

# Molecular Dynamics Simulations Study of the Free Energy of Association of 9-Methyladenine and 1-Methylthymine Bases in Water

Liem X. Dang and Peter A. Kollman\*

Contribution from the Department of Pharmaceutical Chemistry, School of Pharmacy, University of California, San Francisco, California 94143. Received April 10, 1989

**Abstract:** The results of molecular dynamics simulations on the free energy of association of adenine and thymine bases in water are presented. The influence of the solvent on the free energy surfaces (potentials of mean force) of the hydrogen-bonded and stacked interactions of these base pairs is determined with the use of molecular dynamics in conjunction with the statistical perturbation theory of Zwanzig.<sup>1</sup> The stacked-configuration potential of mean force in water is found to be more stable than the hydrogen-bonded configuration, in agreement with previous theoretical and experimental studies on these systems. The calculated  $\Delta G_b$  for base-pair association by stacking is  $-0.8$  kcal/mol, in good agreement with the experiment value of  $-1.15$  kcal/mol.

## I. Introduction

It is well-known that two of the main factors responsible for the stabilization of the DNA double helix are the hydrogen-bonding and stacking interactions between nucleic acid bases. Consequently, there have been many theoretical and experimental studies on these DNA constituents. Theoretical studies on these systems were initiated by Sinanoglu and Abdunur;<sup>2</sup> they concluded that hydrophobic forces play a dominant role in the stacking interaction of nucleic acid bases and thereby contribute to the overall stability of DNA helices in aqueous solution. The Monte Carlo simulations of Pohorille et al.<sup>3</sup> were aimed at comparing the influence of solvent on nucleic acid-base associations in nonpolar solvents ( $\text{CCl}_4$ ) and in water. These studies focused on the study of structures and energetics of the solvent-solvent and solvent-solute interactions; but no calculation of the free energies of base-pair association was carried out. Recently, Cieplak and Kollman<sup>4</sup> have carried out an extensive calculation on the binding free energies of adenine-thymine (AT) and guanine-cytosine (GC) base pairs in vacuo and in solution using thermodynamic cycle perturbation and molecular dynamics methods. They concluded that, in vacuo, the hydrogen-bonded configuration of base pairs is more favorable, whereas, in aqueous solution, the stacked configuration is more favorable. These results were in agreement with available experimental measurements.<sup>5</sup> The experimental evidence on associations of the base pairs in water, gas phase, and nonpolar solution<sup>6</sup> suggested that in nonpolar solutions such as  $\text{CCl}_4$ <sup>7</sup> and  $\text{CHCl}_3$ <sup>8,11</sup> and in the gas phase,<sup>12</sup> bases associate mainly by hydrogen bonding, whereas in water stacked associations are preferred.<sup>13,14</sup> Experimental studies have included the NMR study

of Sarma et al.,<sup>15</sup> which indicated that there are several broad local conformational minima available to the stacked configuration in water. This and the other studies have contributed a great deal to the understanding of base-pair associations in the gas phase and in solution. As noted by Cieplak et al.,<sup>4</sup> their calculation had very large statistical errors and only gave the net free energy of association, not the free energy of association as a function of coordinate. In order to evaluate the free energy as a function of coordinate, one needs to employ a "coordinate coupling" approach in conjunction with statistical perturbation theory.<sup>1</sup> This has been done by Jorgensen et al.<sup>16</sup> and Brooks et al.<sup>17</sup> and has been applied to the study of the conformation of *n*-butane in the gas phase and in water and also of the potential of mean force (PMF) of an argon pair in liquid argon. Here, we present results on the studies of the PMF of hydrogen-bonded and stacked base-pair interactions in water using this approach. The main goal of this study is to determine quantitatively the solvation effects on the association between base pairs. In section II we outline the computational methodology. The results are presented and discussed in section III, and section IV summarizes our conclusions.

## II. Methodology

The calculations used the statistical perturbation theory originally developed by Zwanzig;<sup>1</sup> the details of the theory have been presented in the literature.<sup>17,18</sup> However, we briefly review the approach here.

According to Zwanzig,<sup>1</sup> the potential energy of the system can be decomposed into two parts

$$U_1 = U_0 + \Delta U \quad (1)$$

where  $U_0$  is the potential energy of an unperturbed system and  $\Delta U$  is the perturbation. Then the Helmholtz free energy difference between the two systems is

$$A_1 - A_0 = -kT \ln \frac{Q_1}{Q_0} \quad (2)$$

(1) Zwanzig, R. W. *J. Chem. Phys.* **1954**, *22*, 1420, and references cited therein.

(2) Sinanoglu, O.; Abdunur, S. *Photochem. Photobiol.* **1964**, *3*, 333.

(3) Pohorille, A.; Pratt, L. R.; Burt, S. K.; MacElroy, R. D. *J. Biomol. Struct. Dyn.* **1984**, *1*, 1257. Pohorille, A.; Burt, S. K.; MacElroy, R. D. *J. Am. Chem. Soc.* **1984**, *106*, 402.

(4) Cieplak, P.; Kollman, P. A. *J. Am. Chem. Soc.* **1988**, *110*, 3734.

(5) Nakano, N. I.; Igarashi, S. *J. Biochemistry* **1970**, *9*, 577.

(6) Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag: New York, 1984.

(7) Kucher, E.; Derkosch, J. Z. *Naturforsch., B: Anorg. Chem., Org. Chem. Biochem., Biophys., Biol.* **1966**, *B21*, 209.

(8) Pitha, J.; Jones, R. N.; Pithova, P. *Can. J. Chem.* **1966**, *44*, 1045.

(9) Miller, J. H.; Sobell, H. M. *J. Mol. Biol.* **1967**, *24*, 345.

(10) Katz, L.; Penman, S. *J. Mol. Biol.* **1966**, *15*, 220.

(11) Binford, J. S., Jr.; Holloway, D. M. *J. Mol. Biol.* **1968**, *31*, 91.

(12) Yanson, I. K.; Teplitsky, A. B.; Sukhodub, L. F. *Biopolymers* **1979**, *15*, 1149.

(13) Ts'o, P. O. P.; Melvin, I. S.; Olson, A. C. *J. Am. Chem. Soc.* **1963**, *85*, 1149.

(14) Imoto, T. *Biochim. Biophys. Acta* **1977**, *475*, 409.

(15) Dhingra, M. M.; Sarma, R. In *Biomolecular Structure, Conformation Function, and Evolution*; Srinivasan, R., Yathindra, N., Subramanian, E., Eds.; Pergamon Press: New York, 1981; Vol. 2, pp 221-250.

(16) Jorgensen, W. L.; Buckner, J. K. *J. Phys. Chem.* **1987**, *91*, 6083.

(17) Tobias, D. T.; Brooks, C. L., III. *Chem. Phys. Lett.* **1987**, *142*, 472; *J. Chem. Phys.* **1988**, *89*, 5115.

(18) McQuarrie, D. A. *Statistical Mechanics*; Harper and Row: New York, 1976.

(19) Singh, U. C.; Weiner, P. K.; Caldwell, J. W.; Kollman, P. A. AMBER(UCSF), version 3.0; Department of Pharmaceutical Chemistry, University of California: San Francisco, 1986.

(20) True, J. E. *J. Chem. Educ.* **1969**, *46*, 12. Justice, M. C.; Justice, J. C. *J. Solution Chem.* **1976**, *5*, 543.

(21) Jorgensen, W. L. *J. Am. Chem. Soc.* **1989**, *111*, 3770.

(22) Weiner, S. J.; Kollman, P. A.; Nguyen, D.; Case, D. A. *J. Comput. Chem.* **1986**, *7*, 230.

(23) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D. *J. Chem. Phys.* **1983**, *79*, 926.

in which  $Q_1$  and  $Q_0$  are the configuration integrals of the perturbed and unperturbed systems, respectively. Substituting the expressions for  $Q_0$  and  $Q_1$  into eq 2 and after performing some algebraic manipulation, we get

$$A_1 - A_0 = -kT \ln \frac{\int \exp(-\beta U_1) dr^N}{\int \exp(-\beta U_0) dr^N} \quad (3)$$

$$A_1 - A_0 = -kT \ln \frac{\int \exp(-\beta U_0) \exp(-\beta \Delta U) dr^N}{\int \exp(-\beta U_0) dr^N} \quad (4)$$

or

$$A_1 - A_0 = -kT \ln \langle \exp(-\beta(U_1 - U_0)) \rangle_0 \quad (5)$$

Here,  $\langle \dots \rangle_0$  indicates the average is calculated corresponding to the potential energy describing state 0,  $k$  is the Boltzmann constant,  $T$  is the absolute temperature, and  $\beta = 1/kT$ . Furthermore, let  $U_0 = U(r)$ , and

$$U(r) = U_{\text{BB}}(r) + U_{\text{WB}}(r) + U_{\text{WW}}(r) \quad (6)$$

where  $U_{\text{BB}}$ ,  $U_{\text{WB}}$ , and  $U_{\text{WW}}$  are the potential energies of base-base, water-base, and water-water, respectively.

From eq 5 and 6, the free energy changes when the base-pair separations are changed during the dynamics simulations from  $r$  to  $r + dr$  and  $r$  to  $r - dr$  are

$$A(r \pm dr) - A(r) = -kT \ln \langle \exp(-\beta(U_{\text{BB}}(r \pm dr) + U_{\text{WB}}(r \pm dr) - U_{\text{BB}}(r) - U_{\text{WB}}(r))) \rangle \quad (7)$$

For each simulation for a given  $r$ , eq 7 gives two free energy values for movement along the coordinate  $r$ . When a series of such calculations are performed at different values of  $r$ , the PMF's as a function of the coordinate  $r$  can be obtained.

In calculating the PMF's, we have chosen the distance between N3 of thymine and N1 of adenine and the distance between base centers of mass as the reaction coordinates for the hydrogen-bonded and stacked configurations, respectively. The orientations of these base pairs in water were chosen graphically to maximize stacking and were not changed as the bases were separated. They are shown in Figures 3 and 4. During the dynamics simulations, both base pairs are moved along the reaction coordinate by  $\pm 0.125$  Å; thus, a distance of 0.5 Å is covered in each simulation. The simulated time consisted of 6-ps equilibration followed by 12 ps of data collection for each base-pair separation. A timestep of 1.5 fs was used, and the SHAKE<sup>24</sup> procedure was adapted to constrain all the bond lengths to their equilibrium values. For these simulations, a nonbonded cutoff of 9 Å was used to reduce calculation complexity, and the simulations were carried out with the canonical ensemble (constant  $T$ ) and a temperature of 300 K. A total of 12 simulations were carried for each base-pair configuration.

Subsequently, the PMF's are obtained by joining the results of each simulation at the end point and used to calculate the base-pair association constants and the binding free energies via the following equations<sup>20</sup>

$$K_a = N \int_0^{r_c} 4\pi r^2 \exp(-W(r)/kT) dr \quad (8)$$

and

$$\Delta G_b = -kT \ln K_a \quad (9)$$

Here,  $W(r)$  is the base-pair potential of mean force,  $r_c$  is the geometric limit for association, and  $N$  is the Avogadro number. Although eq 8 was originally derived to determine the association constants for the spherically symmetric ion-pair complexes in solution, it has been used successfully recently by Jorgensen<sup>21</sup> in the calculations of the association constants for amides in solution.

To implement the computational method into the study of the base-pair associations in water, we have chosen a model that consists of an adenine and thymine base pair immersed in a periodic box of roughly 900 water molecules. The system was minimized and then equilibrated with molecular mechanics and molecular dynamics methods, respectively, within AMBER 3.0. The force field of Weiner et al.<sup>22</sup> was used for the bases and base pairs while the TIP3P<sup>23</sup> model was used for the water-water interactions.

An issue is what to use for water-base hydrogen-bonded interactions. In the Weiner et al.<sup>22</sup> force field hydrogen-bonding distances  $\text{H}\cdots\text{X}$  are

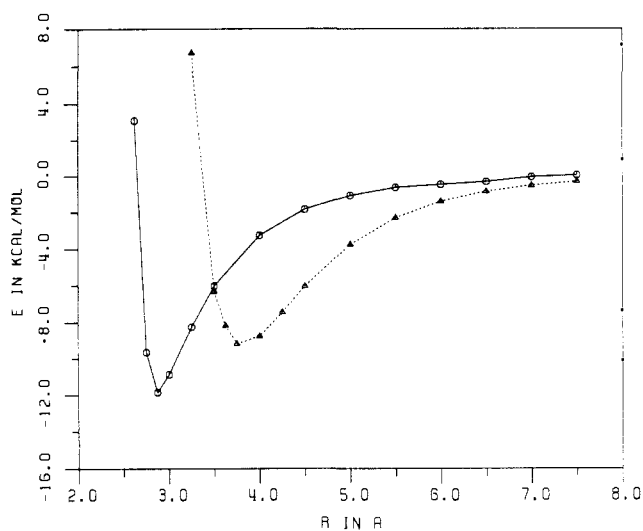


Figure 1. Gas-phase intermolecular potentials of an AT base pair as a function of the reaction coordinate. The solid line is the hydrogen-bonded and the dotted line is the stacked base-pair configuration.

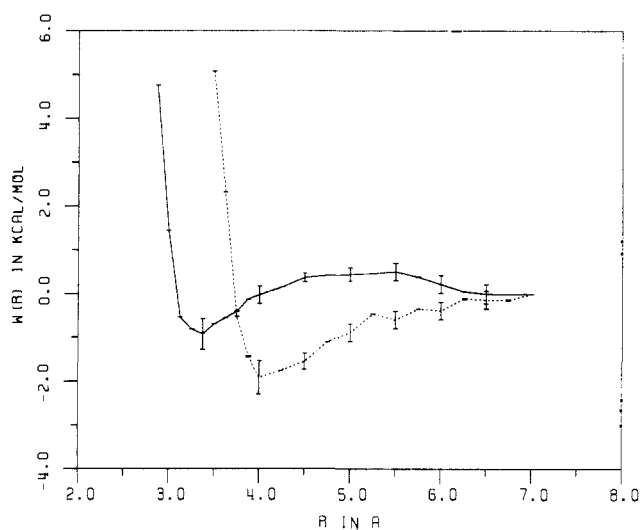


Figure 2. Potentials of mean force of an AT base pair obtained from molecular dynamics simulations in water at 300 K. Error bars were obtained by averaging over the first and last half of the data for each simulation. The solid line is the hydrogen-bonded and the dotted line is the stacking base-pair configuration.

calibrated by a function with inverse 10 and 12 powers for attractive and repulsive interactions in addition to the  $\text{H}\cdots\text{X}$  electrostatic attraction. In the TIP3P model,  $\text{H}\cdots\text{O}$  interactions have only electrostatic interactions. In these calculations, we used no 10-12 parameters for water-base interactions. Since the TIP3P water hydrogen HW has zero 6-12 parameters, the hydrogen-bonded interactions between water hydrogen HW and the base hydrogen-bond acceptor O and N atoms were handled analogously to water-water interactions. But the hydrogen-bonding hydrogen of the bases, the thymine N3-H (atom type H) and the adenine N(6) hydrogens (atom type H2) have small but nonzero 6-12 van der Waals parameters ( $R^* = 1.00$  Å,  $\epsilon = 0.02$  kcal/mol) in the Weiner et al. force field.<sup>22</sup> We first carried out simulations using these 6-12 parameters on the hydrogens (MOD1). Our expectation was, however, that such a model would lead to a too weak hydrogen bond with water as a proton acceptor to these two types of hydrogen. Molecular mechanics calculations of water hydrogen bonding to adenine ( $\text{N6-HN6}\cdots\text{OH}_2$ ) led to an  $\text{H}\cdots\text{O}$  distance of 1.86 Å and  $\Delta E = -11.17$  kcal/mol; water hydrogen bonding to thymine ( $\text{N3-H3}\cdots\text{OH}_2$ ) led to an  $\text{H}\cdots\text{O}$  distance of 2.28 Å and  $\Delta E = -6.87$  kcal/mol. To examine the sensitivity of the potential function, we repeated the calculations with  $\epsilon = 0.0$  kcal/mol for H and H2 atom types, which is as done for hydrogen-bonding hydrogens in the TIP3P model. This led to hydrogen bond distances of 1.79 and 1.89 Å and energies of  $-11.17$  and  $-7.66$  kcal/mol for  $\text{N6-HN6}\cdots\text{OH}_2$  and  $\text{N3-H3}\cdots\text{OH}_2$ , respectively. We feel this latter model (MOD2) is more balanced, given the way water-water hydrogen bonds are handled, so we report the results in detail for MOD2 below. We also discuss the former

(24) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* 1977, 23, 326.

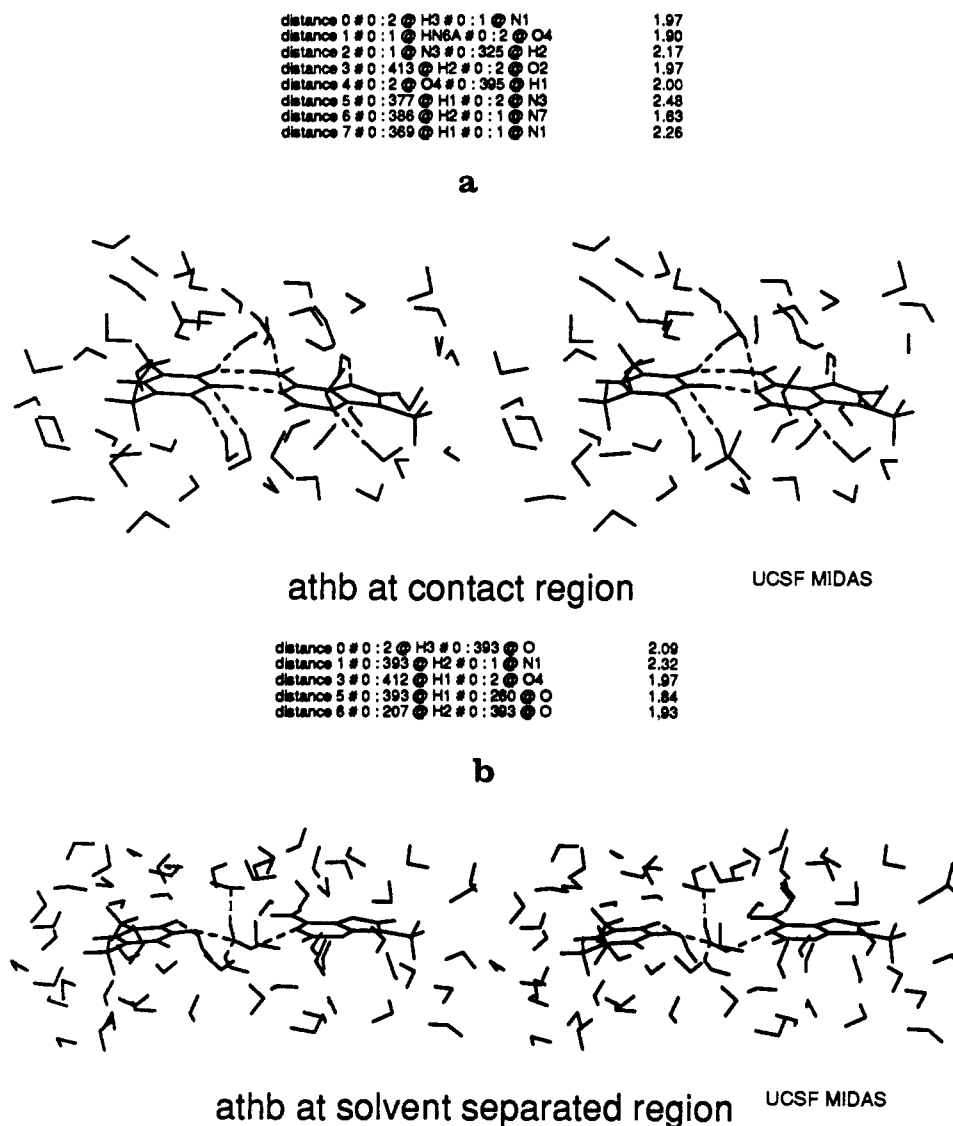


Figure 3. Snapshot of water molecules within 4.5 Å of base pair for the hydrogen-bonded simulations: (a) at the contact region, (b) at the solvent-separated region.

model (MOD1) for comparison, since the hydrogen bonds it forms are also reasonable.

### III. Results and Discussion

We present in Figure 1 the gas-phase intermolecular potentials for the hydrogen-bonded and stacked configurations as a function of reaction coordinate. The minimum in the hydrogen-bonded configuration comes mainly from stabilizing electrostatic interactions,<sup>25</sup> whereas the stabilization of the stacked structure is primarily due to van der Waals interactions. The base-pair PMF's in water for the hydrogen-bonded and the stacked configurations determined from molecular dynamics simulations obtained by joining the results of each window at the endpoint are shown in Figure 2. In displaying the PMF's in the figure we have chosen  $W(7.0 \text{ \AA}) = 0$ , instead of  $W(\infty) = 0$ . It is a reasonable approximation since the energy differences in the region between 6.5–7.0 Å are found to oscillate around zero and their values are within the numerical errors of the calculations. As one can see, the PMF's clearly indicate that the stacked configuration is more favorable than the hydrogen-bonded configuration in solution. This result is understandable, because water molecules have more access to the stacked base pair in solution than the corresponding hydrogen-bonded configuration. Since the water molecules can interact more favorably with the hydrogen-bonding groups of the stacked

base pair, this allows the stacked base pair to retain its hydrogen bonding in water as the base pairs approach, but water–base hydrogen bonds must be broken as the base–base hydrogen bonds form. On the other hand, the hydrogen-bonded configuration is more favorable in the gas phase, as seen in Figure 1. We reiterate again the conclusion of ref 4 that it is the van der Waals stabilization of the bases (through both dispersion and exchange repulsion/hydrophobic effects)<sup>2</sup> that stabilizes the stacked configuration over the hydrogen bonded in water.

On analyzing the individual PMF's, we noted that the contact minimum of both base-pair PMF's is shifted to the right upon going from the gas phase to solution. The effect is more pronounced in the hydrogen-bonded PMF and is probably due to the water destabilizing the hydrogen-bonding interaction of the base pairs. In Figures 3 and 4, we present several snapshots of water around the contact and the solvent-separated regions of the base pair. We indicate in these figures the hydrogen-bonding interactions between base pairs and base–water. In the hydrogen-bonded configuration, there are two hydrogen bonds between the base pairs and they are C4–O4...HN6A and N3–H3...N1 (see Figure 3a). These hydrogen bonds stabilize the base-pair PMF at the contact region. However, they break during the dissociation process, but these atoms formed hydrogen bonds with water molecules as seen in the solvent-separated region (i.e., C4–O4...HOH and HOH...N1). Moreover, we observed a layer of water molecules between the base pairs when they are separated by 6.5 Å (see Figure 3b). On the other hand, in the stacked structure,

(25) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghlo, C.; Alagona, G.; Profeta, S.; Weiner, P. *J. Am. Chem. Soc.* **1984**, *106*, 765.

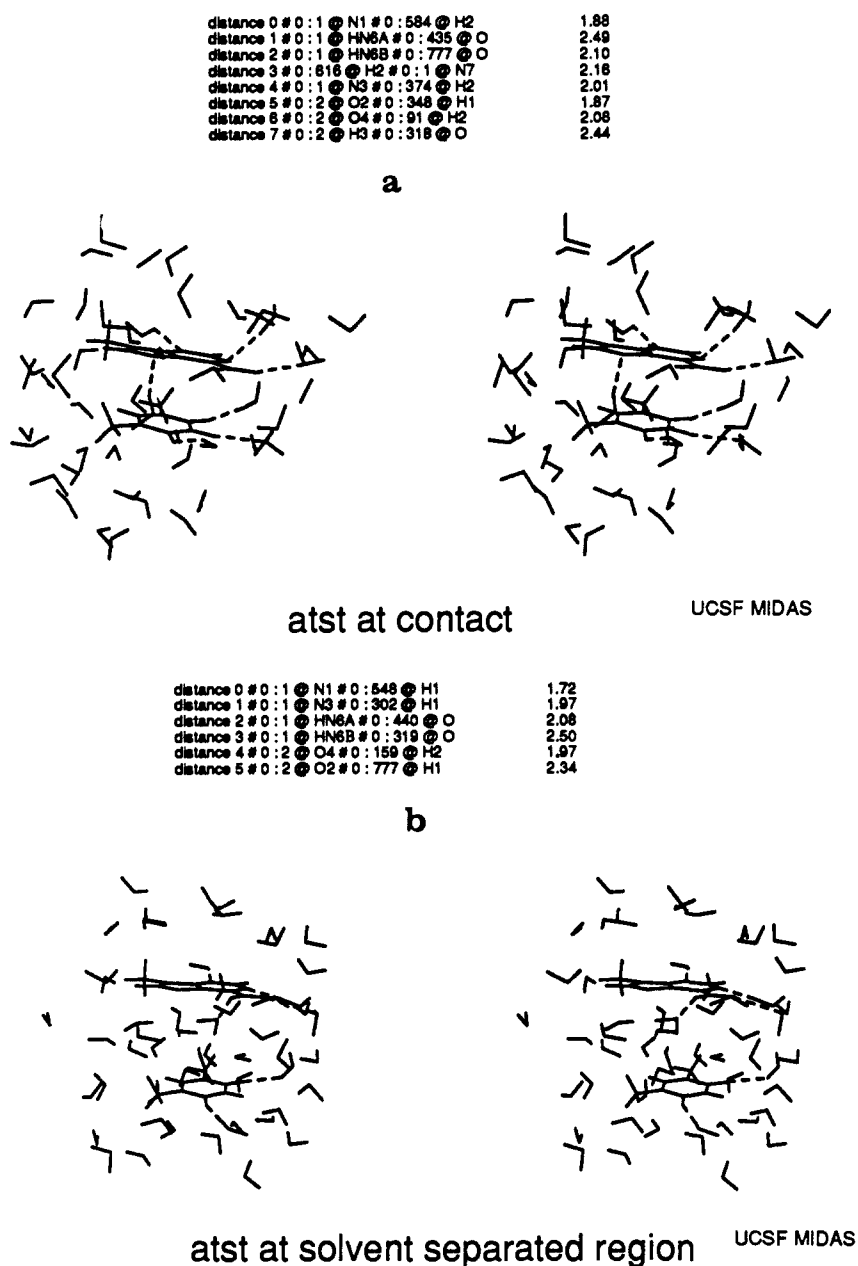


Figure 4. Same as Figure 3 for the stacked base-pair simulations.

waters form hydrogen bonds with the base atoms such as HOH...N1 and C4-O4...HOH (see Figure 4a) in the contact region, and most of these hydrogen bonds remain stable throughout the entire simulation as seen in Figure 4b. We did not see a clear minimum for the stacked configuration in the solvent-separated region as had been found in the study of the benzene-benzene PMF by Beveridge et al.<sup>26</sup> In that study, the stacked benzene-benzene PMF is found to have two equivalent minima in the contact and solvent-separated regions. Thus, there may be a fundamental difference between the hydrophilic base associations studied here and the more hydrophobic benzene-benzene association. There is a hint of a maximum in the  $W(r)$  for the hydrogen-bonded configuration at 5.5 Å (see Figure 2), but this is within the statistical noise of the calculations.

The base-pair association constants and binding free energies for both models are tabulated in Tables I and II. As one can see in both models, the association constants and binding free energies for the stacked configuration are nearly independent of the choice of the integration limit  $r_c$ . On the other hand, the values

Table I.  $\Delta G_b$  and  $K_a$  for MOD1

no.	$r_c$ (Å)	$\Delta G_b$ (kcal/mol)	$K_a$ (L/mol)
(a) Hydrogen-Bonded Configuration			
1	4.0	0.22	0.70
2	4.5	0.16	0.77
3	5.0	0.13	0.81
4	5.5	0.09	0.86
5	6.0	0.05	0.92
6	6.5	-0.02	1.05
7	7.0	-0.12	1.22
(b) Stacked Configuration			
1	5.0	-0.71	3.24
2	5.5	-0.75	3.50
3	6.0	-0.80	3.80
4	6.5	-0.83	4.00
5	7.0	-0.86	4.20
6	expt	-1.15	6.80

of  $K_a$  and  $\Delta G_b$  for the hydrogen-bonded configuration appeared to depend on the  $r_c$ . In any case, the values obtained here are in qualitative agreement with the calculations of Cieplak and Kollman<sup>4</sup> ( $-0.42 \pm 3.0$  and  $-1.86 \pm 2.0$  kcal/mol) and with the

(26) Mezei, M.; Mehrotra, P. K.; Beveridge, D. L. *J. Am. Chem. Soc.* **1985**, *107*, 2239.

Table II.  $\Delta G_b$  and  $K_a$  for MOD2

no.	$r_c$ (Å)	$\Delta G_b$ (kcal/mol)	$K_a$ (L/mol)
(a) Hydrogen-Bonded Configuration			
1	4.0	0.77	0.27
2	4.5	0.65	0.34
3	5.0	0.56	0.39
4	5.5	0.48	0.45
5	6.0	0.38	0.53
6	6.5	0.23	0.68
7	7.0	0.08	0.87
(b) Stacked Configuration			
1	5.0	-0.61	2.76
2	5.5	-0.68	3.10
3	6.0	-0.74	3.43
4	6.5	-0.78	3.67
5	7.0	-0.81	3.85
6	expt	-1.15	6.80

stacked configuration, in good agreement with experiment. This agreement suggests that the binding energies obtained for the other base pairs (i.e., GC and AA) by Cieplak and Kollman<sup>4</sup> are also qualitatively reliable. Furthermore, there are small but noticeable differences in  $\Delta G_b$  and  $K_a$  for the hydrogen-bonded configuration in the two models. This result is qualitatively reasonable because the model MOD1 has weaker water-base hydrogen bonding compared to water-water and base-base hydrogen bonding than the model with zero 6-12 parameters on H and H2. Thus, the association of bases, in which base-base hydrogen bonds replace water-base hydrogen bonds, would be more favorable in the model with nonzero 6-12 parameters. Nonetheless, either model gives results that are in qualitative agreement with experiment and the calculations of Cieplak et al.<sup>4</sup>

Both MOD1 and MOD2 favor the stacked configuration by roughly 1.0 kcal/mol. This is likely to be a lower bound for this energy difference, because we did not allow the bases to have different angular orientation in the stacked conformation. Using eq 8 and 9 has led to a calculated  $\Delta G_b$  association in water for AT base pairs of -0.7 to -0.8 kcal/mol, in good agreement with the experimental value of -1.15 kcal/mol. Any discrepancies in the comparison with experiment likely come from two sources: first, inaccuracies in the force field, and, second, approximation inherent in the use of eq 8 and 9, which were derived assuming spherical ion pairs. Obviously, a more complete sampling of

different stacked configuration might reveal one or more configurations more stable than the assumed one and thus lead to a more negative  $W(r)$ . On the other hand, it is likely that some angular directions for "stacking" are more unfavorable, and considering this in the angular integration leading to eq 8 would make  $\Delta G_b$  less negative. Although we cannot rule out large errors that may cancel fortuitously, the difference between calculated and experimental  $\Delta G_b$  values is small enough to be encouraging. The hydrogen-bonded configuration has a more well-defined orientation. Thus, considering conformational flexibility would lower the energy of the stacked conformation more than the hydrogen-bonded one.

#### IV. Conclusion

The PMF's for the associations of bases in water were calculated with the "coordinate coupling" approach in conjunction with the statistical perturbation theory of Zwanzig.<sup>1</sup> The binding energies were in qualitative agreement with experimental data and confirmed the previous calculations using same Hamiltonian but a different approach to calculate the free energy. More importantly, the calculations reproduced the pattern observed experimentally that the stacked configuration is more favorable than the hydrogen-bonded configuration in water. Thus, the combination of the force field of Weiner et al.<sup>22</sup> and the TIP3P<sup>23</sup> model used here is a reasonable set. We also compared two hydrogen-bonding models for base...OH<sub>2</sub> interactions and show that they gave semi-quantitatively similar answers for the PMF's. We plan to use these techniques to obtain the PMF's for the GC base pair in water. Further research includes the study of the PMF's of base pairing in a nonpolar solvent such as CCl<sub>4</sub>. Experimental evidence indicated that the hydrogen-bonded configuration is more favorable than the stacked configuration, as has been found by Jorgensen<sup>21</sup> in PMF simulations of *N*-methylacetamide dimerization in CHCl<sub>3</sub>. We will report these results in the forthcoming communication.

**Acknowledgment.** Most of the calculations were carried out at the San Diego Supercomputer Center and the Pittsburgh Supercomputer Center, supported by the National Science Foundation. Research support from the NIH (Grant CA-25644) to P.A.K. is acknowledged. L.X.D. thanks Dr. Piotr Cieplak for many useful discussions during the course of this work.

**Registry No.** H<sub>2</sub>O, 7732-18-5; 9-methyladenine, 700-00-5; 1-methylthymine, 4160-72-9.