

Simulation of Water around a Model Protein Helix. 1. Two-Dimensional Projections of Solvent Structure

Mark Gerstein^{†‡} and R. M. Lynden-Bell^{†‡}

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

The water structure around a model α -helix in solution is investigated by molecular simulation. A marked contrast is seen between the hydration of the α and β carbons on one side of the helix axis and of the carbonyl oxygen on the other side. Three well-defined peaks in the oxygen density, loosely corresponding to three layers of water, are found around the two hydrophobic atoms. In contrast, around the hydrophilic carbonyl oxygen the first two peaks merge into a single large peak. Analysis of the water orientation shows that on average the water dipoles are parallel to the surface of the helix. However, near the carbonyl oxygen they are slightly tipped inward and also have a significant component in the helical direction. The overall dipole orientation is consistent with the charges on the helix. Analysis of the hydrogen distribution reveals an inward pointing hydrogen near the carbonyl oxygen with one clearly preferred position for hydrogen bonding. In contrast, water molecules around the two hydrophobic atoms point their hydrogens outward and adopt the clathrate arrangement found in the hydration of small apolar solutes. Particular emphasis is put on developing a representation for the results in terms of two-dimensional projections, displayed in "gray-level". These projections exploit the symmetry of the helix to show the complex, three-dimensional structure of the water distribution better than one-dimensional profiles.

I. Introduction

The interaction of macromolecules, in particular proteins, with water has been studied both theoretically and experimentally. X-ray crystallography is the principal experimental technique for studying the structural aspects of water-protein interaction. It provides a detailed three-dimensional picture of the protein and some of the waters bound to it.^{1,2} Recently, NMR spectroscopy has been able to provide complementary information about the time scale of water-protein interactions, which are usually found to be in the subnanosecond range.³

The principal theoretical techniques used to investigate water-protein interactions are molecular dynamics and Monte Carlo simulations.⁴⁻¹³ In principle, these calculations can provide as detailed a picture of the water around a protein as crystallography provides of the protein itself. In reality they do not. They usually consist of taking a complex three-dimensional structure, running a computer for a long time on it, and presenting the results as a sequence of one-dimensional profiles, in which information from different positions is averaged. For instance, calculations have been done of self-diffusion constants and water densities as a function of distance from the protein. Most structural information in these one-dimensional profiles is averaged away, making detailed geometric comparisons between simulation and crystallography difficult.

A number of more structural representations have been used for the results of water simulations. These can be divided into three categories depending on whether they depict the instantaneous, vibrationally averaged, or diffusionally averaged structure—the I, V, and D structures.¹⁴ The I structure is shown by snapshots at various points in the simulation.^{10,12} These are often in stereo and show a richness of three-dimensional detail. Furthermore, the water molecules in these snapshots can be color-coded on the basis of some property such as their interaction energy with the protein.⁷ However, the instantaneous structure of a liquid, much more so than a solid, is far from representative of its behavior. This difficulty is overcome to some extent with

composite pictures showing superimposed snapshots at a regular intervals, e.g., every 50 ps.^{15,16}

The V structure removes the effect of the high-frequency hindered motions that can be distracting in viewing sequences of snapshots. It can be shown by averaging the trajectory over 0.2-ps intervals to give a sequence of vibrationally averaged structures.^{17,18} Alternatively, the whole trajectory can be shown in a "movie" with frequency filtering.^{19,20}

The D structure can be represented by a probability distribution for the positions and orientations of water molecules averaged over a whole simulation. This distribution provides much more representative information about a liquid than I- or V-structure snapshots. Moreover, it potentially contains as much three-dimensional structural information as the snapshots do. However, because this information is in the form of slowly varying probability density, it is difficult to represent comprehensibly. A number of techniques have been employed to overcome this difficulty. In the statistical state solvation site analysis technique, a single three-dimensional contour encloses regions where the probability distribution is above a certain threshold.^{21,22} Two-dimensional contour maps of solvent density have been used to depict the ordered water around lysozyme²³ and an alanine dipeptide (*N*-acetylalanyl-*N*-methylamide).²⁴ However, in comparison to one-dimensional profiles, averaging is done over fewer variables in calculating two- and, especially, three-dimensional contours, so very long simulations are required to obtain reasonable statistics. This is particularly true for the water found around hydrophobic groups.

Here an attempt is made to circumvent the difficulties of visualizing the results and of obtaining a reasonable statistics by using a model system with high symmetry—a polyaniline α -helix in a bath of 321 water molecules (Figure 1a). Because of the symmetry useful results can be obtained from two-dimensional projections after calculations of reasonable length. Although this approach cannot be applied to real proteins, which have little symmetry, model systems can be used to gain insight into their behavior.

Chemically, the symmetry of a polyaniline α -helix manifests itself in an interesting double-helical arrangement of hydrophobic

[†] MRC Laboratory of Molecular Biology.

[‡] University Chemical Laboratory.

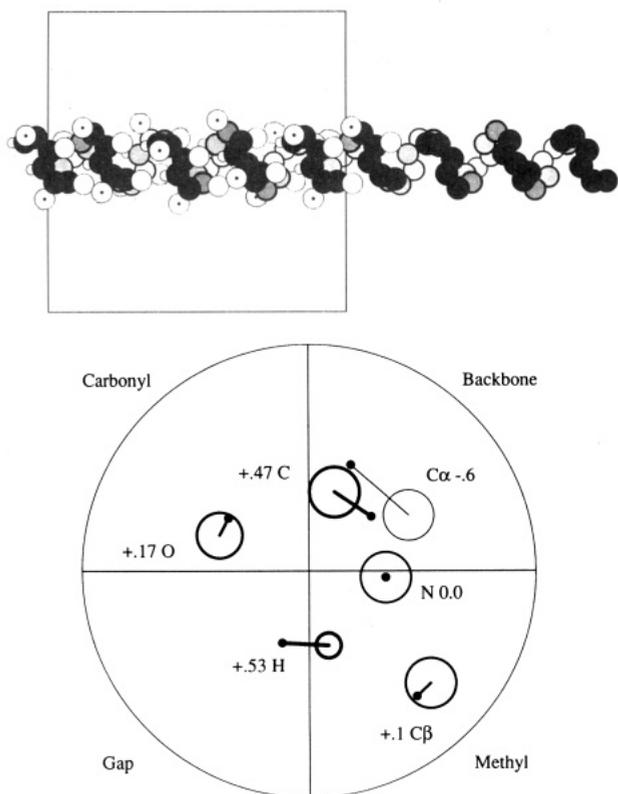


Figure 1. (a, top) Model system of a single polyaniline helix embedded in a $22.219 \times 22.219 \times 20.93$ Å cell. The boundaries of the cell are shown. The helix is oriented along z axis, and the view is down x axis. Backbone atoms (C, N, $C\alpha$) are shown in shades of gray with a thick line. The hydrogen bonding carbonyl oxygen (O) and amide hydrogen (H) are indicated by large and small unfilled white circles. The β -carbon (methyl group) is shown by a large white circle filled with a black dot. Four turns (14 residues) of the helix perfectly fit the cell, but for clarity additional turns are shown. To highlight the twist of the helix, only the backbone atoms are shown for the last few turns. This figure was drawn with MOLSCRIPT.⁶³ (b, bottom) Effect of the helical projection. The figure shows the six atoms forming an alanine residue that are closest to the $z = 0$ plane. The view is down the z axis, the numbers next to the atom names are the z coordinates of the atoms. The small black dots indicate the positions of these atoms in a straight projection. The white circles indicate the positions of these atoms after they have been helically projected (as described in the text). The displacement of the white circles from the black dots indicates the distortion produced by the helical projection. The radius of the large enclosing circle is 4.5 Å, and the division of the helical projection into four quadrants is also indicated.

and hydrophilic groups. That is, the backbone atoms (N, $C\alpha$, C) and the methyl group of the β -carbon form a helical band of hydrophobic atoms. On the opposite side of the helix axis from the β -carbon, there is a gap, and the carbonyl oxygen and amide hydrogen form a hydrogen bond across it. The ladderlike arrangement of the two hydrogen bonding atoms, in turn, forms a hydrophilic band that twists the helix axis in complementary fashion to the hydrophobic band.

To investigate this interesting juxtaposition of hydrophilic and hydrophobic regions, Monte Carlo simulation was used to generate many configurations of water molecules. The analysis of these configurations focused on the average distribution of oxygen and hydrogen atoms and the average orientation of the water molecules. Two ways of projecting this information down to two dimensions were considered. The first is the obvious straight projection along the axis of the helix. The second involves projecting along a helical path that follows the polypeptide backbone. In this helical projection all molecules near the hydrophilic band (i.e., around the carbonyl oxygen) are superimposed and separated from molecules near the hydrophobic band (i.e., around the $C\beta$ atom), so that properties of water molecules near different groups can be easily disentangled.

II. Method

A. Details of the Simulation. The z axis was chosen along the axis of the helix, and cylindrical coordinates R, ϕ, z were used to describe the position of a water molecule. Periodic boundary conditions with a $22.219 \times 22.219 \times 20.93$ Å³ cell were used. The helix was built with "ideal" torsion angle values²⁵ $\phi = 57^\circ$ and $\psi = -41^\circ$, which differ slightly from the average torsion angle values found in helices of refined protein structures.^{26,27} The periodicity of the cell in the z direction closely matched that of 14 residues of α -helix, so the system essentially consisted of an infinite helix.

As summarized in Table I, the water molecules interacted with each other by the TIP3P potential,²⁸ which has partial charges on each atomic site and a Lennard-Jones interaction between oxygen atoms. The protein-water interaction was modeled using the standard CHARMM parameters.²⁹ The electrostatic interaction was cut off at 7.5 Å as is appropriate for the TIP3P model. As the β -carbon is uncharged, the cutoff implied that water molecules further than 9.8 Å from the helix axis will have no electrostatic interaction with the helix.

The initial configuration was prepared by embedding the helix in a box of 375 water molecules already equilibrated at 300 K (a density of 27.55 Å³/water). Water molecules making bad contacts with the helix were removed by standard procedures,²⁹ leaving 321 water molecules in the box. The separation between the outside of one helix and that of its image in the neighboring box is ~ 16 Å, allowing about seven intervening layers of water molecules, only four of which are strongly influenced by the helix. Long runs were needed to obtain good statistics, so a Monte Carlo program^{30,31} was written to exploit the parallel nature of the computer used, an Alliant FX-2800. It incorporated a preferential sampling scheme,^{32,33} in which water molecules within 7 Å of the helix axis were moved twice as often as those outside. To increase the sampling efficiency, the calculation was divided into about fifty independent parts. Each part was initiated by applying a thermal shock to randomize the configuration. The temperature was set to 1000 K and reduced to 300 K over 2000 Monte Carlo cycles. The system was then equilibrated for 10^4 cycles after which averages were accumulated every five cycles for 10^4 cycles. Consequently, each heat-cool-run part is comparable to roughly 44 ps of molecular dynamics with a 2-fs time step. (One Monte Carlo cycle, which consists of attempting moves on all the waters, is considered to be equivalent to a single molecular-dynamics step.) Up to 20 heat-cool-run parts could be executed concurrently. The entire calculation required a total of 3.24×10^5 cycles for cooling and equilibration and 5.52×10^5 cycles for accumulating averages. The total number of cycles is comparable to roughly 2 ns of molecular dynamics.

A number of approximations were made to expedite the calculation and to simplify the analysis. Most importantly, the internal coordinates of the helix were kept rigid and, as evident in Table I, the nonpolar hydrogens were subsumed into neighboring carbons with a united-atom representation. The absence of large-scale motions (due to the rigidity of the helix) is not expected to affect the results, since these motions would take much longer than one 44-ps heat-cool-run calculation.^{13,34} However, water hydrogen bonding to α -helices has often been associated with local distortions from helical geometry that occur on a shorter time scale.^{11,13,35,36}

To test some of these simplifications, an 88-ps molecular-dynamics simulation was done with a partially mobile helix and an all-atom parameter set that explicitly included the protons attached to the α and β carbons. The CHARMM package²⁹ was used. The overall conformation of the helix was fixed by the periodic boundary conditions, but local motions, particularly involving the carbonyl oxygen, were possible. A much larger cell, $46.23 \times 31.1 \times 20.92$ Å³, was used with 909 TIP3P waters. Because of the rapid vibrations of protons attached to nonpolar

TABLE I: Nonbonded Parameters Used in the Simulation^a

atom	ϵ (kJ/mol)	σ (Å)	charge (electrons)
carbonyl carbon	0.5023	3.7418	0.550
α -carbon (incorporating 1 hydrogen)	0.2034	4.2140	0.100
β -carbon (incorporating 3 hydrogens)	0.7581	3.8576	0.000
amide nitrogen	0.9979	2.8509	-0.350
amide hydrogen	0.2085	1.4254	0.250
carbonyl oxygen	0.6660	2.8509	-0.550
water oxygen (in interactions with the helix)	0.6660	2.8509	-0.834
water hydrogen (in interactions with the helix)	0.2085	1.4254	0.417
water oxygen (in interactions with other waters)	0.6367	3.1506	-0.834
water hydrogen (in interactions with other waters)	0.0000	0.0000	0.417

^a ϵ and σ are the two Lennard-Jones parameters and the charge is for standard Coulomb electrostatics. The water-water parameters are from the original TIP3P potential,²⁸ the water-helix interaction is taken directly from recent parameter sets in version 21.3 of CHARMM²⁹ and version 2.1 of X-PLOR.⁶² A number of other parameter sets were tried—in particular those for TIP4P water²⁸—but they were not found to affect the results significantly.

atoms, a short time step (0.5 fs) was used. Such a calculation is considerably more expensive than the Monte Carlo simulations used for the results. No significant differences in the distribution and orientation of water molecules were seen in comparison to the Monte Carlo simulations, indicating that neither the box size, united atom representation, nor rigidity of the helix affected the results meaningfully. In addition, a number of short Monte Carlo simulations were done with longer cutoffs for the nonbonded interactions (up to 9.5 Å) and with the TIP4P potential.²⁸ Again no significant differences were found apart from the obvious lengthening of the calculation.

B. Distribution Averaging. The quantities of interest are the average positions and orientations of the water molecules relative to the helix. All of this information can be extracted from the full molecular distribution function $g(\mathbf{r}\omega)$, which describes the probability of finding a water molecule with position \mathbf{r} and orientation ω relative to the helix. It is similar to the two-particle distribution function used in the theory of liquids³⁷ but describes a water molecule relative to the fixed helix rather than to another mobile water molecule.

The full distribution function contains complete information about the diffusively averaged water structure around the helix. However, it is a function of six variables, so it is difficult to represent comprehensively. The tactic taken here is to average the distribution carefully over four of its six variables, in both straight and helical fashion, and then to show the oxygen and hydrogen distribution and components of the average dipolar and quadrupolar ordering in two-dimensional projections.

The first variable averaged over is the angle describing the orientation of a water molecule relative to its dipole moment. Consequently, the orientation of a water molecule will only be described in terms of a unit vector in the direction of its dipole moment. The resulting five-variable distribution $g(\mathbf{r}\alpha\beta)$ can, in turn, be expanded in spherical harmonics:

$$g(\mathbf{r}\alpha\beta) = (1/2\pi\sigma) \int_0^{2\pi} g(\mathbf{r}\omega) d\gamma = \sum_{lm} g(lm;\mathbf{r}) Y_{lm}(\alpha\beta) \quad (1)$$

where the factor $\sigma = 2$ corrects for the indistinguishability of water hydrogens and the two spherical polar angles α and β are chosen in a coordinate system with the principal axis along the radius vector, \mathbf{R} , from the helix axis to a water oxygen.

Averaging over the two orientational variables α and β in an unweighted fashion, here represented by $\langle \rangle_{\alpha\beta}$, gives the first term in the expansion, $g(00;\mathbf{r})$, which except for a constant factor is the distribution of the centers of mass, $g(\mathbf{r})$:

$$g(\mathbf{r}) = (4\pi)^{1/2} g(00;\mathbf{r}) = \langle g(\mathbf{r}\alpha\beta) \rangle_{\alpha\beta} = (1/4\pi) \int g(\mathbf{r}\alpha\beta) \sin \beta d\alpha d\beta \quad (2)$$

The centers of mass distribution is almost the same as oxygen distribution, and both will be used interchangeably here. The hydrogen distribution is also just a projection of the full distribution

function:

$$g_H(\mathbf{r}) = \langle g(\mathbf{r} - \mathbf{d}_{\text{OH1}}, \alpha\beta) \rangle_{\alpha\beta} + \langle g(\mathbf{r} - \mathbf{d}_{\text{OH2}}, \alpha\beta) \rangle_{\alpha\beta} \quad (3)$$

where \mathbf{d}_{OH1} and \mathbf{d}_{OH2} are vectors from the oxygen to each of the hydrogens.

In the spherical polar coordinate system chosen, the components of the dipole orientation are given by

$$\mu_{\text{rad}} = \hat{\mu} \cdot \hat{\mathbf{R}} = \cos \beta$$

$$\mu_{\text{tan}} = \hat{\mu} \cdot (\hat{\mathbf{R}} \times \hat{\mathbf{z}}) = \cos \alpha \sin \beta$$

$$\mu_{\text{vert}} = \hat{\mu} \cdot \hat{\mathbf{z}} = \sin \alpha \sin \beta \quad (4)$$

where \mathbf{R} and \mathbf{z} are unit vectors in the axes directions and $\hat{\mu}(\alpha\beta)$ is a unit vector in the direction of the dipole moment—i.e., from the oxygen to the midpoint of the hydrogens. With this dipole definition a water molecule whose protons point away from the helix has a positive dipole. Furthermore, the vertical and tangential components of the dipole are defined so that they increase in the same direction as the twist of the helix.

To describe the average dipole moment at a particular point, it is necessary to perform an average over the two orientational angles weighted by the distribution of water orientations. Furthermore, it is advantageous to put this weighted average on a per molecule basis by dividing by the oxygen distribution. The whole operation is denoted by $\langle \rangle_{\omega}$ and for the radial dipole component $\mu_{\text{rad}}(\mathbf{r}\alpha\beta)$ it would be expressed by

$$\langle \mu_{\text{rad}}(\mathbf{r}\alpha\beta) \rangle_{\omega} = \langle g(\mathbf{r}\alpha\beta) \mu_{\text{rad}}(\mathbf{r}\alpha\beta) \rangle_{\alpha\beta} / \langle g(\mathbf{r}\alpha\beta) \rangle_{\alpha\beta} \quad (5)$$

This particular average ranges from -1 to 1 and represents the average radial component of a unit vector in the direction of the water dipole at \mathbf{r} . The average tangential and vertical components of the dipolar ordering $\langle \mu_{\text{tan}} \rangle_{\omega}$ and $\langle \mu_{\text{vert}} \rangle_{\omega}$, and average radial component of the quadrupolar ordering, $\langle P_2(\mu_{\text{rad}}) \rangle_{\omega}$, are similarly obtained. All the dipolar and quadrupolar ordering terms can be directly related back to the full molecular distribution function:

$$g(10;\mathbf{r}) = g(00;\mathbf{r})^{3/2} \langle \mu_{\text{rad}} \rangle_{\omega}$$

$$g(20;\mathbf{r}) = g(00;\mathbf{r}) (5^{1/2}/2) \langle P_2(\mu_{\text{tan}}) \rangle_{\omega}$$

$$g(1, \pm 1; \mathbf{r}) = -g(00;\mathbf{r}) (5^{1/2}/2) (\langle \mu_{\text{tan}} \rangle_{\omega} \pm i \langle \mu_{\text{vert}} \rangle_{\omega}) \quad (6)$$

Note the difference between $g(10;\mathbf{r})$ and $\langle \mu_{\text{rad}} \rangle_{\omega}$. The first describes the polarization of the medium in the radial direction at point \mathbf{r} , and the second, the average dipolar ordering of a molecule situated at the same point. If there are few molecules at a point, this ordering can be appreciable although the total polarization is negligible.

Thus far, six quantities have been presented that depend on the three positional variables, $\mathbf{r} = (R, \phi, z)$: the number of oxygens and hydrogens, the three components of the average dipolar ordering, and the principal component of the average quadrupolar ordering. To complete the picture, all that remains is to average down to two variables. The straightforward way to accomplish this is to project parallel to the z axis. For instance, for oxygen distribution

$$\langle g(\mathbf{r}) \rangle_z = (1/c) \int g(\mathbf{r}) dz \quad (7)$$

where c is the dimension of the periodic cell in the z direction and $\langle \rangle_z$ denotes the straight projection.

C. Helical Projection. However, it is also possible to project helically with the twist of the polypeptide chain:

$$\langle g(\mathbf{r}) \rangle_{\text{HLX}} = \langle g(R, \phi, z) \rangle_{\text{HLX}} = \langle g(R, \phi + 2\pi z/p, z) \rangle_z \quad (8)$$

where $p = 5.4 \text{ \AA}$ is the rise of an α -helix per turn and $\langle \rangle_{\text{HLX}}$ denotes the helical projection.

As will be evident, the helical projection provides more detail than the straight projection. However, it is not as straightforward to interpret. Applying the helical projection to a continuous helix (e.g., a spring made from metal tubing) would be identical to looking at its cross section. However, an α -helix is not continuous: atoms appear at discrete intervals in z . Consequently, as shown in Figure 1b, for an α -helix the positions of atoms in helical projection are only approximately the same as in cross section. The figure shows the six atoms in an alanine residue with a helical conformation closest to the $z = 0$ plane. These six atoms are projected onto the $z = 0$ plane in straight and helical fashion (i.e., replacing the ϕ coordinate of an atom by $\phi + 2\pi z/p$). In the radial (R) direction both projections are identical. However, in the axial (ϕ) direction the helical projection distorts the distances between atoms. This distortion is the main drawback of the projection. It causes a residue to occupy more than 100° axially as is usual in helix with 3.6 residues/turn.³⁸ In terms of the molecular distribution, water molecules above or below the carbonyl group are rotated relative to the position of the oxygen atom. Note, the axial distortion introduced by the helical projection does not affect the dipole orientation at all since the two dipole orientation angles α and β are defined without recourse to ϕ . The straight projection can be approximately regenerated from the helical projection by overlaying seven copies of the helical projection, each rotated by 51° with respect to the previous.

Despite the complexities introduced by the helical projection, it is the best way to take into account the symmetry of the system. In this projection the atoms of all the residues coincide, and as shown in Figure 1b, the methyl group (β -carbon) is in the bottom right; the backbone atoms (C, C α , N) and the hydrogen, top right; and the carbonyl oxygen, top left. This division can be made more precise by dividing the projection into four quadrants, which will be denoted methyl, backbone, carbonyl, and gap. In a normal α -helix, the hydrogen bonds between the carbonyl oxygen and amide proton occur all around the helix axis, but in the helical projection they are superimposed and span the gap quadrant. As there are no atoms in the gap quadrant, the helical projection provides an exceptionally clear picture of the hydration of the C=O...H—N hydrogen bond. That is, all features of the water distribution in the gap quadrant are solely related to the hydration of this hydrogen bond.

As the gap quadrant just contains a hydrogen bond, it is considered hydrophilic. It and the carbonyl quadrant together constitute the hydrophilic half of the projection. The backbone and methyl quadrants form the hydrophobic half.

III. Results

A. Format of the Results. It is worthwhile to consider the organization of the results. Six quantities derived from the full

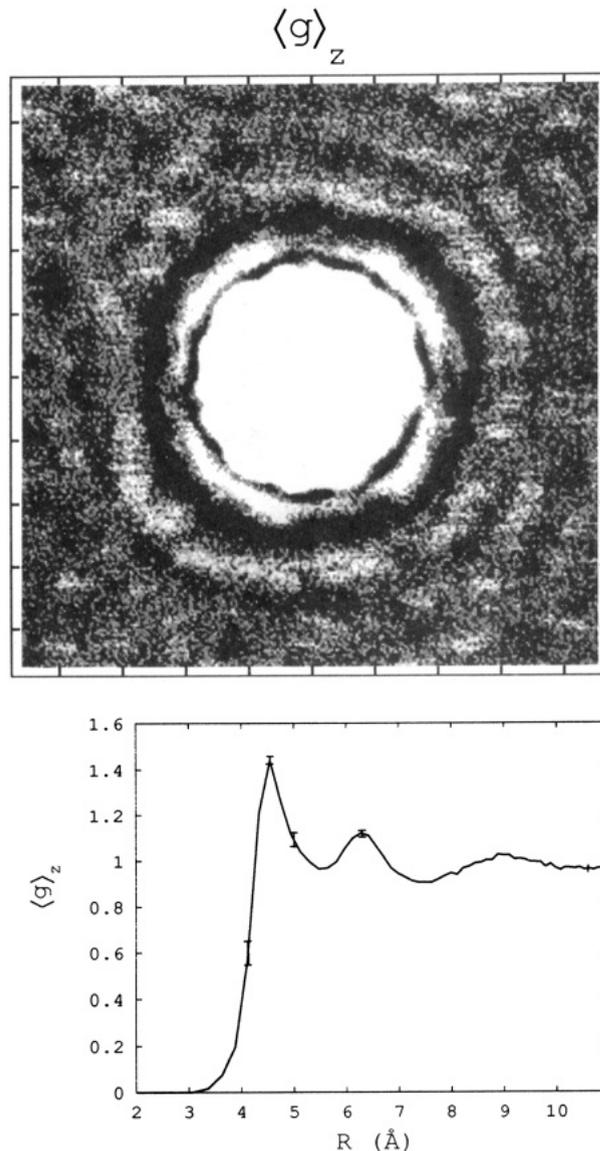


Figure 2. Straight projection of oxygen distribution, $\langle g(\mathbf{r}) \rangle_z$, for water around a polyalanine helix. The helix is centered at $R = 0$ and has approximate 7-fold symmetry in ϕ . The top half of figure shows the distribution in "gray-level", where black is high. Tick marks are spaced 2.5 \AA apart. The innermost ring of density (at $R = 4.6 \text{ \AA}$) has a maxima of ~ 1.7 and a minima of ~ 0.4 . The bottom half of the figure shows the average of the two-dimensional projection around the helix as a function of the distance from axis (R). Error bars were estimated by deviations from 7-fold symmetry.

distribution function are of interest. These six quantities are, in turn, projected in both straight and helical fashion to give 12 functions of R and ϕ . Throughout the simulation, each of these 12 projections was accumulated on a 232×232 grid, which has squares $\sim 0.1 \text{ \AA}$ on a side. Accumulated grids were stored in standard crystallographic formats³⁹ and analyzed with an image-display package written by J. M. Smith. False-color tables⁴⁰ were extremely useful in their interpretation. However, here they are presented in "gray-level", where black is high. The "gray-level" representation gives one a good view of the intricate spatial variation in the projections. However, it is deficient in representing their magnitudes quantitatively. To overcome this difficulty, the projections were averaged over ϕ to give a number of radial profiles. For the helical projections, the averaging was done separately for each quadrant.

B. Oxygen Distribution. 1. Distribution Description. The oxygen distribution is most central to the analysis. It is shown in straight and helical projection in Figures 2 and 3. In the straight projection, there are three peaks at 4.6 , 6.3 , and 8.9 \AA from the

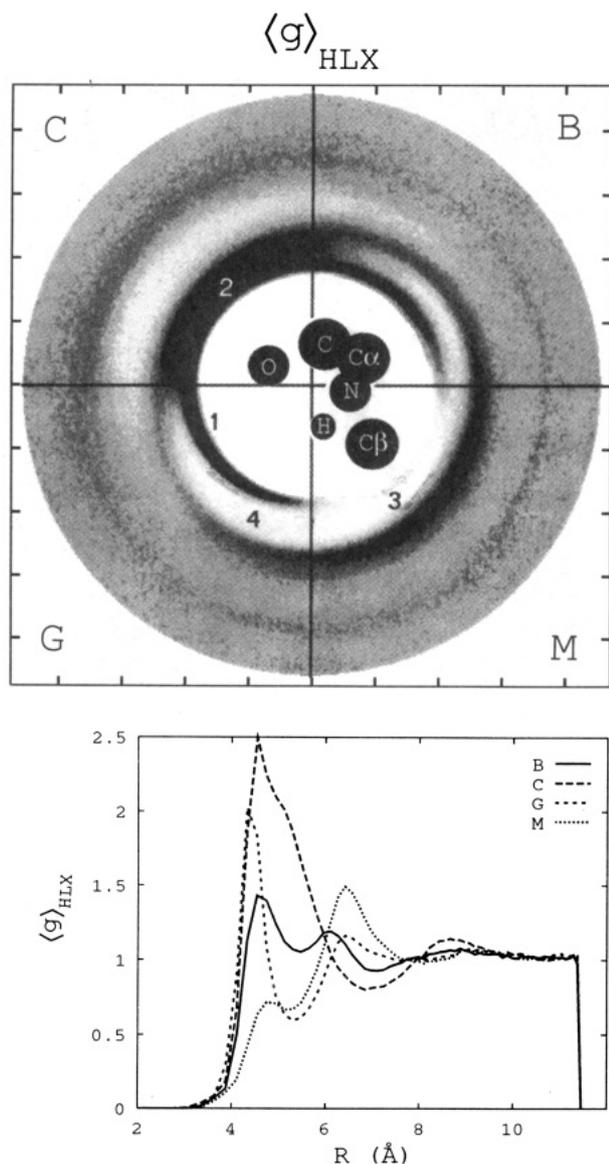


Figure 3. Helical projection of oxygen distribution, $\langle g(r) \rangle_{HLX}$, for water around the polyalanine helix. (a, left) The top half of the figure shows the distribution in "gray-level", where black is high. Tick marks are spaced 2.5 Å apart. Positions of the helically projected helix atoms 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. Beneath the "gray-level" display are averages of the two-dimensional projection in each of the four quadrants as a function of the distance from the helix axis (R). These quadrants are described in the text and indicated here by B (backbone), C (carbonyl), G (gap), and M (methyl). (b, right) Three-dimensional plot of the helical projection of oxygen distribution, $\langle g(r) \rangle_{HLX}$, as a function of R and ϕ . The maximum in the distribution occurs near the carbonyl oxygen and is indicated by 1 in the previous part of the figure. This view of the oxygen distribution shows how the two inner layers of water around the helix merge into one large peak near the carbonyl oxygen.

helix axis. The three peaks can be associated with three shells of water. The second and third of these shells are essentially continuous in that there is little variation in the oxygen distribution around the helix. In contrast, some axial structure can be seen in the innermost shell. The source of the structure is revealed in the helical projection, which shows marked differences between the water around the carbonyl oxygen and around the more hydrophobic atoms. In the backbone and methyl quadrants, there are three well-defined peaks in the oxygen distribution. In contrast, in the carbonyl quadrant the inner peaks merge into a single large peak, leaving only two peaks in the distribution. This merging can be seen in plots of the distribution in each quadrant and is particularly clear on the three-dimensional surface plot shown in Figure 3b.

The inner peak is more than twice as high around the carbonyl oxygen as around the β -carbon (maximum 3.5 versus 1.4). The larger carbonyl peak implies that oxygen centers are more likely to be clustered around the hydrophilic carbonyl oxygen than the hydrophobic β -carbon. It does not mean that the water around

the carbonyl is necessarily "denser" in the sense of being compressed into a smaller volume. Levitt and Sharon⁷ well explained this distinction. They calculated volumes of water molecules around a protein using a Voronoi-polyhedra procedure⁴¹ and found that the average Voronoi-derived density of a water molecule was approximately constant (within 5%) as a function of distance from the protein. In contrast, the probability density derived from counting with cubic grid boxes had a similar variation to that observed here.

2. Comparison with Experiment. The oxygen distribution is what is most accessible experimentally, and the results reported here correspond to what has been observed in crystal structures.

A recent neutron diffraction study of the hydration of carbonmonoxymyoglobin has determined the scattering density of disordered solvent around a protein.⁴² The 1.7-Å separation between the first two peaks in the oxygen distribution found in the study matches the results presented here. The peak heights, however, are not in the same proportion. The discrepancy probably reflects the fact that the crystallographic work averaged

over a very different protein surface containing charged residues. The carbon monoxymyoglobin study is much better suited for comparison with the results presented here than the many previous analyses of protein hydration that focus on the ordered (bound) water around the protein.⁴³

Thanki et al.⁴⁴⁻⁴⁶ superposed the water structure around each functional group in 17 very high resolution crystal structures to get experimental oxygen distributions. They found that water molecules are usually located 2.7 Å from the carbonyl oxygen. Here the carbonyl oxygen is 1.8 Å from the helix axis, so the carbonyl quadrant peak at 4.5 Å agrees with their results. They also reported that, in accordance with sp^2 bonding geometry, water oxygens are arranged on the circular edge of a 120° cone coaxial with C=O bond. It is clear from inspection of Figure 3 (and correction for axial distortions) that waters are spread $\sim 120^\circ$ around the carbonyl oxygen. However, as the results here are two-dimensional projections, it was not possible to verify the low probability for finding a water oxygen near the center of the cone. The higher density seen here on one side of the carbonyl oxygen is not present in their results and perhaps can be attributed to constraints imposed by the α -helical geometry.

Also in accord with the results presented here, Thanki et al. reported no preferred direction for water molecules around the alanine β -carbon. However, they found that a water oxygen is usually located 4 Å away from the β -carbon while here there is only 3.4 Å between the second peak in the methyl quadrant and the center of the β -carbon. This discrepancy is not that serious as apolar hydration is not nearly as well resolved as polar hydration in crystal structures, and Thanki et al. indicated that it was the least conclusive part of their analysis.

C. Water Orientation. All quantities pertaining to orientation are presented here as averages per molecule. This results in statistical problems very close to the helix, where averaging is done over few molecules. To avoid these problems, the orientation results are only shown from 4.0 Å outward.

The quadrupolar ordering is dominated by its radial component, $\langle P_2(\mu_{\text{rad}}) \rangle_\omega$, which is shown in helical projection in Figure 4. The quadrupolar ordering does not vanish until ~ 7 Å, so orientational order extends into the second shell relatively far from the helix. There is not much variation between quadrants. The negative values of the quadrupolar ordering all around the helix indicate that the water dipoles are preferentially oriented perpendicular to the radius vector. This perpendicular orientation is in agreement with the results of other protein-water simulations.⁶ An exception to the perpendicular orientation occurs at the boundary between the carbonyl and gap quadrants, where there is a small positive peak (indicated by a 1 in the figure) from dipoles strongly oriented by the carbonyl oxygen.

Superposed onto the general perpendicular orientation, the dipoles tend to be tilted slightly inward or outward depending on what chemical group they are near. As suggested by Table I, there is a strong electric field across the helix from the positively charged carbonyl carbon and α -carbon to the negatively charged carbonyl oxygen. The radial component of the average dipolar ordering, $\langle \mu_{\text{rad}} \rangle_\omega$, is arranged in opposition to this electric field (Figure 5). That is, there are inward pointing dipoles in the carbonyl and gap quadrants and outward pointing dipoles in the backbone quadrant. The inward orientation is greater than the outward orientation by a factor of ~ 3 . The dipolar ordering is mostly confined to the first shell, extending only to 5 Å. From there to 6.5 Å there is a slight tendency for all the dipoles to point away from the helix, and after that the dipolar ordering vanishes. Comparing the oxygen distribution with the dipolar and quadrupolar ordering (i.e., Figures 3-4) highlights the poorer statistics for the higher-order terms in the harmonic expansion.

In the helical projection the dipolar ordering has vertical and tangential components, which are smaller than the radial component but near the helix are still above the noise level. The

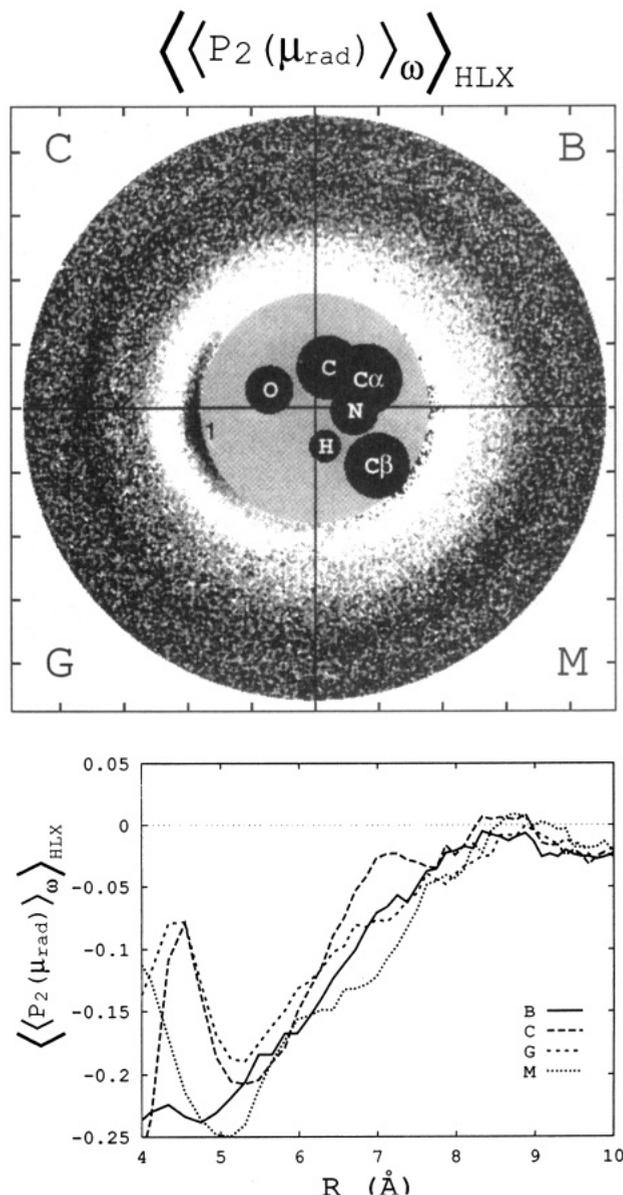


Figure 4. Helical projection of average quadrupolar ordering, $\langle \langle P_2(\mu_{\text{rad}}) \rangle_\omega \rangle_{\text{HLX}}$. The same conventions as in Figure 3a are used. However, because of the numerical uncertainties in averaging over the few water molecules very close to the helix, 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. Maximum 1 near the carbonyl oxygen is -0.05 .

maximum magnitudes of these three components (radial, 0.55, tangential, 0.3, and vertical 0.08) roughly indicate that the radial component is double the tangential component, which, in turn, is 4 times larger than the vertical component. The tangential and vertical components have a similar spatial variation, and only the tangential component is shown (Figure 6). For first-shell water molecules it has a positive peak in the carbonyl quadrant and a negative peak of equal magnitude in the gap quadrant. It has no other appreciable structure. Thus, the nonradial component of the dipolar ordering (i.e., the sum of tangential and vertical components) is oriented along the direction of the helix. Dipoles of water molecules immediately above the carbonyl oxygen point back down along the chain towards it, and those below, point up the chain in the opposite direction.

Many of the features of the water orientation disappear in the straight projection. Because of the axial symmetry in the straight projection, one might expect the tangential component of dipolar ordering, $\langle \langle \mu_{\text{tan}} \rangle_\omega \rangle_z$, to vanish. In practice it is approximately zero and was used to judge the statistics of the simulations. As shown in Figure 7, the radial component, $\langle \langle \mu_{\text{rad}} \rangle_\omega \rangle_z$, does not

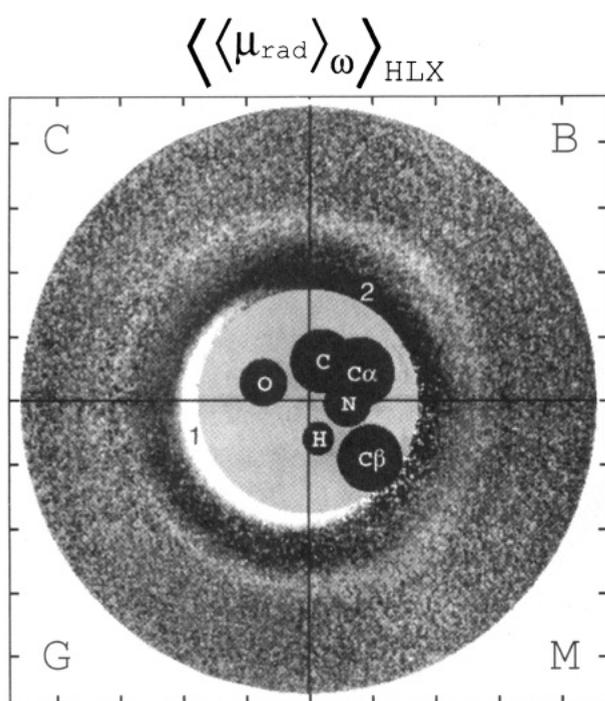


Figure 5. Helical projection of average radial component of the dipolar ordering, $\langle\langle\mu_{\text{rad}}\rangle\rangle_{\omega}$. The same conventions as in Figure 4 are used. Minimum 1 near the carbonyl oxygen is -0.55 , and maximum 2 near the α -carbon is 0.2 .

vanish. Highlighting the symmetry of the system, it has seven sharp peaks around the helix from the inward pointing dipoles near the carbonyl oxygen.

D. Hydrogen Distribution. Additional information about water-molecule orientation is obtained by considering the hydrogen distribution. Shown in Figure 8, the helical projection of the hydrogen distribution is markedly different on the hydrophobic and hydrophilic sides of the helix. On the hydrophilic side (carbonyl and gap quadrants) water molecules point a hydrogen in toward the carbonyl oxygen to make a hydrogen bond. This inward pointing hydrogen, an oxygen, and then a second hydrogen sandwiched between the first two water layers are evident in the hydrogen distribution peaks in the carbonyl quadrant at 3.6 and 5.1 \AA , which bracket the oxygen distribution peak at 4.3 \AA . In contrast, on the hydrophobic side of the helix (methyl and backbone quadrants) water molecules tend to have both hydrogens pointing away from the helix, and the first peak in the hydrogen distribution (at a distance of 5 or greater) is further from the helix than the first peak in the oxygen distribution.

The position marked by a 1 in Figure 8 is the overall maximum in the hydrogen distribution. It occurs in the first shell near the boundary between the carbonyl and gap quadrants. Relative to

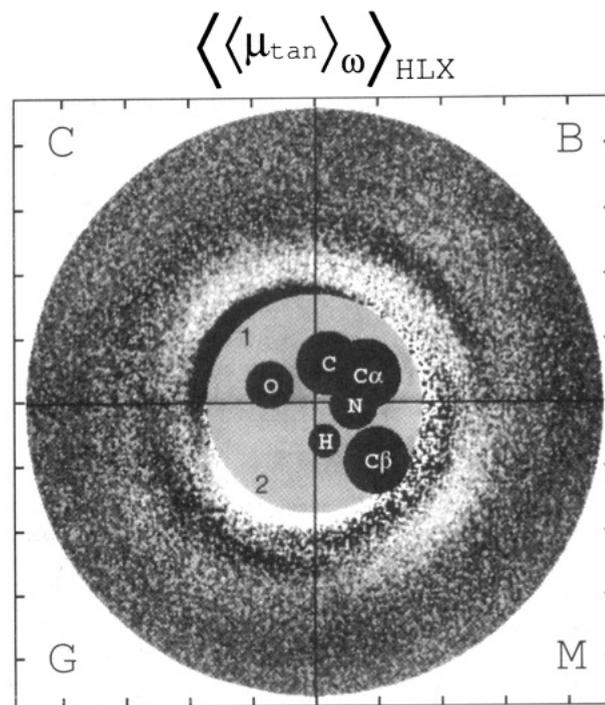


Figure 6. Helical projection of average tangential component of the dipolar ordering, $\langle\langle\mu_{\text{tan}}\rangle\rangle_{\omega}$. The same conventions as in Figure 4 are used. Maximum 1 near the carbonyl oxygen is 0.3 , and minimum 2 in the gap quadrant is -0.15 .

the carbonyl-amide hydrogen bond spanning the gap quadrant, it is much closer to the carbonyl oxygen. After correcting for the distortion of the helical projection, a line connecting this position with the center of the carbonyl oxygen reflects the standard distances and angles for protons hydrogen bonding to sp^2 hybridized oxygens.²⁶ If an additional correction is made for the inward displacement of the water hydrogen relative to the oxygen, this position is approximately where there is a peak in the oxygen distribution (marked by a 1 in Figure 3) and in the radial components of the dipolar and quadrupolar ordering (marked by a 1 in Figures 4 and 5). It is also where the vertical and tangential components of the dipolar ordering change sign (Figure 6); thus, by all indications it is where water molecules interact most strongly with the helix.

IV. Conclusion

In summary, a simple polyaniline α -helix presents a varied chemical environment to the solvent. It has bands of hydrophobic and hydrophilic atoms arranged in a double helix. That is, at any position along the helix axis, there is a carbonyl oxygen on one side and methyl group (β -carbon) on the opposite side. The average water structure reflects this contraposition.

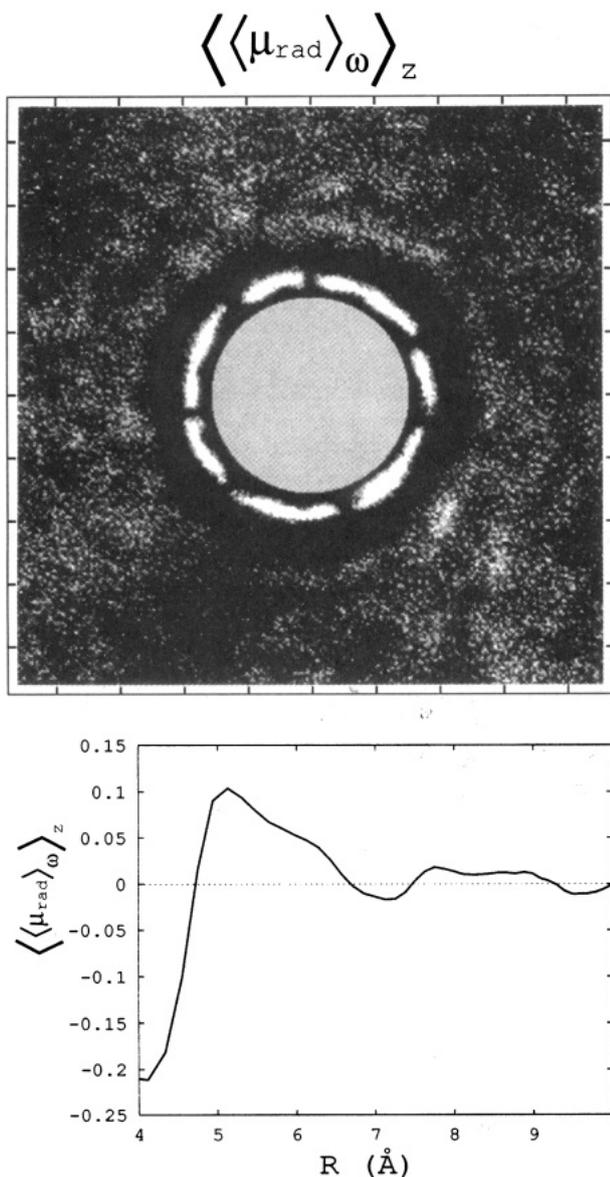


Figure 7. Straight projection of average radial component of the dipolar ordering, $\langle\langle\mu_{\text{rad}}\rangle_{\omega}\rangle_z$. The same conventions as in Figure 2 are used. However, because of the numerical uncertainties in averaging over the few water molecules very close to the helix, 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. The seven symmetrical minima are approximately -0.5 .

Around most of the helix water dipoles are orientated perpendicular to the radius vector from the helix axis. However, near the carbonyl oxygen, they point inward toward the negative charge and also have small components parallel or antiparallel to the helically twisting polypeptide chain. Consistent with this dipolar ordering, water molecules point a proton inward to hydrogen bond with the carbonyl oxygen. The sharp peak in the hydrogen distribution indicative of this hydrogen bonding coincides with notable features in the dipolar and quadrupolar ordering and locates the position where the influence of the helix on the water structure is strongest.

Around the β -carbon the water distribution has a different character. There are three water layers as opposed to the two layers around the carbonyl oxygen. The oxygen-distribution peaks corresponding to these layers are small, and the chance that a first-shell water molecule is around the β -carbon is only half compared to it being around the carbonyl. In the backbone and methyl quadrant, and they face in the opposite direction with outward pointing dipoles and hydrogens.

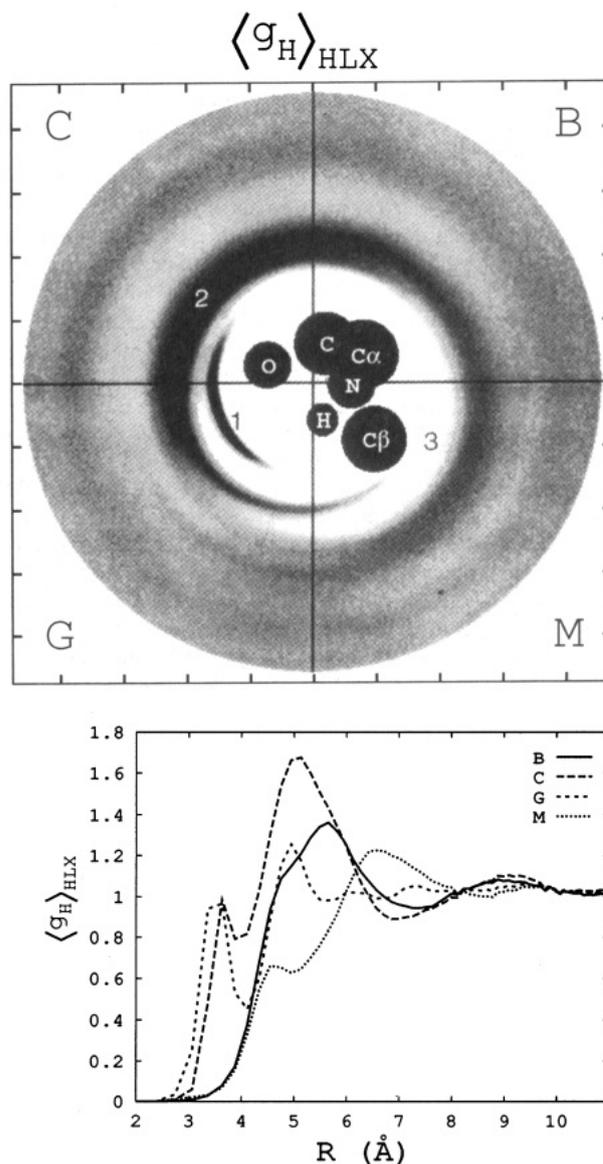


Figure 8. Helical projection of hydrogen distribution, $\langle g_H \rangle_{\text{HLX}}$. The same conventions as in Figure 3a are used. Maximum 1 near the carbonyl oxygen is 2.3; maximum near 2 is 1.6; and minimum 3 near the β -carbon is 0.4.

Hydrophobic hydration is very contingent on the structure of the interface.⁴⁷ Previous simulations of water around hydrophobic surfaces have found two extremes of behavior. In simulations of water around small, nonpolar solutes,^{4,17,48-50} it was found that water molecules pointed their hydrogens outward (i.e., away from the solute) to maintain their hydrogen bonding. However, in studies of extended hydrophobic surfaces just the opposite happens.⁵⁰⁻⁵³ To optimize packing at the surface, the maintenance of hydrogen bonding is sacrificed, and water molecules point a hydrogen inward. Between these two extremes, the hydration of the β -carbon, which protrudes out into solution like a spherical knob, resembles that of a small solute.

The competition between packing and hydrogen bonding⁵¹ accounts for the different water structures on either side of the helix. On the hydrophobic side, in order to satisfy their hydrogen-bonding requirements, water molecules are constrained to adopt a clathrate-like arrangement, which results in less than optimal packing and accounts for the lower water density around the β -carbon and the backbone atoms. In contrast, around the carbonyl oxygen, water molecules do not have to adopt as constrained a geometry to maintain their hydrogen bonding. This

allows for better packing and accounts for the higher water density around the carbonyl oxygen.

Particular importance is attached here to presenting the results in two-dimensional projections, displayed in 'gray-level'. A common criticism of molecular simulation as applied to proteins is that it reduces a complex three-dimensional structure to an over-simplified one-dimensional profile. The two-dimensional projections address this criticism to some degree. It is hoped they will prove useful in representing other quantities than the geometrical ones presented here. For instance, they could be used to show the spatial variation in dynamic quantities, such as velocities and diffusion constants.^{4,49,54,55}

A major problem with two-dimensional representations is that the statistics are necessarily much poorer than for one-dimensional ones. This problem of bad statistics in water simulations has occurred in the past,^{33,52} and the only general strategy to overcome the problem is to run the simulation longer or at a higher temperature. In the model system investigated here it was possible to improve statistics by exploiting the symmetry of the helix. The improvement in statistics between straight and helical projections was more than a factor of seven. The reason for this dramatic improvement is evident in comparing the straight and helically projected oxygen distribution. The helical distribution has a maximum near the carbonyl oxygen at $\phi = 180^\circ$ and a minimum near the β -carbon at $\phi = 290^\circ$. The angular difference between these extrema is close to the 100° periodicity of residues around a helix. Consequently, the straight projection averages incoherently—peak with trough—while the helical projection averages coherently.

Besides the helical projection, a number of other averaging schemes were attempted. For instance, water positions and orientations were averaged with respect to the centroid of each residue. Then the residue distributions for each of the 14 residues were superposed and averaged. However, none of these other averaging schemes was found to be as effective as the helical projection in capturing the symmetry of the helix and providing a representation that can be reproduced on a printed page.

Here the helical projection was applied to an α -helix. However, helical symmetry is probably the most common form of symmetry found in biological macromolecules, and it is possible that the projections developed here may be applied to other systems. For instance, a number of protein assemblies—such as keratin, fibrinogen, myosin, and tobacco mosaic virus—have overall helical symmetry.^{56,57} DNA, which can arrange itself in three helical structures (A, B, and Z) is an even better example. The helical projection could easily be applied in a simulation of water around DNA,^{58,61} and its use in this capacity is under present investigation.

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) Saenger, W. *Annu. Rev. Biophys. Biophys. Chem.* **1987**, *16*, 93–114.
- (2) Teeter, M. M. *Annu. Rev. Biophys. Biophys. Chem.* **1991**, *20*, 577–600.
- (3) Otting, G.; Liepinsh, E.; Wüthrich, K. *Science* **1991**, *254*, 974–80.
- (4) Rossky, P. J.; Karplus, M. *J. Am. Chem. Soc.* **1979**, *101*, 1913.
- (5) Brooks, C. L. I.; Karplus, M. *Methods Enzymol.* **1986**, *127*, 369–400.
- (6) Ahlström, P.; Teleman, O.; Jönsson, B. *J. Am. Chem. Soc.* **1988**, *110*, 4198–4203.
- (7) Levitt, M.; Sharon, R. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 7557–7561.
- (8) Levitt, M. *Chem. Scr.* **1989**, *29A*, 197–203.
- (9) Ahlström, P.; Teleman, O.; Kördel, J.; Forsén, S.; Jönsson, B. *Biochemistry* **1989**, *28*, 3205–3211.
- (10) Triado-Rives, J.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1990**, *112*, 2773–2781.
- (11) DiCapua, F. M.; Swaminathan, S.; Beveridge, D. L. *J. Am. Chem. Soc.* **1990**, *112*, 6768–6771.
- (12) Avbelj, F.; Moul, J.; Kitson, D. H.; James, M. N. G.; Hagler, A. T. *Biochemistry* **1990**, *29*, 8658–8676.
- (13) Daggett, V.; Levitt, M. *J. Mol. Biol.* **1992**, *223*, 1121–1138.
- (14) Eisenberg, D.; Kauzmann, W. *The Structure and Properties of Water*; Clarendon Press: Oxford, 1969.
- (15) Raghavan, K.; Foster, K.; Berkowitz, M. *Chem. Phys. Lett.* **1991**, *177*, 426–32.
- (16) Schlenkrich, M.; Nicklas, K.; Brickmann, J.; Bopp, P. *Ber. Bunsen-Ges. Phys. Chem.* **1990**, *94*, 133–145.
- (17) Zichi, D. A.; Rossky, P. J. *J. Chem. Phys.* **1985**, *83*, 797–808.
- (18) Hirata, F.; Rossky, P. J. *J. Chem. Phys.* **1981**, *74*, 6867.
- (19) Levitt, M. *J. Mol. Biol.* **1991**, *220*, 1–4.
- (20) Sessions, R. B.; Dauber-Osguthorpe, P.; Osguthorpe, D. *J. Mol. Biol.* **1989**, *209*, 617–633.
- (21) Mehrotra, P. K.; Marchese, F. T.; Beveridge, D. L. *J. Am. Chem. Soc.* **1981**, *103*, 672–3.
- (22) Mezei, M.; Beveridge, D. L. *Methods Enzymol.* **1986**, *127*, 21–47.
- (23) Hagler, A. T.; Moul, J. *Nature* **1978**, *272*, 222–226.
- (24) Hagler, A. T.; Osguthorpe, D. J.; Robson, B. *Science* **1980**, *208*, 599–601.
- (25) Arnott, S.; Dover, S. D. *J. Mol. Biol.* **1967**, *30*, 209–212.
- (26) Baker, E. N.; Hubbard, R. E. *Prog. Biophys. Mol. Biol.* **1984**, *44*, 97–179.
- (27) Chothia, C. *Ann. Rev. Biochem.* **1984**, *53*, 537–72.
- (28) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (29) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187–217.
- (30) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A.; Teller, E. *J. Chem. Phys.* **1953**, *21*, 1087–92.
- (31) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Clarendon Press: Oxford, 1987.
- (32) Owicki, J. C.; Scheraga, H. A. *J. Am. Chem. Soc.* **1977**, *99*, 7413.
- (33) Kincaid, R. H.; Scheraga, H. A. *J. Comput. Chem.* **1982**, *3*, 525–547.
- (34) Daggett, V.; Kollman, P. A.; Kuntz, I. D. *Biopolymers* **1991**, *31*, 1115–1134.
- (35) Blundell, T. B.; Barlow, D.; Borkakoti, N.; Thornton, J. *Nature* **1983**, *306*, 281–283.
- (36) Sundarlingam, M.; Sekharudu, Y. C. *Science* **1989**, *244*, 1333–1337.
- (37) Gray, C. G.; Gubbins, K. E. *Theory of molecular fluids I: Fundamentals*; Clarendon Press: Oxford, 1984.
- (38) Dickerson, R. E.; Geis, I. *The Structure and Action of Proteins*; London, 1969.
- (39) CCP4 *The Cooperative Computing Project 4: Protein Crystallography*; SERC Daresbury Laboratory: Warrington, England, 1991.
- (40) Thomas, D. J. *J. Appl. Cryst.* **1989**, *22*, 498–99.
- (41) Richards, F. M. *Annu. Rev. Biophys. Bioeng.* **1977**, *6*, 151–76.
- (42) Cheng, X.; Schoenborn, B. *Acta Crystallogr.* **1990**, *B46*, 195–208.
- (43) Kossiakoff, A. A. *Annu. Rev. Biochem.* **1985**, *54*, 1195–227.
- (44) Thanki, N.; Thornton, J. M.; Goodfellow, J. M. *J. Mol. Biol.* **1988**, *202*, 637–657.
- (45) Thanki, N.; Umrana, Y.; Thornton, J. M.; Goodfellow, J. M. *J. Mol. Biol.* **1991**, *221*, 669–691.
- (46) Goodfellow, J.; Thanki, N.; Thornton, J. M. *Mol. Sim.* **1989**, *3*, 167–182.
- (47) Franks, F. *Water*; The Royal Society of Chemistry: London, 1983.
- (48) Pangali, C.; Rao, M.; Berne, B. J. *J. Chem. Phys.* **1979**, *71*, 2975–2981.
- (49) Rapaport, D. C.; Scheraga, H. A. *J. Phys. Chem.* **1982**, *86*, 873–880.
- (50) Wallqvist, A.; Berne, B. J. *J. Phys. Chem.* **1985**, *145*, 26.
- (51) Lee, C. Y.; McCammon, J. A.; Rossky, P. J. *J. Chem. Phys.* **1984**, *80*, 4448–4455.
- (52) Valleau, J. P.; Gardner, A. A. *J. Chem. Phys.* **1987**, *86*, 4162–4170.
- (53) Wallqvist, A. *Chem. Phys. Lett.* **1990**, *165*, 437–442.
- (54) Geiger, A.; Rahman, A.; Stillinger, F. H. *J. Chem. Phys.* **1979**, *70*, 273–6.
- (55) Zichi, D. A.; Rossky, P. J. *J. Chem. Phys.* **1986**, *84*, 2814–2822.
- (56) Stryer, L. *Biochemistry*; W. H. Freeman: New York, 1988.
- (57) Branden, C.; Tooze, J. *Introduction to Protein Structure*; Garland Publishing: New York, 1991.
- (58) Hirshberg, M. *Simulations of the Static and Dynamic Properties of the DNA Double Helix*. Ph.D. Dissertation, Weizmann, Institute of Science, 1990.
- (59) Laaksonen, A.; Nilsson, L. G.; Jönsson, B.; Teleman, O. *Chem. Phys.* **1989**, *129*, 175.
- (60) Guldbbrand, L.; Forester, T. R.; Lynden-Bell, R. M. *Mol. Phys.* **1989**, *67*, 473–93.
- (61) Forester, T. R.; McDonald, I. R. *Mol. Phys.* **1991**, *72*, 643–660.
- (62) Brünger, A. T.; Kuriyan, J.; Karplus, M. *Science* **1987**, *235*, 458–60.
- (63) Kraulis, P. J. *J. Appl. Crystallogr.* **1991**, *24*, 946–950.