

Accurate Interaction Energies of Hydrogen-Bonded Nucleic Acid Base Pairs

Jiří Šponer,^{*,†,‡} Petr Jurečka,[‡] and Pavel Hobza^{*,‡}

Contribution from the Institute of Biophysics, Academy of Sciences of the Czech Republic, Kralovopolska 135, 612 65 Brno, Czech Republic and Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, Flemingovo nam. 2, 166 10 Prague, Czech Republic

Received March 18, 2004; E-mail: sponer@ncbr.chemi.muni.cz; pavel.hobza@uochb.cas.cz

Abstract: Hydrogen-bonded nucleic acids base pairs substantially contribute to the structure and stability of nucleic acids. The study presents reference ab initio structures and interaction energies of selected base pairs with binding energies ranging from -5 to -47 kcal/mol. The molecular structures are obtained using the RI-MP2 (resolution of identity MP2) method with extended cc-pVTZ basis set of atomic orbitals. The RI-MP2 method provides results essentially identical with the standard MP2 method. The interaction energies are calculated using the Complete Basis Set (CBS) extrapolation at the RI-MP2 level. For some base pairs, Coupled-Cluster corrections with inclusion of noniterative triple contributions (CCSD(T)) are given. The calculations are compared with selected medium quality methods. The PW91 DFT functional with the 6-31G** basis set matches well the RI-MP2/CBS absolute interaction energies and reproduces the relative values of base pairing energies with a maximum relative error of 2.6 kcal/mol when applied with Becke3LYP-optimized geometries. The Becke3LYP DFT functional underestimates the interaction energies by few kcal/mol with relative error of 2.2 kcal/mol. Very good performance of nonpolarizable Cornell et al. force field is confirmed and this indirectly supports the view that H-bonded base pairs are primarily stabilized by electrostatic interactions.

Introduction

The structure and dynamics of nucleic acid molecules are influenced by a variety of contributions. Among those, the interactions present between the nucleic acid base heterocycles are of particular importance. In DNA and RNA, the bases are involved in two qualitatively different mutual interaction types: hydrogen bonding and aromatic base stacking. The H-bonded base pair geometries observed at high resolution in crystal structures of DNA fragments correspond to the (local) minima on potential energy surfaces determined with minimal basis set of atomic orbitals, step by step structure optimization, and empirical London dispersion correction.¹ Similarly extended empirical potential study was presented by Poltev and Shlyupina.² The next improvement in the theoretical description of base pairing was achieved in 1994 when Gould and Kollman reported the first medium quality calculations on four base pairs with inclusion of electron correlation effects.³ This work was followed by extensive electron correlation study on dozens of base pairs by Šponer et al.⁴ These results provided reference

data for verification and parametrization of other computational techniques for almost a decade. The calculations also revealed intrinsic nonplanarity of many base pairs.⁵ Hobza et al. demonstrated good performance of well-parametrized empirical force fields, while semiempirical techniques were considerably less successful.⁶ A number of other QM studies contributed to our knowledge of nucleic acids base pairing including their interactions with metal cations,⁷ base triples, quadruples, and water inserted pairs,⁸ thio-base pairs and other modified base pairs,⁹ protonated base pairs,¹⁰ proton-transfer in base pairs,¹¹ base pair radical cations and anions,¹² base pairing in aqueous solution and micro-hydrated environments,¹³ gas-phase clusters,¹⁴ base pairs in excited state¹⁵ and many others.¹⁶ It is

[†] Institute of Biophysics, Academy of Sciences of the Czech Republic.

[‡] Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic.

(1) Hobza, P.; Sandorfy, C. *J. Am. Chem. Soc.* **1987**, *109*, 1302–1307.

(2) Poltev, V. I.; Shlyupina, N. V. *J. Biomol. Struct. Dyn.* **1986**, *3*, 739–765.

(3) (a) Gould, I. R.; Kollman, P. A. *J. Am. Chem. Soc.* **1994**, *116*, 2493–2499. (b) 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*, 5179–5197.

(4) (a) Šponer, J.; Leszczynski, J.; Hobza, P. *J. Phys. Chem.* **1996**, *100*, 1965–1974. (b) Hobza, P.; Šponer, J. *Chem. Rev.* **1999**, *99*, 3247–3276.

(5) (a) Šponer, J.; Florián, J.; Leszczynski, J.; Hobza, P. *J. Biomol. Struct. Dyn.* **1996**, *13*, 827–833. (b) Šponer, J.; Mokdad, A.; Šponer, J. E.; Špačková, N.; Leszczynski, J.; Leontis, N. B. *J. Mol. Biol.* **2003**, *330*, 967–978.

(6) Hobza, P.; Kabeláč, M.; Šponer, J.; Mejzlík, P.; Vondrášek, J. *J. Comput. Chem.* **1997**, *18*, 1136–1150.

(7) (a) Burda, J. V.; Šponer, J.; Leszczynski, J.; Hobza, P. *J. Phys. Chem. B* **1997**, *101*, 9670–9677. (b) Šponer, J.; Burda, J. V.; Sabat, M.; Leszczynski, J.; Hobza, P. *J. Phys. Chem. A* **1998**, *102*, 5951–5957. (c) Munoz, J.; Gelpi, J. L.; Soler-Lopez, M.; Subirana, J. A.; Orozco, M.; Luque, F. J. *J. Phys. Chem. B* **2002**, *106*, 8849–8857. (d) Moroni, F.; Famulari, A.; Raimondi, M.; Sabat, M. *J. Phys. Chem. B* **2003**, *107*, 496–500. (e) Guerra, C. F.; Bickelhaupt, F. M.; Snijders, J. G.; Baerends, E. J. *J. Am. Chem. Soc.* **2000**, *122*, 4117–4128. (f) Gresh, N.; Šponer, J. *J. Phys. Chem. B* **1999**, *103*, 11415–11427. (g) Munoz, J.; Šponer, J.; Hobza, P.; Orozco, M.; Luque, F. J. *J. Phys. Chem. B* **2001**, *105*, 6051–6060. (h) Pelmenchikov, A.; Zilberberg, I.; Leszczynski, J.; Famulari, A.; Sironi, M.; Raimondi, M. *Chem. Phys. Lett.* **1999**, *314*, 496–500.

entirely outside the scope of the introduction to mention all valuable studies.

Substantial computer advances in recent years allowed us to further improve quantum-chemical calculations of base pairing. Second-order Moeller–Plesset (MP2) calculations with extended basis sets of atomic orbitals indicated a systematic convergence of the results starting with inclusion of f-polarization functions.¹⁷ The full convergence for H-bonded base pairs is expected to occur with basis sets of a cc-pV5Z quality.¹⁷ Such calculations are not routinely feasible with standard MP2 method. However, Jurečka et al. verified that resolution of identity MP2 (RI-MP2)¹⁸ procedure provides essentially identical results as the MP2 method with a fraction of computational resources.¹⁹ This opens a feasible route to large-scale MP2-level electron correlation calculations of base pairing. Importantly, the Coupled Cluster

calculations carried out so far indicate that higher-order electron correlation corrections for H-bonding energies are usually small.²⁰ It does not guarantee that the higher-order terms are negligible, but there is a substantial degree of compensation of errors for hydrogen bonding. Thus, the key step in the accurate description of base pairing is the expansion of the MP2-level evaluations to the basis set limit. Similar advances in quantum chemical studies were reported also for the aromatic base stacking.^{21,22} Studies of base stacking are more complicated due to substantial magnitude of higher-order electron correlation effects²² and lack of well-defined minima on the potential energy surface.²¹

In this paper, structures and interaction energies of a number of different nucleic acid base pairs are presented. The molecular structures are obtained using the RI-MP2 method with extended cc-pVTZ²³ or TZVPP^{18,19} basis sets of atomic orbitals (both basis sets provide essentially identical geometries). The RI-MP2 method provides close to identical results compared with the standard MP2 method, within 0.03 kcal/mol for the interaction energies.¹⁹ Thus, the present structures are equivalent to those that would be obtained at the MP2/cc-pVTZ level. The interaction energies are calculated using the Complete Basis Set extrapolation at the RI-MP2 level with aug-cc-pVDZ, aug-cc-pVTZ, and in some cases aug-cc-pVQZ basis sets of atomic orbitals. For some base pairs, Coupled-Cluster corrections with inclusion of noniterative triplet contributions (CCSD(T)) are given.

The studied base pairs include a wide range of base pairing patterns. In contrast to preceding studies, we do not evaluate all possible combinations of the standard bases. Instead, we include many modified base pairs. Thus, we consider a much wider range of distinct interbase H-bonds. Yet, our study is far from being complete as H-bonding of bases may occur in hundreds of combinations and very often, especially in RNA, involves sugar OH groups, inserted waters and other interactions.

As explained below, the present data supports the prevailing view that base pairing is dominated by the electrostatic interactions. The calculations are approaching complete inclusion of intermolecular electron correlation (dispersion) effects. Not surprisingly, the relative role of this term increases significantly compared to studies with medium basis sets.⁴ Last but not the least, the RI-MP2/cc-pVTZ optimization (compared to SCF and especially force field optimizations) increases the importance of nucleobase intramolecular deformations upon the base pairing. This leads to the improvement of the intermolecular energy terms due to intramolecular geometry adjustments. The most important deformation is the elongation of the N–H bonds participating in the H-bonding. The associated energy improvements, however, are largely canceled by the monomer deformation energies. Nevertheless, proper relaxation of monomers is important to obtain a balanced description of the interactions. Thus gradient optimization is preferred over rigid monomer

- (8) (a) Šponer, J.; Burda, J. V.; Mejzlík, P.; Leszczynski, J.; Hobza, P. *J. Biomol. Struct. Dyn.* **1997**, *14*, 613–628. (b) Brandl, M.; Meyer, M.; Suhnel, J. *J. Am. Chem. Soc.* **1999**, *121*, 2605–2606. (c) Meyer, M.; Steinke, T.; Brandl, M.; Suhnel, J. *J. Comput. Chem.* **2001**, *22*, 109–124. (d) Peters, M.; Rozas, I.; Alkorta, I.; Elguero, J. *J. Phys. Chem. B* **2003**, *107*, 323–330. (e) Gu, J. D.; Leszczynski, J. *J. Phys. Chem. A* **2002**, *106*, 529–532. (f) Louit, G.; Hocquet, A.; Ghomi, M.; Meyer, M.; Suhnel, J. *Phys. Chem. Commun.* **2003**, 15–23.
- (9) (a) Šponer, J.; Leszczynski, J.; Hobza, P. *J. Phys. Chem. A* **1997**, *101*, 9489–9495. (b) Meyer, M.; Suhnel, J. *J. Biomol. Struct. Dyn.* **1997**, *15*, 619–624. (c) Barsky, D.; Kool, E. T.; Colvin, M. E. *J. Biomol. Struct. Dyn.* **1999**, *16*, 1119–1134. (d) Podolyan, Y.; Rubin, Y. V.; Leszczynski, J. *J. Phys. Chem. A* **2000**, *104*, 9964–9970. (e) Contreras, J. G.; Madariaga, S. T. *Bioorganic Chem.* **2003**, *31*, 367–377.
- (10) Šponer, J.; Leszczynski, J.; Vetterl, V.; Hobza, P. *J. Biomol. Struct. Dyn.* **1996**, *13*, 695–707.
- (11) (a) Florián, J.; Leszczynski, J. *J. Am. Chem. Soc.* **1996**, *118*, 3010–3017. (b) Bertran, J.; Oliva, A.; Rodríguez-Santiago, L.; Sodupe, M. *J. Am. Chem. Soc.* **1998**, *120*, 8159–8167. (c) Bertran, J.; Noguera, M.; Sodupe, M. *Afinidad* **2002**, *59*, 470–478.
- (12) (a) Colson, A. O.; Sevilla, M. D. *Int. J. Radiat. Biol.* **1995**, *67*, 627–645. (b) Li, X. F.; Cai, Z. L.; Sevilla, M. D. *J. Phys. Chem. A* **2002**, *106*, 9345–9351. (c) Richardson, N. A.; Wesolowski, S. S.; Schaefer, H. F. *J. Am. Chem. Soc.* **2002**, *124*, 10 163–10 170. (d) Richardson, N. A.; Wesolowski, S. S.; Schaefer, H. F. *J. Phys. Chem. B* **2003**, *107*, 848–853. (e) Hutter, M.; Clark, T. *J. Am. Chem. Soc.* **1996**, *118*, 7574–7577. (f) Reynisson, J.; Steenken, S. *Phys. Chem. Chem. Phys.* **2002**, *4*, 5346–5352. (g) Reynisson, J.; Steenken, S. *Phys. Chem. Chem. Phys.* **2002**, *4*, 5353–5358.
- (13) (a) Florián, J.; Šponer, J.; Warshel, A. *J. Phys. Chem. B* **1999**, *103*, 884–892. (b) Zhanpeisov, N. U.; Leszczynski, J. *J. Phys. Chem. A* **1998**, *102*, 6167–6172. (c) Kabeláč, M.; Hobza, P. *Chem. Eur. J.* **2001**, *7*, 2056–2074. (d) Sivanesan, D.; Babu, K.; Gadre, S. R.; Subramanian, V.; Ramasami, T. *J. Phys. Chem. A* **2000**, *104*, 10 887–10 894.
- (14) (a) Kratochvíl, M.; Šponer, J.; Hobza, P. *J. Am. Chem. Soc.* **2000**, *122*, 3495–3499. (b) Nir, E.; Janzen, C.; Imhof, P.; Kleineremanns, K.; de Vries, M. S. *Phys. Chem. Chem. Phys.* **2000**, *2*, 732–739. (c) Müller, A.; Losada, M.; Leutwyler, S. *J. Phys. Chem. A* **2004**, *108*, 157–165. (d) Nir, E.; Kleineremanns, K.; de Vries, M. S. *Nature* **2000**, *408*, 949–951. (e) Müller, A.; Talbot, F.; Leutwyler, S. *J. Am. Chem. Soc.* **2002**, *124*, 14 486–14 494.
- (15) (a) Shukla, M. K.; Leszczynski, J. *J. Phys. Chem. A* **2002**, *106*, 1011–1018. (b) Shukla, M. K.; Leszczynski, J. *J. Phys. Chem. A* **2002**, *106*, 4709–4717.
- (16) (a) Pranata, J.; Wierschke, S. G.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1991**, *113*, 2810–2819. (b) Guerra, C. F.; Bickelhaupt, F. M.; Snijders, J. G.; Baerends, E. J. *Chem. Eur. J.* **1999**, *5*, 3581–3594. (c) Guerra, C. F.; Bickelhaupt, F. M.; Baerends, E. J. *Cryst. Growth Des.* **2002**, *2*, 239–245. (d) Brameld, K.; Dasgupta, S.; Goddard III, W. A. *J. Phys. Chem. B* **1997**, *101*, 4851–4859. (e) Sherer, E. C.; York, D. M.; Cramer, C. J. *J. Comput. Chem.* **2003**, *24*, 57–67. (f) Kryachko, E. S.; Sabin, J. R. *Int. J. Quantum Chem.* **2001**, *82*, 193–204. (g) Kryachko, E. S.; Sabin, J. R. *Int. J. Quantum Chem.* **2003**, *91*, 695–710. (h) Kawahara, S.; Uchimaru, T.; Taira, K.; Sekine, M. *J. Phys. Chem. A* **2002**, *106*, 3207–3215. (i) Kawahara, S.; Uchimaru, T.; Taira, K.; Sekine, M. *J. Phys. Chem. A* **2002**, *106*, 3207–3212. (j) Kawahara, S.; Kobori, A.; Sekine, M.; Taira, K.; Uchimaru, T. *J. Phys. Chem. A* **2001**, *105*, 10 596–10 601. (k) Kawahara, S.; Uchimaru, T. *Eur. J. Org. Chem.* **2003**, 2577–2584. (l) Czernek, J.; Fiala, R.; Sklenar, V. *J. Magn. Reson.* **2000**, *145*, 142–146. (m) Dingley, A. J.; Feigon, J.; Grzesiek, S. *J. Am. Chem. Soc.* **2001**, *123*, 4014–4022. (n) Joubert, L.; Popelier, P. L. A. *Phys. Chem. Chem. Phys.* **2002**, *4*, 4353–4359. (o) Asensio, A.; Kobko, N.; Dannenberg, J. J. *J. Phys. Chem. A* **2003**, *107*, 6441–6443. (p) Forde, G.; Flood, A.; Salter, L. Hill, G.; Gorb, L.; Leszczynski, J. *J. Biomol. Struct. Dyn.* **2003**, *20*, 811–817.
- (17) Šponer, J.; Hobza, P. *J. Phys. Chem. A* **2000**, *104*, 4592–4597.
- (18) (a) Feyereisen, M.; Fitzgerald, G.; Komornicki, A. *Chem. Phys. Lett.* **1993**, *208*, 359–363. (b) Weigend, F.; Häser, M. *Theor. Chem. Acc.* **1997**, *97*, 331–340.
- (19) Jurečka, P.; Nachtigall, P.; Hobza, P. *Phys. Chem. Chem. Phys.* **2001**, *3*, 4578–4582.
- (20) (a) Hobza, P.; Šponer, J. *Theochem-J. Mol. Struct.* **1996**, *388*, 115–120. (b) Jurečka, P.; Hobza, P. *Chem. Phys. Lett.* **2002**, *365*, 89–94.
- (21) (a) Šponer, J.; Leszczynski, J.; Hobza, P. *J. Phys. Chem.* **1996**, *100*, 5590–5596. (b) Šponer, J.; Gabb, H. A.; Leszczynski, J.; Hobza, P. *Biophys. J.* **1997**, *73*, 76–87. (c) Hernandez, B.; Luque, F. J.; Orozco, M. *J. Comput. Chem.* **1999**, *20*, 937–946. (d) Biot, C.; Buisine, E.; Kwasigroch, J. M.; Wintjens, R.; Rooman, M. *J. Biol. Chem.* **2002**, *277*, 40 816–40 822.
- (22) (a) Šponer, J.; Hobza, P. *Chem. Phys. Lett.* **1997**, *267*, 263–270. (b) Leininger, M. L.; Nielsen, I. M. B.; Colvin, M. E.; Janssen, C. L. *J. Phys. Chem. A* **2002**, *106*, 3850–3854. (c) Hobza, P.; Šponer, J. *J. Am. Chem. Soc.* **2002**, *124*, 11 802–11 808.
- (23) Dunning, T. H. *J. Chem. Phys.* **1989**, *90*, 1007–1023.

approaches. This is especially evident for strong and nonplanar base pairs where the recently suggested rigid monomer approach²⁴ would be insufficient.

Comparison with other methods confirms that well selected medium-level techniques are reasonably accurate for base pairing and can be safely used to study larger H-bonded model clusters of biochemical and biological relevance. Also well parametrized force fields (such as the Cornell et al. force field^{3b}) provide meaningful estimates of base pair H-bonding energies. Interestingly, the overall agreement of the new reference data with the most widely used Cornell et al. molecular mechanical force field is improved compared with the medium-level QM calculations.⁴ Thus, the base–base terms belong to the most successfully approximated contributions in explicit-solvent molecular dynamics simulations.

Methods

The TURBOMOLE 5.6²⁵ program was used for geometry optimizations and all single point RI-MP2 calculations. Geometry optimizations were carried out using cc-pVTZ or TZVPP basis sets without inclusion of any BSSE (basis set superposition error) correction. Both extended basis sets provide essentially identical geometries, and any subsequent differences in the interaction energies are negligible, much smaller than those caused by the uncorrected BSSE during optimization and other factors including the remaining minor uncertainties in the extrapolation techniques. (The cc-pVTZ basis set was not available for the RI-MP2 method at the beginning of the project). RI-MP2 interaction energies were calculated with aug-cc-pVXZ ($X = D, T, Q$) basis sets and corrected for BSSE according to Boys and Bernardi.²⁶ Extrapolation to the complete basis set limit was employed to overcome the very slow convergence of the correlation energy. Two different extrapolation schemes were used. In the first scheme, HF and MP2 single point energies were extrapolated according to Helgaker and co-workers^{27,28}

$$E_X^{\text{HF}} = E_{\text{CBS}}^{\text{HF}} + Ae^{-\alpha X} \text{ and } E_X^{\text{corr}} = E_{\text{CBS}}^{\text{corr}} + BX^{-3}$$

where E_X is the energy obtained with the basis set with cardinal number X ($X = 2$ for DZ, 3 for TZ, ...), E_{CBS} are energies at the basis set limit and A , B , and α are parameters ($\alpha = 1.43$ and 1.54 for D→T and T→Q extrapolations were taken from the literature²⁷ whereas A and B are to be fitted).

In the second scheme by Truhlar,²⁹ HF and MP2 energies are extrapolated according to forms

$$E_X^{\text{HF}} = E_{\text{CBS}}^{\text{HF}} + BX^{-\alpha} \text{ and } E_X^{\text{corr}} = E_{\text{CBS}}^{\text{corr}} + BX^{-\beta}$$

where $\alpha = 3.4$ and $\beta = 2.2$ for D→T²⁹ and the meaning of remaining symbols remains unchanged.

The interaction energy ΔE of a dimer $A...B$ is defined as the electronic energy difference between the dimer ($E^{A...B}$) and the isolated monomers (E^A , E^B). The monomer energies are computed in the basis set of the dimer (dimer-centered basis set) and assuming the geometries of the optimized dimer.^{4,26a} Thus, the results are corrected for the mathematical artifact called basis set superposition error (BSSE).^{26a} When calculating the binding energies it is important to further add the deformation energy E_{Def} .⁴ The deformation energy is a repulsive

contribution due to changes of the monomer geometries upon the complex formation. It is evaluated as the energy difference between the monomers adopting the final deformed geometry (as adjusted in the complex) and relaxed isolated monomers, all evaluated with the monomer basis set.⁴ In summary, the interaction energy is defined in the following way.⁴

$$\Delta E^{A...B} = E^{A...B} - (E^A + E^B) + E_{\text{Def}}$$

In this study, we calculate the deformation energy based on the CBS extrapolation where relevant and with respect to the *nonplanar*, i.e., fully optimized monomers, even for base pairs having the C_s symmetry. Note that in some preceding studies, the deformation energy has been calculated with respect to the planar monomers, thus neglecting the amino group nonplanarity.⁴ These two numbers differ simply by the difference between energies of planar and nonplanar monomers and thus can be easily compared when needed. It is to be noted that it is common in some studies to formally include the deformation energy as a part of the BSSE correction.^{26b} However, although it might look more consistent mathematically, we do not suggest this approach for larger systems such as base pairs and other fragments of biopolymers. The reasons are as follows. First, the integrated expression^{26b} is, after formal rearrangements, *entirely identical* to our expression.^{4a} Second, while BSSE is a mathematical artifact, monomer deformations represent real effects related to fundamental properties of the studied clusters including their vibrational spectra and polarization/charge-transfer effects. Thus, it is quite useful to evaluate the magnitude of the monomer deformations explicitly. Further, for flexible systems (for example many non-Watson–Crick interactions involving also the sugar–phosphate atoms) the changes of geometries upon complex formation include substantial conformational changes and thus any formal inclusion of the deformation term into the BSSE correction would be meaningless.

Higher-order contributions to the correlation energy were taken into account by adding $\Delta\text{CCSD(T)}$ correction^{4b} to the RI-MP2 CBS energy. $\Delta\text{CCSD(T)}$ term is the difference between the MP2 and CCSD(T) interaction energies (CCSD(T) stands for Coupled Clusters Singles and Doubles with perturbative treatment of Triple excitations). This term was evaluated using relatively small basis set (6-31G*) due to its enormous computer requirements. This correction nevertheless is relatively insensitive to the size of the basis set and the 6-31G* basis set is thus sufficient for the purpose of this study. Use of small basis set is based on our previous results²⁰ showing that while MP2 and CCSD(T) interaction energies depend strongly on the size of AO basis set their respective difference is practically basis set independent. Thus quite reliable values of this correction are obtained with small basis sets. All CCSD(T) calculations were carried out using MOLPRO 2002.6³⁰ program. DFT energies were calculated using medium size 6-31G** basis set with Gaussian98 program package and standard B3LYP and PW91 functionals.³¹ Amber 6.0³² program with parm94

(24) Toczyłowski, R. R.; Cybulski, S. M. *J. Phys. Chem. A* **2003**, *107*, 418–426.

(25) Ahlrichs, R.; Bär, M.; Häser, M.; Horn, H.; Kölmel, C. *Chem. Phys. Lett.* **1989**, *162*, 165–169.

(26) (a) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553–575. (b) Xantheas, S. S. *J. Chem. Phys.* **1996**, *104*, 8821–8824.

(27) Halkier, A.; Helgaker, T.; Jørgensen, P.; Klopper, W.; Koch, H.; Olsen, J. *Chem. Phys. Lett.* **1999**, *302*, 437–446.

(28) Halkier, A.; Helgaker, T.; Jørgensen, P.; Klopper, W.; Koch, H.; Olsen, J.; Wilson, A. K. *Chem. Phys. Lett.* **1998**, *286*, 243–252.

(29) Truhlar, D. G. *Chem. Phys. Lett.* **1998**, *294*, 45–48.

(30) MOLPRO, a package of ab initio programs designed by H.-J. Werner and P. J. Knowles, version 2002.1, R. D. Amos, A. Bernhardsson, A. Berning, P. Celani, D. L. Cooper, M. J. O. Deegan, A. J. Dobbyn, F. Eckert, C. Hampel, G. Hetzer, P. J. Knowles, T. Korona, R. Lindh, A. W. Lloyd, S. J. McNicholas, F. R. Manby, W. Meyer, M. E. Mura, A. Nicklass, P. Palmieri, R. Pitzer, G. Rauhut, M. Schütz, U. Schumann, H. Stoll, A. J. Stone, R. Tarroni, T. Thorsteinsson, and H.-J. Werner.

(31) (a) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B. G.; Chen, W.; Wong, M. W.; Andres, J. L.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98*; Revision A.7, Gaussian, Inc., Pittsburgh, PA, 1998. (b) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648–5652. (c) Perdew, J. P.; Wang, Y. *Phys. Rev. B* **1992**, *45*, 13 244–13 249.

parameters was used to optimize the isolated base pairs and to compute the empirical potential interaction energies.

In this study, the following base pairs were considered. 13 out of 29 possible combinations containing the standard bases A, C, G, and either U or T. The abbreviations are taken from preceding studies.^{1,4} The U...U and G...G base pairs were included in order, among other reasons, to have more data to compare with thiobase pairs. All G...A neutral mismatches were included, as they play multiple roles in DNA and RNA.^{5,33,34} Further, G...A mismatches are especially challenging for computations as all of them are substantially internally nonplanar.⁵ G...U wobble base pair is the third most recurrent base pair in RNA molecules.

Further, we considered Watson–Crick inosine...cytosine, 2-aminoadenine...thymine and 8-oxoguanine...cytosine base pairs (I...C, 2-aminoA...T, 8-oxoG...C). The I...C base pair is the key isosteric mimic of A...U base pair (or A...T when I is paired with 5-methylcytosine). This base pair has two H-bonds, whereas its overall electrostatic interaction is similar to the G...C WC base pair.⁴ The 2-aminoA...T base pair is the isosteric mimic of G...5-methylC base pair, it has three H-bonds but due to a low polarity of the monomers its strength is only marginally larger compared with the A...T base pair. The 8-oxoG...C base pair is one of the strongest base pairs and its stability exceeds the G...C WC base pair. I...C and 2-aminoA...T base pairs are the leading modified base pairs in experimental and computational studies aimed for example to reveal the importance of the exocyclic base groups and electrostatic stacking interactions on binding and other key properties of DNA duplexes.³⁵ The 8-oxoG...C WC and 8-oxoG...G base pairs are relevant to the radiation damage.³⁶

The Calcutta U...U base pair is an intrinsically weak base pair occurring in RNA and was considered due to its rather unique C–H...O H-bond.³⁷

The number of base pairs containing thiobases was also included in this study. Thiobases are often present or incorporated in RNA and DNA³⁸ and the description of the sulfur atom represents an important

test for the quality of the QM techniques.^{9a} Thus, the thiobase pairs are of interest from the basic chemistry point of view and we considered number of them.

The nonpolar adenine...difluorotoluene (A...F) base pair is an isosteric mimic of A...T base pair with essentially no H-bonding.^{9b,c, 39}

Finally, C...CH⁺ triple bonded base pair was included as an example of protonated base pairs. Protonated base pairs are stabilized by very strong molecular ion – molecular dipole contributions. The C...CH⁺ base pair is the single canonical base pair in i-DNA tetraplex motif formed for example by telomeric DNA sequences as the counterpart of guanine quadruplexes.⁴⁰

The technique described was first used for DNA base pairs in our previous paper⁴¹ where we determined accurate stabilization energies of selected H-bonded and stacked structures of adenine...thymine and guanine...cytosine base pairs.

Results

Molecular Structures. The base pairs were optimized using the RI-MP2 procedure with cc-pVTZ or TZVPP basis sets of atomic orbitals. For practical reasons these two basis sets provide identical results and thus we do not specify in the text which basis set was used for a given base pair.

Many nucleic acid base pairs are intrinsically nonplanar.^{4,5} The actual geometry of a given base pair is a balance of planar contributions (primary H-bonds) and out-of-plane contributions: intrinsic pyramidalization of the amino groups of bases, electrostatic (secondary) interactions^{16a} and other contributions such as the steric effects.^{4,5} The electronic structure of the amino groups is modulated by the molecular interactions and in-plane H-bonds tend to stabilize the sp² arrangement of the amino group.^{4,5}

The base pairs are divided into three groups for the purpose of this paper: Entirely planar base pairs (majority of the base pairs, for example, the A...T WC base pair), weakly nonplanar base pairs (e.g., G...U wobble where the rings are coplanar and only the unpaired guanine amino group is pyramidal) and substantially nonplanar base pairs (it concerns mainly all the G...A base pairs). The molecular interactions in the C_s and C₁ symmetry structures can be considered as close to identical for weakly nonplanar base pairs while some noticeable differences may exist for substantially nonplanar base pairs. Obviously, there is no clear-cut border between substantially and weakly nonplanar base pairs.

We carried out structure optimizations using C_s symmetry initial structure (as indicated below) for some base pairs that in fact could be weakly nonplanar. The starting geometry can bias the outcome in favor of the C_s symmetry. Nevertheless, in case of weakly nonplanar base pairs the C_s structures and interaction energies are fully representative and do not differ significantly from data based on unconstrained optimizations. The use of C_s structures instead of the nonplanar ones has several advantages. For example, the C_s symmetry simplifies the CCSD(T) calculations and is more suitable as the reference structure for a practical comparison with other methods such as the force fields,

- (32) Case, D. A.; Pearlman, D. A.; Caldwell, J. W.; Cheatham, T. E.; Ross, W. S.; Simmerling, C. L.; Darden, T. A.; Merz, K.; Stanton, R.; Cheng, A. L.; Vincent, J. J.; Crowley, M.; Ferguson, D. M.; Radmer, R. J.; Siebel, G. L.; Singh, U. C.; Weiner, P. K.; Kollman, P. A. AMBER 6; University of California, San Francisco, CA, 1997.
- (33) (a) Chou, S. H.; Zhu, L. M.; Reid, B. R. *J. Mol. Biol.* **1997**, *267*, 1055–1067. (b) Allawi, H. T.; SantaLucia, J. *Biochemistry* **1998**, *37*, 2170–2179. (c) SantaLucia, J.; Turner, D. H. *Biochemistry* **1993**, *32*, 12 612–12 623.
- (34) Leontis, N. B.; Westhof, E. *Q. Rev. Biophys.* **1998**, *31*, 399–455.
- (35) (a) Lankauš, F.; Cheatham, T. E. III; Špačková, N.; Hobza, P.; Langowski, J.; Šponer, J. *Biophys. J.* **2002**, *82*, 2592–2609. (b) Mauro, S. A.; Pawlowski, D.; Koudelka, G. B. *J. Biol. Chem.* **2003**, *278*, 12 955–12 960. (c) Bailly, C.; Waring, M. J. *Nucl. Acids Res.* **1998**, *26*, 4309–4314. (d) Mollegaard, N. E.; Bailly, C.; Waring, M. J.; Nielsen, P. E. *Nucl. Acids Res.* **1997**, *25*, 3497–3502. (e) Crothers, D. M.; Haran, T. E.; Nadeau, J. G. *J. Biol. Chem.* **1990**, *265*, 7093–7096. (f) Diekmann, S.; Mazzarelli, J. M.; McLaughlin, L. W.; Vonkitzing, E.; Travers, A. A. *J. Mol. Biol.* **1992**, *225*, 729–738. (g) Sherer, E. C.; Harris, S. A.; Soliva, R.; Orozco, H.; Laughton, C. A. *J. Am. Chem. Soc.* **1999**, *121*, 5981–5991.
- (36) (a) Thiviyathan, V.; Somasunderam, A.; Hazra, T. K.; Mitra, S.; Gorenstein, D. G. *J. Mol. Biol.* **2003**, *325*, 433–442. (b) Cho, B. P.; Evans, F. E. *Nucleic Acids Res.* **1991**, *19*, 1041–1047. (c) Kuchino, Y.; Mori, F.; Kasai, H.; Inoue, H.; Iwai, S.; Miura, K.; Ohtuska, E.; Nishimura, S. *Nature* **1987**, *327*, 77–79. (d) McAuleyhecht, K. E.; Leonard, G. A.; Gibson, N. J.; Thomson, J. B.; Watson, W. P.; Hunter, W. N.; Brown, T. *Biochemistry* **1994**, *33*, 10 266–10 270. (e) Shibusatani, S.; Takeshita, M.; Grollman, A. P. *Nature* **1991**, *349*, 431–434. (f) Ninaber, A.; Goodfellow, J. M. *J. Biomol. Struct. Dyn.* **1998**, *16*, 651–661. (g) Aida, M.; Nishimura, S. *Mutat. Res.* **1987**, *192*, 83–89. (h) Cysewski, P. *J. Mol. Struct.-Theochem.* **1999**, *466*, 59–67.
- (37) (a) Wahl, M. C.; Rao, S. T.; Sundaralingam, M. *Nat. Struct. Biol.* **1996**, *3*, 24–31. (b) Hobza, P.; Šponer, J.; Cubero, E.; Orozco, M.; Luque, F. J. *J. Phys. Chem. A* **2000**, *104*, 6286–6292.
- (38) (a) Štefl, R.; Špačková, N.; Berger, I.; Koča, J.; Šponer, J. *Biophys. J.* **2001**, *80*, 455–468. (b) Somerville, L.; Krynetski, E. Y.; Krynetskaia, N. F.; Beger, R. D.; Zhang, W. X.; Marhefka, C. A.; Evans, W. E.; Kriwacki, R. W. *J. Biol. Chem.* **2003**, *278*, 1005–1011. (c) Marathias, V. M.; Sawicki, M. J.; Bolton, P. H. *Nucleic Acids Res.* **1999**, *27*, 2860–2867. (d) Alhambra, C.; Luque, F. J.; Esterlich, J.; Orozco, M. *J. Org. Chem.* **1995**, *60*, 969–976. (e) Dubler, E. *Met. Ions Biol. Syst.* **1996**, *32*, 301–338. (f) Rao, T. S.; Durland, R. H.; Seth, D. M.; Myrick, M. A.; Bodepudi, V.; Revankar, G. R. *Biochemistry* **1995**, *34*, 765–772. (g) Nair, T. M.; Myszkka, D. G.; Davis, D. R. *Nucl. Acids Res.* **2000**, *28*, 1935–1940.

- (39) Moran, S.; Ren, R. X. F.; Rumney, S.; Kool, E. T. *J. Am. Chem. Soc.* **1997**, *119*, 2056–2057. (b) Ryjacek, F.; Kratochvil, M.; Hobza, P. *Chem. Phys. Lett.* **1999**, *313*, 393–398.
- (40) (a) Gehring, K.; Leroy, J. L.; Gueron, M. *Nature* **1993**, *363*, 561–565. (b) Berger, I.; Kang, C.; Fredian, A.; Ratliff, R.; Moyzis, R.; Rich, A. *Nat. Struct. Biol.* **1995**, *2*, 416–429. (c) Špačková, N.; Berger, I.; Egl, M.; Šponer, J. *J. Am. Chem. Soc.* **1998**, *120*, 6147–6151. (d) Malliavin, T. E.; Gau, J.; Snoussi, K.; Leroy, J. L. *Biophys. J.* **2003**, *84*, 3838–3847.
- (41) Jurečka, P.; Hobza, P. *J. Am. Chem. Soc.* **2003**, *125*, 15 608–15 613.

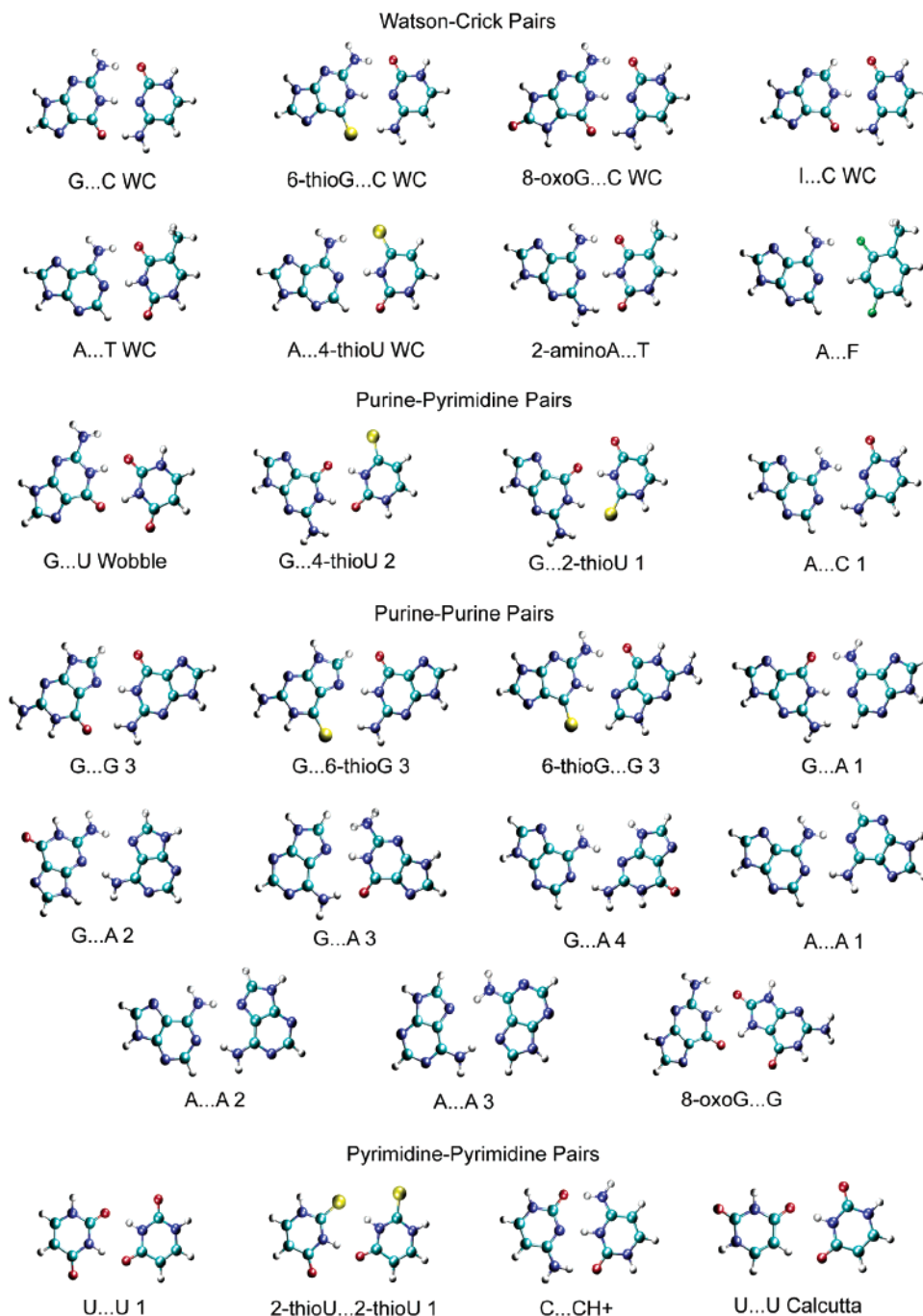


Figure 1. Nucleic acid base pairs structures.

as these methods often for example do not include the amino group nonplanarity. We tried to locate a nonplanar minimum for some of these base pairs but in most cases the structure returned to the planar arrangement. Thus, although the level of calculations does not permit to carry out harmonic vibrational analysis and verify the nature of the optimized structures we assume that all base pairs are represented by relevant structures and no substantial nonplanarity is missed.

For substantially nonplanar base pairs, we report either their genuine nonplanar minima or both C_1 and C_s structures. The intrinsic nonplanarity of G...A 1 base pair is clearly visible in high-resolution DNA and RNA structures and is important to rationalize the conservation patterns of this base pair in ribosomes.^{5b}

Table 1 summarizes the reference primary H-bonding distances for all base pairs. Since all optimized geometries are published as the Supporting Information we do not provide any additional structural data. The RIMP2/cc-pVTZ level shortens the donor–acceptor distances between the heteroatoms by ca. 0.15 Å compared to the HF/6-31G** method (not shown). On the other side, the Becke3LYP optimization provides almost identical donor–acceptor distances as the RI-MP2/cc-pVTZ method with a typical elongation around 0.03 Å. The data thus confirms that geometries of base pairs are much less sensitive to the level of calculations compared with the base pairing energies.

Interaction Energies. The main energy data are included in Table 2. The second, third and fourth column of Table 2 show

Table 1. H-Bond Lengths of the Optimized Base Pairs Obtained at the RI-MP2/TZVPP or RI-MP2/cc-pVTZ Level (all geometries are listed in the Supporting Information. cf. also Figure 1)

base pair ^a	H-bond length	H-bond length	H-bond length
guanine...cytosine WC ^b	N2(G)–O2(C) 2.89	N1(G)–N3(C) 2.90	O6(G)–N4(C) 2.75
6-thioguanine...cytosine WC ^c	N2(G)–O2(C) 2.74	N1(G)–N3(C) 3.11	S6(G)–N4(C) 3.23
8-oxoguanine...cytosine WC ^d	N2(oxoG)–O2(C) 2.89	N1(oxoG)–N3(C) 2.89	O6(oxoG)–N4(C) 2.76
inosine...cytosine WC ^d	N1(I)–N3(C) 2.82	O6(I)–N4(C) 2.82	
adenine...thymine WC ^d	N1(A)–N3(T) 2.83	N6(A)–O4(T) 2.86	
adenine...4-thiouracil WC ^d	N1(A)–N3(thioU) 3.04	N6(A)–S4(thioU) 3.60	
2-aminoadenine...thymine WC ^b	N2(2aminoA)–O2(T) 2.91	N1(2aminoA)–N3(T) 2.91	N6(2aminoA)–O4(T) 2.89
adenine...difluorotoluene WC ^d	N1(A)–C3(DFT) 3.35	N6(A)–F4(DFT) 3.11	
guanine...uracil wobble ^e	O6(G)–N3(U) 2.79	N1(G)–O2(U) 2.77	
guanine...4-thiouracil wobble ^e	O6(G)–N3(thioU) 2.82	N1(G)–O2(thioU) 2.81	
guanine...2-thiouracil wobble ^f	O6(G)–N3(thioU) 2.77	N1(G)–S2(thioU) 3.28	
adenine...cytosine 1 ^d	N1(A)–N4(C) 2.93	N6(A)–N3(C) 2.91	
guanine...guanine 3 ^e	N1(G)–N7(G) 2.87	N2(G)–O6(G) 3.15	
guanine...6-thioguanine 3 ^e	N1(G)–N7(thioG) 2.89	N2(G)–S6(thioG) 3.55	
6-thioguanine...guanine 3 ^e	N1(G)–N7(thioG) 2.92	N2(thioG)–O6(G) 2.95	
guanine...adenine 1 ^{f,g}	O6(G)–N6(A) 2.83	N1(G)–N1(A) 2.86	
guanine...adenine 2 ^{f,h}	N3(G)–N6(A) 3.02	N2(G)–N7(A) 2.99	
guanine...adenine 3 ^f	O6(G)–N6(A) 2.86	N1(G)–N7(A) 2.85	
guanine...adenine 4 ^f	N2(G)–N1(A) 2.95	N3(G)–N6(A) 2.96	
adenine...adenine 1 ^d	N6(A)–N1(A) 2.95	N1(A)–N6(A) 2.95	
adenine...adenine 2 ^d	N6(A)–N1(A) 2.95	N7(A)–N6(A) 2.96	
adenine...adenine 3 ^b	N6(A)–N7(A) 2.96	N7(A)–N6(A) 2.96	
8-oxoguanine...guanine ^b	O8(oxoG)–N1(G) 2.74	N7(oxoG)–O6(G) 2.74	
uracil...uracil 1 ^d	N3(U)–O4(U) 2.83	O2(U)–N3(U) 2.83	
2-thiouracil...2-thiouracil 1 ^d	N3(thioU)–O4(thioU) 2.80	S2(thioU)–N3(thioU) 3.32	
H3-cytosine ⁺ ...cytosine ^d	N4(CH)–O2(C) 2.66	N3(CH)–N3(C) 2.78	O2(CH)–N4(C) 2.90
uracil...uracil "calcutta" ^d	O4(U)–N3(U) 2.87	C5(U)–O4(U) 3.21	

^a Nonplanar structures used where available. The numbers 1–4 labeling A...C, A...A, G...A, and G...G base pairs are consistent with those in ref 4. For other possible nomenclatures see for example <http://www.imb-jena.de/RNA.html>. ^b Weakly nonplanar base pair. ^c C_s structure, but weak nonplanarity is not ruled out. ^d Assumed to be planar, in many cases second optimization attempted to locate the nonplanar structure resulted into planar structure. ^e Only one unpaired amino group is nonplanar, C_s data is given for all G...G 3 type structures as the amino group is away the other guanine. ^f Substantially nonplanar base pair. ^g Commonly labeled as cis-Watson–Crick/Watson–Crick or anti-anti G...A base pair. ^h Commonly labeled as sheared G...A base pair.

the RI-MP2/aug-cc-pVDZ (aDZ), RI-MP2/aug-cc-pVTZ (aTZ) and aDZ→aTZ interaction energies extrapolated according to Helgaker. All numbers include the monomer deformation energies. The next two columns of Table 2 give the RI-MP2/aug-cc-pVQZ (aQZ) and aTZ→aQZ Helgaker's extrapolated data for seven base pairs. The values in parentheses in the aTZ→aQZ column are obtained with Truhlar's aDZ→aTZ (aDZ→aTZ^T) extrapolation for all base pairs. These data reveal that actually the aDZ→aTZ Helgaker's extrapolation stops short of reaching the infinite basis set value and rather predicts the aQZ values. This minor underestimation of binding (0.0–0.7 kcal/mol, less than 2.5%) has a negligible effect on the relative base pair energies.

The Truhlar's aDZ→aTZ (aDZ→aTZ^T) extrapolation is aimed to directly predict the infinite basis set values. It provides numbers that are very close to the Helgaker's extrapolation. The range of differences of relative base pairing energies between the aTZ→aQZ and aDZ→aTZ^T methods is 0.6 kcal/mol, similar to the relative energy difference range of two Helgaker's extrapolations. Note that there are no coefficients available to extend the Truhlar's extrapolation for the aTZ→aQZ case. The extrapolation scheme of Truhlar leads to more negative interaction energies by 0.1 to 1.1 kcal/mol compared to the aDZ→aTZ extrapolations of Helgaker.

Due to the lack of an appropriate auxiliary RI-MP2 basis set we could not perform the aQZ calculations for the thio base pairs.

The MP2/aug-cc-pV5Z (a5Z) H-bond strength values are already very close to the basis set limit,^{17,41,42} and thus the aTZ→aQZ extrapolation aimed to predict the a5Z values can

be considered as being close to converged. The aDZ→aTZ^T extrapolations by Truhlar are also close to the reference values, with difference compared to the Helgaker's aTZ→aQZ extrapolation in the range of –0.3 to +0.3 kcal/mol. This difference range might increase modestly when all base pairs are included.

The seventh column of Table 2 shows the difference between the CCSD(T) and MP2 values, i.e., the higher-order electron correlation correction $\Delta\text{CCSD(T)}$. This correction is in the range of 0.0 to –0.6 kcal/mol. Since we have the $\Delta\text{CCSD(T)}$ data for only a rather small subset of base pairs the actual range of the $\Delta\text{CCSD(T)}$ corrections could be somewhat wider. Nevertheless, this correction is small, mostly negative (increasing stability) and its magnitude tends to increase with the stability of the base pair. Thus the $\Delta\text{CCSD(T)}$ correction has only a marginal effect on the relative base pair stability. The correction has opposite sign compared to base stacking calculations where this term is always decreasing the stabilization energy and is much larger in absolute values.^{22,41} Relatively small values of the $\Delta\text{CCSD(T)}$ term are in full agreement with data on small H-bonded complexes^{20,43} as well as on Watson–Crick and Hoogsteen structures of adenine...thymine and guanine...cytosine base pair and their methylated analogues reported recently by us.⁴¹

The differences between the aDZ→aTZ (Helgaker) values and the aTZ→aQZ data corrected for CCSD(T) are in the range of –0.3 to –1.3 kcal/mol while their absolute values typically increase with the base pair strength. Thus, the aDZ→aTZ extrapolation is highly reliable low-cost estimate (when using

(42) Hobza, P.; Zahradník, R. *Chem. Rev.* **1988**, *88*, 871–897.

(43) Tsuzuki, S.; Uchimaru, T.; Matsumura, K.; Mikami, M.; Tanabe, K. *J. Chem. Phys.* **1999**, *110*, 11 906–11 910.

Table 2. Interaction Energies of Base Pairs (kcal/mol)

structure ^a	ΔE_{MP2}^b					$\Delta \text{CCSD(T)}^c$	$\text{aDZ} \rightarrow \text{aTZ}^a$		
	aDZ	aTZ	aDZ \rightarrow aTZ	aOZ	aTZ \rightarrow aOZ (aDZ \rightarrow aTZ ^T)		ΔE^{SCF}	ΔE^{cor}	E^{def}
Watson–Crick Pairs									
G...C WC	-25.6	-27.0	-27.5	-27.7	-28.2(-27.9)	-0.6	-20.0	-7.4	3.6
6-thioG...C WC pl	-23.6	-25.0	-25.5		(-25.9)		-18.3	-7.0	4.0
8-oxoG...C WC pl	-26.3	-27.7	-28.2	-28.4	-29.0(-29.3)	-0.4	-21.3	-6.9	3.9
L...C WC pl	-20.6	-21.6	-22.0	-22.1	-22.5(-22.7)	-0.2	-15.4	-6.6	2.2
A...T WC	-13.8	-14.7	-15.0	-15.1	-15.4(-15.3)	0.0	-7.0	-8.0	1.5
A...4-thioU WC	-12.8	-13.1	-13.2		(-13.4)		-6.9	-6.3	1.0
2-aminoA...T	-16.1	-17.1	-17.6		(-17.9)		-7.7	-9.9	1.9
2-aminoA...T pl	-15.8	-16.9	-17.3		(-17.6)		-8.8	-8.5	2.4
A...F	-4.75	-4.9	-4.9		(-5.1)		-1.0	-4.0	0.3
purine–pyrimidine pairs									
G...U wobble	-14.4	-15.4	-15.8	-15.7	-15.8(-16.0)	-0.3	-9.7	-6.1	3.0
G...4-thioU2	-14.7	-15.6	-15.9		(-16.2)		-9.3	-6.6	1.9
G...2-thioU1	-13.1	-14.2	-14.6		(-14.9)		-7.1	-7.5	2.0
A...C 1 pl	-14.9	-15.6	-15.9		(-16.1)		-8.1	-7.7	1.7
purine–purine pairs									
G...G 3 pl	-17.4	-18.1	-18.4		(-18.5)		-14.1	-4.3	2.9
G...-6-thioG 3 pl	-17.5	-18.6	-19.0		(-19.2)		-14.3	-4.8	2.8
6-thioG...G 3 pl	-18.3	-19.3	-19.6		(-19.9)		-14.5	-5.2	3.1
G...A 1	-16.4	-17.2	-17.5		(-17.8)		-8.2	-9.3	1.9
G...A 1 pl	-14.8	-15.8	-16.1		(-16.3)		-8.7	-7.4	2.8
G...A 2	-9.7	-10.5	-10.9		(-11.3)		-3.1	-7.7	3.5
G...A 2 pl	-9.5	-10.2	-10.5		(-10.6)		-4.4	-6.1	2.3
G...A 3	-15.8	-16.5	-16.8		(-17.0)		-7.6	-9.1	2.0
G...A 4	-11.0	-11.8	-12.1		(-12.2)		-4.8	-7.3	1.4
A...A 1 pl	-12.2	-12.8	-13.1		(-13.2)		-5.1	-8.0	1.4
A...A 2 pl	-11.5	-12.0	-12.3		(-12.4)		-4.5	-7.8	1.4
A...A 3 pl	-10.2	-10.7	-10.9		(-11.0)		-3.5	-7.4	1.3
8-oxoG...G	-18.3	-19.2	-19.6		(-19.9)		-11.1	-8.5	3.2
pyrimidine–pyrimidine pairs									
U...U 1 pl	-11.4	-12.1	-12.4		(-12.7)	-0.2	-6.8	-5.7	1.1
2-thioU...2-thioU 1 pl	-10.2	-11.2	-11.6		(-12.0)		-4.0	-7.6	1.0
C...CH+ pl	-44.3	-45.7	-46.4	-46.4	-47.0(-47.2)	-0.1	-36.5	-9.9	4.9
U...U Calcutta pl	-8.9	-9.3	-9.5	-9.6	-9.7(9.8)	-0.1	-4.9	-4.5	0.5

^a For abbreviation see Figure 1; pl means that the base pair has been optimized under *C*_s symmetry. Attempts to locate nonplanar geometries for L...C, 8-oxoG...C, G...G 3, A...A 1, A...C 1 A...A 2 and A...T WC resulted in planarization. G...C WC is weakly nonplanar at this level while the planar structure is essentially isoenergetic. All studied pyrimidine...pyrimidine base pairs and some others are assumed to be intrinsically planar. For all base pairs that are significantly nonplanar with provide the unconstrained minimum and in most cases also the planar structure. Cf. also Table 1. ^b aXZ stands for the RI-MP2 interaction energy values with aug-cc-pVXZ (*X* = 2,3,4) basis set with inclusion of extrapolated (*X* = 2,3) deformation energies, aDZ \rightarrow aTZ and aTZ \rightarrow aOZ are the respective CBS extrapolations by Helgaker while aDZ \rightarrow aTZ^T stands for extrapolation by Truhlar. ^c The difference between CCSD(T) and MP2 values with the 6-31G* basis set. ^d ΔE^{SCF} and ΔE^{cor} stand for SCF and correlation parts of the aDZ \rightarrow aTZ interaction energies including the respective deformation terms, E^{def} is the extrapolated monomer deformation energy.

RI-MP2 method instead of MP2) of base pairing energies and the aDZ \rightarrow aTZ^T data appears to be even more accurate.

The next two columns decompose the interaction energy (the aDZ \rightarrow aTZ values) into the SCF and MP2 correlation contributions to the interaction energy. The correlation part of the interaction energy includes mainly two contributions: the dispersion attraction and the intramolecular correlation interaction energy. The second term primarily reflects the reduction of the dipole moments of bases when including the electron correlation and is typically positive, especially for strong base pairs. Thus, the electron correlation interaction energy is not just the pure dispersion attraction as often assumed in the literature. Due to the increase of the size of the basis set the magnitude of the electron correlation term increases compared to the older data.⁴ The other reason for enhancement of the electron correlation term is the use of MP2 gradient optimization. This brings the monomers closer to each other, thus enhancing (in absolute value) both short-range exchange repulsion and the dispersion attraction compared to calculations based on the HF-optimized geometries. It is nevertheless likely that the interbase distance is marginally underestimated with the RI-

MP2/cc-pVTZ optimization since the presently used gradient procedure is not corrected for the basis set superposition error. In summary, for weaker base pairs, the electron correlation stabilization often exceeds the HF stabilization. The A...F nonpolar base pair is dominated by the correlation term. The electron correlation stabilization also increases when the H-bonds directly involve the sulfur atom, as evident for example for the thioU...thioU1 base pair compared with U...U1.

As in all of our older studies we do not list the BSSE-uncorrected interaction energies, as we consider them to be biased. Nevertheless, just to give an example of the BSSE artifact at the RI-MP2/cc-pVTZ level, the pure BSSE correction (see method for explanation) is 3.5 kcal/mol for the thioG...C WC base pair and 2.7 kcal/mol for the G...C WC base pair. The BSSE corrections are smaller for weaker base pairs with reduced interbase overlap and with two H-bonds (not shown). Evidently, uncorrected calculations would exaggerate the base pair strength and bias the relative energies.

The last column of Table 2 presents the deformation energies of monomers. This term is negligible for few weak base pairs but becomes rather significant for strong polar base pairs and

Table 3. Comparison of QM Data (aDZ→aTZ Helgaker Values) with Other Methods (kcal/mol)^a

structure	QM	DFT1	DFT2	MP2/HF	AMBER ^b
G...C WC	-27.5	-25.5 (4.0)/2.0	-27.7/-0.2	-23.4 (2.5)/4.1	-28.0/-0.5
6-thioG...C WC	-25.3	-23.4 (4.4)/1.9	-25.8/-0.5	-22.5 (2.5)/2.8	-25.1/0.2
8-oxoG...C WC	-28.2	-26.8 (3.4)/1.4	-28.7/-0.5	-24.0 (2.3)/4.2	
I...C WC	-22.0	-19.3 (2.5)/2.7	-21.0/1.0	-18.0 (1.4)/4.0	-22.0/0.0
A...T WC	-15.0	-12.3 (1.7)/2.7	-14.5/0.8	-11.8 (0.7)/3.2	-12.8/2.2
A...4-thioU WC	-13.2	-11.0 (1.6)/2.2		-11.2 (0.6)/2.1	
2-aminoA...T WC	-17.6	-14.9 (2.2)/2.7	-17.5/-0.1	-13.7 (1.4)/3.9	-15.8/1.8
A...F WC	-4.9	-3.1 (0.2)/1.8	-4.1/0.8		
G...U wobble	-15.8	-13.4 (2.3)/2.4	-14.8/1.0	-12.7 (1.6)/2.9	-16.0/-0.2
G...4-thioU wobble	-15.9	-12.3 (2.2)/3.6	-13.7/1.8	-12.1 (1.3)/3.8	
G...2-thioU wobble	-14.6	-11.4 (2.1)/3.2	-13.4/1.2	-12.1 (1.2)/2.5	
A...C 1	-15.9	-13.7 (1.8)/2.2	-14.9/1.0	-13.5 (0.9)/2.4	-13.5/2.4
G...G 3	-18.4	-15.8 (2.6)/2.6	-17.4/1.0	-16.3 (1.6)/2.1	-19.4/-1.0
6-thioG...G 3	-19.0	-16.4 (2.3)/2.6	-18.2/0.8		
G...6-thioG 3	-19.6	-17.1 (2.9)/2.5	-19.1/0.5		
G...A 1	-17.5	-14.5 (1.9)/3.0	-16.8/0.7	-14.2 (1.1)/3.3	-14.7/2.8
G...A 2	-10.9	-9.4 (1.7)/1.5	-11.7/-0.8	-9.7 (0.9)/1.2	-11.4/1.2
G...A 3	-16.8	-14.4 (2.0)/2.4	-15.8/1.0	-13.5 (1.2)/3.3	-15.2/1.6
G...A 4	-12.1	10.5 (1.6)/1.6	-12.8/-0.7	-10.3 (0.8)/1.8	-10.7/1.4
A...A 1	-13.1	-10.6 (1.6)/2.5	-12.3/0.8	-11.0 (0.6)/2.1	-10.8/2.3
A...A 2	-12.3	-10.1 (1.6)/2.2	-11.5/0.8	-10.3 (0.6)/2.0	-10.9/1.4
A...A 3	-10.9	-9.1 (1.4)/1.8	-11.4/-0.5	-9.2 (1.0)/1.7	-10.9/0.0
8-oxoG...G	-19.6	-16.2 (3.0)/3.4	18.0/1.6		
U...U 1	-12.4	-10.2 (1.3)/2.2		-10.0 (0.6)/2.4	-12.1/0.3
2-thioU...2-thioU 1	-11.6	-8.1 (1.0)/3.5		-8.8 (0.5)/2.8	
C...CH ⁺	-46.4	-44.3 (5.6)/2.1	-46.6/-0.2	-41.8 (3.1)/4.6	-41.7/-4.7
U...U Calcutta	-9.5	-7.5 (0.7)/2.0	-8.7/0.8	7.4 (0.4)/2.1	

^a DFT1 – Becke3LYP/6-31G** method, DFT2 – PW91/6-31G**//Becke3LYP/6-31G** method, MP2/HF – MP2/6-31G*(0.25)//HF/6-31G** level, AMBER – Cornell et al. force field.^{3b} The first number is the binding energy with inclusion of the monomer deformation and the second number after “/” is the difference with respect to the aDZ→aTZ data. The values in parentheses separately list the deformation energies where relevant. DFT2 has the same deformation energy as DFT1 while the AMBER values are commented on in the text. ^b Majority of the AMBER values are taken from ref 6.

especially for the protonated C...CH⁺ base pair. Large deformation energies indirectly indicate that the base pair is getting substantial stabilization by polarization effects, thus we notice a rather significant difference in the nature of base pairing between weak and strong base pairs. Note that each guanine contributes by 1 kcal/mol to the deformation energy (for C_s structures) due to the guanine amino-group planarization. This energy contribution would be 0.1 kcal/mol for adenine and cytosine, 0.9 kcal/mol for 8-oxoguanine and 6-thioguanine and 0.7 kcal/mol for 2-aminoadenine.

Table 3 compares the reference ab initio data with some other methods. As the reference value we take the aDZ→aTZ Helgaker's extrapolated binding energies including the monomer deformation energies.

The DFT1 values in Table 3 were calculated with the Becke3LYP/6-31G** method (gradient optimizations and energy evaluations are done at the same level) corrected for BSSE (interaction energies) and deformation energies with respect to the nonplanar monomers. The DFT1 method underestimates the reference aDZ→aTZ data systematically by 1.4–3.6 kcal/mol. The largest inaccuracy regarding the relative order of stability (difference of interaction energy between two base pairs) is 2.2 kcal/mol with respect to both aDZ→aTZ and aDZ→aTZ^T reference sets. Not surprisingly, the largest discrepancy is seen for thio base pairs and this reflects the well-established inaccuracy in the DFT description of intermolecular correlation (dispersion) effects. Note that the method fails completely for base stacking calculations.^{4b}

The PW91 DFT functional was suggested recently to perform for H-bonding better than the Becke3LYP method.^{14c,44} Thus the next column (DFT2) provides the PW91/6-31G** interaction

energies obtained at the Becke3LYP/6-31G** optimized geometries, again with inclusion of the monomer deformation energies at the Becke3LYP/6-31G** level. The PW91 functional indeed improves the agreement with the reference aDZ→aTZ values by lowering the interaction energies compared to the DFT1 numbers. The difference with respect to the aDZ→aTZ data is in the range of -0.8 to +1.8 kcal/mol. The largest error in relative base pair stability is 2.6 kcal/mol. Thus, the relative stability is not improved compared to the Becke3LYP method. As usual with DFT the method has a poorer performance for weaker and thio base-containing base pairs. We have also re-evaluated more than a dozen base pairs using the PW91/6-31G** geometries (not shown). The PW91/6-31G**/PW91/6-31G** method further slightly increases stability of the base pairs, however, in the expense of increased relative error (above 3.0 kcal/mol). In fact, it appears that the PW91/6-31G** optimization slightly exaggerates the monomer deformation energies (5.4 kcal/mol for the G...C WC base pair, 7.3 kcal/mol for the C...CH⁺ base pair, etc., cf. with Tables 2 and 3). Note nevertheless that despite the improved description of the base pairing the PW91 DFT method remains to fail for base stacking. Let us give just a single example, the PW91/aug-cc-pVDZ stacking energy of antiparallel undisplaced cytosine dimer is -1.3 kcal/mol, completely away from the MP2/aug-cc-pVDZ value of -10.2 kcal/mol.

We did not test the DFT methods with extended basis sets of atomic orbitals since the DFT methods do not show a systematic improvement of the base pairing energies with the size of the basis set.¹⁷ Further, considering the efficiency of the RI-MP2 method compared to MP2, we see no need to perform the DFT calculations with extended basis sets where the accuracy of the RI-MP2 method is superior. We reiterate

(44) Tsuzuki, S.; Luthi, H. P. *J. Chem. Phys.* **2001**, *114*, 3949–3957.

that while DFT is a very useful method to obtain fast estimates of energies for H-bonded systems it is not a tool for reference calculations and can fail significantly.

The MP2//HF abbreviation stands for the MP2/6-31G*(0.25)//HF/6-31G** method used almost a decade ago for an extended set of reference calculations.⁴ In contrast to the preceding study the data given here in Table 3 include deformation energies of monomers evaluated with respect to the nonplanar isolated bases (see method). The MP2//HF numbers are 1.2–4.6 kcal/mol below (in absolute values) the aDZ→aTZ interaction energies. Taking into consideration that the aDZ→aTZ data still do not match the MP2 basis set limit we conclude that the absolute error of the MP2/6-31G*(0.25)//HF/6-31G** base pair calculations is somewhat larger than we suggested before.⁴⁵ The largest error in the relative stability of different base pairs is 3.2 kcal/mol (3.8 kcal/mol with respect to the aDZ→aTZ^T data), thus the MP2//HF method is less accurate than the DFT methods. The MP2//HF calculations show a different trend compared to the DFT1 and DFT2 numbers as the poorest performance of the MP2//HF level occurs for the strong base pairs. This is a consequence of insufficient flexibility of the monomers at the HF level of theory used for optimization. It penalizes strong base pairs due to their larger monomer geometry adjustments upon base pairing. Note that in contrast to the DFT methods the MP2/6-31G*(0.25) method performs very well for base stacking and thus allows a balanced description of the whole range of base – base interactions.⁴

The last column of Table 2 shows the Cornell et al. (AMBER) force field values.^{3b} These numbers also include monomer deformation. In absolute values, the force field data are amazingly close to the reference QM aDZ→aTZ values, in the range of –1.0 to + 2.8 kcal/mol (except of the C...CH⁺ base pair). The largest error in relative base pairing stability is 3.8 kcal/mol, i.e., larger than for the medium-quality QM data. Nevertheless, taking into account the simplicity of the force field it is an excellent agreement. Further, the force field performs equally well for base stacking and thus allows a balanced description of the base – base interactions. The pattern of difference between the QM and AMBER data is very consistent. The force field performs better for stronger base pairs with considerable electrostatic interaction where sometimes its stabilization energy even exceeds the reference QM data (e.g., the G...G base pair). It is because the AMBER charges are derived at the HF level, and thus overestimate the dipole moments (polarity) of bases. This increases stability of the strong base pairs. On the other side the force field, monomers are quite rigid and do not allow enough monomer deformation which compensates for the overestimated electrostatic contribution. For weak base pairs the AMBER force field usually underestimates the reference QM data.

The only exception is the C...CH⁺ base pair as its stability is underestimated by AMBER by almost 5 kcal/mol. This is because H-bonding of protonated base pairs is influenced by polarization and charge-transfer effects more than in the case of neutral base dimers. The difference indicates that the magnitude of the "ionic" nonadditive effects in the C...CH⁺ pair (relative to neutral base pairs) is of the order of 5 kcal/mol. This is consistent with estimated 2–4 kcal/mol of "ionic"

polarization effects in stacked protonated base dimers¹⁰ and complexes of nucleobases with charged intercalators.⁴⁶

We wish to point out that the AMBER data are included only for a subset of base pairs and the force field parameters are standardized only for standard bases.^{3b} Note that AMBER parameters for modified bases were derived in the past and utilized in simulations^{10,30,39,40,47} but there were numerous minor differences how different groups derived their modified parameters (bonded terms, van der Waals parameters, and the charges). The results of a force field calculation on isolated base pair is also slightly affected by adjustments of the charge set after the backbone unit is deleted. Thus, the present results should not be considered as any kind of a reference AMBER force field values, the purpose of the calculations was merely to show that the overall performance of the force field is very good. That is why we also did not test other force fields used for molecular modeling of nucleic acids and more force field data can be found elsewhere.⁶

On the basis of the referee suggestion, we have also in detail compared the ability of AMBER force field to reproduce the reference RI-MP2 H-bonding distances. The calculations were carried out using two sets of RESP charges, HF/6-31G* and Becke3LYP/6-31G* ones, evaluated in the corresponding QM optimized geometries. The results are summarized in Table S1 in the Supporting Information. In summary, the AMBER performance remains satisfactory and with the HF charges (corresponding to those used in AMBER simulations) majority of H-bonded heteroatom distances are reproduced with accuracy 0.1 Å or better. This is a better agreement with the reference QM data than achieved with the HF/6-31G*-optimized geometries (cf. with Table 2 in ref 4a). Large differences around 0.3 Å are observed for some (but not all) XH...S hydrogen bonds. This may be due to insufficient refinement of the sulfur parameters though it may also reflect the anisotropic nature of the large sulfur atom or some other missing term. More importantly, the C5(U)–O4(U) distance in the U...U Calcutta base pair is overestimated by 0.3 Å which possibly may affect the accuracy of description of the C–H...O H-bonds in simulations.

The AMBER force field by definition neglects all polarization and charge transfer effects. To our opinion, the excellent correlation between the AMBER and QM values confirms that the main stabilizing contribution of H-bonded base pairs originates in the electrostatic interactions well reproduced by the atom-centered electrostatic-potential fitted charges. Although we do not show any direct proof of that statement (such as energy decomposition), we suggest that no other interpretation of the data is viable. On the other hand, it does not rule out some still significant contribution from polarization and charge transfer effects. It is especially apparent for the stability of the C...CH⁺ base pair which, evidently due to a neglect of the additional polarization related to the charge of +1, is underestimated by AMBER by almost 5 kcal/mol. Intuitively, one could assume that the polarization/charge transfer in neutral base pairs is smaller than or comparable to the "ionic" effect seen in the C...CH⁺. This indirectly suggests that the polarization/charge transfer effects do not dominate the base pairing. It is to be noted that the empirical potential base pairing energies are almost exclusively consisting of the electrostatic term while the

(45) Sponer, J.; Leszczynski, J.; Hobza, P. *Biopolymers* **2001**, *61*, 3–36.

(46) Reha, D.; Kabelac, M. Ryjacek, F.; Sponer, J.; Sponer, J. E.; Elstner, M.; Suhai, S.; Hobza, P. *J. Am. Chem. Soc.* **2002**, *124*, 3366–3376.

van der Waals (Lennard–Jones) term is rather small at the force field minima (not shown). The van der Waals term in the potential primarily aims to mimic the sum of the dispersion attraction and the short-range repulsion.

Polarization and charge transfer effects are indirectly correlated with the monomer deformations (such as stretching of N–H bonds) and are of key importance to explain spectroscopic properties of base pairs and many other properties of the base pairs. The polarization and charge transfer effects are partly hidden when considering the total interaction energies as their attractive contribution to stabilization is partly canceled by the repulsive monomer deformation energies. In relation to this, we wish to note that the monomer deformation energies are rather negligible in case of the force field calculations, less than 1.0 kcal/mol except of the C...CH⁺ base pair where the monomer deformation energy amounts to 1.5 kcal/mol. Thus, the excellent agreement between the pair-additive force field and the reference QM data is partly due to a cancellation of errors: overestimation of the electrostatic term by the force field due to the use of HF-fitted charges, neglect of attractive polarization and charge-transfer effects by the force field and, finally, underestimation of the repulsive monomer deformation energies by the force field.

Concluding Remarks

RI-MP2 gradient optimizations of dozens of nucleic acid base pairs were carried out with extended cc-pVTZ (or TZVPP) basis set of atomic orbitals. The accurate structures were utilized to obtain complete basis set MP2 extrapolated interaction energies with aug-cc-pVDZ and aug-cc-pVTZ basis sets, in some cases extended by aug-cc-pVQZ data further corrected for the higher-order electron correlation terms with the 6-31G* basis set. Such nonempirical calculations are assumed to be very close to the fully converged data.

The present calculations provide a new reference set of data for intrinsic (gas phase) structures and binding energies of H-bonded nucleic acids base pairs, ranging from the weakest nonpolar ones (–5 kcal/mol) up to the strongest protonated ones (–47 kcal/mol). Although we obviously could include only a fraction of known base pairing combinations they were selected in order to consider dimers with very different balance of the individual contributions. This is seen by the range of the deformation energies which spans from 0.3 to 5.0 kcal/mol. In absence of relevant gas-phase interaction energy experiments, this set of data reveals the nature of the interactions and is suitable for parametrization, calibration, and verification of other computational methods.

The calculations show still a nonnegligible increase of predicted base pair stabilities compared with the preceding calculations.⁴ Increased stabilization is attributed mainly to a substantial improvement of the description of intermolecular correlation (dispersion) effects. Nevertheless the calculations confirm that medium-quality ab initio methods provide rather satisfactory estimates of the base pairing energies, sufficiently accurate for most applications, especially regarding the relative

stability of base pairs. The PW91 DFT method with medium quality basis sets appears to be one of the best choices^{14c,44} although one has to keep in mind that the DFT methods (in contrast to the MP2 method) are suitable neither for base stacking calculations nor for true reference calculations on H-bonding (see above). On the other side, to guarantee accuracy of 0.5–1 kcal/mol for H-bonded base pairs the CBS MP2 extrapolation should be based on aug-cc-pVQZ data and the ΔCCSD(T) correction should be included.

The Cornell et al. atom–atom pair-additive force field with atom-centered electrostatic potential fitted charges shows modest differences compared to the reference data, and actually its performance is improved compared to the preceding estimates with less complete QM data.^{4b,6} The agreement between the reference quantum-chemical data on base pairing energies and the pair-additive force field indicates that the dominating contribution to base pairing stability is the electrostatic term well described by atom-centered point charges. Thus, we see no imminent need for more complex description of the electrostatic term in molecular modeling, such as the use of distributed multipoles or additional charges. On the other hand, polarization and charge-transfer effects for strong base pairs are certainly not negligible,^{16b,c} as indirectly seen from the monomer deformation energies upon the base pairing. Thus, inclusion of polarization term⁴⁸ into the force field would be very profitable and would further improve the description of relative stability of strong and weak base pairs. Nevertheless, on the basis of the magnitude of the reevaluated electron correlation contributions to the interaction energies we suggest that the second most important contribution to base pair stabilization is the dispersion term.

Acknowledgment. This study was supported by grant LN00A032 by Ministry of Education of the Czech Republic and Wellcome Trust International Senior Research Fellowship in Biomedical Science in Central Europe GR067507 (JS).

Supporting Information Available: RI-MP2/cc-pVTZ (or RI-MP2/TZVPP) optimized reference geometries for all base pairs, and a table with H-bond lengths at the QM, AMBER-HF, and AMBER-DFT levels. This material is available free of charge via the Internet at <http://pubs.acs.org>.

JA048436S

- (47) (a) Gallego, J.; Golden, E. B.; Stanley, D. E.; Reid, B. R. *J. Mol. Biol.* **1999**, *285*, 1039–1052. (b) Soliva, R.; Laughton, C. A.; Luque, F. J.; Orozco, M. *J. Am. Chem. Soc.* **1998**, *120*, 11 226–11 233. (c) Cubero, E.; Laughton, C. A.; Luque, F. J.; Orozco, M. *J. Am. Chem. Soc.* **2000**, *122*, 6891–6899. (d) Florian, J.; Goodman, M. F.; Warshel, A. *J. Phys. Chem. B* **2002**, *104*, 10 092–10 099. (e) Sherer, E. C.; Harris, S. A.; Soliva, R.; Orozco, M.; Laughton, C. A. *J. Am. Chem. Soc.* **1999**, *121*, 5981–5991. (48) Halgren, T. A.; Damm, W. *Curr. Opin. Struct. Biol.* **2001**, *11*, 236–242. (b) Kaminski, G. A.; Stern, H. A.; Berne, B. J.; Friesner, R. A.; Cao, Y. X. X.; Murphy, R. B.; Zhou, R. H.; Halgren, T. A. *J. Comput. Chem.* **2002**, *23*, 1515–1531. (c) Garmer, D. R.; Gresh, N.; Roques, B. P. *Proteins Struct. Funct. Genet.* **1998**, *31*, 42–60. (d) Energy decomposition schemes and definitions of charge transfer are in general not unambiguous, thus it is not straightforward to separate charge-transfer and polarization from the other contributions and some arbitrary steps always need to be done. The conclusions of the present paper are based mainly on the very good empirical correlation between full ab initio treatment and pair-additive force field relying on the simple Coulombic electrostatic model.