lated reaction profile is somewhere between the gas phase and solution profiles represented in Figure 9.

Monte Carlo Simulations of the Solvation of the Dimethyl Phosphate Anion

Giuliano Alagona,[†] Caterina Ghio,[†] and Peter A. Kollman*

Contribution from the Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94143. Received July 6, 1984

Abstract: Monte Carlo (MC) simulations have been carried out for the (DMP) dimethyl phosphate anion $(CH_3O)_2PO_2^-$ in water, using TIPS2 and TIP4P potentials for the H_2O and analogous potentials for the anion. The simulations employed 216 H_2O molecules and 1 DMP, using periodic boundary conditions in an NPT ensemble. Preferential sampling and "umbrella sampling" techniques were employed to analyze the conformational dependence of anion solvation in H_2O. Analysis of the solvation energetics suggested a ΔE solvation for DMP of −65 to −95 kcal/mol, in reasonable agreement with expectations based on analogous small solutes. The partial molar volume of solvation of DMP is calculated to be $\sim 60 \ cm^3/mol$. Analysis of the water structure around DMP suggests four types of water: (1) those tightly bound to the $O^{\delta-}$; (2) H-bonding waters to the ester oxygen; (3) hydrophobically bound waters near the CH_3 group; and (4) bulk waters. The conformational dependence of the solvation has been analyzed in separate simulations on gauche,gauche (g,g) and gauche,trans (g,t) conformations.

One of the most challenging problems in simulations of complex molecules is the determination of an accurate representation of the effect of solvent on the properties of the molecules being studied. In the long term, the goal is to have an accurate description of the ensemble average properties of solute conformations and solvent configurations for a complete representation of the system.

There have been important advances in two avenues which lead us toward this goal. In the first, simulations of the conformations of macromolecules using molecular dynamical methods have begun to allow us to sample conformational heterogeneity in these molecules.[1] Much more recently, such simulations have been carried out, including the crystal environment.[2] Although such simulations only give us relatively short (subnanosecond) representations of macromolecular configurations, the development of Langevin dynamical methods to "average" over less important degrees of freedom and specialized computer hardware to use in simulations should allow significant progress toward the goal to be made in the near future.

The second avenue toward the goal has been to develop more precise models of water and aqueous solutions and to use this information to gain a new qualitative understanding of the nature of solvation as well as perhaps helping to create simple solvation models which can then be used in simulations of macromolecules.

Important progress in the development of models of water and aqueous solutions has been made because of the molecular dynamical simulations of Rahman and Stillinger[3] and the Monte Carlo simulations of Owicki and Scheraga,[4] Beveridge et al.,[5] Clementi,[6] Berne et al.,[7] Alagona and Tani,[8] and Jorgensen.[9] The above simulations have focused on the nature of water and very simple solutes, such as CH_4, Ar, Cl^-, and Na^+. These simulations have been more successful at describing the nature of water and the hydrophobic effect than they have been in accurately representing solvation energies. Monte Carlo and molecular dynamical

simulations on more complex systems have also been illuminating. The papers of Rossky and Karplus[10] and Hagler et al.[11] have shown that one can derive interesting information on conformational dependent solvation energies on simulations of the alanyl dipeptide. More recently, the use of umbrella sampling techniques has allowed the simulation of the conformational equilibrium of *n*-butane in water.[12,13]

There have been important recent technical developments which should allow more accurate simulations of solvation energies and properties. Even though the initial Owicki and Scheraga papers used in the NPT ensemble,[4] this method has been much less used than the NVT method. Particularly in cases where the partial molar volume of solvation of a species is unknown, it seems more physically reasonable to carry out simulations in an NPT ensemble, which allows the volume of the system as well as its interparticle coordinates to change during the simulation, even though this costs a modest additional amount of computer time. Jorgensen has carried out a large number of simulations on water liquid and has determined potentials (TIPS2[14] and TIP4P[15]) which give an ex-

(1) Karplus, M.; McCammon J. A. *Annu. Rev. Biochem.* **1983**, *52*, 263.
(2) von Gunsteren, W. F.; Berendsen, H. J. C.; Hermans, J.; Hol, W. G.; Postma, J. P. M. *Proc. Natl. Acad. Sci. U.S.A.* **1983**, *80*, 4315.
(3) Stillinger, F. H.; Rahman, A. *J. Chem. Phys.* **1974**, *60*, 1545.
(4) (a) Owicki, J. C.; Scheraga, H. A. *J. Am. Chem. Soc.* **1977**, *99*, 7403. (b) *Ibid.* **1977**, *99*, 7413.
(5) (a) Swaminathan, S.; Beveridge, D. L. *J. Am. Chem. Soc.* **1977**, *99*, 8392. (b) Swaminathan, S.; Harrison, S. W.; Beveridge, D. L. *Ibid.* **1978**, *100*, 5705.
(6) Clementi, E. "Liquid Water Structure"; Springer Verlag: New York, 1976 and references cited therein.
(7) Pangali, C.; Rao, M.; Berne, B. J. *Chem. Phys. Lett.* **1978**, *55*, 413.
(8) Alagona, G.; Tani, A. *J. Chem. Phys.* **1980**, *72*, 580.
(9) Jorgensen, W. L. *J. Am. Chem. Soc.* **1979**, *101*, 2016; **1981**, *103*, 335.
(10) Rossky, P. J.; Karplus, M. *J. Am. Chem. Soc.* **1979**, *101*, 1913.
(11) Hagler, A. T.; Osguthorpe, D. J.; Robson, B. *Science Washington, D. C.* **1980**, *208*, 599.
(12) Rosenberg, R. O.; Mikkilineni, R.; Berne, B. J. *J. Am. Chem. Soc.* **1982**, *104*, 7647.
(13) Jorgensen, W. L. *J. Chem. Phys.* **1982**, *77*, 5757.
(14) (a) Jorgensen, W. L. *J. Chem. Phys.* **1982**, *77*, 4156. (b) Chandrasekhar, J.; Jorgensen, W. L. *J. Chem. Phys.* **1982**, *77*, 5080.

[†]Permanent address: Istituto di Chimica Quantistica ed Energetica Molecolare del C.N.R.—Via Risorgimento 35, 56100 Pisa, Italy.

Table I. Parameters Employed in the MC Simulations[a]

	atom type	$A^2 \times 10^{-3}$	C^2	q
H_2O[b]–TIPS2	O	695	600	0
	M[c]	0	0	-1.070
	H	0	0	0.535
H_2O[b]–TIP4P	O	600	610	0
	M[c]	0	0	-1.04
	H	0	0	0.52
DMP[d]	CH_3	7950	2400	0.109
	OS	500	600	-0.410
	P	6025.894	2195.6	0.912
	O2	230.584	429.5	-0.655

[a] Units: electrons for q, kcal $Å^{12}$/mol for A^2; kcal $Å^6$/mol for C^2. [b] Experimental geometry.[15] [c] M is a point between the H's along the bisector of HOH angle, 0.15 Å from oxygen. [d] DMP charges obtained from the best fit to the electrostatic potential[30] produced by an STO-3G* basis set. CH_3 and OS values are TIPS parameters for CH_3 and O (in ROR'). P and O2 parameters are taken from AMBER. Geometry from Newton.[31]

cellent representation of both the energy (enthalpy) of vaporization and molar volume of pure water.

With these two advances, one can begin to carry out more realistic simulations on fragments of biologically important molecules, and we have chosen for a starting point the dimethyl phosphate anion (DMP)

$$(-) \quad O{\overset{\Phi_1}{\diagdown}}\overset{O}{\underset{P}{\diagup}}\text{-}CH_3$$
$$O{\diagup}\underset{\Phi_2}{\diagdown}O\text{--}CH_3$$
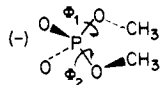
This molecule is interesting in its own right and contains solvation "sites" that are hydrophobic (CH_3), polar (ester O (OS)), and ionic (O^δ-(O2)). Furthermore, it has two torsional degrees of freedom ($\Phi_1 = CH_3$–OS–P–OS, $\Phi_2 = $ OS–P–OS–CH_3), which represent one of the main sources of flexibility of nucleic acids. The question of what is the global conformational minimum of DMP in vacuo has not been definitively settled, but quantum mechanical calculations[16] suggest that the g^+,g^+ ($\Phi_1 \simeq 70°$, $\Phi_2 \simeq 70°$) is the global minimum, with the g,t ($\Phi_1 \simeq 75°$, $\Phi_2 = 180°$) conformations slightly higher in energy and the t,t ($\Phi_1 = 180°$, $\Phi_2 = 180°$) significantly higher. These results are qualitatively consistent with the observations of phosphodiester linkages $ROPO_2OR$ determined by X-ray crystallography, where the observed conformations have been either g^+,g^+ (~80%) or g,t (~20%), with no t,t observed.[17] There is to our knowledge no definitive NMR data on the solution conformation of DMP, but the question remains how solvation will affect the conformational equilibrium of DMP. This question is especially important to address since quantum mechanical calculations using the "supermolecule" approach[18] suggest that, as one adds waters to DMP, the g,t conformation is preferentially stabilized relative to g,g.

To our knowledge, there have been no previous "rigorous" studies of small molecule solvation in which all these types of sites exist and which have been simulated while varying the internal degrees of freedom, using umbrella sampling techniques. A previous, more limited, study of the solvation of the phosphate group with 4 and 10 waters using Monte Carlo techniques has been carried out by Clementi et al.[19] As we will see below, we have been only partially successful in reaching our goal of a "complete" understanding of the conformationally dependent solvation.

(15) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.

(16) Gorenstein, D. J.; Findlay, J. B.; Luxon, B. A.; Kar, D. *J. Am. Chem. Soc.* **1977**, *99*, 3473.

(17) Murray-Rust, P., unpublished analysis of the Cambridge crystal file structure for $ROPO_2OR'$.

(18) (a) Pullman, B.; Pullman, A.; Berthod, H.; Gresh, N. *Theor. Chim. Acta* **1975**, *40*, 93. (b) Pullman, A. In "Environmental Effects on Molecular Structure and Properties"; Pullman, B., Ed.; Reidel: Dordrecht, Holland, 1976.

(19) Clementi, E.; Corongiu, G.; Lelj, F. *J. Chem. Phys.* **1979**, *70*, 3726.
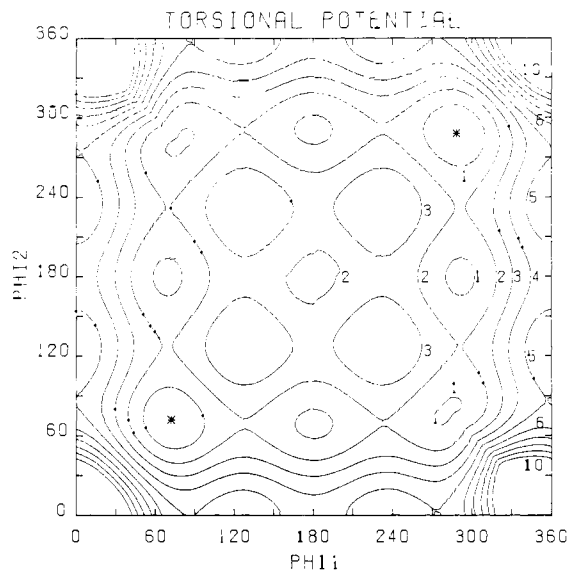
Figure 1. Analytical torsional potential employed in the attempted torsionally variable MC simulation. Contours indicate kilocalories/mole above the minimum energy, marked by stars.

Methods

We began with the final configuration of the simulation of n-butane in 216 TIPS2 waters as reported by Jorgensen and replaced the n-butane with the DMP anion, moving the overlapping waters to the corners. We used the average X-ray (g^+,g^+) internal (R, Φ) geometry for DMP and the experimental geometry for water. The intermolecular potential function between monomer m and n, including Coulomb and Lennard-Jones terms, is

$$\epsilon_{mn} = \sum_{i}^{on\,m} \sum_{j}^{on\,n} \left(\frac{q_i q_j e^2}{r_{ij}} + \frac{A_i A_j}{r_{ij}^{12}} - \frac{C_i C_j}{r_{ij}^6} \right) \quad (1)$$

and the van der Waals and electrostatic parameters are described in Table I.

Due to the presence of additional images of the solute for the use of periodic boundary conditions, we used a solute–solvent and solvent–solvent cutoff distance $R = 8.5$ Å, shorter than half the edge of the box (~9.5 Å). In any case, the solvent molecules interact only with the anion in the central cube.

Our initial goal was to carry out the simulation, varying the two dihedral angles Φ_1 and Φ_2 of DMP, so we carried out molecular mechanics simulations of DMP, using the potential function described in ref 20, constraining the dihedral angles Φ_1 and Φ_2 to various values while optimizing the other internal coordinates of the molecule. We then fit these ($n = 181$) energy points to an analytical function of the following form:

$$\sum_{j=1}^{3} (V_j/2)[(1 - \cos j\Phi_1) + (1 - \cos j\Phi_2)] +$$
$$C_1(1 + \cos \Phi_1 + \cos \Phi_2)^4 \exp(-C_2|\Phi_1 + \Phi_2|) + C_3 \quad (2)$$

The standard deviation for the fit to the molecule mechanics points with Φ_1 or $\Phi_2 > 30° < 330°$ was 0.15 kcal/mol for the following values of the parameters: $V_1 = -0.715\,276$, $V_2 = -1.615\,006$, $V_3 = -2.129\,268$, $C_1 = 0.173\,996$, $C_2 = 0.832\,287$, $C_3 = 7.2922$. The analytical torsional potential shown in Figure 1 represents the intramolecular potential function of eq 2. Thus, the above form of the potential was found after much trial and error, since we needed to represent the steric $CH_3\cdots CH_3$ interaction in terms of internal (dihedral angle) variables. In addition, we use the molecular mechanics energies in which the internal geometry has been allowed to vary to represent the effective energy for internal rigid rotation.

We then began the MC simulation. At first, we kept the DMP rigid in the gauche⁺,gauche⁺ ($\Phi_1 = 75°$, $\Phi_2 = 75°$) geometry, found to be the minimum energy in quantum mechanical calculations and the most frequently observed conformation in crystal structures of dialkyl phosphates. After an initial period of 400K steps, allowing the volume of the unit cell to change from the n-butane value of 8500 to 7000 $Å^3$, we continued the simulation for an additional 200K steps by using a fixed volume.

We then allowed the internal geometry of DMP to vary during the simulation. Because the internal energy is >12 kcal/mol for geometries

Monte Carlo Simulations of Dimethyl Phosphate

J. Am. Chem. Soc., Vol. 107, No. 8, 1985 2231

in which *both* Φ_1 and Φ_2 are between $-30°$ and $30°$, we had not used these geometries in fitting our analytical potential, and so we did not allow torsional motions into these regions of conformational space. However, we chose to sample the remaining torsional angle space by using the method of umbrella sampling, which corrects the statistics on a chopped potential to reduce the barrier height corresponding to unfavorable conformational changes. This approach has been used successfully by Rosenberg et al.[12] and Jorgensen[13] to increase the frequency of g $\rightleftharpoons$ t conformational changes in *n*-butane simulations neat and in H_2O and, thus, to allow the calculation of the equilibrium conformation distribution of th e molecule. We began by chopping the true torsional potential (eq 2) to use the true calculated value for $E < 1.5$ kcal/mol and the value of 1.5 kcal/mol for the remaining Φ_1 and Φ_2 values outside the disallowed region ($-30 < \Phi_1$ and $\Phi_2 < 30°$), where $E = 12$ kcal/mol has been used. After some 400K steps in which there was a reasonable sampling of conformational space, the molecule settled into the higher torsional energy g^+, g^- regions of conformational space and stayed there. We tried various changes in our procedure; e.g., volume moves every 1000 rather than every 300 steps, changes from 1.5 to 6.0 kcal/mol for the energy in the conformations where either Φ_1 or Φ_2 is 0° and the other angle between 30° and 330°, and solute moves every 50 rather than every 100 moves of solvent. Such trials (each of which lasted 100–200K steps) totaled about 1500K. We then switched the conformation of DMP to trans,trans and, after 100K steps with the solute fixed, gradually allowed solute moves and torsional variations; at first, there was sampling of conformational space, but the molecule eventually settled into and remained at cis,gauche and g^+, g^- conformations, even though the total energy of the sytem (-2200 to -2225 kcal/mol) was higher than we later found with fixed molecule simulations. At this point, we decide to abandon the hope of getting a complete sampling of conformational probabilities in DMP and decided to focus our attention on the two major conformations of DMP: g^+, g^+ and g,t. By this time, we had carried out about 3000K Monte Carlo steps. We also learned at this time that TIP4P was a small improvement over TIPS2 and, concurrently switched over to using this potential, which differs from TIPS2 only slightly in van der Waals parameters and partial charges (see Table I). At this point we fixed the conformation at g^+, g^+ ($\Phi_1 = 75°$, $\Phi_2 = 75°$) and at g,t ($\Phi_1 = 75°$, $\Phi_2 = 180°$), carrying out separately two different MC simulations. For the g,g solution, after 1500K steps for equilibration, we evaluated average properties for 1000K steps. The 1500K equilibration appeared to be an adequate length, even if TIP4P is as similar to TIPS2, because changing the conformation from its g^+, g^- value, when we abandoned the torsionally variable simulation, to g^+, g^+ caused some solvent perturbation and rather large energy variations. Moreover, the individual solute group–solvent group radial distribution functions were somewhat different, even though they should be, at least locally, identical by symmetry. In addition, the average energy over the final 1000K steps was significantly lower than over the previous 1000K ($E = 2235.5$ kcal/mol, rms $= 2.8$). We used the same starting water configuration for the g,t conformation. After 1000K steps of equilibration, at which point the energy was oscillating about a mean value, we determined averages by using the next 1000K steps.

Even though the two chains seem to have slightly different lengths (2000K for the g,t simulation and 2500 K for the g,g one), we have taken into account the last 1000K steps for both, when the standard deviations on the energy are very close (2.6 and 2.1 for the g,g and g,t solutions, respectively). However, it appears that to ensure complete convergence of the energy for both simulations, substantially more steps than those we have carried out may be required.

Results

(1) Solute–Solvent Radial Distribution Functions (rdfs). The most widely employed way to analyze the structure of a solution is the perusal of the radial distribution functions around the various atoms. For a pair of atoms i and j, the radial distribution can be defined

$$g_{ij}(r) = (4\pi\rho_j r^2)^{-1} \, dn_{ij}(r)/dr \qquad (3)$$

in which n_{ij} is the number of j atoms (i.e., O or H of water) enclosed within a sphere of radius r centered on i (one of the seven atoms relevant to the solute) and ρ_j is the bulk density of atoms j.

We talk about seven atoms in the solute, because for the CH_3 group a united atom approximation has been used, as one can see from Table I which has all the parameters used in these simulations.

The radial distribution functions have been recorded, using the coordinates saved each 2000 MC steps, by scanning the space
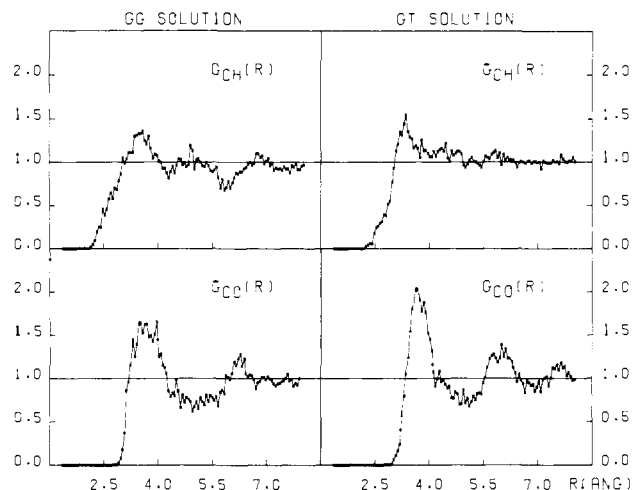
Figure 2. Radial distribution functions $g_{CO}(r)$ and $g_{CH}(r)$ between the first CH_3 group in DMP g,g solution (left part) and g,t solution (right part) and water O (lower part) and H (upper part).

surrounding the molecule with a $\Delta r = 0.055$ Å. As noted above, we found a lack of high symmetry between the equivalent groups in the g,g conformation, likely due to the fact that the bulky solute, while moving inside the cube, might assume positions that are not symmetric with respect to the different surfaces of the cube of solution considered. However, the radial distribution functions for "equivalent" groups are quite similar, even in the case of the g,t conformation where there is no symmetry requirement that they be so. To avoid artifacts due to small differences in the rdfs and to save space, we present only one of the rdfs due to equivalent groups rather than an average, except in special cases.

Let us begin by examining the rdfs of the hydrophobic group, reported in Figure 2. The trend of our CH_3O rdfs is very similar to that found in ethanol[21] and alanyl dipeptide,[10] because of the presence of the other groups in the solute. The center of the broad first peak occurs at about 3.7 Å, close to what was previously found also for spherical solutes.[5b,8,22,23] The noticeable difference between spherical and nonspherical solutes lies in the lower height of the first peak of the latter due to the shielding effect of the other groups present in the molecule that strongly reduce the solid angle available to solvent molecules. An interesting feature to point out is the difference between the $g_{CO}(r)$'s in the g,g solution and in the g,t one as far as the g and t CH_3 are concerned. We can notice a remarkable difference between the height of the first peak in the two cases. As said above, we display the trans CH_3 for the g,t solution, which is located near the two $O^{\delta-}$ halfway between them. Its radial distribution function accounts for this fact, looking more similar to the rdf around the CH_2 that is close to the alcoholic oxygen in ethanol (see Figure 2, ref 21).

As far as the $g_{CH}(r)$'s are concerned, the g,t solution shows an anomalous behavior, exhibiting just one peak slowly sloping toward increasing distance. This behavior is again ascribable to the waters coordinated by the nearby $O^{\delta-}$, which provide a considerably large number of H at a distance where the $g_{CH}(r)$'s would present a minimum.

Another interesting question to address is the geometrical orientation of the waters around the hydrophobic group. To examine this we have recorded the cos θ distribution of waters within 4.7 Å around the methyl groups for both the g,g and the g,t solution. θ is defined as the angle formed by each H with the axis joining the methyl group to the water O, as displayed in Figure 3. It must be remarked that, because of the rather large radius considered, corresponding to the first minimum of $g_{CO}(r)$, to avoid interferences we have not taken into account the waters (see Figure

(20) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. *J. Am. Chem. Soc.* **1984**, *106*, 765.
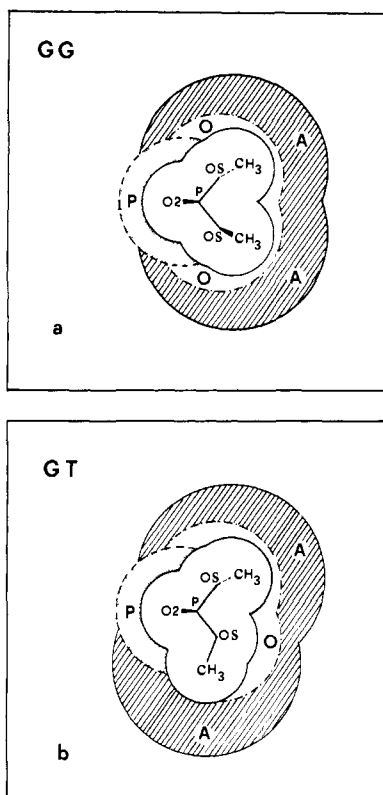(21) Alagona, G.; Tani, A. *Chem. Phys. Lett.* **1982**, *87*, 337.
(22) Bolis, G.; Clementi, E. *Chem. Phys. Lett.* **1981**, *82*, 147.
(23) Okazaki, S.; Nakanishi, K.; Touhara, H.; Adachi, Y. *J. Chem. Phys.* **1979**, *71*, 2421.
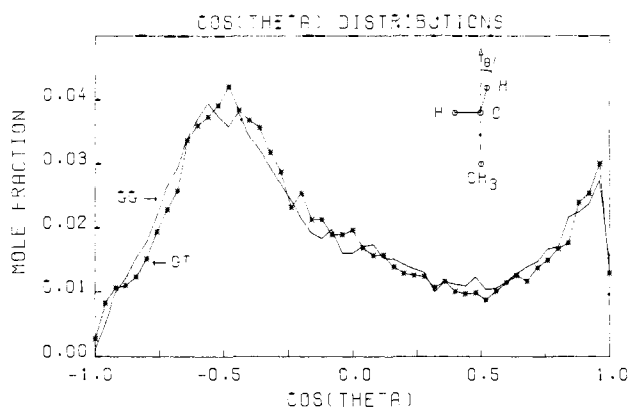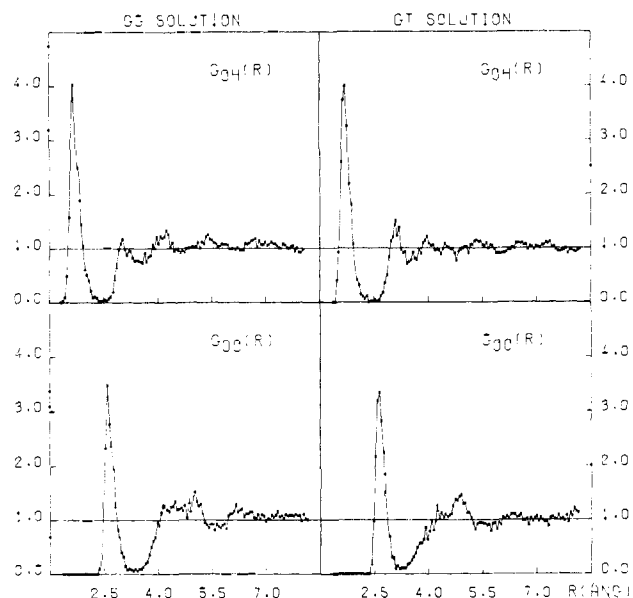
Figure 3. Cos θ distributions for the waters belonging to the apolar region. θ angle defined as depicted. Units for the ordinate are mole fraction per 0.04 cos θ.
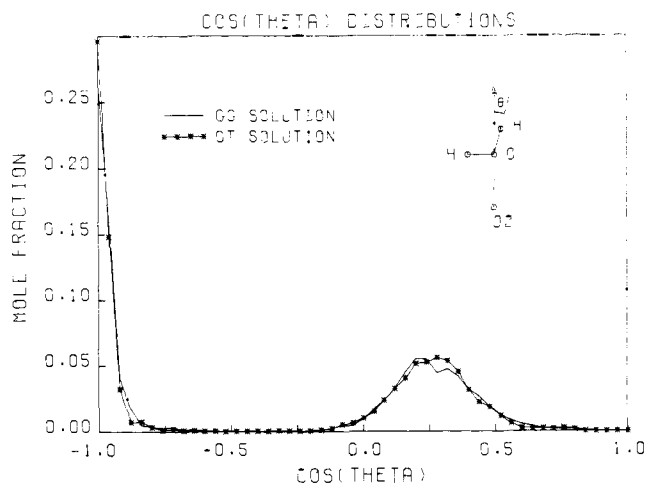

Figure 4. Schematic representation ((a) for the g,g solution; (b) for the g,t one) of the various solvation regions: ionic "P", polar "O", nonpolar "A", bulk all the rest. The blank area immediately surrounding the solute corresponds to its excluded volume.

Figure 5. Radial distribution functions $g_{OO}(r)$ and $g_{OH}(r)$ between the first O2 group in DMP g,g solution (left part) and g,t solution (right part) and water O (lower part) and H (upper part).
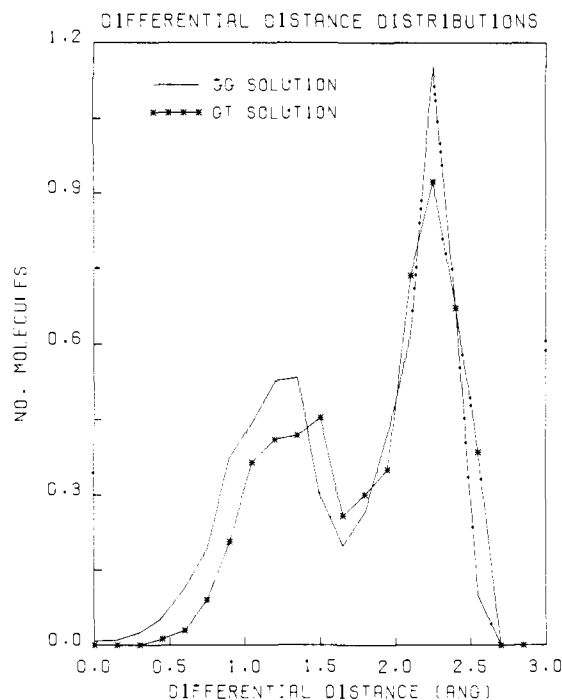

Figure 6. Cos θ distributions for the waters belonging to the polar O2 region. θ angle defined as depicted. Units for the ordinate are mole fraction per 0.04 cos θ.

4) included within 3.2 Å from either O2 or OS. The distribution shows a broad maximum about 120°, which indicates a higher probability to find the water H at angles varying from 105° and 130°. We have another relative maximum near 0°. Thus, we confirm that neither H nor a lone pair of the O points directly toward the methyl group. From the broadness of the peak about 120° with respect to that about 0°, we can assess that the case in which both hydrogens make 120° with the CH$_3$···O axis, and one of the O lone pairs is directed along it, is the most frequent. Both solute conformations (g,g and g,t) show the same trend, which remains analogous even if we consider the angle formed with each H separately.

Further interesting features come from the examination of the rdfs around O$^{\delta-}$ (Figure 5). The maximum of $g_{OO}(r)$ is at about 2.7 Å from O2, the corresponding one of $g_{OH}(r)$ is ~1 Å closer to O2, and since both peaks integrate to about 3, the three waters nearby point one of their hydrogens toward O2. As we will see later on, this fact is also indicated by the analysis of the cos θ

distributions. Also, as expected from the strong partial charge on the O2, all the waters that are close to the O2 are tightly bound, so that the first peak coincides with the rdf for H-bonded water pairs.

Consequently, the first peak is unaffected by the molecular remainder which does, however, affect the rest of the rdf. This is evident for the g,t conformation, where the waters around the methyl group are closer to O2 and mainly oriented with both hydrogens toward it (as we have seen by discussing the cos distribution around the CH$_3$). This is suggested also by the fact that the second peak of the $g_{OH}(r)$ is higher for the g,t solution than for the g,g one and by the second peak of $g_{OO}(r)$ vanishing in the g,t case in a continuous slope. As already noted, the waters interacting with O2 point one of their H's toward O2. This fact turns out clearly from the cos θ distributions (Figure 6) as well.

Both distributions show a sharp peak (maximum) at about –1 and another equivalent maximum at about 0.25 (the different shape is an artifact of the cosine). These cos θ values correspond to $\theta = 180°$ for one H and $\theta = 75.5°$ for the other, exactly consistent with the HOH angle employed. The oscillation about these values is 10–15°, which is as large as the maximum angular variation allowed to water moves during the MC simulations. Since in this solute we have two such O2 groups, it is intriguing to know whether one of the three water molecules makes a bi-

Figure 7. Differential O2–O distance distributions for the waters belonging to the polar O2 region. Units for the ordinate are number of molecules per 0.15 Å.

furcated H bond, that is, the minimum energy position in the "supermolcule" approach for the adduct DMP⁻·H₂O, from an "ab initio" calculation on a 4-31G basis set.[24] In order to check this point, we have computed for each water O belonging to the O2 solvation shell its distance from both O2's and we have recorded the absolute value of the difference between these distances (Figure 7). It can be seen in this way that the probability to find a water molecule forming bifurcated H bonds is zero for the g,t conformation and vanishingly small for the g,g one. However, in its random movement, a water molecule might assume that position but certainly it is not able to hold it for a significant number of steps.

Since the average distance between a water O and the closest O2 obtained from $g_{O2O}(r)$ is ~2.7 Å, the two distinct peaks in the differential distance distributions corresponding to angles O2–O2–O of 100° and 140°, respectively. Such angles correspond to an almost tetrahedral arrangement of the three water O's with respect to the P–O2 bond. The first peak is slighly less populated with respect to the second in the g,t solution ($\Delta n = 0, 44$), while in the g,g the difference is higher ($\Delta n = 1.13$). Since the 140°
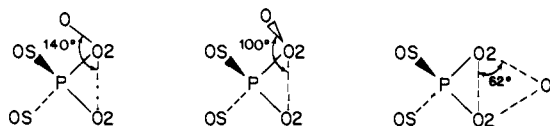
value roughly corresponds to a staggered arrangement of the O2–O with respect to the angle OS–P–OS, while the 100° corresponds to an eclipsed one (~62° would correspond to a bifurcated H bond), we can state that staggered arrangements prevail. Notice that the area below the curves is slightly different because the O2 coordination number is 6.0 for the g,t solution and 5.61 for the g,g.

The group most affected by the molecular environment is the OS, compressed, as it is, between the methyl group and the O2 (Figure 4). Consequently, its rdfs with both O and H (Figure 8) has a very complex pattern, because the waters coordinated by O2 and CH₃ are very close to it. However, the first peak mainly corresponds to a water molecule directly coordinated to OS, be-

(24) Alagona, G.; Ghio, C.; Kollman, P. *J. Chem. Soc.* **1983**, *105*, 5226.
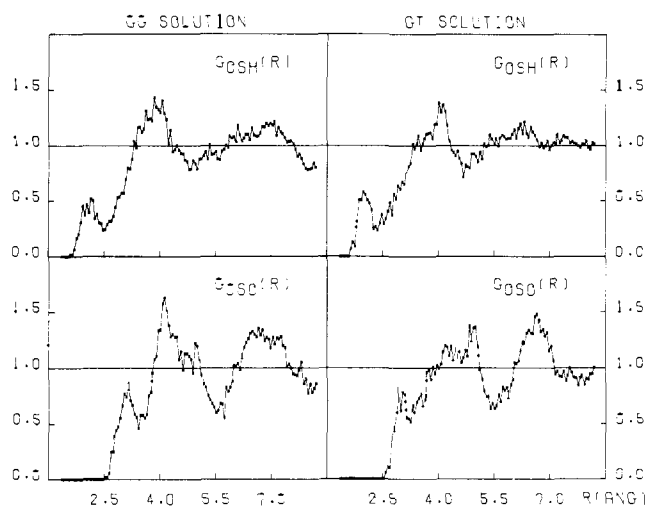
Figure 8. Radial distribution functions $g_{OSO}(r)$ and $g_{OSH}(r)$ between the first OS group in DMP g,g solution (left part) and g,t solution (right part) and water O (lower part) and H (upper part).
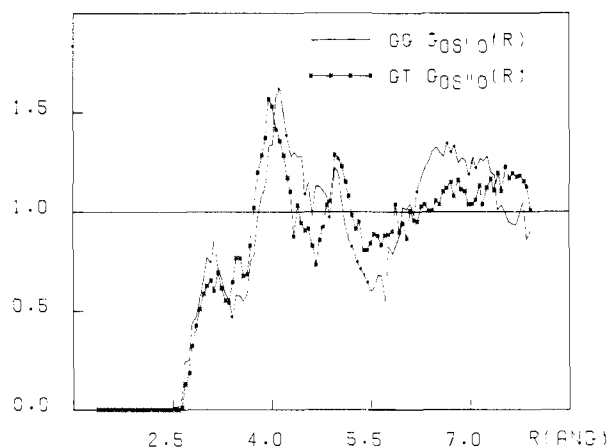
Figure 9. Radial distribution functions $g_{OS'O}(r)$ between the first group in DMP g,g solution and $g_{OS'O}(r)$ between the second OS group in DMP g,t solution and water O.

cause the $g_{OSO}(r)$ up to the first minimum ($R = 3.24$ Å) integrates to 1.25 and up to the second ($R = 5.4$-5.5 Å) 18-19, approximately the total number of the first shell coordinated molecules.

The first peak of $g_{OSH}(r)$ lies 1 Å closer to OS, indicating that, in this case also, the water points one of its H's toward OS. Thus, here we are dealing with an H bond a little weakened by the larger distance between the partners, because of the presence of the other groups and chiefly because of the strong water interaction of the neighboring O2. The difference between the $g_{OS'O}(r)$ of the g,g and g,t conformations is ascribable to the trans conformation of the methyl group near the OS in g,t. If we take a look at the $g_{OS'O}$ (g,g) and $g_{OS'O}$ g,t where both OS come from gauche methyl groups (Figure 9), we can see that the main features of both are similar.

The same considerations done for the rdfs have been repeated for the cos θ distributions (Figure 10). Moreover, because of the shielding effect of the neighboring groups and the excluded volume effect of the strongly coordinated waters, it seems that there is a subtly different arrangement of the solvation water in the g,g solution and in the g,t one, viz.

We have used computer graphics to visualize many of the features described above. Obviously we could not obtain statistics in this way, but it was possible to note many of features that we
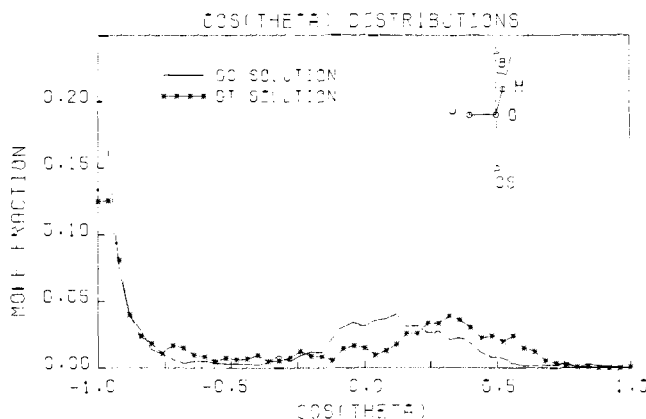
Figure 10. Cos θ distributions for the waters belonging to the polar OS region. θ angle defined as depicted. Units for the ordinate are mole fraction per 0.04 cos θ.

derived later after a long perusal of all the distributions and the thermodynamical results.

First of all, we colored the waters according to their solute–solvent interaction energies. The most strongly interacting waters were placed in all the snapshots around the basis of a cone with the vertex on the O2 (three for each O2), and no water had a bifurcated position between the two O2. There was one water with a little lower interaction energy than the previous ones, taking variable positions near each OS. The large number of interacting waters around the CH₃ prevented us from any analysis.

With the Molecular Surface program of Connolly,[25] we could generate the molecular surface of the system and scan the space searching for empty spaces with a "probe". We have been able to see two small lentil spaces between the methyl group excluded volume and waters accounting for the hydrophobic interaction.

Then, with the movie option of CHEM,[26] we could also simulate "motions" inside the solution by interpolating between various structures.

In the course of such visualization, we noticed one interesting example of a "ligand displacement reaction"; i.e., between 500K and 750K in the g,t simulation, one water coordinated to an O2 moved away from this geometry as an initially "second shell" water took its place. Figure 11 is stereoviews of the coordinates for those waters in the immediate neighborhood of DMP, and, as one can see by comparing them, water 150 is initially coordinated to the anionic O2 and, after 250K steps, water 11 has taken its place near that oxygen. One can also visualize representative configurations of the waters around the OS and methyl groups from this figure.

(2) Solute–Solvent Energy Distributions. The total potential energy E of the solution can be broken down as

$$E = \bar{E}_{SX} + E_{SS} \qquad (4)$$

where, assuming pairwise additivity, the solute–solvent ($\bar{E}_{SX}$) and the solvent–solvent (E_{SS}) contributions are given by

$$\bar{E}_{SX} = \sum_{i=1}^{N} E_{S_iX} \text{ and } E_{SS} = \sum_{i=1}^{N}\sum_{j>i}^{N} E_{S_iS_j} \qquad (5)$$

Once the total energy of an equal number of water molecules as a pure liquid (E°_{SS}) is known, we can calculate the energy of solution (or partial molar internal energy) as

$$\Delta \bar{E}_{sol} = \bar{E}_{SX} + E_{SS} - E^{\circ}_{SS} = \bar{E}_{SX} + \Delta \bar{E}_{SS} \qquad (6)$$

$\Delta \bar{E}_{SS}$ is the variation induced by the solute in the internal energy of the solvent and can be considered as the solvent reorganization energy.

Another interesting property is the partial molar volume

$$\Delta \bar{V}_{sol} = V - V^{\circ} \qquad (7)$$

that is, the difference between the volume of the dilute solution

(25) Connolly, M. QCPE Program No. 427, Bloomington, IN. **1982**.
(26) CHEM, program written by Dearing A. at the UCSF Computer Graphics Lab, 1980–1982.

Table II. Calculated Properties for DMP at 1 atm and 25 °C[a]

property[b]	solution	
	g,g	g,t
E	−2256.8 ± 2.6	−2228.9 ± 2.1
$\bar{E}_{SX}$	−142.7 ± 1.1	−147.3 ± 0.7
E_{SS}[c]	−2114.1 ± 2.9	−2081.5 ± 2.4
$\Delta \bar{E}_{SS}$	50.3 ± 3.7[d]	82.8 ± 3.3[d]
$\Delta \bar{E}_{sol}$	−92.5	−64.5
V[c]	6691.5 ± 13.3	6632.7 ± 8.5
$\Delta \bar{V}_{sol}$	116.1	57.3
$\Delta \bar{H}_{sol}$	−93.1	−65.1
ρ[c]	0.9967	1.0055

[a] Energies are in kcal/mol; V is in Å³; $\Delta \bar{V}_{sol}$ is in Å³/molecule; ρ is in g/cm³. [b] As defined in text (eq 4–8). [c] Pure water TIP4P value[15] E°_{SS} = −2164.4 ± 2.3, V° = 6575.4, ρ° = 0.9827. [d] The standard deviations of the calculated difference quantities are the square roots of the sums of the component variances.

V and that of the pure solvent V°. All the partial molar quantities are indicated by the bar.

From ΔE_{sol} and $\Delta \bar{V}_{sol}$; it is possible to calculate $\Delta \bar{H}_{sol}$

$$\Delta \bar{H}_{sol} = \Delta \bar{E}_{sol} + P\Delta \bar{V}_{sol} - RT \qquad (8)$$

where RT is the PV contribution for the solute in the ideal gas. At 1 atm, $\Delta \bar{H}_{sol} \approx \Delta \bar{E}_{sol} - RT$.

In Table II the calculated quantities for g,g and g,t solutions are reported. The g,g conformer is calculated to be more stable than the g,t one by ∼28 kcal/mol in solution, despite the fact that the solute–solvent interaction energy is ∼5 kcal/mol more favorable for the g,t conformation. This is due to the larger solvent reorganization energy of the g,t solution, which turns out from the greater solvent disruption consequent to the more attractive solute–solvent interaction.

We report in Figure 12 the solute–solvent energy distributions for both solutions. It is evident that the g,t conformation has an average solute–solvent interaction energy more favorable than the g,g one.(Table II), though the g,g conformation has a higher probability of energies enclosed in the range −170 ÷ −162, but their contribution is outweighed by the much higher probability of energies above −142 kcal/mol. On the whole, the g,t solute–solvent interaction has less spread than the g,g one.

The examination of the solute–solvent energy pair distribution (Figure 13) shows that the interaction in the O2 first shell is much better for the g,t than for the g,g: we have to integrate to −7.5 kcal/mol to have all the O2 coordinated water, while it is sufficient to integrate to −8 kcal/mol to encompass all the waters in the g,t solution.

The OS coordinated waters give the small contribution up to −5.5 kcal/mol (g,g) and −6.0 kcal/mol (g,t). The CH₃-coordinated waters fall into the peak about −1.5 kcal/mol that integrates to about 50, whereas the total number of the CH₃ coordinated waters is ∼15.

The small shoulders about 2 kcal/mol integrate to nearly 14 for both the solutions: because of the solvent–solvent H bonding, some molecules of the second shell are forced in an unfavorable position with respect to the solute, and it seems that the better is the interaction with the first shell, the worse it is with the second one.

To have a closer insight of this distribution, we have decomposed the solute–solvent interaction energy, depending on the positions of the solvent molecules (Figure 4). Molecules that lie within a radius of 3.2 Å from each O2 atoms are considered in the polar shell (P), those within 3.2 Å from OS are considered in the less polar OS shell (O), and those with 4.7 Å from CH₃ in the apolar shell (A). The values of the radii coincide with the first minimum of the relevant rdfs.

This decomposition is reported in Table III together with the average contribution by each water molecule (for the waters external to the first shell we have not considered the ∼130 waters that give a zero contribution to the interaction energy).

The more favorable solute–solvent interaction for the g,t solution is mainly due to the O2-coordinated waters and partly to the apolar
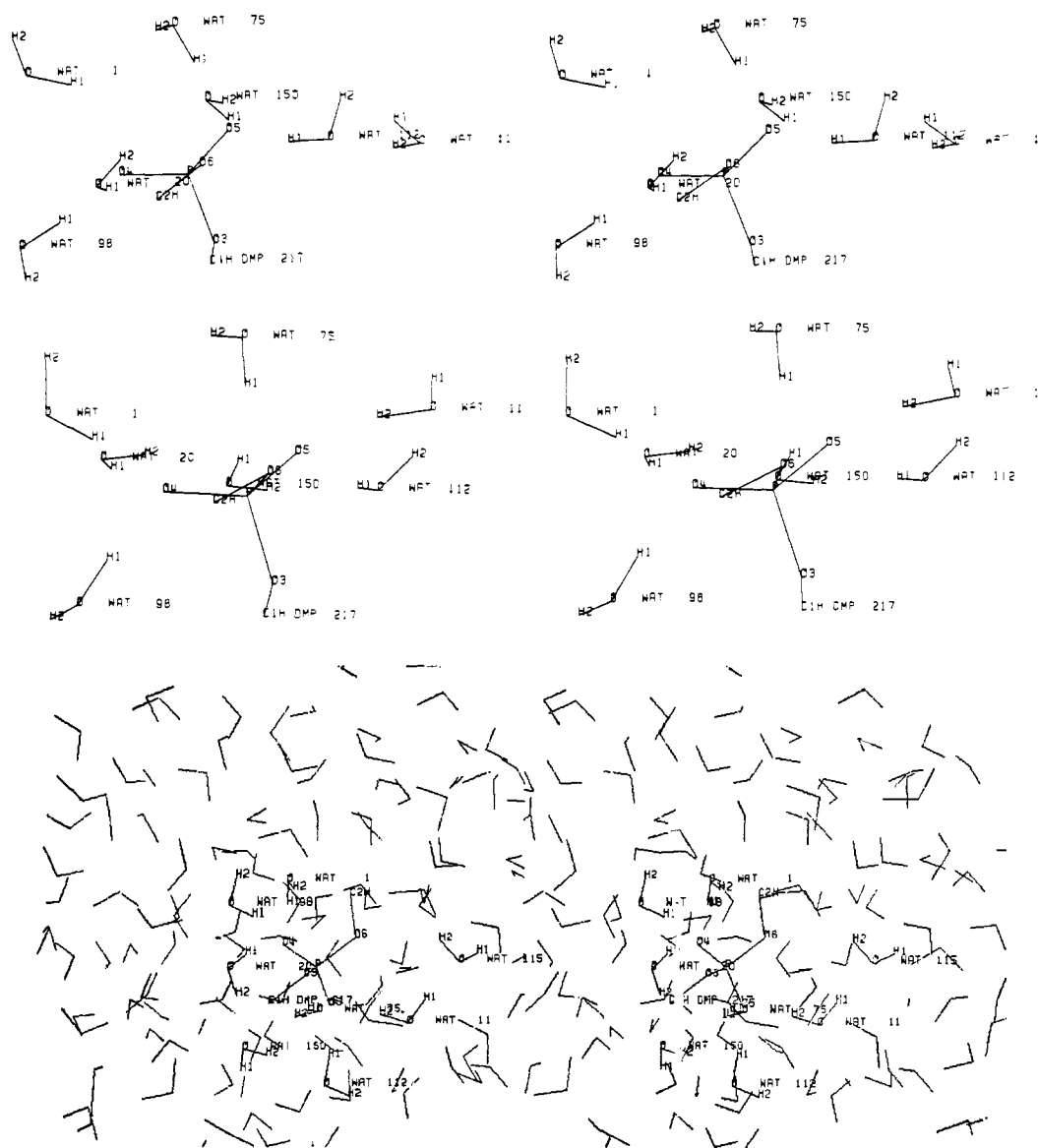
Figure 11. (a) View of the waters near the anionic oxygens of DMP (g,t) after 500K steps; (b) same as (a) after 750K steps; (c) same as (b) with a different view and all the waters included. Water 115 is the one nearest the ester oxygen 06, and one can see the large "gap" in the water structure near the hydrophobic methyl C2H.
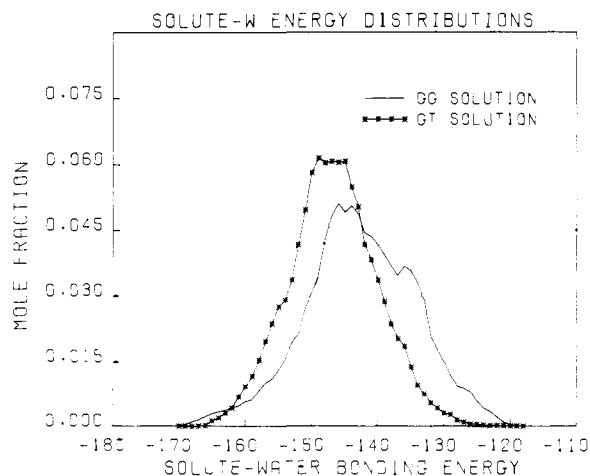
Figure 12. Solute–water energy distributions for g,g and g,t solutions. Units for the ordinate are mole fraction per kilocalories/mole.

Table III. Decomposition of the Solvent–Solute (S–X) Interaction Energy within the First Shell[a]

	g,g	g,t	$(E)_{g,g}$	$(E)_{g,t}$	$N_{g,g}$	$N_{g,t}$
E'_{SXP}	−63.03	−69.70	−11.24	−11.62	5.61	6.00
E'_{SXO}	−8.72	−8.17	−5.82	−5.33	1.47	1.53
E'_{SXA}	−16.66	−17.93	−1.05	−1.19	15.88	15.09
$E'_{SXI}{}^{b}$	−88.41	−95.80			22.96	22.62
$E'_{SXE}{}^{c}$	−54.39	−51.55	−0.86[d]	−0.81[d]	63.04	63.38
			(−0.28)	(−0.27)	193.04	193.38
$E_{SX}{}^{e}$	−142.80	−147.35			216	216

[a] Energies are in kcal/mol; N is the number of water molecules in each zone. [b] The solvent–solute interaction energy with the water molecules in the first shell (E'_{SXI}) has been decomposed in E'_{SXP}, E'_{SXO}, and E'_{SXA}, where SXP, SXO, and SXA indicate the contributions due to waters belonging to the strongly polar or ionic P, polar O, and apolar A regions; $2 \leqslant R \leqslant 3.2$ Å around O2 and OS; $2 \leqslant R \leqslant 4.7$ Å around CH₃. [c] E'_{SXE} is the interaction energy between solute and waters external to the first shell. [d] Average with the exclusion of water molecules (∼130) having 0 kcal/mol interaction energy with the solute. In parentheses the total average is reported. [e] E_{SX} is the overall solvent–solute interaction energy.

interaction. The better interaction with the external- and OS-coordinated waters in the g,g solution is not sufficient to make equal the solute–solvent energies for g,t and g,g.

In Table IV, the solute–solvent interaction energies for the regions just outside the first shell are reported, divided accordingly to the first shell. These interactions account for the trend of the
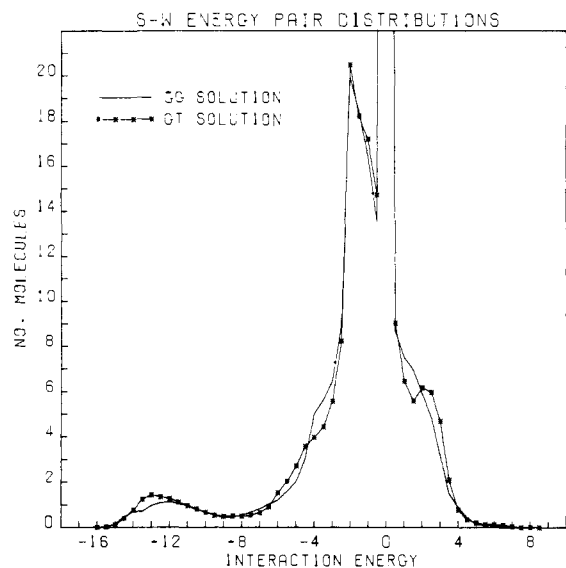
Figure 13. Solute–water pair distributions for g,g and g,t solutions. Units for the ordinate are number of molecules per kilocalories/mole.
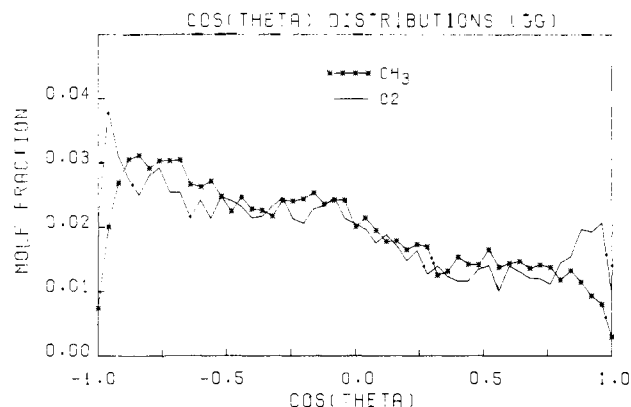
Figure 14. Cos θ distributions for the waters in the bulk g,g solution just outside the O2 and CH$_3$ regions. θ angle defined as in Figures 5 and 7. Units for the ordinate are mole fraction per 0.04 cos θ.

Table IV. Decomposition of the Solute–Solvent Interaction Energy within a Layer in the Bulk, just Outside the First Shell[a]

	g,g	g,t	$(E)_{g,g}$	$(E)_{g,t}$	$N_{g,g}$	$N_{g,t}$
E''_{SXP}	−20.21	−18.15	−1.48	−1.56	13.68	11.63
E''_{SXO}	−0.28	3.40	−0.13	1.25	2.20	2.73
E''_{SXA}	−15.77	−23.08	−0.69	−0.82	22.99	28.19
E''_{SX2}[b]	−36.26	−37.83	−0.93	−0.89	38.86	42.54
E''_{SXE}	−18.13	−13.72	−0.75[c]	−0.66[c]	154.18	150.83
			(−0.12)	(−0.09)		
E'_{SX1}	−88.41	−95.80	−3.85	−4.23	22.96	22.62
E_{SX}	−142.80	−147.35	−1.66[c]	−1.71[c]	216	216
			(−0.66)	(−0.68)		

[a] Energies are in kcal/mol; N is the number of water molecules in each zone. [b] $E''_{SX2} = E''_{SXP} + E''_{SXO} + E''_{SXA}$. See notes in Table III. Here around O2 and OS $3.2 \leqslant R \leqslant 5.5$ Å and around CH$_3$ $4.7 \leqslant R \leqslant 7.0$ Å. [c] Average with the excitation of the nearly 130 waters that have a 0 kcal/mol interaction with the solution.

S–W energy pair distributions, also their population is consistent with the area below the curve, and the nearly 130 water molecules that do not interact with the solute fall into the spike about 0 kcal/mol. A few waters have positive interaction energies and likely they are located just outside the first shell in the O2 region of space.

In these same regions we have also recorded the cos θ distributions to check whether a fixed positioning was observable outside the first shell or not. In Figures 14 and 15, we report these distribution for both of the solutions. It is evident that no angular position prevails, in contrast to that observed for the first shell.

Figure 15. Cos θ distributions for the waters in the bulk g,t solution just outside the O2 and CH$_3$ regions. θ angle defined as in Figure 5 and 7. Units for the ordinate are mole fraction per 0.04 cos θ.
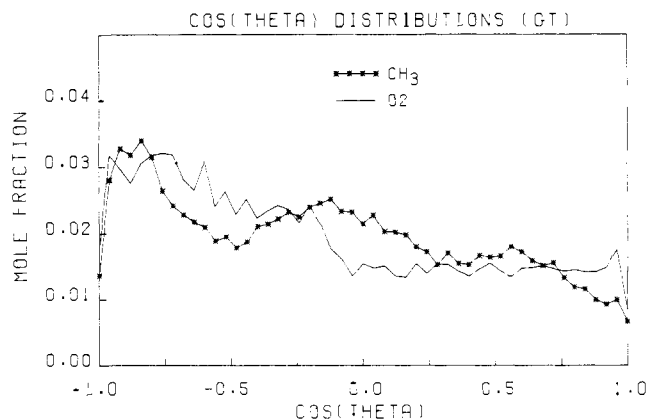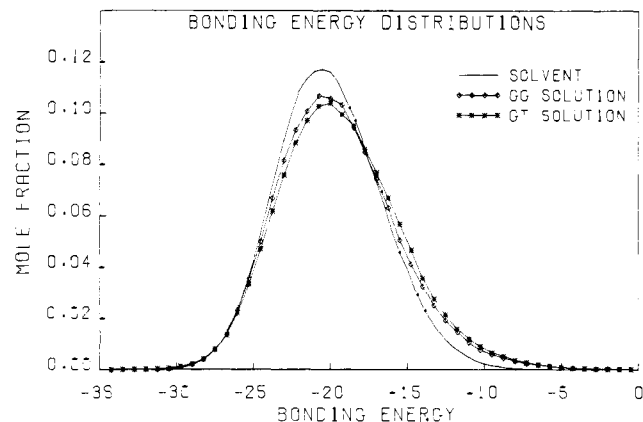

Figure 16. Bonding energy distributions for pure TIP4P water, g,g solution, and g,t solution. Units in ordinate are mole fraction per kilocalories/mole.

We do not show the OS cos θ distribution, because the small occupation number (Table IV) prevents appropriate averaging.

(3) Solvent–Solvent rdfs and Energetics. The solvent–solvent rdfs were also computed in the simulations and compared to the results for pure TIP4P[15] water. To save space we do not report them, because the solvent–solvent structural properties in the solution are very similar to those of pure water, since the number of bulk waters is large with respect to the shell waters: these are just a 13% of the total amount.

The first peak in the g,t solution is slightly affected by the presence of the solute. However, the close similarity between the rdfs with those of the pure solvent further supports our choice of doing the MC calculation with the periodic boundary conditions, instead of with a cluster, that should have contained a larger number of water molecules because of the large size of the solute.

In Figure 16, the bonding energy distributions are reported. Again we can see a slightly different behavior between the g,g and g,t solutions. On the whole they show a higher percentage of less attractive interactions with respect to the pure solvent, ascribable to the first shell waters; we can see that this percentage increases for the g,t solution, where the solute–solvent interactions are stronger.

For the energy pair distribution (Figure 17), the situation is similar to that of rdfs with just negligible differences. This confirms that the interaction between water molecules is mainly of H-bond type. The position of the minimum in the energy pair distribution provides a useful energetic definition, as Jorgensen noted,[9] of an H bond; in the following we will consider H-bonded each pair of monomers bound by at least 2.75 kcal/mol.

In Table V the decomposition of the water–water interaction energy is reported as an energy difference with an equal number of water molecules in pure TIP4P water. The interactions among
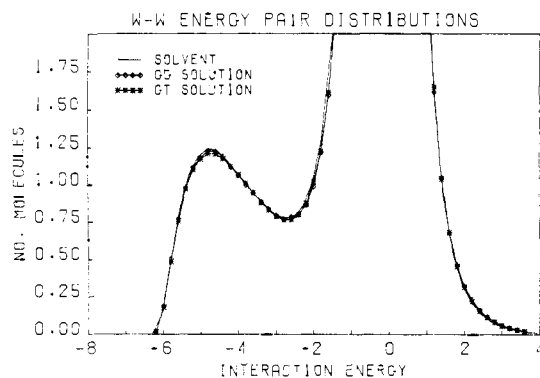
Figure 17. Solvent–solvent energy pair distributions for g,g and g,t solutions. Units in ordinate are number of molecules per kilocalories/mole.

Table V. Decomposition of Water–Water Interaction Energy within the Various Zones, Given as Difference with Respect to an Equal Number of Pure Water Molecules[a]

	g,g	g,t	$(E)_{g,g}$	$(E)_{g,t}$
ΔE_{PP}	4.56	6.65	0.35	0.44
ΔE_{OO}	−0.04	0.02	−0.10	0.05
ΔE_{AA}	−44.04	−37.37	−0.37	−0.35
ΔE_{PO}	1.96	1.15	0.24	0.13
ΔE_{PA}	−11.20	−11.02	−0.13	−0.12
ΔE_{OA}	−11.60	−8.55	−0.50	−0.37
ΔE_{11}	−60.36	−49.12		

[a] Energies are in kcal/mol; for their definitions see the text.

waters belonging to the polar $O^{\delta-}$ shell (ΔE_{PP}) is worse than in pure water because the strong orientation induced by the polar oxygens forces the coordinated waters in less favorable mutual positions. The interaction among waters belonging to the less polar OS region (ΔE_{OO}) is almost the same as in pure water, while the interaction within the apolar region around the methyl groups (ΔE_{AA}) is much better than in pure water, due to the highly ordered (clathrate-like) structure that waters take in this region. For the same reasons, the mixed interactions are better than in pure water when the waters in the apolar region are involved (ΔE_{PA} and, mainly, ΔE_{OA}), while the interaction between waters around the polar $O^{\delta-}$ with the ones around the less polar OS (ΔE_{PO}) is rather worse.

On the whole, the overall interaction among the waters belonging to the first solvation shell (ΔE_{11}) is much better than in pure water, with a large improvement in the g,g solution with respect to the g,t one. It is sufficient, the first shell total contribution $\Delta E'_{solv}$ given by $E'_{SX1} + \Delta E_{11}$, to obtain a more favorably energy of solution for g,g than for g,t (−148.77 vs. −144.92 kcal/mol).
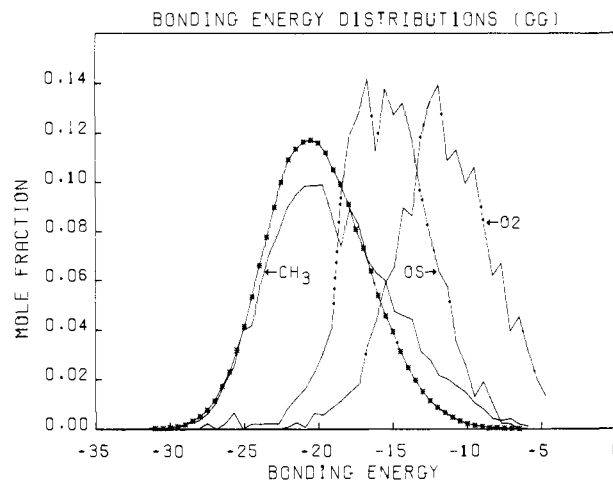

Figure 18. Bonding energy distributions for shell waters in the vicinity of O2, OS, and CH$_3$ for g,g solution. Units in ordinate are mole fraction per kilocalories/mole.

Interestingly, the interactions among waters on a per water basis within the polar shell are worse than in pure water by nearly the same amount as those within the apolar shell are better. As already remarked, they are worse in the g,t solution that has a much better solute–solvent interaction in this region. Interestingly, the overall change in the water–water interaction, ΔE_{SS} (Table II), inside the bulk g,g solution is nearly twice that in the bulk g,t one.

We may have additional information from the bonding energy distributions for waters belonging to the first solvation shell in the vicinity of O2, OS, and CH$_3$, respectively, with all the other water molecules enclosed within a radius of 3.5 Å and from the H-bond analyses in the same regions. In Figure 18 for the g,g solution and in Figure 19 for the g,t one, these distributions are displayed, together with the pure water one, as a comparison. As we have already noted, a stronger solute–solvent interaction goes along with a weaker solvent–solvent interaction, because of the unfavorable positions the waters assumed with respect to each other.

In Table VI the H-bond analysis results are reported for the bulk waters, the shell waters as overall contribution and partitioned according to the region of space surrounding the solute they belong to. As one can see, the waters near the CH$_3$ have H-bond properties quite similar to bulk water, but the ones near O2 and OS, as one might expect, have weaker solvent–solvent interactions and significantly fewer H bonds, due to their stronger interaction with solute. This fact is also evident if we examine the percentage of monomers in N H bonds. About 30% monomers are involved in four H bonds in the bulk solution and around the methyl groups,

Table VI. H-Bond Analysis[a]

	bulk		shell		O2		OS		CH$_3$	
	g,g	g,t	g,g	g,t	g,g	g,t	g,g	g,t	g,g	g,t
av no. of H bonds	3.11	3.12	2.79	2.74	2.59	2.51	2.49	2.52	3.02	2.98
S–S bonding energy	−19.50	−19.32	−16.37	−16.07	−10.69	−11.01	−14.43	−14.38	−18.33	−18.56
$E_{H\ bonds}$[b]	−4.23	−4.22	−4.30	−4.33	−4.44	−4.41	−4.14	−4.36	−4.28	−4.37
$E_{Coulomb}$	−5.63	−5.57	−5.71	−5.76	−5.85	−5.96	−5.56	−5.78	−5.61	−5.85
E_{L-J}	1.40	1.35	1.41	1.43	1.41	1.55	1.42	1.42	1.33	1.48
θ, deg[c]	159.9	159.0	159.9	159.6	160.7	160.4	161.3	160.3	159.1	159.3
Φ	100.3	100.2	100.3	100.6	101.4	99.7	103.6	104.2	98.9	100.5
% Monomers in H Bonds										
$N = 0$	0.2	0.2	0.5	0.5	0.1	0.3	0.5	0.2	0.4	0.1
$N = 1$	3.6	3.6	7.6	8.1	7.0	9.7	8.2	7.3	5.5	5.5
$N = 2$	19.3	19.0	27.5	29.0	30.4	33.1	34.3	33.2	21.1	22.0
$N = 3$	41.6	40.6	42.8	42.6	58.7	52.4	56.2	58.8	40.7	40.6
$N = 4$	32.9	33.8	20.4	18.5	3.8	4.4	0.8	0.5	30.1	28.9
$N = 5$	2.4	2.8	1.2	1.3	0.0	0.1	0.0	0.0	2.3	2.0
$N = 6$	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.1	0.0

[a] Energies are in kcl/mol. [b] Total H-bond energies, which have been decomposed into electrostatic ($E_{Coulomb}$) and Lennard-Jones (E_{L-J}) terms. [c] H-bond angle O–H⋯O. [d] H-bond angle H–O⋯H.

2238 *J. Am. Chem. Soc., Vol. 107, No. 8, 1985*
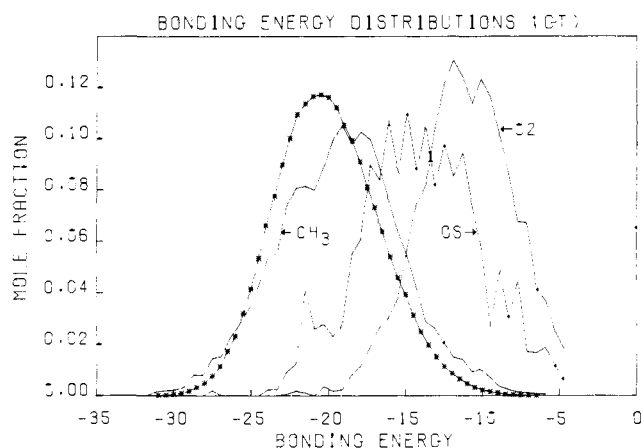
Alagona, Ghio, and Kollman

Figure 19. Bonding energy distributions for shell waters in the vicinity of O2, OS, and CH₃ for g,t solution. Units in ordinate are mole fraction per kilocalories/mole.

Table VII. H-Bond Analysis[a]

	TIP4P	g,g bulk	g,t bulk
av no. of H bonds	3.57	3.52	3.52
S–S bonding energy	−20.03	−19.62	−19.48
$H_{\text{H bonds}}$	−4.17	−4.14	−4.15
E_{Coulomb}	−5.52	−5.45	−5.45
$E_{\text{L–J}}$	1.35	1.32	1.29
θ, deg[c]	158	157.3	156.9
Φ, deg[d]	99	99.1	99.1

[a] Energies are in kcal/mol. [b] Total H-bond energies (see note in Table VI). [c] H-bond angles O–H⋯O. [d] H-bond angle H–O⋯H.

while in the polar regions this percentage falls down to very low values, because of the interaction with the solute that prevents the formation of one H bond with other waters for each monomer.

In Table VII we report the results of the H-bonding analysis for the bulk waters and TIP4P pure water with the same cutoff (−2.25 kcal/mol), to demonstrate that bulk waters behave in a manner quite similar to pure water.

Discussion and Conclusions

We have presented the results and analyses on a MC simulation of the solvation of the dimethyl phosphate anion, the prototype molecule for the linkage between nucleic acid monomer units. To our knowledge, this is the first extensive MC simulation on a small solute with polar, apolar and ionic groups all on the same molecule which has been carried out by using periodic boundary conditions.

It is clear from an examination of the average structural properties and their convergence that we have succeeded in characterizing the structural properties of water solvation of both the g,g and g,t conformations in a reasonably definitive fashion. There is no large conformational dependence of radial distribution functions and angular dependencies of solute group–water geometries but some very interesting differences depending on the nature of the group. The coordination number of each of the anionic oxygens is close to three, consistent with both prior "supermolecule" calculations[18] and with more limited MC simulations.[19] These waters have a strong preference for near tetrahedral coordination around the oxygens, with water O–H bonds pointing directly at the O. There is a negligibly small probability of a water molecule bridging the two anionic oxygens, despite the fact that we have shown[24] that such a geometry is the optimum one for the DMP–single H₂O interaction. An earlier study by Berthod and Pullman[27] had shown that such a bifurcated geometry was optimal for H₂PO₄⁻, and the reference to that work was inadvertently left out of ref 24. Water typically forms slightly more than one (weaker) H bond to the phosphate ester oxygen, despite the fact that this oxygen has a significant negative partial charge and one might expect it to be able to accept closer to two H bonds. The competition of both the anionic oxygen and the

bulk of the methyl groups both apparently contribute to making the OS a weaker H-bond acceptor than one might expect. There are no large differences in average H-bonding properties of the OS whose methyl is trans from the other where it is gauche, but we stress that our use of the same OS partial charge may be part of the cause of this. The waters near the nonpolar methyl group have preferential orientations which point their O–H groups away from the methyl group and one lone pair pointing outward, generally along the CH₃⋯O line. In this way, these waters are able to maintain net H bonding comparable to that in bulk water, despite the excluded volume effect of the methyl group. Rossky and Karplus[10] found in alanyl dipeptide that there was a preference for hydrogens and lone pairs to point away from the methyl groups; their analysis did not distinguish hydrogens and lone pairs. Our analysis (Figure 3) suggests that a lone pair is more likely to point away from the methyl group than a hydrogen, because both hydrogens in this case are strongly oriented toward O2.

One of the goals of our calculation was to assess whether the solvation energy of DMP was more favorable for the gauche,trans than gauche,gauche conformations, as inferred from quantum mechanical "supermolecule" calculations. Our findings do not support the previous studies even though the finding that the solute–solvent energies are more favorable for g,t than g,g by ~5 kcal/mol is consistent with those studies. We find that more favorable solvent–solvent interactions in g,g more than compensate for these solute–solvent energies and, thus, the solvation energy for g,g is calculated to be considerably more favorable than for g,t (the intrinsic gas-phase energy difference favoring g,g is ~1 kcal/mol). Unfortunately, it appears that our solvation energies are not completely converged, and so this prediction of strong intrinsic DMP preference for g,g over g,t in solution remains tentative. It is likely that we have greatly overestimated the solvation energy difference between g,g and g,t in these calculations. However, it appears that at least on order of magnitude longer simulations with a larger box of H₂O molecules will be required to begin to have assurance of energy convergence, simulations beyond the capabilities of our own current computer resources. We are also not sure whether the g,t solvation energy is converged, given the results of Chandrasekhar et al.[28] on F⁻ and Cl⁻ solvation, where they found that the loss of solvent–solvent energy was ~45% of the solute–solvent energy for both anions. On the other hand, the g,t simulation has a loss of H₂O–H₂O energy (Table II) of ~56% of the solute–solvent energy, while the g,g simulation has a loss of 35%. We don't think such a solvent–solvent energy loss will be a universal factor for all solutions, but it may be that the g,t simulation is stuck in a high-energy region that will take many steps to move away from. The energy component analyses also suggest how difficult it will be to ensure convergence, in that the energy component most strongly favoring g,g is the bulk water (Tables V and II), which, if one wishes to use preferentially sampling techniques, is the least sampled in the simulations. On the positive side, we note that our calculated solvation enthalpies for DMP (−65 (g,t) and −93 kcal/mol (g,g)) are of the right magnitude, given that the solvation enthalpy for DMP is likely to be near that of Cl⁻ (−81.3 kcal/mol)[29–31] and significantly less than that of F⁻ (−113 kcal/mol[29]), given the intrinsic ion–single water energies;[15] the water can approach closer to the O2 groups in DMP than they can to the Cl⁻ but cannot solvate the anion as completely because of the "steric" effect of the OS and CH₃. We are not aware of experimental data for the partial molar volume for DMP, but the value of ~63 cm³/mol is qualitatively reasonable, given the size of the molecule.

We did not succeed in a more complete umbrella sampling of the Φ_1 and Φ_2 potential surface of DMP, presumably because of the difficulties of getting out of local minima when one has the

(27) Berthod, H.; Pullman, A. *J. Comput. Chem.* **1981**, *2*, 87.

(28) Chandrasekhar, J.; Spellmeyer, D. C.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1984**, *106*, 903.

(29) Friedman, H. L.; Krishnan, C. V. In "Water: A Comprehensive Treatise"; Franks, F., Ed.; Plenum Press: New York, 1973; Vol. 3.

(30) Singh, U. C.; Kollman, P. *J. Comput. Chem.* **1984**, *5*, 129.

(31) Newton, M. D. *J. Am. Chem. Soc.* **1973**, *95*, 256.

large methyl groups rotating near the highly ionic O2 groups, although our derived analytical potential function (Figure 1) can be used if more clever approaches to get around the sampling problem are found. Such umbrella sampling approaches have worked nicely in studying internal rotation in *n*-butane, and it is likely that the presence of the strong ionic forces causes the difficulties in preventing a similar study of DMP.

Registry No. Dimethyl phosphate anion, 7351-83-9.

Monte Carlo Determination of the Free Energy and Internal Energy of Hydration for the Ala Dipeptide at 25 °C

M. Mezei, P. K. Mehrotra, and D. L. Beveridge*

Contribution from the Department of Chemistry, Hunter College of the City University of New York, New York, New York 10021. Received June 11, 1984

Abstract: The differences in the free energy and internal energy of hydration of the Ala dipeptide in the C_7, α_R, C_5, and P_{II} conformations were computed with the Monte Carlo method in the (T,V,N) ensemble at 25 °C. The free-energy differences were obtained by determining the relative probabilities of the conformations that lie on a line that connects two conformations in the (ψ,ϕ) torsion angle space. The determination of one free-energy difference required three to five separate Monte Carlo runs using non-Boltzman sampling. The results indicate that both the α_R and P_{II} conformations are preferentially stabilized by hydration. The major contributing factor for the stability of internal energy of hydration for these conformations can be traced to the hydration of the carbonyl group.

I. Introduction

The structure of the Ala dipeptide, *N*-acetylalanyl-*N*-methylamide (AcAlaNHMe), in aqueous solution is a matter of considerable current interest in structural biochemistry. The sterically allowed regions of conformation for AcAlaNHMe are prototypical of the polypeptide backbone with all amino acid residues except Gly and Pro, and thus knowledge of the conformational stability of this molecule is highly relevant to fundamental aspects of protein and enzyme structure.[1] The intramolecular energetics of AcAlaNHMe and related molecules have been extensively studied by means of empirical energy functions and molecular quantum mechanics. However, our knowledge of the effect of solvent, especially aqueous hydration, is less complete, and diverse experimental and theoretical results are at variance with one another.

We present herein a new theoretical determination of the thermodynamics of hydration for AcAlaNHMe at 25 °C, based on statistical thermodynamic liquid-state computer simulation. Direct calculations on the internal energy and the free energy of hydration are reported, and new methodological details for free-energy calculations are described. The results obtained are compared with those of previous theoretical calculations and available experimental data.

II. Background.

The conformation of AcAlaNHMe can be specified in terms of the angles of torsion ψ(N–C–C–N) and ϕ(C–N–C–C), as defined in ref 2. Structural studies to date indicate four conformations from the allowed regions of (ψ,ϕ) space for specific consideration:[2] C_7 (90°, –90°), C_5 (150°, –150°), α_R (–50°,

Table I. Calculated Intramolecular Energies (kcal/mol) of the C_5, C_7, α_R, and P_{II} Conformations of AcAlaNHMe Relative to C_7

authors	method	C_7	C_5	α_R	P_{II}
Brant and Flory[7]	EPF	0	0	1	0
Momany et al.[8]	CNDO	0	5	5	5
Hoffman and Imamura[9]	EHT	0	–2	0	0
Pullman et al.[1]	PCILO	0	2	3–4	4
Robson et al.[10,11]	3s2p	0	–3.2	7.8	
	4-31G	0	–2.1	7.5	
	CFF	0	1.6	7.7	
Karplus et al.[12]	EPF	0	6	8	6.5
Scheraga et al.[13]	ECEPP	0	0.38	1.13	1.0

–70°), and P_{II} (150°, –80°). The C_7 and C_5 structures, shown in Figures 1 and 2, respectively, are characterized by seven- and five-atom ring structures completed by an intramolecular hydrogen bond between a CO and NH group. The C_7 form can exist in both axial and equatorial forms; we consider herein the equatorial form. The α_R and P_{II} conformers, shown in Figures 3 and 4, respectively, have ψ and ϕ angles similar to those found in the right-handed polypeptide α-helix and the poly(L-proline)-II helix. However, the α_R and P_{II} forms of AcAlaNHMe are not fully representative of helical polypeptides, which are further stabilized by intramolecular hydrogen bonds between subunits.

The structure of AcAlaNHMe in CCl₄ was established by Avignon and Lascombe by infrared spectroscopy[3] to be predominantly C_7. The crystal structure, reported by Harada and Iitaka,[4] involved two conformers: P_{II} and the structurally similar poly-(L-proline)-I form. Avignon et al. subsequently extended their study to AcAlaNHMe in water[5] and proposed on the basis of an

(1) B. Pullman and B. Maigret in "Conformation of Biological Molecules and Polymers", E. Bergman and B. Pulman, Eds., Academic Press, New York, 1973.
(2) C. R. Cantor and P. R. Schimmel, "Biophysical Chemistry", Vol. 1, W. H. Freeman, San Francisco, 1980.

(3) M. Avignon and J. Lascombe in "Conformation of Biological Molecules and Polymers", E. Bergman and B. Pulman, Eds., Academic Press, New York, 1973.
(4) Y. Harada and Y. Iitaka, *Acta Crystallogr., Sect. B*, **30**, 1452 (1974).
(5) M. Avignon, C. Garrigou-Lagrange, and P. Bothorel, *Biopolymers*, **12**, 1651 (1973).