Monte Carlo Simulation of the Influence of Solvent on Nucleic Acid Base Associations

A. Pohorille,*[1a] S. K. Burt,[1b] and R. D. MacElroy[1c]

Contribution from the Department of Chemistry, University of California, Berkeley, California 94720, and Extraterrestrial Research Division, Ames Research Center, Moffet Field, California 94035. *Received February 28, 1983*

Abstract: The results of Monte Carlo calculations of the associations between nucleic acid bases in a nonpolar solvent (CCl_4) are described. The influence of the solvent on planar and stacked associations of bases was examined by analyzing the total energy of the system, including solute–solute, solute–solvent, and solvent–solvent contributions. Good quantitative agreement with the available experimental data was obtained. Solute–solvent interactions are primarily determined by dispersion forces; consequently, solute–solvent interactions vertical to the solute plane that maximize dispersion interactions are most favored, and a rough proportionality between solute–solvent energy and the surface of the solute was observed. Analysis of solvent–solvent energy shows that the presence of the solute induces significant changes in the structure of the solvent. As a result, solvent–solvent energy is not necessarily reduced when surface area decreases, contrary to the simple cavity concept. "Single molecule probe" calculations were performed to explain the differences in base associations in H_2O and CCl_4. In CCl_4 dispersion forces dominate and planar complexes are stabilized by maximum exposure of molecular planes to the solvent. In H_2O electrostatic forces dominate so that the most stable structures are stacked associations that allow the maximum number of hydrophilic centers to be exposed to the solvent.

Early experimental studies on nucleic acids indicated that the melting temperature of deoxyribonucleic acid (DNA) from various sources was an increasing linear function of the guanine (G) and cytosine (C) content.[2] Since G and C can form three hydrogen bonds between them, and adenine (A) and thymine (T) pairs can form only two hydrogen bonds, it was assumed that hydrogen bonds were the primary stabilizing factor in the stability of the DNA double helix.[3]

Other experimental evidence, however, cast doubt on this assumption and instead suggested that the solvent plays as additional and very important role in the stability of nucleic acids. In particular, isolated nucleic acid bases in nonpolar solvents, such as carbon tetrachloride (CCl_4)[4] and chloroform ($CHCl_3$),[5-9] and in vacuo[10] associate by hydrogen bonding. In aqueous solution, however, bases form stacked complexes.[11-18] Such model systems have the advantage that factors such as the presence of the backbone and geometrical constraints of double-helical DNA are eliminated and the influence of solvent on molecular associations can be determined directly.

The influence of solvent on base–base associations was first investigated theoretically by Sinanoglu and Abdulnur.[19,20] They concluded that hydrophobic forces play a dominant role in promoting the stacking of nucleic acid bases and thereby contribute to the overall stability in aqueous solutions. Further theoretical studies on the role of solvent in base-pair associations were carried out by Kudritskaya and Danilov,[21] Pullman et al.,[22] Egan et al.,[23] and Cieplak et al.[24] In these studies, either continuum models[21,23] or small cluster methods[22,24] were used to describe solvent effects. Unfortunately, in all of these cases, the simplified models used and the qualitative characters of the calculations could give only limited insight into the influence of the solvent on nucleic acid base associations. In particular, continuum models are not able to deal with specific solute–solvent interactions, and small cluster methods are not able to provide accurate information about changes in solvent structure.

Our goal is to quantitatively understand the basis of the differences in interactions among nucleic acid bases in polar and nonpolar solvents in terms of solute–solvent and solvent–solvent interactions. Such quantitative understanding will only be realizable if a method is employed that is able to provide direct information about ensemble averages of mechanical properties and solvent microstructure. For this reason the Monte Carlo method of statistical mechanics was used.

In this paper we present quantitative results of Monte Carlo calculations on the association between nucleic acid bases in CCl_4 and qualitative results of interactions of nucleic acid bases with CCl_4 and with water. The conclusions drawn from these calculations are then contrasted. Nonpolar solvents are of interest not only as model systems to be compared with water but also because a large number of physicochemical measurements (e.g., IR spectroscopy) are performed in such solvents as CCl_4 or $CHCl_3$. Experimental data from nonpolar solvents are usually interpreted with the implicit assumption that solute–solvent interactions, and changes in solvent structure upon solvation, are negligible. Consequently, such data are often used to verify quantum-mechanical calculations performed for isolated solute molecules. We

(1) (a) Department of Chemistry, University of California at Berkely. (b) National Research Council Associate; Extraterrestrial Research Division, Ames Research Center. (c) Estraterestrial Research Division, Ames Research Center.

(2) Marmur, J.; Doty, P. *J. Mol. Biol.* **1962**, *5*, 109.
(3) Watson, J. D.; Crick, F. H. C. *Nature (London)* **1953**, *171*, 737.
(4) Kuechler, E.; Derkosch, J. *Z. Naturforsch. B: Anorg. Chem., Org. Chem.* **1966**, *21b*, 209.
(5) Kyogoku, Y.; Lord, R. C.; Rich, A. *J. Am. Chem. Soc.* **1967**, *89*, 496.
(6) Pitha, J.; Jones, R. N.; Pithova, P. *Can. J. Chem.* **1966**, *44*, 1045.
(7) Katz, L.; Penman, S. *J. Mol. Biol.* **1966**, *15*, 220.
(8) Miller, J. H.; Sobell, H. M. *J. Mol. Biol.* **1967**, *24*, 345.
(9) Binford, J. S., Jr.; Holloway, D. M. *J. Mol. Biol.* **1968**, *31*, 91.
(10) Yanson, I. K.; Teplitsky, A. B.; Sukhodub, L. F. *Biopolymers* **1979**, 18, 1149.
(11) Ts'o, P. O. P.; Melvin, I. S.; Olson, A. C. *J. Am. Chem. Soc.* **1963**, *85*, 1289.
(12) Schweizer, M. P.; Broom, A. D.; Ts'o, P. O. P.; Hollis, D. P. *J. Am. Chem. Soc.* **1968**, *90*, 1042.
(13) Solie, T. N.; Schellman, J. A. *J. Mol. Biol.* **1968**, *33*, 61.
(14) Van Holde, K. E.; Rosetti, G. P. *Biochemistry* **1967**, *6*, 2189.
(15) Farquhar, E. L.; Downing, M.; Gill, S. J. *Biochemistry* **1968**, *7*, 1224.
(16) Nakano, N. I.; Igarashi, S. J. *Biochemistry* **1970**, *9*, 577.
(17) Imoto, T. *Biochim. Biophys. Acta* **1977**, *475*, 409.
(18) Mayevsky, A. A.; Sukhorukov, B. I. *Nucleic Acids Res.* **1980**, *8*, 3029.
(19) Sinanoglu, O.; Abdulnur, S. *Photochem. Photobiol.* **1964**, *3*, 333.
(20) Sinanoglu, O.; Abdulnur, S. *Fed. Proc., Fed. Am. Soc. Exp. Biol.* **1965**, *24* (2), S-12.
(21) Kudritskaya, Z. G.; Danilov, V. I. *J. Theor. Biol.* **1976**, *59*, 303.
(22) Pullman, B.; Miertus, S.; Perahia, D. *Theor. Chim. Acta* **1979**, *50*, 317.
(23) Egan, T. J.; Nir, S.; Rein, R.; MacElroy, R. *Int. J. Quantum Chem., Quantum Biol. Symp.* **1978**, *5*, 433.
(24) Cieplak, P.; Geller, M.; Les, A. In "Biomolecular Structure, Conformation, Function and Evolution"; Srinivasan, R., Ed.; Pergamon Press: Elmsford, NY, 1981; p 321.

0002-7863/84/1506-0402$01.50/0

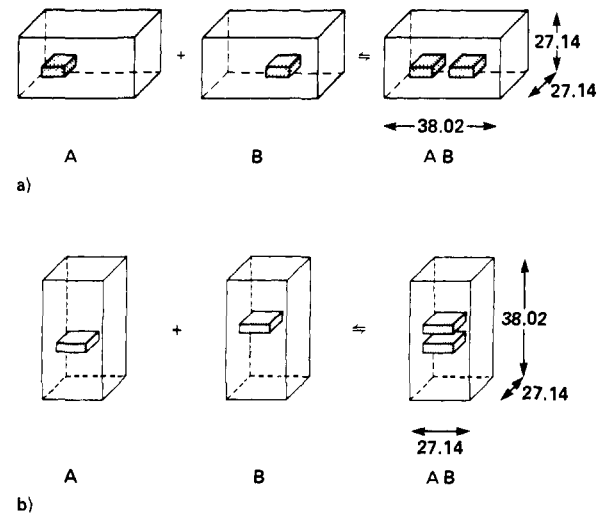

Figure 1. Schematic drawing showing the placement of the nucleic acid bases in the rectangular boxes for the planar (a) and stacked (b) complexes. Note that each individual base is positioned in its box in the orientation that it has in the complex.

will show that nonpolar solvents *do* interact with the solute and that both solute–solvent and solvent–solvent interactions play a significant role in the association of nucleic acid bases.

In this work Monte Carlo calculations were performed for dilute solutions of adenine, uracil (U), guanine, and cytosine in CCl_4. Calculations were also performed on solutions of hydrogen-bonded A-U and G-C pairs in Watson–Crick configuration and on stacked autoassociations, and the relative stabilities of these complexes were compared. Uracil was used instead of thymine to facilitate direct comparison with available experimental data. Analogous calculations for water solutions will be presented in forthcoming papers.[25] It should be noted that our discussion of stability will be limited to internal energy only, since neither free energy nor entropy can be obtained, in a straightforward manner, from Monte Carlo simulation.

It should be recognized that the model that we are investigating is relatively simple from a biological viewpoint. However, we believe that the model not only will provide information on the effect of solvent on the geometry of nucleic-acid base associations in particular, but will also yield useful information about the influence of the solvent on biomolecular associations in general.

Methods

Monte Carlo Procedure. Monte Carlo statistical mechanical simulations were carried out in the standard manner using the Metropolis sampling technique[26] in the canonical (T, V, N) ensemble. Periodic boundary conditions in the first image approximation were employed. The temperature in the system was set at 300 K. All calculations were performed in a rectangular box at experimental density, 1.594 g/cm^3. The edges of the box were 38.02 × 27.14 × 27.14 Å, which corresponds to 175 CCl_4 molecules of pure solvent. As is shown in Figure 1, the long edge was directed along the x axis for hydrogen-bonded pairs and along the z axis for stacked complexes. The rectangular box is preferable to a cube because it corresponds better to the symmetry of the problem in planar cases and can contain a larger number of important vertical solute–solvent interactions in the case of stacked solute associations. The initial configuration was obtained by inserting a solute in a cavity located in the final configuration from the computation for the pure solvent. This cavity was created by removing solvent molecules in such a way that the shape and volume of the cavity corresponded approximately to the shape and volume of the solute. The position of the solute in the box remained fixed throughout the computations. In principle, it would be preferable to allow the solute to move since it reduces anisotropy of molecular correlations induced by periodic boundary conditions. In practice, however, the rate of convergence for solutes as large as those used here would be very slow, and calculations would require an inordinate amount of time. We attempted to reduce the influence of boundaries on energy differences by positioning isolated bases in exactly the same positions that they occupied in the complexes. This procedure, shown in Figure 1, is analogous to the "counterpoise" method[27] used in calculations of intermolecular interactions in quantum chemistry.

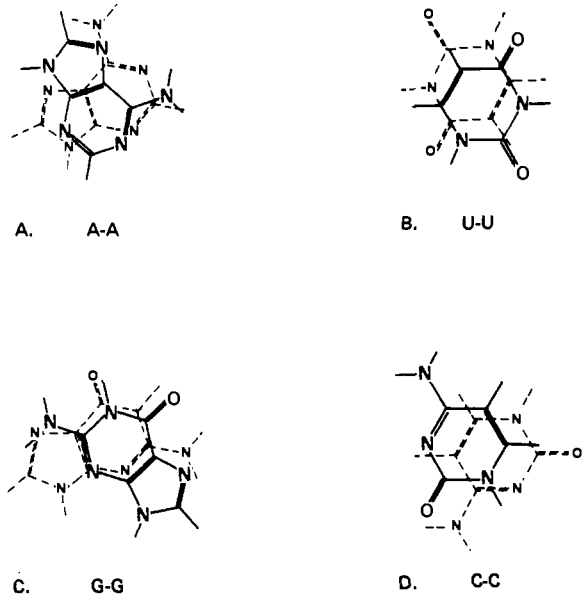

Figure 2. Minimum-energy configuration for stacked adenine-adenine (A), uracil-uracil (B), guanine-guanine (C), and cytosine-cytosine (D) complexes.

New solvent configurations were generated by selecting a molecule at random, translating it along all three Cartesian coordinates, and rotating it by changing three Euler angles, specifying the positions of chlorine atoms about the molecular center. Proper correction for the Jacobian of the polar angle was taken into account.[28] Maximum translational and rotational steps were chosen in such a way that the acceptance ratio was between 0.2 and 0.3.

To speed up convergence, and to increase the variety of structures included in the numerical integration of the configurational integrals, the preferential sampling technique introduced by Owicki and Scheraga[29] was used. This was done by defining a region surrounding the solute consisting of approximately 20% of the total number of solvent molecules and assigning a probability of sampling from this region of 50%. Although there are no theoretical guidelines concerning the optimal choice of the transition matrix in this case, this assignment appears to guarantee sufficient flexibility in both regions of sampling.

Each run consisted of 4×10^6–6×10^6 attempted moves. Initial steps (roughly 1×10^6) were disregarded for equilibration. Every calculation was extended to include as many configurations as were necessary to reduce the statistical error to the level at which calculated energy differences have quantitative significance.

Statistical error was estimated by using the method of subaverages described by Wood.[30] Energy values used in the error analysis were calculated as averages over segments of the Markov chain of configurations of the lengths of $1000N$ attempted moves (N = number of solvent molecules in the sample). Such segments appear to be of sufficient length to ensure that subsequent energy values are not correlated, a necessary condition for the error analysis.

Geometry of Base Complexes. A-U and G-C planar complexes were arranged in the Watson–Crick configuration. Since there are no sufficient experimental data about the geometries of stacked autoassociations of nucleic acid bases, minimum-energy configurations obtained from the advanced quantum-mechanical calculations by Langlet et al.[31] were adopted (Figure 2).

At this point it should be noted that the geometries of the complexes adopted in this study do not necessarily correspond to the lowest energy

(25) See Pohorille, A.; Pratt, L. R.; Burt, S. K.; MacElroy, R. D. *J. Biomol. Struct. Dyn.* **1983**, in press. This paper also includes an overview on solution influence on biomolecular equilibria.

(26) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *J. Chem. Phys.* **1953**, *21*, 1087.

(27) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 533.

(28) Owicki, J. C.; Scheraga, H. A. *J. Am. Chem. Soc.* **1977**, *99*, 7403.

(29) Owicki, J. C.; Scheraga, H. A. *Chem. Phys. Lett.* **1977**, *47*, 600.

(30) Wood, W. W. In "Physics of Simple Liquids"; Temperley, H. N. V.; Rowlison, J. S.; Rushbrooke, G. S., Eds.; North-Holland: Amsterdam, 1968; p 115.

(31) Langlet, J.; Claverie, P.; Caron, F.; Boeuve, J. C. *Int. J. Quantum Chem.* **1981**, *19*, 299.

Table I. Parameters for Potential Energy Calculations

interacting atoms	C, kcal·Å^6/mol	D, kcal/mol	α, Å^{-1}
C···H	84.67	4 874.8	3.582
C···C	618.9	15 465.9	3.117
C···N	686	31 274	3.31
C···O	685	31 274	3.34
Cl···H	388	37 203	3.63
Cl···C	1040	57 319	3.31
Cl···N	1150	115 906	3.50
Cl···O	1150	115 906	3.53
Cl···Cl	1744	212 432	3.51

configurations in solution. In particular, theoretical[31] and experimental[32,33] data indicate that hydrogen-bonded A-U pairs exist in vacuo and in chloroform as an equilibrium of Watson-Crick- and Hoogsteen-type configurations. However, the same studies suggest that the difference in energy between both types of pairing should be very small; thus we do not expect that our energy comparisons will be influenced significantly if the Hoogsteen configuration prevails in CCl_4. The situation for stacked associations is more complicated, since the energy of interaction is a much less sensitive function of molecular orientations and exhibits several weak and broad local minima. Relative stabilities of these minima could be easily influenced by a solvent. Nuclear magnetic resonance measurements for adenine in aqueous solutions,[12,17] where stacked associations are stable, cannot unambiguously determine geometries of complexes but rather can suggest the coexistence of several, vaguely defined, configurations, one of which is very similar to that employed here.[34]

In principle, the potential of mean force or, equivalently, the full solute-solute distribution function could be obtained by allowing the interacting bases to move freely in the solvent. By this procedure the free energy of the complex in solution would be obtained. This approach, however, is computationally impractical because of the inordinate amount of computer time required for equilibration. Moreover, the equilibrated structures obtained would almost entirely contain stable (i.e., planar) complexes of the bases. Since the stacked autoassociations would be found so infrequently, little information about the differences in association between planar and stacked complexes could be obtained. Thus, instead of searching out the most stable configuration in CCl_4 the following question was raised: What is the influence of solvent on the solute associations believed to be the most stable in vacuo?

Intermolecular Potential Functions. The key factor in determining the accuracy of computer simulations is the quality of intermolecular potential functions. These functions are obtained either by empirical methods or from quantum-mechanical calculations, the latter method being used in most of the recent simulations of complex fluids. Unfortunately, although water-solute potentials have been a subject of extensive ab initio molecular orbital studies,[35,36] no such calculations are available for CCl_4. Successful parameterization of interactions involving CCl_4 must correctly describe dispersion energy, which would require calculations carried out beyond the single-determinant Hartree-Fock level. Since the cost of such a procedure would be prohibitive, the empirical approach was adopted.

The interaction energy between two molecules, A and B, was expressed by the pairwise sum of interaction contributions:

$$E_{AB} = \sum_i^A \sum_j^B E_{ij}^{AB} \quad (1)$$

The pair potential function E_{ij} was represented in the form

$$E_{ij}^{AB} = q_i q_j / r_{ij} - C_{ij}/r_{ij}^6 + D_{ij} \exp(-\alpha_{ij} r_{ij}) \quad (2)$$

where q_i and q_j are the net atomic charges on atom i and j, respectively. Atomic charges were obtained from CNDO/2 calculations on isolated molecules.

For CCl_4-CCl_4 interactions, parameters developed by Bates and Busing[37] for crystalline hexachlorobenzene were adopted. These parameters were fitted to obtain the best agreement between observed and calculated lattice energy, lattice parameters, and optically active external mode frequencies and were tested in our study of liquid CCl_4 and $CHCl_3$ (A. Pohorille, S. K. Burt, and R. D. MacElroy, unpublished calculations). The results show that Bates and Busing potentials are superior to other sets of potential functions and are in very good agreement with experimental data. For solute–solvent interactions we used potentials developed by Duchamp (Upjohn Co., unpublished results), who extended the Bates and Busing parameter set to include interactions between chlorine and other atoms of interest. The general description of the parameterization procedure is given by Oie et al.[38] Parameters used in this work are listed in Table I.

In actual application, CCl_4-CCl_4 interactions were truncated at a distance beyond 10 Å between molecular centers. No cutoff was applied for solute-solvent interactions, but the exponential term in (2) was neglected when the interatomic repulsion contribution to the total energy was less than 0.001 kcal/mol.

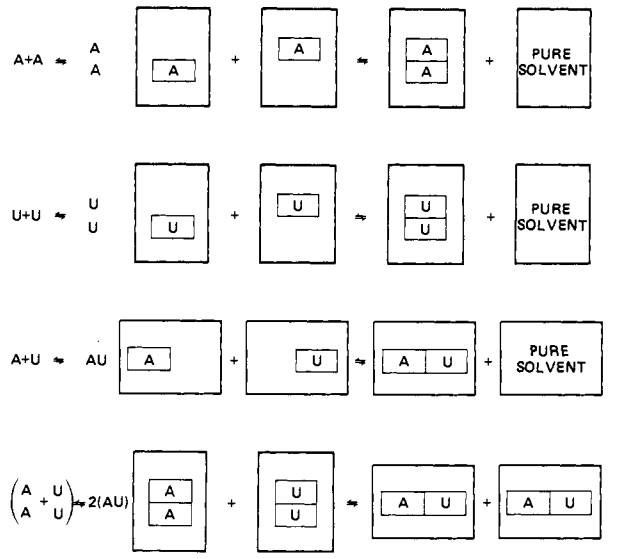

Figure 3. Schemes for various thermodynamical equilibria involving adenine and uracil.

Results and Discussion

General. The average energies per solvent molecule (E_{tot}) calculated from Monte Carlo simulations, as well as the energies of solute–solvent (E_{soln}) and solvent–solvent (E_{solv}) components, are given in Table II. This table also includes the number of solvent molecules N, the total number of MC steps, NSTEP, and the actual number of configurations, $NSTEP_{av}$, used in calculating ensemble averages for every run.

Calculated energy values, as well as various structural parameters, can be further used to analyze complex formations between the nucleic acid bases. The process of association of two solute molecules A and B

$$A + B \rightleftharpoons AB \quad (3)$$

is interpreted in our approach as a transfer of one of the solute molecules from the parent box to the box containing the other solute molecule, leaving pure solvent in the parent box. Similarly, equilibrium

$$AA + BB \rightleftharpoons 2(AB) \quad (4)$$

is expressed as an interchange of molecules A and B between boxes. As an example, various thermodynamical equilibria involving adenine and uracil considered in this study are shown schematically in Figure 3.

The total energy of association, ΔE_{tot}, is defined as

$$\Delta E_{tot} = N_{AB}E_{AB} + N_{ps}E_{ps} - N_A E_A - N_B E_B + \Delta E_{Int} \quad (5)$$

where N_{AB}, N_A, N_B are numbers of solvent molecules and E_{AB}, E_A, and E_B are average energies per solvent molecule for complexes

(32) Katz, L. *J. Mol. Biol.* **1969**, *44*, 279.
(33) Iwahashi, H.; Kyogoku, Y. *J. Am. Chem. Soc.* **1977**, *99*, 7761.
(34) Compare structure A in Figure 2 with structure a in Figure 2 of ref 17.
(35) Clementi, E.; Corongiu, G.; Ranghino, G. *J. Chem. Phys.* **1981**, *74*, 578 (and preceding papers from the same series).
(36) Jorgensen, W. L. *J. Am. Chem. Soc.* **1981**, *103*, 335.
(37) Bates, J. B.; Busing, W. R. *J. Chem. Phys.* **1974**, *60*, 2414.
(38) Oie, T.; Maggiora, G. M.; Christoffersen, R. E.; Duchamp, D. J. *Int. J. Quantum Chem., Quantum Biol. Symp.* **1981**, *8*, 1.

Table II. Summary of Monte Carlo Runs[a]

solute	N (1)	NSTEP (2)	$NSTEP_{av}$ (3)	E_{soln} (4)	E_{solv} (5)	E_{tot} (6)	
CCl_4	175	32000N	22000N		−6.450	−6.450	0.003
stacked complexes							
A	169	29700N	20000N	−0.138	−6.251	−6.389	0.003
A-A	164	26400N	17000N	−0.214	−6.096	−6.310	0.004
U	170	27700N	17000N	−0.114	−6.287	−6.401	0.004
U-U	166	27000N	17000N	−0.190	−6.140	−6.330	0.004
G	168	25500N	15000N	−0.153	−6.209	−6.362	0.004
G-G	163	23200N	15000N	−0.247	−6.000	−6.247	0.004
C	169	26600N	17000N	−0.114	−6.249	−6.363	0.004
C-C	164	26700N	17000N	−0.190	−6.062	−6.252	0.004
planar complexes							
A	169	24900N	15000N	−0.137	−6.249	−6.386	0.004
U	170	26000N	15000N	−0.116	−6.289	−6.405	0.004
A-U	164	35000N	25000N	−0.249	−6.050	−6.299	0.004
G	168	24800N	15000N	−0.254	−6.210	−6.364	0.004
C	169	25100N	15000N	−0.116	−6.249	−6.365	0.004
G-C	163	24700N	15000N	−0.265	−5.911	−6.236	0.004
cavities[b]							
A	169	26700N	17000N		−6.239	−6.239	0.004
U	170	25000N	15000N		−6.283	−6.283	0.004
A-U	164	23500N	15000N		−6.134	−6.134	0.004

[a] All energies in kcal/mol. [b] See Solvent-Solvent Interactions for explanation and discussion.

Table III. Energies of Associations of Nucleic Acid Bases in CCl_4 (in kcal/mol)

equilibrium	ΔE_{soln} (1)	ΔE_{solv} (2)	ΔE_{int} (3)	ΔE_{tot} (4)
A + A ⇌ AA	11.7	−9.9	−6.0	−3.6 ± 1.1
U + U ⇌ UU	7.2	−3.9	−6.3	−3.0 ± 1.3
A + U ⇌ AU	2.2	4.1	−13.2 (−14.5)[a]	−6.9 ± 1.3 (−8.2)[a]
AA + UU ⇌ 2(AU)	−14.5	21.4	−14.1 (−16.7)[a]	−7.2 ± 1.3 (−9.8)[a]
G + G ⇌ GG	11.1	−7.5	−11.7	−8.1 ± 1.3
C + C ⇌ CC	7.4	−4.4	−7.6	−4.6 ± 1.3
G + C ⇌ GC	2.3	3.8	−23.7 (−21.0)[a]	−17.6 ± 1.3 (−14.9)[a]
GG + CC ⇌ 2(GC)	−13.9	19.5	−28.1 (−22.7)[a]	−22.5 ± 1.3 (−17.1)[a]
cavity:[b] A + U ⇌ AU		−12.2		−12.2 ± 1.3

[a] Values in parentheses were obtained by using experimental energies for planar base-pair formation in vacuo.[10] [b] See Solvent-Solvent Interactions for explanation and discussion.

and separated bases. The term E_{ps} is the average energy per molecule in pure solvent; N_{ps}, the number of molecules in pure solvent, is chosen in such a way that

$$N_{ps} = N_A + N_B - N_{AB} \quad (6)$$

This choice of N_{ps} makes the number of solvent molecules on both sides of eq 5 the same. The last term, ΔE_{Int}, represents the energy of interaction between associated bases in the absence of solvent. Since the positions of the solute molecules are kept fixed, ΔE_{Int} remains constant throughout the Monte Carlo process.

The ΔE_{tot} term can also be represented as the sum of energy contributions from solute–solvent (ΔE_{soln}), solvent–solvent (ΔE_{solv}), and solute–solute (ΔE_{int}) interactions:

$$\Delta E_{tot} = \Delta E_{soln} + \Delta E_{solv} + \Delta E_{Int} \quad (7)$$

The total energies of associations, as well as all three components, are given in Table III.

Total Energy. The results shown in the last column of Table III indicate that hydrogen-bonded complexes are preferred over stacked associations and that both are energetically more stable than isolated bases. Not only is this result in qualitative agreement with experiment but also the numerical values are quantitatively consistent with available spectroscopic data (see Table IV). Measurements show that the enthalpy, $\Delta H°$, of planar associations increases, as expected, with increased polarity of the solvent. Thus, $\Delta H°$ in CCl_4 solution should be higher than $\Delta H°$ in vacuo[10] and lower than corresponding values in $CHCl_3$,[5,9] in dimethyl sulfoxide (Me_2SO),[39] and in a Me_2SO/methanol mixture,[40] where presumably competing hydrogen-bond formation between the bases and methanol is observed. This relationship is, indeed, preserved in our calculations.

Table IV. Experimental Enthalpies of Planar Associations of Nucleic Acid Bases (in kcal/mol)

base pair	environment	enthalpy, $\Delta H°$	ref
A-U	gas phase	−14.5	10
A-U[a]	$CHCl_3$	−6.2 ± 0.6	5
A-U	$CHCl_3$	−6.2 ± 0.3	9
A-U	CCl_4	−7.2[b]	4
G-C	gas phase	−21.0	10
G-C	$CHCl_3$	(−10, −11.5)[c]	41
G-C	Me_2SO	−5.8	39
G-C	Me_2SO/methanol	−3.8	40

[a] 1-Cyclohexyluracil and 9-ethyladenine. [b] Estimate based on the assumption that $\Delta S°$ is the same as in $CHCl_3$. [c] Rough estimate based on the assumption of $\Delta S° = 15$ eu.

The only experimental data on CCl_4 solutions come from Kuechler and Derkosch,[4] who found that the association constant for the adenine-uracil dimer is 6 times larger than that obtained in chloroform. If the same entropy of association is assumed for both cases, $\Delta H°$ for the A-U complex in CCl_4 would be equal to −7.2 kcal/mol, a very satisfactory agreement with the value of −8.2 kcal/mol obtained in this work.

Solute–Solvent Interactions. Since CCl_4 is a nonpolar molecule, its interactions with nucleic acid bases, E_{soln}, are dominated by dispersion forces instead of by electrostatics. As a consequence, vertical or "stacked" solute–solvent interactions, where CCl_4 molecules are located above or below the aromatic rings of the bases, are energetically most stable. Such arrangements are strongly favored because they minimize interatomic distances between atoms of solvent and solute molecules. The orientation of a CCl_4 molecule that interacts most strongly with the solute molecule is a "head-to-plane" configuration, in which three chlorine

(39) Newmark, R. A.; Cantor, C. R. *J. Am. Chem. Soc.* **1968**, *90*, 5010.
(40) Petersen, S. B.; Led, J. *J. Am. Chem. Soc.* **1981**, *103*, 5308.
(41) Kyogoku, Y.; Lord, R. C.; Rich, A. *Biochim. Biophys. Acta* **1969**, *179*, 10.

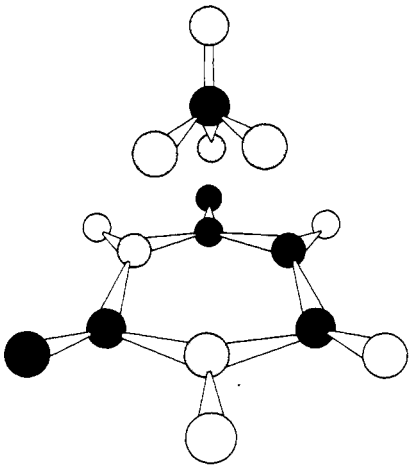

Figure 4. Head-to-plane configuration corresponding to the most favorable orientation of a $ClCl_4$ molecule interacting with uracil.

atoms are directed toward the solute molecule plane (see Figure 4). On the other hand, "in-plane" interactions, in which the carbon atom of CCl_4 molecule remains approximately in the molecular plane of the base, are relatively weak. The difference between energies of vertical and in-plane solute–solvent interactions is shown in Figure 5.

A clear manifestation of the tendency toward vertical interactions between CCl_4 molecules and the aromatic rings of the bases is the rough proportionality between E_{soln} and the planar surface of the solute available to the solvent, as shown in column 4 of Table II. Indeed, E_{soln} for the extended planar A-U and G-C complexes is lower than that for the stacked autoassociations, which, in turn, are lower than E_{soln} for isolated bases. The difference in E_{soln} between stacked complexes and isolated bases is due to an increase in available solute surface when bases are in the stacked configuration. As shown in Figure 2, this increased surface area arises from the fact that the bases do not stack completely but rather are shifted off center, allowing more of the surface to interact with the solvent. It is also worth noting that E_{soln} for purine bases (A, G), which have two conjugated aromatic rings, is lower than the corresponding values for pyrimidine bases (U, C), which have only one ring.

When hydrogen-bonded complexes are formed between separate bases, there is only a small reduction of available solute surface, and the solute–solvent energy increases only slightly (see column 1 of Table III). On the other hand, formation of stacked associations is accompanied by a larger decrease in the available solute surface and, consequently, ΔE_{soln} becomes more unfavorable. This result shows that in CCl_4, solute–solvent interactions favor formation of hydrogen-bonded associations over stacked ones.

Solvent–Solvent Interactions. In studies attempting to calculate energy of association of biological molecules in solution, it is usually assumed that the main contribution to the change of the solvent–solvent free energy or enthalpy comes from the energy of cavitation. This idea originates from estimates of the work that has to be done against the intermolecular forces among the solvent molecules to create a cavity that will accommodate the solute molecule. Therfore, cavitation energy is always positive and usually is considered to be related to the surface tension of a solvent and to decrease with a decrease of the cavity surface area.

Treating cavitation energy as the only factor contributing to the change in the solvent–solvent energy[19-21,23] is an obvious oversimplification for strongly structured solvents like water. The most obvious reason is that the changes in the network of hydrogen bonds must also be taken explicitly into account. The importance of water structure was recently demonstrated by Postma et al.[42] in molecular dynamic simulation carried out for water including cavities of various size. They found that the cavity–water oxygen

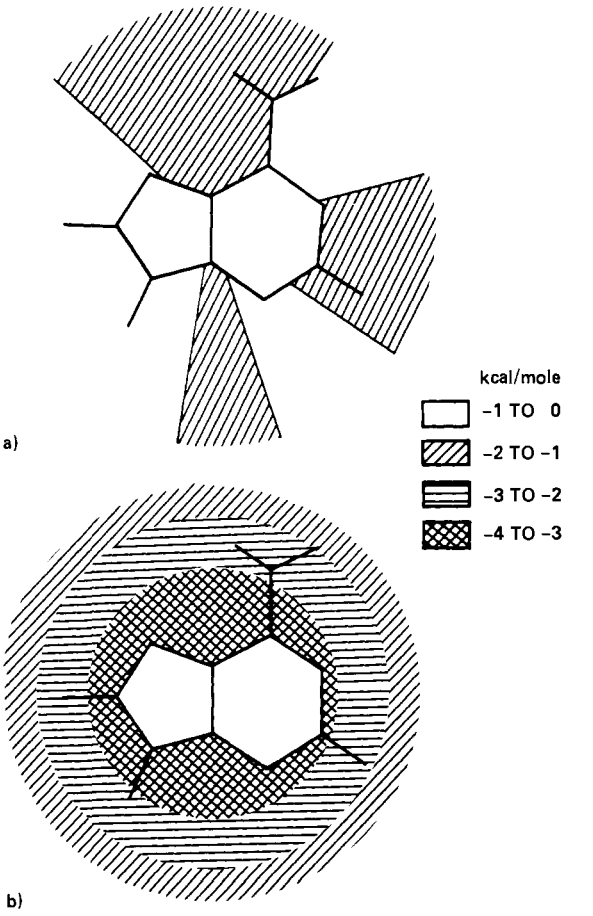

Figure 5. Maps of average in-plane (a) and vertical (b) adenine–CCl_4 interactions obtained from the Monte Carlo calculations in the rectangular box for $N = 169$. In the case of in-plane interactions, solvent molecules within a planar radius of 5 Å were taken into account and the maximum amplitude of the carbon atom was 2 Å from the plane of the solute molecule. For vertical interactions solvent molecules in a cylinder of radius 4.5 Å oriented along the z axis with the maximum $z = 6$ Å were considered.

radial distribution function varies in an irregular manner with the cavity size. This result shows that solvation shell structures are different for different cavity sizes. Another evidence against the cavity approach comes from studies of associations of two argon-like or methane-like molecules in water solution.[43-45] According to the usual view, thermodynamical forces will drive two such species together. Statistical mechanical studies, however, showed that there are two relatively stable positions for the solute pair. In one position, two solutes are nearly touching, whereas in the other position they are separated by a water molecule.

On the other hand, for solvents such as CCl_4, which are not strongly associated and whose structure is determined primarily by short-range packing forces, the neglect of changes in solvent structure may be justified. Monte Carlo simulation gives a unique opportunity to test the validity of this assumption. Results listed in column 2 of Table III indicate that solvent–solvent energy of association ΔE_{solv} favors formation of stacked complexes over isolated bases. This remains in accord with the cavity approach, since the surface area of the stacked bases is smaller than twice the surface area of the separated base. On the other hand, the destabilizing effect of ΔE_{solv} on the formation of hydrogen-bonded complexes cannot be explained the same way, because in this case there is a definite, although small, reduction in the surface area.

To investigate further the relative importance of cavity formation, and of changes in the structure of solvent induced by the

(42) Postma, J. P. M.; Berendsen, H. J. C.; Haak, J. R. *Symp. Faraday Soc.* **1983**, in press.

(43) Pratt, L. R.; Chandler, D. *J. Chem. Phys.* **1977**, *67*, 3683.
(44) Pangali, C.; Rao, M.; Berne, B. J. *J. Chem. Phys.* **1979**, *71*, 2982.
(45) Ravishanker, G.; Mezei, M. Beveridge, D. L. *Symp. Faraday Soc.* **1983**, in press.

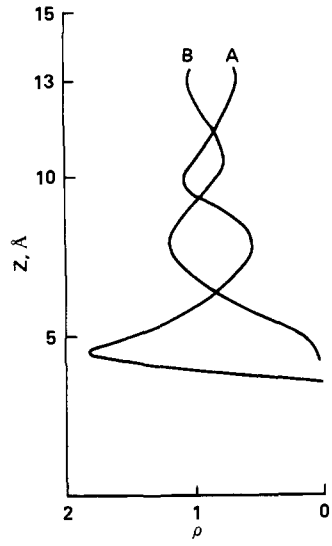

Figure 6. Solvent density fluctuations in a cylinder of radius 4.5 Å oriented along the z axis geometrically centered in the middle of solute. Adenine-uracil pair (A); adenine-uracil cavity (B). Density $\rho = 1$ corresponds to the density of pure solvent.

presence of the solute, additional Monte Carlo calculations were performed. In these cases, adenine, uracil, and an A-U pair were replaced by hard-sphere cavities. The cavities were formed by placing spheres with the appropriate van der Waals radii around every atom in the molecule. The molecular shape generated in this way presented a rigid wall to the solvent molecules. Energies calculated for various cavities are listed in the last three rows of Table II, and the corresponding association energy is given in the last row of Table III. As can be seen, the process of "association" of adenine and uracil cavities is favored by −12 kcal/mol. This result is in contrast to that in which real molecules were placed in the cavities. The difference originates from the fact that placing either a cavity or a real molecule in the pure solvent causes serious local perturbations in the solvent density. This effect is particularly apparent when the density fluctuations along the z axis in a cylinder formed around the molecular plane of the solute are monitored. For example, as shown in Figure 6, presence of the A-U pair (curve A) increases the solvent density in the immediate vicinity of the solute. This increased solvent density arises because of the attractive vertical interactions between CCl_4 and the large A-U plane; such interactions are generally stronger than those between CCl_4 molecules. As a consequence, a local minimum is produced at a distance of 8 Å from the plane, and a secondary maximum is seen at 13 Å.

The pattern for the A-U cavity (curve B) is completely different. Since CCl_4 molecules do not interact with the cavity, they tend to stay inside the solvent rather than near the cavity surface. Consequently, at short distances from the solute plane (where an increase in the number of solvent molecules was observed for the A-U pair), there is a minimum of solvent density. It is worth noting that an increase in solvent density in the vicinity of the solute, which reflects attractive solute–solvent interactions, is usually unfavorable from the point of view of solvent–solvent energy. The reason for this is that a solvent molecule on the surface of the solute has fewer close contacts with other solvent molecules than one in the bulk solvent.

The effect, discussed above, should be also noticeable in solvent–solvent radial distribution functions that uniquely define E_{solv} for pair-additive potentials. Since increased solvent density around the solute diminishes the number of close contacts between solvent molecules, the first peak in the radial distribution functions for the A-U pair should be smaller than the corresponding peak for the A-U cavity. Such a difference, which is small but statistically significant, is shown in Figures 7 and 8 for the Cl–Cl radial distribution function g_{Cl-Cl}, which accounts for most of E_{solv}, and for the C–C radial distribution function g_{C-C}, which represents distribution of molecular centers.

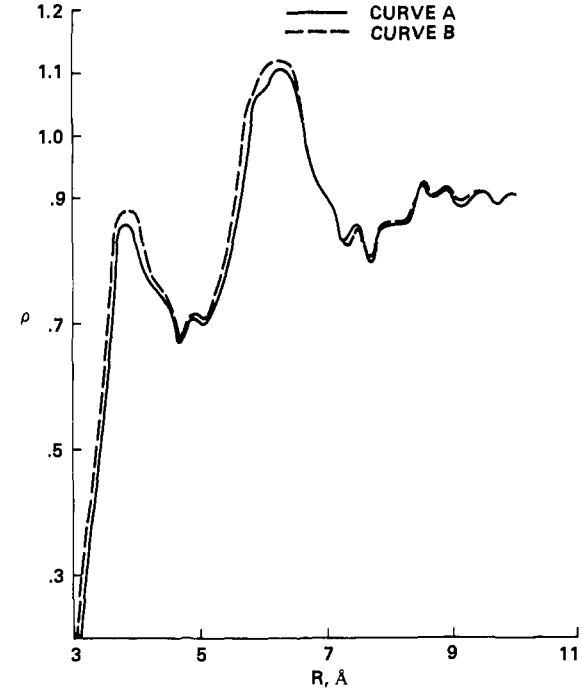

Figure 7. Cl–Cl radial distribution functions for CCl_4 in systems containing 164 solvent molecules and A-U pair (A) or A-U cavity (B).

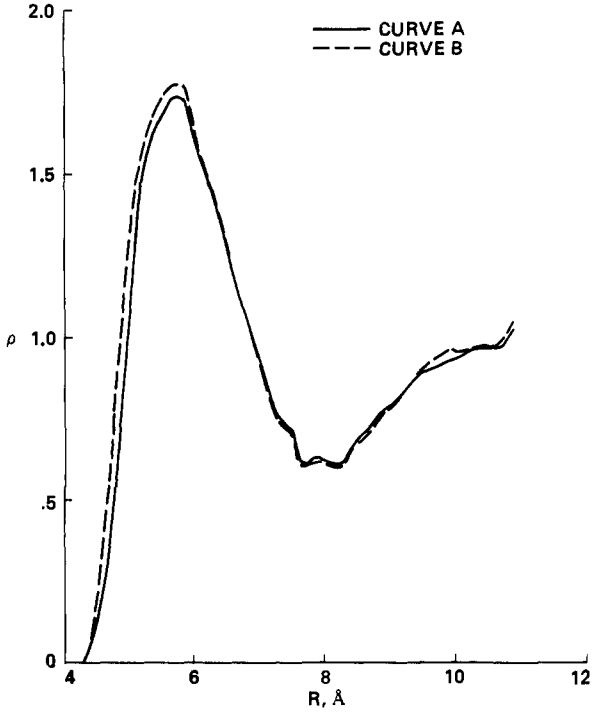

Figure 8. C–C radial distribution functions for CCl_4 in systems containing 164 solvent molecules and A-U pair (A) or A-U cavity (B).

Our analysis suggests that the simple cavity term[46] is insufficient to properly describe changes in E_{solv}. Perturbation of the solvent structure caused by the presence of the solute also gives an important contribution to the total solvent–solvent energy and, therefore, cannot be neglected. This result is of particular interest since in statistical theories it is assumed that correlations between solute and solvent molecules are not very sensitive to inclusion of attractive forces.[47] Although this assumption appears to be justified for small solutes,[47] it does not remain valid when the solute becomes much larger than the size of the solvent.

(46) Sinanoglu, O. *Int. J. Quantum Chem.* **1980**, 18, 381, and references cited therein.
(47) Pratt, L. R.; Chandler, D. *J. Chem. Phys.* **1980**, *73*, 3434.

Solute–Solute Interactions. The solute–solute energy of interaction ΔE_{int} favors formation of hydrogen-bonded base pairs over the stacked associations (see column 3 in Table IV). The values of ΔE_{int} calculated for planar complexes, which are stable in vacuo,[31] are in satisfactory agreement with experimental data obtained by Yanson et al.,[10] shown in Table IV. Attention should be drawn to the fact that solute–solute energy in vacuum can be used as a solvent-independent value, because of the assumption of pairwise additivity of the intermolecular potentials, in which all many-body contributions are neglected.

Qualitative Comparison of CCl_4 and Water–Solute Interactions. The quantitative analysis of all aspects of influence of solvent on base associations is possible only by the means of the "complete" computer simulation when a large number of solvent molecules are explicitly taken into account. However, very instructive qualitative information can be obtained by using the simple method of a single-molecule probe. In this method, the Monte Carlo process is performed for a single solvent molecule enclosed within a certain volume around a solute molecule. Such calculations, which give information about the average energy of interaction between the solute and the solvent molecule located in this region of space, have been performed for molecules of water and CCl_4 around four nucleic acid bases and their complexes. Potentials developed by Clementi[48] were used to calculate solute–water interactions. Solvent molecules were located either in various positions around the molecular plane of the solute (with the maximum amplitude of the central atom 2 Å from the plane), or over the plane. This approach is analogous to that employed by Cieplak et al.[24] in their perturbative study and provides information very similar to that obtained from the single molecule maps used by Clementi et al.[48] and Clementi and Corongiu.[49]

In the case of CCl_4, the results are essentially the same as those shown in Figure 5, indicating a clear tendency for vertical interactions instead of in-plane configurations. For water–nucleic acid base interactions, the situation is completely different. The in-plane interactions are generally strong and vary markedly with the position of the water molecule around a solute. Hydrophilic (around N and O atoms) and hydrophobic (around C–H bonds) regions are very distinctive. In hydrophilic regions a water molecule forms one or two hydrogen bonds with a solute molecule, and consequently its energy is the same or lower than the average energy of water molecules in the bulk solvent. Vertical interactions, on the other hand, are much weaker, and no strong directional specificity of water molecules is observed. As an example, maps of average energies of water molecule for both in-plane and vertical interactions with guanine are shown in Figure 9.

The influence of solvent on molecular associations can be interpreted in the following manner. When interactions between two solute molecules occur, solvent molecules are removed from the region of interaction and released to the bulk solvent. The stronger the solute–solvent interactions, the greater is the loss of solute–solvent energy upon association. Such an association can be energetically favorable only if the solute–solute, and the new solvent–solvent, interactions compensate for a loss in the solute–solvent energy.

In CCl_4, planar interactions between nucleic acid bases are more favored than stacked complexes because they are accompanied by removal of CCl_4 molecules from the region of interaction. Such solvent molecules are relatively weakly bound to the solute. At the same time, CCl_4 molecules, which strongly interact with a solute vertically, remain essentially unperturbed. This contribution, along with the intrinsically higher stability of planar pairs, overcompensates for less favored solvent–solvent interactions. In the case of water solutions, formation of hydrogen-bonded complexes is accompanied by highly unfavorable in-plane dehydration of the main hydrophilic centers of the bases. On the other hand, for stacked associations, no strong solute–water interactions are eliminated. Both competition for hydrogen-bonding centers and the larger surface area make planar complexes less favorable than stacked, despite the fact that solute–solute planar energy is lower than the stacking energy.

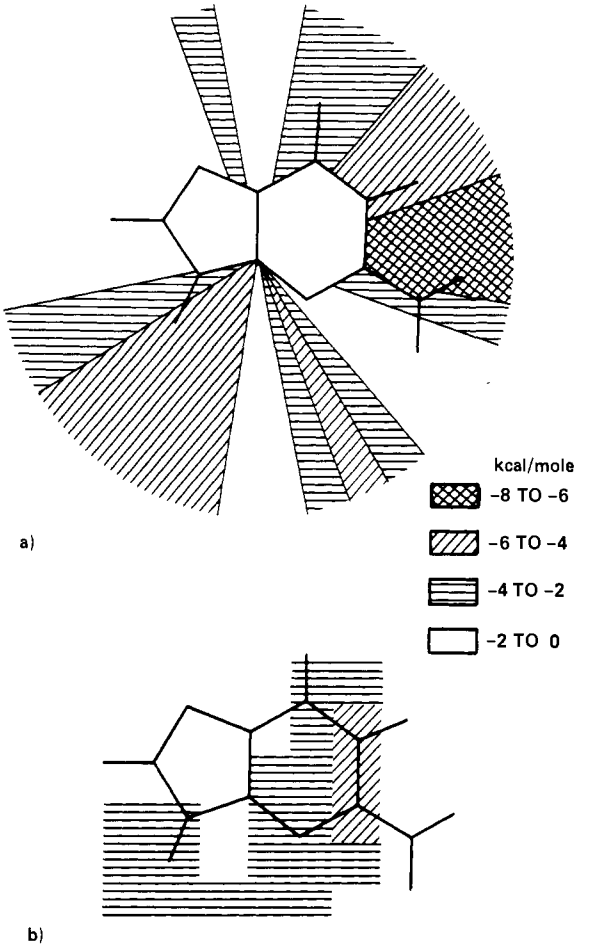

Figure 9. Maps of average in-plane (a) and vertical (b) guanine–water interactions obtained from the single-molecule probe.

Hydrogen-bonded complexes in CCl_4 and stacked complexes in water are more stable than noninteracting bases, because the solute–solvent energy lost by solvent molecules when they are removed from the region of interaction to the bulk solvent is less than the newly gained solvent–solvent energy. Similar arguments for stacking in CCl_4 and hydrogen bonding in water do not lead to such unambiguous conclusions, because the old solute–solvent and the new solvent–solvent energies of solvent molecules removed from the region of interaction are of the same order of magnitude. For the CCl_4 case, the Monte Carlo calculations indicate that stacked bases are slightly more stable than noninteracting bases. Similarly, the problem of stability of hydrogen-bonded pairs in water can be solved only by "complete" computer simulation, wherein various subtle effects contributing to the solute–solvent and solvent–solvent energies are explicitly included.

Accuracy of Calculations. The three most important factors that determine the accuracy of Monte Carlo calculations are the quality of intermolecular potentials, statistical fluctuations of calculated ensemble averages, and the sample-size effect. The first two were briefly discussed under Intermolecular Potential Functions and Monte Carlo Procedure. The third factor arises because locating a limited number of molecules in a box followed by subsequent application of periodic boundary conditions introduces an error into the molecular correlations. For a given system, this effect decreases with an increase in the sample size. Recently, a theory of the influence of periodic boundary conditions on equilibrium properties has been developed[50] and applied to simple liquids.[51] However, in most cases of interest we do not

(48) Clementi, E.; Cavallone, F.; Scordamaglia, R. *J. Am. Chem. Soc.* **1977**, *99*, 5545.
(49) Clementi, E.; Corongiu, G. *J. Chem. Phys.* **1980**, *72*, 3979.
(50) Pratt, L. R.; Haan, S. W. *J. Chem. Phys.* **1981**, *74*, 1864.

know how to choose the size of the system in order to minimize an effect of periodic boundary conditions. The most straightforward test is to perform a series of calculations in which the sample size is systematically increased until calculated values remain unchanged. Since it requires enormous computer time, this method is obviously not of practical use for complex solutions. On the other hand, sample size effects are a reason for legitimate concern in such quantitative computer simulations as this one. To assess the magnitude of the error caused by periodic boundary conditions, calculations were performed for pure CCl_4 in cubic boxes containing 125 and 343 molecules, as well as for A and A-A stacked complexes placed in these boxes. The results were compared with those given in Tables II and III. This comparison showed that solute–solvent energies in boxes of various sizes and shapes are virtually identical. Solvent–solvent energies, on the other hand, wer influenced to a small, yet significant, degree by sample size. In the case of pure solvent, the difference in energy per molecule between $N = 125$ and $N = 175$ was found to be 0.03 kcal/mol. However, the error caused by periodic boundary conditions appears to cancel when boxes of the same size are used throughout a series of calculations. For stacked A-A complexes in boxes of $N = 125$ and $N = 343$, $\Delta E_{tot} = -3.9$ and -3.0 kcal/mol, respectively, and remains in a good agreement with the value of -3.6 kcal/mol obtained for $N = 175$. Thus, it seems justifiable to assume that the effect of periodic boundary conditions on ΔE_{tot} can be neglected in our case, providing that energies for the same sample size are consistently used in (5).

Conclusions

The difference in base-pair associations between nonpolar solvents and aqueous solutions can be understood in terms of the interplay between solute–solvent and solvent–solvent interactions. In particular, it appears that solute–solvent interactions in CCl_4 and water are of similar orders of magnitude but that the interactions favor different base association schemes. These differences are caused by the different nature of the dominant forces acting in both solvents. In nonpolar solvents, in which dispersion forces dominate, planar complexes stabilized by the maximum exposure of molecular planes to the solvent are favored. In polar solvents, where the dominant contribution to energy comes from electrostatics, the most stable complexes are compact, stacked, associations that allow the maximum number of hydrophilic centers to be exposed to the solvent.

The presence of a solute strongly influences the structure of solvent, even for simple nonpolar liquids like CCl_4. In consequence, for solutes that are not strictly solvophobic, solvent–solvent energy is not necessarily reduced when surface area decreases. Thus, the simple cavity concept is insufficient to explain changes in solvent–solvent energy.

Our analysis of solvent effect on base associations differs markedly from that of Egan et al.,[23] who dealt with the same problem. Although their predictions of preferred conformations of base complexes were in qualitative agreement with experiment, it should be pointed out that according to their results, all complexes are less stable than separated bases. On the other hand, several of our conclusions agree with observations made by Cieplak et al.[24] in their simple but very insightful study. In particular, they were the first to point out that although the in-plane solute–solvent interactions are preferred in water, vertical interactions are more favored in nonpolar solvents.

Computer simulations of solutions, relatively new in theoretical biochemistry, still remain an expensive and imperfect method. In particular, further work is needed to improve the accuracy of intermolecular potentials and to investigate the effect of sample size on various properties of the system. On the other hand, results of simulations offer unique insight into the microscopic structure of solutions and into the role of various forces in formation of stable structures.

Acknowledgment. We thank Dr. Lawrence Pratt for many helpful discussions and Dr. Ignacio Tinoco for his comments. This work was supported in part by the NASA Ames-University Consortium NCA2-OR050-204.

Registry No. Adenine, 73-24-5; uracil, 66-22-8; guanine, 73-40-5; cytosine, 71-30-7; CCl_4, 56-23-5.

(51) Pratt, L. R.; Haan, S. W. *J. Chem. Phys.* **1981**, *74*, 1873.