within the adsorbate layer. Experiments are underway to evaluate further the capability to image these and various other hydrocarbon-based monolayers.

Acknowledgment. M.D.P. gratefully acknowledges the support of a Dow Corning Assistant Professorship. Acknowledgment is also made to the donors of the Petroleum Research Fund, administered by the American Chemical Society, for support of this work. Ames Laboratory is operated for the U.S. Department of Energy by Iowa State University under Contract No. W-7405-eng-82. This work was supported by the Office of Basic Energy Sciences, Chemical Science Division.

OPLS Potential Functions for Nucleotide Bases. Relative Association Constants of Hydrogen-Bonded Base Pairs in Chloroform

Julianto Pranata, Scott G. Wierschke, and William L. Jorgensen*

Contribution from the Department of Chemistry, Purdue University, West Lafayette, Indiana 47907. *Received March 5, 1990*

Abstract: Potential functions in the OPLS format have been developed for the nucleotide bases and 2,6-diaminopyridine by fitting to the results of ab initio 6-31G(d) calculations for numerous base–water complexes. These potential functions yield dipole moments and base pair interaction energies in good agreement with available experimental data. The potential functions were tested further in Monte Carlo simulations with statistical perturbation theory to calculate the relative free energies of binding in chloroform for 9-methylguanine with 1-methylcytosine (G–C) versus 9-methyladenine with 1-methyluracil (A–U), and for G–C versus 1-methyluracil with 2,6-diaminopyridine (U–DAP). The calculations predict the G–C complex to be more stable than both the A–U and U–DAP complexes by about 5 kcal/mol. The similar stabilities for complexes like A–U and U–DAP are observed experimentally, though the quantitative enhancement in going to G–C appears to be exaggerated in the simulations. The large difference in association constants between G–C and the similarly triply hydrogen-bonded U–DAP is traced to the gas-phase interaction energies, which favors G–C by about 10 kcal/mol. This in turn is caused by the different arrangement of hydrogen bond donor and acceptor sites in the two complexes, which leads to secondary electrostatic interactions that disfavor U–DAP relative to G–C. The general importance of such secondary interactions for understanding variations in association is discussed.

Introduction

The interaction between nucleotide bases is an important element in the structure of DNA. Consequently, there have been numerous studies, experimental[1-8] and computational,[9-16] concerned with the association of nucleotide base pairs. The computational studies range from gas-phase energy minimizations[9-12] to Monte Carlo and molecular dynamics simulations in solution.[13-16] The work of Pohorille, Kollman, and co-workers is particularly notable.[15,16] Pohorille et al. performed seminal Monte Carlo simulations of stacked and hydrogen-bonded base pairs in CCl_4 and in water.[15] However, only the interaction energies (ΔE's) were calculated, while the more relevant measure of association is the free energy, ΔG. The bases were not allowed to move relative to one another, and the interaction energies were calculated by computing the differences in the total energies for the complex in solution and for the individual bases. This involves computing a small difference between large fluctuating numbers which leads to difficulties with precision. Subsequently, Cieplak and Kollman carried out molecular dynamics calculations for the A–T and G–C base pairs in vacuo and in water.[16a] Free energy changes were now calculated using statistical perturbation theory. The calculations featured arduous series of simulations in which each base and the complexes were made to vanish in water. It was correctly predicted that stacked structures were more stable than hydrogen-bonded ones in water, though the nature of the perturbations and simulation times led to significant uncertainties for the free energy changes.

Recently, Dang and Kollman calculated the free energy of association of 9-methyladenine and 1-methylthymine in water using a different approach.[16b] In this case the potential of mean force (PMF) of the base pair was obtained from a series of simulations in which the bases were gradually perturbed apart. Calculation of the association constant, K_a, then involves an integration of the PMF to a cutoff value that defines association.[17] The relative orientation of the bases was forced to remain constant as they were separated. Though the preference for stacking

* Address correspondence to this author at the Department of Chemistry, Yale University, New Haven, CT 06511.

(1) Kyogoku, Y.; Lord, R. C.; Rich, A. *Proc. Natl. Acad. Sci. U.S.A.* **1967**, *57*, 250.
(2) Kyogoku, Y.; Lord, R. C.; Rich, A. *Biochim. Biophys. Acta* **1969**, 179, 10.
(3) Iwahashi, H.; Kyogoku, Y. *J. Am. Chem. Soc.* **1977**, *99*, 7761. Watanabe, M.; Sugeta, H.; Iwahashi, H.; Kyogoku, Y.; Kainosho, M. *Eur. J. Biochem.* **1981**, *117*, 553. Iwahashi, H.; Sugeta, H.; Kyogoku, Y.; *Biochemistry* **1982**, *21*, 631.
(4) Newmark, R. A.; Cantor, C. R. *J. Am. Chem. Soc.* **1968**, *90*, 5010.
(5) Petersen, S. B.; Led, J. J. *J. Am. Chem. Soc.* **1981**, *103*, 5308.
(6) Williams, L. D.; Chawla, B.; Shaw, B. R. *Biopolymers* **1987**, *26*, 591.
(7) Williams, N. G.; Williams, L. D.; Shaw, B. R. *J. Am. Chem. Soc.* **1989**, *111*, 7208.
(8) Yanson, I. K.; Teplitsky, A. B.; Sukhodub, L. F. *Biopolymers* **1979**, *18*, 1149.
(9) Pullman, B.; Claverie, P.; Caillet, J. *Proc. Natl. Acad. Sci. U.S.A.* **1966**, *55*, 904.
(10) (a) Kudritskaya, Z. G.; Danilov, V. I. *J. Theor. Biol.* **1976**, *59*, 303. (b) Langlet, J.; Claverie, P.; Caron, F.; Boeuve, J. C. *Int. J. Quantum Chem.* **1981**, *19*, 299.
(11) Hobza, P.; Sandorfy, C. *J. Am. Chem. Soc.* **1987**, *109*, 1302.
(12) Aida, M. *J. Comput. Chem.* **1988**, *9*, 362.
(13) Danilov, V. I.; Tolokh, I. S.; Poltev, V. I.; Malenkov, G. G. *FEBS Lett.* **1984**, *167*, 245. Danilov, V. I.; Tolokh, I. S.; Poltev, V. I. *FEBS Lett.* **1984**, *171*, 325. Danilov, V. I.; Tolokh, I. S. *FEBS Lett.* **1984**, *173*, 347.
(14) Egan, J. T.; Nir, S.; Rein, R.; MacElroy, R. *Int. J. Quantum Chem. Quantum Biol. Symp.* **1978**, *5*, 433.
(15) 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.
(16) (a) Cieplak, P.; Kollman, P. A. *J. Am. Chem. Soc.* **1988**, *110*, 3734. (b) Dang, L. X.; Kollman, P. A. *J. Am. Chem. Soc.* **1990**, *112*, 503.
(17) Jorgensen, W. L. *J. Am. Chem. Soc.* **1989**, *111*, 3990.

0002-7863/91/1513-2810$02.50/0

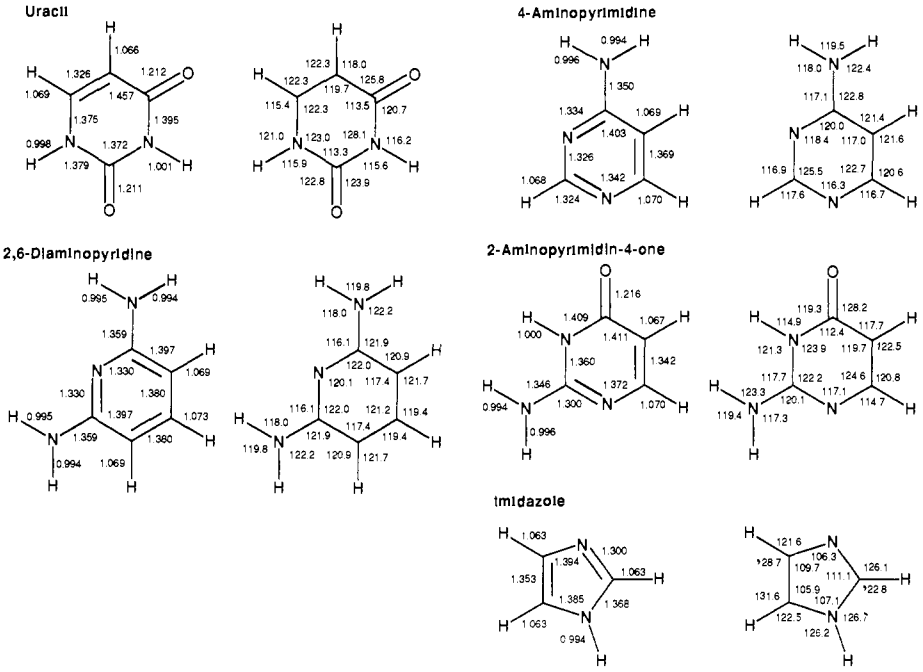


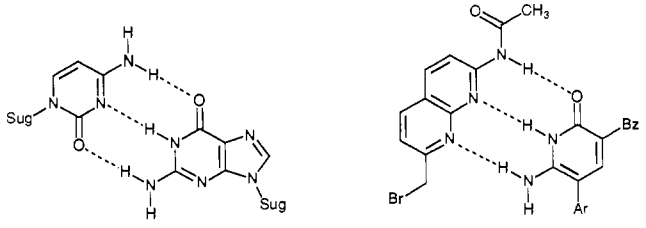
Figure 1. Optimized (3-21G) bond lengths and bond angles for uracil, 4-aminopyrimidine, 2-aminopyrimidin-4-one, imidazole, and 2,6-diaminopyridine. See ref 12 for the corresponding geometries of adenine, thymine, guanine, and cytosine.

Table I. OPLS Lennard-Jones Parameters for Nucleotide Bases and DAP

atom	σ, Å	ϵ, kcal/mol
O	2.96	0.210
N	3.25	0.170
C in C=O	3.75	0.105
other C	3.50	0.080
H on N	0.00	0.000
H on C	2.50	0.050

reappeared, the conclusion and the computed K_a's are not rigorously valid in the absence of complete orientational averaging which was not done.

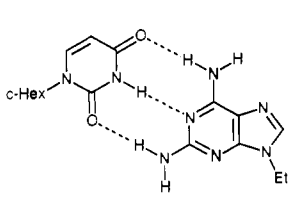
Our initial efforts at studying the association of nucleotide bases are summarized here. To begin, we have extended the OPLS (optimized potentials for liquid simulations) functions[18] to include parameters for nucleotide bases.[19] In order to test these parameters, Monte Carlo simulations were performed for the hydrogen-bonded base pairs of 9-methylguanine with 1-methylcytosine (G–C) and 9-methyladenine with 1-methyluracil (A–U) in chloroform. The availability of experimental association constants for these systems allows us to assess the reliability of our potential functions. In addition, we have noted an apparent anomaly in triply hydrogen-bonded systems.[19] Systems like G–C (**1**) and **2** have reported association constants of 10^4–10^5 M^{-1} in



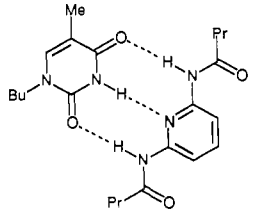
1

2

chloroform.[2,20] However, some triply hydrogen-bonded complexes of uracils and thymines exhibit much weaker association; e.g., the observed K_a's for **3** and **4** are only 170 and 90 M^{-1}.[1,21,22] These



3



4

values are more typical of *doubly* hydrogen-bonded systems; for example, a K_a of 100 ± 20 M^{-1} is found for 9-ethyladenine with 1-cyclohexyluracil. To address this issue, Monte Carlo simulations were performed to calculate the relative K_a's between G–C and 1-methyluracil with 2,6-diaminopyridine (U–DAP), a triply hydrogen-bonded system with a bonding pattern similar to **3** and **4**. Reproduction of the observed binding preference for G–C provided a basis for in-depth analysis of the origin of the differences for triply hydrogen-bonded complexes.

Parameter Development

Geometries. For the nucleotide bases, geometries optimized by ab initio calculations using the 3-21G basis set were adopted. These have been reported for adenine, thymine, guanine, and cytosine.[12] Geometry optimizations were performed in this work with the 3-21G basis set for uracil and DAP, and for the monocyclic fragments of adenine and guanine that are used as described below.[23] The resultant structures are shown in Figure 1. The 3-21G optimized bond lengths and angles typically differ from X-ray crystal structures by about 0.02 Å and 2°.[24] In all subsequent calculations, these geometries were kept constant.

(18) (a) Jorgensen, W. L.; Tirado Rives, J. *J. Am. Chem. Soc.* **1988**, *110*, 1657. (b) Jorgensen, W. L.; Briggs, J. M.; Contreras, M. L. *J. Phys. Chem.* **1990**, *94*, 1683.

(19) A preliminary account of this work is in: Jorgensen, W. L.; Pranata, J. *J. Am. Chem. Soc.* **1990**, *112*, 2008.

(20) Kelly, T. R.; Zhao, C.; Bridges, G. J. *J. Am. Chem. Soc.* **1989**, *111*, 3744.

(21) Hamilton, A. D.; Van Engen, D. J. *J. Am. Chem. Soc.* **1987**, *109*, 5035.

(22) Additional examples of imides and 1,6-diaminopyridines binding with K_a's of 125–150 M^{-1} are available: Kelly, T. R.; Maguire, M. P. *J. Am. Chem. Soc.* **1987**, *109*, 6548. Askew, B. C., Jr. Ph.D. Thesis, University of Pittsburgh, 1988, p 60.

(23) All ab initio calculations were performed using the GAUSSIAN 86 program: Frisch, M. J.; Binkley, J. S.; Schlegel, H. B.; Raghavachari, K.; Martin, R. L.; Stewart, J. J. P.; Bobrowicz, F. W.; DeFrees, D. J.; Seeger, R.; Whiteside, R. A.; Fox, D. J.; Fluder, E. M.; Pople, J. A. GAUSSIAN 86, Release C; Carnegie-Mellon University: Pittsburgh, PA, 1984.

(24) Taylor, R.; Kennard, O. *J. Am. Chem. Soc.* **1982**, *104*, 3209.

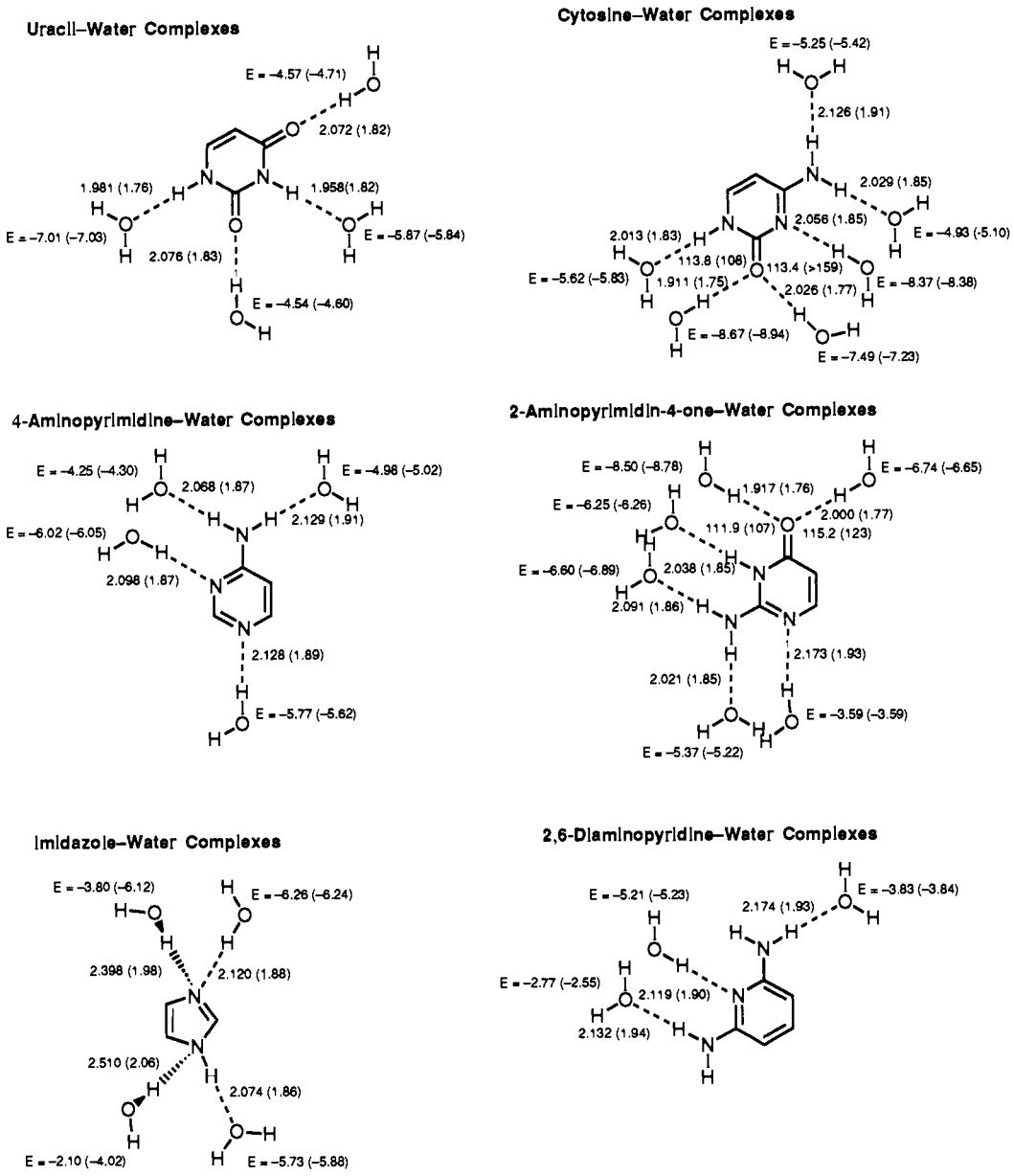


Figure 2. Base–water complexes used to derive the OPLS parameters. 6-31G(d) values for the interaction energies in kcal/mol, distances in Å, and angles in degrees are followed by the OPLS values in parentheses.

Intermolecular Potential Functions. In the OPLS model the intermolecular interaction energies are represented by Coulomb and Lennard-Jones terms between sites centered on nuclei (eq 1).[18]

$$\Delta E_{ab} = \sum_{i}^{\text{on a}} \sum_{j}^{\text{on b}} \left(\frac{q_i q_j e^2}{r_{ij}} + \frac{A_{ij}}{r_{ij}^{12}} - \frac{C_{ij}}{r_{ij}^{6}} \right) \qquad (1)$$

Thus, the interaction energy between molecules a and b is given by the sum of the interactions between the sites in the two molecules. For the nucleotide bases and other aromatic molecules, all-atom potential functions are preferred;[25] i.e., all hydrogens including those attached to carbon atoms are explicit. The combining rules used for the Lennard-Jones interactions are $A_{ij} = (A_{ii}A_{jj})^{1/2}$ and $C_{ij} = (C_{ii}C_{jj})^{1/2}$. In terms of the more familiar σ's and ϵ's, $A_{ii} = 4\epsilon_i\sigma_i^{12}$ and $C_{ii} = 4\epsilon_i\sigma_i^{6}$.

Previously OPLS parameters were obtained for various organic functional groups and neutral protein residues primarily via Monte Carlo simulations of many pure organic liquids.[18] The parameters were chosen to reproduce experimental thermodynamic properties of the liquids, particularly the densities and heats of vaporization. For the nucleotide bases, however, this approach is not feasible, since the bases are not liquids near room temperature. Therefore, it was decided to follow the procedure that was adopted for obtaining parameters for ionic systems.[18] In these cases the fitting was to results of ab intio calculations for ion–molecule complexes.

To derive parameters for uracil, cytosine, and DAP, ab initio calculations were performed on complexes of these bases with one water molecule. Several orientations of the water molecule around each base were considered; all involved some form of hydrogen bonding between the base and the water molecule (Figure 2). In most cases only one parameter, the distance between the hydrogen-bonded atoms, was optimized, although occasionally a key angle was also optimized. This partial optimization and calculation of the interaction energy were done using the 6-31G(d) basis set. This basis set is known to yield good descriptions of hydrogen-bonded complexes.[26] In view of the number of basis functions in these calculations, ca. 150, limitations on the number of optimized variables were necessary. Parameters for thymine were then based on those for uracil by adding a standard methyl group.

(25) Jorgensen, W. L.; Severance, D. L. *J. Am. Chem. Soc.* **1990**, *112*, 4768.

(26) Dill, J. D.; Allen, L. C.; Topp, W. C.; Pople, J. A. *J. Am. Chem. Soc.* **1975**, *97*, 7220. Jorgensen, W. L.; Gao, J. *J. Phys. Chem.* **1986**, *90*, 2174. Gao, J.; Garner, D. S.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1986**, *108*, 4784.

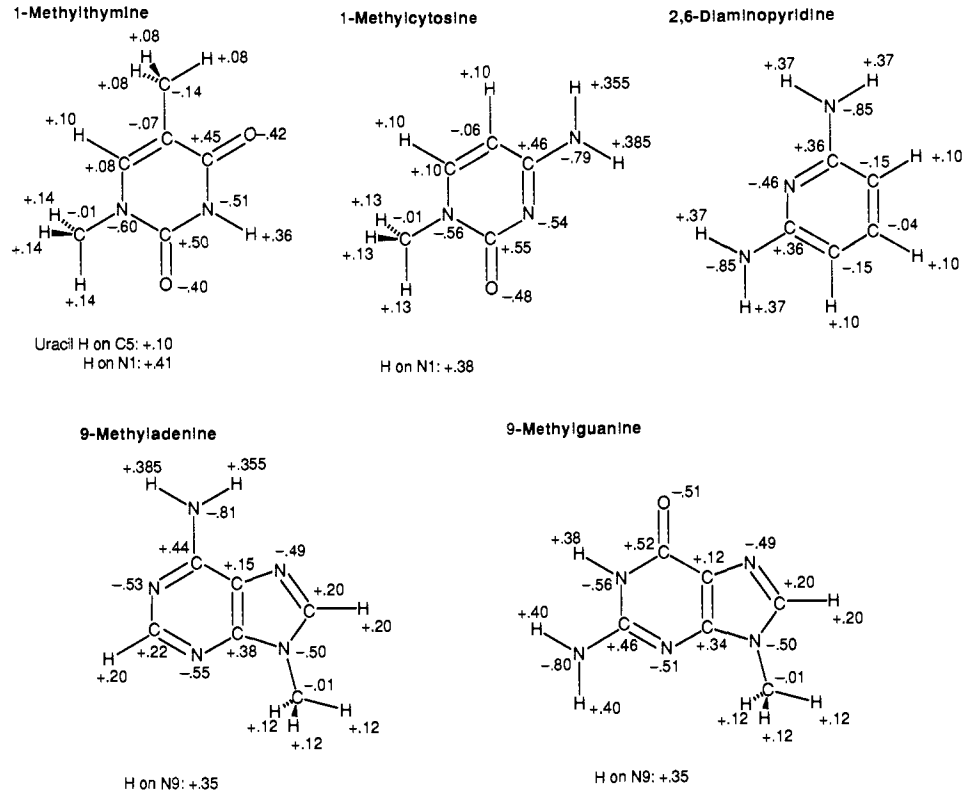


Figure 3. OPLS charge parameters for nucleotide bases and DAP.

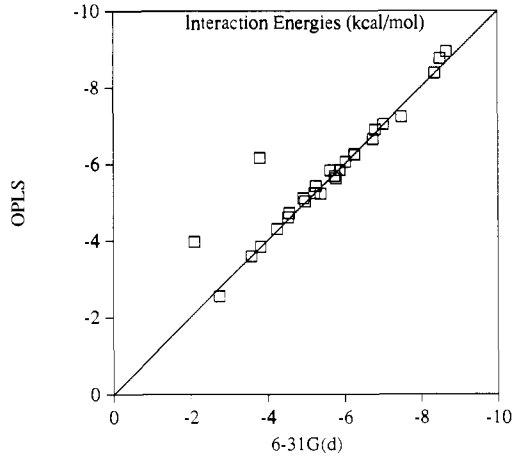


Figure 4. Comparison between ab initio 6-31G(d) and OPLS interaction energies.

For the purine bases, the ab initio calculations were performed on monocyclic fragments, i.e., 4-aminopyrimidine (adenine fragment), 2-aminopyrimidin-4-one (guanine fragment), and imidazole, and their complexes with a water molecule (Figure 2). As before, several hydrogen-bonded orientations were considered, and partial optimization of the intermolecular geometry was performed.

The OPLS parameters were chosen to reproduce the ab initio results for the geometries and interaction energies. The TIP4P model was used for the water molecule.[27] Only the charge parameters (q's) for the bases were adjusted. Standard values were adopted for the Lennard-Jones parameters, as listed in Table I.[18,25] The final charge parameters that were obtained are shown in Figure 3.

The accord for the interaction energies between the OPLS and ab initio results is summarized in Figure 4. Overall, the root-mean-square (rms) deviation is 0.58 kcal/mol. The only significant discrepancies are for the two out-of-plane complexes of imidazole; without them, the rms deviation would be 0.13 kcal/mol. With the simple point charge model used, it was not possible to reproduce simultaneously the very different interaction energies for the in-plane ($\sim$ −6 kcal/mol) and out-of-plane ($\sim$ −3 kcal/mol) complexes (Figure 2). The lower energy structures were emphasized in the fitting since they are expected to be more populated.

Table II. Dipole Moments (D) for Nucleotide Bases

base	OPLS	exp[a]	Singh and Kollman[b] STO-3G	Singh and Kollman[b] Clementi
uracil	3.71	3.86, 3.9	3.37	3.72
thymine	4.14	3.58	3.20	3.54
cytosine	7.20	7.10	5.67	6.14
adenine	2.54	3.16, 3.0	2.17	2.31
guanine	6.44	6.76	6.14	6.21

[a] Reference 29. [b] Reference 30.

The interaction distances from the OPLS calculations are uniformly 0.2–0.3 Å shorter than the ab initio results (Figure 2). This feature is necessary to obtain correct liquid densities from the potential functions, as discussed previously.[18,28]

Another requirement of the OPLS charge distribution is that they give reasonable dipole moments for the nucleotide bases. The dipole moments calculated using the OPLS charge distributions are shown in Table II, along with experimental values[29] and other theoretical results.[30] The latter are from Singh and Kollman and were obtained by fitting to ab initio electrostatic potentials calculated using the STO-3G basis set and the minimal basis set of Clementi.[31] In view of the fact that the dipole moments were not used in the parameter fitting, the agreement between the OPLS and experimental values is surprisingly good. Perfect

(27) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926. Jorgensen, W. L. Madura, J. D. *Mol. Phys.* **1985**, *56*, 1381.

(28) Jorgensen, W. L.; Swenson, C. J. *J. Am. Chem. Soc.* **1985**, *107*, 569, 1489.

(29) Pullman, B. In *Molecular Association in Biology*; Pullman, B., Ed.; Academic: New York, 1968; p 1.

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

(31) Clementi, E.; Andre, J. M.; Andre, M. C.; Klint, D.; Hahn, D. *Acta Phys. Chem.* **1969**, *27*, 493.

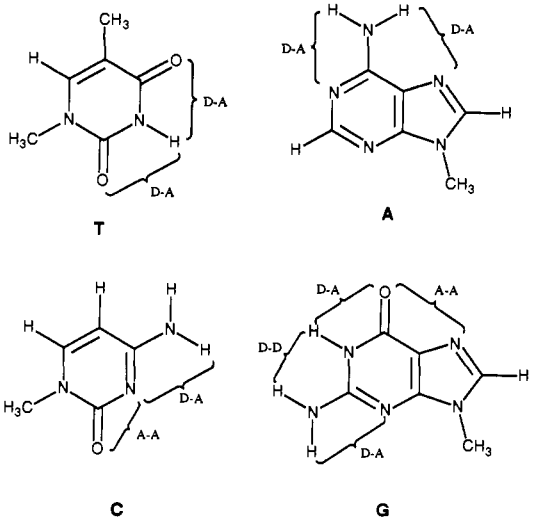


Figure 5. Pairs of hydrogen-bonded donor and acceptor sites on the nucleotide bases.

agreement is not sought, though serious discrepancies would be cause for concern.

Base Pair Interaction Energies

As an initial test of the new OPLS parameters, gas-phase interaction energies were calculated for numerous hydrogen-bonded base pairs between 9-methyladenine, 9-methylguanine, 1-methylthymine, and 1-methylcytosine. *All* possible multiply hydrogen-bonded orientations were considered. These may be identified by considering all the hydrogen-bond donor and acceptor sites that are 1,3 or 1,4 to each other (Figure 5). A D–A site can form two hydrogen bonds with another D–A site. Since there are seven D–A sites in the four bases, there is a total of 28 ($7 \times 8/2$) base pair orientations with this pattern. A D–D site can form two hydrogen bonds with an A–A site, and it can do so in two orientations. There are two D–D sites and only one A–A site, giving a total of 4 (2×2) base pair orientation with this pattern. However, the triply hydrogen-bonded G–C base pair uses both a D–A:D–A and a D–D:A–A pattern. Thus, there is a net of 31 possible multiply hydrogen-bonded orientations for the base pairs.

Intermolecular geometries were fully optimized, although in most cases the base pairs remained coplanar or nearly so. Some of these complexes are shown in Figure 6; the complete set can be found in the supplementary material. The 31 OPLS interaction energies are given in Table III, along with prior theoretical results for some cases. Kudritskaya and Danilov's values were calculated with an atomic dipole approximation using CNDO–Cl wave functions.[10a] Langlet et al.'s results came from a multipole expansion for the electrostatic energy up to the quadrupole-quadrupole term with charge distributions from ab initio SCF calculations.[10b] The experimental ΔH's were obtained using mass spectrometric methods; the actual orientations of the bases in the complexes were not determined.[8] The accord between the OPLS interaction energies and Langlet's values is better than that for the results of Kudritskaya and Danilov, though there are still significant differences as in the case of adenine dimers. The interaction energies can also be compared to the results of Hobza and Sandorfy, who calculated the interaction energies of 28 of these base pairs using ab initio calculations with a minimal basis set, plus a correction for basis set superposition errors and an empirically calculated dispersion term.[11] Although their calculated interaction energies are consistently more negative than the OPLS values, the general trends are very similar. Particularly encouraging is the agreement between the OPLS predictions and the four experimental values. It can be seen that there is substantial variation in the 31 interaction energies. The strongest interactions are observed in base pairs involving G and C, which do have the largest dipole moments.

Kyogoku et al. have measured the association constants for the base pairs shown in Figure 6 in chloroform using infrared spectroscopy.[1,2] Their measured K_a's for G–C, G–G, A–T, C–C, T–T, and A–A are 10^4–10^5, 10^3–10^4, 130, 28 ± 3, 3.2, and 3.1 ± 0.3 M^{-1}, respectively. While it is not appropriate to make a direct comparison between our gas-phase interaction energies and their

Table III. Optimized Interaction Energies (kcal/mol) for Base Pairs in the Gas Phase

base pair	H-bonds[a]	OPLS[b]	Langlet et al.[c]	Kudritskaya and Danilov[d]	exp ΔH[e]
A–A	N6→N1′	−8.3	−14.0	−5.25	
	N6′→N1				
	N6→N7′	−6.0	−12.1	+0.35	
	N6′→N7				
	N6→N1′	−7.9	−12.9	−5.60	
	N6′→N7				
C–C	N4→N3′	−15.0	−20.0	−10.73	−16.0
	N4′→N3				
G–G	N1→O6′	−21.0	−20.8	−16.04	
	N1′→O6				
	N2→N3′	−10.8	−14.9	−6.36	
	N3′→N2				
	N2→O6′	−12.1			
	N1′→N3				
	N1→O6′	−16.2			
	N2→N7′				
	N1→N7′	−17.3			
	N2→O6′				
T–T	N3→O4′	−9.1	−9.5	−5.21	−9.0
	N3′→O4				
	N3→O2′	−9.0	−9.6	−3.73	
	N3′→O2				
	N3→O2′	−9.0			
	N3′→O4				
A–T	N6→O4′	−10.4[f]	−12.9	−7.00	−13.0
	N3′→N1				
	N6→O2′	−10.3	−12.4	−6.78	
	N3′→N1				
	N6→O4′	−10.6	−13.6	−6.44	
	N3′→N7				
	N6→O2′	−10.6	−13.3	−6.67	
	N3′→N7				
G–C	N1→N3′	−22.4[f]	−23.7	−16.79	−21.0
	N2→O2′				
	N4′→O6				
	N1→O2′	−16.0			
	N2→N3′				
	N2→N3′	−12.2			
	N4′→N3				
A–C	N6→N3′	−10.1		−7.32	
	N4′→N1				
	N6→N3′	−9.1		−8.30	
	N4′→N7				
G–T	N1→O2′	−11.5		−7.74	
	N3′→O6				
	N1→O4′	−12.4		−9.86	
	N3′→O6				
	N2→O2′	−11.6			
	N3′→N3				
	N2→O4′	−11.9			
	N3′→N3				
G–A	N1→N1′	−12.2		−9.40	
	N6′→O6				
	N2→N1′	−10.3			
	N6′→N3				
	N1→N7′	−11.5		−5.86	
	N6′→O6				
	N2→N7′	−9.9			
	N6′→N3				
C–T	N4→O4′	−9.3		−3.67	
	N3′→N3				
	N4→O2′	−8.8		−4.40	
	N3′→N3				

[a] Hydrogen-bond donor → acceptor; primed atom belongs to the second of the bases listed in the first column. [b] Methyl-substituted base pairs were used for the OPLS calculations. [c] Reference 10b. [d] Reference 10a. [e] Reference 8. [f] Watson–Crick orientations. Interaction energies for the parent (nonmethylated) base pairs are −10.6 (A–T) and −22.1 (G–C).

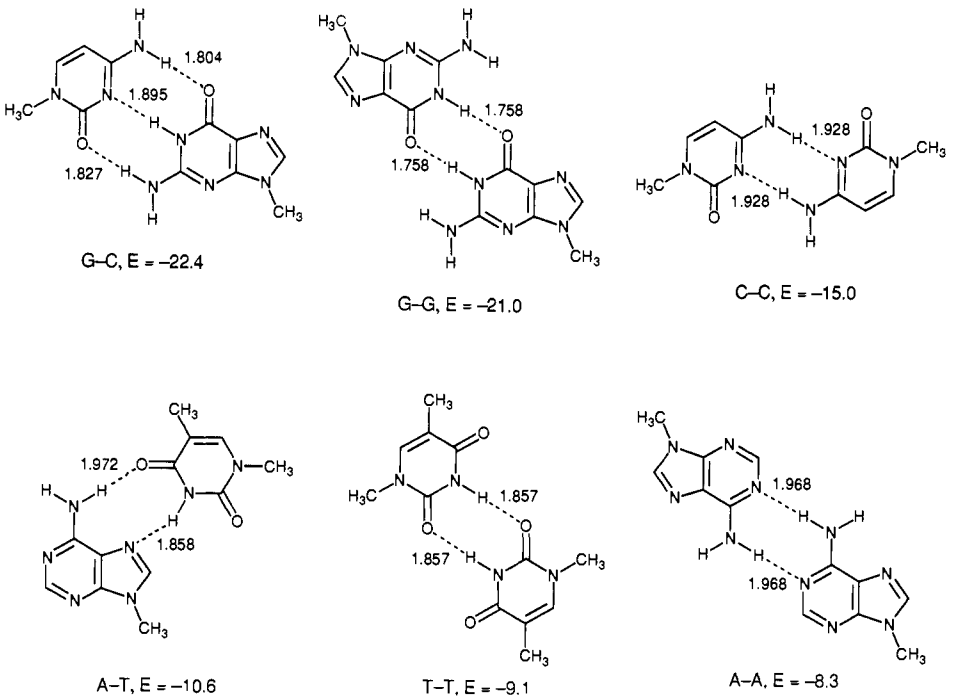
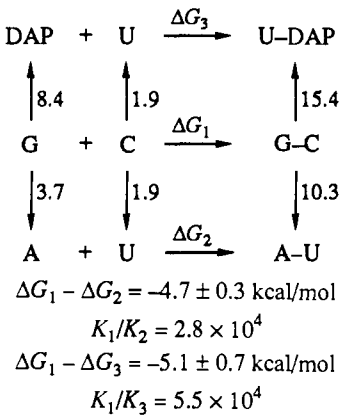


Figure 6. Most stable orientations of the base pairs G–C, G–G, C–C, A–T, T–T, and A–A. Note that the Hoogsteen orientation is the most stable for A–T; however, other orientations including the Watson–Crick have essentially the same interaction energies.

Scheme I



$$\Delta G_1 - \Delta G_2 = -4.7 \pm 0.3 \text{ kcal/mol}$$

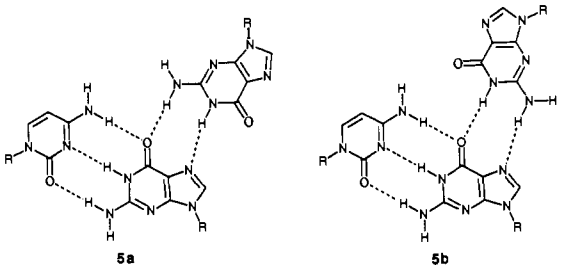
$$K_1/K_2 = 2.8 \times 10^4$$

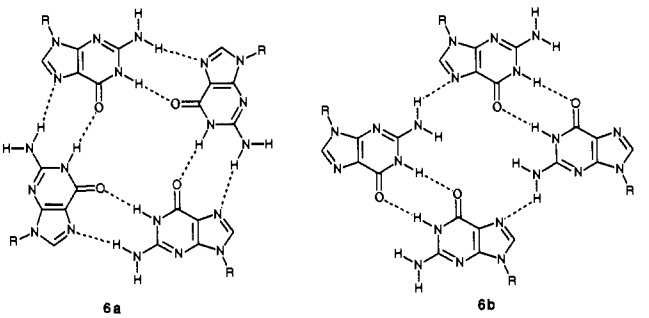
$$\Delta G_1 - \Delta G_3 = -5.1 \pm 0.7 \text{ kcal/mol}$$

$$K_1/K_3 = 5.5 \times 10^4$$

solution-phase association constants, the observed trends are quite similar. As illustrated, the most strongly bound orientations of these base pairs have predicted interaction energies of –22.4, –21.0, –10.6, –15.0, –9.1, and –8.3 kcal/mol. The only rearrangement is between A–T and C–C. However, there are four nearly isoenergetic orientations for the A–T base pair and only one for C–C. If it is assumed that all four orientations contribute equally, the association constant for each orientation is 32.5 M^{-1}, and the discrepancy is lessened.

Interestingly, Williams et al. have studied the formation of trimers and tetramers of G and C.[7] Of the two possible C–G–G trimers (**5a** and **5b**), **5a** was deemed to be the more stable on the basis of NMR spectra. This is consistent with the present results, since the G–G portion of this trimer has an interaction energy of –17.3 kcal/mol compared to –13.6 kcal/mol for **5b**. Nevertheless, the G–G orientation in **5b** is the same as that found in the guanine quartet, proposed to occur in telomeric DNA.[32] It is interesting to note that this is *not* the most stable orientation of the guanine dimer; in fact, it is the second *least* stable of the five possible orientations. However, it is perfectly suited to form the tetramer in **6a**. This tetramer has a total of eight hydrogen bonds. Another reasonable structure is **6b**, which is a dimer of the most stable guanine dimer. There are only six hydrogen bonds, and two of them are substantially nonlinear. However, there may also be partial hydrogen bonds between the amino groups and the carbonyl oxygens. Thus, this structure could be energetically competitive with **6a**.

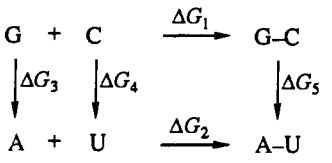




Monte Carlo Simulations

The relative association constant in chloroform between the base pairs G–C and A–U was calculated using Monte Carlo simulations.[33,34] As before, the methyl-substituted bases were used in these calculations. The thermodynamic cycle below was considered, and the relative free energy of association ($\Delta G_1 - \Delta G_2$) was obtained by computing $\Delta G_3 + \Delta G_4 - \Delta G_5$. Statistical perturbation theory was used to obtain the ΔG's.[35] Thus, G was gradually converted to A using a coupling parameter, λ, which

(32) Sen, D.; Gilbert, W. *Nature* **1988**, *334*, 364. Sundquist, W. I.; Klug, A. *Nature* **1989**, *342*, 825. Williamson, J. R.; Raghuraman, M. K.; Cech, T. R. *Cell* **1989**, *59*, 871.
(33) Tembe, B. L.; McCammon, J. A. *Comput. Chem.* **1984**, *8*, 281.
(34) For reviews, see: (a) Jorgensen, W. L. *Acc. Chem. Res.* **1989**, *22*, 184. (b) Beveridge, D. L.; DiCapua, F. M. *Annu. Rev. Biophys. Biophys. Chem.* **1989**, *18*, 431.
(35) Zwanzig, R. *J. Chem. Phys.* **1954**, *22*, 1420.



represents the linear admixture of G and A as λ goes from 0 (G) to 1 (A). Similarly, C was gradually converted to U and the complex G–C to A–U. The complexes started out in the Watson–Crick orientation, but the bases were allowed to move independently. Numerous simulations, each at different values of λ, were performed for each conversion. In each simulation λ was perturbed to both smaller and larger values; thus, each simulation yielded two incremental ΔG's. The incremental ΔG's are given by eq 2 which expresses the free energy change for perturbing

$$G_j - G_i = -k_B T \ln \langle \exp[-(H_j - H_i)/k_B T] \rangle_i \qquad (2)$$

λ_i to λ_j as a function of the energy difference between the systems at λ_i and λ_j, $H_j - H_i$, with the average obtained by configurational sampling for the system at λ_i.

The simulations were carried out at constant temperature (25 °C) and pressure (1 atm). In the G to A and C to U mutations, the bases were surrounded by 125 chloroform molecules in a cubic cell with dimensions ca. 26 × 26 × 26 Å^3, while in the G–C to A–U mutation the number of solvent molecules was 185 and the cell dimensions were ca. 26 × 26 × 39 Å^3. Periodic boundary conditions and the OPLS four-site model for chloroform[18b] were employed. Intermolecular interaction energies were calculated using a molecule-based cutoff distance of 12 Å; i.e., if any intermolecular atom–atom distance was less than 12 Å, the entire intermolecular interaction was included. The cutoffs were smoothed by quadratic feathering over the last 0.5 Å.

Each simulation consisted of 5 × 10^5 to 10^6 configurations for equilibration, followed by averaging for 2 × 10^6 configurations. The configurations were chosen using Metropolis and preferential sampling.[34,36]

For the G–C and A–U mutation, an additional set of simulations was performed. A simulation was carried out at λ = 0 (corresponding to G–C), in which the free complex was perturbed into a system in which weak harmonic constraints (k = 20 kcal/mol·Å^2, r_{eq} = 3 Å) were imposed between N1 of G and N3 of C and between O(C6) of G and N(C4) of C. These constraints were maintained at the intermediate points (0 < λ < 1), in order to enforce proximity of the bases. In the final simulation, performed at λ = 1, the constraints were removed to yield the free A–U complex.

An analogous set of Monte Carlo simulations was performed to calculate the relative association constant for G–C versus U–DAP. The computations were carried out with the BOSS program, version 2.8, on Silicon Graphics 4D and Sun workstations in our laboratory.

Results and Discussion

Figures 7–9 show representative configurations of the base pairs G–C, A–U, and U–DAP in chloroform. These configurations were generated without the harmonic constraints mentioned above, yet the bases remained hydrogen bonded to one another throughout the simulations. This is consistent with the well-known fact that nucleoside bases form hydrogen-bonded complexes in nonpolar solvents.[1-7]

Numerical results of the various series of simulations are shown in Scheme I below. Details on the incremental ΔG's are presented in the supplementary material. The error bars ($\pm 1\sigma$) were obtained by calculating separate averages over blocks of 2 × 10^5 configurations during each Monte Carlo simulation. In all except one of the mutations, four to six separate simulations were used to span the range of the coupling parameter λ from 0 to 1. The exception was the unconstrained G–C to A–U mutation, which was covered in 13 simulations. The smaller increments for λ were beneficial in keeping the bases from drifting irreparably apart in the intermediate stages of the mutation.

The 15.4-kcal/mol free energy difference between G–C and U–DAP includes a necessary correction, RT ln 2, that reflects the changes in symmetry numbers for the two complexation processes.[37] Another, perhaps more obvious, way of looking at this is to realize that there are two ways for DAP to bind to U using three hydrogen bonds, while for G–C there is only one way. Similarly, there are four ways of forming the A–U base pair, which have nearly identical interaction energies (Table II); however, only the Watson–Crick form was sampled. For G–C, the triply hydrogen-bonded Watson–Crick orientation is much more stable than any other orientations, so it is safe to assume that only this orientation is significant. Therefore, the correction for the free energy difference between G–C and A–U is approximately RT ln 4. This factor is included in the 10.3 kcal/mol listed in Scheme I.

The relative K_a of 2.8 × 10^4 between G–C and A–U is somewhat larger than the value suggested from experiment.[1,2] However, there is much uncertainty in the experimental K_a for G–C. For such strongly bound complexes, determination of the association constant requires very dilute solutions and hence fairly long IR cells. This leads to complications caused by heating of the solution. Additionally, water absorption interferes with the absorption bands of the uncomplexed bases. Hence, the experimental K_a of 10^4–10^5 M^{-1} for G–C is only a "rough estimate".[2] However, the calculated relative K_a between G–C and U–DAP (5.5 × 10^4) is essentially the same as that between G–C and A–U, which is consistent with the observation of similar K_a's (ca. 10^2) for systems like A–U or U–DAP.[1,2,21]

The values shown in Scheme I are those obtained from the unconstrained calculations. When the harmonic constraints were included in the simulations, the values obtained are $\Delta G_1 - \Delta G_2$ = −7.2 ± 0.3 kcal/mol (K_1/K_2 = 1.8 × 10^5) and $\Delta G_1 - \Delta G_3$ = −7.0 ± 0.4 kcal/mol (K_1/K_3 = 1.4 × 10^5). These K_a ratios are larger than the ones obtained from the unconstrained simulations. However, comparing the relative K_a between G–C and A–U to that of G–C and U–DAP leads to the same conclusion, similar K_a's for A–U and U–DAP. In the absence of sampling problems, the constrained and unconstrained result should be the same. It is not possible from the present calculations to determine which of the procedures leads to the "correct" prediction. However, calculations of the absolute association constants (see below) are consistent with the unconstrained results.[38]

That G–C is more strongly bound than A–U is in accord with expectations, since three hydrogen bonds should be better than two. That G–C is preferred over the triply hydrogen-bonded U–DAP to a similar extent is surprising, but this is fully consistent with the experimental data for **1–4**. The success of the simulation method in reproducing this unusual result provides confidence in the computations and underlying force field.

Another test and challenge would be to compute the absolute, rather than relative, association constants. There are currently two reasonable approaches for this. One could compute the PMF for the base pair as the bases are gradually separated.[17] The problem is that complete orientational averaging is required which may be difficult to achieve without greatly extending the simulations. Nevertheless, this route to computing the absolute K_a's was investigated for uracil–DAP and G–C in chloroform. The central hydrogen-bonded N–N distance was used as the reaction coordinate; otherwise, the individual molecules moved freely. The computations were well-behaved for uracil–DAP, and a smooth PMF was obtained as shown in Figure 10. The parent uracil (i.e., not 1-methyluracil) was used in this calculation. A single minimum is found at a uracil N3 – DAP N1 separation of 3 Å corresponding to the triply hydrogen-bonded complex. Integration to 5–6 Å over this free energy profile,[17] which has been zeroed at 7.1 Å, yields a K_a of 143 M^{-1}. This is in the range of the

(36) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *J. Chem. Phys.* **1953**, *21*, 1087. Owicki, J. C. *ACS Symp. Ser.* **1978**, *86*, 159.

(37) Bailey, W. F.; Monahan, A. S. *J. Chem. Educ.* **1978**, *55*, 489.

(38) Pranata, J.; Jorgensen, W. L. *Tetrahedron*, in press.

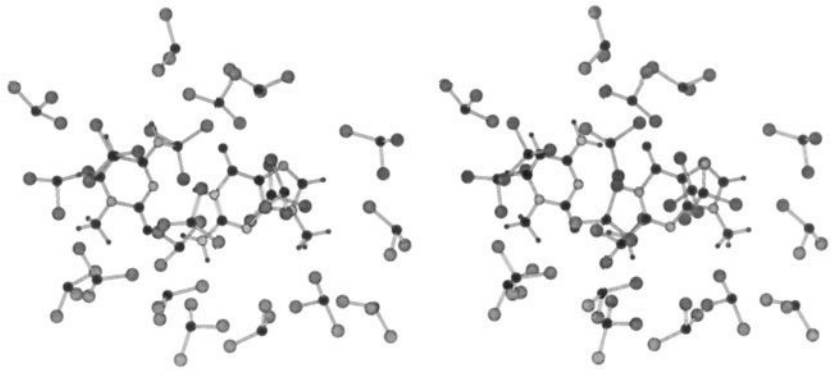


Figure 7. Stereoplot of a typical configuration of G–C obtained during the Monte Carlo simulations. Only chloroform molecules with an atom within 4 Å of any solute atom are shown in Figures 7–9. Note that there is an implicit hydrogen in the OPLS four-site model for chloroform.

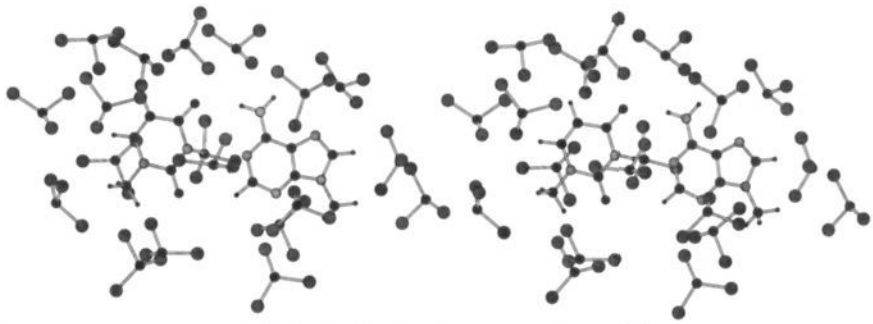


Figure 8. Stereoplot of a typical configuration of A–U obtained during the Monte Carlo simulations.

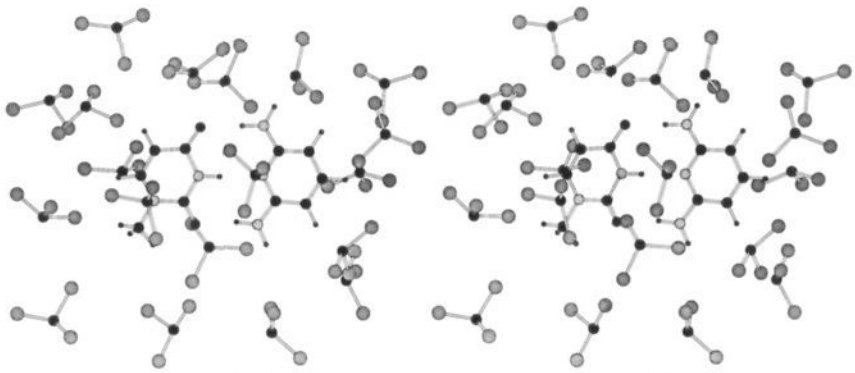
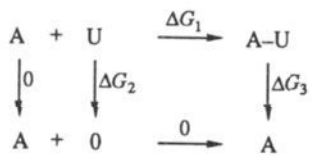


Figure 9. Stereoplot of a typical configuration of U–DAP obtained during the Monte Carlo simulations.

experimental results for **3** and **4**, 170 and 90 M^{-1}, respectively.[2,21] However, the PMF calculation for G–C was problematic because of the existence of varied hydrogen-bonded orientations for this base pair. At different separations, different forms of hydrogen-bonded complexes tend to predominate. The simulations are essentially trapped in these numerous local minima. It becomes impossible to sample enough of configuration space to obtain proper averaging and a bumpy PMF is obtained. Further study is required to develop fully this approach to computing K_a's.

The alternative approach to calculating absolute K_a's is illustrated in the thermodynamic cycle below for A–U.[39] Two sets



of simulations are necessary. In the first, U is made to disappear in the solvent. In the other, U is made to disappear while bound to A. The free energy of association (ΔG_1) is calculated as $\Delta G_2 - \Delta G_3$. These calculations have been performed for G–C and A–U, and the results are presented in full elsewhere.[38] The calculated K_a's with the present potential functions are ca. 5.7 $\times 10^5$ M^{-1} and 4.4 $\times 10^2$ M^{-1} for G–C and A–U in chloroform.

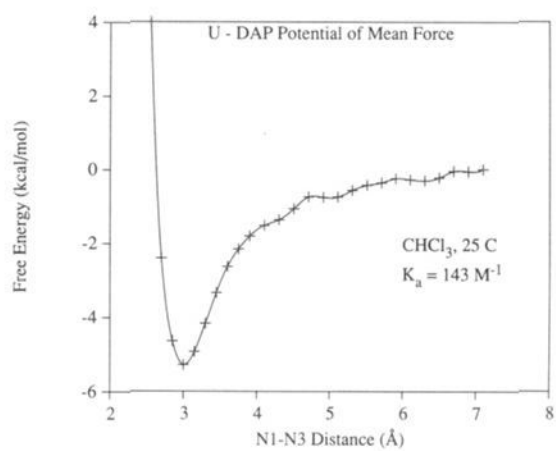


Figure 10. Potential of mean force computed for the U–DAP complex.

(39) Jorgensen, W. L.; Buckner, J. K.; Boudon, S.; Tirado-Rives, J. *J. Chem. Phys.* **1988**, *89*, 3742.

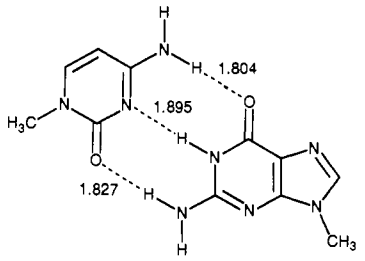


Figure 11. Optimized gas-phase structures of the G–C and U–DAP complexes from the OPLS potential functions.

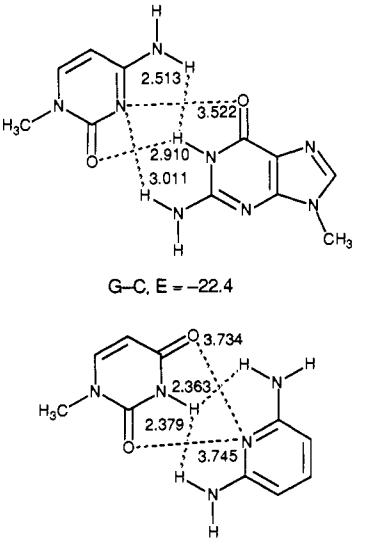


Figure 12. Secondary hydrogen-bonding interactions in the G–C and U–DAP complexes.

These values are both only a little higher than the experimental results. Also, their ratio of 10^3 is closer to the present unconstrained results of 10^4 than the constrained result of 10^5.

The stronger binding of G–C than U–DAP can be traced to the intersolute interaction energy which is 10.5 kcal/mol more attractive for G–C in chloroform. In the isolated complexes, the optimal interaction energies are –22.4 kcal/mol for G–C and –11.1 kcal/mol for U–DAP, as shown in Figure 11. However, the primary hydrogen-bonding interactions are the same, there are two $NH_2 \rightarrow O$ and one $NH \rightarrow N$ hydrogen bonds in both complexes. More subtle effects must be present to account for the 11.3-kcal/mol difference.

A simple interpretation is provided by considering the secondary interactions indicated in Figure 12. These separations are still short (2.3–3.7 Å) owing to the proximal hydrogen bonds and involve substantial electrostatic interactions. In view of the partial positive charge on H and partial negative charges on N and O, two of the secondary interactions are attractive and two are repulsive for G–C, while all four interactions are repulsive for U–DAP. If each primary hydrogen bond is assigned –7.5 kcal/mol and each secondary interaction contributes ±2.5 kcal/mol, the interaction energies are fit at –21.5 kcal/mol for G–C and –11.5 kcal/mol for U–DAP, consistent with the OPLS values.

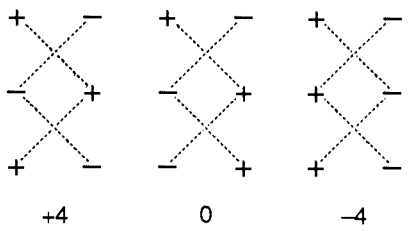
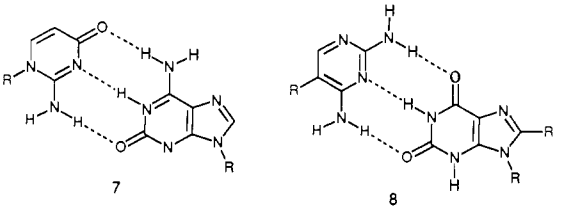


Figure 13. Possible patterns of secondary interactions in triply hydrogen-bonded complexes.

The three possible arrangements of the partially charged sites for triply hydrogen-bonded systems are shown in Figure 13. The worst situation is when the positive and negative sites alternate, as in DAP and imides, resulting in a total of four destabilizing secondary interactions. The intermediate case is represented by G–C, with two attractive and two repulsive interaction resulting in no net secondary effect. The most favorable situation has all the hydrogen-bond donor sites on one molecule and all the acceptor sites on the other, resulting in a total of four attractive secondary interactions. Naturally, the magnitudes of the partial charges as well as their arrangement are important to the overall strength of the interaction. The large dipole moments for G and C (Table 11) reflect the greater charge separation for these bases than for U and DAP (Figure 3). Stronger hydrogen bonding for G and C is also indicated by the results in Figure 2.

The consideration of secondary electrostatic interactions is relevant in many contexts. For example, Benner and co-workers have been "expanding the genetic alphabet" by incorporating the new base pairs isoguanosine–isocytidine (isoG–isoC, **7**) and xanthosine–2,4-diaminopyrimidine (X–κ, **8**) into duplex DNA and



RNA using DNA and RNA polymerases.[40] From the perspective of Figure 13, **7** has a hydrogen-bonding pattern identical with G–C, while **8** has a pattern identical with U–DAP. Neither of these base pairs, nor the others proposed by Benner and his co-workers,[40] have the pattern that could lead to the strongest hydrogen-bonded pair.

Consideration of secondary hydrogen-bonding interactions also helps explain the variability of interaction energies shown in Table III and Figure 6. With the exception of G–C, all the complexes in Figure 6 are doubly hydrogen-bonded. Yet, G–G and C–C have substantially stronger interaction energies than A–T, A–A, or T–T. The secondary interactions between the atoms involved in primary hydrogen bonding are the same, with two repulsive interactions in each complex. However, in G–G there are additional attractive secondary interactions between the amino groups on C2 (which do not participate in primary hydrogen bonding) with the carbonyl oxygens on C6. The situation is similar in C–C, where there are attractive interactions between the oxygens on C2 with the amino groups on C4. No such interactions are present in the other base pairs. In fact, there are *repulsive* secondary interactions in T–T between the C2 and C4 oxygens and perhaps between N7 and the C2 oxygen in A–T. Thus, recognition of the secondary electrostatic interactions is useful in understanding trends in base pair association energies. Of course, their applicability extends to hydrogen-bonded host–guest systgems in general. For example, such secondary interactions were recently invoked to explain the differences in the intramolecular associations between dilactams, diimides, and lactam–imides.[41]

(40) Switzer, C.; Moroney, S. E.; Benner, S. A. *J. Am. Chem. Soc.* **1989**, *111*, 8322. Piccirilli, J. A.; Krauch, T.; Moroney, S. E.; Benner, S. A. *Nature* **1990**, *343*, 33.

Conclusions

OPLS parameters have been developed for the nucleotide bases and DAP. Interaction energies were obtained using these parameters for 31 base pair geometries that are consistent with the limited, related experimental data.

Relative association constants between the G-C and A-U base pairs and between G-C and U-DAP in chloroform were then calculated via Monte Carlo simulations. The results are qualitatively consistent with experimental data, although the calculated preferences for G-C are larger than the experimentally based estimates. Comparison of the two results, however, leads to similar association constants for A-U and U-DAP, a result in agreement with experiment.

The much stronger association of G-C relative to the similarly triply hydrogen-bonded U-DAP is explained by consideration of secondary electrostatic interactions. The arrangement of hydrogen-bond donor and acceptor sites in the two molecules forming a complex can lead to a substantial attractive or repulsive secondary interactions. Such considerations have general applicability in understanding variations in hydrogen-bonding complexation in many contexts.

Acknowledgment. Gratitude is expressed to the National Institutes of Health and the National Science Foundation for support of this work.

Registry No. DAP, 141-86-6; 9-methylguanine, 5502-78-3; 1-methylcytosine, 1122-47-0; 9-methyladenine, 700-00-5; 1-methyluracil, 615-77-0; uracil, 66-22-8; cytosine, 71-30-7; 4-aminopyrimidine, 591-54-8; 2-aminopyrimidin-4-one, 71-30-7; imidazole, 288-32-4; water, 7732-18-5; 1-methylthymine, 4160-72-9.

Supplementary Material Available: Optimized geometries and interaction energies for the 31 complexes of the DNA base pairs and tables showing the incremental free energy changes for the perturbation calculations (13 pages). Ordering information is given on any current masthead page.

(41) Jeong, K. S.; Tjivikua, T.; Rebek, J., Jr. *J. Am. Chem. Soc.* **1990**, *112*, 3215. Jeong, K. S.; Tjivikua, T.; Muehldorf, A.; Deslongchamps, G.; Famulok, M.; Rebek, J., Jr. *J. Am. Chem. Soc.* **1991**, *113*, 201. Jorgensen, W. L.; Severance, D. L. *J. Am. Chem. Soc.* **1991**, *113*, 209.

Activation of Carbon Dioxide by Electron Transfer and Transition Metals. Mechanism of Nickel-Catalyzed Electrocarboxylation of Aromatic Halides

Christian Amatore* and Anny Jutand

Contribution from the Ecole Normale Supérieure, Laboratoire de Chimie, URA CNRS 1110, 24 rue Lhomond, F-75231 Paris Cedex 05, France. *Received October 29, 1990*

Abstract: In the presence of stoichiometric amounts of carbon dioxide, and catalytic amounts of $Ni^{II}(dppe)Cl_2$, electrolysis of bromobenzene results in the nearly quantitative formation of benzoic acid with negligible production of benzene or biphenyl. The mechanism of the nickel-catalyzed electrocarboxylation is shown to proceed through a chain reaction involving Ni(0), Ni(I), Ni(II), and Ni(III) intermediates, very reminiscent of that previously established for the nickel-catalyzed coupling of bromobenzene. Based on a detailed kinetic analysis of the propagation of this catalytic chain and of its competition with the biphenyl chain, all the key steps of the catalytic chain are identified and their rate constants determined.

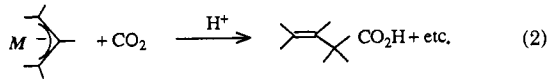
Introduction

A large amount of work has been devoted to reductive activation of carbon dioxide,[1] because of the possibility of energy storage or of using this ubiquituous molecule as a C_1 building block in organic chemistry.[2] Direct electrochemical reduction of CO_2 occurs at rather negative potentials (more negative than −2 versus SCE in most solvents). Depending on the exact experimental conditions, CO_2 reduction affords oxalate, formate, or equimolar amounts of carbon monoxide and carbonate.[1r,s] Therefore, the search for catalysts able to decrease the relatively high overpotential and to increase the selectivity of the reductive process has become an important challenge. Several groups have demonstrated that transition metal catalysts can be used advantageously in this search.[3] Indeed, carbon dioxide can bind to transition metal complexes in a variety of ways[4-8] that may allow its selective activation at rather low potentials.

In the inorganic or organometallic fields the importance of carbon dioxide coordination by metal complexes has been well recognized,[4] with a special emphasis on the creation of carbon-carbon bonds.[4,9,10] For example, CO_2 insertion into carbon-metal bonds:[10]

$$M\text{-}R + CO_2 \rightarrow M\text{-}O\text{-}C(O)\text{-}R \qquad (1)$$

or CO_2 addition to unsaturated hydrocarbons coordinated to a metal center,[4,9,11,12] as, e.g., to η^3-allyl complexes:[11]



(2)

or to η^2 complexes:[12]

$$\{M, CH_2{=}CHC_nH_{2n+1}\} + CO_2 \xrightarrow{H_2} C_{n+2}H_{2n+5}\text{-}CO_2H \qquad (3)$$

result in the overall formation of carboxylic acids or their de-

* To whom any correspondence should be addressed.

(1) See, e.g.: (a) Teeter, T. E.; van Rysselberghe, P. *J. Chem. Phys.* **1954**, *22*, 759. (b) Jordan, J.; Smith, P. T. *Proc. Chem. Soc.* **1960**, 240. (c) Roberts, J. L.; Sawyer, D. T. *J. Electroanal. Chem.* **1965**, *9*, 1. (d) Haynes, L. V.; Sawyer, D. T. *Anal. Chem.* **1967**, *39*, 332. (e) Bewick, A.; Greener, G. P. *Tetrahedron Lett.* **1969**, 4623. (f) Paik, W.; Andersen, T. N.; Eyring, H. *Electrochim. Acta* **1969**, *14*, 1217. (g) Bewick, A.; Greener, G. P. *Tetrahedron Lett.* **1970**, 391. (h) Udupa, K. S.; Subramanian, G. S.; Udupa, H. V. K. *Electrochim. Acta* **1971**, *16*, 1593. (i) Ryn, J.; Andersen, T. N.; Eyring, H. *J. Phys. Chem.* **1972**, *76*, 3278. (j) Tyssee, D. A.; Wagenknetch, J. H.; Baizer, M. M.; Chruma, J. L. *Tetrahedron Lett.* **1972**, 4809. (k) Gambino, S.; Silvestri, G. *Tetrahedron Lett.* **1973**, 3025. (l) Kaiser, U.; Heits, E. *Ber. Bunsenges. Phys. Chem.* **1973**, *77*, 818. (m) Aylmer-Kelly, A. W. B.; Bewick, A.; Cantrill, P. R.; Tuxford, A. M. *Discuss. Faraday Soc.* **1973**, *56*, 96. (n) Russel, P. G.; Kovac, N.; Srinivasan, S.; Steinberg, M. *J. Electrochem. Soc.* **1977**, *124*, 1329. (o) Lamy, E.; Nadjo, L.; Savéant, J. M. *J. Electroanal. Chem.* **1977**, *78*, 403. (p) Williams, R.; Crandall, R. S.; Bloom, A. *Appl. Phys. Lett.* **1978**, *33*, 381. (q) Gressin, J. C.; Michelet, D.; Nadjo, L.; Savéant, J. M. *Nouv. J. Chim.* **1979**, *3*, 545. (r) Amatore, C.; Savéant, J. M. *J. Am. Chem. Soc.* **1981**, *103*, 5021. (s) Amatore, C.; Savéant, J. M. *J. Electroanal. Chem.* **1981**, *125*, 22. (t) Fisher, J.; Lehmann, T.; Heitz, E. *J. Appl. Electrochem.* **1981**, *11*, 743. (u) Hori, Y.; Suzuki, S. *Bull. Chem. Soc. Jpn.* **1982**, *55*, 660. (v) Eggins, B. R.; McNeill, J. *J. Electroanal. Chem.* **1983**, *148*, 17. (w) Kapusta, S.; Hakerman, N. *J. Electrochem. Soc.* **1983**, *130*, 607. (x) Canfield, D.; Frese, K. W., Jr. *J. Electrochem. Soc.* **1983**, *130*, 1772.

0002-7863/91/1513-2819$02.50/0