DOI: 10.1002/chem.200902068

Theoretical Studies on the Intermolecular Interactions of Potentially Primordial Base-Pair Analogues

Judit E. Šponer,*[a] Álvaro Vázquez-Mayagoitia,[b] Bobby G. Sumpter,[c] Jerzy Leszczynski,[d] Jiří Šponer,*[a] Michal Otyepka,[e] Pavel Banáš,[e] and Miguel Fuentes-Cabrera*[f]

Abstract: Recent experimental studies on the Watson–Crick type base pairing of triazine and aminopyrimidine derivatives suggest that acid/base properties of the constituent bases might be related to the duplex stabilities measured in solution. Herein we use high-level quantum chemical calculations and molecular dynamics simulations to evaluate the base pairing and stacking interactions of seven selected base pairs, which are common in that they are stabilized by two N−H···O hydrogen bonds separated by one N−H···N hydrogen bond. We show that neither the base pairing nor the base stacking interaction energies correlate with the reported pK_a data of the bases and the melting points of the duplexes. This suggests that the experimentally observed correlation between the melting point data of the duplexes and the pK_a values of the constituent bases is not rooted in the intrinsic base pairing and stacking properties. The physical chemistry origin of the observed experimental correlation thus remains unexplained and requires further investigations. In addition, since our calculations are carried out with extrapolation to the complete basis set of atomic orbitals and with inclusion of higher electron correlation effects, they provide reference data for stacking and base pairing energies of non-natural bases.

Keywords: ab initio calculations · base pairing · molecular dynamics · origin of life · prebiotic chemistry

[a] Dr. J. E. Šponer, Prof. J. Šponer
Institute of Biophysics
Academy of Sciences of the Czech Republic
Královopolská 135, 61265, Brno, (Czech Republic)
Fax: (+420) 541-211-293
E-mail: judit@ncbr.chemi.muni.cz
sponer@ncbr.chemi.muni.cz

[b] Dr. Á. Vázquez-Mayagoitia
Chemistry Department, University of Tennessee
1416 Circle Drive, 552 Dabney-Buehler Hall
Knoxville, TN 37996-1600 (USA)

[c] Dr. B. G. Sumpter
Center for Nanophase Materials Sciences and
Computer Sciences and Mathematics Division
Oak Ridge National Laboratory, Oak Ridge
P. O. Box 2008, Oak Ridge, TN 37831-6494 (USA)

[d] Prof. J. Leszczynski
Department of Chemistry
Computational Center for Molecular Structure and Interactions
Jackson State University, Jackson, Mississippi 39217 (USA)

[e] Dr. M. Otyepka, Dr. P. Banáš
Department of Physical Chemistry, Faculty of Science
Palacky University Olomouc, tr. 17. Listopadu 12
77146 Olomouc (Czech Republic)

[f] Dr. M. Fuentes-Cabrera
Joint Institute for Computational Sciences, University of Tennessee
Center for Nanophase Materials Sciences and
Computer Sciences and Mathematics Division
Oak Ridge National Laboratory, Oak Ridge
P. O. Box 2008, Oak Ridge, TN 37831-6494 (USA)
Fax: (+1) 865-241-0381
E-mail: fuentescabma@ornl.gov

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/chem.200902068. It contains detailed descriptions of the computational methods and geometry definitions, Cartesian coordinates for all of the optimized base pairs; a table summarizing all HF/aug-cc-pVDZ, HF/aug-cc-pVTZ, RIMP2/aug-cc-pVDZ and RIMP2/aug-cc-pVTZ, MP2/CBS, and CBS(T) interaction energies, CCSD(T) correction terms, solvent corrections, deformation energies obtained at B3LYP/6-31G** level; a graph illustrating the variation of the RIMP2/aug-cc-pVDZ interaction energies as a function of the propeller twist and the vertical separation of the bases for the stacking of two $AP^{OO}\cdots T^{NN}$ base pairs; figures comparing the ring positions in the quantum chemical and MD stacking geometries, parameters used in the MD simulations; and averaged structures obtained from MD simulations.

 © 2010 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim WILEY InterScience DISCOVER SOMETHING GREAT

Introduction

Efforts aimed at understanding the evolution of the genetic alphabet have evoked intense experimental research on synthetic base pair analogues. Beyond the strategy used by nature to establish base pairing in nucleic acids (i.e., hydrogen bonding) some of these studies have used different principles to link together the bases.[1–4] Among them, shape-complementarity[3] and hydrophobic interactions[4] have been documented as a viable alternative of standard base pairing. It was shown that base pairs created in this way can be incorporated into DNA without significant structural distortion, and thus may give rise to an alternative genetic code.

A different strategy has been applied in the search of the origin of the first genetic material by Eschenmoser's group.[5a] Over the years they have synthesized several heterocycles, which can hydrogen bond not only with each other, but also with natural nucleobases.[5,6] These synthetic analogues posses very similar steric dimensions and hydrogen-bonding patterns as their natural counterparts. Thus, they can be considered as the closest relatives of the four nucleic acid bases occurring in the genetic alphabet of the living material.

Recently, Mittapalli et al. reported on the base pairing properties of triazines and 5-aminopyrimidines.[6] Using a PNA-like oligopeptide (PNA = peptide nucleic acids[7]) backbone they created homogenous sequences consisting of 6, 8, 12, and 16 identical consecutive bases in the strands (such sequences are also known as oligo- or poly-N tracts). They examined the stability of the duplexes formed between strands with complementary alternative bases as well as duplexes having complementary pairing between alternative and natural bases (both DNA and RNA strands were used). They used melting point data to assess the duplex stabilities and reported a qualitative correlation between the melting point and the differences of the pK_a values (hereafter denoted with ΔpK_a) of the interacting bases. They suggested that the duplex stability increases with increasing ΔpK_a. The maximum ΔpK_a has been found for the AU/AT and GC WC (Watson–Crick) base pairs, which led to the conclusion that nature selected the five naturally occurring nucleic acid bases because of their highly different pK_a values.

The backbone used by Mittapalli et al. represents a fully compatible steric substitute of common DNA/RNA backbones, and, in contrast to the charge neutral PNAs, it is built up of dipeptide subunits each carrying a net charge of -1. Therefore, one can expect that it exhibits very similar thermodynamics to naturally occurring nucleic acids. Note that it is well established that the thermodynamics of PNA and PNA/DNA hybrids is much more complex than that of DNA.[8] Nevertheless, it is important to point out that, in general, the basic physicochemical relationship between the melting point and duplex stabilities is rather complex; therefore, there is no reason to expect a clear correlation between them.[9] In addition, the duplex stabilities are simultaneously influenced by several factors, such as base pairing, base stacking interactions, solvation effects, interactions with counter ions, etc.

The ΔpK_a values have long been suspected to be related to the hydrogen-bonding strengths. The principle called "pK_a equalization" predicts that the driving force of hydrogen bonding is the reduction of the ΔpK_a values of the hydrogen-bond donor and acceptor.[10] The pK_a equalization principle is very general and is applicable to a wide range of systems stabilized by hydrogen bonds, starting from very strong "low barrier" hydrogen bonds relevant to enzymatic catalysis up to medium strength N(O)−H···N(O) hydrogen bonds, representing the vast majority of hydrogen bonds occurring in biomolecules. The theoretical basis of the pK_a equalization principle is the notion that the covalent component of the hydrogen-bonding interactions increases with decreasing ΔpK_a and eventually reaches its maximum when ΔpK_a becomes 0. For example, the impact of the covalent character of the hydrogen bonds on the binding strength is nicely illustrated by Shan and Herschlag on substituted salicylic acid monoanions.[11] They have found that the binding strength of the intramolecular hydrogen bonds decreases with increasing ΔpK_a both in DMSO and water.

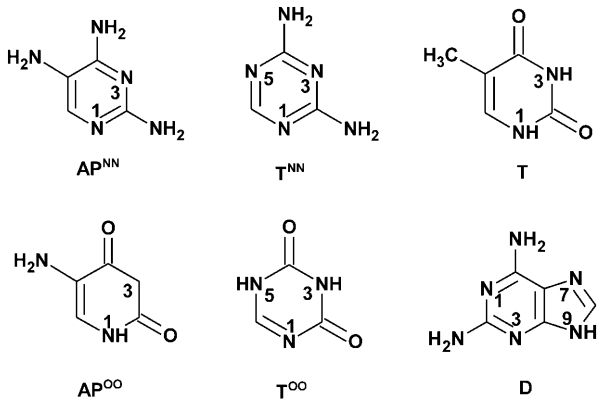
The findings of Mittapalli et al. seem to contradict the p*K*a equalization rule. Reference [6] associates the highest duplex stability with the highest melting points and proposes the strongest duplex stability for the case with the largest ΔpKa. This motivated us to re-investigate the base pairs of reference [6] using high level ab initio quantum chemical calculations. This method is especially suitable to capture molecular orbital effects (associated with the covalent character of the chemical bonds) on the stabilities of hydrogen bonds.

In particular, we will address whether the experimentally observed correlation[6] between the melting point of the duplexes and the pK_a values of the constituent bases is or is not related to the base pairing properties. In addition, knowing that the thermodynamics of the duplex formation might be modulated by the stacking interactions, we will investigate stacking interaction patterns composed of two consecutive base pairs (in other words, stacking in base-pair steps). Our computations will thus account for the intrinsic base pairing and stacking forces acting in these systems. By intrinsic base pairing and stacking energies we mean the results of direct forces between the interacting bases, which reflect the mutual interactions of their electronic structures and which are accurately captured by quantum chemistry.[12] In order to shed light on the complex network of dynamic effects accompanying the duplex formation, we supplement our quantum chemical calculations with explicit solvent molecular dynamics simulations. Thus, our study focuses on both direct (intrinsic) base pairing and base stacking interactions acting at the level of base pairs and base-pair steps as well as the steric compatibility of such base pairs with standard duplex architectures.

 © 2010 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

Computational Methods

In this study, we investigated seven base pairs documented in reference [6]. These base pairs are composed of the following bases: 5-amino-2,4-dioxopyrimidine (hereafter abbreviated as AP^{OO}), 2,4,5-triaminopyrimidine (AP^{NN}), (2,4-diamino)triazine (T^{NN}), (2,4-dioxo)triazine (T^{OO}), thymine (T), and 2,6-diaminopurine (D). All studied base pairs contained two N–H···O and one N–H···N hydrogen bonds. To simplify the subsequent construction of model stacking base-pair steps, C_s symmetry was assumed for all base pairs. Although the base pairs in nucleic acids are commonly propeller twisted, the planar structure is still representative, since the energy changes associated with propeller twisting are small. The optimized geometries of the base pairs are depicted in Figure 1. These geometries were used to generate base-pair steps representing the stacking of two coplanar base pairs in a B-DNA fashion (see Figure 2). For the sake of completeness, as will be shown below, we tested the effect of propeller twisting of the bases on the base stacking energies for the AP^{OO}···T^{NN} base pair.



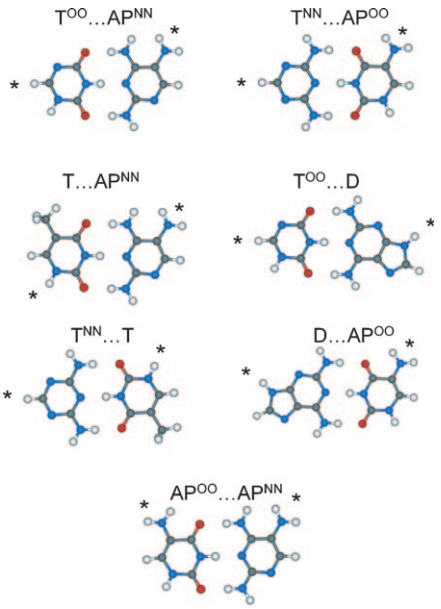


Figure 1. Optimized geometries of the studied base pairs obtained from B3LYP/6-31G** calculations in the gas phase. Connections to the dipeptide based PNA-like backbones (AP^{OO}, AP^{NN}, $T^{OO}$and T^{NN}) and DNA/RNA (T and D) backbones are indicated with *.

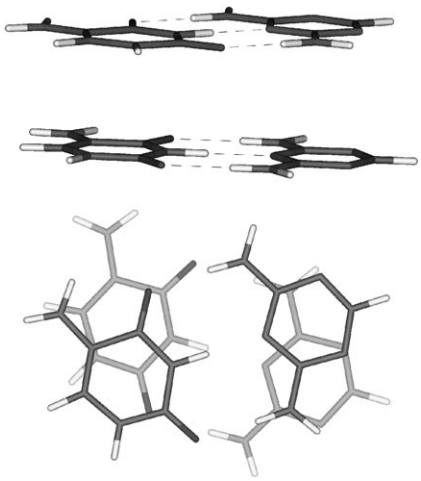


Figure 2. Model of the stacked AP^{OO}···T^{NN} base pairs (base-pair step). The base pair geometry was optimized at the B3LYP/6-31G** level. The base-pair step stacking configuration was then constructed to mimic idealized B-DNA stacking as described in the Supporting Information. Hydrogen-bonding contacts are indicated with scattered lines in the left part of the figure.

Further information on the definition of the stacking geometries, technical details of the computations, such as the computational platforms used for geometry optimization, interaction energy calculations and molecular dynamics (MD) simulations are described in the Supporting Information.[13–34]

Results

Base pairing: Optimized geometries of all base pairs considered in this study are depicted in Figure 1. All studied base pairs form two N–H···O and one N–H···N type hydrogen bonds. The RIMP2/CBS interaction energies (hereafter abbreviated with ΔE_{bp}, see Supporting Information for all definitions and explanations) are quite similar in these systems and range from −17.5 to −19.5 kcal mol^{-1} and from −9.1 to −10.5 kcal mol^{-1} in the gas phase and with inclusion of the continuum solvent correction (using the COSMO approximation to represent the aqueous solution, see Supporting Information), respectively (see the first two columns of Table 1). The corresponding RIMP2/CBS interaction energy of the GC WC base pair is remarkably higher (in absolute value), −31.9 kcal mol^{-1} and −14.8 kcal mol^{-1} in the gas phase and within the COSMO approximation, respectively.[37] Solvent correction with the COSMO method ($\varepsilon = 78.4$, corresponding data are listed in the second column of Table 1) attenuated part of the huge energy difference between the GC base pair and the rest of the base pairs, but did not change the fact that the GC WC pair still remains more stable, albeit by only 4 kcal mol^{-1}. This, however, is still a large difference, taking into account the energy scale of the thermodynamics of the corresponding molecular interactions in water and in nucleic acids.[9b–d,38] Thus, we can conclude that all of the studied base pairs are less stable than the canonical GC WC base pair. Furthermore, they are

 © 2010 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

Table 1. Computed base pairing energies (ΔE_{bp} [kcal mol^{-1}]), base-pair step stacking energies (ΔE_{stack} [kcal mol^{-1}]), melting points (T_m [°C]) and pK_a data for the base pairs considered in this study. The base pairing energies were obtained from RIMP2/CBS calculations, and the stacking energies were computed at the CBS(T) level both in the gas phase and in solution.[a]

Base 1	Base 2	ΔE_{bp}		ΔE_{stack}		T_m[b]	Ref	pK_{a1}	Ref	pK_{a2}	Ref	ΔpK_a[c]
		gas phase	water (COSMO)	gas phase	water (COSMO)							
AP^{OO}	T^{NN}	−19.5	−10.5	−11.5	−9.1	15.5–16.0	[6b]	8.9	[6b]	4.5	[6b]	3.4
AP^{NN}	T^{OO}	−18.2	−9.2	−8.7	−6.9	< 0	[6b]	6.0	[35]	7.2	[36]	1.2
AP^{NN}	T	−17.5	−9.1	−11.1	−7.4	< 10	[6b]	6.0	[35]	9.7	[6b]	3.7
T^{OO}	D	−18.5	−9.7	−13.8	−9.0	14–19.4	[6a]	7.2	[36]	4.4	[6b]	2.8
T^{NN}	T	−19.2	−10.3	−12.1	−6.5	35.8–53.8	[6a]	4.5	[6b]	9.7	[6b]	5.2
AP^{OO}	D	−18.9	−10.3	−14.0	−10.3	41.9–54.3	[6b]	8.9	[6b]	4.4	[6b]	4.5
AP^{NN}	AP^{OO}	−18.1	−9.9	−11.6	−8.1	15.5	[6b]	6.0	[35]	8.9	[6b]	2.9

[a] An extended version of the Table including the BSSE corrected RIMP2/aug-cc-pVDZ and RIMP2/aug-cc-pVTZ interaction energies and other data can be found in the Supporting Information. [b] Melting points were measured for 12-mer duplexes in 1 M NaCl, 10 mM aq. phosphate buffer.[6] [c] $\Delta pK_a = |pK_{a1} - pK_{a2}|$.

intrinsically close to isoenergetic, although their energy is comparable to another triply bonded base pair, 2-aminopurine···thymine.[16]

Base stacking: Stacking geometries were generated from the optimized base pair geometries as explained in the method section and Supporting Information. A representative model used for the AP^{OO}···T^{NN} base pair stacking step is depicted in Figure 2. All sequences correspond to homogenous sequences, as used in the experiments (they have identical bases in one strand).

The stacking energies (calculated using the CBS(T) method and hereafter abbreviated as ΔE_{stack}, see Supporting Information) vary from −8.7 to −14.0 kcal mol^{-1} and from −6.5 to −10.3 kcal mol^{-1} in the gas phase and within the COSMO approximation, respectively (see the data in the third and fourth columns of Table 1).

Considering the gas-phase data, out of the seven studied alternative base-pair steps, the one containing the AP^{NN}···T^{OO} pairs exhibits the weakest stacking (interaction energy is −8.7 kcal mol^{-1}). The AP^{OO}···T^{NN}, AP^{NN}···T, AP^{NN}···AP^{OO} and T^{NN}···T base pairs lead to very similar stacking energies (from −11.1 to −12.1 kcal mol^{-1}) as the step consisting of two GC base pairs with the analogous homogenous 5′-GG-3′ sequence (−11.2 kcal mol^{-1}).[17e] Note that the 5′-GG-3′ stacking is known to be rather weak due to unfavorable intrastrand electrostatic interactions.[17b,e] Among the studied alternative base-pair steps, the strongest gas-phase stacking was observed for those two systems containing purine bases, AP^{OO}··· D and T^{OO}···D. In these two cases, the stacking energies (−14.0 and −13.8 kcal mol^{-1}, respectively) moderately exceed the corresponding gas-phase value obtained for the base-pair step formed by two AT pairs with the homogeneous 5′-AA-3′ sequence (−13.1 kcal mol^{-1}).[17e]

For comparison, we have computed the COSMO solvent corrections (for methodological details see the computational details section in the Supporting Information) for the 5′-AA-3′ and 5′-GG-3′ steps and corrected the gas-phase values reported in ref. [17e]. In this approximation, the stacking energies are −9.1 and −9.6 kcal mol^{-1} for the 5′-AA-3′ and 5′-GG-3′ steps. Among the studied alternative base-pair steps, the T^{OO}···D, AP^{OO}···D, and AP^{OO}···T^{NN} base pairs lead to similar stacking energies (see column 4 of Table 1), all other systems exhibit weaker stacking interactions in the frame of the COSMO solvation model. In other words, those base-pair steps that contain larger purine bases show larger base stacking, as expected from their larger van der Waals overlap. Indeed the gas-phase MP2/6-31G** interaction energies computed for the intrastrand stacking of two D and two T^{OO} bases (−1.5 and −1.6 kcal mol^{-1}, respectively) in the base-pair step consisting of two T^{OO}···D base pairs as well as that of two D in AP^{OO}···D (−1.5 kcal mol^{-1}) are the largest among the computed base-base stacking terms. (The inclusion of a polar solvent into the computational scheme drastically attenuates the electrostatic component to stacking, but the stabilizing effects of the large dispersion forces acting in the purine-containing base-pair steps are not affected.) Thus, the stacking stabilization is ultimately determined by the van der Waals overlap, perhaps with exception of the base-pair step with the AP^{OO}···T^{NN} base pairs. For AP^{OO}···T^{NN}, the gas-phase stacking stabilization (−11.5 kcal mol^{-1}) is noticeably weaker than that in the two purine containing base-pair steps. Nevertheless, the sequence has a relatively low positive solvent correction to the gas phase stacking energy (+2.4 kcal mol^{-1}), which, for AP^{OO}···T^{NN}, in total results in a stacking energy similar to the one computed for T^{OO}···D.

Overall, the stacking energy calculations do not indicate anything unusual compared to standard base pairs. An extended version of Table 1 containing all numerical results can be found in the Supporting information (Table S1).

Note, that our computational models assumed C_s symmetry, which means that the bases within the base pair were coplanar. For completeness, we also considered the effect of a propeller twist of the bases on the base stacking energies. The propeller twist was introduced as a counter-rotation of the bases around the long base pair axis (which involves C6 of triazines and pyrimidines as well as C8 of purines). We calculated RIMP2/aug-cc-pVDZ interaction energies for these systems in the gas phase by fixing the propeller twist at 0, 15 and 30° and the vertical separation of the base pairs at 3.15, 3.30 and 3.45 Å (calculated for the midpoints of the long base pair axes). We have found that the combination of

 © 2010 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

0° of propeller twist and 3.3 Å of vertical separation gives the best interaction energy. Moreover, the interaction energies change only marginally when the propeller twist is 0 or 15°, irrespective of the vertical separation of the bases (see the diagram in Figure S1 of the Supporting Information). Thus, all these geometries are well within the low-energy region. This justifies our approach, in which 0° of propeller twist and 3.3 Å of vertical separation in the stacking energy calculations is assumed.

MD simulations: The classical MD simulations by AMBER[20,22–24] were carried out for the B-DNA homoduplexes with the A ···T pair as well as with the isosterically modified variants of the $AP^{OO}\cdots T^{NN}$ and $AP^{NN}\cdots T^{OO}$ base pairs (hereafter referred to as $mAP^{OO}\cdots mT^{NN}$ and $mAP^{NN}\cdots mT^{OO}$, respectively) on the 20+ ns time scale. The base pair modifications were necessitated by the fact that neither the $AP^{OO}\cdots T^{NN}$ nor the $AP^{NN}\cdots T^{OO}$ base pair could be attached to the DNA sugar-phosphate backbone (for more details see the Supporting Information). Figure 3 documents the above mentioned base pair modifications and the attachment of nonstandard base pairs to the DNA backbone.

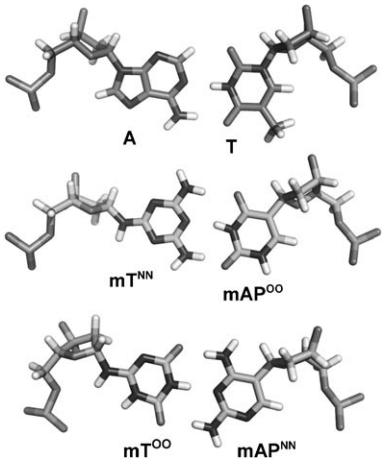


Figure 3. A···T, modified $AP^{OO}\cdots T^{NN}$ ($mAP^{OO}\cdots mT^{NN}$) and modified $AP^{NN}\cdots T^{OO}$ ($mAP^{NN}\cdots mT^{OO}$) model geometries used for MD simulations of the DNA duplexes.

All B-DNA duplexes were stable except towards the opening or frying of terminal base pairs, which is common in DNA duplex simulations. The base pairing and stacking were well preserved in all simulations, indicating that both tested alternative base pairs may well be incorporated in B-DNA type duplexes. Distortions of the double helix on the present simulation time scale (see Figure 4) were not observed. Furthermore, averaged MD geometries of stacked base-pair steps only modestly differ from the idealized geometries used in QM calculations. For example, the distance of the center of masses of the bases in the averaged simulated geometries vary from 5.54 to 5.56 Å and from 5.56 to 5.61 Å for $mAP^{OO}\cdots mT^{NN}$ and $mAP^{NN}\cdots mT^{OO}$ duplexes, respectively. These data are in sound agreement with the corresponding data of the ab initio optimized geometries, which are 5.35 Å and 5.41 Å for $AP^{OO}\cdots T^{NN}$ and $AP^{NN}\cdots T^{OO}$ base pairs, respectively. Furthermore, the vertical separation of the base-pair steps varies from 3.30 to 3.33 Å and from 3.32 to 3.34 Å for the simulated geometries of $mAP^{OO}\cdots mT^{NN}$ and $mAP^{NN}\cdots mT^{OO}$ duplexes, respectively. This is again in perfect agreement with the vertical separation used to construct our ab initio models, 3.3 Å. The only noteworthy difference between the idealized and MD geometries is a modest slide of the base pairs in the MD geometries. This, however, in our opinion should have only a small effect on the energetics of the stacking interactions. A graphical comparison of the aromatic ring positions in the MD geometries and in the idealized geometries used for quantum chemical calculations is presented in Figure S2 (in the Supporting Information). Thus, the averaged MD geometries justify our theoretical stacking model described in the computational details section of the Supporting Information and, further prove that the tested unnatural base pairs are sterically quite compatible with standard duplexes. PDB files containing the averaged simulated geometries (created by averaging the structures from the last two nanoseconds of MD simulations) are available from the Supporting Information as well.

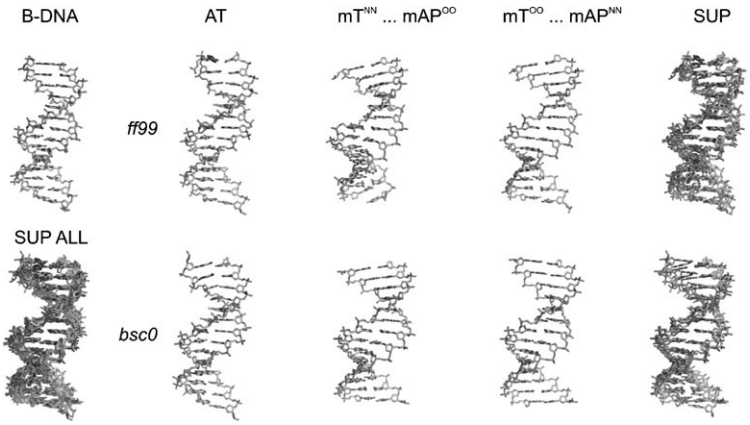


Figure 4. Duplex geometries from MD simulations. From top left to bottom right: the starting structure (terminal base pairs are not shown) in B-DNA form, the averaged structure (30-31 ns) of an AT duplex (ff99 force field), an averaged structure of a $mT^{NN}\cdots mAP^{OO}$ duplex (ff99), an averaged structure of a $mT^{OO}\cdots mAP^{NN}$ duplex (ff99), superimposition (SUP) of all preceding ff99 structures, superimposition (SUP ALL) of all averaged structures, averaged structures of AT, $mT^{NN}\cdots mAP^{OO}$, $mT^{OO}\cdots mAP^{NN}$ duplexes with bsc0 force field and their superimposition. See Supporting Information for all methodological details.

 © 2010 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

Comparison with experimental data

Correlation between the melting points, ΔpK_a and ΔE_{stack}: In the last column of Table 1 we list the ΔpK_a values of the studied base pairs. Mittapalli et al.[6] reported a correlation between the ΔpK_a values and the melting points of the corresponding duplexes. They then suggested that the stability of the helices may reflect stability of the base pairing, thus implying a correlation between ΔpK_a and base pairing strength. However, there are several issues that need to be considered. First, melting points do not always have a clear correspondence to stability as described by free energies.[9] Second, and more importantly, the stability of base-paired helices reflects not only the strength of base pairing, but also stacking interactions and other contributions such as hydration, in particular the entropic and enthalpic effect of the explicit water molecules coordinated to the duplexes.[39] In addition, special sequences such as the homogenous oligo-N tracts can have specific patterns of stacking energetics which may affect stabilities and can be accompanied with distinct structural and dynamical features.[9,18] For example, guanine tracts (consecutive guanines in one strand) are known to be shifted to B-A intermediate structure[18a,d], whereas adenine tracts adopt specific structures that are significantly propeller-twisted and rather stiff.[18b,d] To gain a better insight, Figure 5 shows five correlation diagrams.

Clearly, systematic trends cannot be recognized between ΔpK_a and ΔE_{bp} as well as between ΔpK_a and ΔE_{stack}. Similarly, there is no correlation between T_m and ΔE_{stack}. Conversely, there is a fair correlation between ΔpK_a and T_m, which can be fitted with a linear fitting function (goodness of fitting expressed by the adjusted R^2 is 0.5435). A much better fitting (adjusted R^2 is 0.8118) was obtained when ex-

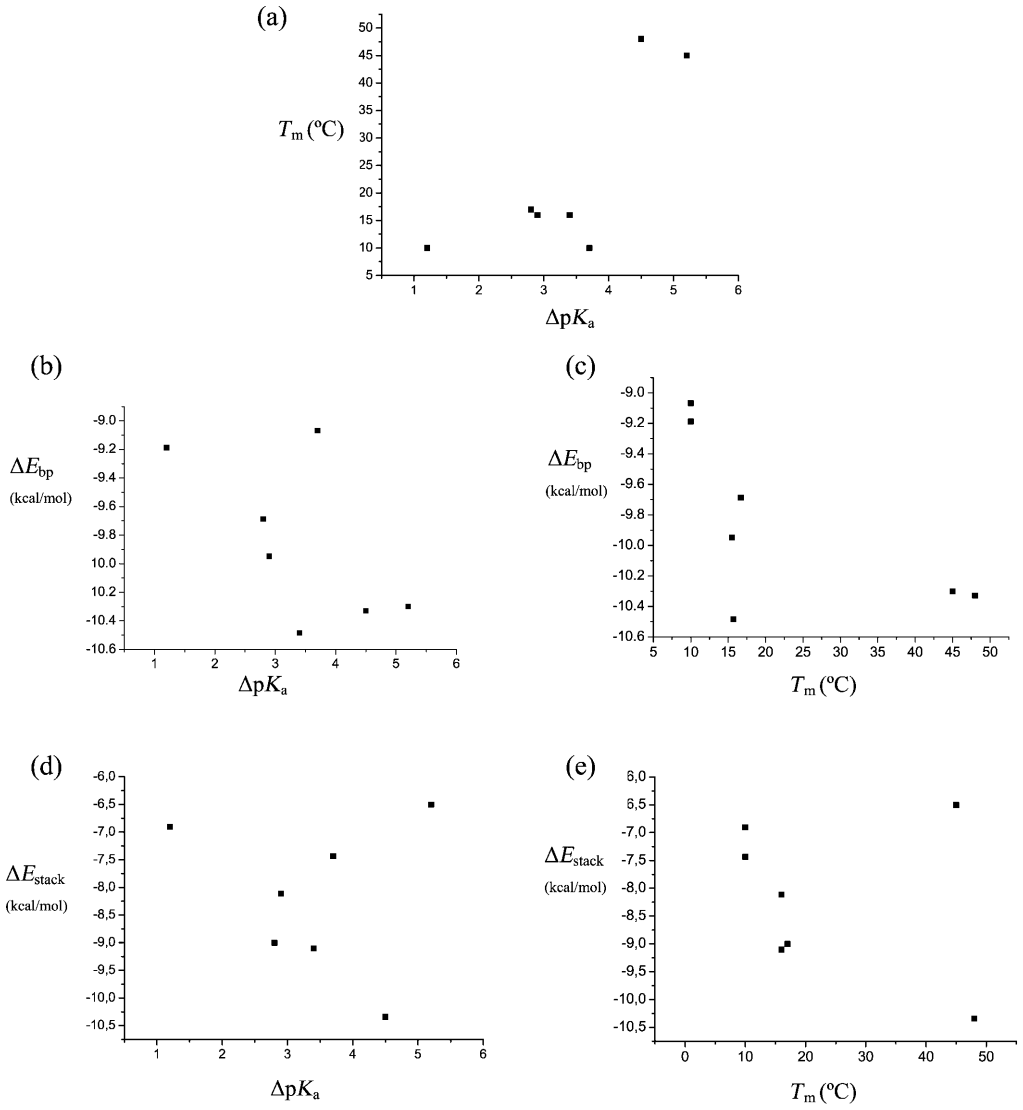


Figure 5. Correlation diagrams: a) ΔpK_a versus T_m (°C); b) ΔpK_a versus ΔE_{bp} (kcal mol^{-1}); c) T_m (°C) versus ΔE_{bp} (kcal mol^{-1}); d) ΔpK_a versus ΔE_{stack} (kcal mol^{-1}); e) T_m (°C) versus ΔE_{stack} (kcal mol^{-1}). The T_m of the two unstable duplexes ($AP^{NN}\cdots T^{OO}$ and $AP^{NN}\cdots T$) are arbitrarily chosen to be 10 °C. ΔE_{bp} and ΔE_{stack} were computed with the COSMO continuum solvent approximation, for details see the text and the Supporting Information.

 © 2010 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

cluding the two unstable duplexes ($AP^{NN}\cdots T^{OO}$ and $AP^{NN}\cdots T$) from the fit. At first glance, it seemed to us that there was a qualitative correlation between the base pairing energies and T_m. Nevertheless, we were not able to fit the available data with a linear fitting function with a sufficient confidence (adjusted R^2 is 0.3164). We repeated the fitting without the data of the two unstable duplexes (for which the T_m was set to be 10 °C in the graph). However, in this case the adjusted R^2 of the fitting was −0.0809. Negative values of the adjusted R^2 represent the situation when the data can be fitted with a horizontal straight line, which indicates that the fitted value of the base pairing energy is constant over the entire studied melting point range. Thus, statistical analysis could not confirm a correlation between the intrinsic base pair strengths and T_m.

From a careful evaluation of the correlation diagrams we conclude that the duplex stabilities, as established from melting point measurements, are dictated neither by the intrinsic strength of the hydrogen bonding nor by that of the stacking interactions. Perhaps, they could be determined by the energetic contribution of the duplex formation and the associated solvation effects. This does not exclude that the duplex stability correlates or is pre-determined by the pK_a values of the interacting bases, since this process certainly involves complex solvation/desolvation effects and eventually some complicated protonation equilibria which both could be pK_a dependent. Note that many of the alternative bases posses pK_a values close to neutrality, which standard nucleobases strictly avoid. We just stress that the correlation between the ΔpK_a and the melting points is not a straightforward consequence of the intrinsic base pairing and stacking interactions and is rather the result of the overall complex balance of all molecular forces. Note that the thermodynamics of the present systems is also affected by the use of the dipeptide-based backbone, in addition to all factors that are in effect also in DNA. Interestingly, even our calculated base pairing energies for isolated hydrogen-bonded base pairs do not reveal the expected trends suggested by the basic pK_a equalization principle. Perhaps, it may be due to the fact that the systems studied by Gilli et al.[10] were rather simple systems with isolated hydrogen bonds, which may not be fully representative for base pairs with multiple electronically coupled hydrogen bonds. Evidently, when going from the intrinsic base pairing stabilities calculated by theory to the real, thermodynamic, experimental data for duplexes of Mittapalli et al.,[6] any comparison becomes further complicated by all the other numerous contributions (such as stacking, solvation, etc., see above) that affect the final thermodynamics properties. This makes simple explanations of the observed correlations between various experimental quantities difficult.

We have found that the base pairing energy is practically the same for all of the studied systems and lower than the one computed for the GC WC pair. This is indeed not surprising since the two N−H···O=C hydrogen bonds are *anti* in the G···C pair, whereas they adopt a *cis* orientation in the rest of the studied systems. Jorgensen and Pranata demonstrated that the *trans* orientation always results in a stronger hydrogen bonding than the *cis* orientation,[40] which can be explained by the alternating distribution of the partial charges brought about by the hydrogen bonding. This popular explanation is also known as the effect of secondary interactions in multiple hydrogen-bonded systems, and it is related to alternative explanations based on consideration of dipole moments and complementarity of electrostatic potentials. All three explanations are related, since alteration of donors and acceptors on the nucleobase edge obviously reduces the dipole moment and softens the electrostatic potentials.[16,41]

Let us also note that the correlation between the melting points and the stacking energies can be ruled out for one more reason. Our interaction energy calculations both in the gas phase and in solution conclusively show that, due to the increased dispersion interaction, with one exception ($AP^{OO}\cdots T^{NN}$), stacking in the two purine containing base-pair steps ($AP^{OO}\cdots D$ and $T^{OO}\cdots D$) is stronger than in the pyrimidine and triazine containing base pairs. On the contrary, the melting point data (41.9–54.3 and 14–19.4 °C, respectively) are markedly different for these two systems.

Conclusion

In the current study, we carried out quantum chemical analyses of the strength of base pairing and stacking interactions in a series of base pair analogues, which occur in the duplexes of 12-mer dipeptide based PNA-like strands, with each other and as complementary single stranded oligomers of the DNA-backbone. These systems are close structural relatives of the canonical DNA and RNA base pairs, and, thus, could be components of a potentially primordial information polymer, as suggested by Mittapalli et al.[6] The nucleobases participating in the base pairs were selected on the basis of their first pK_a values, which fall in the range of 4–10. It was suggested that the difference in the pK_a values of the participating bases might be related to the duplex stabilities measured in aqueous solution. All studied base pairs were stabilized by two N−H···O=C hydrogen bonds separated by an N−H···N hydrogen bond.

MD simulation of the isosterically substituted $AP^{OO}\cdots T^{NN}$ and $AP^{NN}\cdots T^{OO}$ 16-mers revealed that both systems are capable of forming stable duplexes with a B-DNA backbone and justified the theoretical model used for stacking energy calculations. We calculated RIMP2/CBS base pairing energies as well as estimated CBS(T) stacking energies for selected base pairs from this series, in order to evaluate whether the intrinsic hydrogen bonding and stacking interactions of the base pairs may contribute to the above mentioned relationship between the pK_a values of the participating bases and the duplex stabilities.

The computed interaction energies correlated neither with the melting points nor with the different pK_a values. This conclusively reveals that the intrinsic stability of the hydrogen bonding and stacking interactions is not the primary

 © 2010 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

factor in determining the stability of the duplexes studied by Mittapalli et al.[6] Instead, the thermodynamics of duplex formation likely results from a complex mixture of energy contributions, which also include the effects associated with the backbone, solvation effects, etc. In other words, the stability of the duplexes cannot be deduced from the intrinsic properties of base pairs and stacks in any straightforward manner. We think that this is a general conclusion which would also be valid for other types of base pairs. This puts forward the idea that if the different pK_a values of the nucleobases ever had a role in the selection of the nucleobases, then it is most likely because it might influence the solvation and protonation equilibria of the oligonucleotide sequences forming the duplex (and other) NA architectures. For example, since some of the alternative nucleobases posses pK_a values substantially shifted to neutrality compared to standard nucleobases, their pK_a values would likely substantially affect pH-dependence and stability of the folding of the earliest biopolymers (presumably the RNA) compared to the behavior of NA with standard nucleobases with all pK_a values far from neutrality. However, more effort will be needed to understand the physical chemistry that may be responsible for the above noted correlation.

Acknowledgements

This work was supported by the Ministry of Education of the Czech Republic (grant numbers AVOZ50040507, AVOZ50040702, MSM0021622413, LC06030, MSM6198959216, LC512), by the Grant Agency of the Academy of Sciences of the Czech Republic (grant numbers 1QS500040581, IAA400550701 and IAA400040802) and Grant Agency of the Czech Republic (grant numbers 203/09/1476 and 203/09/H046). Work at Oak Ridge National Laboratory (ORNL) was supported by the Center for Nanophase Materials Sciences, sponsored by the Division of Scientific User Facilities, U.S. Department of Energy (USDOE) and used resources of the National Center for Computational Sciences, ORNL, supported by the Office of Science, USDOE. This research used resources of the National Energy Research Scientific Computing Center, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DEAC02-05CH11231. It also used an allocation of advanced computing resources supported by the National Science Foundation; these computations were performed on Kraken (a Cray XT5) at the National Institute for Computational Sciences (http://www.nics.tennessee.edu/). A.V.M. thanks the financial support provided by the USDOE, offices of Basic Energy Science and Advanced Scientific Computing Research as part of the SciDac program. J.E.Š. and J.Š. thank Zdeněk Salvet for the maintenance of the computing facilities of the Brno group.

[1] I. Hirao, *Curr. Opin. Chem. Biol.* **2006**, *10*, 622–627.
[2] a) S. A. Benner, *Acc. Chem. Res.* **2004**, *37*, 784–797; b) C. R. Geyer, T. R. Battersby, S. A. Benner, *Structure* **2003**, *11*, 1485–1498.
[3] E. T. Kool, *Curr. Opin. Chem. Biol.* **2000**, *4*, 602–608.
[4] Y. Kim, A. M. Leconte, Y. Hari, F. E. Romesberg, *Angew. Chem.* **2006**, *118*, 7973–7976; *Angew. Chem. Int. Ed.* **2006**, *45*, 7809–7812.
[5] a) A. Eschenmoser, *Tetrahedron* **2007**, *63*, 12821–12844; b) Z. J. Wang, H. K. Huynh, B. Han, R. Krishnamurthy, A. Eschenmoser, *Org. Lett.* **2003**, *5*, 2067–2070; c) X. L. Wu, G. Delgado, R. Krishnamurthy, A. Eschenmoser, *Org. Lett.* **2002**, *4*, 1283–1286.
[6] a) G. K. Mittapalli, K. R. Reddy, H. Xiong, O. Munoz, B. Han, F. De Riccardis, R. Krishnamurthy, A. Eschenmoser, *Angew. Chem.* **2007**, *119*, 2522–2529; *Angew. Chem. Int. Ed.* **2007**, *46*, 2470–2477; b) G. K. Mittapalli, Y. M. Osornio, M. A. Guerrero, K. R. Reddy, R. Krishnamurthy, A. Eschenmoser, *Angew. Chem.* **2007**, *119*, 2530–2536; *Angew. Chem. Int. Ed.* **2007**, *46*, 2478–2484.
[7] a) P. E. Nielsen, M. Egholm, R. H. Berg, O. Buchardt, *Science* **1991**, *254*, 1497–1500; b) M. Egholm, O. Buchardt, P. E. Nielsen, R. H. Berg, *J. Am. Chem. Soc.* **1992**, *114*, 1895–1897.
[8] a) A. Sen, P. E. Nielsen, *Biophys. J.* **2006**, *90*, 1329–1337.
[9] a) B. S. Alexandrov, V. Gelev, Y. Monisova, L. B. Alexandrov, A. R. Bishop, K. Ø. Rasmussen, A. Usheva, *Nucleic Acids Res.* **2009**, *37*, 2405—2410; b) S. M. Freier, N. Sugimoto, A. Sinclair, D. Alkema, T. Neilson, R. Kierzek, M. H. Caruthers, D. H. Turner, *Biochemistry* **1986**, *25*, 3214–3219; c) D. H. Mathews, J. Sabina, M. Zuker, D. H. Turner, *J. Mol. Biol.* **1999**, *288*, 911–940; d) J. SantaLucia, D. Hicks, *Annu. Rev. Biophys. Biomol. Struct.* **2004**, *33*, 415–440; e) D. H. Turner in *Nucleic Acids: Structures, Properties, and Functions* (Eds.: V. A. Bloomfield, D. M. Crothers, I. Tinoco, Jr.), University Science Books, Sausalito, **2000**, pp. 259–334; f) J. L. Mergny, L. Lacroix, *Oligonucleotides* **2003**, *13*, 515—537.
[10] a) P. Gilli, L. Pretto, V. Bertolasi, G. Gilli, *Acc. Chem. Res.* **2009**, *42*, 33–44; b) P. Gilli, V. Bertolasi, L. Pretto, L. Antonov, G. Gilli, *J. Am. Chem. Soc.* **2005**, *127*, 4943–4953.
[11] S. O. Shan, D. Herschlag, *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 14474–14479.
[12] J. Šponer, J. Leszczynski, P. Hobza, *Biopolymers* **2001**, *61*, 3–31.
[13] Gaussian 03, Revision C.02, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, J. A. Pople, Gaussian, Inc., Wallingford CT, **2004**.
[14] A. D. Becke, *J. Chem. Phys.* **1993**, *98*, 5648–5652.
[15] C. T. Lee, W. T. Yang, R. G. Parr, *Phys. Rev. B* **1988**, *37*, 785–789.
[16] J. Šponer, P. Jurečka, P. Hobza, *J. Am. Chem. Soc.* **2004**, *126*, 10142–10151.
[17] For theoretical evaluation of the stacking interactions of natural nucleobases and base pairs refer to the following works: a) J. Šponer, J. Leszczynski, P. Hobza, *J. Phys. Chem.* **1996**, *100*, 5590–5596; b) J. Šponer, H. A. Gabb, J. Leszczynski, P. Hobza, *Biophys. J.* **1997**, *73*, 76–87; c) J. Šponer, P. Hobza, *Chem. Phys. Lett.* **1997**, *267*, 263–270; d) P. Hobza, J. Šponer, *J. Am. Chem. Soc.* **2002**, *124*, 11802–11808; e) J. Šponer, P. Jurečka, I. Marchan, F. J. Luque, M. Orozco, P. Hobza, *Chem. Eur. J.* **2006**, *12*, 2854–2865; f) J. Šponer, K. E. Riley, P. Hobza, *Phys. Chem. Chem. Phys.* **2008**, *10*, 2595–2610; g) M. Pitonak, K. E. Riley, P. Neogrady, P. Hobza, *ChemPhysChem* **2008**, *9*, 1636–1644.
[18] a) H. L. Ng, R. E. Dickerson, *Nucleic Acids Res.* **2002**, *30*, 4061–4067; b) H. C. M. Nelson, J. T. Finch, B. F. Luisi, A. Klug, *Nature* **1987**, *330*, 221–226; c) S. Warmlander, J. E. Šponer, J. Šponer, M. Leijon, *J. Biol. Chem.* **2002**, *277*, 28491–28497; d) F. Lankaš, T. E. Cheatham, N. Špačková, P. Hobza, J. Langowski, J. Šponer, *Biophys. J.* **2002**, *82*, 2592–2609; e) J. Šponer, J. Florián, H. L. Ng, J. E. Šponer, N. Špačková, *Nucleic Acids Res.* **2000**, *28*, 4893–4902.
[19] W. K. Olson, A. V. Colasanti, Y. Li, W. Ge, G. Zheng, V. B. Zhurkin in *Computational Studies of RNA and DNA* (Eds.: J. Šponer, F. Lankaš, Springer, Dordrecht, **2006**, pp. 235–257.

 © 2010 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

[20] A. Pérez, I. Marchán, D. Svozil, J. Šponer, T. E. Cheatham III, C. A. Laughton, M. Orozco, *Biophys. J.* **2007**, *92*, 3817–3829.
[21] H. R. Drew, S. Samson, R. E. Dickerson, *Proc. Natl. Acad. Sci. USA* **1982**, *79*, 4040–4044.
[22] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, P. A. Kollman, *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.
[23] J. M. Wang, P. Cieplak, P. A. Kollman, *J. Comput. Chem.* **2000**, *21*, 1049–1074.
[24] D. A. Pearlman, D. A. Case, J. W. Caldwell, W. S. Ross, T. E. Cheatham, S. Debolt, D. Ferguson, G. Seibel, P. Kollman, *Comput. Phys. Commun.* **1995**, *91*, 1–41.
[25] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys.* **1983**, *79*, 926–935.
[26] a) A. Halkier, T. Helgaker, P. Jørgensen, W. Klopper, H. Koch, J. Olsen, *Chem. Phys. Lett.* **1999**, *302*, 437–446; b) A. Halkier, T. Helgaker, P. Jørgensen, W. Klopper, H. Koch, J. Olsen, A. K. Wilson, *Chem. Phys. Lett.* **1998**, *286*, 243–252.
[27] a) K. Eichkorn, O. Treutler, H. Oehm, M. Haeser, R. Ahlrichs, *Chem. Phys. Lett.* **1995**, *242*, 652–660; b) F. Weigend, M. Haeser, *Theor. Chem. Acc.* **1997**, *97*, 331–340; c) F. Weigend, M. Haeser, H. Patzelt, R. Ahlrichs, *Chem. Phys. Lett.* **1998**, *294*, 143–152.
[28] P. Jurečka, P. Nachtigall, P. Hobza, *Phys. Chem. Chem. Phys.* **2001**, *3*, 4578–4582.
[29] S. F. Boys, F. Bernardi, *Mol. Phys.* **1970**, *19*, 553–566.
[30] a) P. Hobza, J. Šponer, *THEOCHEM* **1996**, *388*, 115–120; b) P. Jurečka, P. Hobza, *Chem. Phys. Lett.* **2002**, *365*, 89–94.
[31] a) M. O. Sinnokrot, E. F. Valeev, C. D. Sherrill, *J. Am. Chem. Soc.* **2002**, *124*, 10887–10893; b) M. O. Sinnokrot, C. D. Sherrill, *J. Phys. Chem. A* **2004**, *108*, 10200–10207; c) M. O. Sinnokrot, C. D. Sherrill, *J. Phys. Chem. A* **2006**, *110*, 10656–10668.
[32] a) MOLPRO, version 2008.1, a package of ab initio programs, H.-J. Werner, P. J. Knowles, R. Lindh, F. R. Manby, M. Schütz, and others, see http://www.molpro.net; b) C. Hampel, K. A. Peterson, H.-J. Werner, *Chem. Phys. Lett.* **1992**, *190*, 1–12, and references therein.; c) The program to compute the perturbative triples corrections has been developed by M. J. O. Deegan, P. J. Knowles, *Chem. Phys. Lett.* **1994**, *227*, 321–326.
[33] a) V. Barone, M. Cossi, *J. Phys. Chem. A* **1998**, *102*, 1995–2001; b) M. Cossi, N. Rega, G. Scalmani, V. Barone, *J. Comput. Chem.* **2003**, *24*, 669–681; c) A. Klamt, G. Schuurmann, *J. Chem. Soc. Perkin Trans. 1* **1993**, *2*, 799–805.
[34] C. L. Zirbel, J. E. Šponer, J. Šponer, J. Stombaugh, N. B. Leontis, *Nucleic Acids Res.* **2009**, *37*, 4898–4918.
[35] B. Roth, J. Z. Strelitz, *J. Org. Chem.* **1969**, *34*, 821–836.
[36] J. Jonas, J. Gut, *Collect. Czech. Chem. Commun.* **1962**, *27*, 716.
[37] The gas-phase value, $-31.9\ \mathrm{kcal\,mol^{-1}}$, was taken from J. Šponer, M. Zgarbová, P. Jurečka, K. E. Riley, J. E. Šponer, P. Hobza, *J. Chem. Theory Comput.* **2009**, *5*, 1166–1179; the solvent correction, $+17.1\ \mathrm{kcal\,mol^{-1}}$, was taken from J. E. Šponer, K. Réblová, A. Mokdad, V. Sychrovský, J. Leszczynski, J. Šponer, *J. Phys. Chem. B* **2007**, *111*, 9153–9164.
[38] a) J. Florián, J. Šponer, A. Warshel, *J. Phys. Chem. B* **1999**, *103*, 884–892.
[39] a) J. D. Dunitz, *Science* **1994**, *264*, 670; b) J. D. Dunitz, *Chem. Biol.* **1995**, *2*, 709–712.
[40] W. L. Jorgensen, J. Pranata, *J. Am. Chem. Soc.* **1990**, *112*, 2008–2010.
[41] a) J. Šponer, J. Leszczynski, P. Hobza, *J. Phys. Chem.* **1996**, *100*, 1965–1974; b) J. Šponer, P. Jurečka, P. Hobza in *Computational Studies of RNA and DNA* (Eds.: J. Šponer, F. Lankaš), Springer, Dordrecht, **2006**, pp. 343–388.

Received: July 24, 2009
Revised: November 4, 2009
Published online: January 29, 2010

 © 2010 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim