Proc. Natl. Acad. Sci. USA
Vol. 79, pp. 4977–4979, August 1982
Biophysics

Cooperative effects in water–biomolecule crystal systems

(Monte Carlo computer simulation/water–protein interactions/potential energy functions/nonpair additive effects)

JULIA M. GOODFELLOW

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

Communicated by John T. Edsall, May 24, 1982

ABSTRACT **Monte Carlo computer simulation techniques have been used to model non-pair-additive (cooperative) effects in the water organization around several biomolecules. Although most models for water assume pair-additive potentials, both quantum mechanical calculations and experimental data indicate that cooperative effects are not negligible in hydrogen-bonded systems such as water. The many-body polarizable electropole (PE) model for water is used to examine the extent and the consequences of this cooperative behavior in several biomolecule hydrate crystals. Increases in the dipole moments of water molecules are predicted in all systems studied so far and can be as much as 50% more than the monomer value of 1.855 debyes. The average value of the individual dipole moments for any one system differs from that of another system and, therefore, should be considered a property of the system and not of the water molecule itself. When this previously calculated average value of the dipole moment for water molecules in a given system is used as a fixed parameter in the simulation, we find differences between this fixed calculation and the original unfixed simulation. An alternative procedure, which allows for a spread in dipole moments and is not dependent on a predetermined average value, has been developed to make simulations of large water–protein systems, including cooperative effects, computationally feasible.**

Most interatomic potential functions used in energy calculations are of a pairwise form in which it is assumed that many-body effects are either negligible or that they can be accounted for by an effective pairwise function (1). However, it has been found that many-body effects in water are not negligible (2). Both Hankins *et al.* (3) and Del Bene and Pople (4) found there were non-pair-additive contributions in hydrogen bond formation in the water trimer compared with the water dimer. Campbell and Mezei (5) have developed a quantum mechanical non-pair-additive model with which they find large deviations from pairwise additivity in their investigations of various forms of ice. More recently, Clementi *et al.* (6), using sophisticated basis sets, have studied the nonadditivity of interactions in the water trimer and have found that the nonadditive contribution is smaller than that calculated previously by Del Bene and Pople (4).

Non-pair-additivity is a consequence of many-body interactions at the molecular level and depends on both the properties of the molecule and its environment. The polarizability of the electronic charge is one molecular property that is important in the interactions between water molecules and gives a large contribution to the high value for the macroscopic property of the static dielectric constant of bulk water. One result of the polarization of water molecules in bulk water is that the dipole moment of each molecule is increased above the value found in the monomer. For example, in hexagonal ice I, the dipole moment was found experimentally to be between 2.6 and 3.0 debyes (D) (7), which agrees with the simulations using the polarizable electropole (PE) model which give a value of 2.88 D (2). Calculations on liquid water with the PE model result in an average value of the dipole moment of 2.5 D (2). This enhancement of 50% over the monomer value and the considerable spread in values from 1.9 to 3.1 D indicate that non-pair-additivity is important in this system.

In previous work which investigated structural aspects of water–biomolecule hydrate crystals, we extended the PE model for water (8) to describe water–protein interactions. In this paper, we investigate the extent of non-pair-additive effects, using Monte Carlo simulations on a variety of water–biomolecule systems. We also study the effects of computational approximations to the full polarizable model in order to make simulations of large water–protein systems computationally feasible, while still taking into account cooperative effects.

METHODS

PE Model. The PE model of water has been developed in order to model many-body effects by a polarizable dipole formalism (9). The intermolecular energy is a function of three terms: an electropole expansion, nonbonded Lennard–Jones interaction, and the dipole polarizability. This latter term is due to the fields of the surrounding molecules acting on each individual water molecule to polarize it and, hence, enhance its dipole moment above the monomer value. This enhancement is calculated iteratively for all molecules in the system, leading to a self-consistent set of fields and induced dipole moments. This is in contrast to pair-additive models in which the dipole moment is fixed at a constant value that is used not only for individual water molecules within one system but also for water molecules in different systems. The magnitude of the induced dipole moment and its dispersion are indications of the importance of non-pair-additive effects in the system being simulated.

Intermolecular Potentials. The PE model has been extended previously to describe water–amino acid interactions by Goodfellow *et al.* (8). Each solvent–amino acid interaction is described by a set of nonbonded (Lennard–Jones) coefficients together with a set of partial atomic charges centered at each nonhydrogen atom of the amino acid residue. The Lennard–Jones coefficients were determined from an extensive series of Monte Carlo simulations on small amino acid hydrate crystals in which the coefficients were varied until optimal agreement between the predicted and experimental crystal structure was obtained. The partial atomic charges were obtained from quantum mechanical calculations on overlapping molecular fragments at the complete neglect of differential overlap (CNDO)/2 level, a method which may lead to inaccurate values for individual atomic charges (10, 11). The methods described in refs. 10 and 11 may be preferable, but we have not yet applied them to the systems discussed here.

Monte Carlo Simulations. Our Monte Carlo computer simulation program, implemented on the CRAY-1 computer at Daresbury Nuclear Physics Laboratory, uses the usual Metrop-

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

Abbreviations: D, debye(s); PE, polarizable electropole.

olis algorithm (12) for generating a Markov chain of accessible configurations. Simulations were carried out on a variety of biomolecule crystal hydrates (13–18) with a large range in the number of water molecules as shown in Table 1. All simulations were carried out at 20°C, and in each system a full unit cell of molecules was used together with periodic boundary conditions. Averages were calculated after each system had equilibrated, as shown by the oscillation of the potential energy within 1–2% of its average value. Because a whole unit cell was simulated, mean values for any parameter were calculated by averaging over all asymmetric units.

RESULTS

Enhancement of Dipole Moments. In all systems simulated, an increase was found in the average dipole moment of water molecules compared with the monomer value of 1.855 D. Table 1 summarizes the results for several systems, which include amino acids, carbohydrates, and nucleotides and which show a wide range of hydration. The lowest value was 2.11 ± 0.04 D, which was found in the simulation of azidopurine monohydrate. The largest enhancement was seen in the serine monohydrate crystal, with the resultant dipole moment being 2.77 ± 0.13 D.

Range of Dipole Moments. In the simulation of bulk water, a large range of dipole moments was found from 1.9 to 3.1 D, with an average value of 2.5 D. In L-arginine dihydrate (the smallest system studied with more than one unique water position), the values of the dipole moments were not significantly different from each other, being 2.43 ± 0.09 D and 2.49 ± 0.09 D. In homoproline tetrahydrate, we found that the values for each unique water varied between 2.45 and 2.58 D, but there was still no significant difference when the standard deviations over the four asymmetric units were taken into account. In α-cyclodextrin, with six unique water molecules, the differences among values of the dipole moments were just significant. When the size of the solvent network was increased, we found a range of values (as for dCpG–proflavin complex) in which the dipoles varied from 1.9 to 2.7 D. A similar range (1.9–2.9 D) was found in the simulation of vitamin B_{12} coenzyme, which has 68 water molecules in the unit cell (H. F. C. Savage, private communication).

Table 1. Dipole moments for water molecules in several biomolecule hydrate crystals

System*	Water molecules, no.†	Dipole moments, D
Water monomer	1	1.855
Water bulk	216	2.5 (range 1.9–3.1)
Serine monohydrate (13)	4	2.77 ± 0.13
Azidopurine monohydrate (17)	4	2.11 ± 0.04
L-Arginine dihydrate (14)	8	2.43 ± 0.09
		2.49 ± 0.09
Homoproline tetrahydrate (15)	16	2.58 ± 0.03
		2.45 ± 0.23
		2.48 ± 0.23
		2.46 ± 0.20
α-Cyclodextrin hexahydrate (16)	24	2.32 ± 0.11
		2.15 ± 0.04
		2.07 ± 0.11
		2.18 ± 0.08
		2.15 ± 0.09
		2.09 ± 0.10
dCpG–proflavin complex heptadecahydrate (18)	100	2.41 (range 1.9–2.8)

*The numbers in brackets refer to the references in which each structure is described.

†Number of water molecules per unit cell for simulation.

Fixed Dipole Moments. One of the characteristics of an effective pair potential is the use of an average dipole moment for the water molecule that is independent of the system being studied. The differences in the mean value of the dipole moment found for water molecules in the range of systems shown in Table 1 indicates that this is not a possible approximation to make in our model. One approximation which may be valid is the use of an average dipole moment for each system. In this case one assumes that the average value is shown in advance of the simulation and that the spread in the dipole moments is not important. In order to look at the consequences of such an approximation, calculations were carried out on previously studied systems but with the dipole moment as a fixed parameter to be set at the average value relevant to a given system.

The results of fixing the dipole moments at their average value for each system is shown in Table 2. For serine, with only one unique water molecule position, there was, as expected, no difference in the fixed energy (i.e., energy with the dipoles fixed at average values) compared with the unfixed energy (i.e., the original simulations with solvent dipoles at their individual values). For L-arginine dihydrate, the energy difference (i.e., difference in energy with fixed and unfixed dipoles as a percentage of the unfixed energy) was 2.24%, which is hardly significant when compared with the standard deviation of 0.09% around the final energy. However, the energy differences for α-cyclodextrin and dCpG–proflavin were 9.6% and 4.45%, respectively. Both values were greater than the standard deviations (see Table 2). In comparison, for bulk water with 216 water molecules, the energy difference was 1 kcal $mole^{-1}$ or 13% of the final energy.

Updating the Dipole Moments. As the use of fixed average dipole moments for each system requires a knowledge of their value before the calculation and also ignores any effect due to the spread in dipole values, an alternative procedure was developed that considerably increases the speed of the energy calculation but also takes into account non-pair-additive effects. This procedure was based on the observation that the individual dipole moments did not vary much from configuration to configuration, even in the initial stages of the simulation. This led to a method in which the individual dipole moments were calculated initially, were fixed at their individual values for several configurations, but then were updated again according to their environment after this set number of "fixed" configurations. Table 3 summarizes a series of simulations in which the frequency of updating the dipole moments was altered until there was no significant difference between this updating procedure and the original simulation.

For serine monohydrate with no water–water molecular interactions at less than 3.4 Å, there was no difference between the predicted solvent organization when the dipoles were updated at a frequency as low as 1 in every 500 configurations and the original unfixed simulation (i.e., updating the dipoles every configuration). The simulations on L-arginine dihydrate showed

Table 2. Effect of fixed dipoles

	Energy, kcal $mole^{-1}$		%
	Unfixed	Fixed	difference*
Serine	−3.47 ± 0.05	−3.42	1.44
L-Arginine	−2.23 ± 0.02	−2.18	2.24
α-Cyclodextrin	−0.405 ± 0.016	−0.366	9.6
dCpG–proflavin	−2.47 ± 0.01	−2.36	4.45
Bulk water	−7.904	−6.874	13.03

*Difference in energy with fixed and unfixed dipoles as a percentage of the unfixed energy.

Table 3. The effect of updating individual dipole moments

Reciprocal frequency of updating	Energy, kcal mole^{-1}	Dipole, debye	d,* Å	B,† Å	Coordination no.‡
(a) Serine monohydrate					
	−3.47 ± 0.05	2.77 ± 0.13	0.33	0.21	Correct
50	−3.47 ± 0.06	2.73 ± 0.14	0.28	0.17	Correct
200	−3.53 ± 0.04	2.77 ± 0.04	0.27	0.17	Correct
500	−3.50 ± 0.04	2.74 ± 0.16	0.33	0.16	Correct
(b) L-Arginine dihydrate					
	−2.23 ± 0.02	2.45 ± 0.09	0.57	0.21	Correct
50	−2.18 ± 0.05	2.52 ± 0.12	0.51	0.21	Correct
200	−2.13 ± 0.03	2.51 ± 0.14	0.32	0.21	Correct
500	−2.09 ± 0.02	2.41 ± 0.14	0.84	0.30	Incorrect
(c) α-Cyclodextrin hexahydrate					
	−0.41 ± 0.02	2.11 ± 0.15	0.44	0.22	Correct
200	−0.43 ± 0.01	2.20 ± 0.14	0.63	0.36	Incorrect
500	−0.45 ± 0.01	2.20 ± 0.13	0.75	0.29	Incorrect

Updating occurred once in every n moves, where the value of n is given in the first column.

*d is the root-mean-square deviation between predicted and experimental structure.

†B is the root-mean-square deviation around the final predicted position.

‡Coordination number and type of hydrogen-bonded neighboring atoms at less than 3.5 Å.

little or no difference regardless of whether the updating frequency was either 1 in every 50 or 1 in every 200 configurations compared with the original simulation. However, the results for α-cyclodextrin hexahydrate show that the predicted coordination of the water molecules is not the same for an updating frequency of either 1 in every 500 or 1 in every 200 configurations. This change in coordination is also shown by the increase in the root-mean-square deviation between the predicted and experimental structures. It appears that the larger the water network, the more frequently the water molecule dipole moments must be updated to obtain the same predicted structure as the original unfixed coordination. A frequency of 1 in every 50 configurations has been found to be acceptable for all systems studied so far.

DISCUSSION

A polarizable water model has been used to model the effect of electronic molecular polarization and, hence, to attempt to describe the extent of many-body interactions in several water–biomolecule crystal systems. The enhancement of the water molecule dipole moments has been used as an indication of the significance of these non-pair-additive effects. It has been found that this enhancement of the dipole moments above the monomer value is a property of the system as a whole rather than that of the water molecule itself. This is a reasonable conclusion when one considers that the induced dipole moment is a measure of each molecule's environment and its effect on the electronic charge distribution. It follows from this that the use of one value for the dipole moment of all water molecules in all systems may lead to incorrect conclusions about the systems being simulated.

The use of an average dipole moment for water molecules, with this average being different for different systems, presents two major problems. First, it is necessary to know the value of the average dipole moment in advance of the simulations. Second, this method ignores the possible importance in the spread in the values of the individual dipole moments, which we have found to become significant in the larger systems. Our results show that there is a discrepancy between the final energy of the predicted structure calculated with fixed dipoles and that of the original unfixed simulation for systems with more water molecules than in L-arginine dihydrate. The maximum difference in energy with and without fixed dipole moments was found to be 13% of the original unfixed energy for bulk water with 216 molecules per unit.

Because of the problems encountered when attempts are made to use an average value of the dipole moment of the water molecule, we have developed a new approach that simplifies the energy calculations but that still includes the effects of non-pair-additivity. This method involves fixing the individual dipole moments for several configurations but then allowing them to be updated in the usual fashion every n Monte Carlo moves. By using this procedure, no prior knowledge of their average value is required, and the individual values (not the average) are calculated and used in the simulation. The frequency of updating, required to maintain no significant difference from the original simulation, depends on the size of the water network, but a value of 1 in every 50 configurations was found to be generally acceptable. This updating procedure considerably decreased the computational time required to perform the simulations so that it becomes comparable to that of the more usual pair potentials.

In conclusion, we have used the water dipole moment induced by the field of surrounding molecules as an indication of the significance of many-body effects in water networks in biomolecular crystals. The use of an average dipole moment is problematic because we do not know its value *a priori*, and it does not always lead to correct predictions. An alternative method of updating the dipoles has been developed that increases the speed of the energy calculations and makes the study of large water–protein systems computationally feasible with a potential that includes cooperative effects.

I thank Dr. J. L. Finney for his advice and criticism of this paper and Dr. P. Barnes who wrote the original version of the program used in this study. The work was supported by a grant from the Science and Engineering Research Council.

1. Ben-Naim, A. & Stillinger, F. H. (1972) in *Water in Aqueous Solutions*, ed. Horne, R. A. (Wiley-Interscience, New York), pp. 295–330.
2. Barnes, P., Finney, J. L., Nicholas, J. D. & Quinn, J. E. (1979) *Nature (London)* **282**, 459–464.
3. Hankins, D., Moskowitz, J. W. & Stillinger, F. H. (1970) *J. Chem. Phys.* **53**, 4544–4554.
4. Del Bene, J. & Pople, J. A. (1973) *J. Chem. Phys.* **58**, 3605–3608.
5. Campbell, E. S. & Mezei, M. (1977) *J. Chem. Phys.* **67**, 2338–2344.
6. Clementi, E., Kolos, W., Lie, G. C. & Ranghino, G. (1980) *Int. J. Quantum Chem.* **17**, 377–398.
7. Whalley, E. (1978) *Chem. Phys. Lett.* **53**, 449–451.
8. Goodfellow, J. M., Finney, J. L. & Barnes, P. (1982) *Proc. Roy. Soc. London Ser. B* **214**, 213–228.
9. Barnes, P. (1978) in *Progress in Liquid Physics*, ed. Croxton, C. A. (Wiley, Chichester, England), pp. 391–428.
10. Smit, P. H., Derissen, J. L. & van Duijneveldt, F. B. (1977) *J. Chem. Phys.* **67**, 274–282.
11. Halgren, T. A., Kleier, D. A., Hall, J. H., Brown, L. D. & Lipscomb, W. N. (1979) *J. Am. Chem. Soc.* **100**, 6595–6608.
12. Metropolis, M., Metropolis, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. (1953) *J. Chem. Phys.* **21**, 1087–1092.
13. Frey, M. N., Lehmann, M. S., Koetzle, T. & Hamilton, W. C. (1973) *Acta Crystallogr. Sect. B* **29**, 876–884.
14. Lehmann, M. S., Verbist, J. T., Hamilton, W. C. & Koetzle, T. F. (1973) *J. Chem. Soc. Perkin Trans.* **2**, 133–137.
15. Bhattacharjee, S. K. & Chacko, K. K. (1979) *Acta Crystallogr. Sect. B* **35**, 396–398.
16. Klar, B., Hingerty, B. & Saenger, W. (1980) *Acta Crystallogr. Sect. B* **36**, 1154–1165.
17. Glusker, J. P., van der Helm, D., Love, W. E., Minlein, J. A. & Patterson, A. L. (1968) *Acta Crystallogr. Sect. B* **24**, 359–366.
18. Shieh, H.-S., Beriman, H. M., Dabrow, M. & Neidle, S. (1980) *Nucleic Acids Res.* **8**, 85–97.