Comparative study of three systems of potential functions for simulation of nucleic acid hydration

V. I. Poltev,[a]* E. J. Gonzalez,[ab] A. V. Teplukhin,[c] G. G. Malenkov,[d]* and J. Miller[e]

[a]Institute of Theoretical and Experimental Biophysics, Russian Academy of Sciences,
142292 Pushchino, Moscow Region, Russian Federation
[b]Autonomous University of Puebla, Puebla, Mexico
[c]Institute of Mathematical Problems of Biology, Russian Academy of Sciences
142292 Pushchino, Moscow Region, Russian Federation
[d]Institute of Physical Chemistry, Russian Academy of Sciences,
31 Leninsky prosp., 117915 Moscow, Russian Federation
[e]Pacific Northwest Laboratories, Richland, WA, USA

Three systems of potential functions (PF) which are used in simulations of hydration of bioorganic molecules have been compared: Poltev and Malenkov (PM), Weiner and Kollman (WK) (used in the widespread AMBER program), and OPLS (optimized potentials for liquid simulations) of Jorgensen (J). The values of interaction energies of individual water molecules with single molecules of nucleic bases calculated *via* PM potentials are in somewhat better accord with mass spectrometric data than those calculated *via* WK PF. OPLS give much smaller energy values for all compounds considered; therefore they were not used in further computations. Monte Carlo simulation of hydration of 9-methyladenine, 1-methyluracil, and 1-methylthymine in systems with 300 water molecules and periodic boundary conditions have been performed. Simulations with PM potentials give better agreement with the experimental data on hydration energies than those with WK PF that allows to prefer PM PF for simulation of hydration of nucleic acids.

Key words: hydration; computer simulation; Monte Carlo; potential functions; intermolecular interactions; nucleic acids.

Computer simulation of hydration of organic molecules is associated with some methodological problems, incorrect solutions of which may give even qualitatively inadequate results. The search of potential functions and boundary conditions for adequate description of intermolecular interactions in a liquid and generation of Markov chains long enough to achieve the required accuracy of calculations are among these problems.

Nevertheless, computer simulation makes it possible to obtain data frequently unavailable by experimental methods. Thus, simulation of hydration of DNA and its fragments is necessary to elucidate the structural aspects of the functions of this important biopolymer. In this connection, it is necessary to mention the well-known dependence of the conformation of the double helix on the water content in a sample[1] and the X-ray study of crystals of oligonucleotide duplexes[2] indicating a specific hydration of the minor groove of AT-sites of DNA (the stability of the B'-conformation of poly-dA : poly-dT is explained sometimes by the "water spine" formation[3]). In conjunction with that, the hydration of fragments and separate components of DNA is also interesting from the methodological viewpoint. There is a set of experimental results that may be used for quantitative testing a calculation procedure. The energies of interactions of the individual water molecules with nucleic bases and their derivatives obtained in mass spectrometric studies[4] and the energies of hydration of bases[4,5] (calculated as differences in energies of dissolution and sublimation of the corresponding compound) are the required data.

In 1984, we have refined the system of atom-atom potential functions (PF) for the calculation of interactions between water molecules in "liquid water" and water molecules with the components of nucleic acids.[6] These PF (designated as Poltev—Malenkov potentials, PM) were used in several works on simulation of hydration of DNA fragments (*e.g.*, see Ref. 7 and references therein), thus makes possible to reveal and describe the peculiarities of structure of the hydration shells of separate bases and double-helix fragments of nucleic acids.

Another system of PF frequently used in simulation of the structure of biopolymers was proposed by Weiner and Kollman (PF WK).[8] Together with the Jorgensen's TIP3P water model,[9] this system is used in the widespread computer program AMBER for simulation of biopolymers.[10]

This work deals with comparison of two systems of potential functions, *viz.*, PM and WK, and calculations with OPLS potentials (Optimized Potentials for Liquid

---

Translated from *Izvestiya Akademii Nauk. Seriya Khimicheskaya,* No. 7, pp. 1381—1386, July, 1995.

1066-5285/95/4407-1333$12.50

Simulations) designed by Jorgensen[11,12] (PF J), in which his TIP4P model of water was used.[9]

Procedure of Calculations

The calculations of energies of intermolecular interactions are performed within the framework of the method of the atom-atom potential functions using three systems of PF: PM, WK, and J. The averaged crystallographic data on several related compounds were taken for obtaining geometric parameters of nucleic bases. The positions of force field centers of water molecules for all of the system of potentials were taken according to the model accepted for this system.

The simulation of hydration by the Monte-Carlo method (Metropolis sampling) was carried out for 300 K using procedures previously described.[6,7] The periodic boundary conditions were imposed only to the water molecules. Water—base and base—base interactions between the neighboring unit cells were neglected. The parameters of the unit cell in the system of 300 water molecules (25.0 Å×25.0 Å×14.4 Å) are in accord with the density of 0.98 g/cm^3. As to water—base systems, one of the dimensions of the cell (height) was changed to take into account the intrinsic volume of a molecule of a base. The following molecular volumes at 300 K were used: water, 30 Å^3; 9-methyladenine, 168 Å^3; 1-methylthymine, 169 Å^3; 1-methyluracil, 137 Å^3. The molecular volumes of bases were calculated using X-ray data on crystals of 9-methyladenine,[13] 1-methylthymine,[14] and 1-methyluracil.[15] These estimations of the molecular volumes of the bases practically coincide with the data on the various, including mixed, crystals.

Results and Discussion

Interactions of the individual water molecules with nucleic bases. The results of a search for the deepest minima of energies of interactions between a single water molecule and considered purines and pyrimidines are presented in Table 1, and the relative positions of molecules of bases and water in these minima are presented in Fig. 1. All of these minima are correspond to the planar position of the nucleic bases and the oxygen atom of water. As can be seen from Fig. 1 and the nearest interatomic distances presented in Table 1, the positions of water molecules are almost the same in every minimum for all types of PF. In addition, the global minima for all of the bases (except guanine) calculated using these potentials, refer practically to the same position of water molecules near the base. The ranging of energy minima are also practically the same. In calculations with PM potentials for 9-methylguanine, the global minimum corresponds to the formation of H-bonds between the water molecule and the N(7) and O(6) atoms, and in calculations with WK and J potentials, the H-bonds involve the O(6) and H(1) atoms.

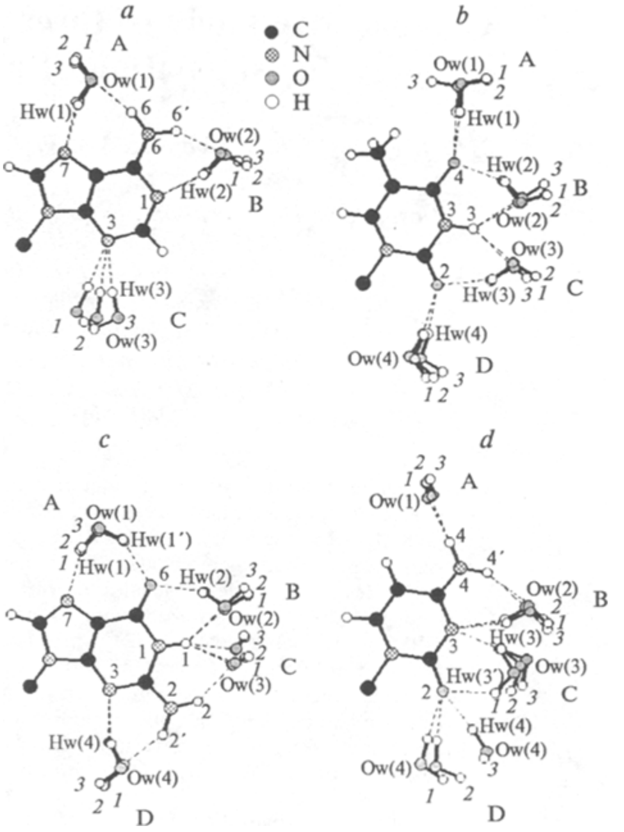

Fig. 1. Positions of water molecules in local interaction energy minima of with 9-methyladenine (*a*), 1-methylthymine (*b*), 9-methylguanine (*c*), and 1-methylcytosine (*d*). Designations of PFs are as follows: PM *1*; WK *2*; J *3*. The minima, whose energies are presented in Table 1, are designated by letters.

These calculations show that the energy values obtained using PM and WK potentials are rather close to each other, but J potentials always give the underestimated values. Since PM potentials are chosen with regatd to (*inter alia*) the energies of interactions of individual water molecules and bases, WK potentials are in good agreement with these experimental data as well.

The detailed comparison of PF using the data on the energies of interactions of a single water molecule with methylated bases is presented in Table 2. The calculations show that the replacement of the H atom by the methyl group in all of the positions, besides those involved in formation of H-bonds, changes the value of energies of water—base interactions in local minima only by 0.1—0.2 kcal mol^{-1}. Hence, one can compare our calculations for 9-methylpurines and 1-methylpyrimidines with the data on polymethylated bases with a satisfactory accuracy. Using the PM potentials, the calculations for other derivatives of bases, for which the experimental data are available, are also possible, but we did not carry out such calculations due

Table 1. Energies of intermolecular interactions and the nearest interatomic distances in local minima of the system involving the base and one water molecule

Local minimum[a]	Parameter[b]	PM		WK		J	
			9-Methyladenine				
A	E	−10.81		−10.80		−7.76	
	$R_{H_w(1)N(7)}$, $R_{O_w(1)N(7)}$	1.92	2.82	1.91	2.86	1.95	2.89
	$R_{O_w(1)H(6)}$, $R_{O_w(1)N(6)}$	1.89	2.77	1.85	2.80	2.02	2.90
B	E	−9.08		−9.78		−7.58	
	$R_{H_w(2)N(1)}$, $R_{O_w(2)N(1)}$	1.89	2.82	1.86	2.82	1.90	2.84
	$R_{O_w(2)H(6')}$, $R_{O_w(2)N(6)}$	1.80	2.78	1.80	2.79	1.90	2.88
C	E	−6.99		−6.01		−5.16	
	$R_{N(3)H_w(3)}$, $R_{O_w(3)N(3)}$	1.89	2.87	1.90	2.86	1.90	2.85
			1-Methylthymine				
A	E	−6.34		−5.46		−6.40	
	$R_{H_w(1)O(4)}$, $R_{O_w(1)O(4)}$	1.85	2.83	1.87	2.84	1.83	2.77
B	E	−7.95		−8.11		−7.48	
	$R_{H_w(2)O(4)}$, $R_{O_w(2)O(4)}$	1.90	2.76	1.94	2.82	1.87	2.78
	$R_{O_w(2)H(3)}$, $R_{O_w(2)N(3)}$	1.92	2.79	1.94	2.82	2.10	2.95
C	E	−6.98		−8.01		−6.46	
	$R_{H_w(3)O(2)}$, $R_{O_w(3)O(2)}$	1.91	2.76	1.94	2.82	1.93	2.82
	$R_{O_w(3)H(3)}$, $R_{O_w(3)N(3)}$	1.93	2.79	1.95	2.83	2.08	2.94
D	E	−5.93		−5.58		−6.28	
	$R_{H_w(4)O(2)}$, $R_{O_w(4)O(2)}$	1.86	2.82	1.86	2.83	1.81	2.77
			9-Methylguanine				
A	E	−12.53		−9.98		−9.37	
	$R_{H_w(1)N(7)}$, $R_{O_w(1)N(7)}$	1.90	2.81	1.92	2.84	2.05	2.93
	$R_{O(6)H_w(1')}$, $R_{O(6)O_w(1)}$	1.87	2.74	1.97	2.75	1.93	2.79
B	E	−10.94		−11.04		−11.11	
	$R_{O_w(2)H(1)}$, $R_{O_w(2)N(1)}$	1.87	2.74	1.91	2.79	1.99	2.85
	$R_{H_w(2)O(6)}$, $R_{O_w(2)O(6)}$	1.90	2.76	1.94	2.84	1.88	2.79
C	E	−10.52		−10.16		−9.62	
	$R_{O_w(3)H(1)}$, $R_{O_w(3)N(1)}$	1.88	2.77	1.96	2.87	1.83	2.81
	$R_{O_w(3)H(2)}$, $R_{O_w(3)N(2)}$	1.84	2.74	1.98	2.88	2.39	3.21
D	E	−10.40		−9.77		−7.81	
	$R_{O_w(4)H(2')}$, $R_{O_w(4)N(2)}$	1.86	2.73	1.89	2.78	1.97	2.86
	$R_{H_w(4)N(3)}$, $R_{O_w(4)N(3)}$	1.92	2.81	1.95	2.87	2.00	2.92
			1-Methylcytosine				
A	E	−6.70		−7.30		−6.24	
	$R_{O_w(1)H(4)}$, $R_{O_w(1)N(4)}$	1.77	2.77	1.80	2.80	1.88	2.88
B	E	−11.81		−11.58		−9.92	
	$R_{O_w(2)H(4')}$, $R_{O_w(2)N(4)}$	1.92	2.79	1.93	2.82	2.04	2.92
	$R_{H_w(2)N(3)}$, $R_{O_w(2)N(3)}$	1.90	2.79	1.89	2.84	1.94	2.86
C	E	−11.80		−9.87		−8.75	
	$R_{H_w(3)N(3)}$, $R_{O_w(3)N(3)}$	1.93	2.77	1.89	2.84	1.89	2.85
	$R_{H_w(3')O(2)}$, $R_{O_w(3)O(2)}$	1.96	2.68	2.47	2.96	2.85	3.19
D	E	−7.57		−6.92		−7.33	
	$R_{H_w(4)O(2)}$, $R_{O_w(4)O(2)}$	1.85	2.83	1.84	2.82	1.78	2.73

[a] Designations of local minima (Fig. 1).

[b] Values of energies are given in kcal/mol, and the distances are given in Å.

to the lack of literature data on atomic charges in these compounds (obtained within the framework of PF WK and J).

The data in Table 2 demonstrate a good agreement of the results obtained using PM potentials with all of the experimental values, except for the value for

Table 2. Experimental enthalpies of formation of methylated base monohydrates and the corresponding calculated energies

Base	$\Delta H_{exp}{}^{4}$	PM	WK	J
m^{1}Thy	−10.4±0.9	−7.9 (B)	−8.1 (B)	−7.5 (B)
m^{1}Cyt	−11.4±0.8	−11.8 (B)	−11.6 (B)	−9.9 (B)
$m^{1}{}_{3}{}^{,4,4}$Cyt	−11.8±0.9	−11.8 (C)	−9.9 (C)	−8.8 (B)
m^{9}Ade	−10.6±1.0	−10.8 (A)	−10.8 (A)	−7.8 (B)
$m^{6}{}_{3}{}^{,6,9}$Ade	−8.3±0.8	−7.0 (C)	−6.0 (C)	−5.1 (C)
$m^{2}{}_{3}{}^{,2,9}$Gua	−14.1±1.0	−12.5 (A)	−11.0 (B)	−11.1 (B)

Note. The designations of minima are given in parentheses after the corresponding calculated values.

1-methylthymine (WK potentials and J also give a disagreement in this case). Slightly lowered absolute values of energy minima of interactions between water and 9-methyladenine and 9-methylguanine *vs.* the experimental data on 6,6,9-trimethyladenine and 2,2,9-trimethylguanine are related with the absence of two additional methyl groups. WK PF give considerably worse, but satisfactory, agreement of calculations with the experimental heats of formation of monohydrates of 1,4-dimethylcytosine, 1,4,4-trimethylcytosine, 6,6,9-trimethyladenine, and 2,2,9-trimethylguanine.

The energies obtained using the J potentials were too underestimated for all of the compounds, in spite of some parameters of these potentials were chosen using the data obtained by *ab initio* calculations for the water—base interactions.[12] Therefore, we did not carry out the further calculations with J PF.

Simulation of hydration of nucleic bases by the Monte Carlo method. In numerous works on simulation of hydration of molecules by the Monte Carlo method the cluster approximation is used, *i.e.*, either the considered system has boundaries, or the location of boundaries makes the density substantially lower than that of liquid water. We began a comparison of the two systems of PF with the calculations for clusters containing a various numbers of water molecules (from 25 to 400) and for the same clusters containing one of the nucleic bases. It appeared that such calculations cannot be used for quantitative comparison with experimental data. Therefore, we report only general, methodologically important conclusions from these calculations.

An increase in a number of water molecules in clusters containing the DNA bases to 100 affords a gradual increase (in absolute value) not only in the total energy and the energies of water—water interactions per one water molecule (this is easy to understand as a result of increase in a portion of non-surface water molecules), but in the water—base interaction energy and in a slight increase in number of H-bonds between water and bases as well. In this procedure, one can generate Markov chains (as a rule, 100—200 thousand configurations per water molecule are enough), giving the reproducible average energies and numbers of H-bonds remaining constant during the further increase in a number of configurations.

Thus, when the number of water molecules in a cluster is relatively small, one can achieve adequate average energies (the water molecules are kept together by the field of a base). However, in this case, the part of surface water molecules is high and we fail to obtain both reliable values of total energy and the energy of the water—base interaction. When the number of water molecules in clusters is large, a convergence of the average energies becomes inadequate, because the water molecules "evaporate" from the surface of a cluster. As is shown in the further calculations with periodic boundary conditions (see below), the average energy of water—base interactions is obtained with a good accuracy. However, this energy is only a part of the energy of hydration of the base. Another part involving a change in the energies of water—water interactions is unavailable in a cluster approximation. Thus, attempts to obtain the energy of hydration as the difference of energies of a cluster of pure water and the cluster of the same number of water molecules with a base gave contradictory results. It is appeared that this difference depends strongly and nonmonotonically on the number of water molecules in the cluster and also on the shapes of boundaries.

More correct calculations of hydration energies were performed using periodic boundary conditions and a larger number of water molecules in the unit cell. The results reported here are related to 300 water molecules (the similar results were also obtained for 400 water molecules in the unit cell). Using the same concept, the volume of the studied system can be taken to be in agreement with the experimental data. It should be noted that the adequate convergence of the average energies requires at least 200000 trials per water molecule: in this case, the average values of total energy of the system and its components are practically constant at the further trials. In Fig. 2, the dependencies of average energy, its components, and a number of H-bonds between water and the base for a system of 1-methylthymine and 300 water molecules obtained by calculations using PM potentials are plotted.

The results of calculations are summarized in Table 3. It appeared that the values of total energy for pure water (E_{pw}) are rather close for calculations both with PM potentials and WK potentials. The calculations indicate that PM potentials afford the adequate hydration energies for all three nucleic bases, for which the experimental data are available. A slightly increased value for 9-methyladenine can be only noted. The WK potentials afford significantly underestimated absolute values of energies for 1-methylthymine and 1-methyluracil. Simultaneously, the hydration energy of uracil calculated using the WK potentials have the higher absolute value than that of thymine, whereas the experimental data point to the opposite situation; this is adequately reproduced in calculations using the PM potentials.

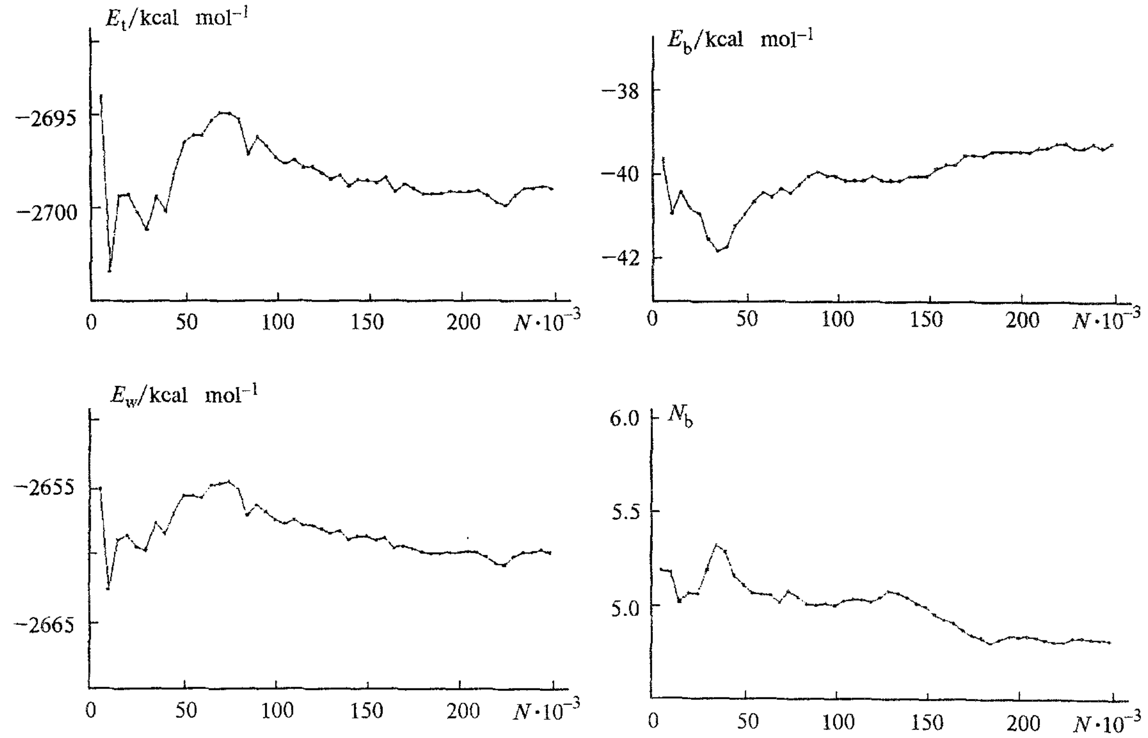

Fig. 2. Dependence of average energy, its components, and the number of H-bonds between water and base for the system containing 1-methylthymine and 300 water molecules on the length of the Markov chain. N is the number of configurations per water molecule.

Table 3. Average energies of the system containing a base and 300 water molecules and its components (at 300 K) in comparison with the experimental enthalpies of hydration

Parameter	9-Methyladenine		1-Methylthymine		1-Methyluracil	
	PM	WK	PM	WK	PM	WK
E_{pw}	−2677.8	−2699.1	−2677.8	−2699.1	−2677.8	−2699.1
E_t	−2702.8	−2721.4	−2702.1	−2713.1	−2698.2	−2713.8
E_b	−50.1	−42.1	−39.3	−30.7	−38.1	−30.8
E_w	−2652.7	−2679.3	−2660.0	−2682.4	−2660.1	−2683.0
N_b	6.78	5.9	4.8	3.8	5.1	3.8
ΔE	−25.0	−22.3	−22.5	−14.0	−20.4	−14.7
ΔH_{exp}	−23.0[5]		−23.4[4]		−21.3[4]	

Note. E_{pw} is the energy of the system consisting of 300 water molecules (without a base molecule); E_t is a total energy of the system containing a base; E_b is the energy of the water—base interactions; E_w is the energy of water—water interactions. N_b is the number of H-bonds between water and base; ΔE is the calculated energy of hydration (E_t-E_{pw}); ΔH_{exp} is the experimental enthalpy of hydration (at 298 K).

It is also to be noted that the PM potentials give the energies of the water—base interactions by *ca.* 10 kcal/mol higher (absolute value) than those of the WK potential. This result also correlates with larger average numbers of the water—base H-bonds obtained using the PM potentials.

Conclusions. Three systems of potential functions used in simulation of hydration of bioorganic molecules are considered; the PM potentials have the best agreement both with the mass spectrometric data on the interaction energies of purines and pyrimidines with individual molecules of water and with the experimental hydration energies (obtained as differences of heats of sublimation and dissolution). The quantitatively reliable hydration energies can be calculated using the Monte Carlo method for systems with periodic boundary conditions. The calculations by cluster approximation afford only rough estimations of hydration energies of organic molecules.

This work was supported by the Russian Foundation for Basic Research (project Nos. 94-03-08996a and 94-04-11438).

References

1. R. Franklin and R. G. Gosling, *Acta Crystallogr.*, 1953, **6**, 673.
2. R. E. Dickerson, M. L. Kopka, and H. R. Drew, in *Structure and Dynamics: Nucleic Acids and Proteins*, Eds. E. Clementi and H. R. Sarma, Academic Press, N.Y., 1983, 149.
3. A. A. Lipanov and V. P. Chuprina, *Nucleic Acids Res.*, 1987, **15**, 5833.
4. B. I. Verkin, I. K. Yanson, L. F. Sukhodub, and A. B. Teplitsky, *Vzaimodei'stviya biomolekul: novye eksperimental'nye podkhody i metody* [*Interactions of Biomolecules: New Experimental Approaches and Methods*], Naukova Dumka, Kiev, 1985, 164 pp. (in Russian).
5. A. Zielenkiewcz, W. Zielenkiewicz, L. F. Sukhodub, O. T. Glukhova, A. B. Teplitsky, and K. L. Wierzchowski, *J. Sol. Chem.*, 1984, **13**, 757.
6. V. I. Poltev, T. I. Grokhlina, and G. G. Malenkov, *J. Biomol. Struct. Dyn.*, 1984, **2**, 413.
7. V. I. Poltev, A. V. Teplukhin, and G. G. Malenkov, *Int. J. Quant. Chem.*, 1992, **42**, 1499.
8. S. J. Weiner, P. A. Kollman, D. T. Nguyen, and D. A. Case, *J. Comp. Chem.*, 1986, **7**, 230.
9. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926.
10. D. A. Perlman, D. A. Case, J. C. Caldwell, G. L. Seibel, U. C. Singh, V. Weiner, and P. A. Kollman, *AMBER 4.0 (UCSF)*, Department of Pharmaceutical Chemistry, University of California, San Francisco, CA, 1991.
11. W. L. Jorgensen and J. Tirado-Rives, *J. Am. Chem. Soc.*, 1988, **110**, 1657.
12. J. Pranata, S. G. Wierschke, and W. L. Jorgensen, *J. Am. Chem. Soc.*, 1991, **113**, 2810.
13. R. K. McMullan, P. Benci, and B. M. Craven, *Acta Cryst.*, 1980, **B36**, 1424.
14. A. Kvick, T. F. Koetzle, and R. Thomas, *J. Chem. Phys.*, 1974, **61**, 2711.
15. D. W. Green, F. S. Mathews, and A. Rich, *J. Biol. Chem.*, 1962, 237, **11**, 2373.

Received December 26, 1994