J Mol. Biol. (1993) **230**, 641–650

What is the Natural Boundary of a Protein in Solution?

Mark Gerstein[1,2] and R. M. Lynden-Bell[2]

[1]*MRC Laboratory of Molecular Biology*
Hills Road, Cambridge CB2 2QH, U.K.

[2]*University Chemical Laboratories*
Lensfield Road, Cambridge CB2 1EW, U.K.

(Received 2 September 1992; accepted 2 December 1992)

At what distance do proteins in solution interact? Molecular simulation of water around two helices is used to address this question. Calculations are done with two ideal, parallel, polyalanine α-helices separated by 9 Å, 11 Å, 13 Å, and 15 Å. The second peak in the oxygen density (or loosely the second shell of water molecules) is used to define a hydration surface around the protein, which separates bulk solvent from water molecules strongly influenced by the protein. The hydration surface is contrasted with the Richards–Connolly molecular surface. It indicates that the helices are not completely separate until 15 Å, while the molecular surface shows complete separation at 13 Å. Suggesting shape-dependent aspects of hydration, the hydration surface only loosely follows the van der Waals outline of the protein surface. In particular, at the 9 Å separation, the van der Waals envelopes of the helices make contact; two narrow crevices are formed on either side of the contact; and the water within the crevices is strongly localized in arrangements bridging the helices. A comparison of these 'normal' water simulations with a simulation of a simple, uncharged solvent highlights the importance of hydrogen bonding in structuring liquid water and further contrasts the molecular surface and the hydration surface.

Keywords: molecular surface; accessible surface; water simulation; hydrophobic effect

1. Introduction

The question of when two proteins in solution are close enough to interact has obvious implications for protein–protein recognition, as in antibodies binding to antigens or serine proteases binding to their inhibitors (see Janin & Chothia, 1990, for a review). It is also important for the assembly of multimeric proteins from individual subunits and for the formation of the molten globule intermediate in the folding of a single subunit (Kim & Baldwin, 1990; Ptitsyn, 1987).

The customary definition of what is together and what is separate involves determining the contact and overlap of protein surfaces. If this is accepted, one regresses to a related question of what is the appropriate definition of a protein surface. Much literature has been devoted to this question. At a basic level the protein surface can be defined in terms of the contours of electron density used in crystallographic structure determination. Alternatively, the surface can be defined as the van der Waals envelope of a known structure. However, in solution a protein surface is clothed by layers of water molecules. Some of these water molecules are strongly perturbed by the protein through hydrogen bonding or hydrophobic interactions. In a sense, they lie within the "region of influence" of the protein and so should be used to help determine the effective protein shape. Furthermore, contours of electron density or van der Waals envelopes are rough in texture and contain regions, such as narrow fissures and cavities, not accessible to water molecules.

To make a protein surface that is more chemically meaningful, an accessible surface derived from rolling a probe water molecule (assumed to be a sphere) on the van der Waals envelope of a protein was developed (Lee & Richards, 1971). Richards (1977) extended the accessible surface into the molecular surface, which has two components. The convex contact surface comprises the part of the van der Waals envelope directly in contact with the probe water molecule. Separate patches of the contact surface are connected by concave and saddle-shaped re-entrant surfaces. The re-entrant surface is the inward facing part of the probe sphere when it makes contact with more than two atoms. Practical calculation of the molecular surface is non-trivial, and the molecular surfaces used here are derived from numerical and analytic algorithms developed by Connolly (1981, 1983*a*,*b*). Other calculation methods have been developed (Finney, 1978; Greer & Bush, 1978; Richmond, 1984).

0022-2836/93/060641-10 $08.00/0

Molecular surfaces have proven to be very useful, particularly in studies involving docking and ligand binding (Kuntz *et al.*, 1982; Shoichet & Kuntz, 1991). However, water does not interact with the protein in the sense of a rolling ball. It has its own structure dictated by highly directional hydrogen bonds. Because they prefer to maintain their hydrogen bonds, water molecules near a protein surface, particularly an apolar one, have fewer allowed orientations than those in solution (Dill, 1990; Franks, 1973, 1983). As a result, they cannot pack as tightly against the protein surface as Lennard-Jones spheres, which have no hydrogen-bonding constraints (Gerstein & Lynden-Bell, 1993*b*). A more realistic definition of the protein surface in solution involves consideration of the structure of water strongly perturbed by the protein as well as the structure of the protein itself. Previous work (Gerstein & Lynden-Bell, 1993*a*) indicated that the second shell of water molecules around a protein provides a convenient boundary between bulk solvent and the water strongly influenced by the protein. Here this second shell is used to define the hydration surface of a protein. The hydration surface is useful for characterizing the interaction of two protein molecules in solution.

In particular, the overlap of hydration surfaces is followed in Monte-Carlo simulations of water around two polyalanine helices. In the simulations the helices are brought together from an initial separation of 15 Å. The hydration surfaces begin to overlap at 13 Å and fuse to form a common envelope at 11 Å. The solvent energetics begin to be affected when the hydration surfaces overlap. At all separations the orientational ordering of the water molecules is contained within the hydration surface. The importance of hydrogen bonding is highlighted by comparing these "normal" water simulations to a simulation with no charges. This comparison highlights the difference between the molecular surface, for which water molecules are modelled as simple, uncharged spheres, and the hydration surface, which is based on normal, charged water.

2. Method

The model system consisted of two parallel, rigid, polyalanine helices at a fixed separation in a bath of water molecules. Monte-Carlo computer simulation was used to determine the structure of the solvent around the helices at 300 K for 4 separations, 9 Å, 11 Å, 13 Å, and 15 Å (between helix axes). This model system is obviously a highly simplified representation for protein structure. However, in assessing its utility it is worth noting that many proteins have α-helices arranged in an almost parallel or antiparallel fashion on their exterior (Chothia & Finkelstein, 1990; Chothia *et al.*, 1981), and the exclusion of water from docking helices is often proposed as a key element in protein folding (Kim & Baldwin, 1982; 1990).

The z axis was chosen along the axis of one helix, and the x axis, along the adjustable separation. Periodic boundary conditions with a $46{\cdot}23\,\text{Å} \times 31{\cdot}1\,\text{Å} \times 20{\cdot}93\,\text{Å}$ cell were used. The cell size in the z-direction closely matched that of 14 residues of α-helix, so the system consisted of a lattice of pairs of infinite, parallel helices separated by an adjustable distance. The water molecules were rigid and interacted with each other by the TIP3P potential (Jorgensen *et al.*, 1983), which has partial charges on each atomic site and a Lennard-Jones interaction between oxygen atoms. The electrostatic interaction was cut off at 7·5 Å as is appropriate for the original TIP3P model. The protein helices were also rigid and the helix–water interaction was modelled using the CHARMM parameter set (Brooks *et al.*, 1983), which puts partial charges only on atomic sites and which subsumes the protons attached to the α and β carbon atoms into a united atom representation.

For each separation the initial configuration was prepared by embedding the helices in a box containing 1000 water molecules, which had been equilibrated at 300 K. Water molecules having van der Waals overlaps with the atoms of either helix were removed by standard procedures with the CHARMM package (Brooks *et al.*, 1983). The number of remaining water molecules varied with the helix separation and is shown in Table 1. Long runs were needed to obtain good statistics, and a Monte-Carlo program (Allen & Tildesley, 1987) was written, which exploited the parallel nature of the computer used, an Alliant FX-2800.† To increase the sampling efficiency, the calculation at each separation was broken into about 30 independent runs, which were executed concurrently. Each run 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 10×10^3 cycles after which averages were accumulated every 5 cycles for 10×10^3 cycles. The results of all the runs at a given separation were simply summed together. As indicated in Table 1 the whole calculation for the 11 Å separation required a total of 324×10^3 cycles for cooling and equilibration and 268×10^3 cycles for accumulating averages. This calculation is comparable to 1·2 ns of molecular dynamics with a 2 fs timestep. (One Monte-Carlo cycle, which consists of attempting moves on all the water molecules, is considered equivalent to a single molecular-dynamics step.)

Figure 1 shows a typical configuration projected onto the x,y plane perpendicular to the axes of the helices. This projection is appropriate for the geometry of the system and helps one to visualize the results. The quantities of interest are the average positions and orientations of the water molecules relative to the helices. To investigate these quantities, the x,y plane was partitioned by a 462×310 grid. In each grid square the number of oxygen atoms, the number of protons, and the x, y, and z components of the water dipole moment were accumulated over the simulation. The oxygen and proton number densities in each grid square, denoted by $g(xy)$ and $g_H(xy)$, were normalized by dividing by the total number of samples and the density of pure water. For water molecules, the oxygen density is essentially the same as the distribution of molecular centers of mass. The mean

† The simulation program was written in ANSI C to run under UNIX. Running on a single Intel i860 processor on the Alliant, the calculation took ~4·5 central processor unit (CPU) s/cycle. As evident from Table 1, total calculation of all 4 separations required about 2×10^6 cycles for heating, cooling, and averaging. This would require ~100 CPU days on a single i860. However, running in parallel and distributing the calculation over 8 to 16 processors decreased actual time for the calculation to about 10 CPU days.

Table 1

Solvent energies as a function of separation

	Helix separation (Å)			
	15	13	11	9
Number of water molecules	908	908	900	909
MC cycles for averages†	260,000	179,900	268,100	165,550
Total energies (kJ/mol)‡				
Helix–water	−1278	−1277	−1304	−1230
Water–water§	777	862	934	825
Energy changes per residue w.r.t. 15 Å‖				
Helix–water		0·0	−0·9	1·7
Water–water		3·1	5·6	1·7
Solvation energy¶		3·1	4·7	3·4

† The number of cycles for cooling and equilibration is roughly 120% of the number quoted for averaging. A single Monte-Carlo (MC) cycle is equivalent to attempting a move on all the water molecules. It is usually considered equivalent to one molecular-dynamics move.

‡ The water–water and water–helix energies reported are totals for the whole system averaged over the whole simulation. The difference between the averages of each half of a run was used to estimate the error, which was ±25 kJ/mol.

§ The water–water energy ΔU_{ww} is reported relative to the energy of an equivalent number of bulk TIP3P water molecules:

$$\Delta U_{ww}=\frac{1}{2}\sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j}}^{n} U_{ij}-nU_{bulk},$$

where n is the number of water molecules in the simulation; U_{ij} is TIP3P pair potential between water molecules i and j (using periodic boundaries and the minimum image convention); and Ubulk is 41·1 kJ/mol, the energy of bulk TIP3P water at 300 K, the temperature for all runs (Jorgensen *et al.*, 1983).

‖ The water–water and helix–water energies are also expressed relative to the 15 Å separation on a per residue basis by subtracting the respective energy at 15 Å and dividing by 28.

¶ The solvation energy is the sum of the water–water and helix–water energies.

dipole moment is a measure of the degree of orientation of the solvent molecules by the helices. Its magnitude, the polarization P, is given by:

$$P=\frac{\sqrt{\langle p_x\rangle^2+\langle p_y\rangle^2}}{p_o\rho}$$

where $\langle p_x\rangle$ and $\langle p_y\rangle$ are the x and y components of the dipole moment per unit volume at a given point x,y averaged over the simulation, p_o is the magnitude of water dipole moment, and ρ is the number density of pure water. In the TIP3P model there is no electronic polarizability, so the polarization reported here results solely from the alignment of water molecules by the helices.

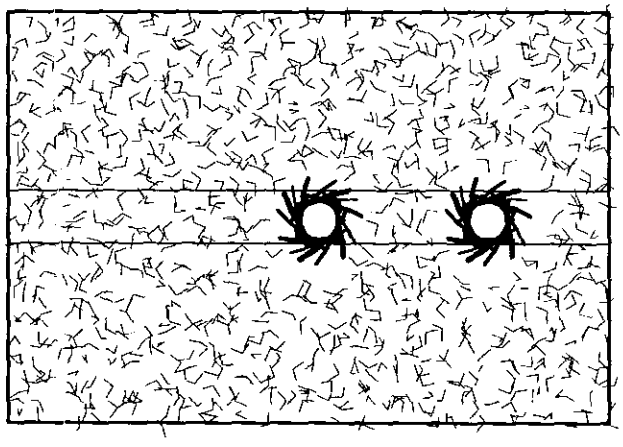

Figure 1. An instantaneous snapshot of the model system. The system consists of 2 parallel, 14-residue polyalanine helices surrounded by 908 TIP3P water molecules. The view is down the axis of the central helix (the z-axis). The other helix is related by a 13 Å displacement in the x direction. Lines enclose the region $-2<y<2$ along the x axis. The density in this region is averaged to make the cross-section plots shown in Fig. 2.

Two additional simulations were done at the 11 Å separation. First, to test some of the simplifications used in the calculation, a short (88 ps) molecular-dynamics simulation was done with partially mobile helices. That is, the helices were fixed at their α-carbon positions, but all other atoms, in particular the carbonyl oxygen atoms, were allowed to move. The simulation was done with the CHARMM package (Brooks *et al.*, 1983) using an all-atom parameter set that explicitly included the protons attached to the α and β carbon atoms. Because of the rapid vibrations of these protons as well as those on water molecules, a short 0·5 fs timestep was used. No differences were seen in comparison to the Monte-Carlo simulations, indicating that the united atom representation and the rigid helices did not significantly affect the results.

To highlight the effect of hydrogen bonding, an additional Monte-Carlo simulation was done with all the charges switched off. As the hydrogen atoms in the original TIP3P model have only electrostatic interactions, the simulation is effectively of a Lennard-Jones liquid with TIP3P oxygen parameters. As discussed elsewhere (Gerstein & Lynden-Bell, 1993*b*), simply turning off the charges gives a high-pressure fluid above the critical point, so it is necessary to adjust the temperature and density of the solvent to values appropriate to a Lennard-Jones liquid. Here a temperature of $0{\cdot}94\varepsilon/k_B$ and a density of $0{\cdot}88/\sigma^3$ were chosen from past work on liquid argon (Verlet, 1967), where σ and ε are the two Lennard-Jones parameters for oxygen. This temperature and density corresponds to doing a simulation with 272 solvent molecules in the same size cell at a temperature of

72 K. To keep the relative sizes and interactions of the helices and the solvent the same, the Lennard-Jones parameters for the helix atoms were unchanged.

As a Lennard-Jones liquid equilibrates more rapidly than water, it was only necessary to run the simulation for $2{\cdot}0 \times 10^5$ cycles. The simulation used the same 7·5 Å non-bonded cutoffs as the normal water simulations.

3. Definition of the Hydration Surface

The hydration surface is defined as the second shell of water molecules around a protein. It forms an approximate boundary between bulk solvent and solvent strongly perturbed by the protein. This definition is based on previous simulations of a single polyalanine helix in water (Gerstein & Lynden-Bell, 1993*a*), where it was found that the second peak in the distribution function for water around a protein formed a continuous surface. Outside this surface water molecules had energies indistinguishable from those of the bulk solvent and were not detectably oriented by the protein. In contrast, molecules within the surface were strongly orientated by the protein and had altered energies relative to the bulk. They were often involved in protein–water hydrogen bonds. All water molecules within the hydration surface will be referred to as belonging to the first-shell.

The distinction between first and second shell water molecules is not specific to the model system studied here. It has been observed in molecular simulations of fully modelled proteins (Ahlström *et al.*, 1988; Levitt & Sharon, 1988), in a neutron diffraction study of carbonmonoxymyoglobin (Cheng & Schoenborn, 1990; 1991), in X-ray crystallographic analyses (Saenger, 1987; Teeter, 1991), and in n.m.r.† experiments (Halle *et al.*, 1981; Otting *et al.*, 1991).

No division between bulk and perturbed solvent is exact, however, and the protein has some small effects on the solvent beyond the hydration surface, such as creating a weak third peak in the distribution function. Furthermore, the hydration surface is based on calculations done with a model system that has polar and non-polar accessible surface but no residues with net charges. A highly charged surface, such as that found for the guanidinium group on an arginine residue, would be expected to more strongly influence water structure. The DNA double helix, which is very highly charged, provides an extreme example of the long-range effect of electrostatics. Experiments (Rau *et al.*, 1984; Rau & Parsegian, 1992) and simulations (Forester & McDonald, 1991; Guldbrand *et al.*, 1989) on DNA fibres have suggested that a DNA molecule may be able to influence water structure out to 20 Å from its surface.

Proteins, however, are not nearly as highly charged as DNA, in terms of either their net charge or the percentage of their accessible surface that is charged. With 5% r.m.s. variations, the average water-accessible surface of both monomeric and oligomeric proteins is 57% non-polar, 24% polar, and only 19% charged (Janin *et al.*, 1988; Miller *et al.*, 1987). Consequently, the effect of long-range electrostatics on water is expected to be much less for proteins than DNA, and the second-shell hydration surface is probably most useful for characterizing the 81% of the protein surface that is not charged.

† Abbreviations used: n.m.r., nuclear magnetic resonance; r.m.s., root-mean-square.

4. Change in the Hydration Surface with Helix Separation

Figure 2 shows the normalized oxygen density around the helices at four separations: 9 Å, 11 Å, 13 Å, and 15 Å. At each separation the density is depicted in two ways. First, a "grey-level" graph shows it projected onto the x,y plane. Because of the noise and constraints of reproduction, it is necessary to show the density thresholded. All the values above 1·2 are black, and all below, white. The central region of each helix where water does not penetrate is shown in grey. This grey-level representation gives one a good feeling for the overall structure of the density in two dimensions. Because of the noise, contour plots are not as effective. It is difficult, however, to represent the magnitude of the density bands in grey-level. To overcome this difficulty, a one-dimensional plot that shows a cross-section through the grey-level graph is presented at the bottom of each Figure. It shows the magnitude of the oxygen density along the x-axis averaged in the region between $-2 < y < 2$ Å. (The region is indicated in Fig. 1.) At each separation, the x,y projection of a Richards–Connolly molecular surface for the two helices is drawn to the same scale for comparison.

In the grey-level graphs there is an inner ring of dark peaks and then a continuous dark band around the helices at all separations. The ring of peaks correspond to localized concentrations of first-shell water molecules, and the continuous outer band, to the hydration surface. For all separations, the oxygen density in the regions that face away from the helix–helix interface ($x<0$ or $x>11$ Å for the 11 Å separation) has essentially the same structure as found previously for an isolated helix (Gerstein & Lynden-Bell, 1993*a*). That is, the peak 6·3 Å from the helix axis with an average height of 1·3 corresponds to the hydration surface. A sharp peak at 4·6 Å corresponds to the isolated water molecules in the first shell. The height of this peak is on average 1·1, but reflecting the discontinuity in the first shell, it varies greatly around the helix. A third maximum at 9 Å shows that the influence of the helix extends beyond the hydration surface to a small degree.

In contrast to the regions that face away from the interface, the oxygen density in the region between the helices varies greatly depending on helix separation. At 15 Å (Fig. 2, bottom right), both hydration and molecular surfaces show that the helices do not interact. One can discern both localized concentrations of first-shell water molecules

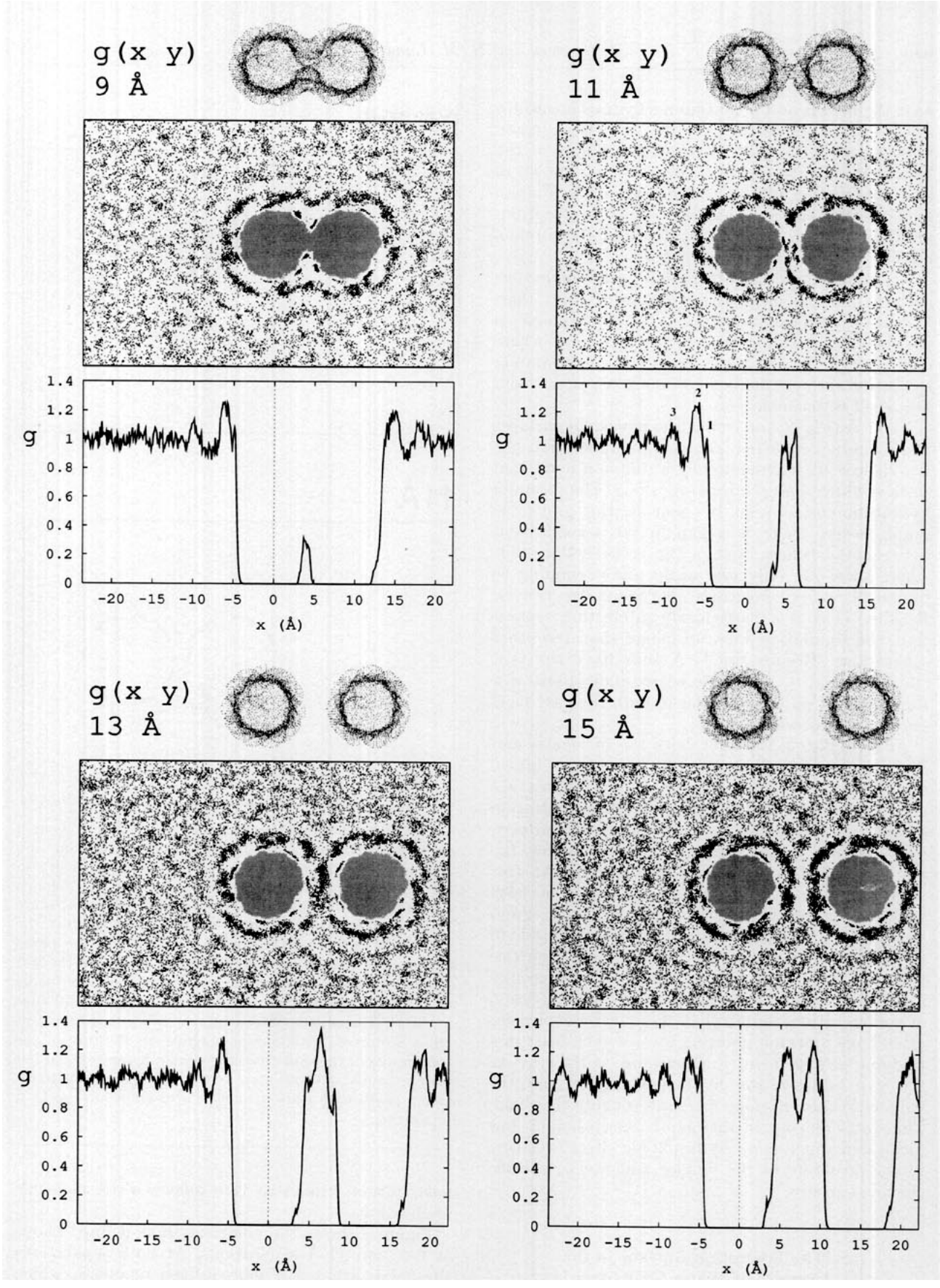

Figure 2. The normalized oxygen density $g(xy)$ for the 2 helices at 4 different separations, 9 Å, 11 Å, 13 Å, and 15 Å. Each of the 4 separations is depicted in 3 ways. Top: A projection of the Richards–Connolly molecular surface of the helices. Middle: The oxygen density projected onto a plane perpendicular to the helix axes (the xy plane) is shown in a grey-level graph. For ease in reproduction it is thresholded at 1·2. That is, all values above 1·2 are black and all below 1·2 are white. The central regions of the helices where water does not penetrate are shown in grey. At all separations one sees an inner ring of dark peaks and then a continuous dark band around the helices. The ring of peaks corresponds to localized concentrations of first-shell water molecules, and the continuous outer band, to the hydration surface. Bottom: A cross-section through the grey-level graph along the x-axis. The region (shown in Fig. 1) between $y = \pm 2$ is averaged. For the cross-section plot at 11 Å separation, the 3 shells of water molecules in 1 region facing away from the helix–helix interface are indicated by numbers.

near the helices and a continuous hydration surface around each of them. In the interface region between the helices, the hydration surfaces of the helices are clearly separate. Moreover, as shown in the cross-section plot at the bottom of the Figure, between the helices the oxygen density is exceptionally low because of the coincidence of the minima outside the two hydration surfaces.

At 13 Å (Fig. 2, bottom left), the molecular surfaces of the helices are separate, but their hydration surfaces overlap. The overlapping is particularly clear in the cross-section plot, where one sees a broad peak from the overlapping hydration surfaces flanked by two sharp peaks from first-shell water molecules.

At 11 Å (Fig. 2, top right), the molecular surfaces make contact at a few points, but in solution the two helices are surrounded by a common hydration surface which does not separate them. The common hydration surface can be seen in the grey-level graph, where there is a gap in the second band between the helices, and in the cross-section plot, where there are only two peaks, corresponding to first-shell water molecules, between the helices. Because 11 Å is a particularly interesting separation, the simulation was run longer than the other separations. Results for 11 Å thus have the best statistics, and it is possible to see indications of a third shell of water molecules in both the grey-level graph and cross-section plot.

At 9 Å (Fig. 2, top left), both the molecular and hydration surfaces indicate that the helices are in contact. In the grey-level graph, the helices are collectively surrounded by a common hydration surface, which follows a smooth course. In contrast, the first-shell water molecules closely follow the van der Waals envelope of the helices. At this separation the van der Waals envelopes of both helices (essentially shown by the region in grey) have merged, forming two crevices on either side of the helix–helix contact. At the base of each crevice there is a large number of first-shell water molecules (density maximum at crevice base is 1·3). Inspection of configurations shows that these water molecules often have a proton pointing toward each helix in a bridging arrangement. Furthermore, at the crevices the gap between the hydration surface and the van der Waals envelope increases roughly threefold. The water density in this gap is exceptionally low (0·23) and suggests a "hydrophobic effect" relating to the structure of the crevice and not to specific chemical groups.

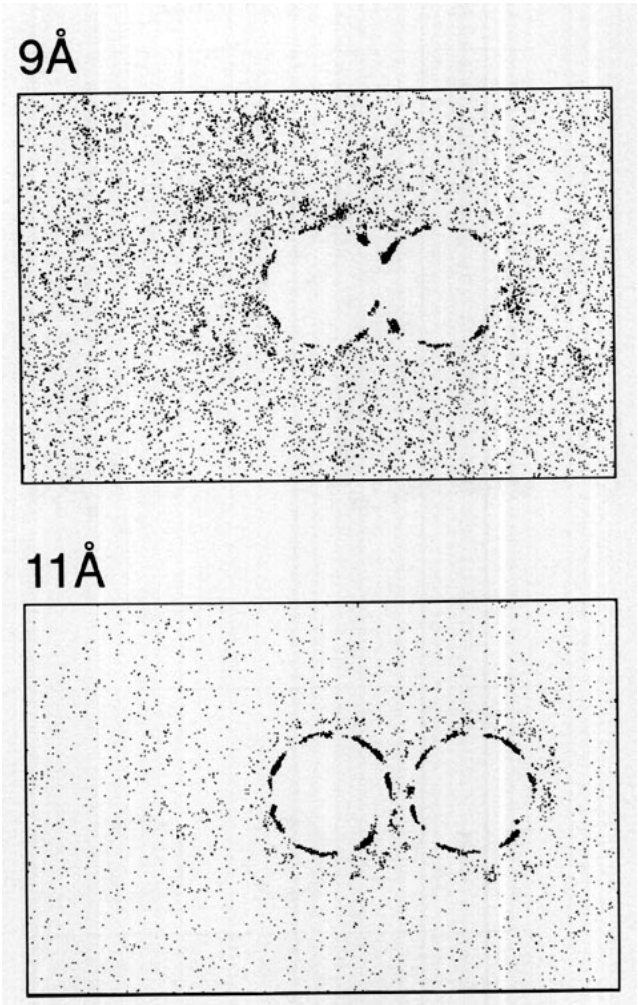

Figure 3. The polarization $P(xy)$ of the 2-helix system at a separation of 9 Å and 11 Å. Both separations are depicted in a similar fashion to the oxygen density in Fig. 2. However, the grey-level graphs are shown without thresholding. At each separation the helices are surrounded by a ring of peaks. In the 11 Å separation, the maximum of most of the peaks is ~0·3, and the peaks between the helices are slightly higher (0·35). In the 9 Å separation, the peaks on the sides of the helices facing away from the interface are the same magnitude (0·3) as in the 11 Å separation. However, the peaks in the crevices are a factor of 2·5 larger in magnitude. Because of the constraints of reproduction, this dynamic range could not be represented without inaccurately diminishing outward-facing peaks with respect to their appearance in the 11 Å separation.

5. The Hydration Surface in the Context of Water Structure

(a) *Strongly oriented water only within the hydration surface*

The second water shell has been used to define a hydration surface. It was identified by looking for maxima in the oxygen density. It is worthwhile to consider how the hydration surface relates to the polarization, a quantity that reflects water-molecule orientation.

Figure 3 shows the polarization as defined above at 9 Å and 11 Å separations. At both separations the polarization is a short-ranged quantity, which dies off rapidly beyond the first shell. As the polarization is the average magnitude of the dipole moment per unit volume, this short-ranged behaviour indicates that dipolar ordering is confined to within the hydration surface. There are, however, other orientational quantities that are not quite as short-ranged. An analogous quantity (not shown)

can be computed for the quadrupolar ordering (Gerstein & Lynden-Bell, 1993*a*). Although it is still contained within the hydration surface, it is longer-ranged than the polarization and indicates that dipoles of water molecules in the first and second shell lie parallel to the protein surface.

The polarization at 9 Å separation has a particularly high peak at the base of each crevice. This peak results from bridging water molecules which are strongly aligned by interactions with both helices. Clear peaks from protons attached to these bridging water molecules can be resolved in the hydrogen density (not shown). At 11 Å, there is a ring of peaks around the helix with approximate sevenfold symmetry. Each peak represents a region of highly orientated water and has a location that is consistent with the geometry of water molecules hydrogen-bonding to the carbonyl oxygen atom (Baker & Hubbard, 1984; Sundarlingam & Sekharudu, 1989). Because of its short-ranged character, the polarization is little changed at 13 Å and 15 Å separation from the 11 Å separation and so is not shown.

(b) *Chargeless simulation and the hydration surface*

In comparison to the polarization, the hydration surface does not directly reflect water molecule orientation. However, it does reflect the hydrogen-bonded structure of water. Figure 4 shows the solvent (oxygen) density for the simulation with all the charges switched off. There is no water–water or water–protein hydrogen-bonding, so packing considerations completely structure the solvent. The solvent structure in this chargeless simulation is radically different from that in the normal water simulations.Within 9 Å from the helix axis, there are at least five discernible solvent shells in the region facing away from the helix–helix interface. These five shells contrast with the three observed in the normal water simulation at 11 Å. The second shell, which forms the continuous hydration surface in the normal water simulation, is not continuous in the chargeless simulation. Furthermore, it is also much closer to the helices in the chargeless simulation. As a result, at the 11 Å separation the helices are separated by two non-overlapping second shells in the chargeless simulation while in the normal water simulation they are enclosed by a common hydration surface.

The Lennard-Jones molecules in the chargeless simulation are similar to the rolling probe spheres used to define the Richards–Connolly molecular surface. As they are not constrained by hydrogen bonding, they can fit more deeply into cavities on the protein surface than normal water molecules. The resulting solvent structure is quite different from the clathrate structures formed around hydrophobic surfaces by normal water (Franks, 1983). It is evident in the many high and localized density peaks near the helices. Because of the great height of these peaks (up to 8), a Lennard-Jones solvent molecule has a greater probability of being very close to the helix than a normally charged water molecule does. Note, however, that it cannot approach the helix any more closely in an absolute sense. That is, the distance between a first-shell peak and the protein is roughly the same in both the normal-water and chargeless simulations. (This is, in turn, a direct consequence of the same Lennard-Jones solvent radius σ being used in both simulations.)

Thus, the comparison between the normal-water and chargeless simulations shows how radically

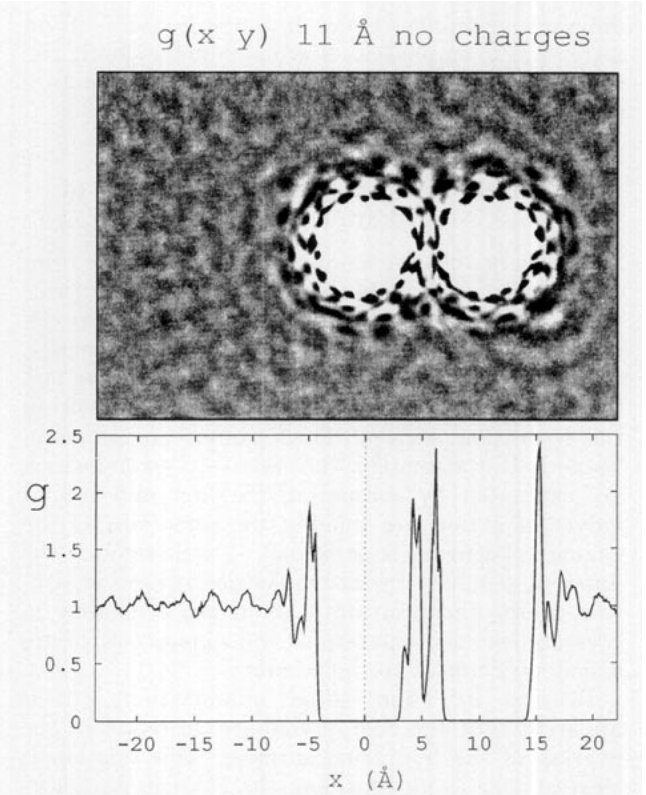

Figure 4. The solvent (oxygen) density $g(xy)$ in a simulation with no charges. The 2 helix system at a separation of 11 Å is shown. This density is depicted in a similar fashion to the oxygen density in Fig. 2. However, the grey-level graph is shown without thresholding. The Figure shows rings of high peaks around the helices. These rings result from solvent molecules packing into crevices in the protein surface, which are inaccessible to normal hydrogen-bonded water. In the innermost ring, each peak has a maximum height of ~8; and in next ring, ~5. The cross-section plot indicates that the average (i.e. integrated) density in the first and second rings is greater than 2. In comparison, the cross-section plot in Fig. 2, top right, shows that the height of the first 2 peaks in the normal-water simulation is less than 1·4. This difference in average peak height implies that a solvent molecule is *on average* closer to the helices in the chargeless simulation. However, in an absolute sense a Lennard-Jones solvent molecule can not approach the helices any more closely than a normally charged water molecule can. That is, since peak location is essentially determined by the Lennard-Jones radius σ, which is the same in both simulations, the location of the first peak, unlike its height, is roughly the same in both chargeless and normal-water simulations.

different a simple, Lennard-Jones liquid is from water and so emphasizes how much more closely the hydration surface is related to real water structure than the molecular surface is.

6. Solvent Energies and The Interaction of Hydration Surfaces

The energy change in moving the two helices from one separation to another is composed of three parts: the change in the direct interaction between the helices, the change in the water–water energy, and the change in the helix–water energy. The last two contributions arise from water molecules within the hydration surface. Their values are given in Table 1. Unfortunately, the statistical errors, which are estimated by comparing the first and second halves of a run, are roughly the same size as the changes between separations. Furthermore, the different number of water molecules at each separation causes an uncertainty about the constancy in pressure between separations and so may contribute a systematic error to the results.

Bearing in mind these qualifications, it is apparent that the energy changes do relate to the overlap of the hydration surfaces. This relation is clearest when one considers the changes in helix–water and water–water energies on reducing the separation from 15 Å. At 13 Å, where the hydration surfaces overlap, the helix–water energy is unchanged. However, the disruption in the hydrogen-bonded structure caused by the overlapping second shells (hydration surfaces) increases the water–water energy (i.e. makes the water–water interaction less strong). At 11 Å where a common hydration surface is established the solvent structure is further disrupted, leading to a maximum water–water energy. In addition, the minimum protein–water energy suggests that there may be some stabilization due to water molecules bridging the two helices (as discussed above for the crevice). At 9 Å, where the van der Waals surface of the proteins touch, each protein is in contact with fewer water molecules, so the helix–water energy is necessarily a maximum. The water molecules, on the other hand, can now more easily form a hydrogen-bonded structure around both helices so that the water–water energy is nearly restored to its value at 15 Å.

Because of the statistical and systematic errors inherent in the calculation, it was decided not to attempt potential of mean force calculations on the helices. Very long runs in the isothermal–isobaric ensemble would be needed to overcome these sources of error and obtain reliable results.

7. Conclusion

The original Lee–Richards accessible surface and its extensions into the Connolly–Richards molecular surface have proven enormously useful in characterizing and measuring protein shape (see, for example, Creighton, 1984; Islam & Weaver, 1991).

The molecular surface is based on a water molecule moving over the protein surface like a rolling ball. That is, the water is modelled like a Lennard-Jones liquid akin to liquid argon. Proteins, and life for that matter, do not exist in liquid argon. Here a suggestion is made for a more appropriate surface for a protein in aequous solution: the locus of second-shell water molecules. This hydration surface takes into account the constraints of hydrogen-bonded water structure as well as that of protein structure. Comparison of the simulations of normal and chargeless water indicates how profoundly these constraints affect solvent structure.

The difference between the molecular surface and the hydration surface can be understood in terms of excluded volumes. Both surfaces exclude the volume in cracks and fissures on the protein surface, which are too small to accommodate a water molecule. However, the hydration surface also excludes cavities that are large enough to fit a water molecule but that greatly disrupt normal water structure. This second type of cavity is apparent in comparing the normal and chargeless water simulations. Furthermore, the hydration surface excludes the volume occupied by first-shell water molecules. These molecules are strongly influenced by the protein and are experimentally different from those in the bulk.

The construction of the hydration surface is illustrated with a two-helix model system. Both the molecular and hydration surfaces indicate that the helices are separate at 15 Å and in contact at 9 Å. However, at 13 Å, the molecular surfaces are separate, but the hydration surfaces overlap. At 11 Å, the molecular surfaces overlap, while the hydration surfaces have fused into a single surface around both helices. Using the hydration surface as a criterion, the helices begin to interact at 13 Å, while using the molecular surface, the critical separation is reduced to 11 Å.

The hydration surface also illuminates shape-dependent aspects of hydration. At 9 Å separation a crevice is found on either side of the helix–helix contact. The hydration surface does not closely follow the van der Waals envelope of the helix around the crevices. There are highly orientated, first-shell water molecules at the bases of both crevices, but the centre of the crevices has a minimum probability of water occupancy. That is, because of its shape and not because of its chemical constitution (which did not change between separations), the centre of each crevice is an unfavorable location for water molecules, i.e. it is hydrophobic. It has been previously observed (Nicolls *et al.*, 1991; Sharp *et al.*, 1991) that the hydrophobicity of a concave surface is inversely proportional to its radius of curvature. This relationship between surface curvature and hydrophobicity is in accordance with the results presented here and perhaps results from the highly directional water structure being unable accommodate sharp clefts as well as a Lennard-Jones liquid.

It is possible to construct a three-dimensional

hydration surface in analogy with the three-dimensional molecular surface. However, for good statistics much more computation than was done here would be necessary. The calculation of the hydration surface can be easily parallelized by doing many Monte-Carlo runs simultaneously and then combining the results. This "coarse-grained" parallelism makes the calculation ideally suited to be distributed over networks of workstations (Carriero & Gelernter, 1989; Gelernter, 1987). Consequently, the calculation of the full three-dimensional hydration surface of real proteins is definitely possible. However, it must be noted that even extremely long simulations will probably never provide good enough statistics to define the hydration surface as precisely as the molecular surface.

Bearing in mind this qualification, construction and measurement of hydration surfaces of real proteins could perhaps improve three customary calculations on protein structure. The solvation energy of a protein has been related to its accessible area (Chothia, 1974; Eisenberg & McLachlan, 1987). The hydration surface may be able to provide more realistic surface areas for these calculations. Second, the calculation of the packing density for a protein requires an accurate demarcation between protein and solvent, for water and protein have different packing densities (Finney, 1978; Richards, 1974; 1977; 1979; Richmond, 1984). Previous calculations represented the water structure on a simple cubic lattice. The hydration surface could provide a more realistic way of partitioning space between protein and solvent. Third, insofar as the hydration surface provides a more accurate shape for the protein in solution, it may be useful for the many geometrical approaches taken in docking and ligand-binding studies (Cherfils *et al.*, 1991; Duquerroy *et al.*, 1991; Jiang & Kim, 1991; Kuntz, 1992; Kuntz *et al.*, 1982; Shoichet & Kuntz, 1991; Wodak *et al.*, 1987).

Thanks are given to Ian McDonald, Cyrus Chothia, and Arthur Lesk for useful discussions; the Herchel-Smith foundation for a scholarship (MG); and the SERC and the Newton Trust for support in the purchase of equipment.

References

Ahlstrom, P., Teleman, O. & Jönsson, B. (1988). Molecular dynamics simulation of interfacial water structure and dynamics in a parvalbumin solution. *J. Amer. Chem. Soc.* **110**, 4198–4203.

Allen, M. P. & Tildesley, D. J. (1987). *Computer Simulation of Liquids*. Oxford: Clarendon Press.

Baker, E. N. & Hubbard, R. E. (1984). Hydrogen bonding in globular proteins. *Progr. Biophys. Mol. Biol.* **44**, 97–179.

Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. & Karplus, M. (1983). CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comp. Chem.* **4**, 187–217.

Carriero, N. & Gelernter, D. (1989). Linda in context. *Comm. of the ACM*, **32**, 444–458.

Cheng, X. & Schoenborn, B. (1990). Hydration in protein crystals: a neutron diffraction analysis of carbonmonoxymyoglobin. *Acta Crystallogr. sect. B*, **46**, 195–208.

Cheng, X. & Schoenborn, B. P. (1991). Neutron diffraction study of carbonmonoxymyoglobin. *J. Mol. Biol.* **220**, 381–399.

Cherfils, J., Duquerroy, S. & Janin, J. (1991). Protein-protein recognition analyzed by docking simulation. *Proteins: Struct. Funct. Genet.* **11**, 271–80.

Chothia, C. & Finkelstein, A. V. (1990). The classification and origins of protein folding patterns. *Annu. Rev. Biochem.* **59**, 1007–39.

Chothia, C., Levitt, M. & Richardson, D. (1981). Helix to helix packing in proteins. *J. Mol. Biol.* **145**, 215–250.

Chothia, C. H. (1974). Hydrophobic bonding and accessible surface area in proteins. *Nature (London)*, **248**, 338–339.

Connolly, M. (1983*a*). Analytical molecular surface calculation. *J. Appl. Crystallogr.* **16**, 548–558.

Connolly, M. (1983*b*). Solvent-accessible surfaces of proteins and nucleic acids. *Science*, **221**, 709–713.

Connolly, M. L. (1981). The MS program. *QCPE Bull.* **1**, 75.

Creighton, T. E. (1984). *Proteins*. Freeman, San Francisco.

Dill, K. A. (1990). Dominant forces in protein folding. *Biochemistry*, **29**, 7133–7155.

Duquerroy, S., Cherfils, J. & Janin, J. (1991). Protein–protein interaction: an analysis by computer simulation. *Ciba Found. Symp.* **161**, 237–49.

Eisenberg, D. & McLachlan, A. D. (1987). Solvation energy in protein folding and binding. *Nature (London)*, **319**, 199–203.

Finney, J. L. (1978). Volume occupation, environment, and accessibility in proteins. Environment and molecular area of RNase-S. *J. Mol. Biol.* **119**, 415–441.

Forester, T. R. & McDonald, I. R. (1991). Molecular dynamics studies of the behaviour of water molecules and small ions in concentrated solutions of polymeric *B*-DNA. *Mol. Phys.* **72**, 643–660.

Franks, F. (Ed.) (1973). *Water: A Comprehensive Treatise*. Plenum Press, New York.

Franks, F. (1983). *Water*. The Royal Society of Chemistry, London.

Gelernter, D. (1987). Programming for advanced computing. *Sci. Amer.*, **257**, 90–98.

Gerstein, M. & Lynden-Bell, R. M. (1993*a*). Simulation of water around a model protein helix. 1. Two-dimensional projections of solvent structure. *J. Phys. Chem.* in press.

Gerstein, M. & Lynden-Bell, R. M. (1993*b*). Simulation of water around a model protein helix. 2. The relative contributions of packing, hydrophobicity, and hydrogen-bonding. *J. Phys. Chem.* in press.

Greer, J. & Bush, B. L. (1978). Macromolecular shape and surface maps by solvent exclusion. *Proc. Nat. Acad. Sci., U.S.A.* **75**, 303–7.

Guldbrand, L., Forester, T. R. & Lynden-Bell, R. M. (1989). Distribution and dynamics of mobile ions in systems of ordered *B*-DNA. *Mol. Phys.* **67**, 473–93.

Halle, B., Andersson, T., Forsén, S. & Lindman, B. (1981). Protein hydration from water oxygen-17 magnetic relaxation. *J. Amer. Chem. Soc.* **1981**, 500–508.

Islam, S. A. & Weaver, D. L. (1991). Variation of folded polypeptide size with probe size. *Proteins: Struct. Funct. Genet.* **10**, 300–314.

Janin, J. & Chothia, C. (1990). The structure of protein–protein recognition sites. *J. Biol. Chem.* **265**, 16027–16030.

Janin, J., Miller, S. & Chothia, C. (1988). Surface, subunit interfaces and interior of oligomeric proteins. *J. Mol. Biol.* **204**, 155–64.

Jiang, F. & Kim, S.-H. (1991). "Soft-docking": matching of molecular surface cubes. *J. Mol. Biol.* **219**, 79–102.

Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935.

Kim, P. S. & Baldwin, R. L. (1982). Specific intermediates in the folding reactions of small proteins and the mechanism of protein folding. *Annu. Rev. Biochem.* **51**, 459–89.

Kim, P. S. & Baldwin, R. L. (1990). Intermediates in the folding reactions of small proteins. *Annu. Rev. Biochem.* **59**, 631–660.

Kuntz, I. D. (1992). Structure-based strategies for drug design and discovery. *Science*, **257**, 1078–1082.

Kuntz, I. D., Blaney, J. M., Oatley, S. J., Langridge, R. & Ferrin, T. E. (1982). A geometric approach to macromolecule ligand interaction. *J. Mol. Biol.* **161**, 269–288.

Lee, B. & Richards, F. M. (1971). The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* **55**, 379–400.

Levitt, M. & Sharon, R. (1988). Accurate simulation of protein dynamics in solution. *Proc. Nat. Acad. Sci., U.S.A.* **85**, 7557–7561.

Miller, S., Janin, J., Lesk, A. M. & Chothia, C. (1987). Interior and surface of monomeric proteins. *J. Mol. Biol.* **196**, 641–56.

Nicolls, A., Sharp, K. A. & Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins: Struct. Funct. Genet.* **11**, 281–296.

Otting, G., Liepinsh, E. & Wüthrich, K. (1991). Protein hydration in aequous solution. *Science*, **254**, 974–80.

Ptitsyn, O. B. (1987). Protein folding: hypotheses and experiment. *J. Protein Chem.* **6**, 272–93.

Rau, D. C., Lee, B. & Parsegian, V. A. (1984). Measurement of the repulsive force between polyelectrolyte molecules in ionic solution–hydration forces between parallel DNA double helices. *Proc. Nat. Acad. Sci., U.S.A.* **81**, 2621–2625.

Rau, D. C. & Parsegian, V. A. (1992). Direct measurement of the intermolecular forces between counterion-condensed DNA double helices–evidence for long-range attractive hydration forces. *Biophys. J.* **61**, 246–259.

Richards, F. M. (1974). The interpretation of protein structures: total volume, group volume distributions and packing density. *J. Mol. Biol.* **82**, 1–14.

Richards, F. M. (1977). Areas, volumes, packing, and protein structure. *Annu. Rev. Biophys. Bioeng.* **6**, 151–76.

Richards, F. M. (1979). Packing defects, cavities, volume fluctuations, and access to the interior of proteins. Including some general comments on surface area and protein structure. *Carlsberg. Res. Commun.* **44**, 47–63.

Richmond, T. J. (1984). Solvent accessible surface area and excluded volume in proteins: analytical equations for overlapping spheres and implications for the hydrophobic effect. *J. Mol. Biol.* **178**, 63–89.

Saenger, W. (1987). Structure and dynamics of water surrounding biomolecules. *Annu. Rev. Biophys. Biophys. Chem.* **16**, 93–114.

Sharp, K. A., Nicholls, A., Fine, R. F. & Honig, B. (1991). Reconciling the magnitude of the microscopic and macroscopic hydrophobic effects. *Science*, **252**, 106–109.

Shoichet, B. K. & Kuntz, I. D. (1991). Protein docking and complementarity. *J. Mol. Biol.* **221**, 327–346.

Sundarlingam, M. & Sekharudu, Y. C. (1989). Water-inserted α-helical segments implicate reverse turns as folding intermediates. *Science*, **244**, 1333–1337.

Teeter, M. M. (1991). Water–protein interactions: theory and experiment. *Annu. Rev. Biophys. Biophys. Chem.* **20**, 577–600.

Verlet, L. (1967). Computer "experiments" on classical fluids. I. Thermodynamic properties of Lennard-Jones molecules. *Phys. Rev.* **159**, 98–103.

Wodak, S. J., De, C. M. & Janin, J. (1987). Computer studies of interactions between macromolecules. *Progr. Biophys. Mol. Biol.* **49**, 29–63.

Edited by F. Cohen