimum (marked by a 4) on the boundary between the gap and carbonyl quadrants presents the opposite situation. It occurs approximately where there is a high peak in the standard simulation (marked with a 1 in Figure 1b) and where the water was previously<sup>1</sup> found to be most strongly oriented because of hydrogen bonding to the carbonyl oxygen. Consequently, this position appears to be favorable in terms of hydrogen bonding but unfavorable in terms of packing.

The helical projection of the solvent distribution in the fluid "argon" simulation has peaks arranged in roughly the same fashion as in the liquid simulation and so is not shown. The heights of these peaks, however, are significantly attenuated—roughly in the same proportion as the peak heights in the straight projection.

### III. Energy of Water Molecules around the Helix

**A. Radial Variation in Water–Water and Water–Protein and Total Energy.** When a protein is dissolved in water, the energy of the solvent molecules changes due to two factors: the direct interaction of a water molecule with the protein and the change in the interaction of water molecules with each other. These quantities can be measured separately in a simulation. The water–water energy of a solvent molecule  $U_{ww}$  is defined as half (to avoid double counting) the sum of its pairwise interactions with neighboring solvent molecules. Here, this energy is averaged over the simulation for molecules at position  $r$  (essentially a weighted average over molecular orientation), put on a per molecule basis, and expressed relative to the bulk to give  $\Delta U_{ww}$ ,



**Figure 4.** Straight projection of solvent distribution,  $\langle g(\mathbf{r}) \rangle_z$ , in the fluid "argon" simulation. Same conventions as in Figure 1a are used. The first ring of peaks has seven maxima of magnitude  $\sim 2.1$ , and the next ring,  $\sim 1.7$ .

the average change in water–water energy relative to the bulk for a molecule at position  $\mathbf{r}$ :

$$\Delta U_{\text{ww}}(\mathbf{r}) = \langle U_{\text{ww}}(\mathbf{r}\omega) \rangle_{\omega} - U_{\text{bulk}} \quad (2)$$

where  $\langle \rangle_{\omega}$  represents a weighted average over orientations (in the notation of the previous paper<sup>1</sup>) and  $U_{\text{bulk}} = -41.3$  kJ/mol ( $-16.6kT$ ) is the energy of a bulk TIP3P water molecule at 300 K at a density of 0.982 g/mL.<sup>19</sup>

The water–water energy is counterbalanced by the interaction energy with the protein  $U_{\text{wp}}$ , which is calculated in a similar fashion by averaging over the simulation. The total energy  $\Delta U_{\text{tot}}$  of a water molecule relative to the bulk is the sum of the water–water and water–protein energies:  $\Delta U_{\text{tot}} = \Delta U_{\text{ww}} + U_{\text{wp}}$ . It is sometimes called the binding energy to emphasize that water molecules with negative total energy are "bound" to the protein. The solvation energy of a protein (relative to a hypothetical gaseous phase) is given by the total energy summed over all water molecules:

$$U_{\text{solv}} = (N/V_A) \int g(\mathbf{r}) \Delta U_{\text{tot}}(\mathbf{r}) \, d\mathbf{r} \quad (3)$$

where  $N$  is the number of waters in the cell and  $V_A$  is the accessible volume of the cell (defined above).

Figure 6 shows the water–water, water–protein, and total energy as a function of distance from the helix axis for all four simulations. Because these energies are averages per molecule, there are statistical problems close to the helix where the averaging is done over few molecules; so the energies are only shown from 4 Å outward. The total energy for a first-shell water molecule in the standard simulation,  $\sim -7.5$  kJ/mol ( $3kT$  at 300 K), is similar to that found in previous simulations.<sup>24</sup> Switching off the helix

charges decreases the water–water energy by  $\sim 1$  kJ/mol but raises the protein interaction energy by  $\sim 7.5$  kJ/mol. Consequently, the total energy in the uncharged-helix simulation increases to  $\sim -1$  kJ/mol. In both simulations the total energy rapidly vanishes with increasing distance from the helix axis, and after the second shell a water molecule has nearly the same energy as in the bulk.

Switching off the charges on the water does not change the protein–interaction energy very much compared to the uncharged-helix simulation. However, it decreases the water–water energy by  $\sim 6$  kJ/mol. This decrease is roughly the same in magnitude as the increase in the protein–interaction energy when the helix charges are switched off. Consequently, the liquid and fluid "argon" simulations have nearly the same total energy *relative to the bulk* as the standard simulation. Note that the water–water (or more properly solvent–solvent) energy for the "argon" simulations is expressed relative to bulk liquid or fluid "argon", which have very different energies ( $U_{\text{bulk}}$ ) than bulk TIP3P water. Consequently, while the *difference*  $\Delta U_{\text{ww}}$  is lower for the "argon" simulations than the normal water simulations, the value of  $U_{\text{ww}}$  is actually higher.

In the standard simulation, the total solvation energy per alanine residue is  $-17$  kJ/mol ( $-6.8kT$ ). In the uncharged helix simulation, it increases to  $-5$  kJ/mol ( $-2kT$ ), indicating much less favorable hydration of this system. In both "argon" simulations it is not much more than in the standard simulation:  $-10$  kJ/mol ( $-4kT$ ) for the fluid one and  $-15$  kJ/mol for the liquid one. Expressed in units of  $kT$  at 72 K, the solvation energy for the liquid simulation becomes  $-25kT$ .

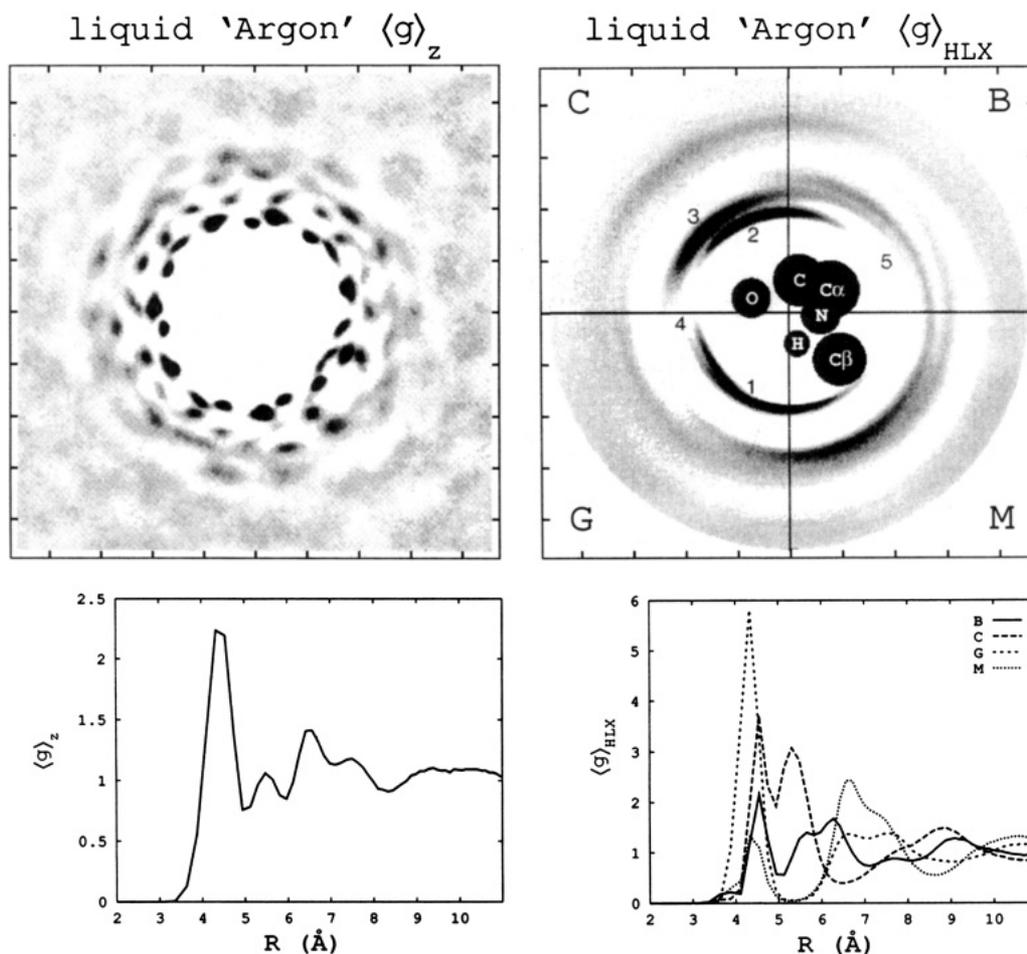
**B. Variation in Energy around the Helix for Normal Water.** Thus far only the energy variation with distance from the helix has been discussed. The helical projection provides additional information on the variation around the helix. Figure 7 shows the helical projection of the water–protein and water–water energy in the standard simulation. The water–protein energy is  $\sim 5$  kJ/mol lower for first-shell water molecules around the hydrophilic half of the helix than around the hydrophobic half (i.e., gap and carbonyl quadrants versus backbone and methyl quadrants). In contrast, the water–water energy is  $\sim 2.5$  kJ/mol higher around the hydrophilic half.

The total energy combines these opposing contributions. As is evident in Figure 6, it is dominated radially by the water–protein energy and so gradually increases with distance from the helix. At a given distance, the smaller variation around the helix is superimposed onto this increasing radial background and so is difficult to see in "gray level". To overcome this difficulty, the average radial background (Figure 6) is subtracted from the total energy in projecting it down to two dimensions. The new projection that results only gives deviations from the mean at a given distance from the helix axis. It will be denoted by  $\langle \rangle_{\delta}$ :

$$\langle \Delta U_{\text{tot}} \rangle_{\delta} = \langle \Delta U_{\text{tot}} \rangle_{\text{HLX}} = (1/2\pi) \int_0^{2\pi} \langle \Delta U_{\text{tot}} \rangle_{\text{HLX}} \, d\phi \quad (4)$$

where  $\langle \rangle_{\text{HLX}}$  represents helical averaging in the notation of the previous paper.<sup>1</sup> This projection of the total energy is shown in Figure 8. Closer than 5 Å to the helix axis, it is dominated by variations in the water–protein energy and so is lowest near the carbonyl oxygen. However, these variations die off rapidly. After 5 Å, the lower water–water energy around the hydrophobic atoms dominates and results in a particularly deep minimum near the  $\alpha$ -carbon (marked by 1 in the figure).

As might be expected, the variation in energy around the helix is much less in the uncharged-helix and "argon" simulations than in the standard simulation. Most of the useful energetic information in these simulations is contained in the radial profiles in Figure 6. In particular, comparison of the standard and uncharged helix simulations shows that the variation in water–water energy around the helix vanishes when the helix charges are switched off. The water–water interactions for first-shell



**Figure 5.** Solvent distribution in the liquid "argon" simulation. Same conventions as in Figure 1 are used. (a, left) Straight projection,  $\langle g(r) \rangle_z$ . The peaks are considerably higher than corresponding peaks in the fluid "argon" simulation in Figure 4 and much higher than those in the normal water simulations (Figures 1 and 2). Each peak in the first shell has a maximum height of  $\sim 7.9$ , and in next shell  $\sim 5.2$ . The radial profile at bottom indicates that the average (i.e., integrated) density in the first two shells is greater than in the standard simulation. This difference in average peak height implies that a liquid "argon" atom is *on average* closer to the helix than a water molecule. However, as it has the same Lennard-Jones radius, an "argon" molecule cannot approach the helix any more closely than a normal water molecule can. Consequently, the location of the first shell, unlike its height, is roughly the same in both the standard and liquid "argon" simulations. (b, right) Helical projection,  $\langle g(r) \rangle_{HLX}$ . Maximum near 1 is 10; maximum near 2, 8; minimum near 3, 4; minimum near 4, 0.8; and minimum near 5, 0.2. Compared to the standard simulation in Figure 1b, there is a minimum rather than a maximum near point 4 and a double rather than single-shell structure in the carbonyl quadrant (peaks 2 and 3).

waters get stronger (by  $\sim 2.5$  kJ/mol) around the carbonyl oxygen and, to a lesser extent, weaker (by  $\sim 1$  kJ/mol) around the  $\alpha$  and  $\beta$  carbons (Figure 9).

The variation in water-water and total energy around the helix is consistent with the traditional, but controversial, view of "iceberg" hydrophobic hydration.<sup>3,15,16,25-27</sup> In this view water molecules around hydrophobic surfaces orient themselves to maintain their hydrogen bonding. The resulting arrangement resembles a clathrate and allows each water molecule to make four hydrogen bonds to neighboring water molecules. Clathrate geometries have been found in simulations of water around small solutes<sup>28-31</sup> but not around extended hydrophobic surfaces.<sup>31-34</sup> In the preceding paper,<sup>1</sup> the orientation of the water molecules around the  $\beta$ -carbon was shown to be similar to that in a clathrate.

Because they maintain hydrogen bonding, clathrate geometries are enthalpically favorable. However, because they restrict water molecule orientations, clathrate geometries are entropically unfavorable. Here the lower water-water and total energy around the  $\alpha$  and  $\beta$  carbons is consistent with the increased water-water hydrogen bonding in such an enthalpically favorable clathrate geometry.

**C. Characterization of Entropic Changes and Hydrophobicity around the Helix.** Entropic effects are more difficult to analyze than enthalpic ones. To some degree, they can be measured by combining information from the total energy and the oxygen

distribution (which reflects the free energy). Here such a combination will be investigated, using the function  $\zeta$  defined by

$$\zeta(\mathbf{r}) = g(\mathbf{r})e^{\beta\Delta U_{\text{ox}}(\mathbf{r})} \quad (5)$$

The function  $\zeta$  is related to the cavity distribution function  $y$  often used in theory of liquids,<sup>35,36</sup> which for the system studied here would be  $y = g \exp(\beta U_{\text{wp}})$ .

Using the potential of mean force, it is possible to interpret  $\zeta$  in terms of entropy. The distribution of water molecules can be related to the potential of mean force  $W(\mathbf{r})$ , which is defined by

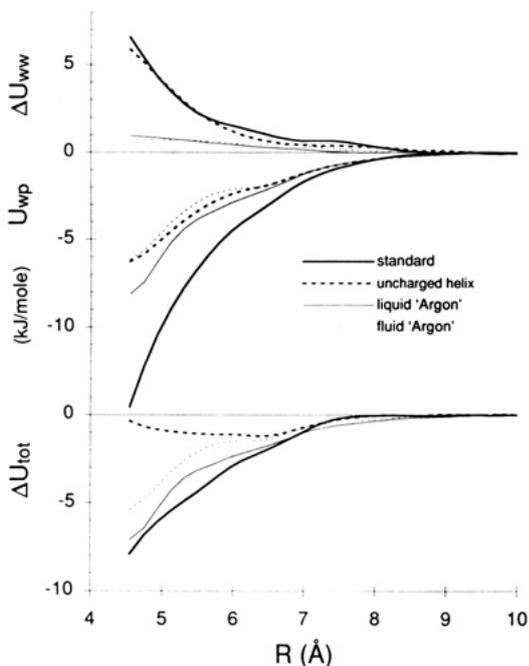
$$W(\mathbf{r}) = -kT \ln g(\mathbf{r}) \quad (6)$$

This quantity is the work done in reversibly bringing a water molecule from the bulk to position  $\mathbf{r}$  at constant temperature. It is equal to the change in Helmholtz free energy  $\Delta A(\mathbf{r})$  of a system constrained to have a water oxygen at position  $\mathbf{r}$  as compared to one with a water molecule constrained to be in the bulk—i.e., far from the helix at position  $\mathbf{r}_0$ , where  $g(\mathbf{r}_0) = 1$ .

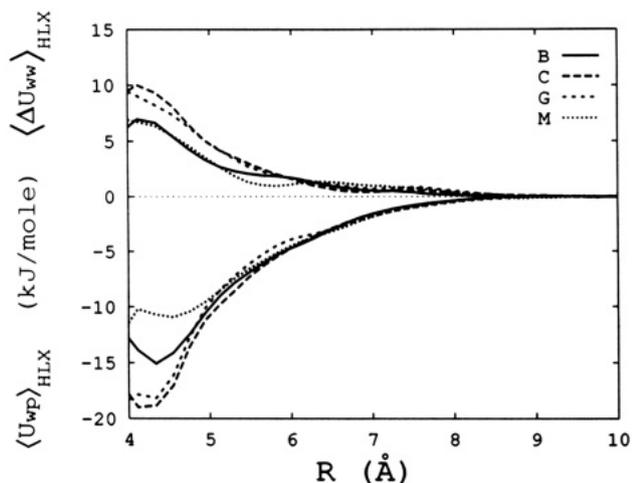
The entropy change of the constrained system can be related to the free energy by the usual relationship ( $\Delta A \equiv \Delta E - T\Delta S$ ), so that

$$\Delta S(\mathbf{r}) = (\Delta E(\mathbf{r})/T) + k \ln g(\mathbf{r}) \quad (7)$$

where  $\Delta E(\mathbf{r})$  is the energy change and  $\Delta S(\mathbf{r})$  is the entropy change of the whole system in moving a water molecule from position



**Figure 6.** Overall energies of the solvent in the four simulations. The water–water ( $\Delta U_{ww}$ , top), water–protein ( $U_{wp}$ , middle), and total energy ( $\Delta U_{tot}$ , bottom) for a solvent molecule at distance  $R$  from the helix axis is shown for the standard, uncharged-helix, fluid “argon”, and liquid “argon” simulations. All energies are expressed relative to that of bulk solvent. The energy of bulk TIP3P water is  $-41.3$  kJ/mol ( $-16.6kT$ ). It is different from the energy of the bulk fluid or liquid “argon”. Consequently, while the difference  $\Delta U_{ww}$  is lower for the “argon” simulations than the normal water simulations, the value of  $U_{ww}$  (see text) is actually higher. The higher  $U_{ww}$  reflects the fact that the water–water hydrogen bonding in normal water is attractive.

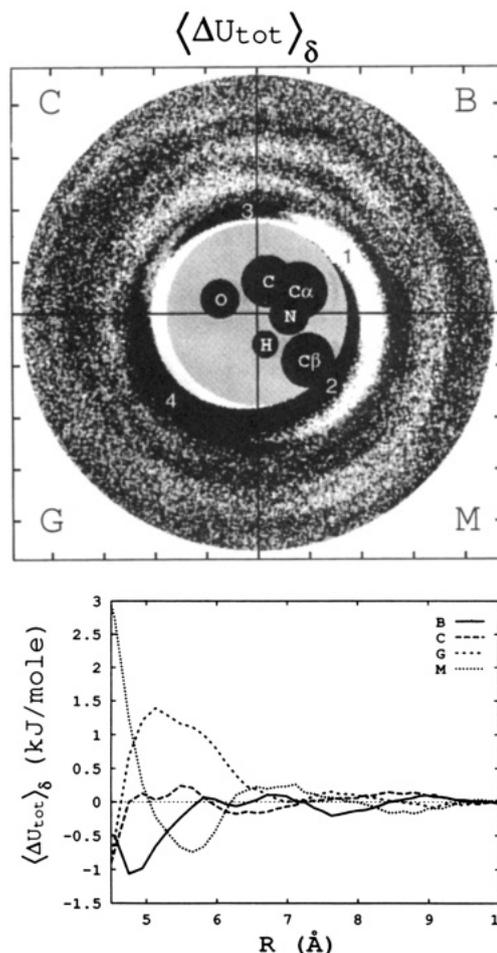


**Figure 7.** Solvent energies in the standard simulation. These are averaged over the quadrants in the helical projection and shown as a function of distance from the helix axis ( $R$ ). The top half of the figure shows the water–water energy  $\langle \Delta U_{ww} \rangle_{HLX}$ , and the bottom half, the water–protein energy,  $\langle \Delta U_{wp} \rangle_{HLX}$ . The quadrants are labeled B (backbone), C (carbonyl), G (gap), and M (methyl). Note that the water–protein energy is lower (more negative) in the carbonyl and gap quadrants than in the backbone and methyl quadrants.

$r_0$  to  $r$ . Insofar as the energy change of the whole system  $\Delta E(\mathbf{r})$  is dominated by the change in energy of the constrained molecule itself,  $\Delta U(\mathbf{r})$ , the entropy change is simply related to  $\zeta$ :

$$\Delta S(\mathbf{r}) \approx k \ln \zeta(\mathbf{r}) \quad (8)$$

However, there is also a contribution to  $\Delta E(\mathbf{r})$  from other water molecules in the system, which may, for example, not be able to hydrogen bond as well in the presence of the constrained molecule. Note that each side of the above equation is interpreted slightly



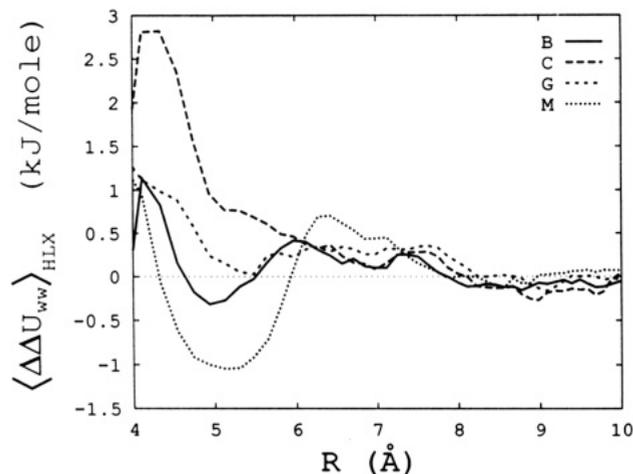
**Figure 8.** Spatial variation of the total solvent energy  $\Delta U_{tot}$ . As explained in the text, it is shown in helical projection as deviation from its average around the helix:  $\langle \Delta U_{tot} \rangle_{\delta}$ . The same conventions as in the helical projection in Figure 1b are used. However, because of the numerical uncertainties in averaging over the few water molecules very close to the helix, the values at points closer than  $4.5$  Å to the helix axis are not shown in “gray-level” at the top of the figure. Instead these points are uniformly filled with 50% gray. Minimum 1 near the  $\alpha$ -carbon is  $-2.5$  kJ/mol ( $-1kT$ ); maximum 2 near the  $\beta$ -carbon,  $5$  kJ/mol; maximum near 3,  $1.2$  kJ/mol; minimum near 4,  $-1.2$  kJ/mol.

differently, particularly with respect to the position  $r$ . That is, the equation implies that the average number of water molecules and their average energy in a grid box at  $r$  (the right-hand side) is closely related to the entropy change of the whole system in fixing a water oxygen at  $r$  (the left-hand side).

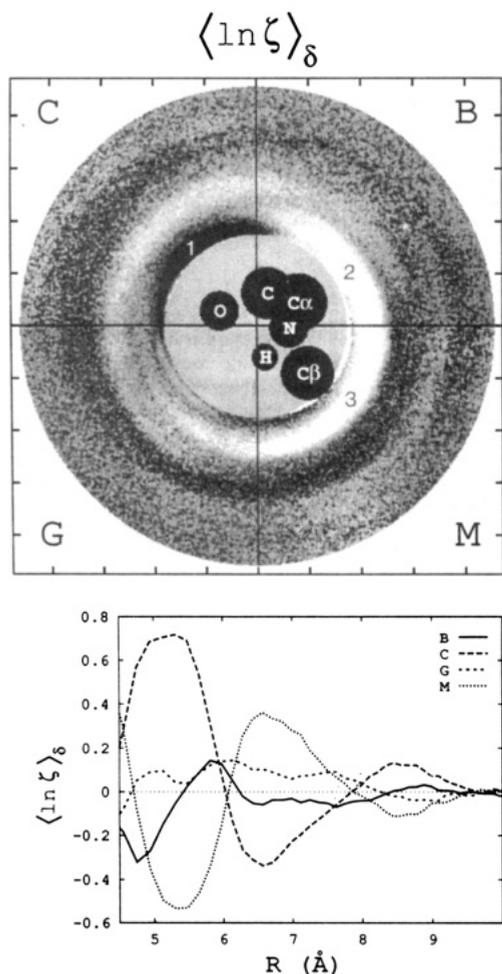
As  $\zeta$  is a synthesis of the oxygen distribution and the total energy, it shares features with both. In particular, like the total energy,  $\zeta$  has small variations around the helix superposed on a large radial increase, so in Figure 10 it is shown as  $\langle \ln \zeta \rangle_{\delta}$ , the deviation of  $\ln \zeta$  from its mean value at a given distance from the helix axis.

Around the helix,  $\langle \ln \zeta \rangle_{\delta}$  has an overall structure that corresponds well with chemical notions of hydrophobicity. It has a sharp peak around the hydrophilic carbonyl oxygen (marked by a 1) and an intermediate value in the gap quadrant. Most importantly, it has a broad minimum around both the  $\alpha$  and  $\beta$  carbons (marked by 2 and 3). This equal treatment of both hydrophobic carbons is not evident in either the oxygen distribution or the total energy. The total energy has a minimum near the  $\alpha$ -carbon and peak near the  $\beta$ -carbon (compare the points marked by 1 and 2 in Figure 8). The oxygen distribution has the opposite variation and has a much deeper minimum near the  $\beta$ -carbon (marked by 3 in Figure 1b) than near the  $\alpha$ -carbon.

Thus, variations in  $\zeta$  can be related to variations in entropy. In the standard simulation they are correlated with hydrophobicity



**Figure 9.** Change in water-water energy when the helix charges are turned off. This is shown in each of the four quadrants of the helical projection as the difference in water-water energy between the standard and uncharged-helix simulations:  $\langle \Delta \Delta U_{ww} \rangle_{HLX} = \langle \Delta U_{ww}^{std} - \Delta U_{ww}^{uncharged} \rangle_{HLX}$ . This difference is positive (i.e., indicating a weaker water-water interaction) in the carbonyl quadrant as water-protein hydrogen bonds are formed in place of water-water hydrogen bonds.



**Figure 10.** Spatial variation of  $\ln \zeta$ . Using the same conventions as Figure 8, it is shown in helical projection as  $\langle \ln \zeta \rangle_{\delta}$ , the deviation in  $\ln \zeta$  from its average around the helix.  $\ln \zeta$  synthesizes information from the oxygen distribution in Figure 1b and the total solvent energy in Figure 8. Note  $\langle \ln \zeta \rangle_{\delta}$  is positive around the hydrophilic carbonyl oxygen (maximum near 1 is 1.0) and negative around the hydrophobic  $\alpha$  and  $\beta$  carbons (minima near 2 and 3 are  $-0.8$  and  $-0.4$ ). It is, thus, better correlated with hydrophobicity than either the oxygen distribution or the total energy.

around the helix. However,  $\zeta$  alone can not completely characterize the entropic effects in all four simulations. As the

uncharged helix is more hydrophobic than the charged helix, it would be expected to order the water around the helix to a greater extent and overall to have lower entropy and hence lower  $\zeta$ . However, in the uncharged-helix simulation,  $\zeta$  is on average greater than in the standard simulation. This large value implies that  $\zeta$  is not useful in an absolute sense—i.e. in comparing two different simulations. However, there is less variation in  $\langle \ln \zeta \rangle_{\delta}$  around the uncharged helix than around the charged one. This agrees with expectations and suggests that  $\zeta$  is primarily useful in a relative sense—for comparing two regions in the same simulation.

#### IV. Conclusion

**A. Water-Water Hydrogen Bonding Is the Dominant Interaction.** The four simulations done here make it possible to analyze systematically the interactions at the helix-water interface. Comparison of the standard and uncharged-helix simulations isolates the contribution of water-protein hydrogen bonding, and comparison of the “argon” and normal water simulations helps to assess the relative contributions of water-water hydrogen bonding and packing. The results suggest that hydrogen bonding is more important than packing. The importance of water-water hydrogen bonding is evident in the similarity of the standard and uncharged-helix simulations in contrast to the liquid “argon” simulation. That is, the water distribution is only slightly perturbed when the helix charges are switched off but adopts a radically different structure when the water charges are turned off. This restructuring is particularly evident in the “gray-level” graphs of the helical projection. It is so radical that some favorable positions for hydrogen bonding to the carbonyl oxygen in the standard simulation are unoccupied in the liquid “argon” simulation, and some favorable packing positions in the “argon” simulation are unoccupied in the standard simulation due to the constraints of water-water hydrogen bonding.

Packing is significant in a secondary sense. Regions where a water molecule can easily maintain its hydrogen bonding—i.e., around hydrophilic atoms—can be packed tightly. In contrast, where a water molecule has difficulty satisfying its hydrogen-bonding requirement—i.e., around hydrophobic atoms—the packing is worse. This packing efficiency is directly reflected in the value of the molecular distribution function at a given point.

The analysis of water energetics further highlights the importance of hydrogen bonding. In the uncharged-helix and “argon” simulations, there is little variation in water-water and water-protein energy around the helix while in the standard simulation there is. This implies that the energy of a water molecule near the helix is more strongly affected by electrostatics and hydrogen bonding than by packing considerations (unless, of course, there are van der Waals overlaps). Consequently, the spatial variation in the total energy is markedly different from that in the molecular distribution function. The function  $\zeta$  is suggested as a synthesis of these two sources of information. Its variation around the helix can be interpreted in terms of entropy changes. Within a particular simulation,  $\zeta$  agrees well—better than the total energy or the oxygen distribution—with common-sense chemical notions of hydrophobicity. Further work is being pursued to better assess the utility of  $\zeta$ .

In all the simulations it was implicitly assumed that the point charges used in the TIP3P and CHARMM parameterizations adequately represent the electrostatic effects in solution. For the essentially qualitative calculations done here, it is felt this assumption is justified. However, there is considerable evidence<sup>37</sup> that distributed multipoles<sup>38</sup> much better account for the detailed electrostatic interactions found in liquid water and would be useful for a more quantitative analysis—especially, one focussing on energetics. Furthermore, in common with nearly all calculations of solvent effects in aqueous solution, the model water molecules used here are not polarizable.

**B. Hydrophobic Crevices Are Less Accessible Than Hydrophilic Ones.** In the four simulations done here, there is a progression

in the extent that solvent molecules "move back" from the helix surface. On average liquid "argon" molecules pack tightly into narrow crevices while normal water molecules "move back" from hydrophobic surfaces. In between these extremes is normal water around hydrophilic surfaces. The implication is that, on average, water-water hydrogen bonding moves water molecules back from the protein surface while water-protein hydrogen bonding and, especially, packing considerations have the opposite effect.

The "moving back" of water molecules from hydrophobic surfaces provides an important qualification in the use of Richards-Connolly molecular surfaces.<sup>7</sup> The molecular surface is defined in terms of rolling a sphere the size of a water molecule on the protein surface. Connolly<sup>39</sup> implemented a definition of this surface, and, subsequently, molecular surface calculations have been used to indicate whether a certain atom or functional group is exposed to solvent. They have also proved valuable in measuring protein surface areas,<sup>40</sup> in characterizing the stability of proteins,<sup>41</sup> and in geometric approaches to ligand docking.<sup>42,43</sup>

However, the molecular surface treats water as if it were a Lennard-Jones liquid, such as argon. The results described here indicate how different "argon" is from normal water in solvent structure. That is, because it is unencumbered by the constraints of water-water hydrogen bonding, an "argon" molecule can fit much more deeply into a narrow hydrophobic crevice on the protein surface than a water molecule can. This greater accessibility of hydrophobic crevices to "argon" suggests that the Richards-Connolly molecular surface may overemphasize the corrugations in the protein surface around hydrophobic atoms and in doing so may overestimate the amount of protein accessible area in aqueous solution. Furthermore, the inaccessibility of hydrophobic crevices to water molecules suggests that to a water molecule—and possibly a water-like ligand—hydrophobic surfaces may appear "smoother" than hydrophilic ones.

**Acknowledgment.** We thank Ian McDonald, Cyrus Chothia, and Arthur Lesk for useful discussions, the Herchel-Smith foundation for a scholarship (M.G.), and the SERC and the Newton Trust for support in the purchase of equipment.

#### References and Notes

- (1) Gerstein, M.; Lynden-Bell, R. M. *J. Phys. Chem.*, preceding paper in this issue.
- (2) Creighton, T. E. *Proteins*; Freeman: San Francisco, 1984.
- (3) Dill, K. A. *Biochemistry* **1990**, *29*, 7133-7155.
- (4) Cybulski, S.; Scheiner, S. *J. Phys. Chem.* **1989**, *93*, 6565.

- (5) Lesk, A. M. *Protein Architecture: A Practical Guide*; IRL Press: Oxford, 1991.
- (6) Richards, F. M. *J. Mol. Biol.* **1974**, *82*, 1-14.
- (7) Richards, F. M. *Annu. Rev. Bioeng.* **1977**, *6*, 151-76.
- (8) Chothia, C. *Annu. Rev. Biochem.* **1984**, *53*, 537-72.
- (9) Janin, J.; Chothia, C. *J. Biol. Chem.* **1990**, *265*, 16027-30.
- (10) Weeks, J. D.; Chandler, D.; Andersen, H. C. *J. Chem. Phys.* **1971**, *54*, 5237-5247.
- (11) Chandler, D.; Andersen, H. C. *J. Chem. Phys.* **1972**, *57*, 1930-1937.
- (12) Chandler, D.; Hsu, C. S. *J. Chem. Phys.* **1977**, *66*, 5231-5234.
- (13) Chandler, D.; Weeks, J. D.; Andersen, H. C. *Science* **1983**, *220*, 787-794.
- (14) Hansen, J. P.; McDonald, I. R. *Theory of Simple Liquids*; Academic Press: Boston, 1986.
- (15) Franks, F. *Water: A Comprehensive Treatise*; Plenum Press: New York, 1973.
- (16) Franks, F. *Water*; The Royal Society of Chemistry: London, 1983.
- (17) Speedy, R. J.; Madura, J. D.; Jorgensen, W. L. *J. Phys. Chem.* **1987**, *91*, 909-913.
- (18) Sciortino, F.; Geiger, A.; Stanley, H. E. *Nature* **1991**, *354*, 218-221.
- (19) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926-935.
- (20) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187-217.
- (21) Chothia, C. *Nature* **1975**, *254*, 304-308.
- (22) Zamyatnin, A. A. *Annu. Rev. Biophys. Bioeng.* **1984**, *13*, 145-65.
- (23) Verlet, L. *Phys. Rev.* **1967**, *159*, 98-103.
- (24) Levitt, M.; Sharon, R. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 7557-7561.
- (25) Franks, F. *Faraday Symp. Chem. Soc.* **1982**, *17*, 7-10.
- (26) Hvidt, A. *Annu. Rev. Biophys. Bioeng.* **1983**, *12*, 1-20.
- (27) Gill, S. J.; Dec, S. F.; Olofsson, G.; Wadso, I. *J. Phys. Chem.* **1985**, *89*, 3758-3761.
- (28) Pangali, C.; Rao, M.; Berne, B. J. *J. Chem. Phys.* **1979**, *71*, 2975-2981.
- (29) Rapaport, D. C.; Scheraga, H. A. *J. Phys. Chem.* **1982**, *86*, 873-880.
- (30) Zichi, D. A.; Rossky, P. J. *J. Chem. Phys.* **1985**, *83*, 797-808.
- (31) Wallqvist, A.; Berne, B. J. *J. Phys. Chem.* **1985**, *145*, 26.
- (32) Lee, C. Y.; McCammon, J. A.; Rossky, P. J. *J. Chem. Phys.* **1984**, *80*, 4448-4455.
- (33) Valleau, J. P.; Gardner, A. A. *J. Chem. Phys.* **1987**, *86*, 4162-4170.
- (34) Wallqvist, A. *Chem. Phys. Lett.* **1990**, *165*, 437-442.
- (35) Ben-Naim, A. *Hydrophobic Interactions*; Plenum Press: New York, 1980.
- (36) Pratt, L. R.; Chandler, D. *J. Chem. Phys.* **1977**, *67*, 3683-3704.
- (37) Finney, J. L.; Quinn, J. E.; Baum, J. O. *Water Sci. Rev.* **1985**, *1*, 93-170.
- (38) Stone, A. J. *Mol. Phys.* **1981**, *43*, 233-248.
- (39) Connolly, M. *Science* **1983b**, *221*, 709-713.
- (40) Islam, S. A.; Weaver, D. L. *Proteins: Struct. Func. Genet.* **1991**, *10*, 300-314.
- (41) Miller, S.; Lesk, A. M.; Janin, J.; Chothia, C. *Nature* **1987**, *328*, 834-6.
- (42) Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. *J. Mol. Biol.* **1982**, *161*, 269-288.
- (43) Shoichet, B. K.; Kuntz, I. D. *J. Mol. Biol.* **1991**, *221*, 327-346.