

# Potential Energy Surfaces of an Adenine–Thymine Base Pair and Its Methylated Analogue in the Presence of One and Two Water Molecules: Molecular Mechanics and Correlated Ab Initio Study

Martin Kabeláč, Lucie Zendlová, David Řeha, and Pavel Hobza\*

*Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, and Center for Biomolecules and Complex Molecular Systems, 166 10 Prague 6, Czech Republic*

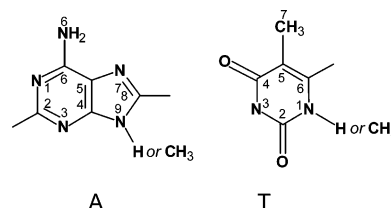
*Received: September 7, 2004; In Final Form: April 25, 2005*

Potential energy surfaces of monohydrated and dihydrated adenine–thymine and 9-methyladenine–1-methylthymine base pairs were examined by the molecular dynamics/quenching technique using the Cornell et al. force field (*J. Am. Chem. Soc.* **1995**, *117*, 5179). Long runs of molecular dynamics/quenching calculations allowed us to evaluate the free energy surface. The most stable and populated structures found were fully reoptimized at the correlated ab initio level employing the resolution of identity Møller–Plesset method. A systematic study of the base pairs' microhydration using both the empirical and the high-level correlated ab initio approaches is presented for the first time. We show that the occurrence of water molecules and their gradually increasing number as well as the methylation of the bases favor stacked structures over the planar hydrogen-bonded ones. These results based on the correlated ab initio calculations are in the excellent agreement with data obtained from our previous empirical potential molecular dynamics study (Kabeláč et al. *Chem.—Eur. J.* **2001**, *7*, 2067).

## Introduction

The DNA molecule is one of the most important biomolecules. DNA, among other functions, stores and transfers genetic information. It is well-known that the presence of solvent plays an important role in the stabilization of the three-dimensional structure of DNA. Depending on the humidity, ionic strength, and ratio of bases presented, DNA adopts multiple conformational forms (e.g., B-DNA, A-DNA, and Z-DNA),<sup>1,2</sup> while the most relevant biological architecture is the antiparallel, right-handed B-DNA double helix. Each DNA structure is the result of a balance of a number of interactions of the individual DNA building blocks. The following interactions are especially important: (i) planar interactions between the bases leading to the formation of the hydrogen-bonded structures stabilized mainly by the electrostatic interactions, (ii) vertical interactions leading to the formation of the stacked structures stabilized mainly by the London dispersion interactions, and (iii) interactions with ions and water molecules, both existing in the natural medium of biomolecules. Numerous X-ray studies have shown that about 20 water molecules are required for a proper description of the first hydration shell of a nucleotide pair.<sup>3</sup> Detailed statistical studies of the hydration sites around the well-refined B-DNA X-ray structures<sup>4–6</sup> predict a maximum of 17 water molecules per nucleotide pair and several local water-binding sites for bases (3 for purines and 2 for pyrimidines).

A large number of theoretical studies devoted to the hydration of isolated adenine and thymine bases have been published. Microhydration of thymine (or uracil) was examined at different levels of theory. Van Mourik<sup>7–9</sup> studied the potential energy surfaces of thymine and uracil with several (2–7) water molecules using the anisotropic atom–atom potential, and in the second step, the structures found were recomputed at the



**Figure 1.** Standard numbering of adenine (A) and thymine (T) atoms and their methylated analogues.

second-order Møller–Plesset level of theory (MP2). A slight preference for the formation of the water clusters instead of a distribution of separate water molecules was found. Similarly, Kryachko<sup>10</sup> found that in trihydrated thiouracil complexes two or three water molecules were bound together to the same binding site. Aamouche et al.<sup>11</sup> have compared the stability of the different uracil dihydrates at the density functional theory (DFT) level using a B3LYP functional, whereas Ghomi et al.<sup>12</sup> obtained the hexahydrated uracil structures with the EPIC potential. Aleman<sup>13</sup> studied the solvation of thymine using the combined discrete/self-consistent reaction-field method, and Chandra<sup>14</sup> calculated the protonation enthalpies of complexes between thymine and water.

Exhaustive studies of the hydration of different tautomers of cytosine, guanine, and adenine were published recently.<sup>15</sup> The most stable hydration site of the canonical form of adenine<sup>15a</sup> was found between the N3 and H9 atoms (for the standard atom labeling of the bases cf. Figure 1). In the most stable structure of dihydrated adenine, the water molecules interact and contact the H9 and N3 atoms of adenine. Similar results were also obtained by Desfrancois.<sup>16</sup> A DFT study of polyhydrated adenine was performed by Sukhanov<sup>17</sup> and Gu.<sup>18</sup> Chandra<sup>19</sup> studied the proton transfer and protonation enthalpies of hydrated adenine.

The base pair solvation has been also studied extensively. It is well-known that in the nonpolar solvent and also in the gas

\* Author to whom correspondence should be addressed. E-mail: hobza@uochb.cas.cz.

phase the base pairs prefer the planar hydrogen-bonded arrangement, whereas in water the stacked configurations are preferred.<sup>20–23</sup> The free energy/molecular dynamics with empirical potential study has shown that stacked complexes are slightly favored over hydrogen-bonded ones.<sup>24</sup> Sivanesan<sup>25</sup> studied the influence of the hydration on both hydrogen-bonded and stacked guanine–cytosine base pairs. He found that the stacked pair hydrates better than the corresponding hydrogen-bonded pair and that the stacked pair can accommodate 5–6 water molecules whereas the hydrogen-bonded pair can accommodate only 4–5 water molecules. In our previous molecular dynamics studies devoted to hydration of base pairs,<sup>26</sup> we showed that a gradual increase in the hydration number of the adenine–thymine base pair can change the base pair orientation. The bare AT pair and the monohydrated pair prefer the planar orientation while in the presence of more water molecules the stacked structures dominated. We also showed<sup>26</sup> that the stacked methylated base pairs, except guanine–cytosine pair, are favored over planar ones even without the presence of water molecules.

The gas-phase interactions of DNA bases were also studied experimentally. However, despite enormous progress in various experimental techniques, relevant experiments determining the relative importance of various energy contributions without the presence of perturbing environmental effects are still rare. This is due to the fact that such experiments are very difficult to perform. Let us mention, for example, the state-of-the-art gas-phase experiments from the laboratories of de Vries and Kleiner<sup>27–31</sup> They obtained unique IR spectra of the selected DNA base pairs using resonance-enhanced multiphoton ionization (REMPI) spectroscopy. The first study of microhydrated base pairs<sup>32</sup> appeared recently. It is important to note, however, that these studies do not allow information on structure and stabilization energy to be extracted. However, the theoretical quantum chemical studies give a direct insight into the physical-chemical origin of the particular interactions and provide, in principle, a full description of the base pairs, including geometry, stabilization energy, electric properties, IR, visible, and UV spectra, selected NMR parameters, etc. However, to obtain reliable data, the calculations must be carried out at a very high level of quantum chemical *ab initio* theory, which minimizes problems with the size of the basis set or the incompleteness of the inclusion of electron correlation effects.<sup>33</sup> These calculations thus represent an important complement to the experimental techniques. It is evident that a close cooperation between the experiment and the theory is essential for a better understanding the function of the DNA itself as well as of the DNA complexes with ions, water, drugs, etc.

An inclusion of electron correlation energy is necessary to obtain reliable data for the various interaction types mentioned above. The first medium-quality correlated *ab initio* calculations for DNA base pairs were performed in the mid-1990s and represented a real breakthrough in the field.<sup>34</sup> Calculations performed with the MP2 method using a medium-sized basis set containing diffuse polarization functions provided significant results for the hydrogen bonding and for the stacking of DNA bases. This method can be sufficiently used for the calculation of the interaction of the DNA bases and the base pairs with water and the metal cations as well. Literature covering the topics can be found in the reviews by Hobza and Šponer.<sup>35,36</sup> An equally important outcome of these calculations was a verification of the quality of several empirical potentials used for molecular modeling that highlighted the excellent performance of the Cornell et al. force field.<sup>37</sup>

Although the performance of computers has increased dramatically for the last several years, routine calculations at the MP2 level still remain impractical. Less expensive Hartree–Fock and DFT correctly describe only the hydrogen-bonded interactions, but they are not able to describe the equally important stacking interactions. Surprisingly, up to now there has been no reliable DFT functional describing the stacking of DNA bases. However, the addition of a term describing dispersion energy seems promising.<sup>38</sup> Recently published calculations of the stacked dimer of the methylated adenine–thymine pairs<sup>39a</sup> have given reasonable results. However, further verifications of this method are necessary. The MP2 level still remains the first acceptable level for the correct description of both hydrogen bonding and stacking interactions with at least a DZ+P basis set. This level provides accurate geometry characteristics, stabilization energies, and other physical-chemical properties of hydrogen-bonded pairs. The only inconsistency is that it underestimates the stabilization energies for the stacking. Reliable values of the stabilization energies of the stacked DNA base pairs were obtained using the MP2/TZ+P level of theory with a basis set containing two sets of first polarization functions and one set of second polarization functions on all the atoms.<sup>39b</sup>

As mentioned, the use of DZ+P or even TZ+P basis sets in combination with the exact MP2 treatment for the DNA base pairs is still difficult if not impractical. Recently, we explored<sup>40</sup> the applicability of the approximate resolution of identity (RI-MP2) method<sup>41,42</sup> for the nucleic acid base pairs and the larger DNA fragments, and we showed that this method is capable of an accurate description of the hydrogen-bonded and stacked DNA base interactions. The results obtained with the RI-MP2 method differ only slightly<sup>40</sup> from those evaluated with the exact MP2 method, while the time saved is as large as 1 order of magnitude.

The aim of the present paper is to characterize mono- and dihydrated complexes of adenine–thymine and 9-methyladenine–1-methylthymine base pairs using the molecular dynamics/quenching technique and the high-level correlated *ab initio* calculations. The latter calculations were employed for the determination of optimized structures and total stabilization energies of all complexes studied.

The geometries of all complexes were calculated at the RI-MP2/DZ+P level and the energies at the RI-MP2/TZ+P level of theory.

## Methods

### 1. Molecular Dynamics/Quenching (MD/Q) Technique.

The potential energy surface (PES) of the monohydrated and dihydrated adenine–thymine as well as the 9-methyladenine–1-methylthymine pairs were investigated using the MD/Q method with the Cornell et al. empirical force field<sup>37</sup> using the standard parameters. We performed all MD/Q simulations in the *NVE* microcanonical ensemble (*N*, *V*, and *E* refer to the amount of particles, volume, and energy, respectively) within the quaternion formalism. The constant total energy corresponded to an average temperature of 298 K, which is high enough to go over relatively high energy barriers between the minima at the potential energy surface.<sup>26a</sup> Every 1 ps MD run was interrupted, the kinetic energy was removed, and the structure of the complex was optimized using the conjugate gradient method. The geometry and energy were stored, while the MD run was restarted from the point where it had been interrupted. Finally, the populations of the various complex structures (related to the  $\Delta G$  of the complex formation) were

determined. The total length of the simulation was 250 ns, which is enough to sample the whole potential energy surface.

**2. Ab Initio Calculations.** After all of the energy minima were located using the MD/Q method, the optimization of the most stable and most populated structures was performed at the high-level correlated ab initio quantum chemical level. We know from our previous studies<sup>43,44</sup> that the Cornell et al. empirical force field describes the hydrogen-bonded and stacked structures of all DNA base pairs at least qualitatively well, so it is thus possible to limit the extremely time-consuming ab initio optimization only to the most stable and populated structures found by the empirical force field.

The most stable and most populated structures of the complexes were further fully reoptimized (conjugate gradient) at the RI-MP2 level using a DZ+P (cc-pVDZ [3s2p1d/2s1p]) basis set with a standard (default) auxiliary basis set. To avoid location of the transition state characteristics by the planar amino group, we changed the geometry before the optimization of the amino group of adenine (corresponding to a minimum in the Cornell et al. force field) to a slightly pyramidal arrangement (minimum at the RI-MP2 level). The stabilization energies were calculated at the RI-MP2/cc-pVDZ and RI-MP2/TZVPP [5s3p2d1f/3s2p1d]/RI-MP2/cc-pVDZ levels of theory. All ab initio calculations were carried out using the Turbomole 5.6 package.<sup>45</sup> In the present paper, we did not search for the transition states.

**3. Separation of Structures into Classes Based on Orientation of Bases.** Three basic base pair arrangements exist in the gas phase: hydrogen-bonded (HB), T-shaped (T), and stacked (S).<sup>34–36</sup> Both HB and S structures can exist in the planar or slightly nonplanar arrangement, while in the T arrangement the bases are oriented almost perpendicularly.

For the identification of structures found, we used our own script combined with the program for trajectory analysis “carnal” (part of AMBER program package). As we have shown previously,<sup>26a</sup> a distance between the centers of masses of the bases can be used for the separation of the HB and S structures. For the HB structures, this distance is about 6 Å, while for the S structures this distance is significantly smaller, about 3 Å. Besides the intermolecular distances, a deviation between the planes of bases was also considered for the more accurate separation of structural types. Finally, six structural types were considered (for details, see Table 1aS of the Supporting Information): (i) planar, hydrogen-bonded, (ii) nonplanar hydrogen-bonded, (iii) T-shaped, (iv) planar stacked, (v) nonplanar stacked, and (vi) structures without any close contact between the bases.

## Results

**1. Potential and Free Energy Surfaces of Hydrated Base Pairs Obtained by the MD/Q Technique.** Table 1 shows a total number of different structural types and their populations. Tables 1a–dS of the Supporting Information present the geometrical properties of the most stable structures of the most stable motifs, which were further reoptimized by ab initio calculations. Figures 1a–eS, 2a–eS, 3a–eS, and 4a–eS of the Supporting Information depict schematic drawings of the first 12 most stable binding motifs with the most preferred position of the water molecule obtained by the MD/Q technique. Figures 5a–dS of the Supporting Information show the 12 most energetically favorable water-binding sites for the most stable planar hydrogen-bonded pattern of the bases.

The PES of monohydrated AT base pair (AT-1W) was mapped by 337 structures. Among the 20 most stable structures

**TABLE 1: Orientation of Bases, Number of Structures, and Total Populations of Structures in a Given Class of Mono- and Dihydrated Adenine–Thymine and 9-Methyladenine–1-Methylthymine Complexes Obtained by the MD/Q Technique Employing the Cornell et al. Force Field<sup>37</sup>**

orientation of bases <sup>a</sup>	number of structures <sup>b</sup>	population (%) <sup>c</sup>	orientation of bases <sup>a</sup>	number of structures <sup>b</sup>	population (%) <sup>c</sup>
AT-1W			mAmT-1W		
p	67	30.9	p	33	5.7
np	38	10.8	np	31	3.4
t	115	30.2	t	86	8.1
ns	20	2.1	ns	18	4.8
s	67	25.6	s	96	77.9
separated	30	0.4	separated	9	0.1
total	337	100	total	273	100
AT-2W			mAmT-2W		
p	281	33.1	p	102	4.1
np	259	10.1	np	138	3.8
t	671	23.8	t	435	11.6
ns	106	2.2	ns	110	5.9
s	402	30.6	s	614	74.5
separated	26	0.2	separated	19	0.1
total	1745	100	total	1418	100

<sup>a</sup> Classes differ by the mutual orientation of bases and the distances between the centers of mass of the bases: p, planar hydrogen-bonded; np, nonplanar hydrogen-bonded; t, T-shaped; ns, nonplanar stacked; s, stacked; separated, in this class the distance between the bases or between water molecule(s) and bases is more than 10 Å. For detailed information about separation into classes, see the Supporting Information. <sup>b</sup> Total number of structures obtained by the MD/Q technique in the given class. <sup>c</sup> Total population of all structures of a given class obtained by the MD/Q technique.

(differing in their stabilization energies by less than 3 kcal/mol), 12 planar HB and 5 planar S structures were found. For the dihydrated AT base pair (AT-2W), as much as 1745 energetically different structures were localized. Of the 20 most stable structures (with difference in stabilization energy less than 0.5 kcal/mol), 17 were planar S structures. For a monohydrated 9-methyladenine–1-methylthymine base pair (mAmT-1W), we localized 273 structures. Only 1 structure out of the 20 most stable structures was not in an S arrangement. The energy difference between the 1st and the 20th structure was almost 4 kcal/mol. Of the dihydrated 9-methyladenine–1-methylthymine base pair (mAmT-2W), 1418 structures were located. All of the first 50 most stable structures were S ones. The difference in stabilization energy between the 1st and 50th structure was less than 2 kcal/mol.

Besides the number of structures localized, we also evaluated their populations. For the AT-1W complexes, the most populated structures were the HB and T ones (30.9% and 30.2%, respectively).

In the free energy surface (FES) of AT-2W, the planar HB structures (33.1%) were also the most populated structures, followed by the planar S ones with almost the same abundance (30.6%). For stacked structures, an increase in the population by about 5% in comparison with AT-1W was observed.

The planar S structures are unambiguously dominant in the whole FESs of mAmT-1W and mAmT-2W complexes (77.9% and 74.5% respectively). The other types of structures presented only sporadically (cf. Table 1). Scans of the PESs and FESs show an unambiguous tendency to prefer S structures with an increasing number of water molecules and methylation.

**2. General Characteristic of Structural Motifs. Hydrogen-Bonded Structures.** In the most stable nonmethylated structures, the bases interact together with two strong hydrogen bonds. The



water molecule(s) lie usually in the plane of the bases. Both RI-MP2/cc-pVDZ and RI-MP2/TZVPP/RI-MP2/cc-pVDZ calculations show that the global minimum of the PES of the AT-1W complex corresponds to the planar HB structure HB1 (cf. Figure 2a) with the water molecule attached to the N6 hydrogen and N7 atoms of adenine. The structure of the global minimum found at the empirical PES was the same. The canonical Watson–Crick structure corresponds to the eighth local minimum at the AT-1W PES, and it is less stable than the global minimum by about 5.5 kcal/mol.

For the dihydrated AT pair, the most stable HB structure HB1 (cf. Figure 2b) corresponds to the first local minimum, while the global minimum corresponds to an S structure. The water molecules in this structure are in close contact between each other, and they also interact with the O2 and N1 atoms of thymine. Generally, structures with two water molecules interacting together are more favorable than those with separated water molecules. A similar situation has been found for isolated bases.<sup>15</sup> Methylated base pairs have less favorable donors or acceptors for binding water molecules; hence, the structures where water molecule(s) are intercalated between the bases become energetically comparable with structures with direct hydrogen bond contact between the bases. The most stable HB structures correspond to the Watson–Crick structure at the PES of mAmT-1W (cf. structure HB1 in Figure 2c) and the reverse Hoogsteen structure at the PES of mAmT-2W (cf. structure HB1 in Figure 2d). Neither of these two is the global minimum at the PES.

For the nonplanar HB structures, an even higher abundance of the water bridge clusters was observed, and the structures with two direct hydrogen bonds between the bases are observed only rarely. The nonplanar HB structures are usually less stable than the planar ones.

**T-Shaped Structures.** We found almost equal populations of the structures with two direct hydrogen bonds between the bases and structures with one hydrogen bond and one water bridge between the bases. The T arrangement was never located as the global minimum of any PES for any complex, and the most stable T structures are approximately less stable by 10 kcal/mol than the structure of the global minimum.

**Stacked Structures.** The water molecule(s) usually create the “clip” (see, for example, the structure S1 in the Figure 2a) between the bases, which means that the water molecules only rarely occur in the plane of the bases. Two possible arrangements of the “clip” exist in the presence of two water molecules. A motif with two water molecules joined together (cf. structure S1 in Figure 2b) is more stable than a pattern where two water molecules lie separately on the opposite sides of the bases (cf. Figure 2b, structure S2). The S structure was found as a global minimum at all PESs, except of AT-1W, where the most stable stacked structure corresponds to the fourth local minimum and it is 4.3 kcal/mol less stable than the global HB minimum.

For all most stable S structures of all PESs, the thymine base rotated by approximately 230° with respect to the adenine molecule (for a more exact definition of the twist angle of rotation see the Supporting Information and ref 47). Such an orientation of stacked bases has been already observed for bare adenine–uracil<sup>47</sup> and adenine–thymine<sup>46</sup> base pairs.

**3. Comparison of Ab Initio and Empirical Calculations.** The correlated ab initio calculations were performed for the 12 most stable binding motifs of adenine–thymine and 9-methyladenine–1-methylthymine complexes having water molecule(s) in the most energetically preferred positions. Most of these structures were planar HB or stacked (cf. Figures 2a–d).

**TABLE 2: Empirical and Ab Initio Interaction Energies of Selected Structures of the Monohydrated Adenine–Thymine Complex**

motif <sup>a</sup>	$\Delta E_{\text{AMBER}}^b$	$\Delta E_{\text{def}}^c$	$\Delta E_{\text{int}}^{\text{DZ}}^d$	$\Delta E_{\text{tot}}^{\text{DZ}}^e$	$\Delta E_{\text{int}}^{\text{TZ}}^f$	$\Delta E_{\text{tot}}^{\text{TZ}}^g$
HB1	−18.98	3.09	−27.19	−24.10	−31.14	−28.05
HB2	−16.87	2.48	−25.19	−22.70	−28.71	−26.23
HB3	−15.12	2.73	−24.72	−21.98	−28.05	−25.31
HB4	−16.08	3.02	−24.07	−21.05	−27.44	−24.42
S1	−17.75	1.99	−18.15	−16.17	−25.70	−23.71
S2	−16.05	2.53	−19.22	−16.69	−25.92	−23.38
HB5	−15.98	4.01	−23.20	−19.20	−27.33	−23.33
S3	−17.61	1.68	−17.99	−16.31	−24.80	−23.12
HB6, WC <sup>h</sup>	−15.35	4.05	−22.49	−18.45	−26.59	−22.54
HB7	−15.42	4.05	−22.29	−18.25	−26.32	−22.27
HB8	−17.51	7.06	−24.55	−17.50	−28.81	−21.75
T1	−14.45	2.29	−16.37	−14.08	−21.98	−19.70

<sup>a</sup> Binding motifs between the bases: HB corresponds to hydrogen-bonded structures, T to T-shaped ones, and S to stacked ones. Structures are ordered by the total interaction energies obtained at the RI-MP2/TZVPP/RI-MP2/cc-pVDZ level of calculation (last column). <sup>b</sup> Empirical interaction energy in kcal/mol calculated as a difference of the energy of the complexes and subunits calculated using the Cornell et al. force field.<sup>37</sup> <sup>c</sup> Total deformation energy in kcal/mol obtained as a sum of the deformation energies of adenine, thymine, and a water molecule calculated at the RI-MP2/cc-pVDZ level. <sup>d</sup> Basis-set superposition error (BSSE)-corrected interaction energies obtained at the RI-MP2/cc-pVDZ level (in kcal/mol) calculated by following equations:  $E_{\text{int}} = E_{\text{AT-1W}} - (E_{\text{A}} + E_{\text{T}} + E_{\text{W}})$  for monohydrated systems and  $E_{\text{int}} = E_{\text{AT-2W}} - (E_{\text{A}} + E_{\text{T}} + E_{\text{W1}} + E_{\text{W2}})$  for dihydrated complexes. <sup>e</sup> Total interaction energy obtained as the sum of the deformation and interaction energy obtained at the RI-MP2/cc-pVDZ level. <sup>f</sup> BSSE-corrected interaction energies calculated at the RI-MP2/TZVPP/RI-MP2/cc-pVDZ level (in kcal/mol) obtained in the same way as in footnote d. <sup>g</sup> Total interaction energy in kcal/mol obtained as the sum of the deformation energy obtained at the RI-MP2/cc-pVDZ level and an interaction energy obtained at the RI-MP2/TZVPP/RI-MP2/cc-pVDZ level. <sup>h</sup> Watson–Crick structure corresponding to the base pair arrangement in DNA.

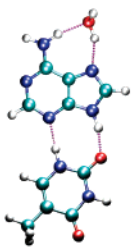
The geometries of the ab initio optimized structures are mostly similar to geometries obtained from the MD/Q simulation (cf. Tables 1a–dS of the Supporting Information). The root-mean-square deviations between them are usually smaller than 0.3 Å. The only exceptions were planar HB structures containing a water bridge between the bases. Several of them changed the mutual orientation of bases to T or S arrangements during the ab initio optimization.

The deviations of the base planes from planarity calculated at the ab initio level are slightly higher than at the empirical level. This difference can be explained by the fact that the Cornell et al. force field describes the amino group of adenine as a planar system, while the correlated ab initio optimization provides for amino group hydrogen atoms a significant deviation from planarity.<sup>48</sup> The deviations of the amino group’s hydrogen atoms from planarity, calculated at the ab initio level, are generally higher for S complexes than for the HB complexes.

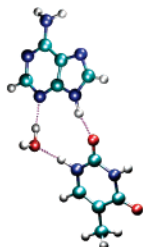
Furthermore, in the empirical potential geometries, all atoms of the water molecules in the planar HB structures lie in the base plane, while in the geometries obtained from the ab initio calculations the water hydrogen atoms rotated outside the planes of the bases and form a spatial net of hydrogen bonds around the bases.

Although the cc-pVDZ basis set gives the reasonable geometries for the S and HB complexes, it is still not sufficiently accurate for obtaining the exact stabilization energies, which can be received by extrapolating to the complete basis set (CBS) limit. This fact is demonstrated in Tables 2–5. These tables show that the stabilization energies obtained from the ab initio calculations with cc-pVDZ are lower by 5 kcal/mol than these

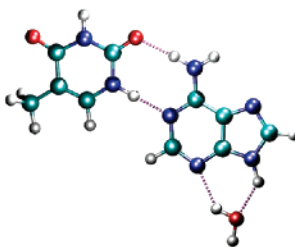
(a)



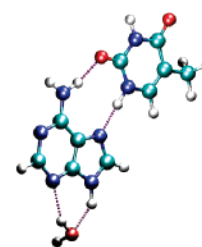
HB1 -28.05 / -18.98



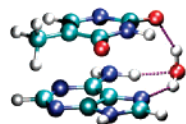
HB2 -26.23 / -16.87



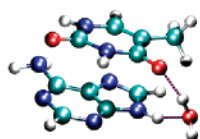
HB3 -25.31 / -15.12



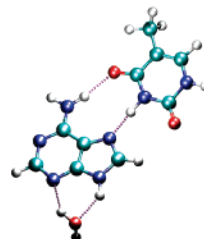
HB4 -24.42 / -16.08



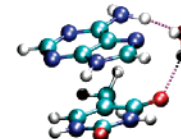
S1 -23.71 / -17.75



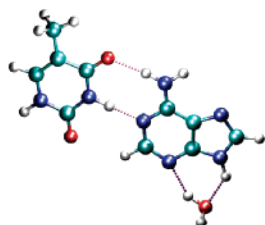
S2 -23.38 / -16.05



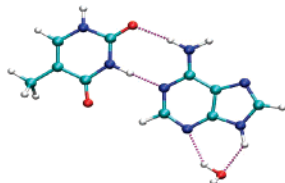
HB5 -23.33 / -15.98



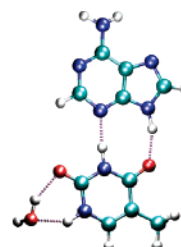
S3 -23.12 / -17.61



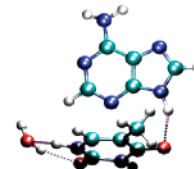
HB6 -22.54 / -15.35



HB7 -22.27 / -15.42

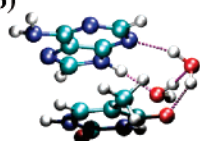


HB8 -21.75 / -17.51

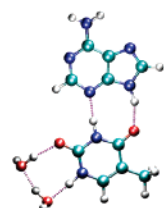


T1 -19.70 / -14.45

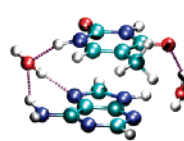
(b)



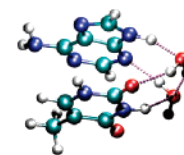
S1 -37.20 / -32.15



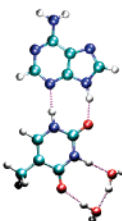
HB1 -36.56 / -32.33



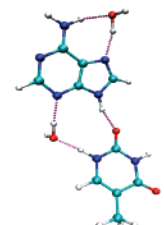
S2 -36.40 / -25.12



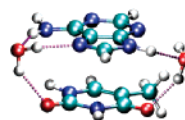
S3 -36.22 / -28.12



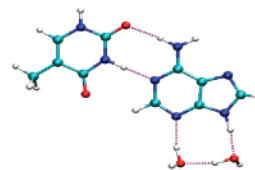
HB2 -36.09 / -30.82



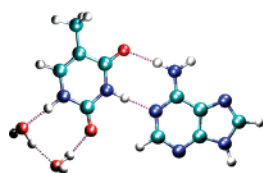
HB3 -36.04 / -29.97



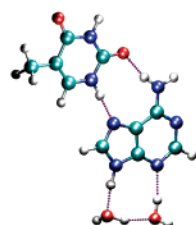
S4 -35.61 / -32.31



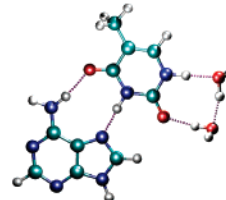
HB4 -35.40 / -29.85



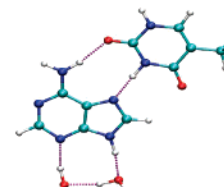
HB5 -34.98 / -26.42



HB6 -34.96 / -32.33

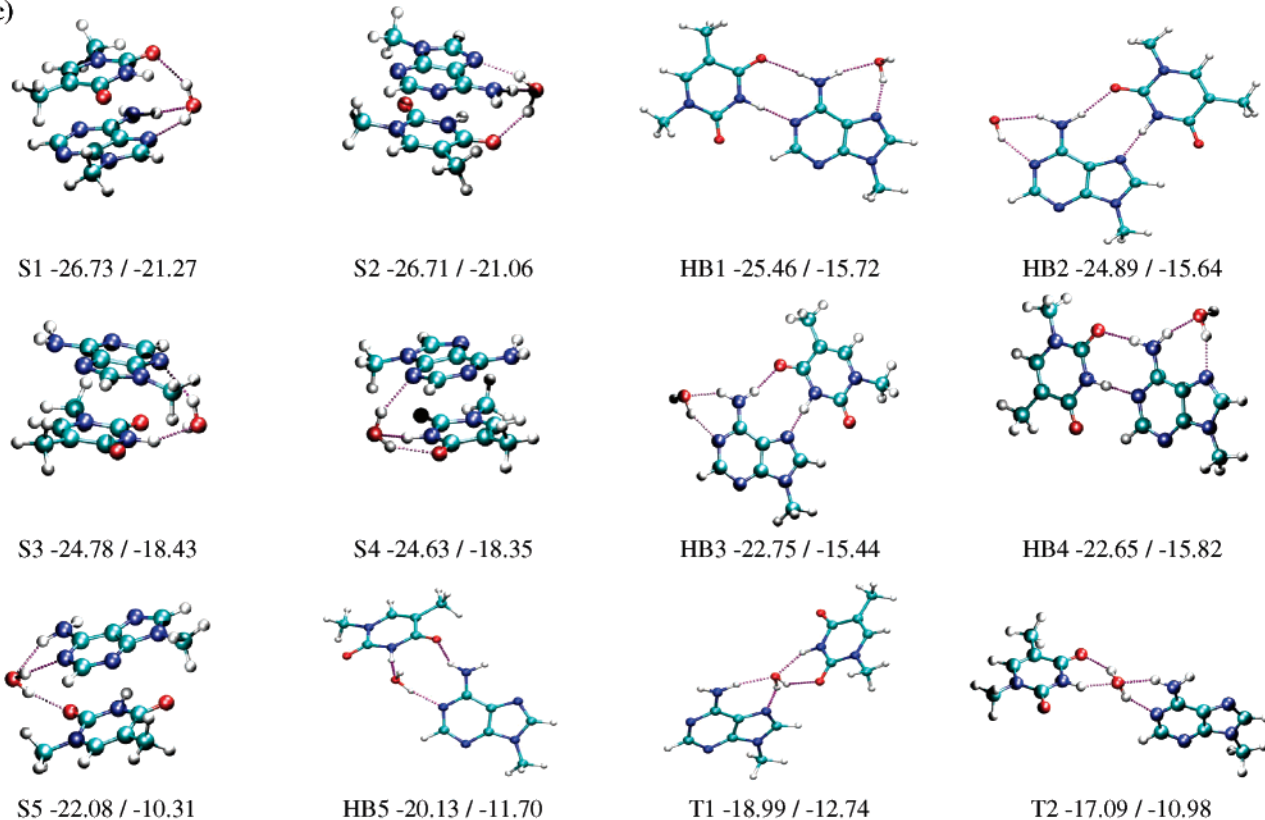


HB7 -34.23 / -30.36

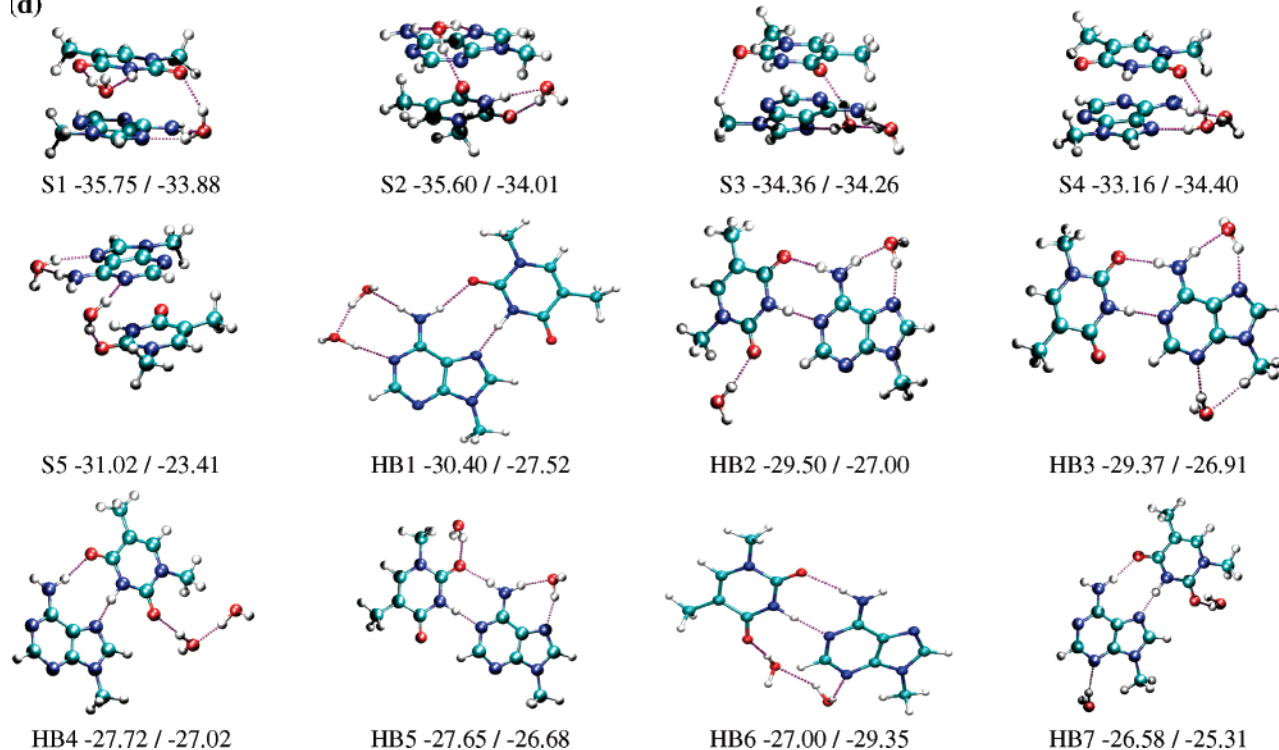


HB8 -34.05 / -30.12

(c)



(d)



**Figure 2.** (a) Monohydrated adenine–thymine complex. (b) Dihydrated adenine–thymine complex. (c) Monohydrated 9-methyladenine–1-methylthymine complex. (d) Dihydrated 9-methyladenine–1-methylthymine complex. Ball-and-stick representations of 12 structures optimized at the ab initio RI-MP2/cc-pVDZ level. The description below the structures shows the binding motif between the bases (HB corresponds to hydrogen-bonded structures, T to T-shaped ones, and S to stacked ones), the interaction energy in kcal/mol (deformation energy included) obtained at the RI-MP2/TZVPP//RI-MP2/cc-pVDZ level, and the empirical interaction energy in kcal/mol obtained using the Cornell et al. force field.<sup>37</sup> Oxygen atoms are colored red, nitrogen atoms are blue, carbon atoms are green, and hydrogen atoms are white. The hydrogen bonds are depicted using dashed purple lines. For an easier comparison of the different structures, the position of adenine is kept in the same orientation except in the structures where such an orientation does not allow a clear view of the spatial arrangement of the bases.



**TABLE 3: Empirical and Ab Initio Interaction Energies of Selected Structures of the Dihydrated Adenine–Thymine Complex**

motif <sup>a</sup>	$\Delta E_{\text{AMBER}}^b$	$\Delta E_{\text{def}}^c$	$\Delta E_{\text{int DZ}}^d$	$\Delta E_{\text{tot DZ}}^e$	$\Delta E_{\text{int TZ}}^f$	$\Delta E_{\text{tot TZ}}^g$
S1	−32.15	3.35	−31.01	−27.66	−40.55	−37.20
HB1	−32.33	3.34	−34.30	−30.96	−39.90	−36.56
S2 <sup>i</sup>	−25.12	3.01	−30.11	−27.10	−39.41	−36.40
S3 <sup>i</sup>	−28.12	3.18	−29.39	−26.21	−39.40	−36.22
HB2	−30.82	3.34	−34.16	−30.82	−39.43	−36.09
HB3	−29.97	3.78	−34.54	−30.76	−39.82	−36.04
S4	−32.31	2.96	−28.45	−25.48	−38.57	−35.61
HB4	−29.85	3.36	−33.40	−30.04	−38.76	−35.40
HB5, WC <sup>h,j</sup>	−26.42	2.67	−32.05	−29.38	−37.66	−34.98
HB6	−32.33	5.27	−35.48	−30.20	−40.23	−34.96
HB7	−30.36	3.77	−32.39	−28.61	−38.00	−34.23
HB8	−30.12	4.87	−33.62	−28.75	−38.92	−34.05

<sup>a–h</sup> For an explanation of the abbreviations, see the footnotes to Table 2. <sup>i</sup> The arrangement of the structure was changed during the optimization from planar to stacked. <sup>j</sup> The arrangement of the structure was changed during the optimization from T-shaped to planar.

**TABLE 4: Empirical and Ab Initio Interaction Energies of Selected Structures of the Monohydrated 9-Methyladenine–1-Methylthymine Complex**

motif <sup>a</sup>	$\Delta E_{\text{AMBER}}^b$	$\Delta E_{\text{def}}^c$	$\Delta E_{\text{int DZ}}^d$	$\Delta E_{\text{tot DZ}}^e$	$\Delta E_{\text{int TZ}}^f$	$\Delta E_{\text{tot TZ}}^g$
S1	−21.27	2.00	−28.73	−15.95	−28.73	−26.73
S2	−21.06	1.65	−28.36	−18.08	−28.36	−26.71
HB1, WC <sup>h</sup>	−15.72	2.69	−21.43	−18.74	−25.62	−25.46
HB2	−15.64	2.39	−20.80	−18.41	−25.04	−24.89
S3	−18.43	1.33	−26.11	−15.76	−26.11	−24.78
S4	−18.35	1.25	−25.88	−15.64	−25.88	−24.63
HB3	−15.44	2.44	−25.20	−18.47	−25.20	−22.75
HB4	−15.82	2.63	−25.29	−18.50	−25.29	−22.65
S5	−10.31	0.93	−23.01	−13.90	−23.01	−22.08
HB5	−11.70	1.74	−17.21	−15.47	−20.46	−20.13
T1 <sup>i</sup>	−12.74	2.96	−16.04	−13.08	−19.41	−18.99
T2 <sup>i</sup>	−10.98	1.19	−18.28	−13.61	−18.28	−17.09

<sup>a–h</sup> For an explanation of the abbreviations, see the footnotes to Table 2. <sup>i</sup> The arrangement of the structure was changed during the optimization from planar to T-shaped.

**TABLE 5: Empirical and Ab Initio Interaction Energies of Selected Structures of the Dihydrated 9-Methyladenine–1-Methylthymine Complex**

mAmT-2W <sup>a</sup>	$\Delta E_{\text{AMBER}}^b$	$\Delta E_{\text{def}}^c$	$\Delta E_{\text{int DZ}}^d$	$\Delta E_{\text{tot DZ}}^e$	$\Delta E_{\text{int TZ}}^f$	$\Delta E_{\text{tot TZ}}^g$
S1	−33.88	2.46	−38.21	−24.65	−38.21	−35.75
S2	−34.01	1.96	−37.56	−24.64	−37.56	−35.60
S3	−34.26	2.38	−36.75	−24.52	−36.75	−34.36
S4	−34.40	3.49	−36.65	−26.92	−36.65	−33.16
S5 <sup>i</sup>	−23.41	2.11	−33.13	−21.13	−33.13	−31.02
HB1	−27.52	4.39	−29.28	−24.89	−34.79	−30.40
HB2, WC <sup>h</sup>	−27.00	2.90	−32.40	−23.64	−32.40	−29.50
HB3	−26.91	2.82	−32.19	−23.75	−32.19	−29.37
HB4	−27.02	2.30	−30.02	−22.16	−30.02	−27.72
HB5	−26.68	3.10	−30.75	−22.10	−30.75	−27.65
HB6	−29.35	2.21	−22.73	−20.51	−29.21	−27.00
HB7	−25.31	2.49	−29.07	−20.92	−29.07	−26.58

<sup>a–i</sup> For an explanation of the abbreviations, see the footnotes to Tables 2 and 3.

obtained with the TZVPP basis set for all HB systems studied. For the S structures, these differences are even higher, by 7 kcal/mol for the monohydrated adenine–thymine and about 10 kcal/mol in all other cases. Evidently, the higher polarization functions are required for getting the converged values of the stabilization energy. Furthermore, while the cc-pVDZ basis set yields relatively reasonable stabilization energies for the HB systems, this basis set (lacking diffuse polarization functions) is less successful in the description of the S complexes.

Comparing the empirical stabilization energies with the ab initio ones, we found that the empirical values reproduce an ab initio stability order fairly well. From the quantitative point of view, the empirical values are close to the DZ ab initio values, which means that they are still underestimated in comparison to the CBS limit data. The remarkable accuracy of the potential, at least from a qualitative point of view, should be stressed here. The computer time for the empirical calculation is negligible, while the times for DZ, and even more so for TZ, calculations are quite substantial.

## Conclusions

Using the MD/Q technique, we located more than 300 different structures for the monohydrated pairs and almost 2000 structures for the dihydrated pairs of adenine–thymine and their methylated derivatives. It was by one or two orders more than we explored for the bare (nonhydrated) base pairs. The differences between the stability of the binding motifs of these complexes are small. The first 50 most stable structures have stability in a narrow range around 5 kcal/mol, which indicates that it will be experimentally difficult to distinguish between them.

Both ab initio and empirical calculations show that the presence of one water molecule is insufficient to change the adenine–thymine HB structures to the S ones. However, if two water molecules are present, then both arrangements have comparable stability. In the presence of more water molecules, the preference for the S structures can be expected. The situation is different for the methylated base pair. We have shown previously that the bare base pair prefers S structures over the HB ones. Monohydration and dihydration leads to the dominant preference of the S structures, since the replacement of hydrogen by a methyl group eliminates the most stable and the most populated HB structures observed for the nonmethylated pairs. Furthermore, the denser network of hydrogen bonds between water and bases present in the S arrangement supports its preference.

**Acknowledgment.** This work was a part of the research project Z40550506 of the Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, and it was supported by grants LC512 (Ministry of Education of the Czech Republic), 203/05/009 (P.H.), and KJB400550518 (M.K.) from the Grant Agency of the Czech Republic. We are grateful to Dr. Jiří Vondrášek for stimulating discussions and careful reading of the manuscript.

**Supporting Information Available:** Geometrical properties of the most stable structures of the most stable motifs, schematic drawings of the first 12 most stable binding motifs with the most preferred position of the water molecule obtained by the MD/Q technique, and the 12 most energetically favorable water-binding sites for the most stable planar hydrogen-bonded pattern of the bases. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## References and Notes

- (1) Stryer, L. *Biochemistry*; W. H. Freeman and Company: New York, 1999.
- (2) Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag: New York, 1983.
- (3) Chalikian, T. V.; Sarvazyan, A. P.; Plum, G. E.; Breslauer, K. J. *Biochemistry* **1994**, *33*, 2394.
- (4) Schneider, B.; Cohen, D.; Berman, H. M. *Biopolymers* **1992**, *32*, 725.

- (5) Schneider, B.; Cohen, D. M.; Schleifer, L.; Srinivasan, A. R.; Olson, W. K.; Berman, H. M. *Biophys. J.* **1993**, *65*, 2291.
- (6) Schneider, B.; Berman, H. M. *Biophys. J.* **1995**, *69*, 2661.
- (7) van Mourik, T.; Price, S. L.; Clary, D. C. *J. Phys. Chem. A* **1999**, *103*, 1611.
- (8) van Mourik, T.; Benoit, D. M.; Price, S. L.; Clary, D. C. *Phys. Chem. Chem. Phys.* **2000**, *2*, 1281.
- (9) van Mourik, T. *Phys. Chem. Chem. Phys.* **2001**, *3*, 2886.
- (10) Kryachko, E.; Nguyen, M. T.; Zeegers-Huyskens, T. *J. Phys. Chem. A* **2001**, *105*, 3379.
- (11) Aamouche, A.; Berthier, G.; Cadioli, B.; Gallinella, E.; Ghomi, M. *THEOCHEM* **1998**, *426*, 307.
- (12) Ghomi, M.; Aamouche, A.; Cadioli, B.; Berthier, G.; Grajcar, L.; Baron, M. H. *J. Mol. Struct.* **1997**, *410*, 323.
- (13) Aleman, C. *Chem. Phys. Lett.* **1999**, *302*, 461.
- (14) Chandra, A. K.; Nguyen, M. T.; Zeegers-Huyskens, T. *J. Phys. Chem. A* **1998**, *102*, 6010.
- (15) (a) Hanus, M.; Kabeláč, M.; Rejnek, J.; Ryjáček, F.; Hobza, P. *J. Phys. Chem. B* **2004**, *108*, 2087. (b) Hanus, M.; Ryjáček, F.; Kabeláč, M.; Hobza, P. *J. Am. Chem. Soc.* **2003**, *125*, 7678. (c) Trygubenko, S. A.; Bogdan, T. V.; Rueda, M.; Hobza, P. *Phys. Chem. Chem. Phys.* **2002**, *4*, 4192. (d) Rejnek, J.; Kabeláč, M.; Hanus, M.; Hobza, P. *Phys. Chem. Chem. Phys.* **2005**, *7*, 2006.
- (16) Desfrancois, C.; Carles, S.; Schermann, J. P. *Chem. Rev.* **2000**, *100*, 3943.
- (17) Sukhanov, O. S.; Shishkin, O. V.; Gorb, L.; Podolyan, Y.; Leszczynski, J. *J. Phys. Chem. B* **2003**, *107*, 2846.
- (18) Gu, J. D.; Leszczynski, J. *J. Phys. Chem. A* **1999**, *103*, 2744.
- (19) Chandra, A. K.; Nguyen, M. T.; Uchimaru, T.; Zeegers-Huyskens, T. *J. Phys. Chem. A* **1999**, *103*, 8853.
- (20) Schweizer, D. M. B.; Broom, A. D.; Ts'o, P. O. P.; Hollis, D. P. *J. Am. Chem. Soc.* **1967**, *90*, 1042.
- (21) Kyogoku, Y.; Lord, R. C.; Rich, A. *J. Am. Chem. Soc.* **1967**, *89*, 496.
- (22) Binford, J.; Holloway, D. M. *J. Mol. Biol.* **1968**, *31*, 91.
- (23) Yanson, I. K.; Teplitsky, A. B.; Sukhodub, L. F. *Biopolymers* **1979**, *18*, 1149.
- (24) Cieplak, P.; Kollman, P. A. *J. Am. Chem. Soc.* **1988**, *110*, 3734.
- (25) Sivanesan, D.; Sumathi, I.; Welsh, W. J. *Chem. Phys. Lett.* **2003**, *367*, 351.
- (26) (a) Kabeláč, M.; Hobza, P. *Chem.—Eur. J.* **2001**, *7*, 2067. (b) Kabeláč, M.; Ryjáček, F.; Hobza, P. *Phys. Chem. Chem. Phys.* **2000**, *2*, 4906.
- (27) Plutzer, C.; Hunig, I.; Kleinerann, K.; Nir, E.; de Vries, M. S. *ChemPhysChem* **2003**, *4*, 838.
- (28) Nir, E.; Grace, L.; Brauer, B.; de Vries, M. S. *J. Am. Chem. Soc.* **1999**, *121*, 4896.
- (29) Nir, E.; Plutzer, C.; Kleinerann, K.; de Vries, M. *Eur. Phys. J. D* **2003**, *20*, 317.
- (30) Nir, E.; Hunig, I.; Kleinerann, K.; de Vries, M. S. *Phys. Chem. Chem. Phys.* **2003**, *5*, 4780.
- (31) Plutzer, C.; Hunig, I.; Kleinerann, K. *Phys. Chem. Chem. Phys.* **2003**, *5*, 1158.
- (32) Abo-Riziq, A.; Crews, B.; Grace, L.; de Vries, M. S. *J. Am. Chem. Soc.* **2005**, *127*, 2374.
- (33) He, Y. G.; Wu, C. Y.; Kong, W. J. *Phys. Chem. A* **2004**, *108*, 943.
- (34) Šponer, J.; Leszczynski, J.; Hobza, P. *J. Phys. Chem.* **1996**, *100*, 5590.
- (35) Šponer, J.; Hobza, P. *Collect. Czech. Chem. Commun.* **2003**, *68*, 2231.
- (36) Hobza, P.; Šponer, J. *Chem. Rev.* **1999**, *99*, 3247.
- (37) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179.
- (38) Elstner, M.; Hobza, P.; Frauenheim, T.; Suhai, S.; Kaxiras, E. *J. Chem. Phys.* **2001**, *114*, 5149.
- (39) (a) Shishkin, O. V.; Elstner, M.; Frauenheim, T.; Suhai, S. *Int. J. Mol. Sci.* **2003**, *4*, 537. (b) Jurečka, P.; Hobza, P. *J. Am. Chem. Soc.* **2003**, *125*, 15608.
- (40) Jurečka, P.; Nachtigall, P.; Hobza, P. *Phys. Chem. Chem. Phys.* **2001**, *3*, 4578.
- (41) Feyereisen, M.; Fitzgerald, G.; Komornicki, A. *Chem. Phys. Lett.* **1993**, *208*, 359.
- (42) Weigend, F.; Haser, M. *Theor. Chem. Acc.* **1997**, *97*, 331.
- (43) Hobza, P.; Kabeláč, M.; Šponer, J.; Mejzlík, P.; Vondrášek, J. *J. Comput. Chem.* **1997**, *18*, 1136.
- (44) Hobza, P.; Hubálek, F.; Kabeláč, M.; Mejzlík, P.; Šponer, J.; Vondrášek, J. *Chem. Phys. Lett.* **1996**, *257*, 31.
- (45) Ahlrichs, R.; Bar, M.; Haser, M.; Horn, H.; Kolmel, C. *Chem. Phys. Lett.* **1989**, *162*, 165.
- (46) Kabeláč, M.; Hobza, P. *J. Phys. Chem. B* **2001**, *105*, 5804.
- (47) Florián, J.; Šponer, J.; Warshel, A. *J. Phys. Chem. B* **1999**, *103*, 884.
- (48) Šponer, J.; Hobza, P. *J. Phys. Chem.* **1994**, *98*, 3161.