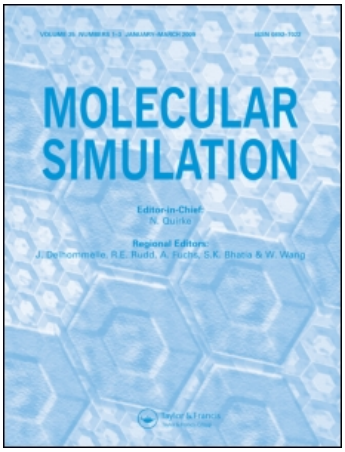

Molecular Simulation
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713644482

Effect of Initial Positions on the Simulation of Water Networks in Crystal Hydrates

P. Lynne Howell[ab]; Julia M. Goodfellow[a]
[a] Department of Crystallography, Birkbeck College, University of London, London, UK [b] Chemistry Department, MIT, Cambridge, MA, USA

To cite this Article Howell, P. Lynne and Goodfellow, Julia M.(1988) 'Effect of Initial Positions on the Simulation of Water Networks in Crystal Hydrates', Molecular Simulation, 1: 5, 333 — 343
To link to this Article: DOI: 10.1080/08927028808080955
URL: http://dx.doi.org/10.1080/08927028808080955

Molecular Simulation, 1988, Vol. 1, pp. 333–343
Photocopying permitted by license only
Reprints available directly from the publisher

EFFECT OF INITIAL POSITIONS ON THE SIMULATION OF WATER NETWORKS IN CRYSTAL HYDRATES

P. LYNNE HOWELL[1] and JULIA M. GOODFELLOW*

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

(*Received August 1987, in final form January 1988*)

In this study we examine the water structure in nucleotide crystal hydrates using the Monte Carlo simulation technique. It is assumed that during such a simulation water molecules can adequately sample configurational space and that their true positions can eventually be found regardless of their starting positions. We have looked at the effect of different initial sites for the water molecules on the final average properties of the water networks using crystal hydrate, guanosine-5′-phosphate trihydrate for which experimental structural data exist. We found that all water moleucles moved from their initial sites regardless of whether these were the experimentally defined crystallographic or random positions. When the water molecules were randomly placed initially, and although they moved on average over 1.2 Å, the water molecules finished further away from the experimental positions than those from the simulation which started with experimentally defined positions. However, an examination of the hydrogen bonding networks, on the basis of oxygen–oxygen distances found in the simulations, showed that the crystallographically determined network was obtained from both simulations.

KEY WORDS: Monte Carlo simulation, solvent networks

INTRODUCTION

Knowledge of the structural and energetic influences of water is a key step to the deeper understanding of the structure and function of proteins [2–5] and nucleic acids [7] in aqueous solution. Computer simulation techniques offer methods for predicting these structural and energetic properties of macromoleucular–solvent systems. The reliability of these methods depends on the use of realistic potential energy functions to model each type of atom–atom interaction and on adequate sampling of configurational space [3].

Computer simulations have been used to predict water structure in liquid water [9] and aqueous solutions of small solutes [10–12]. It has also been used to study the solvent networks in crystal hydrates of amino acids [13], peptides [14], proteins [15,16] and nucleotides [17,18]. In these latter studies, the water oxygen (and sometimes hydrogen) atomic positions have been located from crystallographic data. Thus, one can make a direct comparison of the hydrogen bond networks found during the simulation with those found experimentally.

The aim of this study is to look for the effect of different initial positions for the water molecule sites on the final simulated solvent hydrogen bond network. Thus, we

[1]Present Address: Chemistry Department, MIT, Cambridge MA 02139, USA
*To whom correspondence should be sent

Table 1

Initial[a] *Site*	*Crystal Site*[b]		*Crystal Site*[c]		*Crystal Site*[d]	
	Site	*Dist*[e]	*Site*	*Dist*[e]	*Site*	*Dist*[e]
R1	X1	1.12	X2	3.48		
R2	X2	2.07	X9	3.72		
R3	X3	1.68	X2	3.69		
R4	X4	1.81	X5	3.43		
R5	X5	1.66	X6	3.34		
R6	X6	1.75	X5	2.97		
R7	X7	1.32	X7	2.82		
R8	X8	1.27	X8	2.25	X9	3.55
R9	X9	2.18	X8	3.05		
R10	X10	1.16	X12	3.64	X11	3.73
R11	X10	1.55	X11	2.03		
R12	X12	2.59				

(a) Initial sites for water molecular oxygen positions
(b) Closest Crystal water site to initial random sites
(c) Second closest Crystal water site less than 3.75 Å to initial site
(d) Third closest Crystal water site less than 3.75 Å to initial site
(e) Distance between the initial and crystal sites in Å.

wish to ascertain whether, given arbitary starting positions, the water molecules (in a crystal hydrate) can move during the simulation to their true, experimentally determined positions.

METHODS

For any simulation it is necessary to define an initial set of coordinates for both the solute and solvent molecules. Initial coordinates for the guanosine-5′-phosphate are taken from the experimental crystal structure [1]. For the water molecules, two

Table 2

Atom Type	*Charge*[a]	*Atom Type*	*Charge*[a]
P	0.310	C1′	0.086
N1	−0.452	O1′	−0.035
C2	0.418	C2′	−0.039
N2	−0.279	O2′	−0.425
N3	−0.345	C3′	0.262
C4	0.310	O3′	−0.206
C5	−0.112	C4′	0.077
C6	0.342	C5′	0.138
O6	−0.437	O5′	−0.279
N7	−0.159	O1	−0.458
C8	0.020	O2	−0.406
N9	−0.095	O3	−0.444
HN1	0.379	HC1′	0.023
HN2	0.168	HC2′	0.007
HN2	0.164	HO2′	0.228
HC8	0.131	HC4′	−0.005
		HC5′	−0.016
		HC5′	−0.004

(a) partial atomic charges given as fractions of an electronic charge

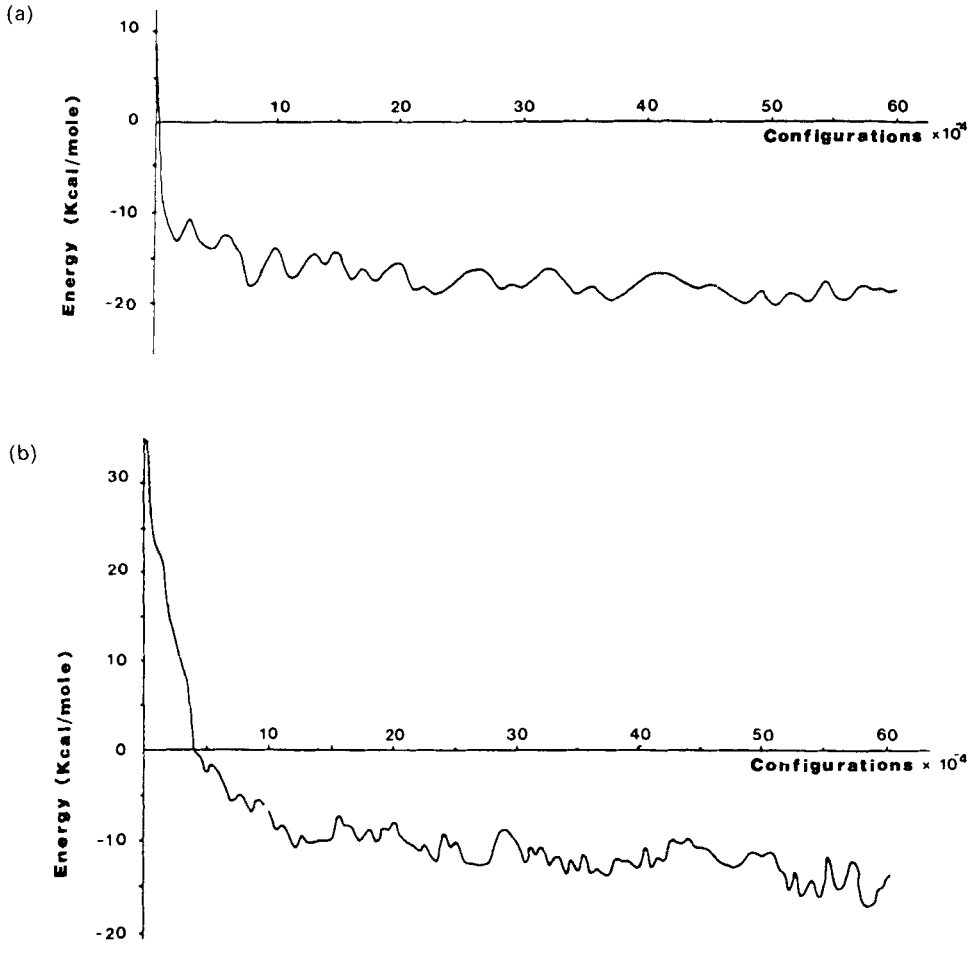

Figure 1 The potential energy profile as a function of the number of configurations for simulations (a) SIMG and (b) SIMR.

alternative sets of initial coordinates are used. These are either those of the experimentally defined oxygen atoms with random hydrogen positions (SIMG) or randomly generated water oxygen and hydrogen positions (SIMR).

The random positions are generated by dividing the unit cell into cubes of side 0.5 Å in order to define a grid. The grid points are then searched sequentially for possible water sites, using the criterion that no water oxygen atom could be found within 2.5 Å of any other atom (water or solute). Table 1 shows the association between each random water position and its closest experimental position (mapping) in order to give the smallest average root mean square deviation (RMSD) of 1.72 Å. Distances between the inital and experimental sites are also given for other mappings which lead to slightly larger deviations. In nearly all cases the mapping with the lowest root mean square deviation is obvious but for water molecules 7 and 8 it is not clear cut an

Table 3

Name	*Initial*[a] *configurations*	*Total number of moves*	*Analysis range*
SIMG1	A	6×10^5	$1–2 \times 10^5$
SIMG2	A	6×10^5	$5–6 \times 10^5$
SIMR1	B	6×10^5	$2–3 \times 10^5$
SIMR2	B	6×10^5	$5–6 \times 10^5$

(a) A: Experimental Crystal Sites
B: Random Sites from Grid

alternative mapping is possible with a slightly higher root mean square deviation procedure has to be taken during the analysis in order to determine the mapping with the smallest RMSD.

Monte Carlo simulations are performed using the Metropolis algorithm [19] on a unit cell of guanosine-5′-phosphate trihydrate in the NVT ensemble. The step sizes for translational and rotational moves were chosen to give an acceptance ratio of approximately 50%. The polarisable electropol (PE) model [20,21] is used to model the water–water interactions and the potential energy functions of Lifson et al [22] are used with the PE model and a geometric mean comination rule for the water–solute interactions. Thus, the water–solute potential energy is represented by a non-bonded Lennard Jones (6/9) interaction and an electrostatic term. The partial charges on the solute (see Table 2) are calculated using the semi-empirical CNDO/2 method as the charges for the phosphorous atoms are not available from the Lifson et al study. The water–water potential energy is a linear sum of a non-bonded Lennard-Jones (6/9) term, an electrostatic contribution obtained by the interaction of electropoles centred on the oxygen atom of the water molecules which includes terms as high as the octupole-dipole interaction and a polarisation energy. Each water molecule dipole experiences the field of all the surrounding atoms, so that it is increased from the monomer value (1.85 Debye) to a value which depends on the environment [23].

Periodic boundary conditions are applied to the simulations in order to simulate an infinite crystal lattice using the minimum image convention along with a distance cutoff of 7.5 Å. The solute molecule has all its atoms defined explicity, using the experimental crystal structure conformation, and is kept fixed and rigid throughout the simulation. The water molecules are placed at their initial positions and allowed to move randomly independent of any symmetry constraints.

The potential energy of the systems is shown in Figure 1a and 1b as a function of the number of configurations. Examination of Figures 1a and 1b shows that the initial equilibration occurred after approximately 100,000 and 200,000 configurations for SIMG and SIMR simulations respectively. Although each simulation is extended to 600,000 configurations (i.e. 50,000 trial moves per water molecule), detailed examination of the plots in Figure 1a and 1b show that the potential energy is still decreasing slightly. Two analysis ranges were chosen for each simulation in order to look for changes in properties as a function of the simulation length as shown in Table 3.

Table 4

Name	*Initial RMSD*[a] *(Å)*	*Final RMSD*[b] *(Å)*	*SIM RMSD*[c] *(Å)*
SIMG1	0.0	0.45	0.45
SIMG2	0.0	0.60	0.60
SIMR1	1.72	1.11	1.21
SIMR2	1.72	1.12	1.33

(a) Root mean square deviation of initial to experimental water molecule oxygen sites.
(b) Root mean square deviation of final to experimental water molecule.
(c) Root mean square deviation of final deviation of final to initial water molecule oxygen sites.

RESULTS

Net Movement

From the comparison of the simulated mean water oxygen positions with the experimental water positions, it is possible to obtain an estimate of how far the water molecules have moved towards or away from the 'true' (i.e. experimental) positions. The root mean square deviation between the simulated and the experimental crystal hydrate water molecule oxygen positions both for the initial configuration (Initial RMSD) and also over the 100 *K* configurations in each analysis range (Final RMSD) are given in Table 4. The root mean square deviation for the water molecules during the simulation from their initial positions, during the simulation, is also shown (Sim RMSD).

During SIMG1 and SIMG2 simulations, the water molecules move 0.45 Å and 0.6 Å respectively away from the initial positions. Because the initial sites are the experimental positions, the water molecules finish 0.6 Å RMSD away from their true positions after 600 K configurations. During the SIMR1 and SIMR2 simulations, the water molecule sites move by 1.21 Å and 1.33 Å respectively away from their initial random positions i.e. considerably more than in SIMG1 and SIMG2. However, the water molecules start 1.72 Å away from the experimental positions and finish only 1.12 Å RMSD away from their true positions after 600 *K* configurations. Although the water molecules move further during the simulations SIMR1 and SIMR2, they still finish further away from the true positions than water molecules in the SIMG1 and SIMG2 simulations.

Fluctuations

The average deviation around the mean simulated oxygen position (the fluctuation) can also be calculated and compared to the isotropic temperature factor obtained

Table 5

Simulation	*Fluctuations* (Å)		
	Water 1	*Water 2*	*Water 3*
Experimental	0.18 ± 0.01	0.18 ± 0.01	0.19 ± 0.01
SIMG1	0.00 ± 0.003	0.12 ± 0.06	0.14 ± 0.03
SIMG2	0.02 ± 0.01	0.19 ± 0.10	0.15 ± 0.04
SIMR1	0.12 ± 0.03	0.14 ± 0.02	0.17 ± 0.05
SIMR2	0.06 ± 0.03	0.18 ± 0.11	0.15 ± 0.09

from the crystallographic data. The fluctuation of each water molecule in the asymmetric unit, averaged over all asymmetric units in the unit cell, is shown in Table 5. In all simulations the fluctuation for water molecule 1 was found to be smaller than that calculated from the isotropic temperature factor. However, for the other two unique water molecule sites the fluctuations are the same, within the estimated error, as the experimental value calculated from the isotropic temperature factor.

Contacts

We define a contact as the presence of an atom (solute or solvent oxygen) within a 3.25 Å radius of any one of the water oxygen atoms within the unit cell within the analysis range under consideration. These contacts can be summed over the whole unit cell for water–water and water–solute contacts separately as shown in Table 6. All simulations show remarkable agreement with each other and with the experimental data for the total number of water–solute contacts. The water–water contacts are also very close to the experimental value of 23 for SIMG1 and SIMG2 and only slightly lower for the random start simulations SIMR1 and SIMR2.

Further analysis shows that this agreement in overall number of contacts is fortuitous. When the type of contact is considered, as in Table 7, it can be seen there is a consistent underestimation of the number of contacts that water molecules make with carbon, nitrogen and polar hydrogen atoms but an overestimation of the contacts to solute oxygen atoms. Finally, the total number of discrepancies in the type of contact are shown in the final column of Table 7 (irrespective of sign). This shows that simulations SIMR1 and SIMR2 lead to considerably greater discrepancies in the number and type of water–solute contacts compared with experiment than those obtained from SIMG1 and SIMG2.

Table 6

Simulations	*Contacts water–solute*	*Contacts water–water*
Experimental	152	24
SIMG1	153	22
SIMG2	152	24
SIMR1	153	18
SIMR2	153	20

Table 7

Simulation	*a*	*b*	*c*	*d*	*e*	*f*	Δ
SIMG1	−3	0	6	−1	0	−1	11
SIMG2	−2	0	6	−3	1	−2	14
SIMR1	−4	1	11	−8	3	−3	30
SIMR2	−1	1	12	−11	5	−5	35

a – all carbons other than carbonyl
b – carbonyl carbon
c – oxygen
d – nitrogen
e – apolar hydrogen
f – polar hydrogen
Δ – total number of differences

Table 8

Simulation	*Agreement factor AG(I)*		*Agreement factor AG(II)*	
	(a)	*(b)*	*(a)*	*(b)*
SIMG1	0.25	0.24	0.22	0.25
SIMG2	0.33	0.34	0.34	0.37
SIMR1	0.59	0.50	0.51	0.32
SIMR2	0.62	0.51	0.53	0.33

(a) water–solute interactions
(b) water–water interactions

Agreement Factors

Agreement factors I (AG(I)) and II (AG(II)) have been devised to give one overall number for the agreement or lack of agreement between the contacts within a 3.75 Å sphere made by the water molecule oxygen atoms. AG(I) refers to the root mean square deviation in contact distances between the simulated water networks in each simulated asymmetric unit and the average simulated structure, i.e. it is a form of internal error estimate. The lower the value of the agreement factor the better the agreement between simulation and experiment. As shown in Table 8, the agreement factors are lower for SIMG1 and SIMG2 than for SIMR1 and SIMR2 indicating that the former two simulations are in better agreement with experimental than the latter. Moreover AG(I) and AG(II) are similar for SIMR1 and SIMR2 in magnitude indicating that the agreement with experiment is the same as that between asymmetric units. This is not the case in SIMR1 and SIMR2 in which the agreement with experiment is much worse than that between asymmetric units.

Hydrogen Bonding

The hydrogen bonding patterns exhibited by the water molecules have also been

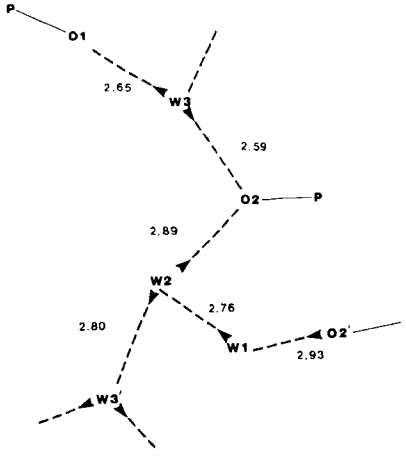

Figure 2 A schematic representation of the solvent hydrogen bonding network obtained from the experimental crystallographic data. The arrows represent the direction of the hydrogen bonds and all distances are given in Å.

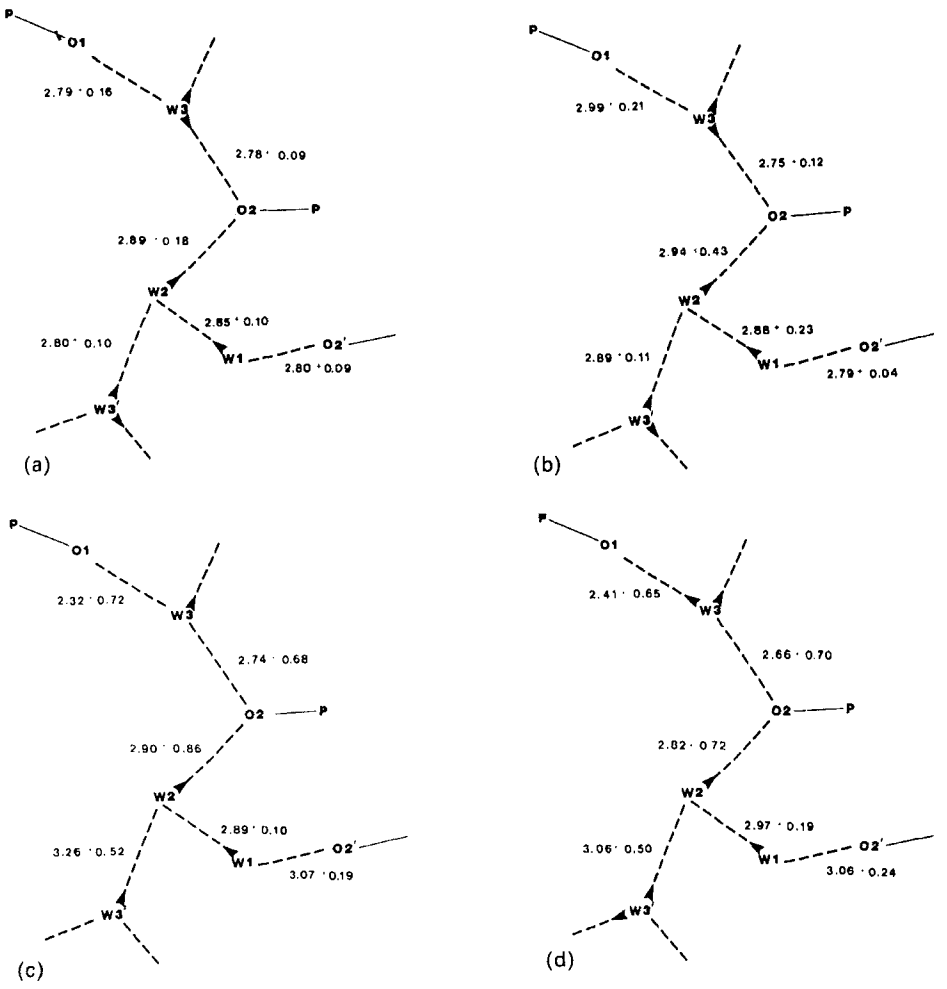

Figure 3 A schematic representation of the solvent hydrogen bonding network from simulations (a) SIMG1, (b) SIMG2, (c) SIMR1 and (d) SIMR2. The arrows represent the direction of the hydrogen bonds and all distances are given in Å.

examined to assess whether they are preserved during the simulations. The experimentally determined solvent hydrogen bond network are shown schematically in Figure 2 whereas Figures 3a–d are schematic diagrams of the water network found in each of the simulation. The simulations SIMG1 and SIMG2 show good agreement with experiment as can be seen by comparing Figures 3a and b with Figure 2. SIMR1 (Figure 3c) leads to only 4 out of 6 of the main hydrogen bonds being reproduced with one of the others being too short and the other too long. However, the network in SIMR2 (Figure 3d) is much improved with all hydrogen bond distances being acceptable.

Arrows have been drawn in Figures 3a–d when the H–O distance is between 1.5–2.5 Å and the OH–O angle is greater than 150°. Comparison of Figures 3a–d with

the experimental data in Figure 2 shows that the hydrogen atom positions are not particularly well reproduced in any of the simulations.

DISCUSSION

This study shows that the effect of random initial positions for the solvent molecules depends on the property of the system under investigation. The overall solvent hydrogen bonding pattern (based on water oxygen to water oxygen or solute non-hydrogen atom distances) could be successfully simulated with either random or experimentally determined initial positions. Little or no difference was seen in either the water molecule fluctuation or the total number of contacts made by the water molecules with the solute atoms. However, a more detailed analysis involving the type, as well as number, of contacts and the solvent hydrogen atom positions showed that the discrepancy with experimental data was greater with random rather than experimental initial solvent positions. Moreover, the agreement factor analysis showed that the comparison between simulated and experimental data (AG(I)) was less that the comparison between the four simulated asymmetric units when the initial solvent positions. Moreover, the agreement factor analysis showed that the comparison between simulated and experimental (AG(I)) was less that the comparison between the four simulated asymmetric units when the initial positions were the experimentally determined ones. This implies that the agreement with experiment is within the error of the simulated results. The analysis of root mean square deviations during the simulations and between experiment and simulation showed that although the water molecules in random start simulations move further than those in the SIMG simulation, they still remain further away, on average, from their true positions.

This study also included data on effect of extending the equilibration period from just over 8300 and 16600 moves per water for SIMG1 and SIMR1 respectively to 50000 moves per water molecule for SIMG2 and SIMR2. Again, we see that some properties such as overall hydrogen bond pattern, water molecule fluctuations and total number of contacts remain unchaged. Other properties such as root mean square deviation from experiment, discrepancies in type of contacts and agreement factors all get worse.

The overestimate of water–oxygen atom contacts and the underestimate of the carbon, nitrogen and polar hydrogen atom contacts are most likely due to errors in the non-bonded potential energy functions. The difficulty in reproducing the correct solvent hydrogen atom positions may well be due to defects in the potentials being used to model water itself. All water models have difficulty in quantitatively reproducing the experimentally determined $g_{O\text{-}O}$ (the partial pair correlation function for water molecule oxygen atoms) and the details of the other partial pair correlation functions $g_{O\text{-}H}$ and $g_{H\text{-}H}$ (from experiment) are still unclear [9].

There are some obvious limitations which may inhibit movement of water molecules during the simulation and hence prevent the randomly positioned water molecules from reaching their true positions. These include the use of fixed solute atomic positions and that the symmetry inherent in the crystal system may impose packing constraints. It was also noted that the potential energy does not appear to equilibrate particularly quickly in the simulation with random starting positions. Obviously, the speed of equilibration could be improved by intially using an energy minimization technique before proceding to the full computer simulation. However, it is likely

that such a method would drive the ensemble into the closest local minimum from which it might have more difficulty in escaping in order to reach the experimental structure.

The simulations described here show that the overall solvent hydrogen bonding patterns can be reproduced even when the water molecules are initially positioned randomly. If more sophisticated criteria are used in the comparison of the simulated and experimental data, then discrepancies become apparent. Therefore, it appears that the simulations can reproduce overall features of solvent hydrogen bonding but that some details are not reproduced satisfactorily. This latter result may be due to defects in the potential energy functions or to hinderance of movement of water molecules in small crystal hydrates.

ACKNOWLEDGEMENTS

We wish to thank the Science and Engineering Research Council for a studentship to P.L.H. and for project grant GR/C/65704.

References

[1] W. Marayama, N. Nagashima & Y. Shimizu. 'The Crystal and Molecular Structure of Guanosine-5'-Phosphate Trihydrate', *Acta Cryst* **B25,** 2236 (1969).
[2] J.T. Edsall & H.A. Mackenzie, "Water and Proteins I', *Adv. Biophysics* **10,** 137 (1978).
[3] J.T. Edsall & H.A. Mackenzie, "Water and Proteins II', *Adv. Biophysics* **16,** 53 (1983).
[4] J.L. Finney, 'Solvent Effects in Biomolecular Systems and Processes', *J. de Physique C7 suppl.* **109,** 197 (1984).
[5] I.D. Kuntz & W, Kauzmann 'Hydration of Proteins and Polypeptides', *Adv. Prot. Chem,* **28,** 239 (1974).
[6] J.L. Finney, J.M. Goodfellow, & P.L. Poole, 'The Structure and Dynamics of Water in Globular Proteins' in Structural Molecular Biology Eds, D. B. Davies et al, Plenum, N.Y. p387 (1982).
[7] J. Texter, 'Nucleic Acid-Water Interactions', *Prog. Biophys. Molec. Biol.* **33,** 83 (1978).
[8] J.L. Finney, J.M. Goodfellow, P.L. Howell & F. Vovelle, 'Computer Simulation of Aqueous Biomolecular Systems', *J. Biomol. Struct. Dyn.* **3,** 599 (1985).
[9] J.L. Finney, J.E. Quinn & J.O. Baum, 'The Water Dimer Potential Surface', *Water Science reviews* **1,** 33 (1986).
[10] P. Rossky & M. Karplus, 'Solvation, A Molecular Dynamics Study of a Dipeptide in Water' *J. Am. Chem. Soc.* **101,** 1913 (1979).
[11] D.L. Beveridge, M. Mezei, P.K. Mehrotra, F.T. Marchese, V. Thirmulai & G. Ravi-Shanker, 'Liquid State Computer Simulations of Biomolecular Solvation Problems' *Ann. N.Y. Acad. Sci.* **367,** 108 (1981).
[12] J.M. Goodfellow, 'Computer Simulation of Hydration Networks around Amino Acids', *Int. J. Biol. Macro.* **9,** 273 (1987).
[13] J.M. Goodfellow, J.L. Finney & P. Barnes, 'Monte Carlo Computer Simulation of Water-Amino Acid Interactions', *Proc. Roy. Soc.* **B214,** 213 (1982).
[14] A.T. Hagler, D.J. Osguthorpe & B. Robson, 'Monte Carlo Simulation of Water Behaviour around a Dipeptide', *Science* **208,** 599 (1980).
[15] A.T. Hagler & J. Moult, 'Computer Simulation of the Solvent Structure around Biological Macromolecules', *Nature* **272,** 222 (1979).
[16] W.F. van Gunsteren, H.J.C. Berendensen & J. Hermans, W.G.J. Hol & J.P.M. Postma, 'Computer Simulation of the Dynamics of Hydrated Protein Crystals and its Comparison with X-ray Data', *Proc. Natl. Acad. Sci.* **80,** 4315 (1983).
[17] E. Clementi & G. Corongiu, 'Simulations of the Solvent Structure for Macromolecules', *Biopolmers* **21,** 763 (1982).
[18] V.C. Singh, S.J. Weiner & P.A. Kollman, 'Molecular Dynamics Simulations of d(C-G-C-G-A), d(T-C-G-C-G) with and without "hydrated" Counterions', *Proc. Natl. Acad. Sci.* **82,** 755 (1985).

[19] M. Metropolis, A.W. Metropolis, M.N. Rosenbluth, A.H. Teller & E. Teller, 'Equation of State Calculations by Fast Computing Machines', *J. chem., Phys* **21,** 1087 (1953).
[20] P. Barnes, J.L. Finney, J.D. Nicholas & J.E. Quinn, 'Cooperative Effects in Water-Biomolecule Crystal Systems' *Nature* **282,** 459 (1979).
[21] B.J. Gellatly, J.E. Quinn, P. Barnes & J.L. Finney, 'Two, Three and Four Body Interactions in Model Water Interactions' *Mol. Phys.* **59,** 949 (1983).
[22] S. Lifson, A.T. Hagler, & P, Dauber, 'Consistent Force Field Studies of Intermolecular Forces in Hydrogen Bonded Crystals' *J. Am. Chem. Soc.* **53,** 4544 (1979).
[23] J.M. Goodfellow, 'Cooperative Effects in Water-Biomolecule Crystal Systems' Proc. *Natl. Acad. Sci.* **79,** 4977 (1982).
[24] J.M. Goodfellow, F. Vovelle, H.F.J. Savage, J.E. Quinn & J.L. Finney, The Effect of the Water Model on the Simulation of Solvent Regions in Vitamin B_{12} Coenzyme Crystal Hydrate', in 'Water and Aqueous Solutions', 37th Colston Symp. Eds. G.W. Neilson & J.E. Enderby, Adam Hilger, Bristol (1986).