JCTC Journal of Chemical Theory and Computation

# Reference Quantum Chemical Calculations on RNA Base Pairs Directly Involving the 2′-OH Group of Ribose

Jiří Šponer,*[,†] Marie Zgarbová,[‡] Petr Jurečka,*[,†,‡] Kevin E. Riley,[§,||] Judit E. Šponer,[†,⊥] and Pavel Hobza[§]

*Institute of Biophysics, Academy of Sciences of the Czech Republic, Královopolská 135, 612 65 Brno, Czech Republic, Department of Physical Chemistry, Palacky University, tr. Svobody 26, 771 46 Olomouc, Czech Republic, Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic and Center of Biomolecules and Complex Molecular Systems, Flemingovo náměstí 2, 166 10 Prague 6, Czech Republic, Department of Chemistry, P.O. Box 23346, University of Puerto Rico, Rio Piedras, Puerto Rico 00931, and National Center for Biomolecular Research, Faculty of Science, Masaryk University, Brno, Czech Republic*

Received December 10, 2008

**Abstract:** The folded structures of RNA molecules and large ribonucleoprotein particles are stabilized by a wide range of base pairs that actively utilize the 2′-OH groups of ribose for base pairing. Such base pairing does not occur in DNA and is essential for functional RNAs. We report reference quantum chemical calculations of base pairing energies for a representative selection of 25 RNA base pairs utilizing the ribose moiety for base pairing, including structures with amino acceptor interactions. All base pairs are evaluated at the MP2 level with extrapolation to the complete basis set (CBS) of atomic orbitals. CCSD(T) correction terms were obtained for four base pairs. In addition, the base pairing is evaluated using the DFT-SAPT perturbational procedure along with the aug-cc-pVDZ basis set, which allows for the decomposition of the interaction energies into separate, physically meaningful, components. These calculations confirm that, compared to canonical base pairs, many RNA base pairs exhibit a modestly increased role of dispersion attraction compared to canonical base pairs. However, the effect is smaller than one would assume based on assessment of the ratio of HF and correlation components of the interaction energies. Interaction energies are further calculated using the SCS(MI)-MP2 and DFT-D methods. Finally, we estimate the effect of aqueous solvent screening on the base pairing stability using the continuum solvent approach.

## Introduction

The principles of RNA base pairing differ strikingly from those of DNA due to the presence of the 2′-hydroxyl group of ribose. Thus, besides the common canonical base pairs, functional RNAs utilize a very wide range of non-Watson−Crick (non-WC) base pairing patterns[1−21] (Figure 1). The most important non-WC RNA base pairs directly involve the ribose 2′-OH groups as either donors or acceptors of H-bonds. These "sugar-edge" base pairs are absolutely essential for building up complex three-dimensional (3D) RNA architectures. They include the leading RNA tertiary interactions, such as the A-minor [10−12] and packing interactions,[13,14] and many important base pairs in the internal RNA loops.[15−19] Some RNA base pairs lack base-base H-bonds,[8,9] many are intrinsically nonplanar,[8,9,22−25] and the conformational space of RNA base pairs can consist

* Corresponding author e-mail: sponer@ncbr.chemi.muni.cz (J.S.),petr.jurecka@upol.cz (P.J.).

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

‡ Palacky University.

§ Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic and Center of Biomolecules and Complex Molecular Systems.

|| University of Puerto Rico.

⊥ Masaryk University.

10.1021/ct800547k CCC: $40.75 
Published on Web 02/25/2009

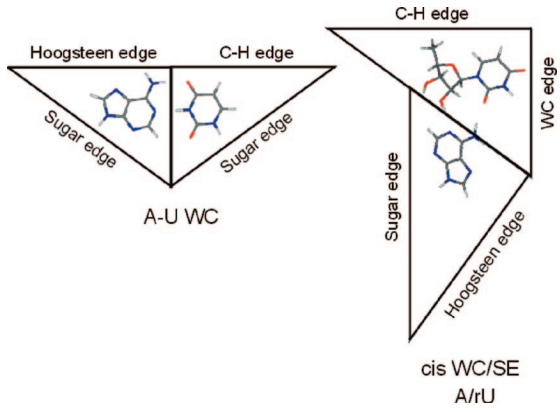


***Figure 1.*** RNA nucleotides create base pairs in a systematic manner using three edges: Watson Crick, Sugar, and Hoogsteen (or C−H for pyrimidines).[9,21]

of several substates including water-inserted structures.[14,19] While many RNA base pairs are strong, with clear minima on the gas phase potential energy surfaces, others are weakly bound and are evidently held together by the overall RNA architecture.[22−26] Even these weak interactions, however, have biochemical relevance.

RNA base pairs have been classified by Leontis and Westhof according to their 3D shapes, considering the span of the base pair and orientation of the attached sugar phosphate backbone as the main structural determinants.[8,9,18] There are rare instances where this classification scheme fails to unambiguously describe an RNA base pair.[19]

The classification describes twelve principal base pair families as well as additional intermediate (bifurcated) families. The families are further divided into isosteric subgroups.[8,9] Each isosteric subgroup can be primarily characterized by the distance between the C1′ atoms and the mutual orientation of the corresponding vectors between C1′ and the glycosidic N (N1 in pyrimidines and N9 in purines). Isosteric base pairs can substitute for each other without affecting the RNA 3D structure. Since the RNA function depends chiefly on the 3D structure, isosteric substitutions are presumed to have a minimal effect on function, i.e., can be considered as functionally neutral substitutions.[8,9,15,16] This is known as the RNA isostericity principle. The isostericity principle is very robust, although it is based exclusively on structures and does not take into account the energetics of base pairing. The RNA 2D structures are considerably less conserved compared with the 3D structures, reflecting the fact that the 3D structure is the primary target of evolutionary constraints.[16] Nevertheless, the basic isostericity principle can be modulated by other factors. In such a case only a subset of possible isosteric substitutions is realized during evolution, or the frequency of occurrence (in the aligned sequences) differs for otherwise isosteric base pairs. A textbook example is the P-interaction tertiary quartet, where a G/U wobble base pair interacts with the C=G Watson Crick base pair while it cannot be substituted with the isosteric A/C base pair. The reason is an unfavorable electrostatic interaction between the A/C and C=G base pairs.[14] (In this paper, the "/" mark indicates non-WC base pairs, while G=C and A-U marks are used for the standard base pairs). The phylogenetic preference to utilize the G=C base pairs as receptors rather than the A-U ones in A-minor tertiary interactions could also be caused by the strength of the interactions, although available biochemical experiments did not indicate any large free energy differences.[20] Similarly, the otherwise isosteric cis-Watson−Crick/Watson−Crick A/G and G/A base pairs do not covary when the amino group of guanine is involved in out-of-plane or tertiary interactions utilizing the unpaired nonplanar guanine amino group.[22]

Despite their biological importance, computational and physical chemistry literature on the RNA base pairing is sparse,[14,19,23−37] contrasting the abundance of studies on simple Watson Crick base pairs. Computations provide a useful complement to experimental techniques and bioinformatics, as they can capture certain physical chemistry features of the RNA base pairing that cannot be directly visualized by experiment. The leading structural methods (X-ray crystallography,[1−7] structural bioinformatics,[8,9,16−18] and cryoelectron microscopy[38,39]) show basically static averaged RNA structures with some resolution limits. They thus provide only limited insights into the dynamics of the molecular interactions. NMR can provide more data on structural dynamics of RNA, but this method has significant resolution limits for RNA.[40−43] These techniques do not reveal the magnitude and nature of the intermolecular forces between the interacting groups and nucleotides. Insights into energetics of molecular interactions in RNA can be obtained by thermodynamics studies, which, however, do not allow an unambiguous analysis of the individual energy contributions.[44,45] Computational methods (molecular dynamics simulations and quantum chemistry) can complement the information available in atomic resolution structures by providing energetics of molecular interactions and structural dynamics.

Recently, we have carried out studies to complement the structural and bioinformatics classification of RNA base pairs using advanced quantum chemical calculations. These studies were specifically aimed at the characterization of RNA base pair families directly utilizing the sugar edges.[23−27] The structures of these complexes were relaxed using the density functional theory method, often with constraints to keep the base pairs in biologically relevant geometries. The interaction energies were derived at the MP2 level with the aug-cc-pVDZ basis set of atomic orbitals, which is sufficient for semiquantitative accuracy.[46] The results were compared with those obtained using the leading molecular modeling force field, AMBER.[47] The force field provides reasonably accurate results for these types of interactions, which is also reflected by its, generally very good, behavior in treating these base pairs in explicit solvent simulations.[19,34−37]

In the present study, we selected a set of 25 RNA base pairs on which to perform reference QM calculations and to obtain better insights into the balance of forces in various types of RNA base pairs. These 25 base pairs were chosen from five key families in which the sugar interaction is important or dominant, to cover as broad a spectrum as possible of RNA interactions. The gas phase interaction energies (see below) of these base pair interactions range from −10 to −31 kcal/mol. Our set also contains alternative substates in which some of the nucleobase amino groups are in amino-acceptor positions with respect to the 2′-OH group

of ribose. Due to the medium to low resolution of RNA X-ray structures, the amino-acceptor interactions would be difficult to prove by experiment, but their presence in RNAs cannot be ruled out.[22]

Most reference data reported in this paper were obtained using the MP2 method with binding energy results extrapolated to the complete basis set limit using the aug-cc-pVDZ and aug-cc-pVTZ basis sets of atomic orbitals (MP2/CBS level, see Methods below). For four base pairs we report the CBS(T) data, i.e., MP2/CBS calculations corrected for the higher electron correlation effects using the CCSD(T) method with a smaller basis set of atomic orbitals (see Methods below). In contrast to our preceding reference study on pure "base to base" nucleic acids base pairs,[48] in this study we were not able to carry out the CBS(T) calculations systematically due to the size of the RNA base pairs with one or two riboses directly involved in the interactions. Nevertheless, the MP2/CBS method is a viable benchmark, as the CCSD(T) corrections for H-bonded base pairs are only modest.

The studied base pairs were further evaluated using the perturbational DFT-SAPT technique,[49−52] which, with large basis sets, should yield results similar to those of the CCSD(T) method. The present SAPT calculations were carried out with the medium aug-cc-pVDZ basis set (see Methods below). The SAPT procedure allows decomposition of the interaction energies into well-defined components, corresponding to physical phenomena such as dispersion and induction. Although energy decompositions should not be overinterpreted, they provide new insights into the complexity of the RNA and DNA[53−59] base pair interactions.

Additional interaction energy computations on these base pairs were carried out with the SCS(MI)-MP2[60] and DFT-D[61,62] methods. The SCS(MI)-MP2 method is a recently proposed semiempirical MP2 technique providing a balanced description of stacked and H-bonded molecular clusters. The DFT-D is a fast method that includes the dispersion forces via a damped empirical correction. Finally, we carried out solvation calculations, which should give us an estimate of the stability of the base pairs studied in solution and provide an alternative energy ordering of the individual structures.

## Methods

**Geometry Optimizations.** Geometry optimizations were carried out at the DFT (Density Functional Theory) level of theory using the Gaussian03 program package. The density functional was built up by Becke's three-parameter exchange and Lee−Yang−Parr's correlation functional (abbreviated as B3LYP). The 6−31G** basis set was used for all geometry optimizations. The B3LYP/6−31G**-optimized structures compare quite well with reference RIMP2/cc-pVTZ data and are entirely sufficient for the subsequent high-quality interaction energy calculations.[48] For many RNA base pairs we had to apply specific constraints to impose the target geometry, as unconstrained optimization would result in a different base pair type or a substantially perturbed geometry. Considering this fact, the flexibility of RNA pairing and the limited accuracy (resolution) of X-ray structures of complex RNAs, such geometry optimization procedure is entirely appropriate for our purpose. Where relevant, the geometries were taken from our preceding studies where more details about the geometries can be obtained.[23−26] All computed structures can be found in the Supporting Information.

**Interaction Energies.** Total interaction energy of a dimer [AB] $\Delta E^{AB}$ is defined as

$$\Delta E^{AB} = E^{AB} - E^{A} - E^{B} \quad (1)$$

where $E^{AB}$ stands for the electronic energy of the whole system in the optimized geometry, and $E^{A}$ and $E^{B}$ are the electronic energies of the isolated subsystems A and B in the dimer geometry. In a few cases (base pairs with water insertion) we evaluated trimers; the extension of eq 1 for trimers is straightforward.[63]

The interaction energy ($\Delta E$) of the standard electron correlation calculations (MP2 and CCSD(T)) has two components: the Hartree−Fock (HF) term ($\Delta E^{HF}$) and the electron correlation term ($\Delta E^{cor}$).

$$\Delta E = \Delta E^{HF} + \Delta E^{corr} \quad (2)$$

The main quantities described by the $\Delta E^{HF}$ term are the electrostatic interaction energy, short-range exchange repulsion, and polarization/charge transfer contributions to the total interaction energy. The $\Delta E^{corr}$ term is dominated by the dispersion attraction and also includes the electron correlation corrections to the other contributions. The correlation correction to electrostatics is usually repulsive since the electron correlation reduces the dipole moments of the monomers and thus also the electrostatic attraction.

All variational interaction energies are corrected for the basis set superposition error (BSSE) using the standard counterpoise procedure but do not include the deformation energies. The deformation energies are disregarded due to substantial structural alterations of the sugar-base segment upon base pairing in many RNA base pairs. Such rearrangements do not reflect the direct (electronic) forces between the interacting monomers. Thus a direct inclusion of monomer deformations into the interaction energies would bias the results. For a detailed discussion regarding the role of the deformation energies in base pairing calculations see refs 23−25 and 48 where we explain why a formal inclusion of deformation energies into the BSSE correction is inappropriate except when dealing with the simplest H-bonded systems.

The interaction energies were evaluated by the following methods:

**Reference Interaction Energies.** The MP2 calculations were performed with aug-cc-pVDZ (aDZ) and aug-cc-pVTZ (aTZ) basis sets and extrapolated to the complete basis set (MP2/CBS) using the technique of Helgaker and co-workers (eq 3).[64,65]

$$E_X^{HF} = E^{HF}_{CBS} + Ae^{-\alpha X} \text{ and } E_X^{corr} = E^{corr}_{CBS} + BX^{-3} \quad (3)$$

$E_X$ and $E_{CBS}$ are energies for the basis set with the largest angular momentum X and for the complete basis set, respectively, and $\alpha$ is a parameter fitted by the authors.[64,65]

In four cases (cWS C/rG, cWS C/rC, tWS A/rA, and tWS A/rG; see below for the abbreviation) the MP2/CBS energies were further corrected for the higher order correlation effects by adding a ΔCCSD(T) correction to the MP2/CBS energy. The ΔCCSD(T) correction is the difference between the CCSD(T) and MP2 interaction energies ($\Delta E^{CCSD(T)} - \Delta E^{MP2}$) calculated in a small basis set (eq 4).

$$\Delta E_{CBS(T)} = \Delta E_{CBS}{}^{MP2} + (\Delta E^{CCSD(T)} - \Delta E^{MP2})_{6\text{-}31+G^*} \quad (4)$$

For more details see elsewhere.[48,66]

All MP2 calculations were performed in TurboMole 5.10 program[67,68] with RI approximation (RI-MP2[68]) and CCSD(T) calculations in the Molpro package.[69] The frozen-core approximation was used throughout the study.

**Decomposition of Interaction Energies via DFT-SAPT.** The symmetry adapted perturbation theory method[70] allows for decomposition of the total interaction energies into physically meaningful components. Recently a computationally less demanding approach, in which the monomer is described by DFT and intermolecular interactions by SAPT, became available.[49−52,71] Here we chose the DFT-SAPT method implemented in the Molpro program package.[69] The total interaction energy is usually calculated as a sum of the following terms (eq 5):

$$E^{SAPT} = E_{el}{}^{(1)} + E_{exch}{}^{(1)} + E_{ind}{}^{(2)} + E_{exch\text{-}ind}{}^{(2)} + E_{disp}{}^{(2)} + E_{exch\text{-}disp}{}^{(2)} + \delta(HF) \quad (5)$$

The individual terms correspond to the electrostatic ($E_{el}{}^{(1)}$), exchange repulsion ($E_{exch}{}^{(1)}$), induction including charge transfer ($E_{ind}{}^{(2)}$) and dispersion ($E_{disp}{}^{(2)}$) contributions and their mixing terms ($E_{exch\text{-}ind}{}^{(2)}$ and $E_{exch\text{-}disp}{}^{(2)}$). $\delta$(HF) is a correction term for the most important higher order contributions (for details see e.g. refs 64−66). In the following we will add the mixing terms $E_{exch\text{-}ind}{}^{(2)}$ and $E_{exch\text{-}disp}{}^{(2)}$ to their corresponding second order components denoting them as $E_{ind}$ ($= E_{ind}{}^{(2)} + E_{exch\text{-}ind}{}^{(2)}$) and $E_{disp}$ ($= E_{disp}{}^{(2)} + E_{exch\text{-}disp}{}^{(2)}$). All calculations are performed with density fitting, asymptotically corrected exchange-correlation density functional PBE0AC recommended by the authors[50] and the aug-cc-pVDZ basis set.

**DFT Calculations with Empirical Dispersion Correction.** DFT-D combines a standard DFT calculation with an empirical correction for long-range correlation, also called dispersion or van der Waals interaction.[61,62] At the typical van der Waals distances the empirical dispersion of the form $-C_6/r^6$ is dampened by a damping function, which corrects for the nonphysical divergence and correlation double counting in the overlap region.[62] Here we used a combination of the TPSS density functional, 6−311++G(3df,3pd) basis set and dispersion parametrization based on radii scaling as described in ref 62. DFT interaction energies were not corrected for the basis set superposition error, and dispersion parameters corresponding to BSSE uncorrected calculation were chosen (by adjusting the dispersion to the BSSE uncorrected DFT energies, the BSSE error is partially mitigated). DFT calculations were performed with the TurboMole 5.10[67,68] program, and dispersion was calculated using our own Fortran code.

**SCS(MI)-MP2 Calculations.** The spin component scaled second order Møller−Plesset perturbation method (SCS-MP2) scales the relative contributions of singlet and triplet states to the correlation energy.[60] In principle any scaling factors can be used, and the SCS-MP2 method can be parametrized against a particular training set in order to obtain accurate results for some given property. The molecular interactions variant of the SCS-MP2 method, SCS(MI)-MP2, was parametrized against the S22 test set of molecular complexes which includes a balanced mixture of hydrogen bonded, dispersion bound and mixed structures.[72,73] This method, when used along with the cc-pVTZ basis set, has shown promise in terms of its ability to accurately describe noncovalent interactions at a relatively low computational cost, including potential energy curves for the hydrogen bonded and stacked uracil dimer.[74]

**Desolvation Energies.** To calculate the desolvation energy upon complex formation, i. e., the energy difference between the solvated monomers and solvated complex, we used the IEFPCM model implemented in program Gaussian 03[75] with UAKS radii, PBE functional, 6−311G** basis set and default parameters for water. Resulting dehydration energies, which are computed as Gibbs energies, were then added to the MP2/CBS interaction energies to obtain an estimate of the base pair stability in water solution. The limitations of this approach are discussed below.

**Abbreviations.** We use standard abbreviation for the base pairs, where "t" and "c" stand for trans and cis, and "W", "H", and "S" stand for Watson Crick, Hoogsteen, and Sugar edges, respectively. When ribose is included in the QM calculations (we included the ribose for those base pairs where it directly participates in the interaction), it is denoted by a preceding "r". Thus, tHS A/rG means trans-Hoogsteen Sugar Edge base pair where adenine interacts with its Hoogsteen edge with guanosine sugar edge. Amino acceptor variants of the base pairs are indicated as "aa". G=C and A-U stand for the canonical base pairs.

## Results and Discussion

Table 1 summarizes the calculated interaction energies, while Figure 2 summarizes the structures. The first three columns give the MP2/aug-cc-pVDZ, MP2/aug-cc-pVTZ, and MP2/CBS data (shown in bold). The numbers in parentheses in the third column are the CBS(T) values. The next three columns give SCS(MI)-MP2, SAPT, and DFT-D data. Finally, the last two columns list estimates of the solvation energies.

The first 22 rows in the Table 1 (rows cWS A/rG to tHS G/G) summarize data for 18 base pairs including four alternative structures with amino acceptor interactions (see above for abbreviations). The next row corresponds to the rC/rU base pair taken from the fully optimized GC/UG P-interaction; the rC/rU base pair is the key interaction in this important RNA tertiary quartet.[14,26] The subsequent six rows list pairwise interaction energies calculated for two systems that were formally optimized as trimers. The first one is the A-minor I triad, where the minor groove side of canonical rG=rC base pair acts as a receptor for adenosine through two SE/SE interactions.[10−12,26] A-minor I is the most common RNA tertiary interaction. The second trimer is a water mediated cSS rU/rC base pair. Finally,

***Table 1.*** Total Interaction Energies and Solvation Energies (kcal/mol) Calculated by Different Methods for RNA Base Pairs (Figure *2*)

| base pair | MP2/aDZ | MP2/aTZ | MP2/CBS (CBS(T)) | SCS(MI)-MP2 | DFT-SAPT aDZ | DFT-D TPSS/LP[c] | $\Delta G_{Solv}$ | MP2/CBS + $\Delta G_{Solv}$ |
|---|---|---|---|---|---|---|---|---|
| cWS A/rG | −16.00 | −17.27 | **−17.78** | −16.90 | −14.62 | −17.99 | 16.38 | −1.40 |
| cWS C/rG | −16.99 | −18.24 | **−18.74** (−17.74) | −18.10 | −15.76 | −18.87 | 17.70 | −1.04 |
| cWS C/rC | −21.84 | −23.16 | **−23.68** (−22.95) | −23.70 | −21.23 | −23.56 | 18.78 | −4.90 |
| tWS A/rA | −9.96 | −10.70 | **−11.01** (−10.62) | −10.10 | −9.01 | −10.95 | 10.30 | −0.71 |
| tWS A/rG | −16.51 | −17.84 | **−18.38** (−17.60) | −17.40 | −14.92 | −18.81 | 15.54 | −2.84 |
| tWS G/rC | −28.47 | −29.97 | **−30.58** | −31.60 | −27.92 | −30.87 | 25.14 | −5.44 |
| cSS rA/rA | −18.54 | −20.18 | **−20.84** | −20.00 | −17.32 | −21.81 | 18.36 | −2.48 |
| cSS rA/rC | −19.02 | −20.53 | **−21.15** | −20.70 | −17.67 | −21.46 | 20.64 | −0.51 |
| cSS rC/rA | −20.00 | −22.15 | **−23.02** | −22.10 | −18.17 | −23.88 | 19.91 | −3.11 |
| cSS rC/rU | −17.23 | −18.85 | **−19.49** | −18.80 | −15.73 | −19.05 | 19.24 | −0.25 |
| cSS rA/rG[d] | −23.63 | −25.55 | **−26.34** | −25.10 | - | −27.32 | 24.24 | −2.10 |
| tSS rA/rG | −21.49 | −23.41 | **−24.20** | −22.40 | −19.25 | −24.22 | 21.71 | −2.49 |
| tSS rG/rC | −13.52 | −14.13 | **−14.38** | −14.50 | −13.27 | −14.10 | 14.69 | 0.31 |
| tSS rG/rC aa[a] | −20.92 | −22.69 | **−23.41** | −22.40 | −19.27 | −23.53 | 21.88 | −1.53 |
| tWS C/rC | −9.72 | −10.19 | **−10.39** | −9.90 | −9.33 | −9.81 | 8.46 | −1.93 |
| tWS C/rC aa[a] | −14.46 | −15.58 | **−16.03** | −15.70 | −13.48 | −15.65 | 16.31 | 0.28 |
| tWS U/rC | −15.90 | −17.03 | **−17.48** | −17.70 | −15.16 | −17.28 | 14.65 | −2.83 |
| tHS A/rA | −10.23 | −11.00 | **−11.33** | −10.40 | −9.26 | −11.13 | 10.60 | −0.73 |
| tHS A/rA aa[a] | −9.96 | −10.97 | **−11.39** | −10.60 | −8.82 | −11.17 | 12.79 | 1.40 |
| tHS A/rG | −15.16 | −16.34 | **−16.82** | −15.80 | −13.64 | −17.22 | 15.69 | −1.13 |
| tHS A/rG aa[a] | −15.65 | −16.86 | **−17.36** | −16.60 | −14.00 | −17.94 | 16.70 | −0.66 |
| tHS G/G | −9.27 | −9.76 | **−9.96** | −10.00 | −9.14 | −9.95 | 9.35 | −0.61 |
| rU/rC pair of the P-motif | −16.98 | −18.30 | **−18.82** | −18.60 | −15.79 | −18.99 | 18.36 | −0.46 |
| rG/rA of A_minor I[d] | −15.22 | −16.52 | **−17.06** | −15.60 | - | −17.22 | 18.98 | 1.92 |
| rG=rC of A_minor I | −29.05 | −30.64 | **−31.29** | −31.70 | −28.38 | −31.72 | 25.08 | −6.21 |
| rA/rC of A_minor I | −17.02 | −18.50 | **−19.10** | −18.70 | −15.69 | −19.17 | 16.66 | −2.44 |
| cSS rU/rC WM[b] | −15.39 | −16.81 | **−17.38** | −17.00 | −14.32 | −15.91 | 17.27 | −0.11 |
| rU_wat WM[b] | −3.50 | −3.83 | **−3.96** | −3.68 | −3.42 | −4.29 | 3.84 | −0.12 |
| rC_wat WM[b] | −9.04 | −10.02 | **−10.42** | −10.19 | −8.71 | −11.14 | 9.44 | −0.98 |
| U-A | −15.30 | −16.52 | **−17.02** | −16.64 | −14.52 | −17.49 | 14.01 | −3.01 |
| C=G | −29.47 | −31.21 | **−31.91** | −32.42 | −28.75 | −32.37 | 24.64 | −7.27 |
| | | | | wrt MP2/CBS | wrt MP2/aDZ | wrt MP2/CBS | | |
| mean signed error | | | | 0.51 | 1.04 | −0.13 | | |
| rmsd | | | | 0.78 | 1.16 | 0.51 | | |

[a] aa: amino-acceptor. [b] WM: water mediated. [c] For the basis set description see Methods. [d] SAPT calculation failed.

the last two rows in Table 1 show data for canonical A-U and G=C pairs calculated with the same methods.

**Nature and Magnitude of the Interaction Energies in Vacuum.** Variations in total interaction energies in vacuum should be governed mainly by the number of common hydrogen bonds, their strength, and the overall complementarity of the electrostatic potentials. The MP2/CBS data support this view. The complexes with the largest interaction energies are characterized by multiple hydrogen bonds and often by strong hydrogen bonds including two oxygen atoms (OH...O) or an oxygen atom as the donor (OH...N) see, e.g., structures of G=C (−32.0 kcal/mol), tWS G/rC (−30.6 kcal/mol), cSS rA/rG (−26.3 kcal/mol), and tSS rA/rG (−24.2 kcal/mol) base pairs. On the other hand, the least stable structure, tHS G/G (−10.0 kcal/mol), contains just one NH...O bond. Thus, the base pairing strength seems to be roughly correlated with the number and character of the hydrogen bonds.

**SAPT Energy Decomposition and the Role of Dispersion Energy.** In our preceding studies we have suggested that the contribution of dispersion to the stabilities of many of the "SE" RNA base pairs is greater than its contribution to the overall stabilities of canonical base pairs.[23−26] This tentative idea was indirectly derived from the large values of the correlation components of the computed MP2/aug-cc-pVDZ interaction energies. The correlation component contains the dispersion energy (intermolecular correlation interaction energy) as the dominant contribution. However, such calculations do not allow an unambiguous decomposition because the correlation component also contains other terms, most notably the correlation correction of the electrostatic term. In addition, the balance of the HF and electron correlation components is sensitive to the intermonomer separation. Thus, in the present study we analyzed this issue using a rigorous evaluation of the dispersion energy by the SAPT procedure (Table 2). Note that the analysis is complicated by the fact that the individual energy terms are very sensitive to the intermonomer separations. The complex RNA base pairs obviously may have locally compressed or extended intermolecular contacts as the result of the overall structure balancing. This complexity will be demonstrated below on selected examples. Therefore, in order to obtain a correct picture, it is vital to analyze the relation of the dispersion, electrostatic, and induction term to the other energy terms. We suggest that the best measure of the relative role of, e.g., dispersion could be the % contribution of the dispersion energy to the overall attractive interaction (i.e., the sum of all attractive terms). This will be marked as Disp/Stab ratio in the following text and

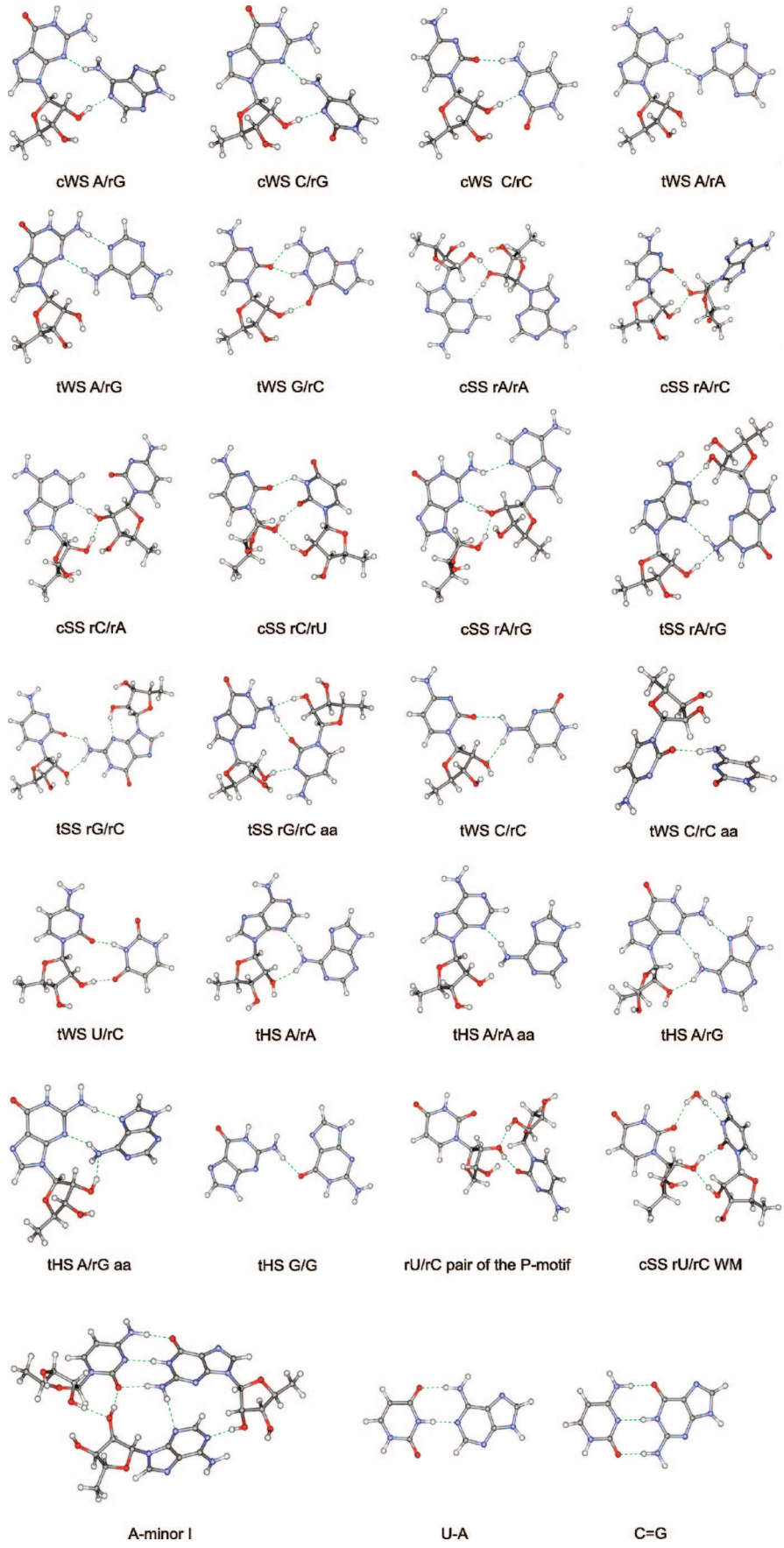


***Figure 2.*** Structures of the calculated base pairs. Outmoded nucleotide order from reference 9 is used for tSS pairs.

analogically for electrostatics and induction Elst/Stab and Ind/Stab, respectively.

The Elst/Stab ratios for the canonical G=C and A-U base pairs are 58.0% and 56.7%. This actually is quite a small

***Table 2.*** Decomposition of the Gas Phase Interaction Energies in RNA Base Pairs (kcal/mol)

| | $E_{el}$[c] | $E_{exch}$[c] | $E_{ind}$[c] | $E_{disp}$[c] | $\delta$(HF)[c] | $E^{SAPT}$ | % $E_{disp}$[d] | $\Delta E^{HF}_{CBS}$ | $\Delta E^{corr}_{MP2/CBS}$ | DFT[e] | DISP[f] |
|---|---|---|---|---|---|---|---|---|---|---|---|
| cWS A/rG | −26.4 | 32.9 | −5.8 | −11.0 | −4.4 | −14.6 | 23.2 | −8.1 | −9.7 | −13.3 | −4.7 |
| cWS C/rG | −25.8 | 31.3 | −6.0 | −11.0 | −4.3 | −15.8 | 23.4 | −9.3 | −9.5 | −14.0 | −4.9 |
| cWS C/rC | −30.6 | 33.8 | −8.0 | −10.8 | −5.6 | −21.2 | 19.6 | −16.4 | −7.3 | −19.7 | −3.9 |
| tWS A/rA | −14.7 | 18.0 | −2.5 | −7.8 | −2.0 | −9.0 | 28.8 | −4.4 | −6.6 | −6.9 | −4.0 |
| tWS A/rG | −27.2 | 33.5 | −5.6 | −11.2 | −4.5 | −14.9 | 23.1 | −9.0 | −9.4 | −13.7 | −5.1 |
| tWS G/rC | −39.6 | 40.6 | −10.4 | −11.9 | −6.7 | −27.9 | 17.3 | −24.5 | −6.0 | −27.0 | −3.9 |
| cSS rA/rA | −32.8 | 44.7 | −8.8 | −14.1 | −6.4 | −17.3 | 22.7 | −9.2 | −11.6 | −15.5 | −6.3 |
| cSS rA/rC | −29.5 | 36.7 | −7.1 | −13.0 | −4.8 | −17.7 | 23.9 | −11.1 | −10.0 | −15.1 | −6.4 |
| cSS rC/rA | −42.3 | 59.9 | −10.9 | −17.3 | −7.6 | −18.2 | 22.2 | −9.6 | −13.4 | −16.9 | −7.0 |
| cSS rC/rU | −28.2 | 37.9 | −7.3 | −13.1 | −5.0 | −15.7 | 24.4 | −9.3 | −10.2 | −13.6 | −5.4 |
| cSS rA/rG[g] | - | - | - | - | - | - | - | −11.4 | −15.0 | −18.7 | −8.6 |
| tSS rA/rG | −38.4 | 51.3 | −8.4 | −17.5 | −6.3 | −19.2 | 24.8 | −8.0 | −16.2 | −16.4 | −7.8 |
| tSS rG/rC | −16.3 | 13.9 | −3.2 | −6.0 | −1.7 | −13.3 | 22.0 | −11.0 | −3.3 | −11.4 | −2.7 |
| tSS rG/rC aa[a] | −33.4 | 42.8 | −8.2 | −14.9 | −5.7 | −19.3 | 23.9 | −11.2 | −12.2 | −17.2 | −6.3 |
| tWS C/rC | −11.5 | 10.4 | −2.1 | −5.0 | −1.1 | −9.3 | 25.5 | −6.6 | −3.7 | −7.4 | −2.4 |
| tWS C/rC aa[a] | −20.8 | 25.6 | −4.9 | −9.9 | −3.5 | −13.5 | 25.2 | −8.5 | −7.6 | −11.3 | −4.3 |
| tWS U/rC | −24.2 | 28.8 | −6.4 | −8.6 | −4.7 | −15.2 | 19.7 | −12.2 | −5.2 | −14.7 | −2.6 |
| tHS A/rA | −15.0 | 18.1 | −2.4 | −8.0 | −1.9 | −9.3 | 29.3 | −4.6 | −6.7 | −7.1 | −4.0 |
| tHS A/rA aa[a] | −17.2 | 23.5 | −3.3 | −9.3 | −2.5 | −8.8 | 28.7 | −3.4 | −8.0 | −6.9 | −4.3 |
| tHS A/rG | −23.2 | 28.0 | −4.5 | −10.4 | −3.5 | −13.6 | 24.9 | −8.0 | −8.8 | −12.2 | −5.0 |
| tHS A/rG aa[a] | −24.5 | 30.7 | −5.1 | −11.3 | −3.8 | −14.0 | 25.3 | −7.5 | −9.9 | −12.4 | −5.6 |
| tHS G/G | −11.7 | 10.9 | −2.3 | −4.7 | −1.5 | −9.1 | 23.3 | −7.1 | −2.9 | −7.9 | −2.0 |
| rU/rC pair of the P-motif | −25.2 | 31.8 | −6.8 | −11.1 | −4.4 | −15.8 | 23.4 | −10.8 | −8.0 | −13.9 | −5.1 |
| rG/rA of A_minor I[g] | - | - | - | - | - | - | - | −4.9 | −12.1 | −10.7 | −6.5 |
| rG=rC of A_minor I | −41.5 | 43.3 | −10.5 | −12.0 | −7.7 | −28.4 | 16.7 | −25.4 | −5.9 | −27.9 | −3.8 |
| rA/rC of A_minor I | −29.2 | 36.3 | −6.9 | −11.4 | −4.5 | −15.7 | 21.9 | −10.7 | −8.4 | −14.7 | −4.4 |
| cSS rU/rC WM[b] | −24.6 | 34.0 | −7.3 | −11.2 | −5.2 | −14.3 | 23.2 | −8.5 | −8.9 | −12.9 | −3.1 |
| rU_wat WM[b] | −6.2 | 7.0 | −1.2 | −2.3 | −0.7 | −3.4 | 22.4 | −2.7 | −1.3 | −3.5 | −0.8 |
| rC_wat WM[b] | −20.8 | 24.5 | −4.0 | −5.8 | −2.6 | −8.7 | 17.4 | −6.4 | −4.0 | −9.7 | −1.5 |
| U-A | −28.1 | 35.0 | −6.5 | −9.5 | −5.5 | −14.5 | 19.1 | −10.6 | −6.4 | −14.4 | −3.1 |
| C=G | −44.2 | 47.5 | −11.3 | −12.2 | −8.5 | −28.8 | 16.1 | −25.8 | −6.1 | −28.8 | −3.5 |

[a] aa: amino-acceptor. [b] WM: water mediated. [c] $E_{el}$: electrostatic component; $E_{exch}$: exchange repulsion; $E_{ind}$: induction (sum of $E_{ind}$ and $E_{exch\text{-}ind}$); $E_{disp}$: dispersion (sum of $E_{disp}$ and $E_{exch\text{-}disp}$); $\delta$(HF): delta HF correction for higher order induction terms (see eq 5). [d] Percent of the dispersion attraction with respect to the total attraction ($100*E_{disp}/(E_{disp} + E_{ind} + E_{el} + \delta(HF))$). [e] DFT term from the DFT-D TPSS/LP calculations. [f] Dispersion term from the DFT-D TPSS/LP calculations. [g] SAPT calculation failed.

difference, taking into account that the A-U base pair is assumed to be less polar than the G=C one due to the considerably smaller polarity of the interacting monomers.[76] When considering all studied interactions, the Elst/Stab ratio ranges from 50% (rU...rC dimer from the corresponding water mediated structure) to 63% (rC interaction with water from the same structure). However, the later two structures are obvious outliers. Waters typically bridge the most polar sites of the base pairs, and then the direct nucleoside-nucleoside contacts in such water mediated structures are less electrostatic.

For the noncanonical RNA base pairs with conventional H-bonding (disregarding the alternative amino acceptor geometries), the Elst/Stab ratio is in the range of 52.6 to 59.9%. Comparing with the data for canonical base pairs this appears to be a significant variability. For most of the base pairs this ratio is smaller than 56%, indicating a weaker relative contribution of the electrostatics compared to the canonical base pairs.

The complexity of the SAPT analysis can be demonstrated for the amino-acceptor geometries of the base pairs. For example, the conventional binding of the tSS rG/rC base pair is associated with the small electrostatic term of −16.3 kcal/mol, which, at first sight, could indicate weak electrostatics. However, the Elst/Stab ratio is 59.9% which is the largest value among all base pairs. Then, the amino-acceptor variant of this base pair has an electrostatic term of −33.4 kcal/mol. This, at first sight, indicates an increase of the electrostatic nature of the base pair. This is obviously counterintuitive, as replacing one standard H-bond by an amino-acceptor interaction should reduce the electrostatic contribution to the total interaction. Indeed, for the amino-acceptor geometry all terms (in absolute value) increase, while the Elst/Stab ratio drops to 53.7%. These numbers nicely illustrate that, due to the sensitivity of the individual SAPT terms to interatomic distances, the decomposition is not straightforward and can be easily misinterpreted. As such, the literature interpretations of these values should be treated with great care.

The induction component (sum of the $E_{ind}$ and $\delta$(HF) terms, Ind/Stab) contributes between 16% for the tHS A/rA base pair and 26% for the rU/C interaction from the water mediated complex. It parallels the trends in the overall strengths (statistically, $R^2 = 0.59$). The induction correlates very nicely with the electrostatics ($R^2 = 0.96$), which in part explains the success of the pair additive empirical force fields, which use systematically overestimated charges (dipoles) to substitute for the missing polarization term.

The dispersion contribution to the sum of all attractive terms, Disp/Stab ($100*E_{disp}/(E_{disp} + E_{ind} + E_{el} + \delta(HF))$), varies between 16% in the G=C base pair and 29% in the tHS A/rA base pair (see Table 2). The relative dispersion contribution is smallest in the strongest complex and largest in one of the weakest complexes, but this is rather accidental, because the strength of the complex does not correlate with its dispersion contribution. In absolute values the dispersion

is largest in the extended complexes in which the aliphatic sugar hydrogen appears close to the interacting molecule, i.e., mostly in the SS rX/rY base pairs.

Closer inspection of the data, nevertheless, reveals a change in the balance between dispersion and electrostatic energies in many RNA base pairs, when these base pairs are compared with the reference canonical A-U and G=C base pairs. While the G=C base pair is known to be a strong electrostatic pair formed by two bases with large dipole moments, the A-U base pair has weaker electrostatics. Thus, we can consider properties of the A-U and G=C base pairs as demarcating the conventional range of the balance of forces in nucleic acids base pairs. To get the full picture, we consider not only the overall trends and correlations but also the individual cases and outliers, since the diversity of molecular interactions provides the variability of RNA molecular structures. RNA base pairs play numerous roles in nature, and trying to characterize them only by "average" base pairing properties and overall correlations could be like mixing apples and oranges.

The SAPT dispersion components of the A-U and G=C base pairs are −9.5 and −12.2 kcal/mol, respectively. The dispersion components in RNA base pairs (considering the first 23 rows of Table 1) vary between −4.7 and −17.5 kcal/mol, with the largest dispersion component in the cSS rC/rA (−17.3 kcal/mol) and tSS rA/rG (−17.5 kcal/mol) base pairs. This is also confirmed by the Disp/Stab ratio, which is 19 and 16% for the A-U and G=C base pairs, but in the range of 20−29% for the vast majority of RNA base pairs. The dispersion vs electrostatics ratio is 0.34:1 and 0.28:1 in the A-U and G=C base pairs and increases to 0.46:1 in the tSS rA/rG base pair and 0.43:1 in the rU/rC dimer from the P-interaction. The tSS rA/rG base pair is the key interaction stabilizing the A-minor I tertiary interaction. The differences are also visualized by the ratio between the dispersion energy and the total SAPT interaction energy, which is 0.42:1, 0.66:1, 0.70:1, and 0.91:1 in G=C, A-U, rC/rU P-interaction, and tSS rA/rG base pairs, respectively. This value is 0.75:1 in cWS A/rG base pair, etc. Therefore, although the data confirm that the RNA base pairs generally exhibit behavior typical of H-bonded systems, some of them clearly profit from an increased role of dispersion energy compared to canonical base pairs. Although the differences appear modest in the gas phase, they might be important in a natural RNA environment, where the primary H-bonds of the base pairs always compete with binding of water molecules. The increased dispersion is likely important mainly for the tertiary interactions which are exposed to solvent much more than Watson−Crick base pairs that are sheltered and stacked inside double helices.

Let us now comment on other options to separate the dispersion energy from the other contributions. A crude estimate is based on the evaluation of the MP2 (correlation) contribution to the intermolecular energy, i.e., the difference between the MP2 and HF interaction energies. Although this difference is often dominated by dispersion, it contains also the MP2 corrections to the electrostatics and induction, which cannot be separated. The data in Table 2 reveal that the MP2 intermolecular contribution is about 2.6 kcal/mol lower than that of the SAPT dispersion term, but it correlates with the DFT-SAPT dispersion values fairly well, with $R^2 = 0.83$. Because these lower dispersion values can be partly explained by the, typically repulsive, MP2 correction to electrostatics, it seems that the main contribution to the MP2 correction indeed comes from dispersion. Note that these conclusions are valid for hydrogen bonded complexes and should not be generalized to other types of bonding interactions without further justification. The studied base pairs show a large degree of variability in the ratio between HF and correlation contributions to the interaction energies. This ratio is 4.2:1 and 1.7:1 for the G=C and A-U base pairs, contrasting the values of 0.7:1 and 0.49:1 obtained for cSS rC/rA and tSS rA/rG RNA base pairs, i.e., for those RNA base pairs with the largest dispersion energy. However, closer inspection of the available data reveals that the actual differences in binding are not as dramatic. The difference between the A-U and G=C base pairs is not surprising, as the strong electrostatics of the G=C base pair is well-known. However, a substantial part of the difference in the ratios of HF and correlation components for these two base pairs is due to the correlation correction of the electrostatic term, caused by intramolecular electron correlation reduction of the dipoles of guanine and cytosine. It reduces mainly the correlation interaction energy of the G=C base pair and is not related to the dispersion energy which is due to intermolecular electron correlation contributions. This effect for the A-U base pair is considerably smaller as their smaller dipoles are much less affected by the electron correlation. Furthermore, close inspection of the SAPT data for the two RNA base pairs reveals that their large dispersion (the largest among systems studied here) and electron correlation energies are associated with the largest exchange repulsion terms of 59.9 and 51.3 kcal/mol. This evidently contributes to the large change in the HF vs correlation interaction energy ratio, especially for the cSS rC/rA base pair. Thus, the increased correlation interaction energy component rather reflects the larger compactness of the base pairs, which (in absolute values) increases magnitude of all the energy components. In other words, although we see that in some H-bonded RNA base pairs the role of dispersion energy is somewhat increased, this effect is smaller than one would guess from the ratio of HF and correlation components of the interaction energies.

Another way to approximate the dispersion contribution is based on empirical formulas, such as those used, along with an appropriate damping function, in the DFT-D method.[62] Table 2 shows that the empirical dispersion is far smaller in magnitude than the DFT-SAPT reference (on average by more than 6 kcal/mol). This is in agreement with our earlier study of DNA bases,[55] in which the undamped empirical dispersion was in fairly good agreement with the DFT-SAPT reference, but damping significantly reduced the magnitude of dispersion, especially in the hydrogen bonded complexes (by about 80%). In other orientations of bases, for instance in stacks, dispersion is damped much less (often by only about 20%). In this respect, the damped empirical dispersion should be understood as a complement to a given DFT functional rather than as an estimate of the dispersion

contribution, and therefore care is needed when assessing the importance of dispersion from the DFT-D data. Still, within a group of similar complexes (hydrogen bonded complexes here), relative dispersion contribution is informative, and the empirical dispersion correlates fairly well with the DFT-SAPT values here. As pointed out in ref 62, damping is in DFT-D used mainly to correct for double counting of the interaction energy at short distances, and it has little physical meaning. Damping is necessary, because current density functionals seemingly cover some portion of the correlation attraction at short distances. However, this attraction is largely spurious as it comes in part from the physically incorrect behavior of the exchange functional (for more details see ref 62). In this respect, decomposing the (supermolecular) DFT interaction energy is dubious. The perturbative (as opposed to supermolecular) DFT-SAPT values are much more meaningful, and they are necessary to obtain a rigorous estimate of the dispersion contribution to intermolecular interactions.

We have also tested whether the strength of the dispersion interaction can be related to the contact area between the bases, estimated as a half of the difference between the surface of the base pair and the surfaces of the isolated bases. The surfaces of the monomers and dimers were taken from the Gaussian 03 solvation calculations with standard UAKS radii of the atoms and default cavity construction algorithm. The contact area indeed correlates with the dispersion component of the interaction obtained by the SAPT method with $R^2 = 0.81$ (intercept forced to 0) and also with the damped dispersion from DFT-D ($R^2 = 0.78$). The average surface dispersion energy is $-0.52$ kcal mol$^{-1}$ Å$^{-2}$ for SAPT and $-0.22$ kcal mol$^{-1}$ Å$^{-2}$ for DFT-D. Although this correlation is statistically significant, dispersion prediction based on the contact area would be rather inaccurate. Also, our set of complexes is fairly homogeneous, and if another type of interaction was included, the correlation would likely be worse.

**Correction for the Higher Order Electron Correlation Effects.** In the four complexes for which we were able to calculate the ΔCCSD(T) correction (cWS C/rG, cWS C/rC, tWS A/rA, and tWS A/rG), the value of this term was smaller than 1 kcal/mol (see Table 1). For purely hydrogen bonded complexes this correction is usually very close to zero, and the value of roughly 1 kcal/mol may be related to the somewhat larger dispersion contribution in the present systems (as indicated by DFT-SAPT analysis). Due to the fact that the systems studied here interact in similar manners, the CCSD(T) corrections of approximately the same magnitude can also be expected for the remaining complexes in the present set. When we take into account the expected underestimation of the MP2/CBS limit by the aDZ/aTZ extrapolation, which should partially cancel out the missing CCSD(T) correction term, the value of approximately ±1 kcal/mol can be viewed as a reasonable estimate of the maximum error in present MP2/CBS interaction energies.

**Comparison of Different QM Methods.** Table 1 also compares the performance of different methods used to calculate interaction energies in vacuum. Comparisons are made with respect to the MP2/CBS energies. For the sake of consistency MP2/CBS energies were used as the reference even for structures where the ΔCCSD(T) correction terms are available.

The partly semiepirical SCS(MI)-MP2 by Distasio and Head-Gordon[72] paired with the cc-pVTZ basis set (column 5 in Table 1) compares fairly well with the MP2/CBS results with an average signed error of 0.51 kcal/mol and a root-mean-square deviation (rmsd) of 0.78 kcal/mol. An average underestimation of 0.51 kcal/mol is a good result, because the reference MP2 values likely on average overestimate the real value, as indicated by the available ΔCCSD(T) corrections. The SCS(MI)-MP2 − MP2/CBS differences cover a range from ∼−1 kcal/mol to +2.2 kcal/mol.

The DFT-D calculations with the TPSS functional and 6−311++G(3df,3pd) (LP) basis set demonstrate the high quality of the DFT-D relative energies (rmsd only 0.51 kcal/mol) and also yield small average errors (−0.13 kcal/mol). If we take into account the above-mentioned MP2 overestimation of the binding energies, it seems that the DFT-D method probably also slightly overestimates the interaction strengths, which is attributable to the well-known overestimation of hydrogen bonding stabilities in DFT-D.[62] Because of its speed/accuracy ratio the DFT-D method seems to be an excellent compromise. In addition, all these discussed differences are so subtle that they are expected to be insignificant in most applications. The DFT-D − MP2/CBS differences cover a range from ∼−1.0 kcal/mol to +1.5 kcal/mol.

The DFT-SAPT/aDZ calculations were compared with the MP2/aDZ results rather than with the MP2/CBS for the sake of basis set consistency. DFT-SAPT interaction energies are on average about 1 kcal/mol smaller (in absolute value), which is partly because MP2 overestimates dispersion and in part because of different basis set convergence behavior of the DFT-SAPT and MP2 methods. rmsd is rather large here, but standard deviation, which is a better measure of accuracy in the case of large systematic error, is quite small (0.53 kcal/mol). Because DFT-SAPT is often considered a more accurate method than MP2 this increases our confidence in the MP2/CBS data as a reference. The largest difference between DFT-SAPT and MP2/aDZ (2.24 kcal/mol) is found for the tSS rA/rG base pair.

**Many-Body Terms.** In the clusters containing three or more molecules the total interaction energy is not simply a sum of all the pair interaction energies. The difference between the pairwise sum and the total interaction energy - the nonadditivity - consists of the sum of the many-body terms. Two of our model complexes consist of three molecules (the A-minor I and the water mediated complex) where the 3-body terms arise. We calculated the 3-body term at the MP2/aug-cc-pVDZ level, and it amounts to −0.37 kcal/mol for the water mediated rU...rC pair and 0.84 kcal/mol for the A-minor I trimer. These terms are rather small, and it is unlikely that omitting them would affect any conclusion derived in this article. Inclusion of many body terms would be needed for quantitative analyses of larger clusters. This result is entirely consistent with calculations of base trimers.[77] Note that the MP2 method does not include eventual dispersion nonadditivity.

**Inclusion of Solvent Effects.** Within an aqueous environment, the electrostatic interaction loses much of its strength, which is reflected in the positive solvation Gibbs energies (column 8 of Table 1). Taken together with the gas phase interaction energies the resulting stabilizations in water (the last column of Table 1) are in line with our expectations for hydrogen bonded complexes. The calculated free energies of A-U and G=C base pairs are −3 and −7 kcal/mol, respectively. It is roughly in agreement with a contribution of −2 to −3 kcal/mol per H-bond suggested by force field calculations;[78] note, however, that these force field calculations include only the solvation effect on the electrostatic interactions.

The corresponding experimental estimates of base pair stability of −1 to −2 kcal/mol[79−81] per base pair are smaller in absolute value than our calculated values. The experiments, however, also include the destabilizing contribution of the configuration entropy including the cost of bringing the monomers together. In addition, the experimental data are derived for base pairs embedded within nucleic acids. It is not straightforward to subtract the net base pair stability from the overall free energy data. Data for H-bonded base pairs in water are in fact not available since base stacking is preferred over H-bonding in water; the free energy of stacking association in water is 0 to −1 kcal/mol, and the base pairing should thus be even less stable.[82−85]

Considering all these data, our calculations most likely exaggerate stabilities of canonical base pairs in water and also the difference in stability of A-U and G=C base pairs. We do not know, however, how much of this exaggeration is due to neglect of the configuration entropy changes. The paucity and complexity of the experimental data do not allow us to make any unambiguous conclusion.

Available theoretical data are also mutually inconsistent. Using MD free energy simulations with the AMBER force field, Stofer et al.[86] predicted values of −4.3 and −5.8 kcal/mol for A-T and G=C base pairs. A Langevine Dipole (LD) study by Florian et al.[87] predicted values of −0.8 and −1.8, apparently being much closer to the range expected based on the experimental data. The LD method was adjusted to reproduce the available experimental values of stacking free energies in water.

The G=C base pair is predicted to be the most stable one among all studied base pairs by our calculations, which is probably a correct result, taking into consideration the key role of this base pair in RNA thermodynamics. The calculations indicate that, among the noncanonical RNA base pairs, the most stable complexes in water are the tWS G/rC and cWS C/rC ones. The least stable RNA base pair is the tSS rG/rC one, when considering conventional binding only and disregarding the trimers. rG/rA of A-minor I and tHS A/rA amino-acceptor structures are predicted to be even less stable. Perhaps surprisingly, the tSS rA/rG base pair, which has the largest dispersion contribution, possesses only medium stability upon inclusion of continuum solvent effects. As noted above, however, we have to keep in mind that the calculated Gibbs solvation energies are of significantly lower accuracy than the gas phase interaction energies due to many approximations in the continuum model calculations. Also, the dehydration energy can be significantly different when a chosen motif is immersed in a certain RNA environment (the present calculations assume that both the pair and the isolated monomers are fully hydrated). Although water is probably a good average representation of the highly hydrated RNA structure, local interactions can significantly influence the stability of a given motif, both through modified solvation exposure and through interplay with additional interactions. Another factor, which is not included in our calculations, is the change of the configuration entropy of the solute (we add Gibbs solvation energies to interaction energies). Configuration entropy is very hard to estimate and depends critically on the local structure through flexibility. Our stabilization energies in the water environment should thus be viewed only as very rough estimates. We presently do not see any straightforward way how to improve the reliability of the RNA base pair solvation calculations or to independently verify them.

The solvent calculations can be perhaps used to get relative stability ranking of the RNA base pairs. However, even here some caution is needed. As the base pairs have quite variable shapes and thus solvent exposures, their stability evaluation can be biased by variable sensitivity to approximations of the solvent calculations. It is known from thermodynamics experiments that non-WC base pairs typically destabilize A-RNA helices; however, these effects are known to be context-dependent. In addition, it is known that double-helices are optimally sterically suited for canonical base pairs. In fact, the base pairing stabilization effects may be much more complicated than usually assumed. The stabilization or destabilization associated with a given base pair may very much depend on its context, on its complementarity to all the other interactions around, including specific water and ion binding. Thus, a given base pair may have very different influence on stabilization depending on where is the base pair inserted. Understanding of this "promiscuity" of molecular interactions would be a major step forward in studies of RNA structure and folding and, in our opinion, is one of the main challenges for computational chemistry of nucleic acids.

## Conclusions

We have carried out reference quantum chemical calculations of base pairing energies for a representative selection of 25 diverse RNA base pairs utilizing the ribose moiety through the 2′-OH group, including structures with amino acceptor interactions. Such extended RNA base pairs are of primary importance for building up the complex three-dimensional structures of RNA molecules, which, thus far, have been largely ignored in the quantum chemical and physical chemistry literature. Since the RNA base pairs bring new interactions not present in standard base pairing, we suggest that some of our complexes should be added to the portfolio of structures considered in parametrization of computational methods that are designed to study RNA.

The base pairs were evaluated at the MP2 level with extrapolation to the complete basis set limit (CBS) of atomic orbitals. CCSD(T) correction terms were obtained for four base pairs. The interaction energy decomposition has been

performed using the DFT-SAPT perturbational procedure with the aug-cc-pVDZ basis set. Many RNA base pairs have a modestly increased role of dispersion attraction compared to canonical base pairs. However, the effect is smaller than one would assume based on assessment of the ratio of HF and correlation components of the interaction energies. The increased role of dispersion energy is nevertheless assumed to be important for stabilization of RNA tertiary interactions. Complexity of the interpretation of the SAPT energy decomposition is discussed.

SCS(MI)-MP2 and DFT-D methods have exhibited very good performance for RNA base pairs involving the ribose 2′-OH interactions as well as amino acceptor interactions. The differences between these methods and the reference data are nevertheless visibly larger than those achieved for canonical base pairs, with uncertainty in the calculated relative energies of ∼3 kcal/mol. The DFT-D method produces results that are generally closer to the reference data. We also roughly estimate the effect of aqueous solvent screening on the base pairing stability using continuum solvent approach.

The RNA base pairs are very diverse molecular interactions which offer a range of structures with different shapes, stabilities, and balance of interaction energy contributions.

**Acknowledgment.** This contribution was supported by the Grant Agency of the Academy of Sciences of the Czech Republic, grants No. IAA400040802 and IAA400550701; the Ministry of Education of the Czech Republic, grants LC06030, LC512, MSM0021622413, and MSM6198959216 (P.J.); and by the Academy of Sciences of the Czech Republic, grants no. AV0Z50040507, AV0Z50040702, and Z40550506. P.H. acknowledges the support from Praemium Academiae, Academy of Sciences of the Czech Republic, dedicated to P.H. in 2007. K.R. gratefully acknowledges the support of the NSF EPSCOR program (EPS-0701525). We thank Zdenek Salvet for kind help with the computer facilities in Brno.

**Supporting Information Available:** Structures of studied base pairs. This material is available free of charge via the Internet at http://pubs.acs.org.

## References

(1) Cate, J. H.; Gooding, A. R.; Podell, E.; Zhou, K.; Golden, B. L.; Kundrot, C. E.; Cech, T. R.; Doudna, J. A. Crystal Structure of a Group I Ribozyme Domain: Principles of RNA Packing. *Science* **1996**, *273*, 1678–1685.

(2) Ban, N.; Nissen, P.; Hansen, J.; Moore, P. B.; Steitz, T. A. The Complete Atomic Structure of the Large Ribosomal Subunit at 2.4 A Resolution. *Science* **2000**, *289*, 905–920.

(3) Wimberly, B. T.; Brodersen, D. E.; Clemons, W. M., Jr.; Morgan-Warren, R. J.; Carter, A. P.; Vonrhein, C.; Hartsch, T.; Ramakrishnan, V. Structure of the 30S Ribosomal Subunit. *Nature* **2000**, *407*, 327–339.

(4) Moore, P. B.; Steitz, T. A. The Structural Basis of Large Ribosomal Subunit Function. *Annu. Rev. Biochem.* **2003**, *72*, 813–850.

(5) Schuwirth, B. S.; Borovinskaya, M. A.; Hau, C. W.; Zhang, W.; Vila-Sanjurjo, A.; Holton, J. M.; Cate, J. H. Structures of the Bacterial Ribosome at 3.5 A Resolution. *Science* **2005**, *310*, 827–834.

(6) Korostelev, A.; Trakhanov, S.; Laurberg, M.; Noller, H. F. Crystal Structure of a 70S Ribosome-tRNA Complex Reveals Functional Interactions and Rearrangements. *Cell* **2006**, *126*, 1065–1077.

(7) Selmer, M.; Dunham, C. M.; Murphy, F.; Weixlbaumer, A.; Petry, S.; Kelley, A. C.; Weir, J. R.; Ramakrishnan, V. Structure of the 70S Ribosome Complexed with mRNA and tRNA. *Science* **2006**, *313*, 1935–1942.

(8) Leontis, N. B.; Westhof, E. Conserved Geometrical Base-Pairing Patterns in RNA. *Q. Rev. Biophys.* **1998**, *31*, 399–455.

(9) Leontis, N. B.; Stombaugh, J.; Westhof, E. The Non-Watson-Crick Base Pairs and Their Associated Isostericity Matrices. *Nucleic Acids Res.* **2002**, *30*, 3497–3531.

(10) Nissen, P.; Ippolito, J. A.; Ban, N.; Moore, P. B.; Steitz, T. A. RNA Tertiary Interactions in the Large Ribosomal Subunit: the A-Minor Motif. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 4899–4903.

(11) Tamura, M.; Holbrook, S. R. Sequence and Structural Conservation in RNA Ribose Zippers. *J. Mol. Biol.* **2002**, *320*, 455–474.

(12) Lescoute, A.; Westhof, E. The A-Minor Motifs in the Decoding Recognition Process. *Biochimie* **2006**, *88*, 993–999.

(13) Gagnon, M. G.; Steinberg, S. V. GU Receptors of Double Helices Mediate tRNA Movement in the Ribosome. *RNA* **2002**, *8*, 873–877.

(14) Mokdad, A.; Krasovska, M. V.; Šponer, J.; Leontis, N. B. Structural and Evolutionary Classification of G/U Wobble Basepairs in the Ribosome. *Nucleic Acids Res.* **2006**, *34*, 1326–1341.

(15) Leontis, N. B.; Westhof, E. The 5S rRNA Loop E: Chemical Probing and Phylogenetic Data Versus Crystal Structure. *RNA* **1998**, *4*, 1134–1153.

(16) Lescoute, A.; Leontis, N. B.; Massire, C.; Westhof, E. Recurrent Structural RNA Motifs, Isostericity Matrices and Sequence Alignments. *Nucleic Acids Res.* **2005**, *33*, 2395–2409.

(17) Mokdad, A.; Frankel, A. D. ISFOLD: Structure Prediction of Base Pairs in Non-Helical RNA Motifs from Isostericity Signatures in Their Sequence Alignments. *J. Biomol. Struct. Dyn.* **2008**, *25*, 467–472.

(18) Sarver, M.; Zirbel, C. L.; Stombaugh, J.; Mokdad, A.; Leontis, N. B. FR3D: Finding Local and Composite Recurrent Structural Motifs in RNA 3D Structures. *J. Math. Biol.* **2008**, *56*, 215–252.

(19) Razga, F.; Koča, J.; Mokdad, A.; Šponer, J. Elastic Properties of Ribosomal RNA Building Blocks: Molecular Dynamics of the GTPase-Associated Center rRNA. *Nucleic Acids Res.* **2007**, *35*, 4007–4017.

(20) Battle, D. J.; Doudna, J. A. Specificity of RNA-RNA Helix Recognition. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 11676–11681.

(21) Nasalean, L.; Stombaugh, J.; Zirbel, C. L.; Leontis, N. B. RNA 3D Structural Motifs: Definition, Identification, Annotation, and Database Searching. In *Non-Protein Coding RNAs*, Walter, N. G., Woodson, S. A., Batey, R. T., Eds.; Springer: Berlin, 2009; Vol. 13, pp 1−26.

(22) Šponer, J.; Mokdad, A.; Šponer, J. E.; Špačková, N.; Leszczynski, J.; Leontis, N. B. Unique Tertiary and Neighbor Interactions Determine Conservation Patterns of

cis Watson-Crick A/G Base-pairs. *J. Mol. Biol.* **2003**, *330*, 967–978.

(23) Šponer, J. E.; Špačková, N.; Kulhánek, P.; Leszczynski, J.; Šponer, J. Non-Watson-Crick Base Pairing in RNA Quantum Chemical Analysis of the cis Watson-Crick/Sugar Edge Base Pair Family. *J. Phys. Chem. A* **2005**, *109*, 2292–2301.

(24) Šponer, J. E.; Špačková, N.; Leszczynski, J.; Šponer, J. Principles of RNA Base Pairing: Structures and Energies of the trans Watson-Crick/Sugar Edge Base Pairs. *J. Phys. Chem. B* **2005**, *109*, 11399–11410.

(25) Šponer, J. E.; Leszczynski, J.; Sychrovský, V.; Šponer, J. Sugar Edge/Sugar Edge Base Pairs in RNA: Stabilities and Structures from Quantum Chemical Calculations. *J. Phys. Chem. B* **2005**, *109*, 18680–18689.

(26) Šponer, J. E.; Réblová, K.; Mokdad, A.; Sychrovský, V.; Leszczynski, J.; Šponer, J. Leading RNA Tertiary Interactions: Structures, Energies, and Water Insertion of A-Minor and P-Interactions. A Quantum Chemical View. *J. Phys. Chem. B* **2007**, *111*, 9153–9164.

(27) Vokáčová, Z.; Šponer, J.; Šponer, J. E.; Sychrovský, V. Theoretical Study of the Scalar Coupling Constants across the Noncovalent Contacts in RNA Base Pairs: the cis- and trans-Watson-Crick/Sugar Edge Base Pair Family. *J. Phys. Chem. B* **2007**, *111*, 10813–10824.

(28) Roy, A.; Panigrahi, S.; Bhattacharyya, M.; Bhattacharyya, D. Structure, Stability, and Dynamics of Canonical and Noncanonical Base Pairs: Quantum Chemical Studies. *J. Phys. Chem. B* **2008**, *112*, 3786–3796.

(29) Bhattacharyya, D.; Koripella, S. C.; Mitra, A.; Rajendran, V. B.; Sinha, B. Theoretical Analysis of Noncanonical Base Pairing Interactions in RNA Molecules. *J. Biosci.* **2007**, *32*, 809–825.

(30) Brandl, M.; Meyer, M.; Sühnel, J. Water-Mediated Base Pairs in RNA: A Quantum-Chemical Study. *J. Phys. Chem. A* **2000**, *104*, 11177–11187.

(31) Oliva, R.; Tramontano, A.; Cavallo, L. $Mg^{2+}$ Binding and Archaeosine Modification Stabilize the G15 C48 Levitt Base Pair in tRNAs. *RNA* **2007**, *13*, 1427–1436.

(32) Oliva, R.; Cavallo, L.; Tramontano, A. Accurate Energies of Hydrogen Bonded Nucleic Acid Base Pairs and Triplets in tRNA Tertiary Interactions. *Nucleic Acids Res.* **2006**, *34*, 865–879.

(33) Schneider, C.; Brandl, M.; Sühnel, J. Molecular Dynamics Simulation Reveals Conformational Switching of Water-Mediated Uracil-Cytosine Base-Pairs in an RNA Duplex. *J. Mol. Biol.* **2001**, *305*, 659–667.

(34) Réblová, K.; Špačková, N.; Štefl, R.; Csaszar, K.; Koča, J.; Leontis, N. B.; Šponer, J. Non-Watson-Crick Basepairing and Hydration in RNA Motifs: Molecular Dynamics of 5S rRNA Loop E. *Biophys. J.* **2003**, *84*, 3564–3582.

(35) Razga, F.; Koča, J.; Šponer, J.; Leontis, N. B. Hinge-Like Motions in RNA Kink-Turns: the Role of the Second A-Minor Motif and Nominally Unpaired Bases. *Biophys. J.* **2005**, *88*, 3466–3485.

(36) Krasovska, M. V.; Sefcikova, J.; Réblová, K.; Schneider, B.; Walter, N. G.; Šponer, J. Cations and Hydration in Catalytic RNA: Molecular Dynamics of the Hepatitis Delta Virus Ribozyme. *Biophys. J.* **2006**, *91*, 626–638.

(37) Špačková, N.; Šponer, J. Molecular Dynamics Simulations of Sarcin-Ricin rRNA Motif. *Nucleic Acids Res.* **2006**, *34*, 697–708.

(38) Frank, J.; Spahn, C. M. T. The Ribosome and the Mechanism of Protein Synthesis. *Rep. Prog. Phys.* **2006**, *69*, 1383–1417.

(39) Mitra, K.; Frank, J. Ribosome Dynamics: Insights from Atomic Structure Modeling into Cryo-Electron Microscopy Maps. *Annu. Rev. Biophys. Biomol. Struct.* **2006**, *35*, 299–317.

(40) Ferner, J.; Villa, A.; Duchardt, E.; Widjajakusuma, E.; Wöhnert, J.; Stock, G.; Schwalbe, H. NMR and MD Studies of the Temperature-Dependent Dynamics of RNA YNMG-Tetraloops. *Nucleic Acids Res.* **2008**, *36*, 1928–1940.

(41) Musselman, C.; Al-Hashimi, H. M.; Andricioaei, I. iRED Analysis of TAR RNA Reveals Motional Coupling, Long-Range Correlations, and a Dynamical Hinge. *Biophys. J.* **2007**, *93*, 411–422.

(42) Réblová, K.; Fadrná, E.; Sarzynska, J.; Kulinski, T.; Kulhánek, P.; Ennifar, E.; Koča, J.; Šponer, J. Conformations of Flanking Bases in HIV-1 RNA DIS Kissing Complexes Studied by Molecular Dynamics. *Biophys. J.* **2007**, *93*, 3932–3949.

(43) Shankar, N.; Xia, T.; Kennedy, S. D.; Krugh, T. R.; Mathews, D. H.; Turner, D. H. NMR Reveals the Absence of Hydrogen Bonding in Adjacent UU and AG Mismatches in an Isolated Internal Loop from Ribosomal RNA. *Biochemistry* **2007**, *46*, 12665–12678.

(44) Chen, G.; Kierzek, R.; Yildirim, I.; Krugh, T. R.; Turner, D. H.; Kennedy, S. D. Stacking Effects on Local Structure in RNA: Changes in the Structure of Tandem GA Pairs when Flanking GC Pairs are Replaced by isoG-isoC Pairs. *J. Phys. Chem. B* **2007**, *111*, 6718–6727.

(45) Mathews, D. H.; Sabina, J.; Zuker, M.; Turner, D. H. Expanded Sequence Dependence of Thermodynamic Parameters Improves Prediction of RNA Secondary Structure. *J. Mol. Biol.* **1999**, *288*, 911–940.

(46) Riley, K. E.; Hobza, P. Assessment of the MP2 Method, along with Several Basis Sets, for the Computation of Interaction Energies of Biologically Relevant Hydrogen Bonded and Dispersion Bound Complexes. *J. Phys. Chem. A* **2007**, *111*, 8257–8263.

(47) 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. A 2nd Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.

(48) Šponer, J.; Jurečka, P.; Hobza, P. Accurate Interaction Energies of Hydrogen-Bonded Nucleic Acid Base Pairs. *J. Am. Chem. Soc.* **2004**, *126*, 10142–10151.

(49) Hesselmann, A.; Jansen, G.; Schütz, M. Density-Functional Theory-Symmetry-Adapted Intermolecular Perturbation Theory with Density Fitting: A New Efficient Method to Study Intermolecular Interaction Energies. *J. Chem. Phys.* **2005**, *122*, 14103.

(50) Hesselmann, A.; Jansen, G. First-Order Intermolecular Interaction Energies from Kohn-Sham Orbitals. *Chem. Phys. Lett.* **2002**, *357*, 464–470.

(51) Hesselmann, A.; Jansen, G. Intermolecular Induction and Exchange-Induction Energies from Coupled-Perturbed Kohn-Sham Density Functional Theory. *Chem. Phys. Lett.* **2002**, *362*, 319–325.

(52) Hesselmann, A.; Jansen, G. Intermolecular Dispersion Energies from Time-Dependent Density Functional Theory. *Chem. Phys. Lett.* **2003**, *367*, 778–784.

(53) Toczylowski, R. R.; Cybulski, S. M. An Analysis of the Interactions between Nucleic Acid Bases: Hydrogen-Bonded Base Pairs. *J. Phys. Chem. A* **2003**, *107*, 418–426.

(54) Hesselmann, A.; Jansen, G.; Schütz, M. Interaction Energy Contributions of H-Bonded and Stacked Structures of the AT and GC DNA Base Pairs from the Combined Density Functional Theory and Intermolecular Perturbation Theory Approach. *J. Am. Chem. Soc.* **2006**, *128*, 11730–11731.

(55) Sedlák, R.; Jurečka, P.; Hobza, P. Density Functional Theory-Symmetry Adapted Perturbation Treatment Energy Decomposition of Nucleic Acid Base Pairs Taken from DNA Crystal Geometry. *J. Chem. Phys.* **2007**, *127*, 075104.

(56) Langner, K. M.; Sokalski, W. A.; Leszczynski, J. Intriguing Relations of Interaction Energy Components in Stacked Nucleic Acids. *J. Chem. Phys.* **2007**, *127*, 111102.

(57) Cybulski, H.; Sadlej, J. Symmetry-Adapted Perturbation-Theory Interaction-Energy Decomposition for Hydrogen-Bonded and Stacking Structures. *J. Chem. Theory Comput.* **2008**, *4*, 892–897.

(58) Guerra, C. F.; Bickelhaupt, F. M.; Snijders, J. G.; Baerends, E. J. The Nature of the Hydrogen Bond in DNA Base Pairs: The Role of Charge Transfer and Resonance Assistance. *Chem.-Eur. J.* **1999**, *5*, 3581–3594.

(59) Mo, Y. R. Probing the Nature of Hydrogen Bonds in DNA Base Pairs. *J. Mol. Model.* **2006**, *12*, 665–672.

(60) Grimme, S. Improved Second-Order Møller-Plesset Perturbation Theory by Separate Scaling of Parallel- and Antiparallel-Spin Pair Correlation Energies. *J. Chem. Phys.* **2003**, *118*, 9095–9102.

(61) Grimme, S. Accurate Description of van der Waals Complexes by Density Functional Theory Including Empirical Corrections. *J. Comput. Chem.* **2004**, *25*, 1463–1473.

(62) Jurečka, P.; Černý, J.; Hobza, P.; Salahub, D. R. Density Functional Theory Augmented with an Empirical Dispersion Term. Interaction Energies and Geometries of 80 Noncovalent Complexes Compared with Ab Initio Quantum Mechanics Calculations. *J. Comput. Chem.* **2007**, *28*, 555–569.

(63) Burda, J. V.; Šponer, J.; Leszczynski, J.; Hobza, P. Interaction of DNA Base Pairs with Various Metal Cations ($Mg^{2+}$, $Ca^{2+}$, $Sr^{2+}$, $Ba^{2+}$, $Cu^{+}$, $Ag^{+}$, $Au^{+}$, $Zn^{2+}$, $Cd^{2+}$, and $Hg^{2+}$): Nonempirical Ab Initio Calculations on Structures, Energies, and Nonadditivity of the Interaction. *J. Phys. Chem. B* **1997**, *101*, 9670–9677.

(64) Halkier, A.; Helgaker, T.; Jørgensen, P.; Klopper, W.; Olsen, J. Basis-Set Convergence of the Energy in Molecular Hartree-Fock Calculations. *Chem. Phys. Lett.* **1999**, *302*, 437–446.

(65) Halkier, A.; Helgaker, T.; Jørgensen, P.; Klopper, W.; Koch, H.; Olsen, J.; Wilson, A. K. Basis-Set Convergence in Correlated Calculations on Ne, N-2, and H2O. *Chem. Phys. Lett.* **1998**, *286*, 243–252.

(66) Jurečka, P.; Hobza, P. True Stabilization Energies for the Optimal Planar Hydrogen-Bonded and Stacked Structures of Guanine. .Cytosine, Adenine. .Thymine, and Their 9- and 1-Methyl Derivatives: Complete Basis Set Calculations at the MP2 and CCSD(T) Levels and Comparison with Experiment. *J. Am. Chem. Soc.* **2003**, *125*, 15608–15613.

(67) Ahlrichs, R.; Bär, M.; Häser, M.; Horn, H.; Kölmel, C. Electronic-Structure Calculations on Workstation Computers - the Program System Turbomole. *Chem. Phys. Lett.* **1989**, *162*, 165–169.

(68) Weigend, F.; Häser, M. RI-MP2: First Derivatives and Global Consistency. *Theor. Chem. Acc.* **1997**, *97*, 331–340.

(69) Werner, H.-J.; Knowles, P. J.; Lindh, R.; Manby, F. R.; Schütz, M.; Celani, P.; Korona, T.; Rauhut, G.; Amos, R. D.; Bernhardsson, A.; Berning, A.; Cooper, D. L.; Deegan, M. J. O.; Dobbyn, A. J.; Eckert, F.; Hampel, C.; Hetzer, G.; Lloyd, A. W.; McNicholas, S. J.; Meyer, W.; Mura, M. E.; Nicklass, A.; Palmieri, P.; Pitzer, R.; Schumann, U.; Stoll, H.; Stone, A. J.; Tarroni, R.; Thorsteinsson, T. *Molpro version 2006.1*; a Package of Ab Initio Programs, 2006. See http://www.molpro.net (accessed month year).

(70) Jeziorski, B.; Moszynski, R.; Szalewicz, K. Perturbation Theory Approach to Intermolecular Potential Energy Surfaces of van der Waals Complexes. *Chem. Rev.* **1994**, *94*, 1887–1930.

(71) Misquitta, A. J.; Podeszwa, R.; Jeziorski, B.; Szalewicz, K. Intermolecular Potentials Based on Symmetry-Adapted Perturbation Theory with Dispersion Energies from Time-Dependent Density-Functional Calculations. *J. Chem. Phys.* **2005**, *123*, 214103.

(72) Distasio, R. A.; Head-Gordon, M. Optimized Spin-Component Scaled Second-Order Møller-Plesset Perturbation Theory for Intermolecular Interaction Energies. *Mol. Phys.* **2007**, *105*, 1073–1083.

(73) Jurečka, P.; Šponer, J.; Černý, J.; Hobza, P. Benchmark Database of Accurate (MP2 and CCSD(T) Complete Basis Set Limit) Interaction Energies of Small Model Complexes, DNA Base Pairs, and Amino Acid Pairs. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1985–1993.

(74) Pitoňák, M.; Riley, K. E.; Neogrady, P.; Hobza, P. Highly Accurate CCSD(T) and DFT-SAPT Stabilization Energies of H-Bonded and Stacked Structures of the Uracil Dimer. *ChemPhysChem* **2008**, *9*, 1636–1644.

(75) Frisch, M. J. ; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03; revision C.02*; Gaussian Inc.: Wallingford, CT, 2004.

(76) Šponer, J.; Leszczynski, J.; Hobza, P. Structures and Energies of Hydrogen-Bonded DNA Base Pairs. A Nonempirical Study with Inclusion of Electron Correlation. *J. Phys. Chem.* **1996**, *100*, 1965–1974.

(77) Šponer, J.; Burda, J. V.; Mejzlík, P.; Leszczynski, J.; Hobza, P. Hydrogen-Bonded Trimers of DNA Bases and Their Interaction with Metal Cations: Ab Initio Quantum-Chemical and Empirical Potential Study. *J. Biomol. Struct. Dyn.* **1997**, *14*, 613–628.

(78) Arora, N.; Jayaram, B. Energetics of Base Pairs in B-DNA in Solution: An Appraisal of Potential Functions and Dielectric Treatments. *J. Phys. Chem. B* **1998**, *102*, 6139–6144.

(79) Freier, S. M.; Sugimoto, N.; Sinclair, A.; Alkema, D.; Neilson, T.; Kierzek, R.; Caruthers, M. H.; Turner, D. H. Stability of XGCGCp, GCGCYp, and XGCGCYp Helixes: an Empirical Estimate of the Energetics of Hydrogen Bonds in Nucleic Acids. *Biochemistry* **1986**, *25*, 3214–3219.

(80) Turner, D. H.; Sugimoto, N.; Kierzek, R.; Dreiker, S. D. Free Energy Increments for Hydrogen Bonds in Nucleic Acid Base Pairs. *J. Am. Chem. Soc.* **1987**, *109*, 3783–3785.

(81) Tanner, N. K.; Cech, T. R. Guanosine Binding Required for Cyclization of the Self-Splicing Intervening Sequence Ribonucleic Acid from Tetrahymena Thermophila. *Biochemistry* **1987**, *26*, 3330–3340.

(82) Nakano, N. I.; Igarashi, S. J. Molecular Interactions of Pyrimidines, Purines, and Some Other Heteroaromatic Compounds in Aqueous Media. *Biochemistry* **1970**, *9*, 577–583.

(83) Ts'o, P. O. P.; Melvin, I. S.; Olson, A. C. Interaction and Association of Bases and Nucleosides in Aqueous Solutions. *J. Am. Chem. Soc.* **1963**, *85*, 1289–1296.

(84) van Holde, K. E.; Rossetti, G. P. A Sedimentation Equilibrium Study of the Association of Purine in Aqueous Solutions. *Biochemistry* **1967**, *6*, 2189–2194.

(85) Solie, T. N.; Schellman, J. A. The Interaction of Nucleosides in Aqueous Solution. *J. Mol. Biol.* **1968**, *33*, 61–77.

(86) Stofer, E.; Chipot, C.; Lavery, R. Free Energy Calculations of Watson-Crick Base Pairing in Aqueous Solution. *J. Am. Chem. Soc.* **1999**, *121*, 9503–9508.

(87) Florián, J.; Šponer, J.; Warshel, A. Thermodynamic Parameters for Stacking and Hydrogen Bonding of Nucleic Acid Bases in Aqueous Solution: Ab Initio/Langevin Dipoles Study. *J. Phys. Chem. B* **1999**, *103*, 884–892.

CT800547K