Thermodynamic Parameters for Stacking and Hydrogen Bonding of Nucleic Acid Bases in Aqueous Solution: Ab Initio/Langevin Dipoles Study

Jan Florián,*,† Jiří Šponer,‡ and Arieh Warshel†

Department of Chemistry, University of Southern California, Los Angeles, California 90089-1062, and J. Heyrovský Institute of Physical Chemistry, Academy of Sciences of the Czech Republic, Dolejškova 3, 182 23 Prague 8, Czech Republic

Received: September 14, 1998

The potentials of mean force (PMF) for the association of purine, adenine, thymine, guanine, cytosine, and uracil in aqueous solution are investigated using ab initio MP2/6-31G(d-0.25) calculations (diffuse d-polarization functions were used) and Langevin dipoles solvation model. The entropy contributions to the free energies for stacking and hydrogen bonding are approximated using the linear relationship between binding enthalpies and entropies determined here from the available experimental data. This methodology is used to evaluate the dependence of PMF, and the gas-phase and solvation energies on the twist angle (Ω) in a number of undisplaced face-to-back stacking complexes. Further, we characterized the vertical association of the parallel (Ω = 0°) and antiparallel (Ω = 180°) stacked cytosine dimers. The results show large compensation between the gas-phase and solvation energetics and an overall preference of the bases in the undisplaced face-to-back stacked complexes for the twist angles near 30°. An important exception from this trend involves the GC and CG complexes, for which the largest stabilization occurs for the twist angle near 180°. In addition, free energies for the formation of 27 hydrogen-bonded base pairs were determined and compared with their stacking counterparts. The calculated standard free energies for the formation of stacked and hydrogen-bonded complexes at 298 K and neutral pH fell in a narrow region between 0.3 and −1.9 kcal/mol. Here, the hydrogen-bonded Watson−Crick guanine·cytosine base pair was found to be the most stable of all studied complexes. In agreement with the previous experimental findings, complexes containing purine bases were calculated to be more stable than their pyrimidine-containing counterparts.

1. Introduction

The stacking and hydrogen-bonding interactions between nucleobases are important forces stabilizing DNA double helix.[1−3] The nature of these forces has been examined by a wide range of computational approaches. The hydrophobic and electrostatic solute−solvent interactions were found to play a significant role in the semiempirical[4] and classical[5−13] simulations. In addition, recent ab initio quantum mechanical studies[14−16] as well as some earlier experimental works[17,18] stressed the importance of the electron correlation (dispersion) contribution to the interaction energy for nucleobase stacking in the gas phase.

To progress further in our understanding of the properties of DNA it is important to integrate and properly balance the classical and quantum mechanical description of the energetics of the base-stacking interactions in water. This goal can be achieved using ab initio calculations coupled with a reliable model of a solute−solvent interface. In this paper, we combine the ab initio Hartree−Fock (HF) and Møller−Plesset (MP2) methods with the Langevin dipoles (LD) solvation model. Our model is calibrated using the enthalpy−entropy compensation deducted from the relevant experimental thermodynamics of nucleoside association in aqueous solution. We first evaluated the free energy profile (potential of mean force) for the vertical dissociation of the stacked cytosine dimer. Further, the twist of the stacked bases with respect to each other around the axis perpendicular to the molecular planes is characterized for all face-to-back combinations of the adenine, guanine, cytosine, and uracil molecules, and thymine−thymine and purine−purine dimers. The twist angles in the 0°−360° range are considered. At the same computational level, free energies for the formation of 27 hydrogen-bonded base pairs are examined. The obtained results enable us to compare consistently the hydrogen-bonding and stacking thermodynamics in aqueous solution.

2. Methods

2.1. Definition of the Twist Angle Ω. According to the generally used convention, the bases are listed in the order corresponding to the 5′ to 3′ direction. In the absence of the sugar−phosphate linkage in our model, the order of bases is determined as follows (Figure 1). First, the bases are placed in two parallel planes, which are perpendicular to the viewing direction. Then, the N9, C8, and N7 atoms of purines, and N1, C6, and C5 atoms of pyridines, are arranged counterclockwise if the bases are viewed in the 5′ to 3′ direction. This type of stacking is denoted as face-to-back. The face-to-back stacking arrangement occurs for bases belonging to the same strand of DNA double helix. On the other hand, the face-to-face arrangement, in which the second base is flipped around its glycosidic bond, is characteristic of interstrand stacking interactions.

The twist angle Ω is defined as the rotation of the second base (with respect to the first base) around the axis perpendicular to the molecular planes and pointing from to the first to the

* Corresponding author. E-mail: florian@usc.edu.
† University of Southern California.
‡ Academy of Sciences of the Czech Republic.

10.1021/jp983699s CCC: $18.00
Published on Web 01/20/1999

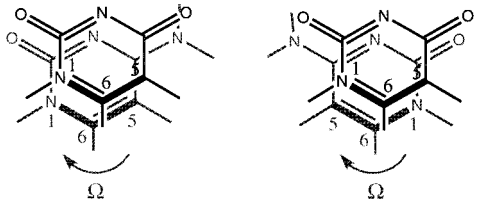

Figure 1. Face-to-back (left) and face-to-face (right) stacking in the UC dimer, $\Omega = 0°$ for the face-to-back structure. Arrows indicate the positive twist directions if cytosine is rotated.

second base, i.e., in the 5′ to 3′ direction. Consequently, positive twist angles correspond to the clockwise rotation. For the face-to-back complexes, the zero twist angle corresponds to the parallel glycosidic bonds, which are approximated here by the N9−H and N1−H bonds for purine and pyrimidine bases, respectively.

2.2. Ab Initio Calculations. The correlated gas-phase ab initio calculations represent the most costly part of the study. Therefore, wherever possible, we have used geometries and energies of stacked and H-bonded complexes which were published in the previous papers.[15,19] Nevertheless, a number of new structures were evaluated in the course of this study.

Stacking complexes of base pairs were studied using one-dimensional potential energy searches with rigid intramolecular geometries of planar isolated bases, obtained at the MP2/6-31G* level.[15] The interaction energies were evaluated using the "frozen core" MP2 procedure with a standard 6-31G basis set augmented by diffuse d-polarization functions with an exponent of 0.25 to all second-row elements (designated 6-31G(d-0.25)). Interaction energies were corrected for the basis set superposition error at both MP2 and HF levels. The use of diffuse polarization functions instead of the standard ones is required to include a sufficient amount of the intermolecular electron correlation effects.[15,16] Consideration of standard d-shells would drastically underestimate the dispersion attraction. The MP2/6-31G(d-0.25) aromatic stacking energies are expected to be very close to the actual values, since certain undervaluation of the dispersion energy due to the size of the 6-31G(d-0.25) basis set is compensated for by the neglect of higher-order electron correlation contributions. These are repulsive for all aromatic stacking clusters (see ref 20 and references therein). We could not use gradient optimization for stacked pairs for the following reasons. It is still too demanding, it is too much spoiled by the basis set superposition error, and some stacking structures would finally converge into H-bonding arrangements since there is no stacking minimum on their respective gas-phase potential energy surfaces. Moreover, recent MP2/6-31G* gradient optimizations of several stacked DNA base dimers show convincingly that a single-point search with rigid monomers provides excellent estimates of gas-phase base stacking energies.[21]

The interaction energies of H-bonded base pairs have been evaluated by means of the frozen-core MP2/6-31G(d-0.25) method (corrected for the basis set superposition error) using gradient-optimized geometries of base pairs assuming C_s symmetry. The optimizations were carried out within the Hartree−Fock approximation using standard 6-31G** basis set of atomic orbitals. For more details see ref 19. All calculations were done using the Gaussian94 program.[22]

2.3. Hydration Free Energies. Contributions of aqueous solvation to the energetics of the formation of stacked complexes were evaluated using the recent version of the Langevin dipoles (LD) solvation model.[23] This method represents the distribution and average polarization of the solvent molecules by the set of Langevin-type point dipoles centered on a cubic grid surrounding the solute. The electrostatic part (ΔG_{ES}) of the solvation free energy is determined from the magnitude of the interaction energy between the electrostatic potential-derived (ESP) atomic charges on the solute, and solvent dipoles, where the solvent dipole−dipole interactions are taken into account by an iterative procedure. The total solvation free energy (ΔG_{solv}) is obtained as a sum of ΔG_{ES} and the terms approximating the van der Waals (ΔG_{vdW}) and hydrophobic (ΔG_{phob}) energy, and the polarization of the solute by the solvent (ΔG_{relax}). The ΔG_{relax} term was evaluated from the solvated atomic charges calculated using the polarized continuum model (PCM)[24,25] implemented in the G94 program. The default Pauling's atomic radii scaled by the factor of 1.2, dielectric constant of water ($\epsilon = 80$) and the HF/6-31G* basis were used in the PCM calculations.

Although the LD model has not been parametrized to provide the enthalpic and entropic parts of ΔG_{solv}, we assume that for stacking of neutral aromatic molecules such as nucleobases, the entropic part of the stacking solvation free energy, $T\Delta S_{stack}$ can be reasonably approximated by the hydrophobic term:

$$T\Delta\Delta S_{solv} = \Delta\Delta G_{phob} \quad (1)$$

$$\Delta\Delta H_{solv} = \Delta\Delta G_{solv} - \Delta\Delta G_{phob} \quad (2)$$

The iterative LD calculations were carried out for the gas-phase HF/6-31G** geometries using the program ChemSol.[26] In these calculations, we employed the default parameters obtained previously.[23] Although the training set for this parametrization did not contain neutral nucleic acid bases (for which experimental values of ΔG_{solv} are not available), ΔG_{solv} values (kcal/mol) calculated by us for uracil (−13.6), thymine (−12.6), adenine (−10.8), cytosine (−18.0), guanine (−20.4), and purine (−9.7) agree well with the results of the recent free energy perturbation (FEP) calculations.[27,28] In addition, our relative solvation free energies are in reasonable agreement with the observed order of the distribution coefficients of butylated bases between water and cyclohexane.[29]

2.3. Equilibrium Constants. The methodology used here for the prediction of equilibrium constants for base stacking in water falls into the category of hybrid ab initio/LD calculations, often denoted as QM(ai)/LD.[30] The quantum mechanical treatment has the advantage of treating accurately the polarization and electron correlation effects, which play an important role in stacking interactions.[14−16] Also, the QM(ai)/LD calculations are computationally less demanding than molecular dynamic (MD) FEP calculations. On the other hand, the disadvantage of the QM(ai)/LD method stems from the fact that it cannot predict the entropic part of the binding free energy. It is therefore essential to calibrate the calculated results on available experimental data of related compounds. This is done here by taking advantage of the linear relationship between the experimental enthalpies and entropies for stacking in dimers of purine, 6-methylpurine, deoxyadenosine, caffeine, cytidine, uridine, and thymidine[1] (Supporting Information, Figure 1S),

$$\Delta S_{bind} = 2.08\Delta H_{bind} - 3.99 \quad (3)$$

To use this equation, which involves total entropy (ΔS_{bind}, cal mol^{-1} K^{-1} and enthalpy (ΔH_{bind}, kcal/mol) changes related to the formation of the stacked complex from the individual monomers in water at 1M concentration, we approximate the binding enthalpy as

$$\Delta H_{bind} = \Delta E_{gas} + \Delta\Delta H_{solv} \quad (4a)$$

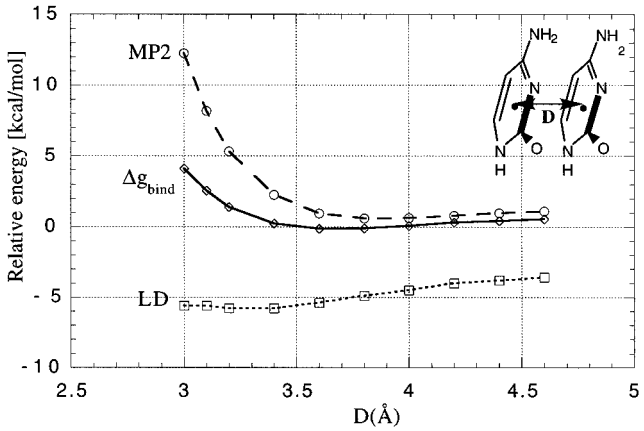

Figure 2. Variation of the gas-phase interaction energy (ΔE_{gas}, MP2/6-31(d-0.25)), solvation free energy ($\Delta\Delta G_{solv}$, LD), and PMF (Δg_{bind}) for the vertical association in the cytosine dimer, $\Omega = 0°$.

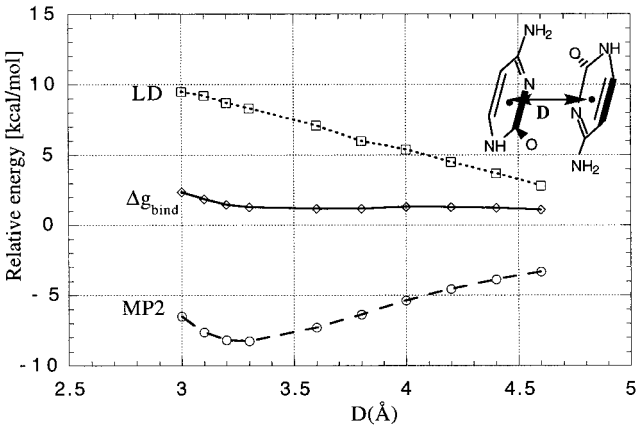

Figure 3. Variation of ΔE_{gas} (MP2), $\Delta\Delta G_{solv}$ (LD), and PMF (Δg_{bind}) for the vertical association in the cytosine dimer, $\Omega = 180°$.

Here, ΔE_{gas} represents the BSSE-corrected gas-phase energy difference obtained at the MP2 level (see section 2.2). The use of ΔE_{gas} in place of the gas-phase enthalpy difference seems to be a plausible approximation for stacking interactions, for which the vibrational energy difference can be expected to be of similar magnitude but opposite sign as the pV term. For hydrogen complexes, the gas-phase enthalpy difference, $\Delta H^0_{gas} = \Delta E_{gas} + \Delta ZPE$, where ΔZPE is the change in the zero-point vibrational energy, was used for the evaluation of ΔH_{bind}:

$$\Delta H_{bind} = \Delta H^0_{gas} + \Delta\Delta H_{solv} \qquad (4b)$$

From eqs 1−4, the potential of mean force $g_{bind}(\mathbf{R})$ at 298 K can be expressed as

$$\Delta g_{bind}(\mathbf{R}) = 0.38[\Delta H^{\circ}_{gas}(\mathbf{R}) + \Delta\Delta H_{solv}(\mathbf{R})] + 1.19 \quad (5)$$

where $g_{bind}(\mathbf{R})$ (kcal/mol) is the free energy change associated with bringing the infinitely separated monomers into a contact configuration characterized by the intermolecular coordinate **R**. In this work, potential of mean force was examined as a function of the intermolecular separation, D, and the twist angle, Ω. For practical reasons, only the face-to-back orientation of the bases and zero displacement (see section 2.2) were considered. From these calculations, the stacking free energy ΔG_{stack} and the corresponding equilibrium constant K_{stack} have been determined as

$$\Delta G_{stack} = \Delta g_{bind}(\Omega = \Omega_{min}, D = D_{min}) - RT \ln 2 \quad (6a)$$

and

$$K_{stack} = \exp[-\Delta G_{stack}/RT] \qquad (7a)$$

where $D_{min} = 3.3$ Å and Ω_{min} is the twist angle, for which $\Delta g_{bind}(\Omega, D_{min})$ potential attains its minimum value. The last term in eq 6a was included to take into account the existence of two equivalent face-to-back and back-to-face configurations. Note that the eq 6a assumes that the face-to-face dimers and configurations with nonzero displacements are less stable than the face-to-back structures considered in this work. To confirm this assumption a more extensive sampling of the configuration space is needed.

Although eq 3 has been derived for the stacked complexes, this equation is, at least in principle, applicable to any association processes occurring in aqueous solution. In this context, eqs 1−5 enable one to compare directly the potentials of mean force for stacking and hydrogen-bonding interactions of nucleic acid bases. Since the solvation and electron-correlation effects on the equilibrium geometry of hydrogen-bonded complexes are much smaller than for stacking, the free energies (ΔG_{hb}) and equilibrium constants (K_{hb}) for the formation of hydrogen-bonded complexes were evaluated as

$$\Delta G_{hb} = \Delta g_{bind}(R = R_{gas}) \qquad (6b)$$

and

$$K_{hb} = \exp[-\Delta G_{hb}/RT] \qquad (7b)$$

respectively. Here, R_{gas} denotes the HF/6-31G* gas-phase geometry of the complex.

3. Results and Discussion

3.1. Vertical Association. The variation of the ΔE_{gas} and $\Delta\Delta G_{solv}$ energies along the reaction coordinate for stacking of nucleic acid bases was studied using the cytosine dimer as a model system. In this calculation, the vertical separation of two cytosine molecules (lying in parallel planes) was varied from 3 to 4.6 Å keeping the geometries of monomers frozen (see the Methods section). Two limiting cases involving parallel and antiparallel arrangement of the dipole moments of the cytosine molecules were examined.

As expected, the gas-phase interaction energy of the parallel monomers is repulsive (Figure 2), with a very shallow minimum at the distance of 3.9 Å. The stacked complex is stabilized significantly by the solvation energy so that the potential of mean force for the association of the parallel cytosine molecules features a flat minimum near 3.6 Å. The predicted stability of the parallel arrangement of the stacked cytosine molecules agrees well with the cytosine crystal structure.[31] In the crystal, however, cytosine molecules are considerably displaced, and this reduces the electrostatic repulsion. Perhaps due to this displacement, which was not considered in our calculations, the interbase separation of 3.36 Å observed in the crystal is somewhat shorter than its calculated counterpart. Let us note, however, that the crystal structures are not always directly comparable with our calculations due to crystal packing forces, large displacements, and limited amount of solvent molecules present in the molecular crystals of nucleic acid bases.[32]

The largest part of the solvation stabilization originates from its electrostatic part (Supporting Information, Figure 2S). This is because, for neutral solutes, this part of the solvation energy is in the first approximation (Onsager model) proportional to

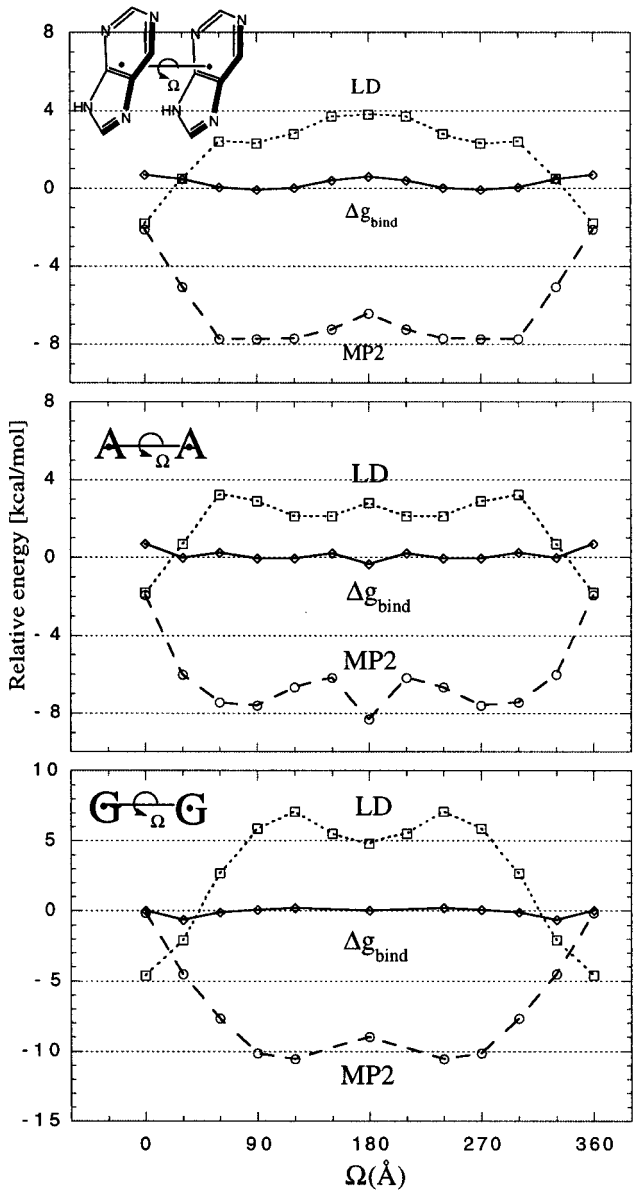

Figure 4. Dependence of ΔE_{gas} (MP2), $\Delta\Delta G_{solv}$ (LD), and PMF (Δg_{bind}) on the twist angle (Ω) for face-to-back self-stacking in purines. Nucleic acid bases are denoted by the first letter of their name.

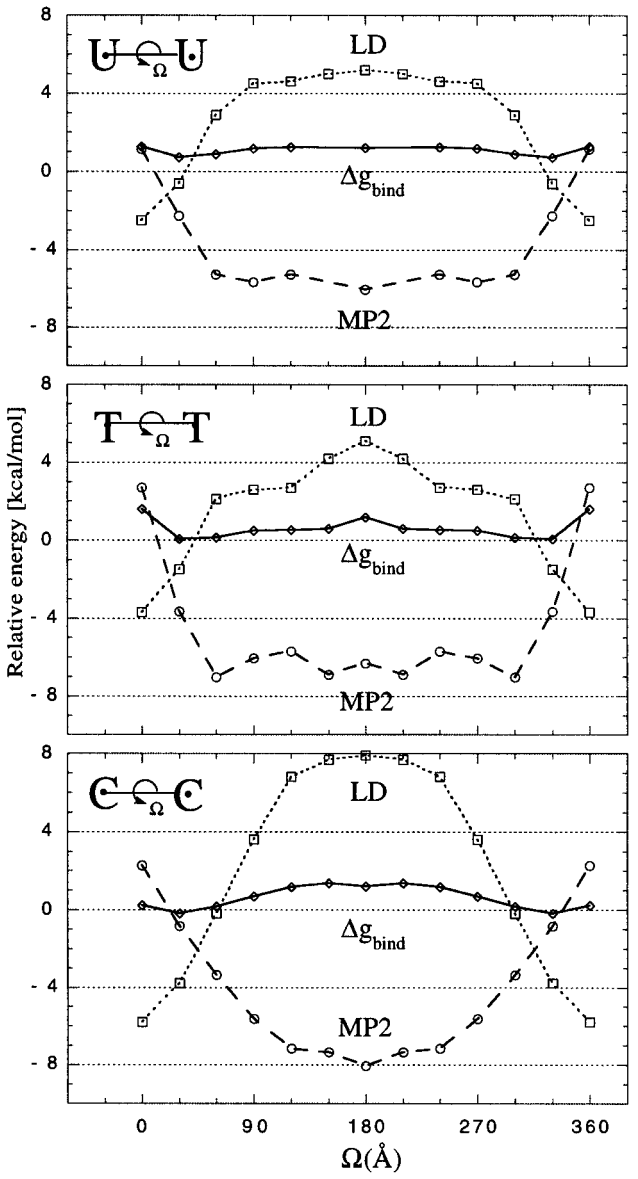

Figure 5. Dependence of ΔE_{gas} (MP2), $\Delta\Delta G_{solv}$ (LD), and PMF (Δg_{bind}) on the twist angle (Ω) for face-to-back self-stacking of pyrimidines.

the square of the dipole moment. As a result, the solvation energy of the complex with parallel dipoles is more negative than the sum of the solvation energies of interactants, and consequently it has a stabilizing effect on the complex. The contribution of the solute polarization to this term was found to increase from about 30% to 45% upon increasing intermolecular distance. The remaining part (about −1 kcal/mol) of the stabilization energy was found to originate from the hydrophobic terms. The electrostatic and hydrophobic forces, which induce stacking, are partly offset by the positive (about 2 kcal/mol) van der Waals (vdW) term simulating vdW interactions between the cytosine and the solvent molecules. The magnitudes of the hydrophobic and vdW energies change very slowly with the monomer separation and these changes tend to cancel each other.

The gas phase and solvation energies interchange their roles when one of the cytosine monomers is rotated by 180° around the axis that is perpendicular to the molecular plane and intersects the centers of mass of both monomers ($\Omega = 180°$). In this arrangement, the gas-phase energy stabilizes the stacked complex, with the minimum for the separation of 3.3 Å. The existence of this minimum is the result of a delicate balance between the Hartree−Fock and electron correlation energies (Supporting Information, Figure 5S). However, this minimum practically disappears in aqueous solution (Figure 3). Again, the analysis of the components of $\Delta\Delta G_{solv}$ shows that the primary role in the solvation energetics is played by the electrostatic and solute polarization terms (Supporting Information, Figure 4S). (For a related early finding see ref 4.)

3.2. Variation of the Twist Angle. Because the study of the vertical association of the cytosine dimer in water indicated that the energetics of this process strongly depends on the twist angle Ω (see above), we evaluated the gas phase and solvation energies of various stacked base pairs as a function of this angle. In these calculations, the intermolecular distance was fixed at 3.3 Å. Geometries of monomers were rigid and we have used the planar MP2 gas-phase optimized structures—see the Methods section.

The results calculated for the self-stacking of substituted purines (Figure 4) show a large compensation of the gas-phase and solvation energies. In purine dimer, the minimum of the potential of mean force, $\Delta g_{bind}(\Omega)$, occurs in the 60−90° region.

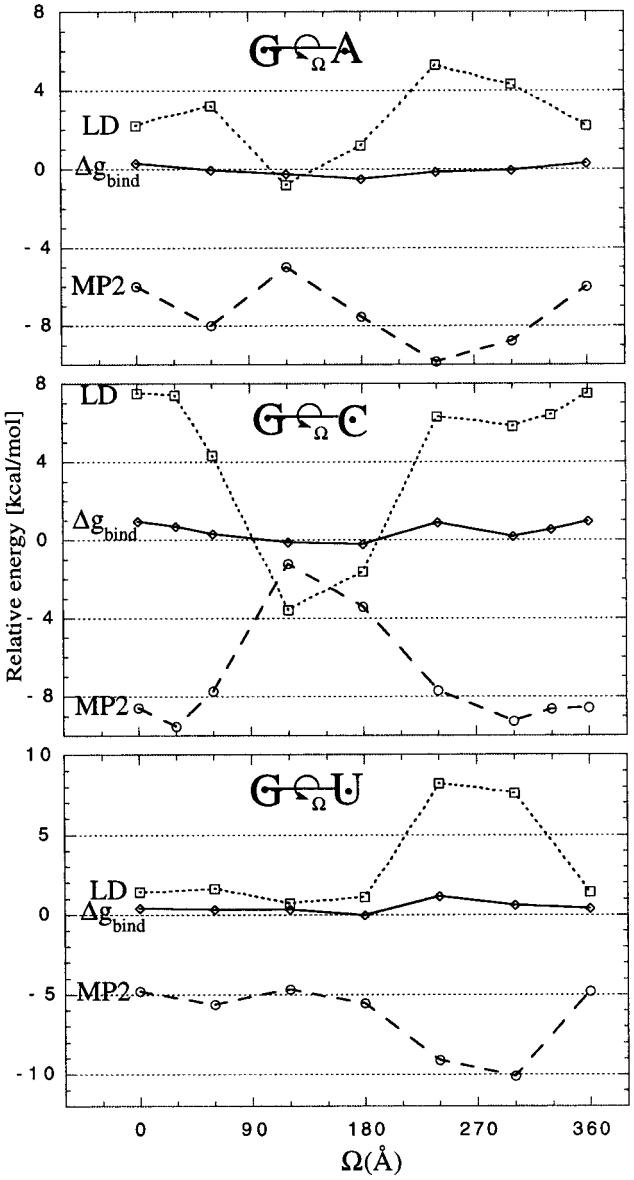

Figure 6. Dependence of ΔE_{gas} (MP2), $\Delta\Delta G_{solv}$ (LD), and PMF (Δg_{bind}) on the twist angle (Ω) for face-to-back stacking of GA, GC and GU dimers.

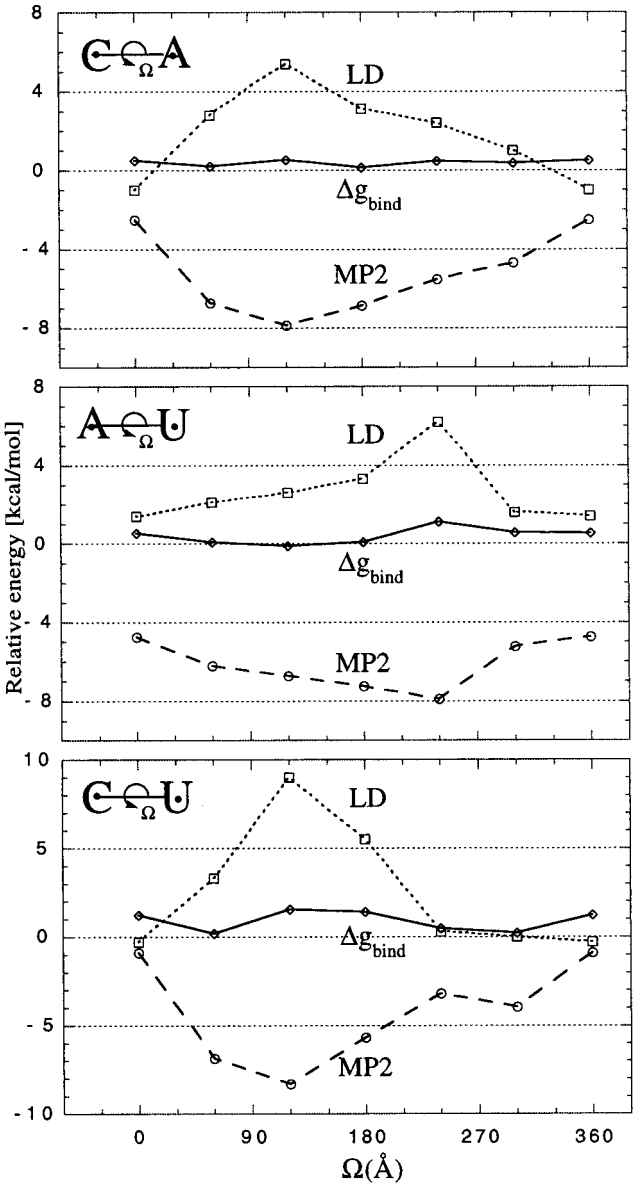

Figure 7. Dependence of ΔE_{gas} (MP2), $\Delta\Delta G_{solv}$ (LD), and PMF (Δg_{bind}) on the twist angle (Ω) for face-to-back stacking of CA, AU and UC dimers.

Interestingly, this is the region of the minimal overlap of the molecular surfaces. The presence of the additional amino group in adenine results in the destabilization of the 60°−90° region in favor of the geometries with the twist angles near 30° and 180°. The stable free-energy regions calculated for adenine are retained in the guanine dimer. Here, however, the stacked structure for $\Omega = 30°$ has the lowest energy in solution.

The twist dependence of the self-stacking energies of pyrimidine bases is presented in Figure 5. Interestingly, for all pyrimidine bases $\Delta g_{bind}(\Omega)$ reaches its minimum value for the twist of 30°. This conformation is further stabilized in the presence of the CH_3 group at the C5 position of the ring due to the steric destabilization of the parallel stack in the thymine dimer. Because also the stabilities of adenine and guanine stacked homodimers are enhanced for $\Omega = \pm 30°$, the stacking interactions are predicted to be a driving force for the formation of ordered helical conformations of homosequences of oligo- and polynucleotide single strands. Such ordered structures have been observed at neutral pH for oligo- and poly(A) and poly(dA),[33,34] and oligo- and poly(C),[35,36] whereas the observation of helical conformations of single strands containing guanine has been obscured by the formation of higly stable quadruple helices.[1,51] The unimolecular coil−helix transition has been detected also for mixed-sequence oligonucleotides.[37] However, the stability of the single-stranded helical structures was found to be smaller than for homosequences of the comparable length.

The energy profiles calculated for the mixed complexes are presented in Figures 6 and 7. The profiles for the remaining face-to-back complexes can be determined from the data in Figures 6 and 7 using the formula

$$g_{bind}{}^{BA}(\Omega) = g_{bind}{}^{AB}(-\Omega) \qquad (8)$$

where indexes A and B denote stacked bases. Naturally, the same relation is valid for gas-phase and solvation energies. For all complexes involving guanine the minimum of $\Delta g_{bind}(\Omega)$ occurs for $\Omega = 180°$ (Figure 6). Other regularities found involve the stabilization of the twist angles in the 30°−60° region in all pyrimidine−pyrimidine face-to-back complexes, and rather small Ω dependence of $\Delta g_{bind}(\Omega)$ for complexes involving adenine.

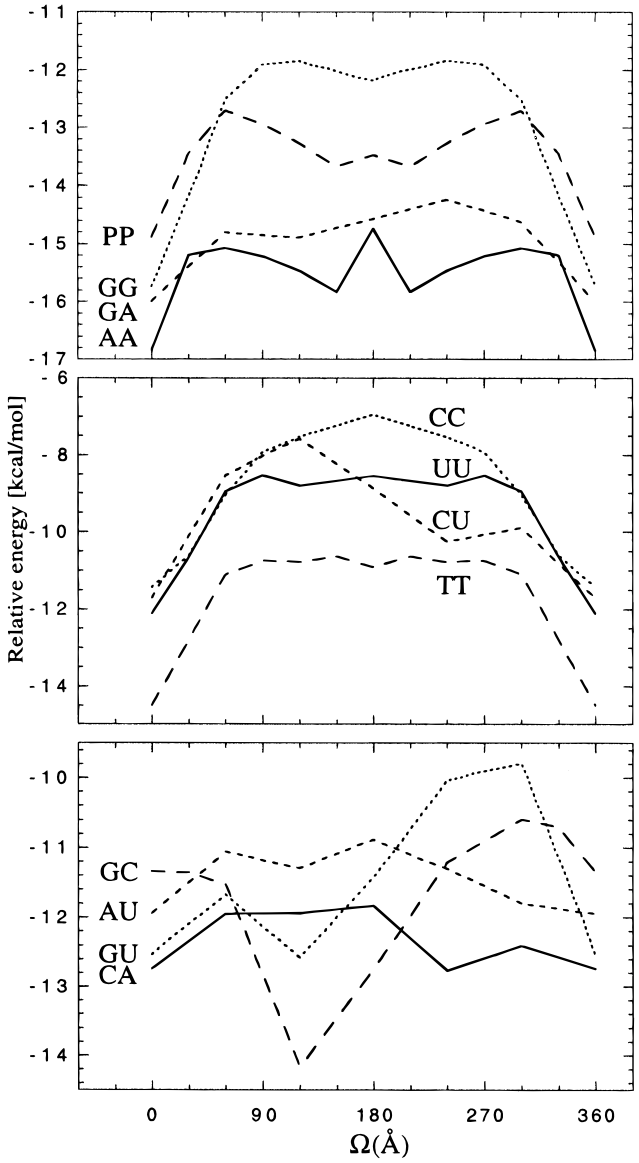

Figure 8. Twist dependence of the electron-correlation part of ΔE_{gas} for face-to-back stacking of nucleic acid bases and the purine·purine dimer (PP).

The electron correlation components of ΔE_{gas} for stacked nucleic acid bases are compared in Figure 8. The contribution of electron correlation, which determines attractive van der Waals−London interactions between bases, was found to stabilize strongly stacked dimers for all Ω values. This stabilization is more pronounced for complexes involving larger and less polar nucleobases. In addition, there is a significant Ω dependence of the correlation contribution, which tends to favor structures with $\Omega = 0°$. The GC and CG pairs, for which the correlation component attains the largest magnitude for $\Omega = 120°$ and 240°, respectively, represent the only exception. Quite unexpectedly, we found that the electron correlation and solvation contributions depend on the twist angle in a similar way. Thus, the fact that in many cases the most stable stacking conformations in water and molecular crystals are characterized by largely repulsive gas-phase electrostatic interactions can be explained by the synergetic action of the electron correlation and solvation forces.

3.3. Comparison of the Stacking and Hydrogen-Bonding Propensities. The predicted free energies for *vertical stacking* of nucleic acid bases in aqueous solution (Table 1) indicate rather small differences in the stability of different dimers. Nevertheless, we can see a clear trends in that the presence of the $-NH_2$, $=O$, and $-CH_3$ functional groups enhances stacking interactions. This tendency agrees well with the results of experimental studies of stacking interactions in water [1] and gas phase.[38] Considering that the replacement of uracil by thymine results in about 0.5 kcal/mol stabilization of the stacked complex, free energies for the stacking of DNA bases are predicted to fall within the range of only 0.7 kcal/mol. As expected, dimers involving two purine bases were found to be more stable then the purine·pyrimidine and pyrimidine·pyrimidine complexes. The calculated magnitudes of the stacking free energies are consistent with observed equilibrium constants, although, besides purine dimer, experimental data presented in Table 1 actually correspond to the association between nucleosides. The use of nucleosides for association experiments was necessitated by very low solubilities of nucleic acid bases in aqueous solution.[1] Because the solubility of guanosine is still very low, the stacking propensity of the guanine·guanine dimer could not be determined experimentally. Also, it should be pointed out that the experimental data obtained by the solid-state solubility [39] method actually reflect the properties of the solid−liquid interface. Furthermore, only for purine dimer it has been shown conclusively that the structure of the formed complex corresponds to the vertical stacking. This is because the vapor pressure osmometry and sedimentation equilibrium techniques do not distinguish the H-bonded and stacked complexes. Conclusive evidence that the stacked purine dimer is actually formed was provided by the concentration dependence of the NMR signal measured for purine CH protons.[40] The analogous upfield chemical shifts were found for adenosine and deoxyadenosine dimers.[1] Unfortunately, because of the low solubility, the same NMR technique was not applied to the association of nucleobases or complexes involving guanosine and cytidine.

TABLE 1: Thermodynamic Properties for Stacking of Nucleic Acid Bases in Aqueous Solution (kcal/mol)[a]

		this work			
complex	expt ΔG	ΔG_{stack}[b]	ΔH_{stack}[c]	ΔS_{stack}[d]	ΔS_{solv}[e]
PP	−0.5[f]	−0.5	−3.3	−10.9	7.0
GG		−1.1	−4.9	−14.2	5.7
GA	−1.5[g]	−0.9	−4.5	−13.2	6.4
AA	−0.9[m]	−0.8	−4.0	−12.2	5.0
GC	−1.0[h]	−0.6	−3.7	−11.7	4.4
CC	0.1[i]	−0.6	−3.7	−11.5	3.4
AU	−0.8[j]	−0.5	−3.5	−11.1	2.4
GU		−0.5	−3.3	−10.7	4.0
TT	0.1[k]	−0.4	−3.1	−10.1	7.0
CA	−0.9[j]	−0.3	−2.8	−9.7	3.4
UC		−0.2	−2.7	−9.5	3.0
UU	0.3[l]	0.3	−1.3	−6.7	5.4

[a] 1 M aqueous solution, 298 K, 1 atm. [b] Equation 6a. [c] Equation 4a. [d] Total entropy change (cal/mol/K, eq 3). [e] Hydrophobic contribution to ΔS_{hb}, (cal/mol/K), eq 1. [f] Average of the values of −0.44 and −0.61 kcal/mol obtained for purine dimer by the vapor pressure osmometry (VPO)[47] and the sedimentation equilibrium technique,[48] respectively. [g] Average of the values of −1.7 and −1.3 kcal/mol determined for adenine·deoxyguanosine and adenosine·deoxyguanosine complexes from the measurements of solubilities of solid-state adenine and deoxyguanosine, respectively.[39] [h] Average of the values of −0.7 and −0.8 to −1.2 kcal/mol determined for deoxyguanosine·cytosine and deoxyguanosine· deoxycytidine complexes by the VPO[49] and solid-state solubility[39] methods, respectively. [i] Cytidine, VPO.[47] [j] Solid-state solubility measurement.[39,47] [k] Deoxythymidine, VPO.[49] [l] Uridine, VPO.[47] [m] Adenosine, VPO.[50]

TABLE 2: Calculated Thermodynamic Parameters for the Formation of Hydrogen-Bonded Complexes in Aqueous Solution (kcal/mol)[a]

complex[b]	ΔH^0_{gas}[b]	$\Delta\Delta G_{solv}$	$\Delta\Delta G_{hb}$[c]	K_{hb}[c]	ΔH_{hb}[d]	ΔS_{hb}[e]	ΔS_{solv}[f]
GCWC	−21.9	13.3	−1.9	24.3	−8.10	−20.8	1.7
GG1	−21.5	17.6	−0.4	2.0	−4.2	−12.7	-1.0
GCNEW	−19.0	14.7	−0.5	2.4	−4.5	−13.3	-0.7
CC	−15.5	11.1	−0.4	2.1	−4.3	−12.9	0.3
GG3	−16.6	9.7	−1.1	6.2	−6.0	−16.4	3.0
GA1	−13.3	7.8	−0.6	3.1	−4.9	−14.1	2.0
GT1	−13.3	6.5	−0.9	4.9	−5.6	−15.6	4.0
GT2	−13.0	6.9	−0.8	4.0	−5.3	−15.0	2.7
AC1	−11.7	5.7	−0.7	3.3	−5.0	−14.4	3.4
GC1	−12.3	6.6	−0.7	3.1	−4.9	−14.1	2.7
AC2	−11.4	4.1	−0.9	4.6	−5.5	−15.4	6.0
GA3	−12.3	5.2	−0.9	4.6	−5.5	−15.4	5.4
TAH	−11.4	4.6	−0.8	4.0	−5.3	−15.0	5.0
TARH	−11.3	4.8	−0.8	3.6	−5.1	−14.6	4.7
TAWC	−10.5	4.0	−0.8	3.6	−5.1	−14.6	4.7
TARWC	−10.4	4.1	−0.7	3.3	−5.0	−14.4	4.4
AA1	−9.3	3.8	−0.6	2.8	−4.7	−13.8	2.7
GA4	−9.9	4.0	−1.0	5.5	−5.8	−16.1	0.3
TC2	−9.5	6.0	−0.0	1.1	−3.2	−10.6	1.0
TC1	−9.5	5.6	−0.0	1.1	−3.3	−10.8	2.0
AA2	−8.8	2.0	−0.8	4.0	−5.3	−15.0	5.0
TT2	−9.1	3.0	−0.8	3.6	−5.1	−14.6	3.4
TT1	−9.2	2.9	−0.8	4.0	−5.3	−15.0	3.4
TT3	−9.1	3.3	−0.8	3.8	−5.2	−14.8	2.0
GA2	−8.9	2.9	−1.0	5.5	−5.8	−16.1	0.7
GG4	−9.4	1.8	−1.2	8.1	−6.4	−17.3	4.0
AA3	−7.8	1.4	−0.6	2.8	−4.7	−13.8	5.7

[a] 1 M aqueous solution, 298 K, 1 atm. [b] The structures, abreviations, and gas-phase enthalpies of the H-bonded complexes are taken from the ref 19. [c] Equations 6b and 7b. [d] Equation 4b. [e] Total entropy change (cal/mol/K, eq 3). [f] Hydrophobic contribution to ΔS_{hb} (cal/mol/K, eq 1).

The tendency for decreasing stability in the order purine·purine > purine·pyrimidine > pyrimidine·pyrimidine and small structure-related free-energy differences were also found for *H-bonded complexes* (Table 2; for the structures of 27 studied base pairs see ref. 19) The Watson−Crick guanine·cytosine base pair (GCWC), which is significantly more stable than other complexes, represents an important exception from this trend. More specifically, there is 1 kcal/mol free-energy difference between GCWC and other base pairs that are sterically compatible with the DNA structure (Figure 9). The exceptional stability of GCWC originates from its large gas-phase interaction enthalpy. Furthermore, this gas-phase attraction is offset by solvation to a lesser extent than in the case of GG1 and GCNEW complexes that are also very stable in the gas phase. In contrast, the canonical Watson−Crick base pairing between adenine and thymine (ATWC) shows no special stability. This fact is best demonstrated by similar equilibrium constants for all four adenine·thymine base pairs. In agreement with our results, crystals containing ATH (Hoogsteen) base pairs were grown from solution of 9-methyladenine and 1-methylthymine in 1:1 stoichiometry.[41] On the other hand, the ATWC base pair is known to stabilize DNA structure in solution. For example, the study of DNA melting thermodynamics that compared free energy changes upon dissociation of various terminal base pairs showed that the AT base pair is about 0.3 kcal/mol more stable than the GT, CT, and TT mispairs.[42] Thus, it appears that the relative ATWC stability is increased due to a combination of hydrogen-bonding and stacking interactions, and the presence of counterions. In addition, the direct comparison of our results with this and other experimental investigations of DNA and RNA duplex stability[43−45] is hampered by the entropic and

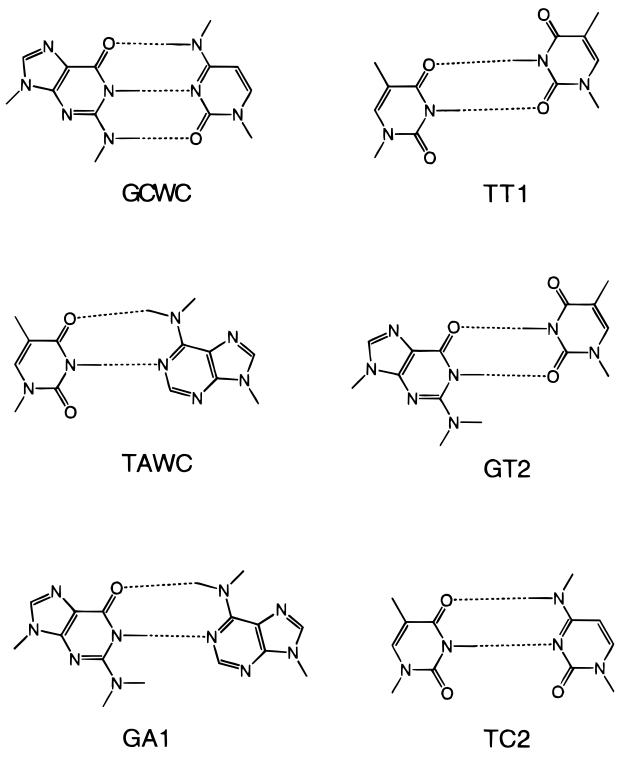

Figure 9. H-bonded base pairs that are sterically compatible with the DNA structure.

enthalpic effects of the sugar−phosphate backbone and interstrand stacking interactions that were not taken into account in this study.

A consistent theoretical description of the association processes presented in this paper enables us to address the *relative propensities of nucleobases for stacking or H bonding in aqueous solution*. Our calculations indicated that, with the exception of the GCWC pair mentioned above, the H-bonded and stacked arrangements are equally stable (cf. Tables 1 and 2). In contrast, previous molecular dynamics study of the association of adenine and thymine, and guanine·cytosine,[9,10] predicted the stacked complexes to be 1.0 to 1.7 kcal/mol (depending on the computational protocol) more stable than their H-bonded counterparts. While it is not clear which estimate is more realistic, we analyze below possible sources of systematic errors in our calculations that could lead either to the stabilization of the stacked complexes, or to the destabilization of H-bonded complexes. First, the stacking free energies presented in Table 1 could be underestimated due to the neglect of the displaced stacking configurations in our calculations. Because of the flat free energy surfaces predicted for twist and vertical separation, we estimate that this configuration-related underestimation of stacking stabilities amounts only to a few tens of kcal/mol. The second factor that should be considered is our assumption that the entropy−enthalpy compensation derived for base stacking is valid also for H-bonded complexes. However, the alternate relationships which were deduced by Petruska et al. from DNA melting thermodynamics[42,46] provide H-bonding free energies that are for most base pairs very similar to those obtained using eq 3. In fact, the only significant change implied by the relationship of refs 42 and 46 involves the association energy of the GCWC pair, which is decreased to −1.2 and −1.5 kcal/mol, respectively. Third, it is possible that the dipolar solvent model used in this study does not provide sufficient solvation destabilization for H-bonded complexes. However, $\Delta\Delta G_{solv}$ energies calculated for the formation of ATWC and

GCWC base pairs by the LD solvation model are by 7.4 and 2.3 kcal/mol more positive, respectively, than those calculated by Cieplak and Kollman using the all-atom solvent model.[9] Consequently, the smaller stabilities of the GCWC and ATWC complexes calculated by Cieplak and Kollman are entirely due to gas-phase energetics. In this area, however, our ab initio gas-phase results are expected to be more accurate than empirical potential energy functions. Overall, our results can be considered as an advance in the continuing effort to predict association free energies for both stacked and H-bonded complexes of nucleic acid bases in aqueous solution.

4. Concluding Remarks

In this paper, we consistently described stacking and hydrogen-bonding interactions of nucleic acid bases in water. We put emphasis on the formulation of the computational model that includes all the major components of the interaction energy, i.e., the electrostatic, inductive, electron correlation, and solvation terms. Problems with the calculation of the entropic contribution were circumvented assuming an entropy−enthalpy compensation derived by the analysis of thermodynamic properties observed for stacking of nucleosides in neutral aqueous solution. Despite significant simplifications involved, the calculated stacking free energies agree well with their experimental conterparts. Also, there is a considerable similarity between twist preferences obtained from our calculations and from the relevant crystal structures.[32] These agreements indicate the plausibility of our model. In addition, we find it promising that our solvation energies tend to almost completely compensate for the changes in gas-phase electrostatics. This behavior is required to obtain stable and reliable potentials of mean force for all orientations of the interacting molecules.

Clearly, the main practical drawback of our methodology is the need to evaluate the electron correlation energies and related basis set superposition errors for many points on the studied surface. Due to the large demands of these ab initio calculations on the computer resources we had to limit ourselves to the examination of undisplaced face-to-back stacking complexes. However, large displacements occur frequently in DNA. Therefore, conclusive links between our results (especially those for the twist dependence of stacking energy) and the DNA structure are not waranted. Nevertheless, there are several features that can be expected to hold even after the configuration space in solution is sampled more completely. These points include the following:

(i) Stabilities of stacking complexes of nucleobases decrease in the order purine·purine > purine·pyrimidine > pyrimidine·pyrimidine.

(ii) Methyl, amino and keto substituents on the purine and pyrimidine rings enforce stacking interactions.

(iii) The differences in free energies for the formation of the hydrogen-bonded and stacked configuration of a given complex are smaller than 0.5 kcal/mol. In specific cases, for example for the association between guanine and cytosine, hydrogen-bonded complexes are predicted to be more stable than the stacked complex between the same interactants.

(iv) Hydrogen-bonded guanine·cytosine complex (Watson−Crick) is notably more stable than other hydrogen-bonded base pairs.

Acknowledgment. This study was partly supported (J.Š.) by a grant 23/97/0029, Grant Agency of the Czech Republic. We would like to thank the USC and Brno supercomputing centers for generous allotment of computer time.

Supporting Information Available: Experimental enthalpy−entropy compensation for the self-association of nucleosides in aqueous solution; the variation of the components of the solvation free energy and MP2 energy with the vertical separation in the stacked cytosine dimer. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Ts'o, P. O. In *Basic principles in nucleic acid chemistry*; Ts'o, P. O., Ed.; Academic Press: New York, 1974; Vol. 1; p 454.

(2) Sweitzer, B. A.; Kool, E. T. *J. Am. Chem. Soc.* **1995**, *117*, 1863.

(3) Guckian, K. M.; Sweitzer, B. A.; Ren, R. X. F.; Sheils, C. J.; Paris, P. L.; Tahmassebi, D. C.; Kool, E. T. *J. Am. Chem. Soc.* **1996**, *118*, 8182.

(4) Warshel, A. *J. Phys. Chem.* **1979**, *83*, 1640.

(5) Sinanoglu, O.; Abdulnur, S. *Photochem. Photobiol.* **1964**, *3*, 333.

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

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

(8) Danilov, V. I.; Tolokh, I. S. *J. Biomol. Struct. Dyn.* **1984**, *2*, 119.

(9) Cieplak, P.; Kollman, P. A. *J. Am. Chem. Soc.* **1988**, *110*, 3334.

(10) Dang, L. X.; Kollman, P. A. *J. Am. Chem. Soc.* **1990**, *112*, 503.

(11) Friedman, R. A.; Honig, B. *Biophys. J.* **1995**, *69*, 1528.

(12) Arora, N.; Jayaram, B. *J. Phys. Chem.* **1998**, *102*, 6139.

(13) Danilov, V. I.; Zheltovsky, N. V.; Slyusarchuk, O. N.; Poltev, V. I.; Alderfer, J. L. *J. Biomol. Struct. Dyn.* **1998**, *15*, 69.

(14) Hobza, P.; Sponer, J.; Polasek, M. *J. Am. Chem. Soc.* **1995**, *117*, 792.

(15) Šponer, J.; Leszczynski, J.; Hobza, P. *J. Phys. Chem.* **1996**, *100*, 5590.

(16) Šponer, J.; Gabb, H. A.; Leszczynski, J.; Hobza, P. *Biophys. J.* **1997**, *73*, 76.

(17) Fetzer, S. M.; Huang, C.-R.; Harvey, R. G.; LeBreton, P. R. *J. Phys. Chem.* **1993**, *97*, 2385.

(18) Yu, C.; O'Donnell, T. J.; LeBreton, P. R. *J. Phys. Chem.* **1981**, *85*, 3851.

(19) Šponer, J.; Leszczynski, J.; Hobza, P. *J. Phys. Chem.* **1996**, *100*, 1965.

(20) Šponer, J.; Hobza, P. *Chem. Phys. Lett.* **1997**, *267*, 263.

(21) Hobza, P.; Sponer, J. *Chem. Phys. Lett.* **1998**, *288*, 7.

(22) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Gill, P. M. W.; Johnson, B. G.; Robb, M. A.; Cheeseman, J. R.; T. Keith; Petersson, G. A.; Montgomery, J. A.; Raghavachari, K.; Al-Laham, M. A.; Zakrzewski, V. G.; Ortiz, J. V.; Foresman, J. B.; Cioslowski, J.; Stefanov, B. B.; Nanayakkara, A.; Challacombe, M.; Peng, C. Y.; Ayala, P. Y.; Chen, W.; Wong, M. W.; Andres, J. L.; Replogle, E. S.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Binkley, J. S.; Defrees, D. J.; Baker, J.; Stewart, J. P.; Head-Gordon, M.; Gonzalez, C.; Pople, J. A. *Gaussian 94*, Revision D.2, Gaussian, Inc., 1995.

(23) Florián, J.; Warshel, A. *J. Phys. Chem. B* **1997**, *101*, 5583.

(24) Miertus, S.; Scrocco, E.; Tomasi, J. *Chem. Phys.* **1981**, *55*, 117.

(25) Miertus, S.; Tomasi, J. *Chem. Phys.* **1982**, *65*, 239.

(26) Florián, J.; Warshel, A. ChemSol, Version 1.0, University of Southern California, 1997. The program can be downloaded free of charge from http://apriori.usc.edu/florian.

(27) Elcock, A. H.; Richards, W. G. *J. Am. Chem. Soc.* **1993**, *115*, 7930.

(28) Miller, J. L.; Kollman, P. A. *J. Phys. Chem.* **1997**, *100*, 8587.

(29) Shih, P.; Pedersen, L. G.; Gibbs, P. R.; Wolfenden, R. *J. Mol. Biol.* **1998**, *280*, 421.

(30) Bentzien, J.; Florián, J.; Glennon, T.; Warshel, A. In *Combined Quantum Mechanical & Molecular Mechanical Methods*; Gao, J., Thompson, M. A., Eds.; American Chemical Society: Washington, DC, 1998; ACS Symposium Series Vol. 712; p 16.

(31) Baker, D.; Marsh, R. E. *Acta Crystallogr.* **1964**, *17*, 1581.

(32) Bugg, C. E.; Thomas, J. M.; Sundaralingam, M.; Rao, S. T. *Biopolymers* **1971**, *10*, 175.

(33) Dewey, T. G.; Turner, D. H. *Biochemistry* **1979**, *18*, 5757.

(34) Brams, J.; Michelson, A. M.; van Holde, K. E. *J. Mol. Biol.* **1986**, *15*, 467.

(35) Freier, S. M.; Hill, K. O.; Dewey, T. G.; Marky, L. A.; Breslauer, K. J.; Turner, D. H. *Biochemistry* **1981**, *20*, 1419.

(36) Pörschke, D. *Biochemistry* **1976**, *15*, 1495.

(37) Vesnaver, G.; Breslauer, K. J. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 3569.

(38) Zegar, I. S.; Prakash, A. S.; Harvey, R. G.; LeBreton, P. R. *J. Am. Chem. Soc.* **1985**, *107*, 7990.
(39) Nakano, N. I.; Igarashi, S. J. *Biochemistry* **1970**, *9*, 577.
(40) Chan, S. I.; Schweitzer, M. P.; Ts'o, P. O. P.; Helmkamp, G. K. *J. Am. Chem. Soc.* **1964**, *86*, 4182.
(41) Hoogsteen, K. *Acta Crystallogr.* **1963**, *16*, 907.
(42) Petruska, J.; Goodman, M. F.; Boosalis, M. S.; Sowers, L. C. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 6252.
(43) Freier, S. M.; Kierzek, R.; Jaeger, J. A.; Sugimoto, N.; Caruthers, M. H.; Neilson, T.; Turner, D. H. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 9373.
(44) SantaLucia Jr., J.; Allawi, H. T.; Senevirante, P. A. *Biochemistry* **1996**, *35*, 3555.
(45) Mathews, D. H.; Andre, T. C.; Kim, J.; Turner, D. H.; Zuker, M. In *Molecular Modeling of Nucleic Acids*; Leontis, N. B., SantaLucia Jr., J., Eds.; American Chemical Society: Washington, DC, 1998; Vol. 682; pp 246.
(46) Petruska, J.; Goodman, M. F. *J. Biol. Chem.* **1995**, *270*, 746.
(47) Ts'o, P. O. P.; Melvin, I. S.; Olson, A. C. *J. Am. Chem. Soc.* **1963**, *85*, 1289.
(48) van Holde, K. E.; Rossetti, G. P. *Biochemistry* **1967**, *6*, 2189.
(49) Solie, T. N.; Schellman, J. A. *J. Mol. Biol.* **1968**, *33*, 61.
(50) Broom, A. D.; Schweizer, M. P.; Ts'o, P. O. P. *J. Am. Chem. Soc.* **1967**, *89*, 3612.
(51) Smith, F. W.; Feigon, J. *Nature* **1992**, *356*, 164.