group exists, with the exception of the phenyl group, which naturally has a greater s character since it is already considered to be sp^2 hybridized. Density plots of the Mn–R bond in **1** for $(CO)_5MnCH_3$, $(CO)_5MnCF_3$, and $(CO)_5MnH$ appear in Figure 2, plots a, b, and c, respectively. An immediate contrast between parts a and b of Figure 2 is apparent. First, the Mn–CH_3 LMO (2a) is noticeably more delocalized toward the Mn atom than is the Mn–CF_3 LMO (2b). Second, delocalization of the alkyl lone pair into neighboring carbonyl $2\pi^*$ antibonding orbitals is clearly present in the Mn–CH_3 LMO, but absent in the Mn–CH_3 LMO. The delocalization of the methyl lone pair into the carbonyl $2\pi^*$ orbital, while the trifluoromethyl group does not, is a manifestation of the greater basicity of the methyl group. The qualitative difference between the Mn–CH_3 and the Mn–CF_3 bonds is in complete agreement with the conclusions[42] reached by Hall and Fenske. The Mn–H LMO (2c) shows both greater delocalization to the metal and the carbonyl $2\pi^*$ orbitals than the methyl group (2a). This is probably due to the shorter Mn–H distance and the greater angular flexibility of the hydrogen s orbital, compared to the sp hybrid of the methyl lone pair.

The LMO's of $(CO)_4CoCH_3$ were also calculated. The atomic populations for Co and C(methyl) in the Co–CH_3 LMO are 0.37 e and 1.60 e, respectively. The corresponding Mn and C(methyl) atomic populations in the Mn–CH_3LMO are 0.40 e and 1.53 e. This comparison further exemplifies the greater acidity of the $(CO)_5Mn$ moiety compared to $(CO)_4Co$. The calculated hybridizations at the metal in these two LMO's are $s^{1.0}p^{1.3}d^{3.5}$ and $s^{1.0}p^{1.2}d^{1.0}$ for the Mn–CH_3 and Co–CH_3 LMO's, respectively. These results provide insight as to why the $(CO)_4Co^+$ moiety is a weaker base than $(CO)_5Mn^+$. The smaller amount of metal 3d character in the Co–CH_3 LMO requires greater utilization of the less energetically favorable 4s and 4p AO's in forming the metal–alkyl bond. The decrease in the participation of the Co 3d AO's in the Co–CH_3 LMO relative to the Mn–CH_3 LMO is a direct result of the d^8 configuration of $(CO)_4CoCH_3$ versus the d^6 configuration of $(CO)_5MnCH_3$.

Conclusions

We have presented a theoretical study of the effects of a large set of different substituents upon the kinetic and thermodynamic behavior of the group migration reaction in alkylpentacarbonylmanganese(I) systems. This study was accomplished by first performing appropriate geometry optimizations of the relevant conformations (**1**–**5**) for each substituent at the PRDDO level of theory. In most instances the PRDDO optimized geometries agreed very well with experiment. More limited geometry optimizations were carried out at the ab initio HF level of theory. The relative energetics obtained at the PRDDO level were then reevaluated by ab initio HF theory and were found to agree qualitatively well with the ab initio calculations. These calculated energetic quantities included the following: the energy of isomerization of **1** to **3** and **1** to **4**, the estimated barrier for the two isomerizations, and the relative energy change for the carbonylation reaction (eq 2). It was found that electron-releasing alkyl groups should facilitate group migration and stabilize the product formed, while electron-withdrawing alkyl groups should retard the migratory step and destabilize the product formed. Traditionally, the observed trends in these substituent effects have often been rationalized in terms of metal–alkyl bond strengths; however, we have presented an alternative and more general rationale, which employs simple Lewis acid–base arguments. More basic alkyl groups tend to interact more favorably with the CO $2\pi^*$ orbitals, while less basic alkyl groups interact to a lesser extent with CO $2\pi^*$ orbitals. This leads to a reduced tendency for electron-withdrawing alkyls to migrate relative to electron-releasing alkyls. This acid–base picture is supported by our calculations of the overlap populations, degrees of bonding, group charges, optimized metal–alkyl bond lengths, and localized molecular orbitals. The kinetic stability of the formyl complex **3** is found to be much less than that of the corresponding acyl complexes.

Acknowledgment. We thank the Robert A. Welch Foundation (Grant Y-743), Cray Research Inc., and the Organized Research Fund of the University of Texas at Arlington for their support of this work. We also thank the University of Texas System Center for High Performance Computing for the use of their computational facilities.

Registry No. **1** (R = CF_3), 13601-14-4; **1** (R = CF_2H), 15022-34-1; **1** (R = CH_2F), 15337-23-2; **1** (R = $CH_2C_6H_5$), 14049-86-6; **1** (R = CH_3), 13601-24-6; **1** (R = CH_2CH_3), 15694-83-4; **1** (R = $CH_2CH_2CH_3$), 15628-53-2; **1** (R = $CH(CH_3)_2$), 113686-65-0; **1** (R = C_6H_5), 13985-77-8; **1** (R = H), 16972-33-1.

---

[42] Hall, M. B.; Fenske, R. F. *Inorg. Chem.* **1972**, *11*, 768.

Calculation of the Free Energy of Association of Nucleic Acid Bases in Vacuo and Water Solution[1]

Piotr Cieplak[†] **and Peter A. Kollman***

Contribution from the Department of Pharmaceutical Chemistry, School of Pharmacy, University of California, San Francisco, California 94143. *Received July 27, 1987*

Abstract: The free energy perturbation/molecular dynamics (FEP/MD) method has been applied to study the solvation effect on the association of the nucleic acid bases in vacuo and in water solution. Five complexes were considered: the adenine–thymine Watson–Crick H-bonded pair, an adenine–thymine stacked pair, an adenine–adenine stacked pair, the guanine–cytosine Watson–Crick H-bonded pair, and a guanine–cytosine stacked pair. The stacked complexes were calculated to be slightly more stable than the H-bonded complexes in water, whereas in vacuo the H-bonded complexes are favored, which is consistent with available experiments.[1] Furthermore, the magnitude of the calculated free energies of nucleic acid base association in water (ΔG = 0 to −2 kcal/mol) is consistent with the magnitude found experimentally. This shows that it is now possible to calculate ab initio the free energies of noncovalent association of complex molecules in solution.

It is well-known that two of the main factors responsible for the stabilization of the DNA double helix are the hydrogen bond and stacking interactions between nucleic acid bases. Several lines of experimental evidence confirm the importance of these types of interactions. Among them are the studies on the analysis of the dependence of melting temperature[2] on base content of helices as well as recent studies on the influence of dangling (unpaired) nucleotide ends and terminal base pairs on stability and helix

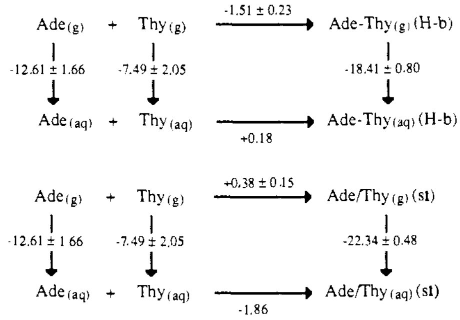
---

[†] Permanent address: Quantum Chemistry Laboratory, Department of Chemistry, University of Warsaw, Pasteura 1, 02-093 Warsaw, Poland.

[1] Nakano, N. I.; Igarashi, S. J. *Biochemistry* **1970**, *9*, 577.

[2] Crothers, D. M.; Zimm, B. H. *J. Mol. Biol.* **1964**, *9*, 1.

0002-7863/88/1510-3734$01.50/0



$\Delta\Delta G_{\text{experimental}} = -1.15$

Figure 1. Thermodynamic cycle for adenine–thymine complex in H-bond and stacked configurations.

formation.[3-5] These latest studies provide useful thermodynamic data that allow the partitioning of the free energy of base pair formations into "stacking" and "pairing" contributions.

There is considerable experimental evidence on the details of association of the base pairs in water, gas phase, and nonpolar solutions.[6] Various studies suggest that in nonpolar solution such as CCl_4,[7] $CHCl_3$[8-11] and in the gas phase,[14] bases associate mainly by hydrogen bonding, whereas in water solution, stacked configurations are preferred.[12-19]

A considerable number of theoretical studies have been published on the subject of base pair association. Simple minimization[20-22] and Monte Carlo simulations[23,24] were performed to study base association. In those studies the discussion of the relative stability of different associates was based on consideration of total interaction energies, i.e. solute–solvent, solvent–solvent, and their differences between different geometries of complexes. In one case where the continuum model was used to model polar (H_2O) and nonpolar (CCl_4) solutions of base pairs, Egan et al.[25] were able to give the description of association in terms of free energy differences. Although those studies correctly predicted the relative tendency of stacked configurations to be favored over H-bonded configurations in water and H-bonded to be favored over stacked in CCl_4, such solvation models are still not as accurate as desirable. According to those theoretical studies, base pairs do not tend to associate in water, because the calculated ΔG for base association in water is positive. This is probably because the continuum model is not able to deal with specific solute–solvent interactions.

Since the perturbation free energy method incorporated into molecular dynamics or Monte Carlo simulation programs is becoming a reliable technique for analyzing solvation effects,[26-31] we present here an approach to studying the base pair association problem using the thermodynamic cycle presented in Figure 1. The application of the molecular dynamics and free energy perturbation method allows the calculation of free energy changes for $X_g + Y_g \rightarrow XY_g$, $Y_g \rightarrow Y_{aq}$, $X_g \rightarrow X_{aq}$ and $XY_g \rightarrow XY_{aq}$. Those data allow one to calculate ΔG for the association processes in water. Thus, a first goal is to compare this calculated ΔG association with experiment. Finding good agreement between theory and experiment would suggest that our potential functions and our method of calculating ab initio free energies of association are reasonable.

A second aim of studying this thermodynamic cycle is to obtain additional information on the energetics of H-bond formation in vacuo and to compare the affinity of base pair and separate bases for water. Those affinities could be measured by the equilibrium constant for transfer from solution to the vapor phase.[32]

Methods

Our calculations were performed with the molecular simulation program package AMBER-UCSF (version 3.0).[33]

In the free energy perturbation method,[29,31] the free energy differences between two states A and B, which were described by the Hamiltonians H_A and H_B, were calculated by summing up the contributions to the free energy (eq 1) calculated during the hypothetical transformation of A (λ = 0) into the B (λ = 1) state. This process could be defined in terms of the set of intermediate states coupled to the dimensionless parameter λ in the range 0–1 (eq 2). The ΔG_λ in eq 1 can be calculated by performing molecular dynamics or Monte Carlo simulation using the windowing and/or slow-growth techniques. In the case of the windowing method, the ΔG_λ is obtained from eq 3 where $\langle\rangle_\lambda$ is a mean value calculated over the state defined by the parameter λ (window) and λ' defines a neighboring state in λ and $\Delta H(\lambda \rightarrow \lambda') = H(\lambda') - H(\lambda)$. For each λ, separate MD simulations with equilibration and data collection stages are performed. In the case of the so-called slow-growth method, i.e. where each individual MD step constitutes a separate intermediate state differing from the previous one, the free energy differences come from summing up the individual free energy differences for each step as in ref 31 (eq 4). The above procedures have been extensively tested for several

$$\Delta G = \sum_{\lambda=0}^{\lambda=1} \Delta G_\lambda \tag{1}$$

$$H_\lambda = \lambda H_B + (1 - \lambda)H_A \tag{2}$$

$$\Delta G_\lambda = -RT \ln \langle \exp(-\Delta H(\lambda \rightarrow \lambda')/kT)\rangle_\lambda \tag{3}$$

$$\Delta G = \sum_{\lambda=0}^{\lambda=1} H_{\lambda'} - H_\lambda \tag{4}$$

---

(3) Freier, S. M.; Alkema, D.; Sinclair, A.; Neilson, T.; Turner, D. H. *Biochemistry* **1985**, *24*, 4533.

(4) Freier, S. M.; Kierzek, R.; Caruthers, M. H.; Neilson, T.; Turner, D. H. *Biochemistry* **1986**, *25*, 3209.

(5) Freier, S. M.; Sugimoto, N.; Sinclair, A.; Alkema, D.; Neilson, T.; Kierzek, R.; Caruthers, M. H.; Turner, D. H. *Biochemistry* **1986**, *25*, 3214.

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

(7) Kuchler, 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) Kyogoku, Y.; Lord, R. C.; Rich, A. *Science (Washington, D.C.)* **1966**, *154*, 518.

(13) Kyogoku, Y.; Lord, R. C.; Rich, A. *Proc. Natl. Acad. Sci. U.S.A.* **1967**, *57*, 250.

(14) Yanson, I. K.; Teplitsky, A. B.; Sukhodub, L. F. *Biopolymers* **1979**, *18*, 1149.

(15) Ts'o, P. O. P.; Melvin, I. S.; Olson, A. C. *J. Am. Chem. Soc.* **1963**, *85*, 1289.

(16) Schweizer, M. P.; Broom, A. D.; Ts'o, P. O. P.; Hollis, D. P. *J. Am. Chem. Soc.* **1968**, *90*, 1042.

(17) Solie, T. N.; Schellman, J. A. *J. Mol. Biol.* **1968**, *33*, 61.

(18) Van Holde, K. E.; Rosetti, G. P. *Biochemistry* **1967**, *6*, 2189.

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

(20) Langlet, J.; Claverie, P.; Caron, F.; Boeuve, J. C. *Int. J. Quantum Chem.* **1981**, *19*, 229.

(21) Poltev, V. I.; Shulyupina, N. V. *J. Biomol. Struct. Dyn.* **1986**, *3*, 739.

(22) Hobza, P.; Sandorfy, C. *J. Am. Chem. Soc.* **1987**, *109*, 1302.

(23) Pohorille, A.; Pratt, L. R.; Burt, S. K.; MacElroy, R. D. *J. Biomol. Struct. Dyn.* **1984**, *1*, 1257.

(24) Pohorille, A.; Burt, S. K.; MacElroy, R. D. *J. Am. Chem. Soc.* **1984**, *106*, 402.

(25) Egan, J. T.; Nir, S.; Rein, R.; MacElroy, R. *Int. J. Quantum Chem. Quantum Biol. Symp.* **1978**, *5*, 433.

(26) Postma, J. P. M.; Berendsen, H. J. C.; Haak, J. R. *Faraday Symp. Chem. Soc.* **1982**, *17*, 55.

(27) Lybrand, T. P.; McCammon, J. A.; Wipff, G. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 833.

(28) Jorgensen, W. L.; Ravimohan, C. *J. Chem. Phys.* **1985**, *83*, 3050.

(29) Singh, U. C.; Brown, F. K.; Bash, P. A.; Kollman, P. A. *J. Am. Chem. Soc.* **1987**, *109*, 1607.

(30) Bash, P. A.; Singh, U. C.; Brown, F. K.; Langridge, R.; Kollman, P. A. *Science (Washington DC)* **1987**, *235*, 574.

(31) Bash, P. A.; Singh, U. C.; Langridge, R.; Kollman, P. A. *Science (Washington D.C.)* **1987**, *236*, 564.

(32) Cullis, P. M.; Wolfenden, R. *Biochemistry* **1981**, *20*, 3024.

(33) 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.

Table I. ΔG_{assoc} (in kcal/mol) for the Gas-Phase Association of the Base Pairs

		ΔG_{assoc} (gas phase)									
		H-bonded A:T		stacked A/T		H-bonded G:C		stacked G/C		stacked A/A	
		A + T → AT	AT → A + T	A + T → A/T	A/T → A + T	G + C → GC	GC → G + C	G + C → G/C	G/C → G + C	A + A → A/A	A/A → A + A
run I	electrostatic part	−7.85	7.85	0.14	−0.13	−25.40	25.40	−3.94	4.01	−0.08	0.03
	nonbonded part	−1.04	1.04	−7.98	7.98	9.71	−9.66	−7.58	7.59	−7.76	7.77
	total	−8.89	8.89	−7.85	7.85	−15.69	15.74	−11.53	11.60	−7.84	7.79
run II	electrostatic part	−8.34	8.34	0.14	−0.14	−25.08	25.08	−3.94	4.01	−0.09	0.12
	nonbonded part	−0.95	0.95	−7.75	8.09	9.65	−9.63	−7.53	7.54	−7.51	7.51
	total	−9.29	9.29	−7.61	7.96	−15.43	15.45	−11.48	11.55	−7.59	7.63
mean $\Delta G_{X+Y\rightarrow XY}$		−9.09 ± 0.23		−7.82 ± 0.15		−15.58 ± 0.16		−11.54 ± 0.05		−7.71 ± 0.12	
corr for vibratnl entropy changes during complex formn		9.43		10.05		10.65		10.78		10.85	
corr due to expressing energy in diff std states, RT ln (1/22.4)		−1.85		−1.85		−1.85		−1.85		−1.85	
resultant ΔG_{assoc}		−1.51 ± 0.23		0.38 ± 0.15		−6.78 ± 0.16		−2.61 ± 0.05		1.29 ± 0.12	

cases where accurate exerimental data are available.[26-31] We stress that since Gibbs free energy is a state function, the computed free energy differences should be path independent (i.e. should not depend on the manner in which the "mutation" from state A to B is performed). We also assume that in our calculation the changes in ΔG are mainly due to the potential energy, and contributions from kinetic energy change cancel out and/or are negligible.

Both techniques, windowing and slow growth, were applied in our studies. To model the association (or rather dissociation) process for the system X–Y in vacuo, the charges and the van der Waals parameters for one of the molecules were decreased to zero during the MD simulation. To model the solvation of X and Y or the X–Y complex, a simulation was done on the system with the solute fully represented and then its electrostatic and van der Waals parameters decreased to zero. In most cases, calculations of the electrostatic contribution to the free energy were done by the windowing procedure, whereas slow growth was applied to obtain the van der Waals contribution. In the case of the in vacuo simulations of the H-bonded complexes, we performed our perturbation from the complete structure into the structure that "disappeared" totally one of the bases (i.e. T or C) in rather short simulations, which covered totally 20 ps of the evolution in time and used slow growth without decoupling[31] the electrostatic and van der Waals contributions. This was done to keep the H-bonded molecules close enough to each other and to not allow such complexes to disperse or bend into the stacked-type associates. In all the other cases the windowing method was used for the electrostatic energy, and the perturbation free energy calculation was carried out in a series of 21 consecutive values of the parameter λ spanning its full range of $\langle 0,1 \rangle$ with 500 steps of equilibration and 500 steps of data collection with a time step 0.002 ps performed in each window. This was followed by a calculation of the van der Waals contribution to the free energy using the slow-growth method over 21 000 steps with $\Delta t = 0.002$ ps.

For the in vacuo MD simulation we checked the dependence of the convergence of the results on the length of the simulation. Comparison of the simulation with different total time of simulation and time step revealed that, for example, for A–T pair in stacking configuration in vacuo, the results change insignificantly when Δt was switched from 0.001 to 0.002 ps and/or the total time of simulation was changed from 42 to 168 ps. The simulations with and without moderate constraints for distances and dihedral angles between bases were also compared. In general, using constraints was not necessary, but when their use was necessary, we carried out the simulations at a series of constraint weights and extrapolated to zero constraint.

The molecular dynamics simulations in solution were carried out at $T = 300$ K [1 atm pressure (NPT ensemble)] in a water bath to which periodic boundary conditions were applied. The TIP3P water model[34] was used. The SHAKE procedure was used to constrain all covalent bonds at equilibrium lengths.[35] In all vacuo and in solution simulations we used a dielectric constant equal to 1 in the molecular mechanics model[36] to calculate the electrostatic interaction energy. The assumption in the solution simulation is that the explicit water model will lead to a correct representation of electrostatic effects. The TIP3P model was derived with $\epsilon = 1$.

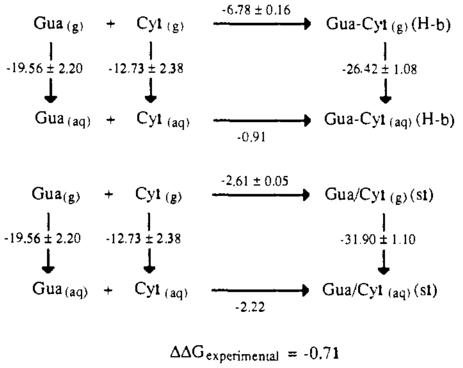


Figure 2. Thermodynamic cycle for guanine–cytosine complex in H-bond and stacked configurations.

The van der Waals and electrostatic parameters for the nucleic acid bases were taken from Weiner et al.[36] In all our calculations, N1- or N9-methylated bases at the all atom levels were used. When combining the TIP3P model for H_2O with the Weiner et al.[36] model for nucleic acids, one could either use zero van der Waals interactions between H-bonding hydrogens and H-bonding accepting heteroatoms (as in TIP3P) or include explicit 10–12 H-bonding interactions (as in Weiner et al.). We have chosen to use the simpler model, i.e., zeroing the H-bond van der Waals interactions. We should note that, in the Weiner et al. and TIP3P models, H-bond energies come exclusively[34] or nearly so[36] from the electrostatic terms in the interaction function. As discussed in ref 36, such an approximation is supported as reasonable based on results from ab initio quantum mechanical calculations.

The starting base pair configurations were taken to be those of a B-DNA double helix. The final configurations after perturbation were checked to ensure that the complex remained in the qualitative structure in which it had begun. Two independent simulations using different velocities and equilibrations were done to check the dependence of starting geometry, at least near a "B-DNA" structure for stacking and H bonding.

We needed to calculate the translational/rotational/vibrational entropy loss upon gas-phase dimerization, since our procedure of zeroing charges and van der Waals parameters for one of the gas-phase monomers does not take this term into account. We approximated this entropy change by minimizing the energy of monomers and dimers to low gradients and then diagonalizing the mass-weighted force constant matrix to determine the normal modes. These were then used to calculate the conformational entropies. The results were evaluated with the thermodynamic cycles in Figures 1–3.

Results and Discussion

In Table I are presented the calculated association free energies for the base pairs in vacuo due to disappearing charges and van der Waals parameters of one of the bases. The total energies and components in this table seem intuitively reasonable, with the

---

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

(35) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327.

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

Table II. Results of the Normal-Mode Analysis for N-Methylated Nucleic Acid Bases and Their Complexes in T = 300 K [Total Entropies for the Systems (cal/mol K)]

molecular system	S monomers	S H-bond complex	S stacked complex	$-T\Delta S_{\text{H bond}} = -T(S_{\text{H bond}} - S_{\text{monomers}})$, kcal/mol	$-T\Delta S_{\text{stack}} = -T(S_{\text{stack}} - S_{\text{monomers}})$, kcal/mol
A	94.52	164.26	162.21	9.43	10.05
T	101.19				
G	101.83	155.62	155.18	10.65	10.78
C	89.29				
A	94.52		152.88		10.85
A					

$$\begin{tikzcd}
\mathrm{Ade}_{(g)} \arrow[d, "-12.61 \pm 1.66"'] & + & \mathrm{Ade}_{(g)} \arrow[r, "+1.29 \pm 0.12"] \arrow[d, "-12.61 \pm 1.66"'] & \mathrm{Ade/Ade}_{(g)}(\mathrm{st}) \arrow[d, "-28.66 \pm 1.45"] \\
\mathrm{Ade}_{(aq)} & + & \mathrm{Ade}_{(aq)} \arrow[r, "-2.16"'] & \mathrm{Ade/Ade}_{(aq)}(\mathrm{st})
\end{tikzcd}$$

$$\Delta\Delta G_{\text{experimental}} = -1.80$$

Figure 3. Thermodynamic cycle for adenine–adenine complex in stacked configuration.

possible exception of those for the H-bonded G:C complex, where the electrostatic term is a very large negative quantity and the van der Waals a very large positive. The cause of this is likely that when the charges and van der Waals parameters for the cytosine disappear simultaneously, the van der Waals parameters have become suitably small, such that the H-bonding distances become very short, allowing very favorable $H^{\delta+}\cdots X^{\delta-}$ electrostatic interactions, compensated by van der Waals repulsions from these short distances. The H-bonding in A:T is sufficiently weaker that such penetration effects do not occur there. In any case, the G:C results suggest that the approach used here may not give an unequivocal separation of electrostatic and van der Waals energies.

To properly calculate the free energy of association, we must include two other terms. The first one is the correction due to changes in vibrational entropies during complex formation at T = 300 K, which is not taken into account during molecular dynamics simulation, and the second one is to convert the gas-phase free energies to a 1 M standard state. The first correction was estimated by performing normal mode analysis for isolated bases and their H-bonded and stacked complexes. The results from this analysis are presented in the Table II. The second correction is simply $RT \ln(1/22.4)$, since the translational entropy from the normal-mode program[33] uses atmosphere units.[37]

Our results (bottom of Table I) show that H-bonded complexes are preferred in the gas phase, which is consistent with experimental observations[7-11] in nonpolar solvents. There is no direct experimental data on the structure of base pairs in the gas phase,[14] but it is reasonable to assume that the structure will be similar to that found in nonpolar solvents. The G–C H-bonded complex is held together most strongly mainly due to more favorable electrostatic interaction. The in vacuo stability of the different complexes decreases in the following order: G:C (H-bonded), G/C (stacked), A/T (H-bonded), A/T (stacked), and A/A (stacked). The important observation is that H-bonded complexes are mostly stabilized by electrostatic contribution to the free energies, whereas the stacked ones are stabilized mainly by the van der Waals type interactions. These results are consistent with other model calculations.[21,22]

In Table III we present the results for the solvation of the base pairs in water obtained from MD simulations. In contrast to the in vacuo situation, the stacked complexes are much better solvated by the water solution than H-bonded pairs. This could be rationalized by the fact that in the stacked complexes the solute atoms that can form hydrogen bonds are exposed to the water solution. This leads to additional stabilization, which is stronger than the hydrogen bonding of the two bases itself. In the case of the H-bonded pairs those atoms that are involved in creating hydrogen bonds with the other solute molecules are prevented from the H-bonding with other water molecules. However, a more complete analysis must consider both base–base and base–water interactions. We can separately add up the base–base and base–water electrostatic and van der Waals contributions from Tables I and III for stacked and H-bonded complexes. When one does this for A/T, one finds that the sum of the van der Waals (base–base and base–water) is −18.8 kcal/mol for the stacked complex and only −10.2 kcal/mol for the H-bonded complex, whereas the corresponding electrostatic energies are −10.9 kcal/mol for the stacked complexes and −17.1 kcal/mol for the H-bonded. The same analysis works for G:C, even though the large and compensating electrostatic + van der Waals energies in the gas-phase complexes make the absolute energies very different from those for the A–T. Thus, it appears that the difference in the net van der Waals energy is the *key* component in favoring the net greater stabilization of the stacked over H-bonded complexes in aqueous solution. It is difficult to separate uniquely this term into dispersion attraction and cavity (exchange repulsion) terms, but it seems intuitively reasonable that the stacked complexes have more favorable dispersion attraction (base–base + base–water) than the H-bonded complexes, and because of its smaller cavity volume, a less unfavorable cavity term.[31]

The following is the order of stabilization of stacking complexes in water solution obtained from our simulations G/C > A/A > A/T. The largest van der Waals interaction energy contribution to ΔG_{sol} is for the A/A stacked pair, whereas the electrostatic contribution is largest for G/C. Both contributions to the free energy of solvation have a similar magnitude for both A/T pairs and for the A/A pair. We expect the results for all base pairs to be similar to those for AT base pairs, since the methyl group solvation will likely be not very different for isolated bases and base pairs. In the gas phase, the A–U structure will favor H-bonded over stacked structures by, if anything, more than for the A–T pair.

In Table IV the association free energies for base pairs obtained by the free energy perturbation/molecular dynamics simulations are collected and compared with the available experimental data.[1] The $\Delta G_{\text{assoc(aq)}}$ was calculated from the thermodynamic cycles by subtracting $\Delta G_{\text{solv(A)}}$ and $\Delta G_{\text{solv(B)}}$ (taken from ref 31) from the sum of the $\Delta G_{\text{assoc(g)}}$ and $\Delta G_{\text{solv(A-B)}}$. For example, the solvation free energy of adenine comes from adding the free energy difference between adenine and methane, −14.73 kcal/mol (Table 2 in ref 31), and between methane and nothing, +2.12 kcal/mol (Table 1 in ref 31) for a net solvation energy for adenine of −12.61 kcal/mol. The solvation free energy for the other bases are calculated in a similar fashion. We note that the value of the $\Delta(\Delta G)$ for cytosine and thymine and adenine and guanine should be −5.24 and −6.95 kcal/mol, with the signs of these values incorrect in Table 2 of ref 31. The "filled out" thermodynamic cycles for different pairs are also presented on Figures 1–3. For the $\Delta G_{\text{assoc(aq)}}$ no error bars are given in Table IV, but we expect these to be of the order of 2 kcal/mol. We should keep in mind that these errors are caused mainly by the limited time evolution (trajectory) of our system during the simulations since they were done over a 42-ps time span. On the other hand the final results for the $\Delta G_{\text{assoc(aq)}}$ were based on adding and subtracting rather large numbers containing their own errors. Despite these large errors, our simulations give reasonable values for the absolute ΔG

---

[37] Ben-Naim, A.; Marcus, Y. *J. Chem. Phys.* **1984**, *81*, 2016.

Table III. ΔG (in kcal/mol) for the Solution of the Base Pairs in Water

		ΔG_{sol}									
		H-bonded A:T		stacked A/T		H-bonded G:C		stacked G/C		stacked A/A	
no. of water molecules in MD simulations		868		875		877		867		873	
	process	gas → water	water → gas	gas → water	water → gas	gas → water	water → gas	gas → water	water → gas	gas → water	water → gas
run I	electrostatic part	-9.18	9.32	-10.37	10.63	-13.96	13.45	-19.99	20.96	-14.10	14.45
	nonbonded part	-8.47	8.48	-11.42	11.44	-11.84	11.86	-10.53	10.54	-15.62	15.64
	total	-17.65	17.80	-21.80	22.07	-25.71	25.31	-30.52	31.49	-29.72	30.09
run II	electrostatic part	-9.00	9.13	-11.96	11.84	-13.44	13.87	-19.68	19.71	-13.07	13.60
	nonbonded part	-10.03	10.04	-10.83	10.85	-13.66	13.69	-13.08	13.10	-14.07	14.98
	total	-19.03	19.17	-22.79	22.69	-27.10	27.56	-32.76	32.81	-27.14	27.69
mean $\Delta G_{gas \to water}$		-18.41 ± 0.80		-22.34 ± 0.48		-26.42 ± 1.08		-31.90 ± 1.10		-28.66 ± 1.47	

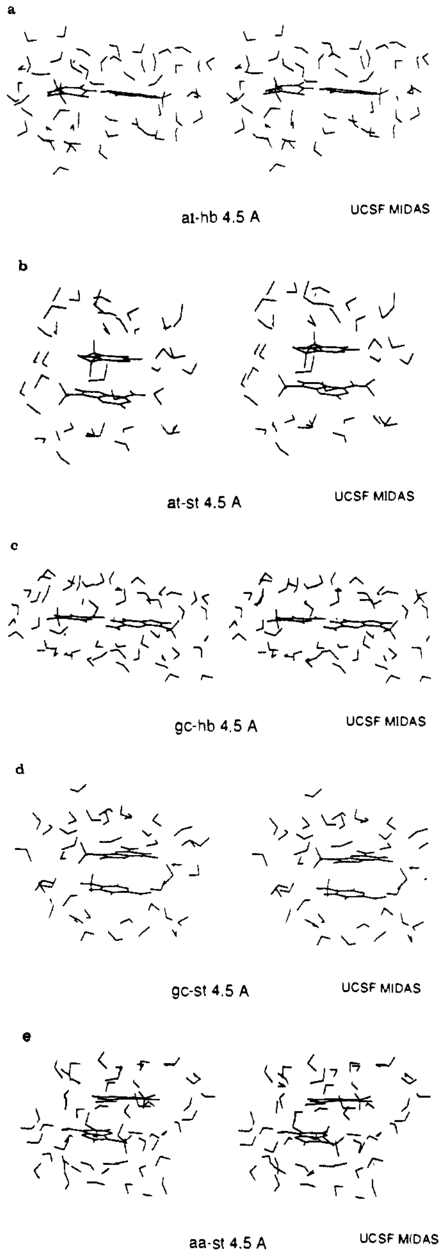


Figure 4. Representative configurations of base complexes in water before perturbation showing water molecules within 4.5 Å of any atom of the base pair: (a) A/T H bonded, (b) A/T stacked, (c) G/C H bonded, (d) G/C stacked, (e) A/A stacked.

of solution association, which are in the range of the experimental values. The resulting free energies for the association of the base pairs in water solution show that the free energy perturbation/molecular dynamics method is able to properly reproduce the fact that H-bonded configurations are preferred in the gas phase for adenine–thymine and guanine–cytosine pairs but that, in aqueous solution, stacked configurations are preferred.

It is interesting that our simulations can not only reproduce association free energies in the right range of the experimental values but also that some quantitative reproduction is achieved. The association free energy for A/A stacked in water solution is greater than for A/T stacked and, in addition, ΔG_{assoc} for A/T stacked is greater than for A:T H-bonded. For the G–C complexes in solution, the stacked complex is calculated to be more stable than the H-bonded.

The one disagreement with experiment is that the calculated free energy of G/C stacking in solution is more negative than those of A/T and A/A, whereas the reverse is observed experimentally. However, given the large error bars in each of these simulated numbers, the level of agreement with experiment found here is quite encouraging.

Table IV. Association Free Energy (kcal/mol) for Nucleic Acid Base Pairs in H-Bonded and Stacking Configurations in the Gas Phase and Water Solution in T = 300 K (Experimental Data Taken from Ref 1)

	A–T H-bond	A/T stack	G–C H-bond	G/C stack	A/A stack
calcd $\Delta G_{\mathrm{assoc(g)}}$	−1.51 ± 0.23	0.38 ± 0.15	−6.78 ± 0.16	−2.61 ± 0.05	1.29 ± 0.12
calcd $\Delta G_{\mathrm{assoc(aq)}}$	0.18	−1.86	−0.91	−2.22	−2.16
exptl $\Delta G_{\mathrm{assoc(aq)}}$		−1.2		−0.7	−1.8

In Figure 4 we present representative configurations from the molecular dynamics trajectories for the bases after they were equilibrated and before they were "perturbed". As one can see, there is reasonable H-bonding observed between many of the base sites and water as well as between water. However, as expected in such simulations, there are also examples of nonideal H-bonding due to compromises between the opposing forces of solvent–solvent and solute–solvent H-bonding.

Using the results of ref 31, one can see that of the total free energy of solvation of isolated adenine and thymine bases, of ∼−20 kcal/mol, about −17 kcal/mol comes from electrostatic/H-bonding and −3 kcal/mol from van der Waals terms. The van der Waals terms can sometimes be separated into the cavity term, which makes the solvation free energy more positive, and the dispersion term, which makes it more negative.[31] When one forms the AT base pairs (either stacked or H-bonded), the electrostatic free energy is far less negative than −17 kcal/mol but the nonbonded term far more negative than −3 kcal/mol. Thus, it is clear that once the price for creating the cavity in water had been "paid" by one base, not much must be paid for placing the second base on top or H-bonding to the first base, and the dispersion part of the van der Waals terms adds considerable net stabilization to base association in water.

There are a number of directions for further refinement of our results. First, we have begun each of these base associates in standard B-DNA geometries and have only sampled configurational space near these minima. This is a greater approximation for, stacked complexes, where the lowest energy configurations for e.g. A/A would likely involve antiparallel rather than nearly parrallel alignment of adenine dipole moments. This probably contributes to the counterintuitive result that A/T stacking is more favorable than A/A stacking in the gas phase. In any case, one needs to more extensively sample the possible configurations of stacked base complexes than has been done here.

Second, if our focus is an accurate calculation of the association free energy in solution only, rather than the analyses of the complete free energy cycles (Figures 1–3), one can use umbrella sampling or other potentials of mean force approaches, such have been applied to simple ion-pair or hydrophobic association in solution in a number of studies.[38] Such direct approaches would, of course, involve much smaller error bars than found here for the solution association process, but it still would be computationally demanding for adequate sampling of different possible base–base alignments to be achieved during the base–base associations.

There is uncertainty in our calculations in our use of normal mode analysis to calculate the translation/low-frequency vibration entropy contribution to base association in the gas phase. We assume that such contributions to the base association are not included when we "disappear" the van der Waals and electrostatic base–base energies. This is certainly true for translational/rotational entropy terms, but what is not certain is how much the free energy contribution of low-frequency vibrations are "counted twice" in this approach.

In view of the approximation used and the limited sampling done in these calculations, it is encouraging that the agreement between the absolute calculated and experimental $\Delta G_{\mathrm{assoc}}$ is so reasonable. We stress that, given the large error bars noted in the figures and the limited sampling in our calculations that one could view the good agreement as fortuitous. Given that we have achieved this agreement for three different complexes, we do not see it as such. We see this agreement with experiment as partially validating the potential functions and the approach chosen here. The single most severe approximation in the calculations reported here employing the perturbation method is the limited sampling of conformational space. When two independent simulations were run, we made a limited study of the dependence of the free energies upon starting geometry. We expect that the calculated free energies for the H-bond geometry is a sharp local minimum in the energy surface. On the other hand, the stacking geometries are characterized by a rather flat potential surface for rotation of one base with respect to the other. Thus, our sampling may be far from complete for this geometry but, because of the similarity of the energies, may be reasonably representative of the low energies of the system. At least the reasonable agreement with the experimental free energies of association suggests is not too unrepresentative.

Conclusions

The free energy perturbation/molecular dynamics method is able to calculate free energies for the association of bases in vacuo and in solution with surprising accuracy, given the comparison with the experimental free energies of A/T, G/C, and A/A association in solution. The calculations reproduce the observed preference of these associates to be stacked rather than H-bonded in water. Furthermore, the absolute calculated free energies of aqueous association are also in reasonable agreement with experiment. There are considerable sampling problems that need to be addressed and further simulations that need to be done to reduce the error bars in the calculation.

Nonetheless, we stress that these calculated free energies were determined ab initio, i.e. with no adjustable parameters in the simulations. The level of agreement is approaching chemical accuracy (<1 kcal/mol with almost all of the trends reproduced) and suggests that the molecular mechanical parameters used here (combining the base parameters of Weiner et al.[36] with the TIP3P water model[33]) is a reasonable and well-balanced set.

Acknowledgment. We are glad to acknowledge research support of the NIH (CA-25644). P.C. thanks the Polish Academy of Science for partial support for this project through Grant CPBP 01.12. Most of these calculations were carried out at the San Diego Supercomputer Center through supercomputer support provided to P.A.K. by the NSF (DMB-84-19883). We also gratefully acknowledge the use of an FPS-264 array processor purchased through grants from the NIH (RR-02441) and NSF (DMB-84-13762) in some of this work, as well as the facilities of the UCSF Computer Graphics Laboratory (supported by RR-1081 to R. Langridge). We thank Allison Howard for drawing Figures 1–3.

Registry No. A, 73-24-5; T, 65-71-4; G, 73-40-5; C, 71-30-7.

---

(38) Jorgensen, W. L.; Buckner, J. K.; Huston, S.; Rossky, P. *J. Am. Chem. Soc.* **1987**, *109*, 1891.