Eur Biophys J (1985) 11 : 225–237

European Biophysics Journal

Solvent structure in vitamin B_{12} coenzyme crystals

Computer simulation study

F. Vovelle[1], J. M. Goodfellow*, H. F. J. Savage[2], P. Barnes, and J. L. Finney

Department of Crystallography, Birkbeck College, University of London, Malet Street. London WC1E 7HX, UK

Received 12 March 1984/Accepted 17 September 1984

Abstract. Both the ordered and disordered solvent networks of vitamin B_{12} coenzyme crystal hydrate have been generated by Monte Carlo simulation techniques. Several different potential functions have been used to model both water-water and water-solute (i.e., water-coenzyme) interactions. The results have been analysed in terms of the structural properties of the water networks, such as mean water oxygen and hydrogen positions, coordination of each water molecule, and maxima of probability density maps in all four asymmetric units of this crystal.

The following results were found: (I) Within each asymmetric unit only one hydrogen bonding network was predicted although there were several hydrogen atom positions for any one solvent molecule (defined as maxima in probability density). (II) Reasonable agreement was obtained between predicted and experimental positions in the ordered solvent region, independent of the potential function used. (III) The positions of the calculated probability density maxima for the disordered channel region were different in different asymmetric units; this led to different simulated hydrogen bond networks which were not always consistent with the experimentally determined alternative (lower occupancy) sites.

The results suggest that it is advisable to simulate more than one asymmetric unit if one wishes to look at disorder in the solvent regions. Probability density maps were qualitatively very useful for picturing these disordered regions. However, there were no significant differences between quantitative results predicted using either average atomic positions or maxima of the probability density distributions.

Problems in quantifying agreement between experimental and predicted disordered solvent networks are discussed. The potential which included hydrogen atoms explicitly (EMPWI) seemed to give the best overall agreement, mainly because it was successful in predicting the unusually short hydrogen bonds which are found in this crystal.

Key words: Solvent structure, hydrogen bond networks, computer simulation, B_{12} coenzyme crystals, probability density

Introduction

Computer simulation techniques offer a unique possibility of examining energetic and structural properties of complex molecular assemblies, provided that the interatomic potential energy functions are adequately described. Moreover, they can be used to predict the structure of solvent networks both in liquids, such as pure water (Wood 1979) and in relatively well-ordered systems such as small, amino acid hydrate crystals (Goodfellow et al. 1982). The latter systems are studied because of the detailed experimental data which are available from X-ray and neutron diffraction. Large biomolecule solvent crystals contain a mixture of ordered and disordered solvent sites which can, with care, be characterised from high resolution crystallographic data, in terms of occupancies (related to static disorder) and temperature factors (related to thermal vibrations about average atomic positions). However, for the majority of large systems of biological interest it is only the ordered solvent regions, often with hydrogen bonds to solute atoms, which are well-defined. Even partially disordered solvent regions cannot be defined because of limited resolution of the data (which limits the data-to-parameter ratio in structure refinement) and the accumulation of errors.

If computer simulation is to be used to understand more about solvent-biomolecule interactions, it is essential to know how reliable these structural

1 Centre de Biophysique Moléculaire, CNRS, 1A Avenue de la Recherche Scientifique, F-45045 Orléans Cedex, France
2 National Bureau of Standards, Washington, DC, USA
* To whom correspondence should be addressed

predictions are. For this reason, it is necessary to simulate crystal hydrates for which experimental data are available and to make detailed comparisons before making conclusions from simulations of aqueous solutions of macromolecules. In small, amino acid hydrate crystals it was found that the predicted networks were very sensitive to the values of the input parameters used in the simulation, i.e., the potential energy functions describing each type of atomic interaction (Goodfellow et al. 1982). Even for pure water, there are a variety of models available in the literature (see review by Beveridge et al. 1983) none of which adequately describe all known properties. These models can be categorised into types such as four-point charge, three-point charge and (polarisable) electropole models. Five models have been used in this study covering this range of potential energy functions.

Previous simulations of solvent biomolecule networks have been described for small amino acid hydrates (Goodfellow et al. 1982), medium sized cyclic peptides (Hagler et al. 1980; Madison et al. 1983) and large protein crystals (Hagler and Moult 1978). For the latter crystals, only a relatively few ordered (and no disordered) water molecule sites had been determined experimentally, whereas the small amino acid hydrates contained only ordered sites for solvent. The medium-sized peptides contained few water molecules and the disorder had been defined only partially by X-ray crystallographic data.

We have chosen to study the solvent regions in vitamin B_{12} coenzyme because it is the only molecule of its size (Mr 1,580 daltons) for which (a) high resolution 0.9 Å data is available from both X-ray and neutron diffraction so that hydrogen/deuterium positions can be unambiguously determined and (b) the structure has been refined using several methods leading to a detailed model of the solvent networks in both the ordered and disordered regions (Savage 1983; H. F. J. Savage and J. L. Finney, unpublished). These hydrogen/deuterium positions are particularly important when evaluating structural properties of hydrogen bonding networks and when considering the possible orientational dependence of the hydrogen bond energy (Gellatly et al. 1983).

Methods

Experimental structure

The structure of vitamin B_{12} coenzyme was originally solved by Lenhert (1968). In order to further define the solvent regions, X-ray and neutron diffraction data have been collected at better than 1.0 Å resolution. This structure is described in detail by Savage (1983). In summary, the structure has been refined using full-matrix least-squares refinement to a residual of around 8.5%. The water occupies two distinct regions, separated by an acetone molecule (from which solvent coenzyme B_{12} is crystallised). An "ordered pocket" of water molecules is formed on one side of the acetone molecule (Fig. 1) and a more "disordered channel" region of water molecules is formed on the other side. However, this separation is complicated by the fact that the occupancy of the acetone molecule is only approximately 0.7. The ordered pocket region exhibits two distinct solvent networks, one of occupancy 0.7 and the other of occupancy 0.3 in the presence and absence of the acetone molecule respectively. The channel region is more disordered in that there are two major networks of occupancy 0.3 and 0.4 which can exist in the presence of acetone as well as several more minor sites which exist when the acetone molecule is not present. A stereo view of the major solvent network can be seen in Fig. 1b. The crystallographic temperature factors of the water molecule sites are generally higher in the channel region than in the pocket region. In comparing predicted and experimental results, the higher occupancy sites are considered first and then the alternative, lower occupancy water molecule positions. Density measurements indicate that there should be a total of 14 water molecules per asymmetric unit in the presence of acetone and about seventeen in its absence.

Potential energy parameters

Several different models are used here to describe water-water and water-solute interactions. Three of these models are "effective" pair potentials, based on point charge representations of each water molecule. They are (i) the ST2 model derived by Stillinger and Rahman (1974), (ii) a modified four point charge model (EMPWI) developed by Vovelle et al. (1981) to reproduce experimental water molecule positions in small peptide crystal hydrates and (iii) a three point charge model (TIPS2) developed by Jorgensen (1982). The final model was the polarisable electropole model (PE) which has been developed to reproduce properties of water in gaseous, liquid, and ice phases (Barnes et al. 1979) and was extended for use in Monte Carlo simulation of small hydrate crystals (Goodfellow et al. 1982). It is a non-pair additive model in which cooperative effects of water interactions are taken into account by dipole polarisability. In this study two different levels of the model have been used including multiple interactions up to (a) quadrupole-quadrupole [PE(QQ)] (Barnes et al. 1979) and (b) dipole octupole [PE($\mu\Omega$)] (Gellatly et al. 1983).

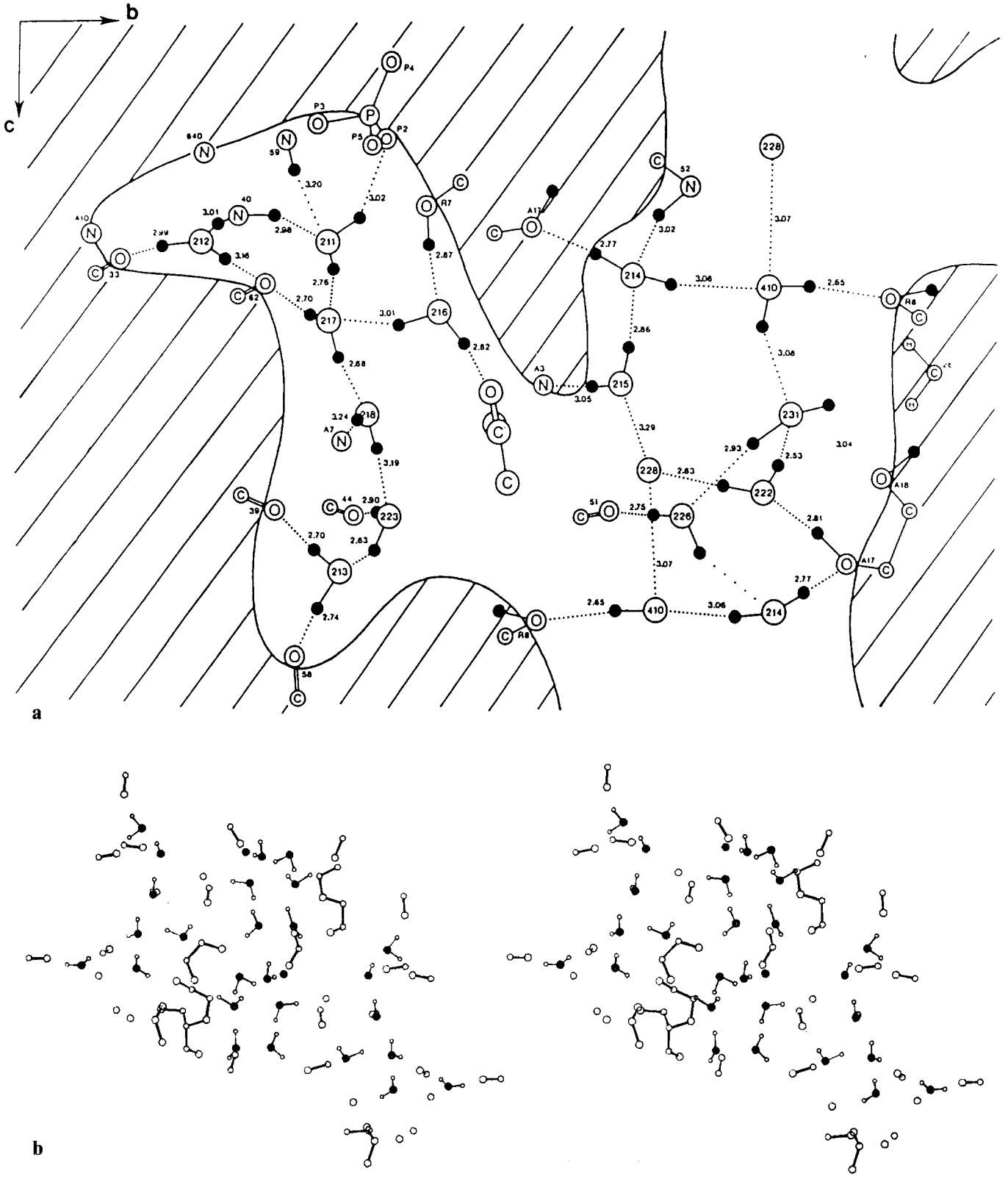

Fig. 1. a A schematic view of the experimentally-determined solvent region in crystals of vitamin B_{12} coenzyme showing the ordered pocket region and the disordered channel region separated by the central acetone molecule (Savage 1983). **b** A stereo view of the experimentally determined solvent networks

Table 1. Characteristics of the potential energy functions

		TIPS2	ST2	PE(QQ)	PE(μΩ)	EMPWI
(a)	*Water-water interactions*					
(i)	Electrostatic					
	charge-charge	√	√	–	–	√
	electropole	–	–	√	√	–
(ii)	Polarisation	–	–	√	√	–
(iii)	Non-bonded					
	O----O 6–12	√	√	√	–	√
	O----O 6–9	–	–	–	√	–
	H----H 6–12	–	–	–	–	√
(b)	*Water-coenzyme interactions*					
(i)	Electrostatic					
	charge-charge	√	√	–	–	√
	charge-electropole	–	–	√	–	
(ii)	Non-bonded					
	all non-hydrogen atoms 6–12	√	√	√	–	√
	all non-hydrogen atoms 6–9	–	–	–	√	–
	hydrogen 6–12	–	–	–	–	√

The main characteristics of the five potential functions are presented in Table 1. The solute non-bonded interaction coefficients were either of 6–12 form (i.e., $B/R^{12}-A/R^6$) which were taken from Momany et al. (1974) or of a 6–9 form (i.e., $B/R^9-A/R^6$) which were taken from Lifson et al. (1979). The partial atomic charges on the coenzyme molecule were calculated using the semi-empirical quantum mechanical CNDO/2 method using overlapping fragments in the experimental conformation (Savage 1983). The coordinates for the coenzyme molecule are available on request.

Computer simulation

Monte Carlo computer simulations of a complete unit cell of atoms were carried out using the standard Metropolis algorithm (Metropolis et al. 1953). Periodic boundary conditions were used to reproduce the translational symmetry of the crystal. The atoms of the B_{12} and acetone molecules were kept fixed in their experimental crystal positions. The experimental solvent network with maximum occupancy was used as the starting configuration (Savage 1983). No symmetry constraints were applied to the positions of the 56 water molecules which were independently randomly rotated and translated within the unit cell. A distance cut off of 7.5 Å was used in the energy calculation in all the simulations. At least 2×10^5 configurations were generated with each potential function. The statistically averaged properties were calculated over the subsequent 3×10^5 configurations.

Probability density

The probability density (PD) map for solvent molecules in the unit cell was determined by dividing the unit cell into 0.4 Å sided cubes and counting the number of configurations for which each cube was occupied during the progress of the simulation. PD maps were calculated separately for oxygen and hydrogen atoms. The hydrogens of any one water molecule were regarded as indistinguishable.

Representation of these results is rather difficult, as the values of the probability density for each cube, obtained in the five simulations, were so scattered. In order to compare our results with the experimental data, we show probability density maps obtained by superposition of the different sections in the crystallographic $\vec{a}$ direction. This projection, in the $\vec{b}$ $\vec{c}$ plane of the probability density data, can be compared to the projection, in the same plane, of the experimentally determined sites of water molecules. Zero probability contours are drawn and the positions of maximum probability are indicated. Of course, this constitutes only a partial representation of the results; the real accuracy of the calculations (along the $\vec{a}$ axis) cannot be assessed from these maps. Further examination of these probability density results will be presented below, in terms of maxima of the probability density.

Results

Several structural and energetic properties were calculated with the aim of comparing the simulated

Table 2. RMSD between experimental and simulated water positions

Potential function	Average RMSD for all waters in unit cell	Average RMSD for waters in each asymetric unit			
		AU1	AU2	AU3	AU4
PE (QQ)	1.22 ± 0.73	1.55	1.31	0.90	1.12
PE ($\mu\Omega$)	1.26 ± 0.6	1.25	1.37	1.31	1.21
ST2	1.02 ± 0.56	1.28	0.74	0.99	1.06
TIPS2	0.86 ± 0.55	0.86	0.99	0.81	0.81
EMPWI	0.75 ± 0.24	0.69	0.82	0.79	0.68

structures with the experimentally defined solvent networks. Because each asymmetric unit has been simulated independently, i.e., without symmetry constraints on the solvent, it is equivalent to undertaking four simulations on the same ensemble. For an infinite simulation, we would expect all prossible configurations to be sampled in each asymmetric unit and the results from each asymmetric unit would be the same on average. Thus, ultimately the same water network(s) would be predicted in each asymmetric unit. In practice, finite simulations are subject to equilibration and sampling problems which may become severe in complex heterogeneous systems such as at water-macromolecular interfaces (Finney et al. 1984). We found that different asymmetric units in these simulations sample different configurations. Thus we have attempted to analyse our results to take into account the total information predicted in all four asymmetric units and to look at the level of agreement between the different asymmetric units.

(i) Average properties

The mean positions of each of the 56 water molecules were calculated from the simulated configurations and were used to obtain the deviation of the solvent molecule oxygen positions from corresponding experimental positions (RMSD) in Å (Table 2). The discrepancy between simulated and experimental data is smallest with the EMPWI model and largest for the PE models. The relatively large standard deviations of the RMSD for the first four models highlights the problems of using this quantity to assess agreement with experiment. The differences between the average RMSD for each asymmetric unit may be an indication that different solvent networks are being sampled in the four asymmetric units.

The RMSD of each water molecule around its simulated average position can be compared with the isotropic root mean square displacements around experimental sites (EXPTL) deduced from the crystallographic temperature factors (Table 3). The predicted results are always greater than the experimental results, reflecting large movements of the water molecules and a possibility of oscillation between alternative water positions during the simulation. This is illustrated by calculating separately these fluctuations for water molecules in two distinct solvent regions of the crystal, namely the ordered pocket region and the disordered channel region. Using the EMPWI potential, a mean value of 0.33 ± 08 Å was found for the RMSD of 28 water molecules in the pocket. This compares well with 0.31 Å ± 0.03 for the average displacement found from the experimental data. In the channel region, this simulation leads to a fluctuation of 0.4 Å. This is higher than the experimental value (0.31 Å) but, as discussed later, the simulation takes no account of alternative, lower occupancy, solvent networks whose fluctuations are calculated separately for the experimental data.

Table 3. Fluctuation around average positions

Potential function	RMSD around average positions for all waters in unit cell in Å
PE (QQ)	0.55 ± 0.16
PE ($\mu\Omega$)	0.46 ± 0.15
ST2	0.52 ± 0.2
TIPS2	0.48 ± 0.16
EMPWI	0.37 ± 0.12
EXPTL	0.31 ± 0.025

(ii) Solvent molecule environment

The root mean square deviations calculated in the previous section do not lead to a clear view of the environment of each solvent molecule. It is, however, difficult to express quantitatively the agreement between experimental and predicted coordination of each solvent molecule. An attempt has been made to assess the discrepancy between water molecule environment by calculation of various agreement factors, the values of which are given in Table 4 (A and B).

Table 4a. Agreement between the coordination of water molecules in experimental and simulated environment

Agreement factor I[a]					
Potential function	Average over all waters in Å	Average values for asymmetric unit AU			
		AU1	AU2	AU3	AU4
(A) Water-water interactions					
PE (QQ)	0.65	0.88	0.90	0.34	0.49
PE ($\mu\Omega$)	0.79	0.66	0.92	0.90	0.69
ST2	0.33	0.32	0.34	0.32	0.32
TIPS2	0.37	0.48	0.35	0.33	0.31
EMPWI	0.26	0.31	0.31	0.20	0.22
(B) Water-coenzyme interactions					
PE (QQ)	0.52	0.61	0.68	0.26	0.52
PE ($\mu\Omega$)	0.78	0.78	0.71	0.96	0.67
ST2	0.38	0.48	0.30	0.33	0.40
TIPS2	0.36	0.26	0.55	0.33	0.30
EMPWI	0.28	0.31	0.32	0.25	0.24

[a] Measure of similarity in environment between predicted and experimental results (large values indicate worse agreement)

Table 4b. Agreement between the predicted coordination of water molecules in each asymmetric unit

Agreement Factor II[b]					
Potential function	Average value over 56 waters	Average values for asymmetric unit AU			
		AU1	AU2	AU3	AU4
(A) Water-water interactions					
PE (QQ)	0.35	0.41	0.28	0.49	0.20
PE ($\mu\Omega$)	0.44	0.58	0.52	0.42	0.27
ST2	0.22	0.19	0.25	0.22	0.22
TIPS2	0.30	0.40	0.24	0.34	0.22
EMPWI	0.12	0.17	0.15	0.08	0.07
(B) Water-coenzyme interactions					
PE (QQ)	0.28	0.23	0.32	0.34	0.21
PE ($\mu\Omega$)	0.39	0.46	0.41	0.42	0.26
ST2	0.21	0.28	0.20	0.20	0.17
TIPS2	0.22	0.23	0.32	0.16	0.19
EMPWI	0.07	0.07	0.10	0.05	0.05

[b] Measure of similarity in environment between the predicted results in each asymmetric unit of the crystals (larger values indicate worse agreement)

The first factor is the RMSD between the predicted and experimental coordination, i.e., solvent oxygen to neighbouring non-hydrogen atom (OSS) *distances within a sphere of 3.5Å radius.* The second agreement factor is defined as the RMSD of the OSS distances between the individual and average values for each asymmetric unit. Thus a lower agreement factor denotes better agreement (I) with experiment and (II) between asymmetric units. These factors are calculated for all interactions and for water-solute and water-water interactions separately. It was found that, in all simulations

(i) the agreement factor I is greater than agreement factor II for all interactions; i.e., agreement with experiment is worse than agreement between simulations of different asymmetric units;

(ii) the agreement factor II for water-water interactions is greater than for water-solute interactions;

(iii) the actual size of the agreement factors depends on the potential used (see Table 4) such that significantly different solvent coordinations are predicted;

(iv) the values for agreement factor I found with the EMPWI potential for this system are of the same order of magnitude as those obtained previously for *small* nucleic acid crystal hydrates (Goodfellow 1984). Independent calculations of agreement factors for "ordered pocket" and "disordered channel"

Fig. 2a and b. Probability density map from simulation using EMPWI potential functions for (**a**) solvent oxygen and (**b**) solvent hydrogen atoms in two asymmetric units. × Indicates the experimental solvent oxygen sites. + Indicates an alternate (lower occupancy) solvent oxygen site. △ Indicates a maximum in the probability density map

water molecules separately lead to a much better agreement for the experimental and simulated data pocket waters e.g., the agreement factor between for PE(QQ) water-water interactions varies from 0.38 Å in the pocket to 1.08 Å in the channel.

(iii) Probability density maps

In order to assess the possibilities of movement of water molecules between alternative positions during the simulation, the probability densities for both water oxygen and hydrogen atoms were recorded for the simulations using the EMPWI potential (Fig. 2a and b) and the PE(QQ) potential. In the pocket region, the oxygen contours are well defined, such that each contour can be attributed to a given water molecule. In the channel region, the contours are much broader so that it is no longer possible to attribute a given area to a specific water molecule. The higher and lower occupancy experimental positions of water molecules are nearly always contained within the contoured areas which indicates that water molecules can move between these different sites. Although there is considerable disorder, it can be seen that the contours presented for the two asymmetric units shown in Fig. 2a and b are related by a crystallographic two fold screw axis. The contours for hydrogen atoms are broader, partly because each water molecule has two hydrogen atoms. However, the maxima are well-defined showing preferential sites for the hydrogen atoms during the simulation. A superposition of the maxima for both the oxygen and hydrogen shows that the solvent oxygen to hydrogen directions (Figure 3) are similar to the experimental results (Figure 1b).

(iv) Hydrogen bond network

(a) From average positions. In any study of solvent interactions, it is the hydrogen bonding networks, and not just the individual atomic positions, which are important. Histograms of water-water and water-solute hydrogen bond (OSS) distances (< 3.5 Å) are drawn in Fig. 4b and c from the average solvent positions of the EMPWI and PE (QQ) simulations respectively and can be compared to the experimental distribution (Fig. 4a) which is much broader with some unusually short (< 2.6 Å) hydrogen bond distances. All simulated distributions exhibit very strong maxima whose positions vary from 3.05 Å for PE ($\mu\,\Omega$) to 2.85 Å for TIPS2.

Examination of the experimental data (Fig. 1b) and one of the predicted networks (Fig. 3) shows that there are no major differences in the hydrogen bonding patterns in the pocket region of the B_{12} crystal although there are differences in the values of hydrogen bond distances. In the channel region, the situation is different. The EMPWI simulation alone reflects the experimental network with reasonable hydrogen bond distances. The predicted PE (QQ) solvent structure is far from the experimental one and does not look like any alternative experimental network partly because water 215 moves more than 2 Å along the $\vec{a}$ axis. Experimentally, this movement is possible and might be correlated with movement of water 228 (Savage 1983).

Average hydrogen bond distances between water hydrogen atoms and neighbouring solute atoms, and hydrogen bond angles are given in Table 5. Although TIPS2 and EMPWI potentials perhaps show closest agreement with experimental values, the standard deviations on all average HSS distance are such that

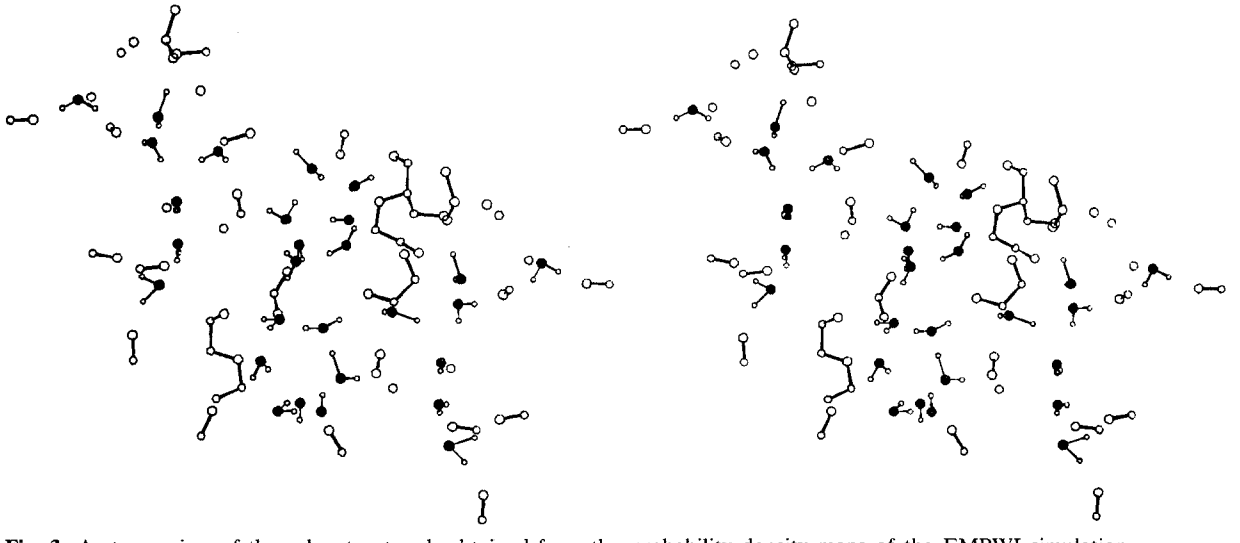

Fig. 3. A stereo view of the solvent network obtained from the probability density maps of the EMPWI simulation

Table 5. Hydrogen bond distances and angles

Hydrogen bond distances	PE (QQ)[a]	EMPWI[a]	EMPWI[b)]	Experimental
Water oxygen———water oxygen [Å]	3.04 ± 0.18	2.88 ± 0.2	2.91 ± 0.24	2.92 ± 0.24
Water oxygen———coenzyme atom [Å]	3.19 ± 0.19	3.04 ± 0.21	3.0 ± 0.21	3.0 ± 0.3
Water hydrogen———coenzyme atom [Å]	2.49 ± 0.69	2.37 ± 0.2	1.99 ± 0.2	1.97 ± 0.2
Hydrogen bond angle				
Subtended at water hydrogen atom [Å]	137 ± 24	112 ± 18	148 ± 17	155 ± 12

[a] Calculated from mean oxygen and hydrogen positions
[b] Calculated from maxima of the probability density

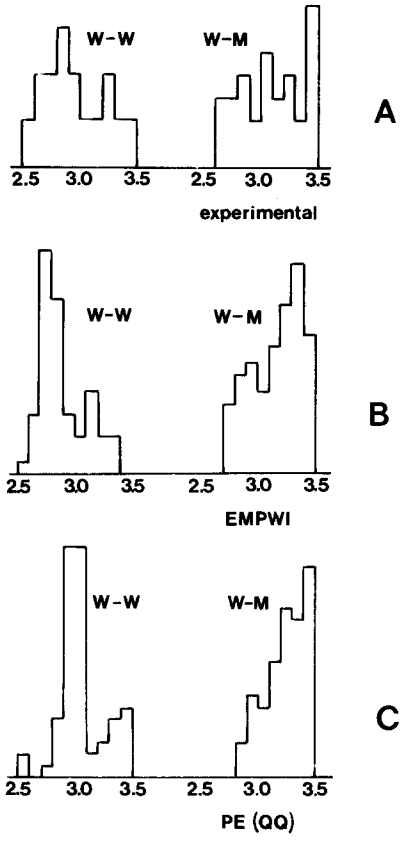

Fig. 4a–c. Histograms of both water–water (W-W) and water–coenzyme (W-M) hydrogen bond distances (OSS). (**a**) For the experimental structure; (**b**) from a simulation using EMPWI potential functions and (**c**) from the simulation using PE (QQ) potential functions

any further significant distinction is difficult to make. The average hydrogen positions, predicted using EMPWI potential (Fig. 5a), compare well with the experimentally determined positions which show no disorder in the pocket region. However, a few slight differences were found which are as follows: (i) In most asymmetric units water 216 is bonded solely to water 218 whereas experimentally the hydrogen bond occurs with water 217 (Fig. 1a) although the 0216. . .0218 distance is short, 3.3 Å. (ii) There is no prediction of a hydrogen bond between H212 and solute oxygen 062. (iii) The coordination of water 218 with nitrogen 97 is not predicted.

In the channel, the simulated solvent hydrogen bond pattern differs from that in the pocket in that different networks are predicted in the different asymmetric units. Five water molecules (214, 215, 222, 226, and 231) have experimentally determined alternative hydrogen/deuterium sites whereas water 228 has no well-defined hydrogen positions. In general, asymmetric units 1, 3, and 4 are similar but 2 is different particularly around the positions of water 222 and 226 which show a high degree of disorder experimentally. However, the simulated hydrogen bond pattern in asymmetric unit 2 is the closest to the experimentally determined pattern. Thus it appears that the experimentally seen disorder shows up as different hydrogen bonding patterns in the different asymmetric units.

(b) From the probability density data. An alternative way to analyse results of simulations is to calculate structural properties from the maxima of the probability density maps for oxygens and hydrogens since they represent the sites where an atom has the highest probability of being found during the simulation. Several maxima of the same order of magnitude were sometimes found corresponding to alternative positions for an atom and these were compared with both main and lower occupancy experimentally determined sites.

The average deviation between the calculated maxima and the highest occupancy crystal oxygen position was found to be 0.76 ± 0.24 Å. This is indistinguishable from the 0.75 ± 0.24 Å obtained for the deviation from the predicted *average* atomic

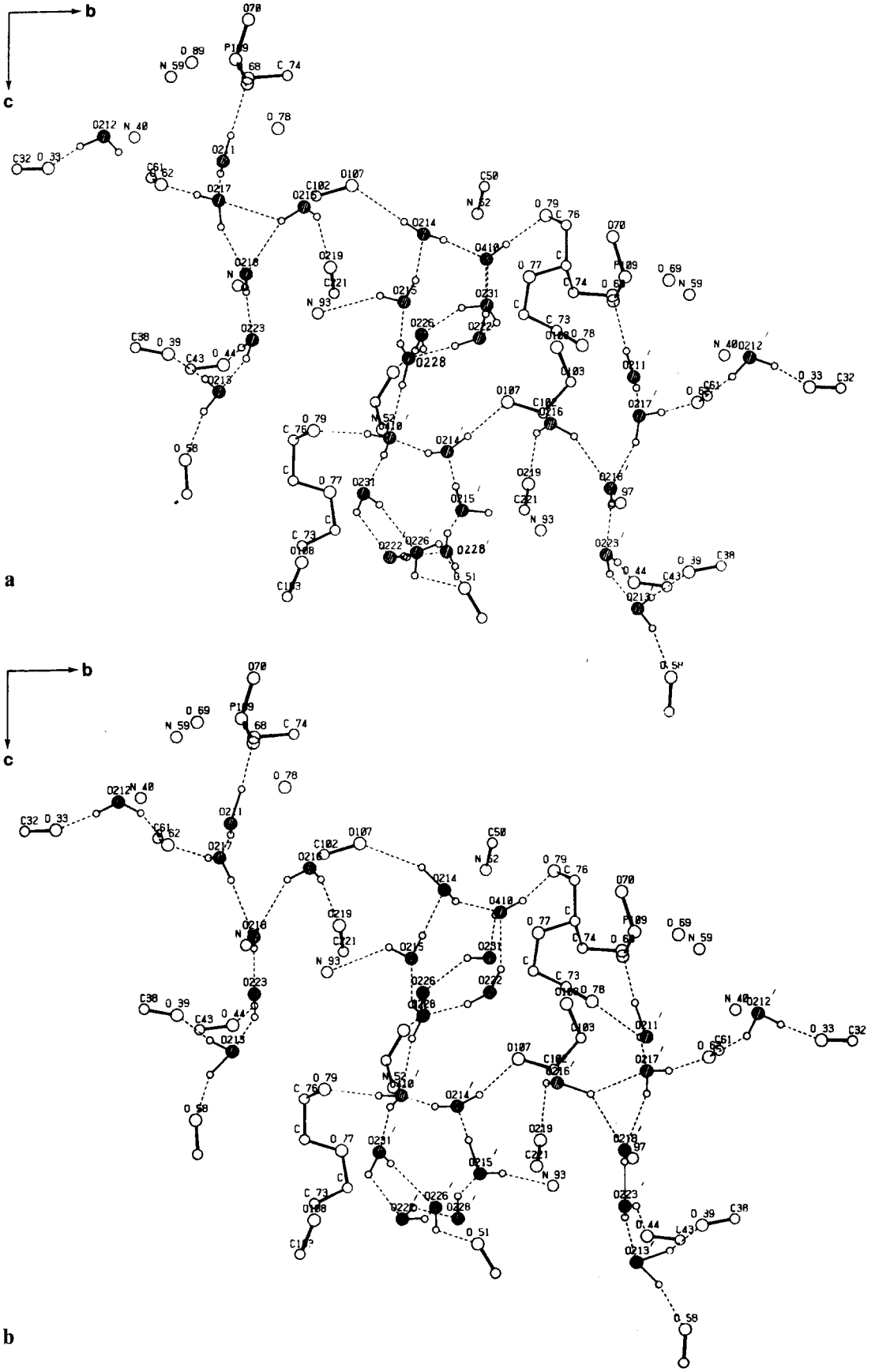

Fig. 5a and b. Projections of the solvent networks showing (**a**) both oxygen and hydrogen/deuterium positions, average atomic positions using EMPWI potential; (**b**) atomic positions assigned from the maxima of probability density obtained using EMPWI potential with dotted lines representing hydrogen bonds. Water molecules in the second asymmetric unit have been designated with a prime e.g., 0211′

positions. Hydrogen bond distances and angles are given in Table 5. The hydrogen bond distances for the EMPWI model are very close to the average experimental values and there appears to be no significant differences between these values calculated from maxima of probability density maps and average positions. The agreement between predicted and experimental hydrogen bond angles is less satisfactory. Both distributions show deviation from linearity which is consistent with data on small hydrate crystals.

Distances between predicted oxygens and alternative water positions show that water 211 is closer to an alternative experimental site in all asymmetric units. In the channel, simulated waters are often close to alternative sites but the pattern is never the same in all symmetric units.

The hydrogen bonding pattern from the probability density maps is presented, in projection, in Fig. 5b. The main differences from the experimental network involve (a) water 211 moving towards an alternative position and (b) water 410 in the channel region hydrogen bonding to water 222 instead of water 228 in asymmetric unit 1. However, this behaviour is not observed in asymmetric unit 2 where the solvent structure is similar to the highest occupancy experimental sites. These differences between asymmetric units are not surprising since there is much disorder in the region of the 410 site correlated with the movement of the water molecules through the channel (Savage 1983).

Although the predicted probability density maxima and the predicted average atomic position are very similar, slight differences can be seen in the hydrogen bonding networks if Fig. 5a and b are compared. These differences include (1) waters 212 and 216 in the pocket region, (2) some hydrogens in the channel region which could not be located as no sharp maxima were seen in the region of the very disordered waters 222, 231, and 228. (3) No alternative sites are predicted for waters 214 and 215.

(c) Summary of hydrogen bonding networks. In summary, it is found that:

(i) Within each asymmetric unit, several hydrogen positions (i.e., maxima of probability density) may be found but they all lead to the same pattern of hydrogen bonding.

(ii) In the pocket region, the hydrogen bonding pattern predicted agrees reasonably with the experimental network.

(iii) In the channel region, the position of these maxima may differ from one asymmetric unit to another and then a different hydrogen bonding pattern may be predicted in each asymmetric unit.

(iv) In the channel region, although different patterns are produced in different asymmetric units these do not always resemble those alternative networks found experimentally.

Discussion

In this study, computer simulations of the network of solvent molecules around a biomolecule have been made using several models for the solvent interactions and each of these simulations has been compared with the well-defined experimental water structure found in the vitamin B_{12} coenzyme crystal. The main problem has been to find qualitative and quantitative parameters which give a useful measure of the success of the simulation, i.e., significance of the agreement or disagreement with the experimental solvent structure. The other feature of this study is the use of various models for the interactions of the solvent molecules both with other solvent and with the biomolecule and whether the differences predicted in the structures of the solvent networks can be related to the models themselves. All computer simulations of complex heterogeneous systems are limited by sampling and equilibration problems. Therefore, it may not be possible to draw firm conclusions about the success of a simulation in reproducing experimental results but useful comparisons can be made between the predictions from simulations using different potential energy functions (Finney et al. 1984).

The quantity most generally used to compare solvent networks is the root mean square displacement between the solvent oxygen atom positions in the predicted and experimental networks. Although this RMSD gives a quantitative measure of agreement between each water molecule position, it has been found not to be a good criterion for either intuitively picturing the level of agreement or for assessing the hydrogen bond network as distinct from the individual positions. In the study of amino acid crystal hydrates (Goodfellow et al. 1982), it was found that lowest values of the root mean square displacement did not necessarily correspond to the best agreement in prediction of the coordination (neighbours within a 3.5 Å sphere) of each solvent molecule. It is this latter parameter which will define the hydrogen bond distances and angles.

The average fluctuation (for all water molecules in the ensemble) of water molecule oxygen positions around their average predicted position is a quantity which can be compared with experimental values (derived from the crystallographic temperature factors). As the coenzyme atoms are kept fixed during the simulation it is expected that the fluctuations of

solvent positions would be less than the experimental values. However, the values for the fluctuations found with all potentials (Table 3) are at least as large as the experimental values. This may be due to the simulated fluctuations accounting not only for movement around individual solvent sites but also for movement between alternative sites where there is disorder in the solvent network. The alternative sites are included explicitly in the experimental structure and therefore the fluctuations are derived for each site. When these considerations are taken into account, the agreement between prediction and experiment is relatively good for all models.

The coordination (defined as nearest neighbour atoms within 3.5 Å sphere) has been defined for each solvent molecule from average oxygen atom coordinates for each simulation. Use of agreement factor I allows us to calculate *one* parameter which includes the distances between all solvent molecules and their neighbours. In order to assess the significance of this number, a related quantity was calculated to show the discrepancy between asymmetric units within each simulation. One would hope that both agreement factors would be similar in magnitude but since we found that agreement factor I was always larger than agreement factor II, we still have room for improvement especially in the determination of the water-solute parameters.

Probability density maps were found to be useful in that they give an easily assimilated qualitative picture of the simulation results and their agreement with experimental solvent sites. Quantitatively, the average oxygen positions calculated over the whole simulation gave similar results to those obtained from the maxima of the probability density distributions for the solvent oxygen and hydrogen atoms. Although the probability density maps led to several hydrogen atom positions within any one asymmetric unit, these gave the same hydrogen bonding networks as those deduced from the average atomic positions.

In the disordered region, different hydrogen bonding patterns were predicted in different asymmetric units from both probability density and average positions. Again, probability density maps are not essential but the importance of simulating more than one asymmetric unit (without crystal constraints on the solvent molecules) is underlined. Therefore, probability density maps are found to be visually useful but not essential for the quantitative study of this hydrate crystal. This qualitative usefulness is limited by the considerably increased computational requirements and data reduction problems of the probability density studies.

Although it is clear that simulations of solvent performed with different potential functions give different predicted results, it is not easy to decide which model is "best". This is due to two main problems which arise when comparisons of hydrogen bond networks are made. First the stereochemical constraints on these networks are weak and thus allow a range of hydrogen bond distances and angles. Secondly solvent networks in crystal hydrates can show a range of order as defined by the crystallographic occupancies (static disorder) and temperature factors (thermal disorder). It appears from this study that EMPWI leads to the overall closest agreement with the experimental solvent network. PE (QQ) appears to have the worst agreement, with ST2 and TIPS2 being intermediate. However, any parameter we use to measure differences in structure has a large standard deviation so that the significance of these differences is not clear. For example, the RMSD for EMPWI and PE (QQ) are significantly different on a statistical (student's *t*-test) basis but the difference between PE (QQ), ST2, and TIPS2 is not significant.

Further analysis of the hydrogen bond networks, rather than just atomic positions, showed that the EMPWI potential reproduces the unusually short hydrogen bonds found in the B_{12} structure better than the other potentials. This may be the cause of this potential yielding the best overall agreement. The main difference between EMPWI and ST2 is in the explicit inclusion in the former of non-bonded energy terms for the solvent and solute *hydrogen atoms*. The importance of hydrogens in maintaining solvent structure has been seen in the experimental B_{12} coenzyme solvent networks (Savage 1983) and in water hydrogen bonds (Gellatly et al. 1983). If this difference between EMPWI and ST2 is the reason for the better agreement with experiment, it is unfortunate because it significantly increases the computational requirements.

If the EMPWI potential is to be used in further simulations of solute-solvent interactions, it is necessary to ascertain the properties of bulk water predicted by this model. Such studies have been initiated.

Acknowledgements. This work was supported by SERC project grant GR/B/73736. JMG and FV wish to thank the SERC for an Advanced Fellowhip award and support by the CNRS/Royal Society exchange programme respectively.

References

Barnes P, Finney JL, Nicholas JD, Quinn JE (1979) Cooperative effects in simulated water. Nature 282: 459–464

Beveridge DL, Mezei M, Mehrotra PK, Marchese F, Ravi-Shanker G, Vasu T, Swaminathan S (1983) Monte Carlo computer simulation studies of the equilibrium properties and structure of liquid water. In: Haile JM, Mansoori GA (eds) Molecular

based study and prediction of fluid properties. Adv Chem American Chemical Society, New York, pp 297–351

Finney JL, Goodfellow JM (1983) Cooperative effects in aqueous biomolecular systems and processes. In: Clementi E, Sarma RA (eds) Structure and dynamics: Nucleic acids and proteins. Adenine Press, New York, pp 81–93

Finney JL, Goodfellow JM, Howell PL, Vovelle F (1985) Computer simulation of aqueous biomolecular systems. J Biomol Struct Dynamics (in press)

Gellatly BJ, Quinn JE, Barnes P, Finney JL (1983) Two, three and four body interactions in model water interactions. Mol Phys 59: 949–970

Goodfellow JM, Finney JL, Barnes P (1982) Monte Carlo computer simulation of water-amino acid interactions. Proc R Soc (London) B214: 213–228

Goodfellow JM (1984) Solvent interactions in nucleic acid crystal hydrates. J Theor Biol 107: 261–274

Hagler AT, Moult JM (1978) Computer simulation of the solvent structure around biological macromolecules. Nature 272: 222–226

Hagler AT, Moult, JM, Osguthorpe DJ (1980) Monte Carlo simulation of the solvent structure in crystals of a hydrated cyclic peptide. Biopolymers 19: 395–418

Jorgensen WL (1982) Revised TIPS for simulations of liquid water and aqueous solutions. J Chem Phys 77: 4156–4163

Lenhert PG (1968) The structure of vitamin B_{12} VII. The X-ray analysis of the vitamin B_{12} coenzyme. Proc R Soc (London) A303: 45–84

Lifson SØ, Hagler AT, Dauber P (1979) Consistent force field studies of intermolecular forces in hydrogen-bonded crystals I. Carboxylic acids, amides and the C=O· · ·H-hydrogen bonds. J Am Chem Soc 101: 5111–5121

Madison V, Osguthorpe DJ, Dauber P, Hagler AT (1983) Monte Carlo simulations of peptide solvation. Biopolymers 22: 27–31

Metropolis M, Metropolis AW, Rosenbluth MN, Teller AH, Teller E (1953) Equation of state calculations by fast computing machines. J Chem Phys 21: 1087–1092

Momany FA, Carruthers LM, McGuire RF, Scheraga HA (1974) Intermolecular potentials from crystal data III. Determination of empirical potentials and application to the packing configurations and lattice energies in crystals of hydrocarbons, carboxylic acids, amines and amides. J Phys Chem 78: 1595–1620

Savage HFJ (1983) A study of the disordered water structure in crystals of vitamin B_{12} coenzyme. Ph D Thesis. University of London

Stillinger FA, Rahman A (1974) Improved simulation of liquid water by molecular dynamics. J Chem Phys 60: 1545–1557

Vovelle F, Geneste M, Ptak M, Maigret B (1981) Empirical models of hydration of small peptides. In: Pullman B (ed) Intermolecular forces. Reidel, Dordrecht, pp 299–315

Wood DW (1979) In: Franks F (ed) Water, vol 6. Plenum Press, New York