
Crystallography Without a Lattice^{a,b}

John L. Finney^c

Abstract

Molecular structure determination in crystals depends on the presence of the crystal lattice. In liquids, there is no underlying lattice, so phase relationships between scattered X-rays or neutrons are not preserved. Hence, only pair-distance distributions can be obtained from simple diffraction experiments. Advances in the past 30 years in radiation sources, instrumentation, and computing have enabled us to go beyond this apparent limitation. We can now obtain detailed structural information on relatively complex liquids and thus see clearly for the first time how molecules actually interact in solution, and how the solvent is perturbed by the presence of the solute. Using as an example recent work on aqueous solutions of tertiary butanol, the ways in which we can obtain high quality structural information in the absence of a lattice are summarized. Not only can we see how intermolecular interactions are influenced by changes in temperature and concentration, we can also begin to see where the structural source of the entropic driving force for the hydrophobic interaction may be found.

Keywords

Liquid structure • Neutron scattering • Aqueous solutions • Hydrophobic interactions

Prologue

In the late summer of 1972, half way across the Pacific Ocean, I got into conversation with David Harker. We were both returning on a flight chartered by the American Crystallographic Association, to facilitate attendance at the 1972 Congress of the International Union of Crystallography in Kyoto, Japan. Many of my memories of this first foray into involve Alan Mackay—including staying in a ryokan with him, and watching him gleefully embarrass several of our hosts with “difficult” Japanese characters.

^a*Structural Chemistry* 2002, 13(3/4):231–246.

^bThis contribution is part of a collection titled *Generalized Crystallography* and dedicated to the 75th anniversary of Professor Alan L. Mackay, FRS.

^cDepartment of Physics and Astronomy, University College London, London WC1E 6BT, UK
e-mail: j.finney@ucl.ac.uk

Reconsidered in retrospect, the discussion with Dave Harker highlights the progress over 30 years in tackling one of the most fascinating aspects—to me—of Generalized Crystallography, to which both Alan and myself were exposed when working with J. D. Bernal in the Birkbeck Crystallography Department in the 1960s. I was just beginning my scientific career in the study of liquids. And liquids, of course, cannot be treated by standard structural techniques as their molecules do not helpfully line up on a lattice as they do in crystals. Hence, the whole standard crystallographic technique armory is not available for the structural study of liquids. Apparently, the best we could do with diffraction techniques was to obtain the radial distribution function—the distribution of pair distances in the liquid. Interestingly, this function has a strong formal relationship with the Patterson function then used extensively by crystallographers in the solution of the phase problem. Harker had worked with Patterson and was, of course, responsible for the Harker

section of the Patterson function that was particularly useful. Talking with Harker, I took the opportunity to discuss the problems of liquid structure determination in the hope that his insight into the development and use of the Patterson function would suggest how we might move forward to try to obtain data on liquid structures that could be comparable in detail to that obtainable on crystals.

In the absence of a lattice, this seemed to me a tall order. Although having no clear ideas at the time as to how the problem of the absence of a lattice could be overcome, I argued with Harker that there must be some way of extracting more information than the rdf and, ultimately, obtain really detailed structural information on those relatively complex liquids that are of biological and chemical importance, for example, in obtaining information on the mechanisms of protein association and enzyme-ligand binding in solution. I felt that making significant progress in this direction would be a scientific career well spent. He very patiently listened, although I think he felt I was on a wild goose chase.

Thirty years on, things have changed dramatically. The major developments we have seen in radiation sources, instrumentation, and computational techniques have transformed the situation. I believe we have now achieved what I then saw as a dream. We really can do liquid state crystallography and obtain remarkably detailed structural information in relatively complex liquid structures. We can see how amphiphilic molecules interact with each other as temperature, pressure, and concentration are varied. We can determine the structure of the hydration shell of a nonpolar molecule in aqueous solution and see how it is perturbed by adding salt. And we are, perhaps, on the threshold of understanding the structural basis of crucially important phenomena, such as, the hydrophobic interaction.

In this paper, I try to summarize how we obtain this detailed structural information in the absence of a lattice. In addition to the technical advances mentioned above, leaves have also been taken out of the crystallographer's technique book to enable the detailed information to be obtained in a way that perhaps parallels standard crystallographic refinement techniques. I take as an example aqueous solutions of an amphiphile—tertiary butanol—which has a large nonpolar head group and a polar tail and is a system whose thermodynamics is classically hydrophobic [1]. Not only can we see details of the structures of the associations of these molecules in solution, we can see also how these differ from those predicted by computer simulations and, hence, begin to see how the standard potential functions used in such simulations might be significantly improved. Finally, we can perhaps glimpse the structural origin of the entropic driving force for the hydrophobic interaction.

Introduction

In the absence of a lattice, our description of the structure of a liquid has to be a local one. For a single component liquid, this structure is described by the *radial distribution function* $g(r)$ (the rdf), which quantifies the ratio between the local density of atoms at a distance r from an atom at the origin to the average number density of atoms in the system. A two-dimensional analogy of the function is shown in Fig. 1 and indicates how useful structural information can, in principle, be obtained from this function.

A diffraction experiment uses the known wavelength of the (usually X-ray or neutron) scattering probe to obtain the information on pair-distance distributions contained in the radial distribution function. The scattering experiment measures the intensity scattered as a function of the scattering vector \mathbf{Q} , defined simply as the difference between the scattered and incident wave vectors, with a magnitude $|\mathbf{Q}| = 4\pi \sin \theta / \lambda$, where 2θ is the scattering angle and λ the wavelength of the radiation used. From the scattered intensity, the structurally significant *structure factor* $S(\mathbf{Q})$ can be extracted,

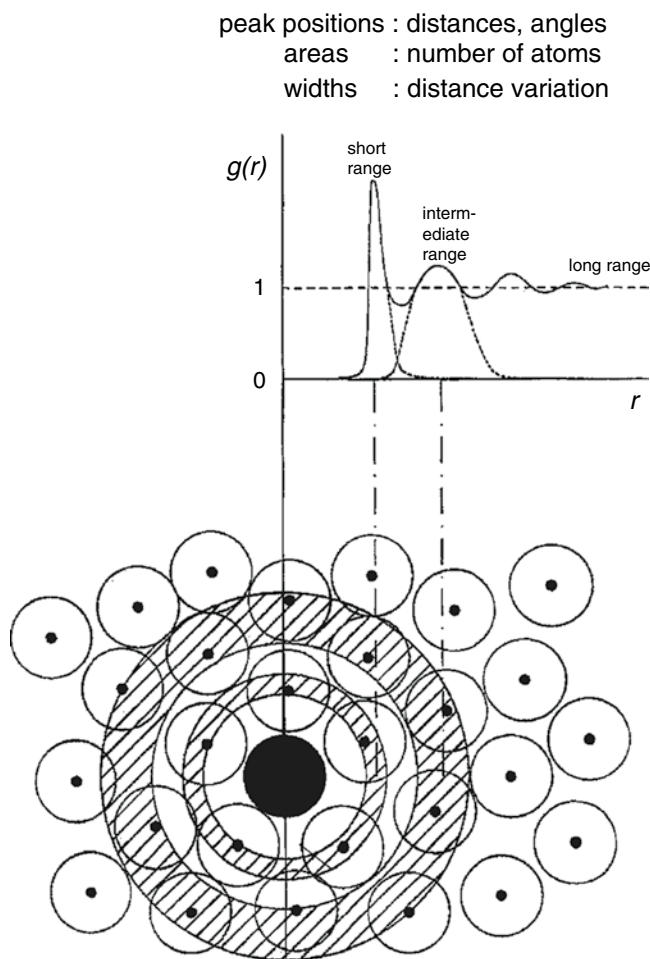


Fig. 1 The radial distribution for a liquid in two dimensions.

which can then be related to the radial distribution function through a Fourier transform relationship:

$$[S(Q)-1] = 4\pi\rho \int_0^\infty r^2 dr [g(r)-1] \frac{\sin Qr}{Qr} \quad (1)$$

Single component liquids are, however, of very limited interest when trying to understand interactions in solution. For any solution, there must be more than one different atomic component, so the above functions must be generalized. Consider a two-component system, with components α and β , with N_α, N_β (where $N_\alpha + N_\beta = N$) atoms of each, and the atomic fractions of each component defined as $c_\alpha = N_\alpha/N, c_\beta = N_\beta/N$. The structure factor can be split into three terms—partial structure factors $S_{\alpha\beta}$ —each relating to different pairs of interacting atoms $\alpha\alpha, \beta\beta$, and $\alpha\beta$. Thus, *partial radial distribution functions* $g_{\alpha\beta}(r)$ are defined similarly to the radial distribution function, but with the added identification of the type of atom both at the local origin and at the distance r . We can then rewrite Eq. (1) as

$$[S_{\alpha\beta}(Q)-1] = 4\pi\rho \int_0^\infty r^2 dr [g_{\alpha\beta}(r)-1] \frac{\sin Qr}{Qr} \quad (2)$$

For a two-component system, we require three partial radial distribution functions to describe the structure of the system. For an n component system, $n(n+1)/2$ partials are needed. Each of these partials will be related to a partial structure factor that is embedded in the measured diffraction data and, taken together, comprise the *total* structure factor

$$F(Q) = \sum_{\alpha,\beta} (2 - \delta_{\alpha\beta}) c_\alpha c_\beta b_\alpha b_\beta [S_{\alpha\beta}(Q)-1] \quad (3)$$

where b_α, b_β are the scattering lengths of the components α and β .

A single diffraction experiment gives us access to only the total radial distribution function $g(r)$, which is a weighted sum of the partial rdfs. This is not very useful information for a solution: we need to try to get closer to determining the partial structure factors. The key to achieving this is given in Eq. (3): the fact that the total structure factor *is* a weighted sum gives us a way forward. One of the weighting factors is the scattering length of each component. If we could perform an experiment on the same system, but somehow change the scattering length, we could perform more than one experiment on chemically similar systems and begin to extract the kind of more detailed information we really want.

That we can do this is made possible by the fact that the neutron is scattered by the nucleus and that many elements have different isotopic forms. If these isotopic forms had different neutron scattering lengths, then we could perform an additional experiment on a system in which we change only

the isotope of one of the components. Referring to Eq. (3), we could then obtain from a second scattering experiment a different structure factor $F'(Q)$ in which, say b_α , is replaced by b'_α . Now performing a third experiment replacing the isotope of the second component, b_β is replaced by b'_β . Alternatively, isotope α could be replaced by either a third isotope α'' (or what is, in effect, the same, a mixture of isotopes α and α'). We now would have three equations with the three partial structure factors as the three unknowns. In principle, these could be solved for and the corresponding partial radial distribution functions obtained.

For relatively simple two-component liquids and glasses, such “full” partial structure factors and, hence, “full” partial radial distribution functions, have been extracted (see, for example, Ref. [2]). For the relatively complex systems in which we are here interested, performing the much larger number of isotopically distinct scattering experiments that would be necessary to extract all the partial rdfs looks a tall order. In our example case of *t*-butanol in water, there are seven chemically distinct atoms in the system (methyl carbon, methyl hydrogen, central carbon, alcohol oxygen, alcohol hydroxyl hydrogen, water oxygen, water hydrogen). We would require $7 \times 8/2 = 28$ isotopically distinct experiments to yield the 28 partial radial distribution functions (or “partials” for short). Setting aside whether an appropriate set of isotopically distinct solutions could be prepared (not all elements have isotopes with significantly different scattering lengths), the requirements for neutron beam time, which is not cheap, would be extremely high. As we shall see below, there are other ways round the problem that enable us to obtain not only all the partials for a system such as this, but even to go *beyond* the partials to obtain even more detailed structural information on relatively complex solutions.

Experimental

For the *t*-butanol–water system, performing 28 isotopically distinct diffraction experiments is not a realistic possibility. The “best” kind of isotopic substitution we can do on this system is to deuterate the hydrogens on (a) the methyl groups of the alcohol head group and (b) the water. In fact, for each substitution, we work with (a) fully deuterated, (b) fully hydrogenated, and (c) a 50:50 hydrogen/deuterium mixture. This gives three values of the neutron-scattering length of the substituted sites and, hence, for each triplet of substitutions, enables us to obtain three sets of (partial) radial distribution functions. We will call these $g_{HH}(r), g_{HX}(r)$, and $g_{XX}(r)$. Here the subscript H refers to a site whose hydrogens are substituted, while the X subscript refers to all the nonsubstituted sites. Thus, for a set of substitutions on the methyl hydrogen sites of *t*-butanol, H refers to all the substituted methyl hydrogen sites, while X refers to all other sites.

The corresponding set of experiments can now be specified as follows.

1. Solvent–solvent distribution functions are probed through substitution on the water hydrogen sites to yield the (partial) radial distribution functions $g_{HH}(r)$, $g_{HX}(r)$, and $g_{XX}(r)$ for the solvent–solvent correlations. The three experiments are made with (a) $(\text{CD}_3)_3\text{COD}$ in D_2O , (b) $(\text{CD}_3)_3\text{COH}$ in H_2O , and (c) 50:50 $(\text{CD}_3)_3\text{COH}/(\text{CD}_3)_3\text{COD}$ in 50:50 $\text{H}_2\text{O}/\text{D}_2\text{O}$.
2. Solute–solvent distribution functions are probed through substitution on the alcohol methyl hydrogen and the water hydrogen sites. In combination with the solute–solute and solvent–solvent distribution functions, this yields the (partial) radial distribution functions $g_{HH}(r)$, $g_{HX}(r)$, and $g_{XX}(r)$ for the solute–solvent correlations. The three experiments are made with (a) $(\text{CH}_3)_3\text{COH}$ in H_2O , (b) $(\text{CD}_3)_3\text{COD}$ in D_2O , and (c) 50:50 $(\text{CD}_3)_3\text{COD}/(\text{CH}_3)_3\text{COH}$ in D_2O .
3. Solute–solute distribution functions are probed through substitution on the alcohol methyl hydrogen and the water hydrogen sites. In combination with the solute–solute and solvent–solvent distribution functions, this yields the (partial) radial distribution functions $g_{HH}(r)$, $g_{HX}(r)$, and $g_{XX}(r)$ for the solute–solvent corrections. The three experiments are made with (a) $(\text{CH}_3)_3\text{COH}$ in H_2O , (b) $(\text{CD}_3)_3\text{COD}$ in D_2O , and (c) 50:50 $(\text{CD}_3)_3\text{COD}/(\text{CH}_3)_3\text{COH}$ in 50:50 $\text{H}_2\text{O}/\text{D}_2\text{O}$.

Thus we appear to have specified nine isotopically distinct experiments. However, we note that samples 1(a), 2(a), and 3(b) are identical. Therefore, we perform only seven experiments to yield the nine (partial) radial distribution functions specified above, namely $g_{HH}(r)$, $g_{HX}(r)$, and $g_{XX}(r)$ for each of the solvent–solvent, solute–solute, and solute–solvent cases considered.

Theory

Despite the fact that we have only performed seven different experiments, we can use these data, together with other known chemical information, to extract the partial radial distribution functions we require. This is not a case of getting something for nothing—it is doing the same sort of thing as a normal crystallographer would do in refining a crystal structure in which the number of independent reflections is—as is usually the case for large molecules—less than the number of structural parameters that need to be refined. That refinement is constrained by what we know of the chemistry of the system (e.g., standard bond lengths and angles, non-overlap of atoms). In effect, we do something similar in the liquid case.

The procedure used—the Empirical Potential Structure Refinement (EPSR) technique developed by Soper [3]—produces model ensembles of molecules that are consistent with the observed scattering. These ensembles can be interrogated to see the detailed geometries of the various intermolecular interactions in a given solution, just as sets of coordinate data from a crystalline refinement can be similarly mined. In summary, the procedure starts by setting up a Monte Carlo simulation of the system using a set of standard potential functions $U_{\alpha\beta}^0(r)$. The Monte Carlo simulation is then run to equilibrium, from which the various partial radial distribution function $g_{\alpha\beta}(r)$ are estimated. This, in general (in all cases we are aware of so far), fails to give adequate agreement with experiment, indicating in passing that the starting potentials are inadequate to reproduce the experimental data. Thus, the initial potential energy function is then modified by adding sets of potentials of mean force between the various different sites. These are derived from comparing the experimental $g_{\alpha\beta}^D(r)$ and the simulated $g_{\alpha\beta}(r)$ radial distribution functions. Thus we obtain a modified potential set:

$$U_{\alpha\beta}^N(r) = U_{\alpha\beta}^0(r) + kT \ln g_{\alpha\beta}(r) / g_{\alpha\beta}^D(r) \quad (4)$$

This new potential is fed into the system and a further Monte Carlo simulation performed. The above procedure is then repeated until $U_{\alpha\beta}^0(r) \approx U_{\alpha\beta}^N(r)$ and $g_{\alpha\beta}(r) \approx g_{\alpha\beta}^D(r)$ for all r and for all pairs of atoms α, β . The resulting experimentally consistent ensembles can then be interrogated to obtain site–site partial radial distributions to examine the detailed structure of the system. In passing, we can note that the way in which the potential function is modified to bring the simulation into agreement with the experiment may be able to guide us in improving the potential functions originally used in the system—or at least give us an idea of what aspects of the starting potentials need modification. Note that what is described above is an earlier implementation of EPSR, which operated in real space, i.e., the comparisons were made between partial radial distribution functions. The normal implementation is now in reciprocal space, i.e., using the partial structure factors $S_{\alpha\beta}(Q)$ to modify the potential.

Results

Partial Radial Distribution Functions

Now that we have these ensembles of molecules, we can extract whatever structural information we want. The 28 partial radial distribution functions for *t*-butanol water at three concentrations (0.06, 0.11, 0.16 mole fractions) [4] are shown in Figs. 2 and 3. The various site labels are: CC, central carbon; C, methyl carbon; M, methyl hydrogen; O, alcohol oxygen; H, alcohol hydrogen; O_w, water oxygen; H_w,

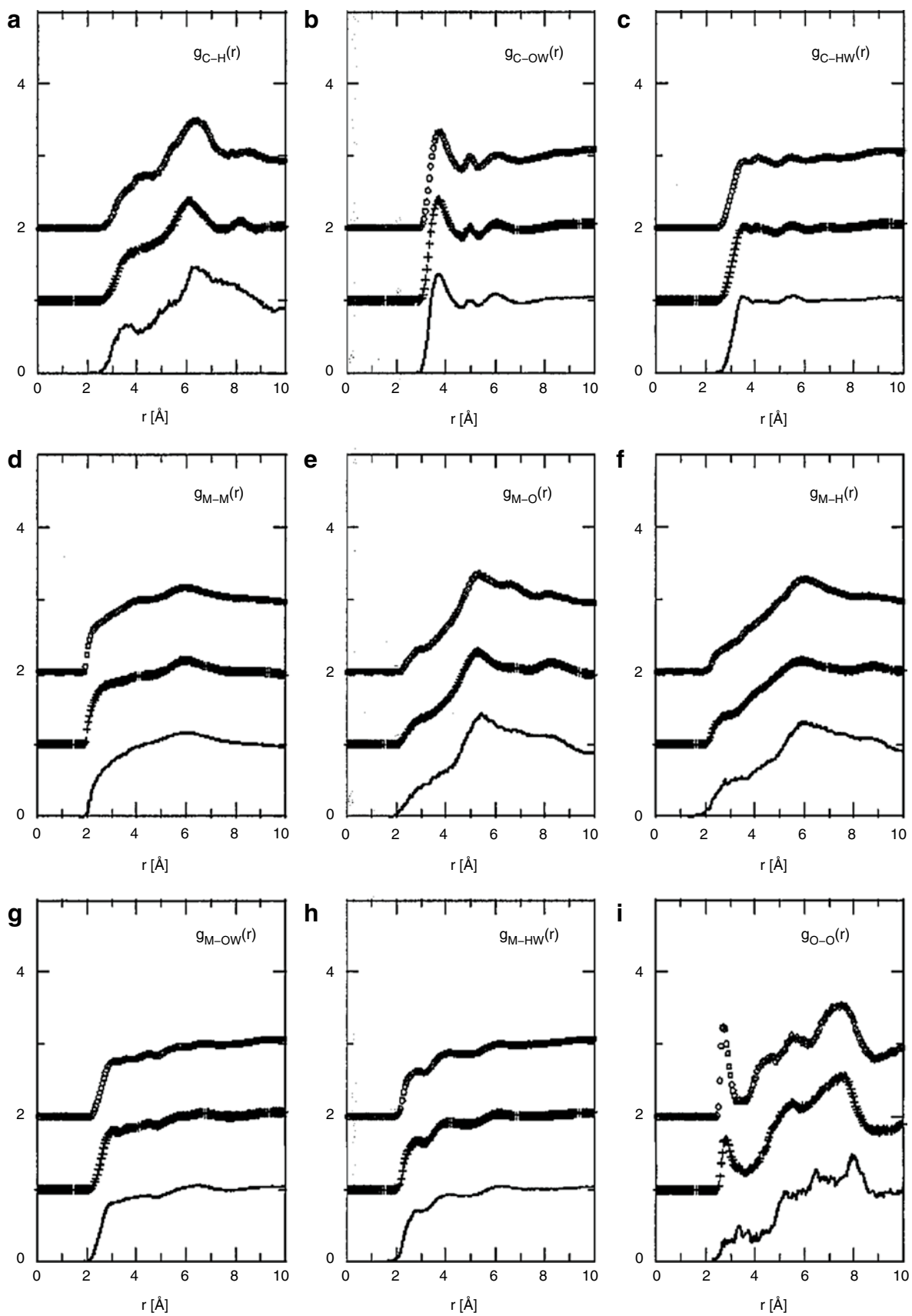


Fig. 2 Intermolecular partial rdfs for 0.16 (O, *top*), 0.11 (+, *middle*), and 0.06 (-, *bottom*) mole fraction *t*-butanol in water.

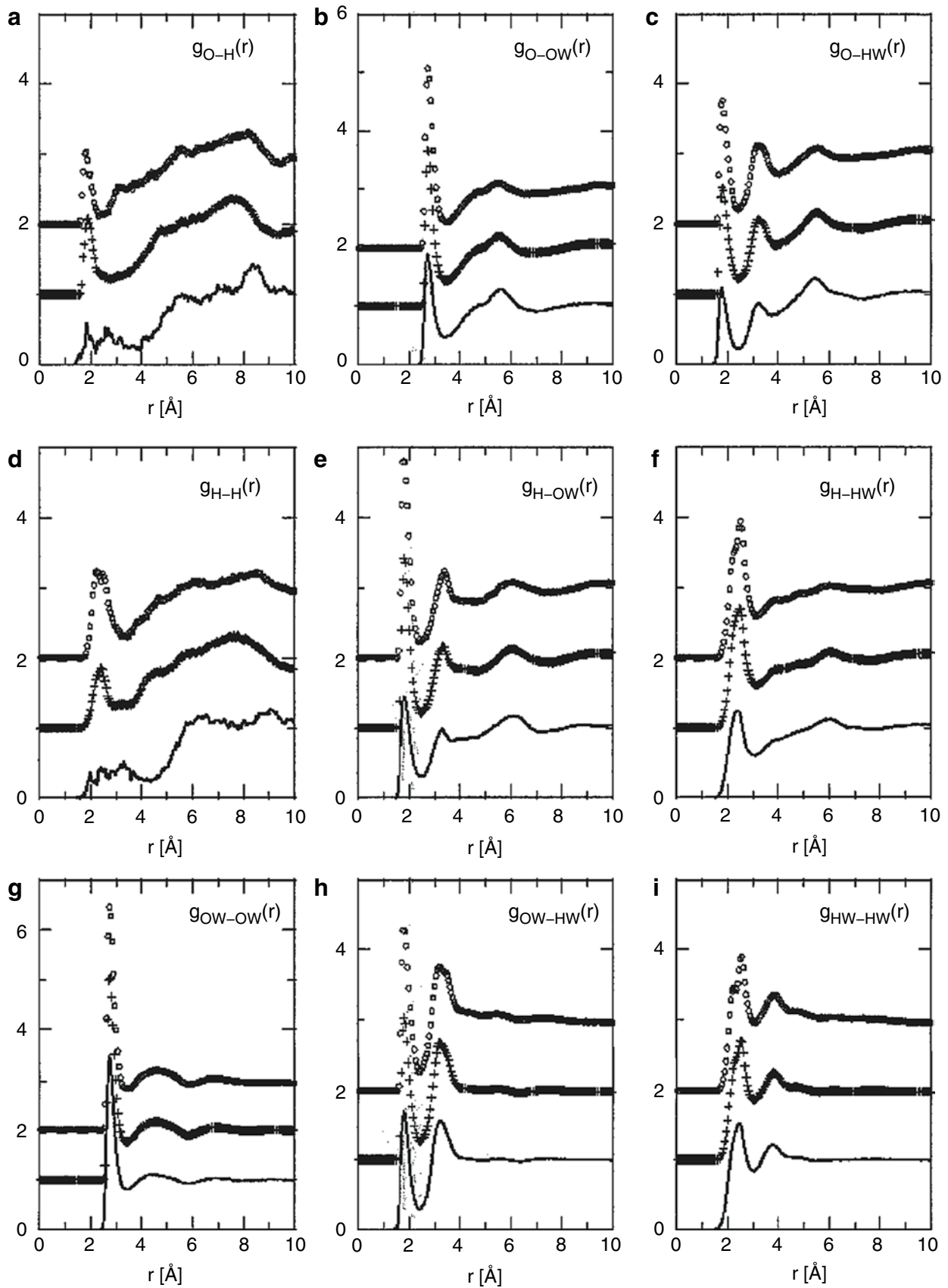


Fig.2 (continued)

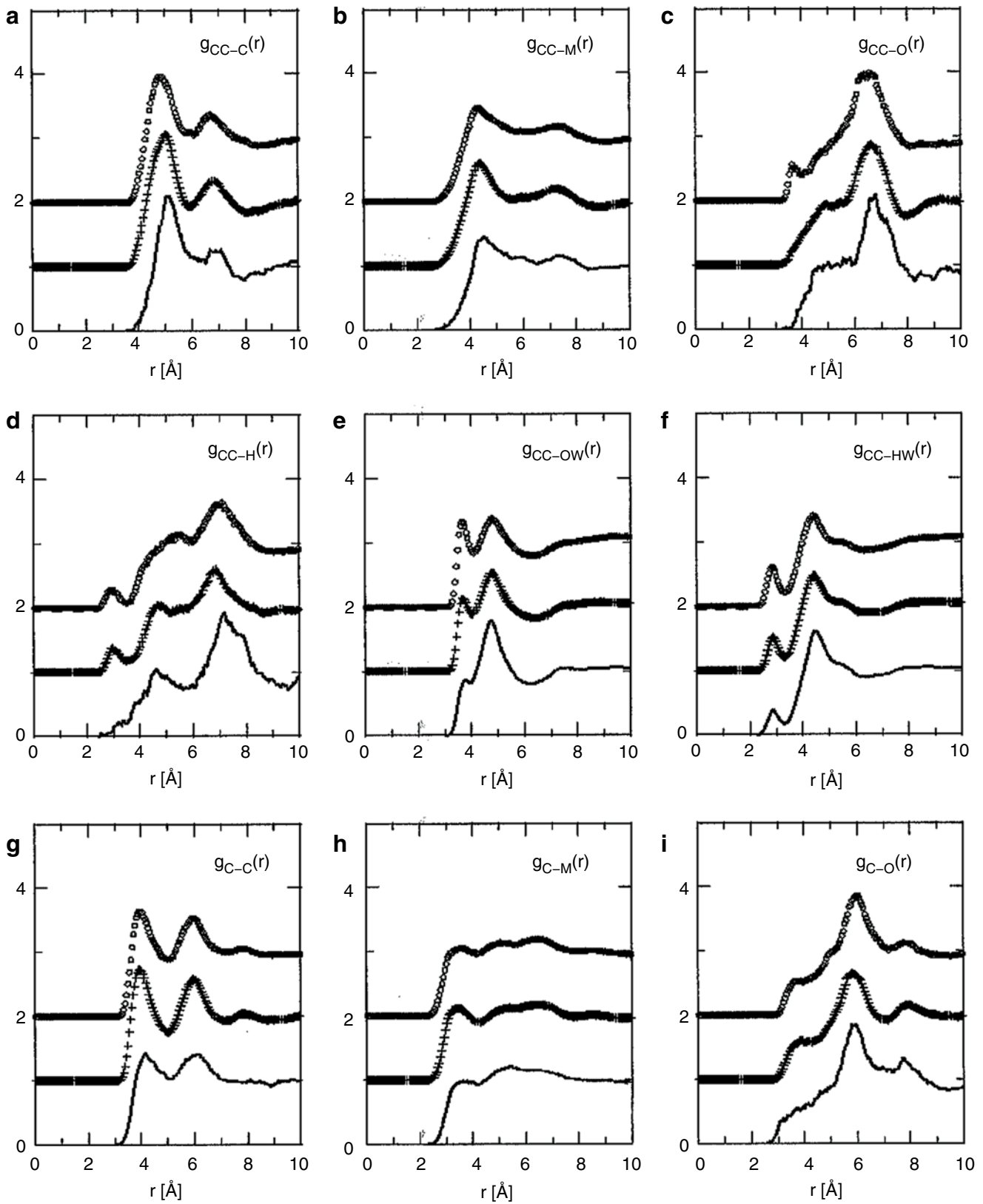


Fig. 2 (continued)

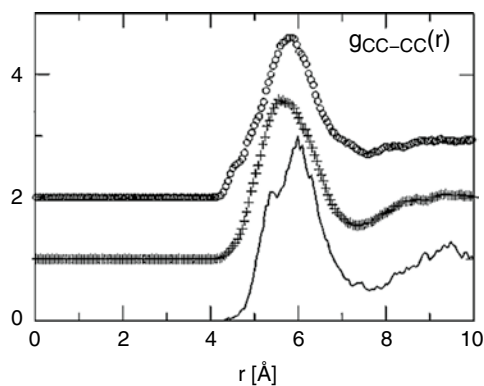


Fig. 3 The molecular centers function for 0.16 (O), 0.11 (+), and 0.06 (–) mole fraction *t*-butanol in water.

water hydrogen. There is clearly a tremendous amount of information in these functions. We now proceed to look at some of this information.

First, looking at the central carbon–central carbon radial distribution function (Fig. 3), we see for all three concentrations a broad peak centered between 5.5 and 6.0 Å. The peak position tells us immediately that even at the lowest concentration there is direct contact between neighboring *t*-butanol molecules: the molecules are not—as suggested in earlier models of the hydrophobic interaction [5]—separated by an intervening solvent layer. The area under this peak tells us how many molecules, on average, surround a central solute molecule. This and other coordination numbers are given in Table 1: it rises from 2.8 ± 0.6 at 0.06 mole fraction, through 4.4 ± 0.6 at the intermediate concentration, to 5.8 ± 0.6 at 0.16 mole fraction. The fact that significant solute association is found at the lowest concentration was perhaps unexpected, suggesting some tendency to microscopic phase separation.

The CC- O_w partial in Fig. 3 gives us information on how the water molecules are arranged around the solute molecule from the point of view of the *t*-butanol’s central carbon. We note two peaks at around 3.7 and 4.8 Å, respectively. The shorter distance is consistent with a water oxygen hydrogen bonding to the hydroxyl group of the alcohol, while the broader peak is at a distance consistent with waters surrounding the nonpolar head group, i.e., those in the hydration shell of the nonpolar head. Integration under this latter peak tells us (see Table 1) that the number of water molecules in this hydration shell falls from about 21 at the lowest concentration to about 13 at the highest—water molecules are, as expected, displaced from the hydration shell as the solute–solute coordination increases with concentration in the manner described above.

Integration under the first, hydrogen-bonding peak gives an essentially concentration invariant figure of 2. This tells us that each alcohol molecule hydrogen bonds to two water

Table 1 Selected coordination numbers for aqueous *t*-butanol solutions

Correlation	R_{\min} (Å)	R_{\max} (Å)	0.06 Mole fraction	0.11 Mole fraction	0.16 Mole fraction
CC-CC	4.0	7.0	2.8 ± 0.6	4.4 ± 0.6	5.8 ± 0.6
CC- O_w	3.0	4.0	2.1 ± 0.4	2.2 ± 0.3	2.0 ± 0.2
	4.0	6.2	21.0 ± 1.2	16.4 ± 0.9	12.8 ± 0.6
O- O_w	2.3	3.3	2.2 ± 0.4	2.3 ± 0.3	2.1 ± 0.3
O- H_w	1.4	2.5	1.2 ± 0.3	1.3 ± 0.2	1.2 ± 0.2
H- O_w	1.4	2.5	0.8 ± 0.2	0.9 ± 0.2	0.9 ± 0.2
O-O	2.4	3.25	–	0.1 ± 0.1	0.2 ± 0.1
O-H	2.5	3.5	–	0.1 ± 0.1	0.1 ± 0.1
H-H	1.6	3.0	–	0.2 ± 0.1	0.3 ± 0.1

molecules through the polar tail. We can compare this with the O- O_w , O- H_w , and H- O_w coordination numbers in Table 1: not only are the values fully consistent with the alcohol hydroxyl group making two hydrogen bonds to water, but they also demonstrate that one of these waters donates a hydrogen to the alcohol oxygen, while the second water accepts a hydrogen from the alcohol.

The bottom three lines of Table 1 give information on the solute–solute hydrogen bonding interactions through the polar tails. For the lowest concentration, as is shown in the appropriate O-O, O-H, and H-H partials in Fig. 2, there is not even a peak visible at the appropriate hydrogen-bonding distance, showing effectively zero intermolecular hydrogen bonding. Even at the two higher concentrations, the peak areas are very small, suggesting that solute–solute hydrogen bonding is extremely limited. These results seem to be telling us that the *t*-butanol alcohol group preferentially bonds to water molecules rather than to other *t*-butanols. Thus, we can conclude quite clearly that the experimental data is not consistent with a significant amount of alcohol–alcohol hydrogen bonding at these concentrations. As the initial Monte Carlo simulations performed as part of the EPSR refinement process suggested a larger degree of intermolecular hydrogen bonding, there are indications here that the model potentials used for this system overemphasize the strength of the intermolecular hydrogen bonding interaction at the expense of the water–alcohol interaction. As the stability of a protein is a delicate balance between protein–protein and protein–water hydrogen bonds, this result indicates that work needs to be done on these hydrogen bonding potentials if they are to be seriously used in modeling biomolecular interactions.

Spatial Distribution Functions

Although all the information we should need to understand structure is, in principle, included in the full set of partial radial distribution functions, these functions are still averages

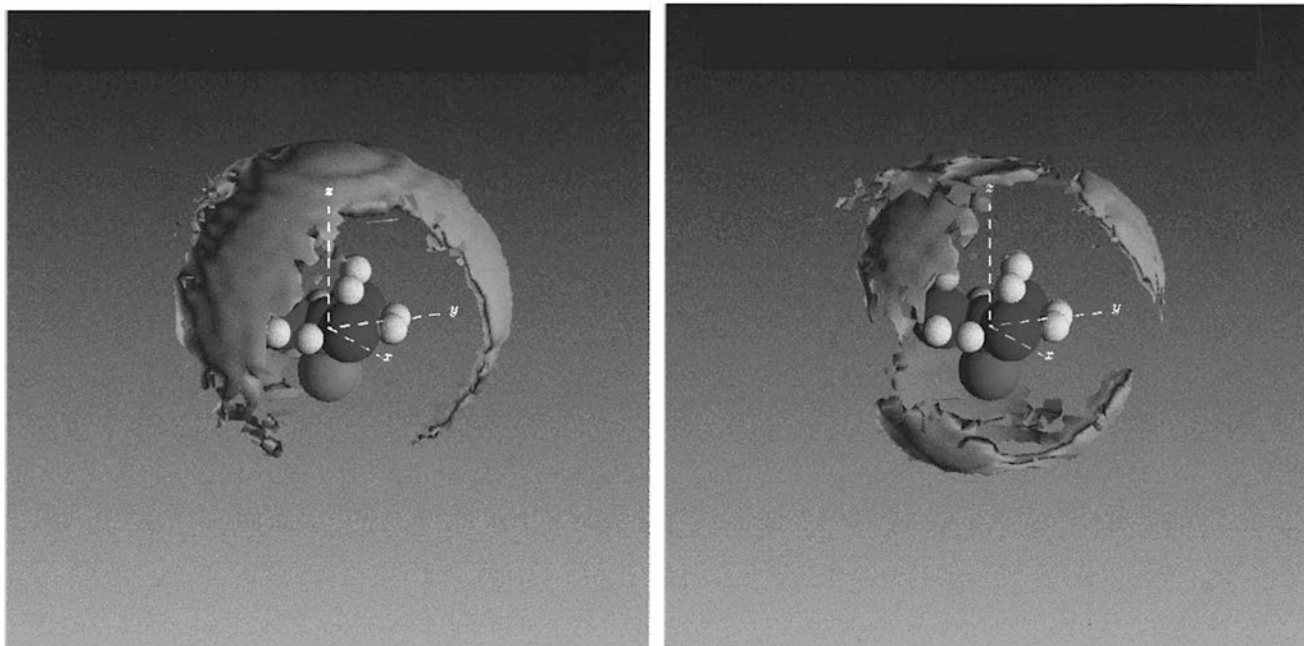


Fig. 4 The spatial density function of 0.06 mole fraction aqueous *t*-butanol at 25 °C (*left*) and 65 °C (*right*).

over orientations of both the central molecule at the origin of the function, and the neighboring molecule at a distance r . As we have complete data on the coordinates of all the sites on all the molecules in the model ensembles we are interrogating, we can construct further functions that enable us to look at the relative orientations of molecules.

First, we imagine ourselves sitting on a molecular site and looking at the various neighbors at different distances. We begin to note that there is a preference for a certain atom not only at a given radial distance corresponding to a peak in the partial radial distribution function, but also that peak tends to be in a particular direction. We thus can construct a further pair-distribution function that tells us the orientational preference of a particular site as a function of distance from the central site. This function, the *spatial density function*, gives us orientational information on the nature of the interactions between the various molecules in our solution. For example, referring again to Fig. 3, we have a peak that tells us that there are neighboring *t*-butanols (specified by the positions of their central carbon atoms) at a distance centered between 5.5 and 6 Å. We do not know from this orientational average the directions relative to the geometry of the central *t*-butanol in which these molecules are to be found. Do they cluster around the head groups or are there some contacts through the tails? What we really want to do is to assign a location on a surrounding sphere—in essence a latitude and a longitude—for each of the neighboring molecules included under the peak shown in Fig. 3. This is what the spatial density function $g\alpha\beta(r, \Omega)$ does [6]. It represents

a three-dimensional (3-D) map of the density of β sites as a function of radial distance, r , and orientation Q about a central site α .

The spatial density function of *t*-butanol molecules around a central *t*-butanol molecule, defined with reference to the central carbon sites of each, is shown in Fig. 4 (*left panel*) for the 0.06 mole fraction concentration at 25 °C. The continuous density above the central *t*-butanol molecule and the three sections curving down show the angular distributions around the central molecule of the molecules that are under the broad peak in Fig. 3. We thus have more information than in that orientationally averaged partial radial distribution function: in addition to the preference for the “capping” position, there is greater density of neighboring molecules in three orientations separated by about 120°. Interestingly, these directions correspond to the positions between the three methyl groups on the head group of the central molecule, which enables a neighboring molecule to approach slightly closer than if it were approaching along a direction in which a methyl group was located. In fact, if we consider the concentration dependence of the peak position in Fig. 3, we see a slight outward move as the concentration is reduced. This outward move—which separates out a low- r shoulder at the lowest concentration—appears to be related to an apparent compressive driving force at the higher concentrations, which “pushes in” the neighboring *t*-butanol molecule. To anticipate the following section, the methyl groups on neighboring molecules seem to be engaging with each other like two cogwheels.

Orientalional Distribution Functions

In the discussion of the previous paragraph, we are able to comment only on the positional distributions of the β sites around the central α site. We do not know how the neighboring molecule is oriented with respect to the central one. This information would obviously be useful in understanding the chemical nature of the interactions between the molecules in the various orientations around the central molecule: for example, are the nonpolar groups of the neighboring *t*-butanol in direct contact with the nonpolar groups of the central *t*-butanol, or are other orientations involved? Moreover, if we increase the temperature to 65 °C, how would an enhancement of the hydrophobic interactions driving the *t*-butanol association show up? Does the nature of the solute–solute interactions change? The spatial density function for 65 °C in the right-hand panel of Fig. 4 already shows that there is a shift in solute–solute interaction away from the head group to the polar tail: what is the chemistry of these interactions in the case?

Just as when moving from the partial radial distribution function to the spatial density function, we added orientational information to the molecules under the one-dimensional peak of the partial, so here we now need to add orientational information to the molecules in the lobes of the spatial density functions such as those in Fig. 4. As we have the detailed coordinate data in the ensembles generated by the EPSR refinement, this information can be accessed. The orientational distribution function reduced disorder in the head-group to head-group interaction, an interesting consequence of the increased temperature. Second, there is a significant feature around 12 o'clock, which is indicative of a polar to nonpolar interaction, i.e., the head group of one molecule is interacting with the polar tail of a neighbor. A further point to note is that this lobe is shifted out to a distance of about 8 Å, so this appears not to be a direct contact between the hydroxyl group and the methyl head groups. Rather there is space for an intervening water molecule. We can, therefore, perhaps consider this as a neighboring *t*-butanol hydrogen bonding to one of the water molecules in the nonpolar hydration shell of the central molecule.

Finally, the bottom panel shows the orientational plot in the direction of the hydroxyl group of the central *t*-butanol molecule. The most obvious change to note is the growth of intensity around the 6 o'clock position at 65 °C. This lobe corresponds to *t*-butanol to *t*-butanol contacts through direct hydrogen bonding, an occurrence that, we have already noted, is much rarer at 25 °C. A second point to note is a splitting of the lobe around 12 o'clock that denotes polar to nonpolar contacts; at 65 °C, the inner lobe comes in closer, the rise in temperature appearing to make easier a close, apparently chemically unfavorable, contact.

Raising the temperature in this system thus seems to do several interesting things. First, the straight nonpolar contacts seem to become more ordered. Second, there is an

increase in the degree of direct hydrogen bonding between *t*-butanol molecules. Third, the temperature increase seems to make classically unfavorable interactions between polar and nonpolar groups easier. The “strengthening” of the nonpolar contacts might be expected from classical ideas of the temperature dependence of the hydrophobic effect. The other two effects seem less easy to relate directly to consequences of a hydrophobic driving force. $g(\mathbf{r}, \omega_1, \omega_2)$ relates the relative position vector \mathbf{r} of two molecules 1 and 2, with their orientations ω_1, ω_2 in the laboratory reference frame. As this is strictly a function of nine variables (three positional, six angular), there are major computer memory limitations to storing this function, even when the variables are reduced by symmetry. An efficient way of dealing with this problem is to expand the orientational correlation function as a spherical harmonic expansion. The correlation function can then be stored as a series of spherical harmonic coefficients [7] $h(l_1 l_2 l; n_1 n_2; r)$, where l_1, l_2, n_1, n_2 relate to the orientational coordinates ω_1 and ω_2 , while l relates to the spatial distribution of molecule 2 about molecule 1.

Figure 5 shows the definition of the coordinate system used for both the orientational position of the neighbor with respect to the geometry of the central molecule and for the orientation of the neighboring molecule with respect to the first. The two angles θ_1 and ϕ_1 describe the direction of the neighbor (the central molecule being rotationally averaged about the Z-axis shown) and the three Euler angles $\phi_m, \theta_m,$ and χ_m define the orientation of the neighboring molecule. As a first example of its use, we show in Fig. 6 correlation maps of θ_m for the 0.06 mole fraction concentration at 25 °C for three values of θ_1 . The $\theta_1=0^\circ$ plot shows high density for θ_m around 180°, i.e., the dominant interaction in this direction is with the two *t*-butanol molecules making contact through their nonpolar head groups. This is classically what we would expect for a system subject to a hydrophobic driving

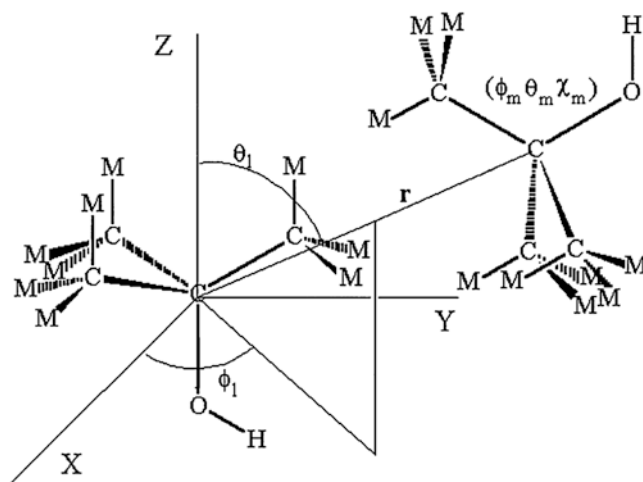


Fig. 5 Schematic diagram of the coordinate system used to define the orientational correlations between two *t*-butanol molecules.

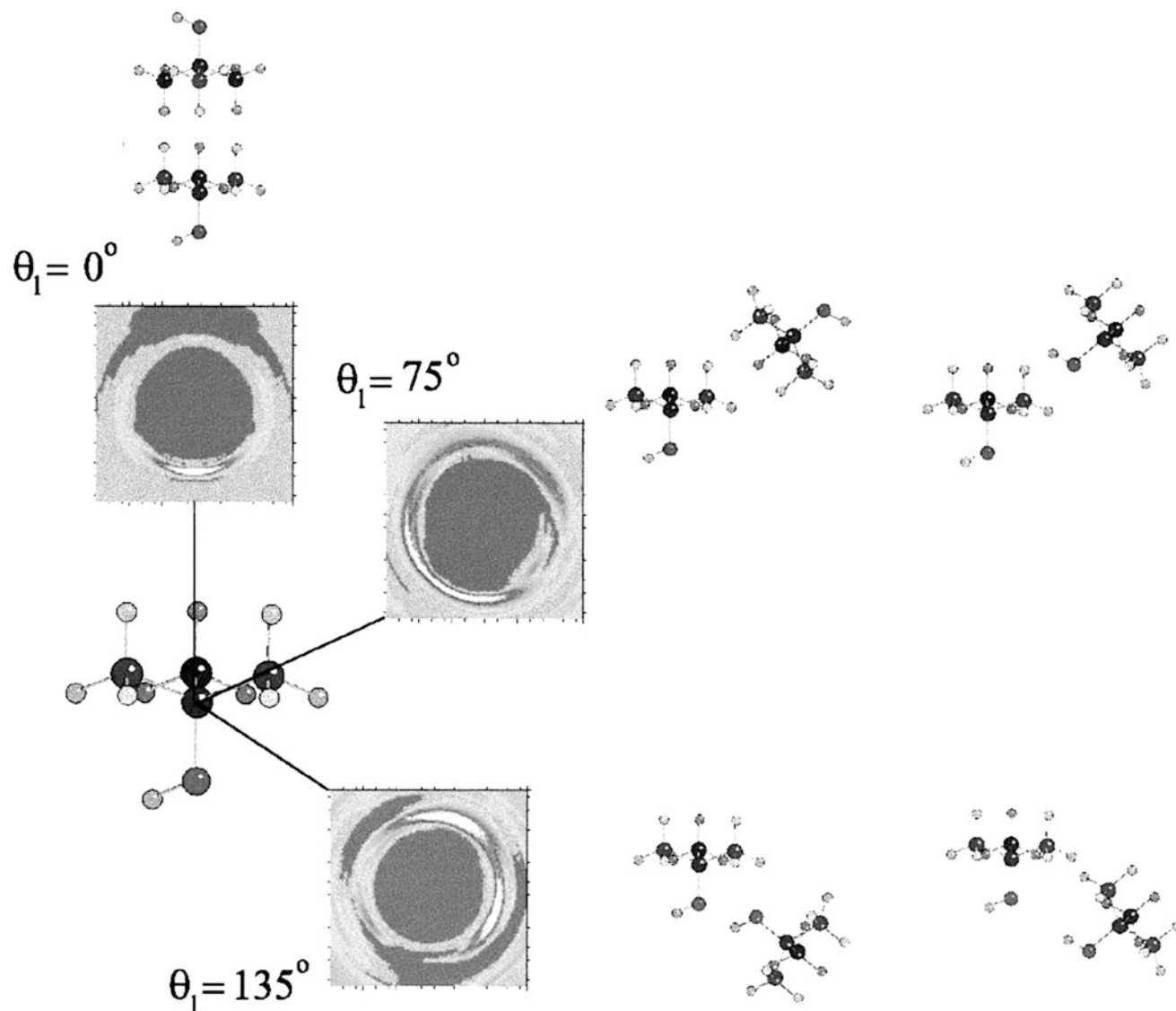
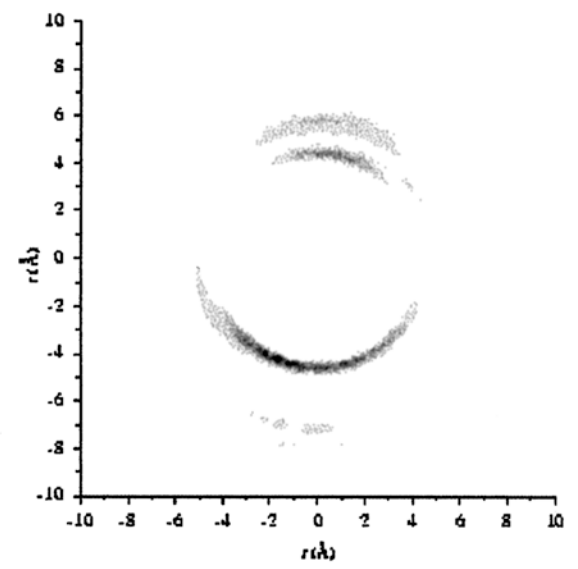
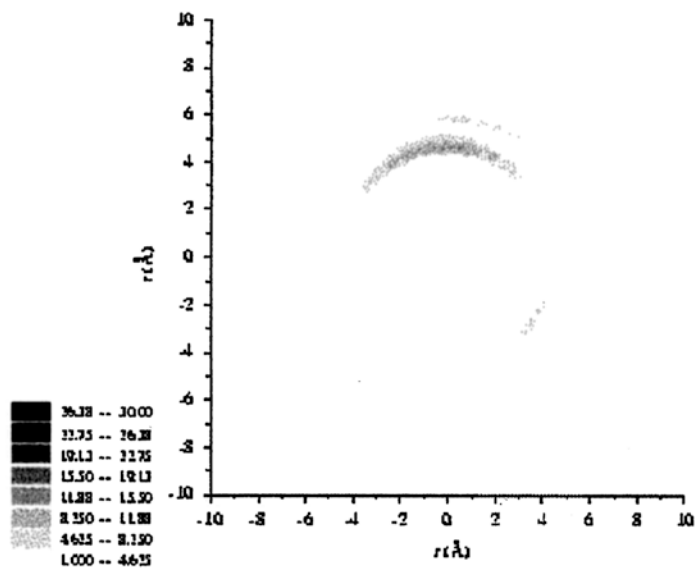
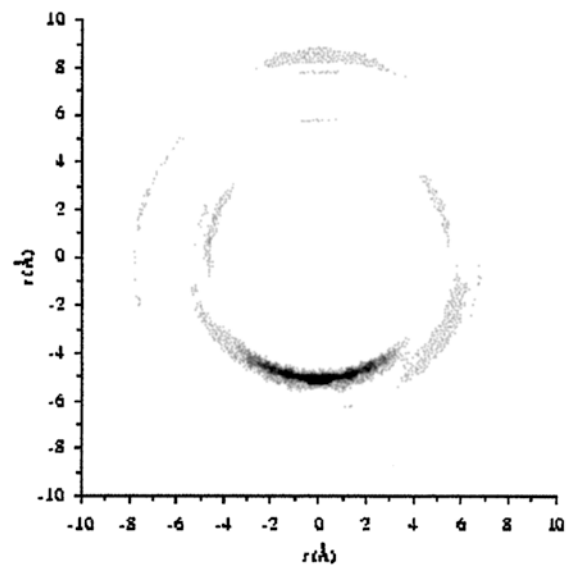
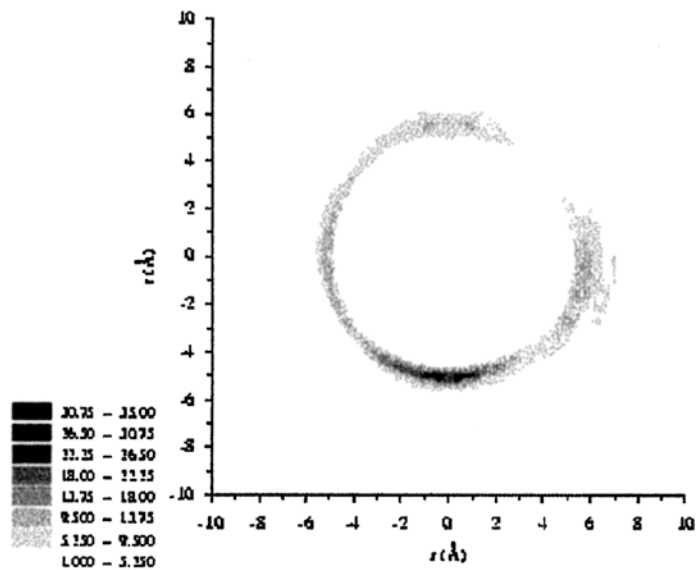
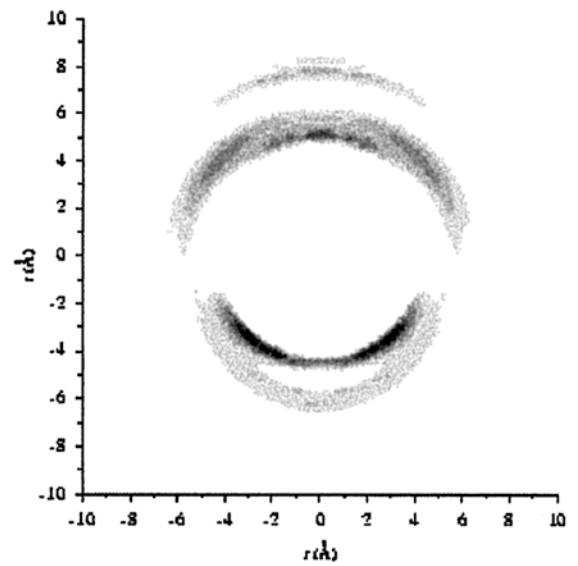
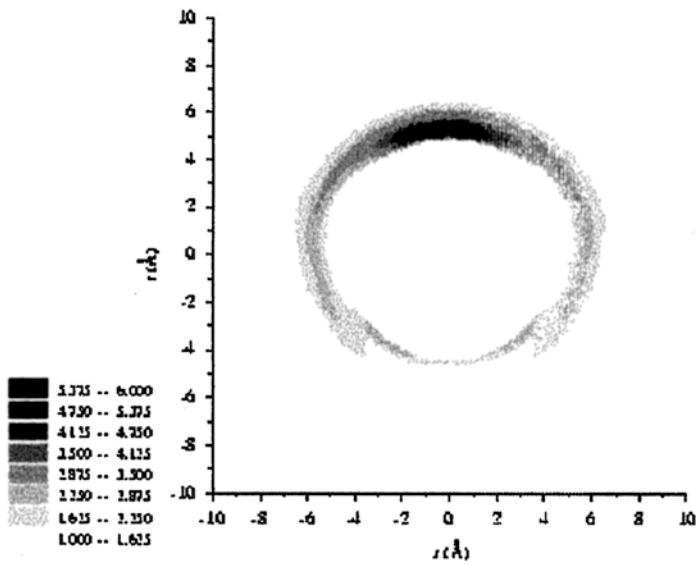


Fig. 6 Sections through the orientational correlation function (correlation maps) of two *t*-butanol molecules in 0.06 mole fraction aqueous solution at room temperature. The molecular arrangements shown are the dominant ones corresponding to the highest contour levels in the plots. For each plot shown, the angular variable plotted is θ_m .

force. Moving to the $\theta_1 = 75^\circ$ plot, we see the highest intensity between the 6 o'clock and 9 o'clock positions, again consistent with a dominant nonpolar to nonpolar contact as shown in the figure. Here, however, there is some—though small—intensity around the 1–2 o'clock position, indicating some contribution from the polar to nonpolar contact, again indicated by the molecules in the figure. Finally, moving to the $\theta_1 = 135^\circ$ plot, we note bright lobes in the two molecular orientations shown—clearly these correspond to intermolecular hydrogen bonding. We have already noted, however, that these contacts occur only rarely in this solution. What the orientational correlation functions tell us is that, even if the intermolecular contacts in this region of the central molecule are rare, when they do occur, they are of a typically hydrogen bonding nature.

A second example of the information in the orientational correlation function refers back to the spatial density functions shown for the 0.06 mole fraction concentration at the two temperatures of 25 and 65 °C (Fig. 4). What can we learn about the change in the nature of these intermolecular interactions as the temperature is raised in such a way as to (we expect) enhance the hydrophobic driving force [8]? We have already seen that at 25 °C the lobes in the spatial density function relate to nonpolar to nonpolar contacts through the head groups of neighboring molecules; at the higher temperature (Fig. 4, right-hand panel), these lobes shrink, and there is significant growth in contacts at the polar end of the molecule. What is the chemical nature of these new contacts?

We can answer all these questions by referring to the plots in Fig. 7. The left-hand panels relate to the 25 °C solution,



while the right-hand column relates to 65 °C. The top two panels are plots through the two spatial density functions of Fig. 4 and show that the dominant head-to-head contact indicated by the bright lobe around 12 o'clock for 25 °C is reduced in favor of increasing intensity around 5 and 7 o'clock in the 65 °C solution. The second panel shows orientational correlation function plots of θ_m . The dominant intensity around 180° again indicates a dominance of nonpolar contacts through the head groups of contacting molecules. Moving right to the 65 °C panel, we find two interesting points. First, the 6 o'clock feature is a little “tighter” indicating.

Structural Basis of the Hydrophobic Interaction

The above examples show that we can now obtain detailed geometrical information on intermolecular interactions in solution. We can see how solute molecules interact and how these interactions might change under changed external conditions. We conclude by addressing some structural issues relating to the perturbation of the water structure in our *t*-butanol solution. These impinge directly on the nature of the solvent ordering that is thought to be the driving force for the hydrophobic interaction.

First, we note the classical view of the nature of the hydration shell of a nonpolar group and how this is thought to relate to the entropic driving force of the hydrophobic interaction. The conventional wisdom on this comes from the classic paper of Frank and Evans in 1945 [9]. In brief, this paper argues that the water in the hydration shell is somehow “more ordered” than that in the bulk. Bringing together two nonpolar surfaces in solution will then displace some of this “ordered” water into the bulk, with a consequent gain in entropy to the system. Hence, the source of the entropic driving force to hydrophobic association.

The detailed structural nature of this solvent ordering has been the subject of much discussion. The prevailing view seeming to prefer a picture something like the clathrate cage structure found in gas hydrate crystals [10]. Earlier work using the kind of techniques discussed here has partly verified this picture, although it also demonstrates that the simple cage model specifies too much order. Rather, we should think of the hydration shell as a *disordered* structure that has some relationship to the clathrate cage; we are, after all, dealing with a

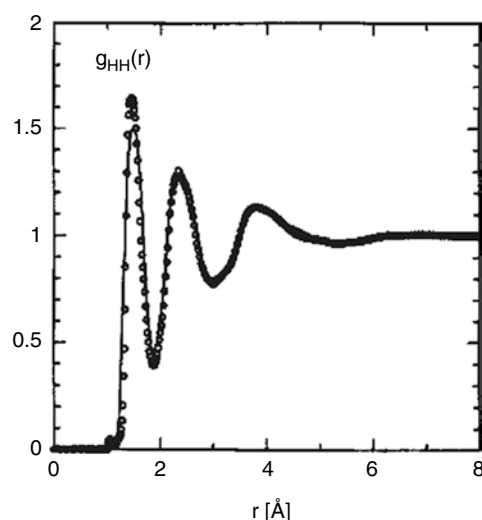


Fig. 8 The hydrogen–hydrogen partial rdf for a 1:9 mole fraction solution of methanol in water (*line*) compared to the same function for pure water.

liquid in which such long-lived ordered structures are unlikely to exist. A more surprising conclusion of that work [10] concerned the structure of the water in the hydration shell. Figure 8 compares the hydrogen–hydrogen partial rdf for bulk water with that of the water in a 1:9 aqueous methanol solution [11]. The conclusion we can draw from the figure is both obvious and unexpected: there is no detectable difference between the two functions. As far as the water is concerned, when its structure is measured through the hydrogen–hydrogen radial distribution function, a sensitive structural measure that includes all-important orientational information, there is no observable perturbation of the structure in the direction of increased—or decreased—order. Similar conclusions have been drawn since in studies of a number of different systems containing nonpolar head groups [4, 8, 12]. Thus, these structural results do not support the traditional view that the water in the first hydration shell of a nonpolar group is structurally more ordered than in the bulk.

The clue to the actual source of the entropic driving force may be in Fig. 9, which shows the spatial density functions of the first (top panels) and second (bottom panels) neighbor shells of water in (from left to right) bulk water at 25 °C, 0.06 mole fraction *t*-butanol at 25 °C, and 0.06 mole fraction *t*-butanol at 65 °C. We emphasize that we are looking here at the neighbourhood of the water, not of the nonpolar group itself.

Fig. 7 The intermolecular orientation correlation maps corresponding to the solute–solute interactions as a function of molecular centers separation in 0.06 mole fraction *t*-butanol–water solution. The *left-hand column* represents the correlations prevalent at 25 °C, while the *right hand column* those at 65 °C. The *top panels* represent the probability of finding a neighboring alcohol molecule about an arbitrarily chosen central alcohol molecule as a function of the angle θ_1 , i.e., the result is averaged over all orientations of the neighboring alcohol molecule. The *middle panels* illustrate the dominant orientation as a function of θ_m of a neighboring alcohol molecule in the direction above the methyl groups of the central molecule, i.e., where $\theta_1 = \phi_1 = \phi_m = \chi_m = 0^\circ$. The *lower panels* show the relative orientations of alcohol molecules θ_m , in the direction of the hydroxyl group interactions where $\theta_1 = 135^\circ$, $\phi_1 = 60^\circ$, and $\phi_m = \chi_m = 0^\circ$.

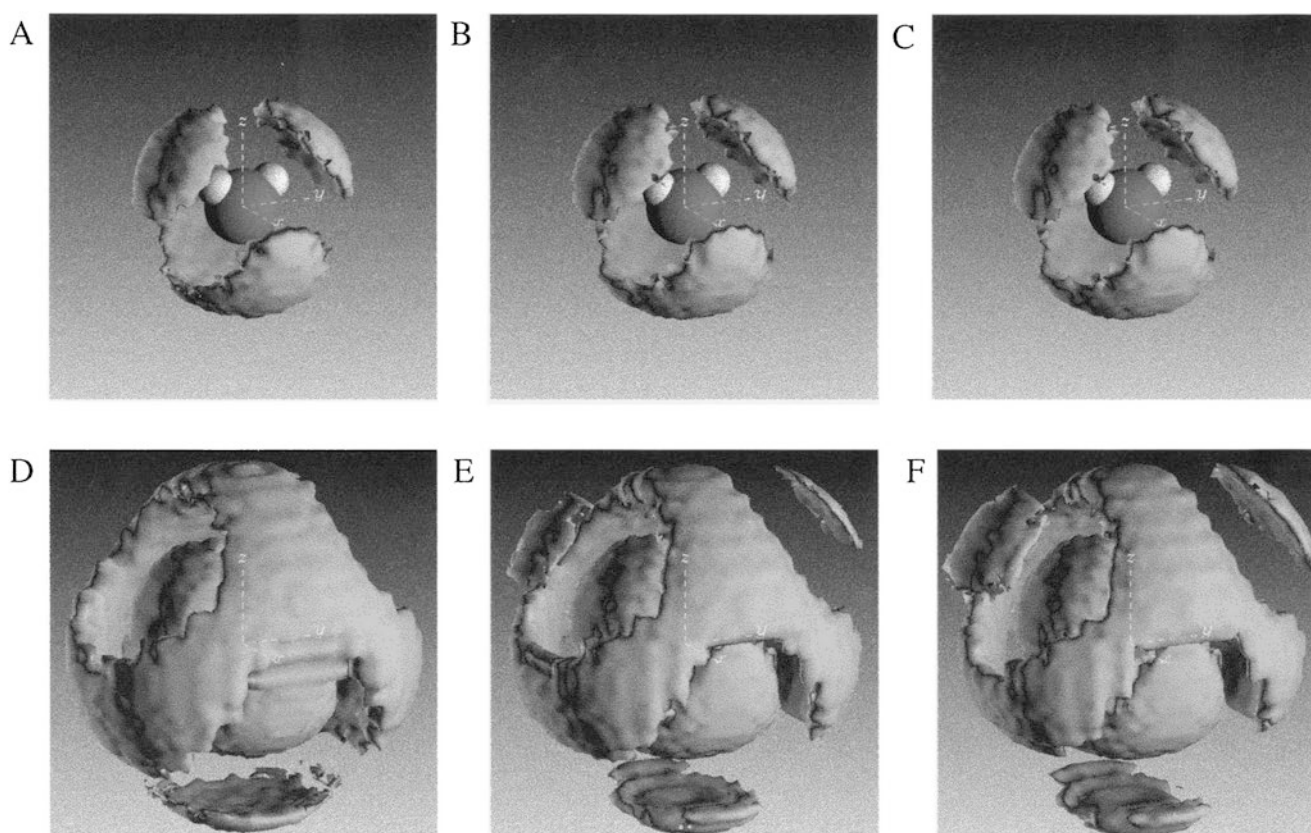


Fig. 9 The spatial density functions of water molecules around a central water molecule for (*top row*) first neighbor shell water molecules and (*bottom row*) first and second shell waters. In turn, the panels show the distributions for (*left*) pure water at 25 °C, (*center*) 0.06 mole fraction *t*-butanol–water solution at 25 °C and (*right*) 0.06 mole fraction *t*-butanol–water solution at 65 °C. The first shell of neighbors (*top row*) illustrates the invariance between the three systems of the short-range correlations that are of a predominantly tetrahedral local water coordination. The appearance of the third neighbor shell in the direction of the central molecule O–H bonds for the alcohol–water mixtures (*bottom center and bottom right*) visibly illustrates the solute-induced compression of the solvent molecular density, which appears to be enhanced by increasing temperature.

Focusing first on the first shells in the top panels, we see the standard spatial density function for the first neighbor shell in water. The two lobes above the central water relate to neighboring waters that accept hydrogens from the central water, while the broad lobe beneath indicates waters donating hydrogens to the lone pair region of the central water. This first neighbor shell appears unchanged in both *t*-butanol solutions.

When we look at the bottom panels, where the contour levels have been lowered to show up the second neighbor shell, we see some clear differences. Although the basic geometry of the second shell is similar in all four cases, there are very significant differences: the second shell is progressively “pulled in” toward the central molecule as we move to the 25 °C solution, and then to the 65 °C case. Consequently, lobes of spatial density corresponding to the third shell begin to appear in the two solution plots.

This “drawing in” of the second shell is accompanied by a significant ordering of the second shell. This is, indeed, visible in the progressive narrowing of the band in the center

of the bottom panels of Fig. 9 as we move from left (bulk water) to right (65 °C 0.06 *t*-butanol). It is perhaps more obvious, however, in the 2-D spatial density maps of Fig. 10, which are obtained by rotationally averaging $g_{\infty}(r, \Omega)$ about the z axis (shown in Fig. 9). The second shell structural tightening induced by the presence of the alcohol molecule can be seen in the breaks of continuity in the 10 and 2 o’clock directions as we move from bulk water to the two solutions. Also visible is the small, but significant, inward shift of the lobes, which is particularly obvious in the 12 o’clock direction. Taking sections through the appropriate lobes allows us to quantify both these inward shifts and the narrowing of the shell widths. In terms of the water molecule’s second neighbor shell, we have clear evidence that the presence of *t*-butanol increases structural order in the system, an ordering that is enhanced with the rise in temperature to 65 °C. Perhaps it is here—in the *water’s* second shell rather than the *alcohol’s* first hydration shell—that the entropic driving force for the hydrophobic interaction is to be found.

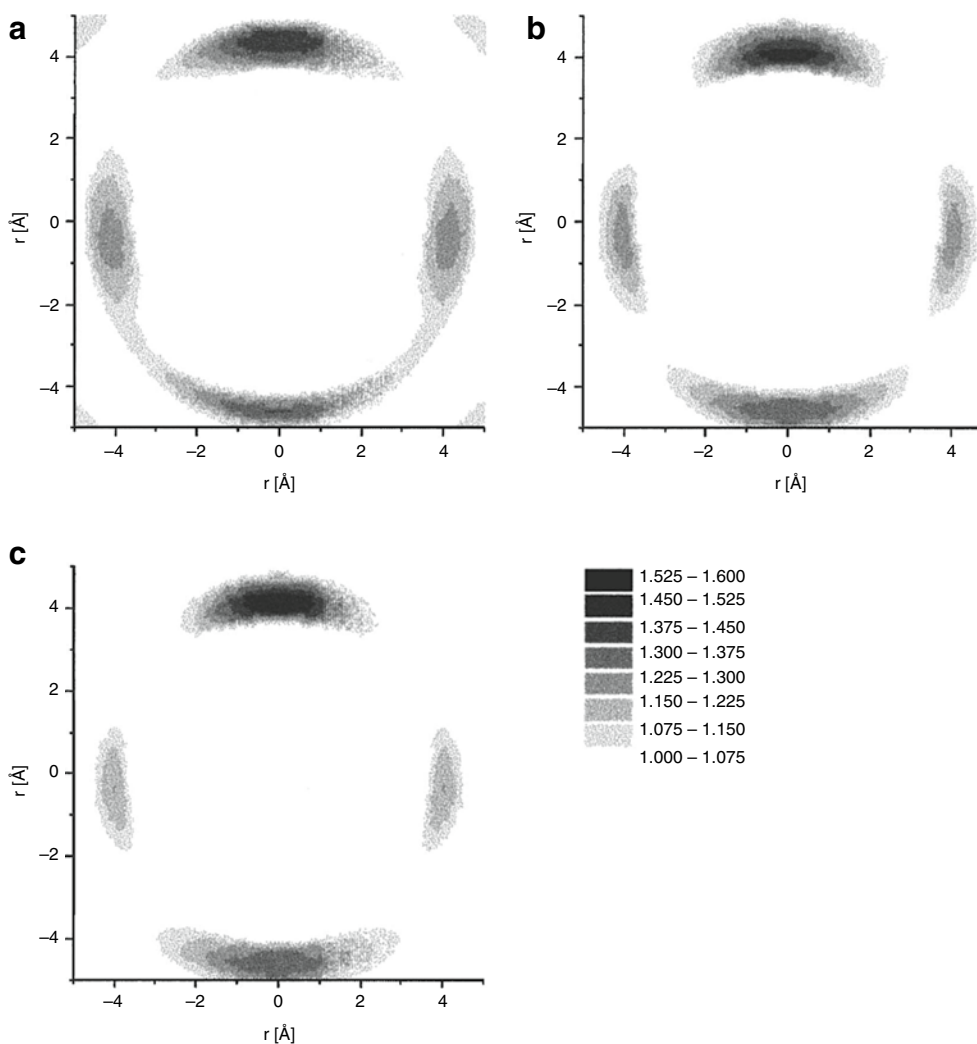


Fig. 10 The upper panels are 2-D maps of the second water shell spatial density functions of Fig. 9. These are obtained by rotationally averaging $g_{oo}(r, \Omega)$ about the z axis, i.e., about the ϕ coordinate. The map in (a) corresponds to pure water at 25 °C, in (b) to 0.06 mole fraction *t*-butanol–water solution at 25 °C, and in (c) to the alcohol solution at 65 °C. The disruption of the second shell water in the presence of the alcohol molecule can be seen in the 2 o’clock direction: this disruption is also marked by the inward migration of the shell, in particular, along the 12 o’clock direction.

Conclusions

“Normal” crystallography exploits the existence of the lattice to maintain phase relationships between scattered beams of X-ray, neutrons, or electrons, which enable us to determine detailed structures of even highly complex molecules within a crystal. Without the crystal lattice, these phase relationships are lost, and we are left with being able to determine only distributions of pair distances. When we try to generalize crystallography to disordered systems, in general, we have to develop ways of obtaining high-quality structural information in the absence of this lattice “crutch.” If we do want to understand process in solution—which are central to much of chemistry and biology—we need to do structure

determination without a lattice being present. We need to develop “no-lattice crystallography.”

Taking advantage of the past 30 years’ developments in neutron sources and instrumentation, allied to advances in both computer hardware and software, we have been able to make major progress in structure determination in the absence of the lattice. We can perhaps make a comparison with the advances in crystallography since the 1920s, when Kathleen Lonsdale’s solution of the structure of hexamethylbenzene and hexachlorobenzene caused Christopher Ingold to comment that “the calculations must have been terrible ... but one structure like this brings more certainty into organic chemistry than generations of activity by us professionals.” From those early molecular structures, we have advanced—taking

advantage of developments in radiation sources, instrumentation, and computing resources—to determining structures of macromolecular assemblies, such as the ribosome.

In the liquid case, we have similarly made major steps forward. Christopher Ingold's comment on molecular structure is paralleled in a way by the chemist David Feakins' 1993 (personal communication) comment on liquids: "(our understanding of solutions is) in the state that our knowledge of molecules would be if we only had, say, the thermodynamics and kinetics of reaction to guide us." Thirty years ago it was unimaginable to either me or to David Harker on that Air China charter flight that we could determine the structural details of the interaction between two amphiphiles in solution, let alone how that interaction might change with temperature or concentration. Or that we would be able to measure either the hydration shell of a nonpolar group, or how the solvent in an aqueous solution of molecules containing large nonpolar groups. We can now do these things, and more. At the turn of the century, liquid state crystallography—crystallography without a lattice—has perhaps come of age.

Acknowledgments The work described here, and building up to it, has been performed in collaboration with a number of colleagues, Daniel Bowron, Alan Soper, and Jacky Turner having been particularly closely

involved. The early ideas developed owe much to many discussions with John Enderby, who pioneered isotope substitution methods in liquid structure studies. Alan Mackay has remained a stimulating colleague throughout.

References

1. Franks, F.; Desnoyers, J. E. *Water Sci. Rev.* **1985**, *1*, 171.
2. Edwards, F. G.; Enderby, J. E.; Howe, R. A.; Page, D. I. *J. Phys. C* **1975**, *8*, 3483.
3. Soper, A. K. *Chem. Phys.* **1996**, *202*, 295.
4. Bowron, D. T.; Finney, J. L.; Soper, A. K. *J. Phys. Chem. B* **1998**, *102*, 3551.
5. Pratt, L.; Chandler, D. *J. Chem. Phys.* **1977**, *67*, 3683.
6. Svishchev, I. M.; Kusalik, P. G. *J. Chem. Phys.* **1993**, *99*, 3049.
7. Gray, C. G.; Gubbins, K. E. *Theory of Molecular Liquids, Vol. I: Fundamentals*; Oxford University Press: New York, 1984.
8. Bowron, D. T.; Soper, A. K.; Finney, J. L. *J. Chem. Phys.* **2001**, *114*, 6203.
9. Frank, H. S.; Evans, M. W. *J. Chem. Phys.* **1945**, *13*, 507.
10. Davidson, D. W. In *Water A Comprehensive Treatise*; Franks, F., Ed.; Plenum Press, New York, 1973; Vol. 2, p. 115.
11. Soper, A. K.; Finney, J. L. *Phys. Rev. Lett.* **1993**, *71*, 4346.
12. Turner, J.; Soper, A. K.; Finney, J. L. *Molec. Phys.* **1990**, *70*, 679.