DOI: 10.1002/chem.200501126

Exceptional Thermodynamic Stability of DNA Duplexes Modified by Nonpolar Base Analogues Is due to Increased Stacking Interactions and Favorable Solvation: Correlated Ab Initio Calculations and Molecular Dynamics Simulations

David Řeha, Michal Hocek, and Pavel Hobza*[a]

Abstract: The geometries of DNA hexamer (5′-GGAACC-3′) and DNA 13-mer (5′-GCGTACACATGCG-3′) have been determined by molecular dynamics (MD) simulations using an empirical force field. The central canonical base pair was replaced by a pair of nonpolar base analogues, 2,2′-bipyridyl and 3-methylisocarbostyril. The stabilization energy of the model system (model A) consisting of a central base pair (base-analogue pair) and two neighboring base pairs was determined by the RI-MP2 method using an extended aug-cc-pVDZ basis set. The geometry of the model was averaged from structures determined by MD simulations. The role of the solvent was covered by the COSMO continuum solvent model and calculations were performed for a larger model system (model B) which also contained a sugar–phosphate backbone. The total stabilization energies of the unperturbed system and the system perturbed by a base-analogue pair (model A) were comparable to the stability of both duplexes experimentally determined. This is due to large stacking interaction energy of the base-analogue self-pair which compensates for the missing hydrogen-bonding energy of the replaced adenine···thymine base pair. The selectivity of the base-analogue pair was reproduced (model B) when their desolvation energy was included with the interaction energy of both strands determined by the approximate SCC-DFTB-D method.

Keywords: ab initio calculations · DNA structures · molecular dynamics simulations · solvent effects · stacking interactions

Introduction

The structure of duplex DNA is based on the complementary Watson–Crick (WC) hydrogen-bonding patterns of A–T and G–C pairs. Replacement of the natural nucleobases by diverse surrogates has become very popular in the last few years.[1] Use of these unnatural nucleobases has been directed towards three main goals: 1) the formation of stabilized duplexes,[2] 2) the design of universal nucleobases[3] that do not discriminate between the complementary bases, and, most importantly, 3) the extension of the genetic alphabet.[4]

The design of artificial base-pair analogues relies on diverse types of molecular interactions which are believed to be dominant: 1) hydrogen bonds between unnatural nucleobases[5] (e.g. isoG–isoC), 2) other noncovalent interactions[6] (mainly of the van der Waals type), or 3) metal chelation.[7] The first generation of unnatural purine and pyrimidine bases and analogues based on hydrogen-bonding patterns similar to WC suffered[8] from rather low selectivity in replication due to stable tautomeric forms that can mismatch pair with natural nucleobases and misinteract with polymerases. Therefore, the second generation was based on nonpolar hydrophobic nucleobase analogues (usually just simple aromatic rings in C-nucleosides). As stacking interactions are the major forces that hold[9] a duplex together, the absence of hydrogen bonds does not destabilize a duplex too much[6] and, in some cases, the increased stacking of the aromatic rings may lead to even more stable duplexes.[10] These hydrophobic pairs are very selective probably due to the energetically disadvantageous desolvation required for mismatch pairing of the hydrophobic nucleobase with a hydro-

[a] Dr. D. Řeha, Dr. M. Hocek, Prof. P. Hobza
The Institute of Organic Chemistry and Biochemistry
Academy of Sciences of the Czech Republic
and Center for Biomolecules and Complex Molecular Systems
Flemingovo nám. 2, 166 10 Prague 6 (Czech Republic)
Fax: (+420) 220-410-320
E-mail: pavel.hobza@uochb.cas.cz

Supporting information for this article is available on the WWW under http://www.chemeurj.org/ or from the author.

philic natural nucleobase. DNA polymerases are quite promiscuous and in many cases do recognize and selectively incorporate[6,11] hydrophobic nucleobases into the DNA duplex using the triphosphates of the particular base-modified nucleosides (it seems that the hydrogen bonds are not crucial even for selectivity in DNA synthesis). These base analogues are usually selective to form self-pairs (the self-pair consists of two same-base analogue molecules). An initial study was also performed to develop a protocol for the enzymatic phosphorylation[12] of hydrophobic unnatural nucleosides to the corresponding triphosphates necessary as substrates for DNA polymerase. Several artificial nucleobase pairs have already been successfully used for in vitro transcription and enzymatic incorporation of noncoded amino acids into proteins.[4,13] However, there is still a great interest in the design and synthesis of novel types of highly selective and biocompatible nucleobase-pair analogues for potential use in artificial organisms with an extended genetic alphabet. Molecular modeling based on ab initio calculations combined with molecular dynamics could be of great advantage in the further design of unnatural base pairs with improved selectivity.

The theoretical procedures used to study base-pair analogues should be, however, carefully tested to avoid the misinterpretation of theoretical results.

The thermodynamic stability of a DNA duplex modified by different base analogues is studied by measuring the melting temperature. These experiments show surprisingly high stability (in comparison with unmodified DNA) and thermodynamic selectivity (difference between the stability of base pairs and mispairs).[6,10] These results indicate that the replacement of strong hydrogen bonds in adenine···thymine or even in guanine···cytosine base pairs is compensated by nonspecific stacking interactions between nonpolar (mostly aromatic) systems. In 1996 using high-level quantum-chemical calculations we showed for the first time the important role of base stacking and in the following years we demonstrated that stacking of DNA base pairs can be as important as hydrogen bonding and in the case of AT-rich DNA the contribution of the stacking interaction to the stability of the system is comparable to that of hydrogen bonding.[14] Stacking plays an even more important role in the interaction of intercalators and drugs with DNA. While in the case of minor-groove binders part of the stabilization can be attributed to hydrogen bonding, in the case of intercalators, practically all the stabilization originates from the stacking interaction. The theoretical study of base stacking is very demanding and today is one of the most difficult tasks of computational biology. This is because stacking originates from London dispersion energy and a theoretical description of this term requires a very large portion of correlation energy to be considered.

The aim of this study was to investigate the nature of the high stability as well as the selectivity for self-pairing of DNA modified by nonpolar base analogues. The analysis will be based on high-level correlated quantum-chemical calculations and molecular dynamics simulations. We suggest that the main contribution to the stability (ΔG) of a modified duplex comes from the interaction energy (basically determined by the dispersion term) between base-pair analogues and neighboring native base pairs. Besides the favorable stacking energy, solvation (desolvation) energies also play a role.

Strategy of Calculations

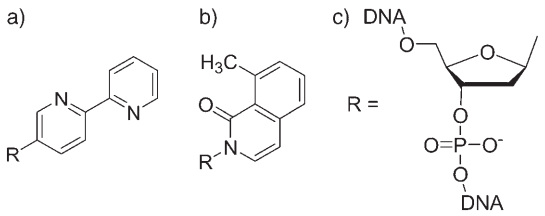
The aim of this study was to determine the change in the free energy of the DNA duplex upon replacement of one canonical base pair by a base-pair analogue. We separately determined the changes in the interaction energy between two strands of a DNA duplex and in the solvation energy of these two strands upon modification of a canonical base pair. The choice of starting geometry is critical and the original idea was to use crystal geometries. However, since crystal structures of DNA duplexes modified by an unnatural base pair do not exist, the relevant structure was generated by performing molecular dynamics (MD) simulations. Two model fragments of DNA duplex were considered: 1) a hexamer (5′-GGA**A**CC-3′) and 2) a 13 mer (5′-GCGTAC**A**-CATGCG-3′). The central AT base pair (shown in bold) of the unperturbed duplexes was replaced by a pair of base analogues. Both duplexes were used for generating geometrical structures and the average geometry of each duplex was determined from 2 ns MD simulation runs. For the subsequent highly accurate quantum chemical calculations of the interaction energies between base analogues and the natural bases, we used a model system (model A) constructed as follows. We considered only three central base pairs and the positions of all six subsystems (determined by MD simulations) were overlaid (using the least-squares fitting method) by their HF/6-31G** optimized geometries. We intentionally overlaid the subsystems and not the base-pair geometries as we wanted to keep the geometrical arrangement of the MD simulations. The subsystem geometries determined at the HF, B3LYP, and MP2 levels differ negligibly. Thus, this model consists of six bases (three base pairs) without sugar–phosphate backbones. A more extended model (model B) was used for calculating the interaction energy between both strands (in this case a lower-level quantum chemical method was utilized) and their solvation/desolvation energies. This model was constructed from the previous one (model A) by adding a sugar–phosphate backbone with an averaged MD geometry. The negative charges of the phosphates were neutralized by adding hydrogen atoms to the phosphate oxygen atoms. This neutralization process is a standard procedure. Its main advantage lies in the fact that the positions of the metal ions are not well-defined since they migrate along the helix. The positions of the bases' atoms were frozen during optimization and only sugar–phosphate backbone atoms were relaxed.

In both cases (2,2′-bipyridyl and 3-methylisocarbostyril) we evaluated the stabilization energies of the native and self-pair structures. Since the conclusions drawn from these

calculations are similar for both duplexes, computationally more demanding and accurate calculations were carried out for both self-pairs and mismatched pairs and in addition to stabilization energies, hydration energies were also determined but only for the 3-methylisocarbostyril base analogue. In this case direct comparison with experimentally determined melting points was also possible.

Methods

Systems studied: In the case of the DNA hexamer, 2,2′-bipyridyl (BP) was used as the base analogue (Scheme 1). The DNA 13-mer was modified by the 3-methylisocarbostyril (MICS) unit[6b] (Scheme 1). Both base analogues are selective in the creation of self-pairs. The 13-mer considered in the latter case was taken directly from the experiments considered, while the hexamer represents the smallest stable duplex model to still reasonably describe the nearest neighborhood of the base analogue investigated.



Scheme 1. Structure of base analogues a) 2,2′-bipyridyl (BP) and b) 3-methylisocarbostyril (MICS).

Molecular dynamics simulations and empirical force field: For the duplexes studied (fragment of DNA + ions + explicit water) MD simulations were performed using the empirical force field of Cornell et al.[15]. Standard parametrization was adopted for natural nucleosides, while for the unnatural nucleosides (which are not covered by standard AMBER parametrization) the missing parameters were derived using the standard (recommended) procedure. The atomic charges were derived using the RESP procedure[16] on the basis of ESP HF/6-31G* calculations performed for isolated unnatural nucleosides. The atomic charges on the deoxyribose moiety were frozen at standard values and only the atomic charges of the base analogue part were calculated. The missing valence and dihedral angle parameters were obtained by standard procedures. After adding Na^+ counterions and explicit water molecules, a standard equilibration procedure was performed. The canonical ensemble MD simulations were then run at 300 K. After 2 ns of simulation the average geometries of all the duplexes were determined.

Correlated ab initio quantum-chemical calculations: The interaction energies of the base pairs and base-analogue pairs were determined for model system A (three base pairs) by means of the approximate resolution of identity MP2 (RI-MP2) method[17–19] using Dunning's aug-cc-pVDZ basis set.[20] It has previously[21] been shown that the absolute as well as the relative RI-MP2 energies of the DNA bases and base pairs differ only marginally from the exact MP2 values, while the time saved is as large as one order of magnitude. The use of an aug-cc-pVDZ basis set (containing diffuse polarization functions) is required for correct description of the stacking interaction. We previously showed that MP2/DZP stabilization energies were significantly underestimated before the first reliable results were obtained using the aug-cc-pVDZ basis set. The basis-set superposition error (BSSE) was systematically removed by using the standard function counterpoise method.[22]

In addition to the RI-MP2 method the popular density functional theory (DFT) method with the B3LYP functional[23] and the 6-31G** basis set was also used.

Several interaction energies within the model used were defined. The total interaction energy (ΔE) was defined as the energy released upon separation of all six subsystems to infinity and in the case of unperturbed complexes, it is the sum of the energies of three hydrogen bonds, four intrastrand and four interstrand stacked contacts. This energy was determined as the difference between the energy of a supersystem consisting of six bases and six isolated subsystems. The basis-set superposition error was removed, that is, the energy of an isolated base was systematically determined using the ghost functions of all the remaining bases. The total interaction energy of the complex covering all two-, three-, four-, five-, and six-body energy contributions is a reliable measure of its stability energy. The interaction energy of the base hexamer was first determined at the correlated ab initio level using an extended basis set. Furthermore, the interaction energies between base pairs and a base pair···base-analogue pair were determined as the energy released when both pairs were separated to infinity. This contribution tells us about the stabilization of both stacked base pairs and it covers all intrastrand and interstrand stacking contributions between the bases. Finally, the interaction energies of hydrogen-bonded or stacked pairs have been determined as the energy released upon separation of two bases to infinity.

SCC-DFTB-D method: The SCC-DFTB-D method[24] is a combination of the approximate tight-binding DFTB method[25] and the empirical dispersion energy. The inclusion of an empirical dispersion term removes the major deficiency of DFT methods, namely the lack of dispersion energy. A more detailed description of this method can be found in the original paper.[24] Our previous studies showed that the DFTB-D method provides surprisingly good estimates of the interaction energies for hydrogen-bonded and stacked structures of DNA base pairs and stacked complexes of base pairs with various intercalators.[26] This method was used for calculating the interaction energies of model A (whose performance was tested by comparing it with accurate quantum chemical data) as well as for calculating the interstrand interaction energy of the larger model B in which accurate quantum chemical calculations are impractical.

Continuum solvent model: The role of solvents was studied for model B containing three base/base-analogue pairs with neutralized sugar–phosphate backbones. Water was represented by a continuum model based on the C-PCM (COSMO) methodology.[27–29] The cavity was described by the united atoms radii optimized at the HF/6-31G** level of theory (UAHF). The COSMO method was used to calculate the interstrand solvation/desolvation free energy. This energy was determined as the difference between the solvation energy of the duplex and the sum of the solvation energies of the isolated strands. The geometry of the duplex and both strands was kept rigid (without optimization) at values obtained from SCC-DFTB-D optimization (see the last paragraph of the section Strategy of Calculations).

Codes: The RI-MP2 and DFT/B3LYP energies were determined using TURBOMOLE 5.6.[30] The MD simulations were performed using AMBER 6.0.[31] SCC-DFTB-D energies were obtained using DFTB codes[25] and the solvation free energies were performed using the continuum solvent model COSMO implemented in Gaussian 03.[32]

Results and Discussion

Geometry of DNA duplexes and model complexes: The final geometries after 2 ns MD simulation of the DNA hexamer and the BP···BP (symbol "···" means that both base analogues are stacked)-modified hexamer are depicted in Figure 1 (structures shown do not correspond to averaged structures, but to just the last snapshot of the respective MD simulations). The DNA hexamer in both simulations was stable [the root-mean-square displacement (RMSD) value

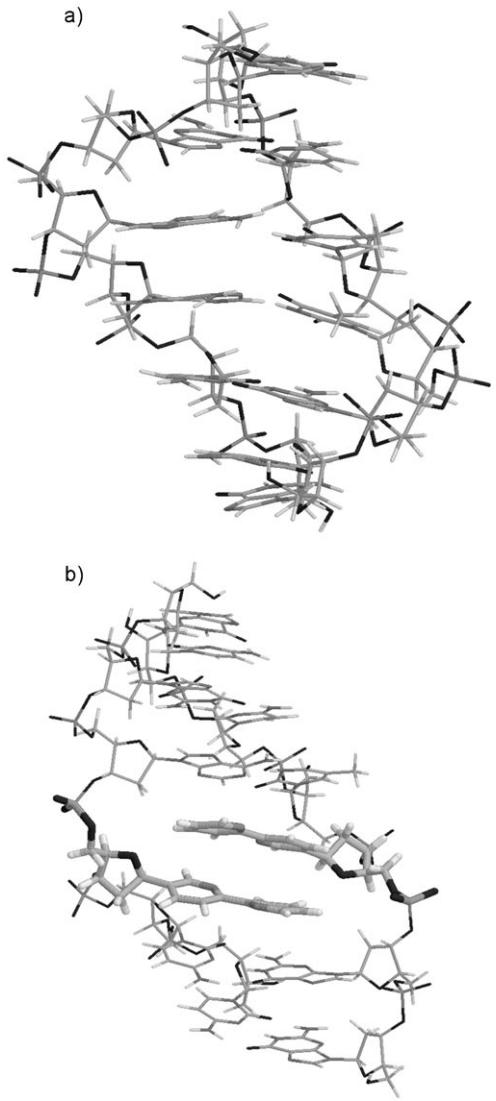


Figure 1. The geometry of the DNA hexamer after 2 ns of MD simulation: a) unperturbed system, b) modified by the BP···BP unnatural base pair.

fluctuated about 1.7 and 2.2, respectively]. The RMSF (root-mean-square fluctuation) per nucleoside (cf. Table S1 in the Supporting Information) show slightly increased fluctuations of the central part of the hexamer after modification by BP. Figure 1a shows that all the base pairs in an unperturbed hexamer are practically planar and that the hexamer possesses a regular double-helical structure. Passing from the natural (central) A···T pair to the unnatural BP pair changes the structure dramatically. Figure 1b shows that the BP···BP pair is overextended and that the DNA hexamer is not flexible enough to accommodate the pair within one plane. The BP base analogues are stacked and the DNA hexamer is partially deformed. The same motif of BP···BP was recently detected experimentally by NMR experiments.[33] However, it must be mentioned that the substituted DNA hexamer (as well as the unperturbed duplex) was stable during 2 ns MD simulations.

In the second step, the geometry of the DNA 13-mer was generated and the final geometry of the unperturbed duplex after 2 ns simulation is shown in Figure 2a (structures shown in Figure 2 do not correspond to averaged structures, but to just the last snapshot of the respective MD simulations). Also, in this case the duplex was stable during simulation (the RMSD value fluctuated about 1.9) and possesses a regular double-helical structure with practically planar base pairs. Replacement of the central A···T base pair by a MICS···MICS base pair (the RMSD value fluctuated about 3.5) leads to significant changes in the duplex geometry (cf. Figure 2b). As in the previous case, the MICS···MICS pair is stacked.[34] The stacked motif of the MICS···MICS system is similar to the stacked adenine···adenine motif formed in the zipper-like DNA duplex.[35]

To investigate the selectivity of MICS self-pairing we also considered the following modification of the native 13-mer. The original DNA was modified by only one MICS molecule. In this case either adenine was replaced by a MICS to form a MICS···T mispair or thymine was replaced by a MICS to create an A···MICS mispair. In both cases the mispairs were planar and were easily accommodated in the DNA duplex. The resulting modified duplexes were stable during 2 ns MD simulations (the RMSD value fluctuated about 3.6 and 3.9, respectively). Also, in this case the fluctuations (see the RMSF values in Table S2 in the Supporting Information) of the central part increased after modification of the 13-mer.

Figure 3 and Figure 4 show model B complexes modified by BP and MICS, respectively. The sugar–phosphate backbones of both duplexes in which the negative charges on the phosphates have been neutralized by the addition of hydrogen atoms were optimized (the positions of the bases and the base analogues were frozen) by the SCC-DFTB-D method. It is seen that in the unperturbed DNA the three base pairs considered are practically planar (Figure 3a and Figure 4a). In the DNA perturbed by one molecule of MICS the central base mispairs (A···MICS and MICS···T) are planar as well (Figure 4b and c) despite the lack of hydrogen bonding between the base and MICS. For the system perturbed by two base analogues (both BP and MICS), the central self-pair is stacked. Stacking of the BP pair (Figure 3b) and the MISC pair (Figure 4d) is comparable and in both cases the unnatural nucleobases overlap significantly.

Interaction energies in model complexes: Table 1 summarizes the interaction energies (see also Figure 5; and Scheme 1) within the AT-AT-CG model A complex. The RI-MP2 total interaction energy of the complex is $-82\,\mathrm{kcal\,mol^{-1}}$. The largest contribution comes from the CG hydrogen-bonded pair ($-26\,\mathrm{kcal\,mol^{-1}}$). The contribution from the AT hydrogen-bonded pairs is smaller ($-13\,\mathrm{kcal\,mol^{-1}}$). The contributions from the stacking interaction between the two pairs have comparable values ($-14\,\mathrm{kcal\,mol^{-1}}$). The stacking energies of the AT/AT and AT/CG pairs are almost identical and

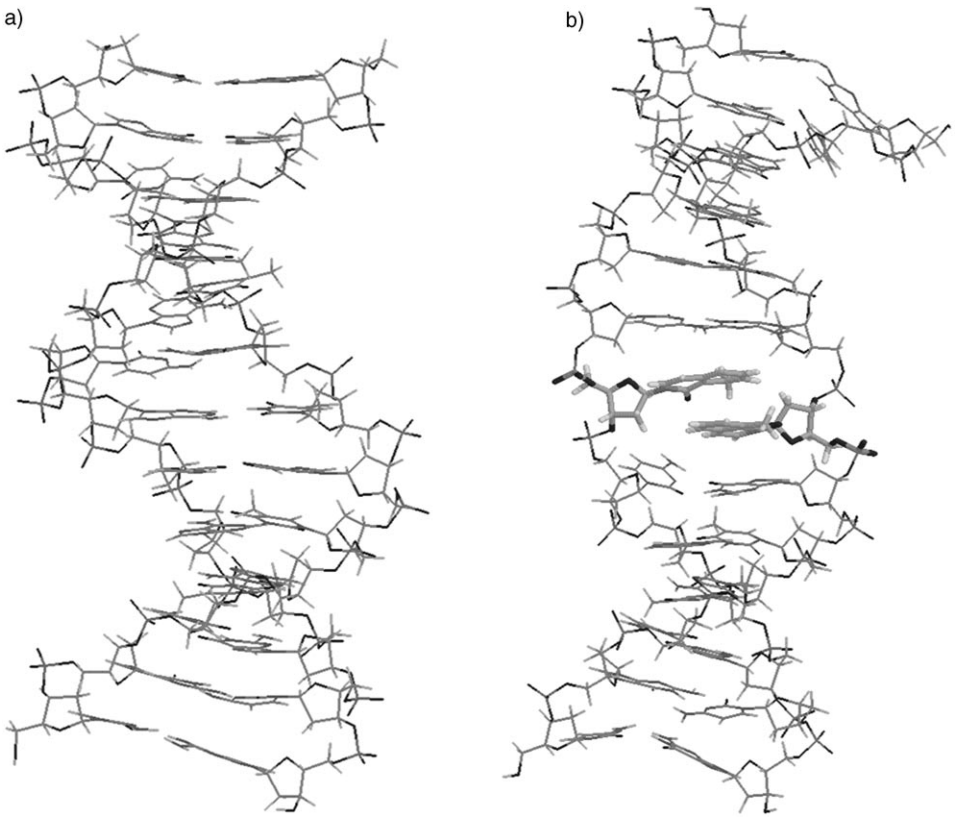


Figure 2. Geometry of the DNA 13-mer after 2 ns of MD simulation: a) unperturbed system, b) modified by the MICS···MICS unnatural base pair.

it should again be noted that these energies are substantial. Combining all the partial interaction energy contributions we obtain $-80.68\,kcal\,mol^{-1}$; the difference between this value and the total interaction energy corresponds to the many-body term. Evidently, this term is small ($1.32\,kcal\,mol^{-1}$) and fully justifies the use of pair interaction energies. By investigating the performance of other lower-level methods, we found that the empirical Cornell[15] and semi-empirical DFTB-D methods yield very good estimates of total as well as partial interaction energies. On the other hand, the nonempirical DFT/B3LYP method fails completely due to the inability of the method to describe base-pair stacking. The absolute error in both cases is very large, approaching $20\,kcal\,mol^{-1}$.

The substitution of the central AT pair by the BP···BP pair resulted (in comparison with the unperturbed system) in a decrease in the total stabilization energy (cf. Table 2). By comparing the RI-MP2 entries of Table 1 and Table 2 we found that the stacking of AT/BP···BP and BP···BP/CG (the symbol "/" dividing the subsystems means that the stacking interaction energy for "AT/BP···BP" was calculated between subsystems AT and BP···BP) is very similar to the stacking in the unperturbed system. Table 2 also shows that the stacking contribution between the AT pair and the remote BP base (cf. Figure 6) is small, less than $0.5\,kcal\,mol^{-1}$. In the case of the CG pair stacked with the remote BP base, this contribution is slightly larger but still less than 1 kcal mol^{-1}. These numbers prove the importance of nearest-neighbor stacking. The AT and CG hydrogen bonding in both clusters is practically identical and the real surprise is the stacking of the BP pair. By comparing the data in Table 1 and Table 2, we found that the stabilization resulting from strong AT hydrogen bonding (unperturbed system) and BP/BP stacking differs by only about $3\,kcal\,mol^{-1}$. This finding also explains why the total stabilization energies of the unperturbed cluster and the BP···BP-modified cluster are rather similar. Furthermore, the data in Table 2 also show that the DFTB-D and Cornell methods yield similar total and partial stabilization energies to the RI-MP2 method (the former method now being closer to the reference data) but the DFT methods fail (again) completely.

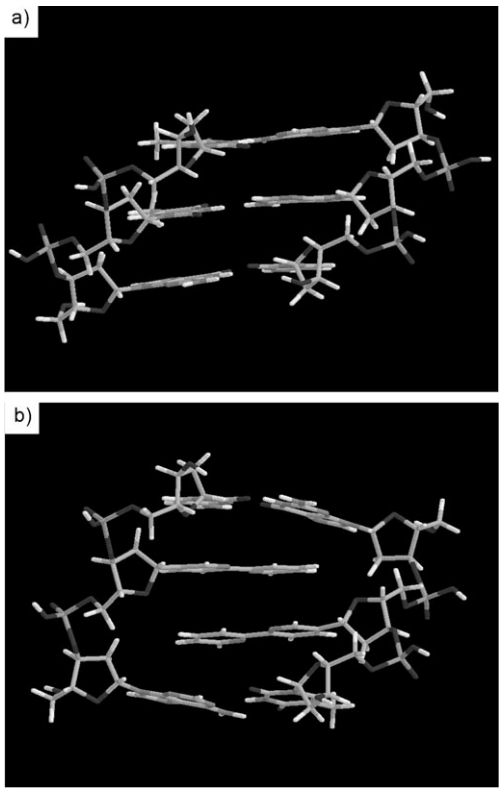


Figure 3. Model of the three central base pairs with neutralized sugar–phosphate backbones (Model B) obtained from the average geometry of the hexamer: a) unperturbed AT-AT-CG , b) self-pair AT-BP···BP-CG.

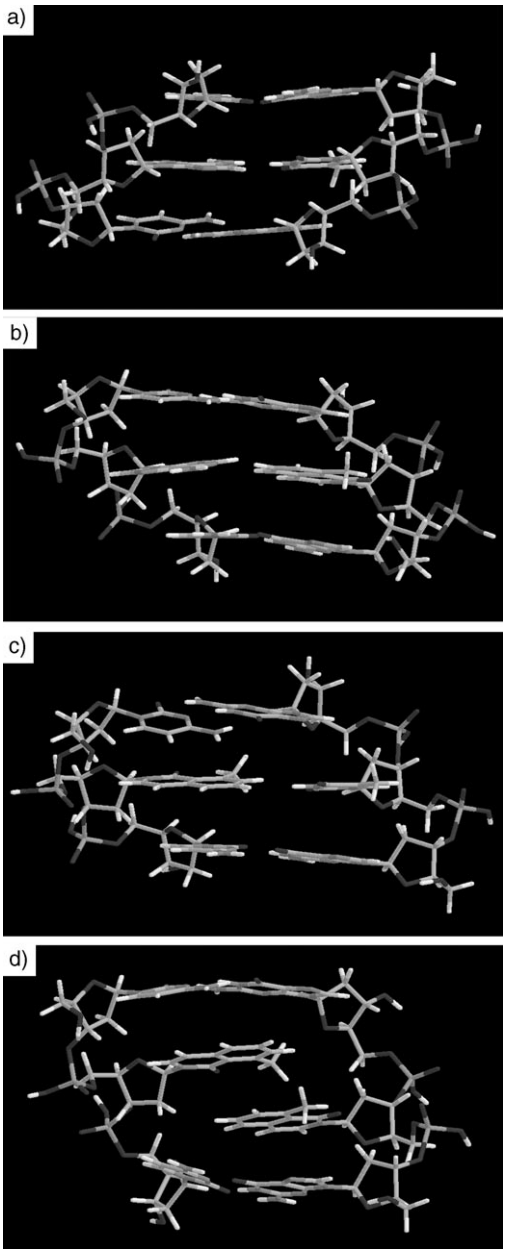


Figure 4. Model of the three central base pairs with neutralized sugar–phosphate backbones (Model B) obtained from the average geometry of the 13-mer: a) unperturbed CG-AT-CG, b) mispair CG-A···MICS-CG, c) mispair CG-MICS···T-CG, d) self-pair CG-MICS···MICS-CG.

Table 1. Interaction energies (in kcal mol^{-1}) for the unperturbed AT-AT-CG system calculated by four different methods.

	AT-AT-CG			
	RI-MP2	DFTB+D	B3LYP	AMBER
total[a]	−82.00	−70.82	−34.9	−78.5
$S1_{(AT/AT)}$[b]	−13.99	−13.42	7.0	−14.6
$S2_{(AT/CG)}$[b]	−14.39	−13.90	6.7	−14.1
$P1_{(AT)}$[b]	−13.10	−10.04	−11.5	−12.0
$P2_{(AT)}$[b]	−13.20	−10.07	−11.6	−12.1
$P3_{(CG)}$[b]	−26.00	−22.50	−25.4	−24.5

[a] Total interaction energy of the system (the difference between the energy of the supersystem containing six bases and the energies of the six isolated subsystems). [b] Specific interaction energies are given in Figure 5.

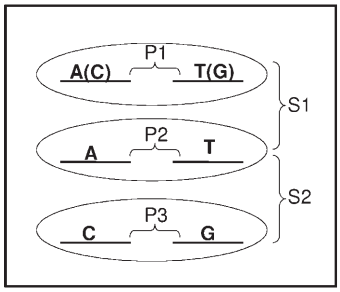


Figure 5. Description of the calculated interaction energies used in Table 1 and Table 3 for the unperturbed systems AT-AT-CG and CG-AT-CG.

Table 2. Interaction energies (in kcal mol^{-1}) for the modified AT-BP···BP-CG system calculated by four different methods.

	AT-BP···BP-AT			
	RI-MP2	DFTB+D	B3LYP	AMBER
total[a]	−76.14	−68.67	−13.8	−65.9
$SX1_{(AT/BP\cdots BP)}$[b]	−13.43	−12.97	6.8	−11.6
$SX2_{(BP\cdots BP/CG)}$[b]	−14.26	−13.56	8.6	−11.0
$S1_{(AT/BP)}$[b]	−13.04	−12.50	6.6	−11.2
$S2_{(BP/CG)}$[b]	−13.41	−13.05	8.4	−10.7
$P1_{(AT)}$[b]	−12.67	−9.90	−11.0	−11.7
$PS2_{(BP/BP)}$[b]	−10.18	−9.39	6.1	−7.3
$P3_{(CG)}$[b]	−25.32	−22.78	−24.9	−24.2

[a] Total interaction energy of the system (the difference between the energy of the supersystem containing six bases and the energies of six isolated subsystems). [b] Specific interaction energies are given in Figure 6.

The total and partial interaction energies in the model A cluster derived from the DNA 13-mer are shown in Table 3. The total stabilization energy of the CG-AT-CG model is 11 kcal mol^{-1} higher than that of the AT-AT-CG model. This increase corresponds to a greater stabilization of the CG pair relative to the AT pair. Other characteristics are similar to the unperturbed AT-AT-CG system. The data in Table 3 show that the interaction energies calculated by the SCC-DFTB-D method are comparable to the RI-MP2 ones.

Substitution of the central AT pair by the MICS···MICS pair decreases the stabilization energy by approximately 12 kcal mol^{-1} (see Table 4). This decrease is due to the smaller stabilization energy of the stacked self-pair (MICS···MICS) compared with the stabilization energy of the hydrogen-bonded AT pair. Furthermore, the smaller stabilization energy of the first CG pair is caused by geometrical deformation of this pair. The base-pair stacking interaction energy and the difference between the interaction of the CG pair with the close and remote MICS base analogue is similar to the case of AT-BP···BP-CG. Also, the SCC-DFTB-D and AMBER methods provide comparable results.

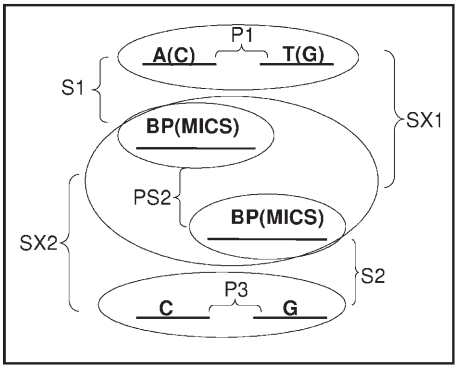


Figure 6. Description of the calculated interaction energies used in Table 2 and Table 4 for the modified systems AT-BP···BP-CG and CG-MICS···MICS-CG.

Table 3. Interaction energies (in kcal mol^{-1}) for the unperturbed CG-AT-CG system (derived from 13-mer) calculated by two different methods.

	CG-AT-CG	
	RI-MP2	DFTB+D
total[a]	−93.64	−83.30
$S1_{(CG/AT)}$[b]	−13.77	−14.04
$S2_{(AT/CG)}$[b]	−14.43	−14.03
$P1_{(CG)}$[b]	−26.03	−22.74
$P2_{(AT)}$[b]	−12.99	−9.73
$P3_{(CG)}$[b]	−26.17	−22.6

[a] Total interaction energy of the system (the difference between the energy of a the supersystem containing six bases and the energies of six isolated subsystems). [b] Specific interaction energies are given in Figure 5

Table 4. Interaction energies (in kcal mol^{-1}) for the modified CG-MICS···MICS-CG (derived from 13-mer) system calculated by two different methods.

	CG-MICS···MICS-CG		
	RI-MP2	DFTB+D	AMBER
total[a]	−81.80	−72.06	−71.4
$SX1_{(CG/MICS\cdots MICS)}$[b]	−13.99	−12.70	−10.4
$SX2_{(MICS\cdots MICS/CG)}$[b]	−13.14	−12.56	−12.6
$S1_{(CG/MCIS)}$[b]	−13.72	−12.31	−10.2
$S2_{(MICS/CG)}$[b]	−12.18	−11.74	−11.7
$P1_{(C\cdots G)}$[b]	−20.87	−15.79	−18.3
$PS2_{(MICS/MICS)}$[b]	−8.24	−8.29	−5.9
$P3_{(C\cdots G)}$[b]	−25.56	−22.72	−24.5

[a] Total interaction energy of the system (the difference between the energy of the supersystem containing six bases and the energies of six isolated subsystems). [b] Specific interaction energies are given in Figure 6.

In the preceding paragraphs we have shown that passing from a canonical base pair to base-analogue pairs reduces the total stabilization energy of a complex. The reduction is, however, only marginal owing to the large stacking stabilization of the base-analogue pair. This conclusion is no doubt important. However, the other experimental finding concerning the selectivity of base-pair analogues, which is comparably significant, remains to be rationalized. To explain it, we evaluated the interstrand interaction and solvation/desolvation energies. We used model B, that is, three base pairs plus sugar–phosphate backbones and we considered the interstrand interaction and desolvation energies. The interstrand interaction energy was determined as the interaction energy between two unrelaxed strands (both strands of model B were separated). Table 5 shows the interstrand interaction and desolvation energies for all four models shown in Figure 4. The interaction energy between the two strands was calculated using the SCC-DFTB-D method whose performance was tested in previous paragraphs. Evidently the unperturbed CG-AT-CG and the perturbed self-pair CG-MICS···MICS-CG system have identical interaction energies, while systems with mispairs are 10 (and 9, respectively) kcal mol^{-1} less stable. This decrease is due to the absence of any interaction between MICS and A (and T, respectively) in the central mispair. Thus, the interstrand interaction energies fully agree with the total interaction energies and show similar stability to the CG-AT-CG and CG-MICS···MICS-CG systems. Furthermore, smaller interstrand stabilization energies for both mispairs support the thermodynamic selectivity of the MICS···MICS self-pair. The interstrand stabilization energies themselves do not, however, explain the experimental finding of a higher melting point (higher stability) for the DNA duplex containing a MICS self-pair.

Table 5. The SCC-DFTB-D interaction and COSMO desolvation energies (in kcal mol^{-1}) between both strands of the duplex for the model of three central bases with neutralized sugar–phosphate backbones.

	ΔE[a]	ΔG_{solv}[b]	$(\Delta E+\Delta G_{solv})$[c]
CG-AT-CG	−66.5	+66.8	+0.3
CG-A···MICS-CG	−56.6	+57.7	+1.1
CG-MICS···T-CG	−57.7	+60.8	+3.1
CG-MICS···MICS-CG	−66.4	+61.7	−4.7

[a] Interstrand interaction energy (the difference between the energy of the duplex and the sum of the energies of the two single strands). [b] Interstrand desolvation energy (the difference between the solvation free energy of the duplex and the sum of the solvation free energies of the two strands). [c] The sum of the interstrand interaction energy and interstrand desolvation energy; estimate of the relative interstrand free energy.

In all the above-mentioned calculations we considered the interaction energy only. Evidently, passing from a highly polar central A···T pair to a practically nonpolar base-analogue pair should result in a different solvation pattern. The interstrand desolvation energy was calculated as the difference between the solvation energy of the duplex and the sum of the solvation energies of the two strands. The geometry of the two single strands was kept rigid (the same as in the duplex). From Table 5 it is clear that the desolvation energy for the unperturbed CG-AT-CG system is 5 kcal mol^{-1} higher than that of the MICS···MICS self-pair system. The explanation is straightforward: MICS, a nonpolar molecule, is less solvated than the polar adenine or thymine. Therefore less energy is required to desolvate strands containing MICS than strands with natural bases. By summing the interaction and desolvation energies (see the last column of Table 5), the CG-MCIS···MICS-CG system be-

comes the most stable. The native CG-AT-CG system is 5 kcal mol^{-1} less stable and the systems with mispairs are 6 and 8 kcal mol^{-1}, respectively, less stable. Note that the final values do not correspond to the absolute changes in the free energies, they merely describe relative values. This means that the CG-MICS···MICS-CG duplex is about 5 kcal mol^{-1} more stable than the CG-AT-CG one. Similarly, positive values for the first three duplexes in Table 5 do not mean destabilization of those duplexes.

Comparison with experimental melting temperatures: The melting-point temperatures of the native 13-mer (59.5 °C), duplexes modified by one MICS (A···MICS and MICS···T, 53.6 and 55.1 °C, respectively), and the MICS self-pair (62.3 °C) were measured by Wu et al.[6b] The trend in the sum of the interstrand and desolvation energies (the last column of Table 5) reproduces the trend in these temperatures reasonably well, especially the highest stability of the duplexes containing the MICS self-pair and the native duplex. The two duplexes containing a single MICS base analogue are less stable and the theoretical data reproduces this trend. (Note that in the experimental system, the MICS-A was measured while we considered a reversed A···MICS system, which we believe would make no significant difference.) Certainly, we cannot expect full agreement between the experimental temperatures and theoretical energies investigated. Probably the most important missing term is the reaction entropy. However, this repulsive term should be similar for all systems considered and cannot qualitatively modify the results.

Conclusion

We have demonstrated that a combination of accurate and approximate quantum chemical calculations and molecular dynamics simulations provides a reliable tool for calculating the interaction energies of short DNA duplexes containing unnatural hydrophobic nucleobases. If the desolvation energy is included, the theoretical results agree well, not only with experimental data concerning the stability of the duplexes (T_m values), but also with the selectivity of base-pair analogues. Thus, the model suggested has a predictive potential and could be used for precalculating the properties of duplexes containing different nucleobases or base pairs and thus help in the rational design of novel nucleobases.

This study has also shown that the self-paired biaryl or bicyclic aromatic unnatural hydrophobic nucleobases do not form planar pairs (i.e. Watson–Crick pairs) but stacked pairs that only slightly distort the duplex. The resulting stacking interaction is substantial and compensates for the missing hydrogen-bonding stabilization. The stability of the modified duplex is further increased (and is comparable or even higher than in the case of canonical pairs) by the considerably lower desolvation energy of hydrophobic nucleobases. In contrast, mismatched pairs of one hydrophobic (MICS) and one canonical (A or T) base stay planar in the duplex with the hydrophobic base distorted slightly out of the stacked system. This distortion and decrease of stacking interaction together with the lack of any other interaction energy causes the dramatic disfavoring of such mismatched pairs and explains why hydrophobic nucleobases are so highly selective and, thus, potentially applicable in the extension of the genetic alphabet.

Acknowledgements

This work is a part of the Research Project Z4055905 and was supported by grants from the Grant Agency of the Czech Republic [Grants 203/05/0043 (to M.H.) and 203/05/0009 (to PH)] and MSMT CR (LC512).

[1] For reviews, see: a) E. T. Kool, *Acc. Chem. Res.* **2002**, *35*, 936–943; b) E. T. Kool, J. C. Morales, K. M. Guckian, *Angew. Chem.* **2000**, *112*, 1046–1068; *Angew. Chem. Int. Ed.* **2000**, *39*, 990–1009; c) A. A. Henry, F. E. Romesberg, *Curr. Opin. Chem. Biol.* **2003**, *7*, 727–733.
[2] J. L. He, F. Seela, *Nucleic Acids Res.* **2002**, *30*, 5485–5496.
[3] a) D. A. Harki, J. D. Graci, V. S. Korneeva, S. K. B. Ghosh, Z. Hong, C. E. Cameron, B. R. Peterson, *Biochemistry* **2002**, *41*, 9026–9033; b) F. Seela, A. Jawalekar, *Helv. Chim. Acta* **2002**, *85*, 1857–1868.
[4] a) L. Wang, P. G. Schultz, *Chem. Commun.* **2002**, 1–11; b) A. A. Henry, F. E. Romesberg, *Curr. Opin. Chem. Biol.* **2003**, *7*, 727–733.
[5] a) J. D. Bain, C. Switzer, A. R. Chamberlin, S. A. Benner, *Nature* **1992**, *356*, 537–539; b) Y. Tor, P. B. Dervan, *J. Am. Chem. Soc.* **1993**, *115*, 4461–4467.
[6] a) A. K. Ogawa, O. K. Abou-Zied, V. Tsui, R. Jimenez, D. A. Case, F. E. Romesberg, *J. Am. Chem. Soc.* **2000**, *122*, 9917–9920; b) Y. Q. Wu, A. K. Ogawa, M. Berger, D. L. Mcminn, P. G. Schultz, F. E. Romesberg, *J. Am. Chem. Soc.* **2000**, *122*, 7621–7632; c) K. M. Guckian, T. R. Krugh, E. T. Kool, *J. Am. Chem. Soc.* **2000**, *122*, 6841–6847; d) J. Parsch, J. W. Engels, *J. Am. Chem. Soc.* **2002**, *124*, 5664–5672; e) J. S. Lai, J. Qu, E. T. Kool, *Angew. Chem.* **2003**, *115*, 6155–6159; *Angew. Chem. Int. Ed.* **2003**, *42*, 5973–5977; f) J. S. Lai, E. T. Kool, *J. Am. Chem. Soc.* **2004**, *126*, 3040–3041; g) S. Matsuda, F. E. Romesberg, *J. Am. Chem. Soc.* **2004**, *126*, 14419–14427.
[7] a) K. Tanaka, M. Shinoya, *J. Org. Chem.* **1999**, *64*, 5002–5003; b) E. Meggers, P. L. Holland, W. B. Tolman, F. E. Romesberg, P. G. Schultz, *J. Am. Chem. Soc.* **2000**, *122*, 10714–10715; c) H. Weizman, Y. Tor, *J. Am. Chem. Soc.* **2001**, *123*, 3375–3376.
[8] a) J. Piccirilli, T. Krauch, S. Moroney, S. Benner, *Nature* **1990**, *343*, 33–37; b) C. Switzer, S. Moroney, S. Benner, *Biochemistry* **1993**, *32*, 10489–10496; c) J. Horlacher, M. Hottiger, V. N. Podust, U. Hubscher, S. A. Benner, *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 6329–6333.
[9] K. M. Guckian, B. A. Schweitzer, R. X. F. Ren, C. J. Sheils, D. C. Tahmassebi, E. T. Kool, *J. Am. Chem. Soc.* **2000**, *122*, 2213–2222.
[10] C. Brotschi, A. Häberli, C. J. Leumann, *Angew. Chem.* **2001**, *113*, 3101–3103; *Angew. Chem. Int. Ed.* **2001**, *40*, 3012–3014.
[11] a) D. L. McMinn, A. K. Ogawa, Y. Q. Wu, J. Q. Liu, P. G. Schultz, F. E. Romesberg, *J. Am. Chem. Soc.* **1999**, *121*, 11585–11586; b) E. L. Tae, Y. Wu, G. Xia, P. G. Schultz, F. E. Romesberg, *J. Am. Chem. Soc.* **2001**, *123*, 7439–7440; c) A. A. Henry, C. Z. Yu, F. E. Romesberg, *J. Am. Chem. Soc.* **2003**, *125*, 9638–9646; d) M. Fa, A. Radeghieri, A. A. Henry, F. E. Romesberg, *J. Am. Chem. Soc.* **2004**, *126*, 1748–1754; e) M. Kimoto, M. Endo, T. Mitsui, T. Okuni, I. Hirao, S. Yokoyama, *Chem. Biol.* **2004**, *11*, 47–55; f) J. C. Delaney, P. T. Henderson, S. A. Helquist, J. C. Morales, J. M. Essigmann, E. T. Kool, *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 4469–4473; g) A. A. Henry, A. G. Olsen, S. Matsuda, C. Yu, B. H. Geierstanger, F. E. Romesberg, *J. Am. Chem. Soc.* **2004**, *126*, 6923–6931.
[12] Y. Q. Wu, M. Fa, E. L. Tae, P. G. Schultz, F. E. Romesberg, *J. Am. Chem. Soc.* **2002**, *124*, 14626–14630.

[13] I. Hirao, T. Ohtsuki, T. Fujiwara, T. Mitsui, T. Yokogawa, T. Okuni, H. Nakayama, K. Takio, T. Yabuki, T. Kigawa, K. Kodama, T. Yokogawa, K. Nishikawa, S. Yokoyama, *Nat. Biotechnol.* **2002**, *20*, 177–182.
[14] J. Šponer, J. Leszczynski, P. Hobza, *J. Biomol. Struct. Dyn.* **1996**, *14*, 117–135.
[15] 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.
[16] C. Bayly, P. Cieplak, W. D. Cornell, P. A. Kollman, *J. Phys. Chem.* **1993**, *97*, 10269–10280.
[17] M. Feyereisen, G. Fitzgerald, A. Komornicki, *Chem. Phys. Lett.* **1993**, *208*, 359.
[18] O. Vahtras, J. Almlöf, M. Feyereisen, *Chem. Phys. Lett.* **1993**, *213*, 514.
[19] D. E. Bernholdt, R. J. Harrison, *Chem. Phys. Lett.* **1996**, *250*, 470.
[20] T. H. Dunning, Jr., *J. Chem. Phys.* **1989**, *90*, 1007.
[21] P. Jureèka, P. Nachtigall, P. Hobza, *Phys. Chem. Chem. Phys.* **2001**, *3*, 4578.
[22] S. F. Boys, F. Bernardi, *Mol. Phys.* **1970**, *19*, 553.
[23] A. D. Becke, *Phys. Rev. A* **1988**, *38*, 3098–3100.
[24] M. Elstner, P. Hobza, T. Frauenheim, S. Suhai, E. Kaxiras, *J. Chem. Phys.* **2001**, *114*, 5149.
[25] M. Elstner, T. Frauenheim, E. Kaxiras, G. Seifert, S. Suhai, *Phys. Status Solidi B* **2000**, *217*, 357.
[26] D. Řeha, M. Kabeláè, F. Ryjáèek, J. Šponer, J. E. Šponer, M. Elstner, S. Suhai, P. Hobza, *J. Am. Chem. Soc.* **2002**, *124*, 3366–3376.
[27] V. Barone, M. Cossi, J. Tomasi, *J. Comput. Chem.* **1998**, *19*, 404–417.
[28] A. Klamt, G. J. P. Krooshof, R. Taylor, *AICHE J.* **2002**, *48*, 2332–2349.
[29] M. Cossi, N. Rega, G. Scalmani, V. Barone, *J. Comput. Chem.* **2003**, *24*, 669–681.
[30] R. Ahlrichs, M. Bär, M. Häser, H. Horn, C. Kölmel, *Chem. Phys. Lett.* **1989**, *162*, 165.
[31] AMBER 6.0, D. A. Case, D. A. Pearlman, J. W. Caldwell, T. E. Cheatham III, W. S. Ross, C. L. Simmerling, T. A. Darden, K. M. Merz, R. V. Stanton, A. L. Cheng, J. J. Vincent, M. Crowley, V. Tsui, R. J. Radmer, Y. Duan, J. Pitera, I. Massova, G. L. Seibel, U. C. Singh, P. K. Weiner; P. A. Kollman, University of California, San Francisco, **1999**.
[32] Gaussian 03, Revision A.1, 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, 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., Pittsburgh, PA, **2003**..
[33] C. Brotschi, G. Mathis, C. J. Leumann, *Chem. Eur. J.* **2005**, *11*, 1911–1923.
[34] The NMR structure of a duplex containing a related PICS···PICS base pair confirms this geometry: see ref. [3b] and F. E. Romesberg unpublished results.
[35] N. Špaèková, I. Berger, J. Šponer, *J. Am. Chem. Soc.* **2000**, *122*, 7564–7572.

Received: September 11, 2005
Revised: December 15, 2005
Published online: February 27, 2006

