Principles of RNA Base Pairing: Structures and Energies of the Trans Watson−Crick/Sugar Edge Base Pairs

Judit E. Šponer,*,† Nad'a Špačková,† Jerzy Leszczynski,‡ and Jiří Šponer*,†,§

Institute of Biophysics, Academy of Sciences of the Czech Republic, Královopolská 135, 612 65 Brno, Czech Republic, Department of Chemistry, Computational Center for Molecular Structure and Interactions, Jackson State University, Jackson, Mississippi 39217, and Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, Flemingovo náměstí 2, 166 10 Prague 6, Czech Republic

Received: March 4, 2005; In Final Form: April 18, 2005

Due to the presence of the 2′-OH hydroxyl group of ribose, RNA molecules utilize an astonishing variability of base pairing patterns to build up their structures and perform the biological functions. Many of the key RNA base pairing families have no counterparts in DNA. In this study, the trans Watson−Crick/sugar edge (trans WC/SE) RNA base pair family has been characterized using quantum chemical and molecular mechanics calculations. Gas-phase optimized geometries from density functional theory (DFT) calculations and RIMP2 interaction energies are reported for the 10 crystallographically identified trans WC/SE base pairing patterns. Further, stable structures are predicted for all of the remaining six possible members of this family not seen in RNAs so far. Among these novel six base pairs, the computations substantially refine two structures suggested earlier based on simple isosteric considerations. For two additional trans WC/SE base pairs predicted in this study, no arrangement was suggested before. Thus, our study brings a complete set of trans WC/SE base pairing patterns. The present results are also contrasted with calculations reported recently for the cis WC/SE base pair family. The computed base pair sizes are in sound correlation with the X-ray data for all WC/SE pairing patterns including both their cis and trans isomers. This confirms that the isostericity of RNA base pairs, which is one of the key factors determining the RNA sequence conservation patterns, originates in the properties of the isolated base pairs. In contrast to the cis structures, however, the isosteric subgroups of the trans WC/SE family differ not only in their H-bonding patterns and steric dimensions but also in the *intrinsic* strength of the intermolecular interactions. The distribution of the total interaction energy over the sugar−base and base−base contributions is controlled by the cis−trans isomerism.

Introduction

The basic principles of base pairing in DNA and RNA are strikingly different. While the DNA duplexes are dominated by standard Watson−Crick (WC) base pairs coding the genetic information, atomic-resolution structures of RNA revealed that the nominally unpaired loop regions of RNAs are in fact involved in highly versatile non-Watson−Crick base pairing patterns (Figure 1).[1−3] The non-WC interactions play unique roles in the evolution and folding of RNA architectures. They stabilize tertiary contacts between remote segments of the RNA macromolecules and form specific non-WC regions called RNA motifs.[1−3] Non-WC interactions often utilize the H-bonding capability of the ribose 2′-OH hydroxyl group. This architectural versatility sets the macromolecule to fulfill highly specific fundamental biocatalytic functions.

Although the large RNA molecules such as ribosomal subunits look at first sight extremely complex, it appears that RNA molecules actually use a limited set of universal and recurrent structural motifs, interactions, and building blocks for the folding and function. For example, the most important tertiary interaction in the RNA world is the A-minor motif, where a patch of two to three consecutive adenines interacts with the minor groove of Watson−Crick (preferably G=C) base pairs.[3d] The A-minor motif is by far the most abundant non-Watson−Crick tertiary interaction in the large ribosomal subunit and apparently also in other RNAs. In 23S and 5S rRNA, 186 adenines participate in A-minor interactions, 68 of which are conserved.[3d] This interaction is even utilized by the small ribosomal subunit to discriminate between cognate and near-cognate tRNA, since universally conserved adenines A1492 and A1493 of the 16S rRNA read the shape of the first two base pairs of the codon−anticodon helix.[4] The discrimination of correct versus incorrect codon−anticodon base pairing is the key step in decoding. Similarly, it seems that ribosomal RNAs utilize a rather limited set of common RNA motifs (building blocks consisting of consecutive non-Watson−Crick interactions), such as the kink turns.[3c] Thus, non-WC interactions are absolutely essential for RNA structure and functions.

The non-Watson−Crick base pairs were in detail examined by Leontis et al., who introduced a classification of the RNA base pairs on the basis of crystal database analysis.[1b] They ranked the up-to-date known approximately 110 different RNA base pairing patterns into 12 families and elaborated a matrix formalism to signify those being in an isosteric relationship. They also formulated the isostericity principle. The isostericity principle assumes that a given RNA base pair can be replaced by another one with a similar (isosteric) shape without any loss

* To whom correspondence should be addressed. E-mail: judit@ncbr.chemi.muni.cz (J.E.Š.); sponer@ncbr.chemi.muni.cz (J.Š.).

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

‡ Jackson State University.

§ Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic.

10.1021/jp051126r CCC: $30.25
Published on Web 05/17/2005

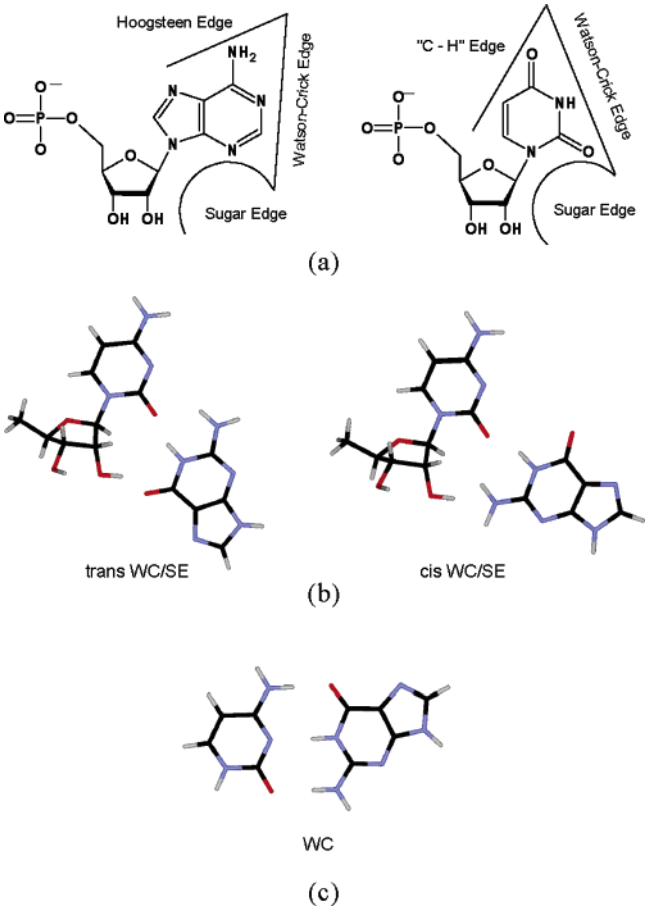

Figure 1. Nomenclature of the RNA base pairs:[1b] (a) classification of the interaction sites in purine and pyrimidine nucleobases; (b) cis and trans isomers of the G/rC Watson−Crick/sugar edge (WC/SE) base pair (idealized unoptimized geometry; G in WC, rC in SE position); (c) standard GC Watson−Crick/Watson−Crick (WC/WC) base pair.

of RNA folding and function. Although the applicability of the isostericity principle is obviously not absolute, it appears to be one of the leading rules governing the RNA base pairing patterns and sequence conservation. A textbook example of the isostericity principle is the 5S rRNA loop E motif.[5] The bacterial loop E motif is a unique helix with seven consecutive non-Watson−Crick base pairs, likely the longest such helix in the ribosome. Chloroplasts are related evolutionarily to bacteria. Chemical and enzymatic probing, isostericity modeling, MD simulations, and finally solution NMR structural analysis have demonstrated that spinach chloroplast 5S rRNA also contains loop E.[5] In addition, spinach chloroplast loop E is isosteric with the *E. coli* one, despite the fact that 5 out of the 14 loop E bases (36%) are mutated, including 4 G-to-A substitutions. All the individual loop E substitutions offer the formation of isosteric base pairs, and the isostericity is used in shaping up the spinach chloroplast loop E.

Various aspects of nucleic acid (NA) base pairing have been studied by quantum chemical methods. However, RNA base pairs other than the standard ones have been largely neglected in the quantum chemical literature, despite their evident biological importance. Let us reiterate that for the biological functions of RNA the non-WC base pairing is in fact more important than the standard base pairs (see Figure 1). To date, only very few quantum chemical studies have been published on RNA base pairing, dealing for example with water-inserted base pairs and G/A cis Watson−Crick interaction.[6] In a recent study, we carried out the first systematic quantum chemical (QM) analysis of RNA base pairs involving the ribose hydroxyl group, dealing with the complete cis Watson−Crick/sugar edge (WC/SE) RNA base pair family, one of the key classes of RNA interactions.[7] We used the density functional theory (DFT) and RIMP2 (resolution of identity variant of the second-order Moeller−Plesset perturbational method) techniques to scrutinize geometrical and energetic aspects of the nucleobase−nucleoside interactions in fully isolated gas-phase structures of these base pairs, to complement the experimental structural data by energy analysis. The calculations revealed that the interaction between the sugar and the nucleobase in the WC position significantly contributed to the stability of the cis WC/SE binding pattern. With two exceptions, the strength of the base pairs was very similar, ranging from −16 to −22 kcal/mol and being thus between the GC WC and AU WC standard base pairs. Stable structures were obtained even for several base pairs not yet observed in RNA molecules. A sound correlation was found between the crystallographically determined and computed geometries. This along with the interaction energy data pointed at the intrinsic stability of the cis WC/SE base pairs, which in significant part can be attributed to the sugar−base contribution. Thus, the observed geometries of the cis WC/SE RNA family are primarily determined by their intrinsic gas-phase properties.

In the present study, we focus on another key RNA base pairing pattern, the trans WC/SE interactions. Similar to the cis WC/SE family, the first base interacts via its Watson−Crick face with the sugar edge of the other nucleotide with a direct participation of its O2′ hydroxyl group. However, the two bases forming the pair are in the trans position, leading to entirely different mutual arrangements of the attached sugar−phosphate backbones. We wish to emphasize that an important advantage of the QM approach is its capability to complement the structural data by reliable information about the strength of the base pairing. In fact, prior to our study, no energy data were available on this kind of base pairing, since neither the X-ray structures nor the crude isostericity modeling provide any information about the base pairing energies, except of the number of H-bonds.

The trans WC/SE family may consist of up to 16 distinct members. Having only 10 crystallographically characterized interaction patterns so far (see Figure 7 in ref 1b), this base pair family is less common in comparison to its cis counterpart. Comparing the crystallographically reported cis and trans WC/SE binding patterns (Figures 6 and 7 of ref 1b), one can recognize a reduction of the sugar−base contact area in the trans structures. This sets one of the main goals of the current study, to evaluate how alteration of the mutual base−nucleoside position modifies the strength and structural parameters of the base pairs. In addition, our aim is to refine the structures of four trans WC/SE base pairs predicted on the basis of a rough isosteric modeling and to determine the structures of the remaining two members of this family that were entirely missing in the preceding isostericity analysis.

Computational

Geometry Optimizations. When available, initial structures were built up on the basis of crystal geometries, using the structures listed in Figure 7 of ref 1b. The majority of them were thus taken from the large ribosomal subunit of *M. marismortui* (NDB code RR0033).[3e] Additional crystal data were obtained from PR0022,[3f] RR0015,[3b] and UR0004.[3g] Sugar rings not involved in the sugar−edge interactions were deleted. Thus, the studied complexes consisted of one base and one nucleoside, both terminated by hydrogen atoms. For those structures not

yet seen in crystallographic studies, we have constructed the initial model geometries according to the H-bonding pattern shown in Figure 7 of ref 1b. No suggested structures are available for G/rA and G/rG. The choice of the initial geometries for these two systems is described in the relevant part of the Results and Discussion. In all cases, the QM calculations resulted in a substantial refinement of the H-bonding patterns.

Geometry optimizations were carried out at the density functional level of theory using the Gaussian98 program package.[8] The density functional was built up by Becke's three-parameter exchange[9] and Lee−Yang−Parr's correlation functional (abbreviated as B3LYP).[10] The 6-31G** basis set was used for geometry optimizations. We have shown recently that the B3LYP/6-31G**-optimized structures compare very well with reference RIMP2/cc-pVTZ data.[11]

Interaction Energies. Interaction energies were computed on the B3LYP-optimized structures using the RIMP2 approach combined with a large diffuse aug-cc-pVDZ basis set of atomic orbitals internally stored in the Turbomole code.[12] The same approach has been used in our former study,[7] where we have shown that the aug-cc-pVDZ basis set gives almost as good energies as the aug-cc-pVTZ basis set. The RIMP2 method for calculating interaction energies has been validated in refs 11 and 13. The level of calculations used in this study is thus substantially better than that in the majority of QM studies published for NA base pairs so far and is entirely sufficient for our purpose, as we do not intend to perform the highest-accuracy reference calculations.

The total interaction energy of a nucleobase−nucleoside dimer (BN) (ΔE^{BN}) is defined as

$$\Delta E^{BN} = E^{BN} - E^{B} - E^{N} \quad (1)$$

where E^{BN} stands for the electronic energy of the whole system and E^{B} and E^{N} are the electronic energies of the isolated subsystems B (nucleobase) and N (nucleoside).

The interaction energy (ΔE) has two components: the Hartree−Fock (HF) term (ΔE^{HF}) and the electron correlation term (ΔE^{cor}).

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

The ΔE^{HF} term includes mainly the electrostatic interaction energy, short-range exchange repulsion, and polarization/charge transfer contributions. The ΔE^{cor} term is dominated by the dispersion attraction and further includes the electron correlation correction to the electrostatic energy, which is mostly repulsive, since the electron correlation reduces the dipole moments of the monomers.

All interaction energies are corrected for the basis set superposition error using the standard counterpoise procedure[14] but do not include deformation energies. The main reason for this is the substantial structural alteration of the sugar−base segment upon base pairing in many structures. Thus, a direct inclusion of monomer deformations into the interaction energies would rather bias the results. For a detailed discussion regarding the role of the deformation energies in base pairing calculations, see ref 11. Deformation energies can be readily obtained considering the optimized geometries available in the Supporting Information.

Separation of the Base−Base and Base−Sugar Contributions. We approximately dissected the base−base and base−sugar interaction energy terms. These energies were derived for the following dimers: (i) the first nucleobase (B) interacting via its WC edge with the second base (B′) after sugar deletion (BB′) and (ii) the first nucleobase interacting with the sugar (BS). The nucleoside (N) was split to sugar (S) and base (B′) along the glycosidic bond. The dangling bonds were saturated with a hydrogen atom. The C1′−H distance of the ribose was assumed to be 1.1 Å, while the N1−H and N9−H distances were fixed at 1.0 Å. No other modifications were made.

The base−base interaction energy (ΔE^{bb}) was computed for subsystems B and B′ according to the following formula:

$$\Delta E^{bb} = E^{BB'} - E^{B} - E^{B'} \quad (3)$$

and the base−sugar contribution (ΔE^{bs}) was computed as follows:

$$\Delta E^{bs} = E^{BS} - E^{B} - E^{S} \quad (4)$$

Molecular Mechanics. Molecular mechanics calculations have been performed using the AMBER7[15] program package

TABLE 1: Main Geometrical Parameters of the Interbase and Ribose−Base Contacts for the Fully Optimized trans WC/SE Base Pairs[a]

isosteric subfamily	base pair	donor atom (X)	acceptor atom (Y)	distance X−Y	distance X−H	Δ(X−H)[b]
I_1	A/rA	N6(A)	O2′	3.04	1.009	0.002
		N6(A)	N3(A′)[d]	3.01	1.020	0.013
	A/rC[c]	N6(A)	O2(C)	2.85	1.022	0.015
		O2′	N7(A)	2.76	0.989	*e*
	A/rG	N6(A)	O2′	2.92	1.012	0.005
		N6(A)	N3(G)	3.04	1.021	0.014
		N2(G)	N1(A)	2.97	1.032	0.022
	A/rU[c]	N6(A)	O2(U)	3.00	1.014	0.007
		O2′	N7(A)	2.73	0.989	*e*
	C/rA	N4(C)	O2′	3.05	1.008	0.002
		N4(C)	N3(A)	2.99	1.023	0.014
	C/rC	N4(C)	O2′	3.09	1.011	0.005
		N4(C)	O2(C′)[d]	2.88	1.010	0.001
	C/rG	N4(C)	O2′	2.93	1.011	0.005
		N4(C)	N3(G)	3.03	1.025	0.016
		N2(G)	N3(C)	2.89	1.035	0.025
	C/rU[c]	N4(C)	O2′	3.04	1.011	0.005
		N4(C)	O2(U)	2.96	1.010	0.001
I_2	G/rC	O2′	O6(G)	2.65	0.992	*e*
		N1(G)	O2(C)	2.84	1.021	0.008
		N2(G)	O2(C)	2.89	1.018	0.008
	G/rU	O2′	O6(G)	2.65	0.992	*e*
		N1(G)	O2(U)	2.92	1.019	0.006
		N2(G)	O2(U)	2.96	1.013	0.003
I_3	U/rA	O2′	N3(A)	2.94	0.977	*e*
		N3(U)	N3(A)	3.13	1.025	0.012
	U/rC	O2′	O4(U)	2.74	0.981	*e*
		N3(U)	O2(C)	2.77	1.036	0.023
	U/rU[c]	O2′	O4(U)	2.82	0.979	*e*
		N3(U)	O2(U′)[d]	2.73	1.034	0.021
		O2(U)	N3(U′)[d]	3.18	1.021	0.008
I_4	U/rG	O2′	O(W)[f]	2.80	0.978	*e*
		O(W)[f]	N3(G)	2.72	1.006	0.040
		N3(U)	O(W)[f]	2.83	1.032	0.019
		N2(G)	O2(U)	2.87	1.019	0.009
predicted as I_2 by us[g]	G/rA[c]	O2′	O6(G)	2.71	0.984	*e*
		N1(G)	N3(A)	3.0	1.034	0.021
	G/rG[c]	O2′	O6(G)	2.73	0.982	*e*
		N1(G)	N3(G′)[d]	3.06	1.026	0.013

[a] Interatomic distances are given in angstroms. The C3′-endo conformation is considered for all structures (see the text). Cartesian coordinates are listed in the Supporting Information for all structures. [b] Elongation of the X−H covalent bond with respect to the isolated optimized monomers. [c] No crystallographic data are available for this structure. [d] The prime sign denotes the base connected to the nucleoside. [e] Reference value is not relevant due to internal H-bond formation inside the fully optimized nucleoside structure. [f] W = water. [g] Not assigned by Leontis et al.[1b]

TABLE 2: Interaction Energies (kcal/mol) for the Optimized trans WC/SE Base Pairs Obtained at the RI-MP2/aug-cc-pVDZ Level Using the B3LYP/6-31G-Optimized Geometries and with the Cornell et al. (AMBER) Force Field Using the Same DFT-Optimized Geometries (For Further Details, See Methods)**

		interaction energy			pairwise terms	
trans WC/SE base pair	isosteric subfamily	RIMP2	HF	AMBER	B···S[a]	B···B′[b]
A/rA	I_1	**−9.9 (−13.8)**[c]	−4.4 (−5.5)	**−10.2 (−13.0)**[c]	−3.8 (−6.0)	−6.5 (−6.3)
A/rC[d]		**−17.2**	−10.5	**−15.1**	−9.1	−6.7
A/rG		**−16.5 (−16.1)**	−8.8 (−7.6)	**−17.1 (−16.4)**	−4.0 (−4.2)	−12.5 (−11.7)
A/rU[d]		**−16.2**	−9.6	**−14.5**	−10.1	−3.4
C/rA		**−10.8 (−6.6)**	−5.8 (−1.8)	**−11.1 (−7.3)**	−4.6 (−4.7)	−6.9 (−1.6)
C/rC		**−9.7 (−14.4)**	−6.7 (−8.3)	**−11.1 (−15.2)**	−5.6 (−5.1)	−5.1 (−8.4)
C/rG		**−20.9 (−17.0)**	−14.0 (−8.9)	**−21.7 (−19.0)**	−4.9 (−3.4)	−15.9 (−13.8)
C/rU[d]		**−8.1**	−4.5	**−9.6**	−5.4	−3.5
G/rC	I_2	**−28.4**	−24.2	**−28.0/−29.2**[e]	−9.1	−15.9
G/rU		**−23.1**	−18.8	**−22.4/−23.8**[e]	−9.2	−10.6
U/rA	I_3	**−9.9**	−3.8	**−11.3**	−0.9	−8.9
U/rC		**−15.9**	−11.8	**−15.5**	−6.7	−7.1
U/rU[d]		**−17.7**	−12.0	**−17.4**	−7.1	−9.2
U/rG	I_4	**−26.4**	−17.5	**−25.4**	*f*	*f*
G/rA[d]	predicted as I_2 by us[g]	**−20.8**	−13.4	**−20.6**		−7.3
G/rG[d]		**−19.9**	−10.5	**−19.3**		−7.9

[a] B = nucleobase, interacting via its WC edge, S = sugar. [b] B′ = nucleobase interacting via its S edge. [c] The data in parentheses refer to the optimized structures with amino-acceptor interactions. [d] No crystallographic data are available for this structure. [e] The second value was calculated for optimized geometries obtained with AMBER. [f] Due to the water bridge, evaluation of the sugar−base and base−base interaction energy terms is not applicable for this structure. [g] Not assigned by Leontis et al.[1b]

in combination with the Cornell et al. force field parameters (parm99).[16,17] AMBER atomic charges have been derived using the Hartree−Fock approximation with the 6-31G* basis set of atomic orbitals via the RESP fitting procedure.[18] For our calculations, the original AMBER atomic charges were slightly modified to keep the monomers neutral. For the nucleobase, the sugar−phosphate segment was replaced by hydrogen and its charge was set to keep the base neutral. For the nucleoside, the neutralization was carried out by smearing the excessive charge over several sugar atoms, resulting in negligible charge modifications of 0.01−0.02 e. We have carried out test calculations with additional slightly modified charge redistributions, which show that the results are not sensitive to such minor changes of the charge distributions. The charges used in our computations are listed in the Supporting Information. Geometry optimizations were carried out starting from the gas-phase optimized structures. Interaction energies have been obtained from single point calculations using either QM or AMBER-optimized geometries according to eq 1 and do not include correction for the deformation energies. The AMBER deformation energies are anyway rather negligible compared to the QM ones (at least considering deformations directly caused by H-bonding).[11]

Results and Discussion

The trans WC/SE base pairs are more compact than their cis counterparts. In our previous study,[7] we have shown that the C1′(nucleotide)−N1(pyrimidine) or C1′−N9(purine) *interbase* distances (later in the text referred to as C1′−N) can be used as a measure of isostericity equally as well as the C1′−C1′ interbase distances commonly used to characterize the base pair diameter.[1b] We use the C1′−N distances simply because one of the sugars was deleted in the computations. The crystallographically determined range of the C1′−N distances of the cis WC/SE base pairs (5.2−9.2 Å) fully includes that of the trans structures (6.5−8.6 Å), indicating that the separation between the individual isosteric subgroups might be smaller in the trans WC/SE family. The original definition of isostericity also includes the directionality of the glycosidic bonds. Viewing the crystal geometries referred to in ref 1b, one recognizes that this criterion is much looser than the one meant by the C1′−C1′ interbase distances. We have calculated the angle of the C1′N(9,1) vectors, representing the direction of the glycosidic bond inside a nucleotide unit, for all available crystal geometries of the trans WC/SE base pair family. The lowest values have been obtained for the I_1 isosteric subfamily (from 52 to 94°). The interval calculated for the I_3 subfamily (100−121°) completely overlaps that of I_2 (102−109°), whereas the one-member I_4 subfamily exhibits the highest value 137°. For a detailed list of these values along with those obtained for the DFT-optimized geometries, refer to Table S1 in the Supporting Information.

TABLE 3: Computed and Crystallographically Determined Values of the N1−C1′ and N9−C1′ Distances (Å) for the trans WC/SE Base Pair Family

			distance	
isosteric subfamily	base pair	measure	crystal[a]	gas-phase QM
I_1	A/rA	N9−C1′	8.66	8.50
	A/rC[b]	N9−C1′	∼8.06[c]	6.63
	A/rG	N9−C1′	7.81	8.16
	A/rU[b]	N9−C1′	∼8.06[c]	6.55
	C/rA	N1−C1′	7.88	7.96
	C/rC	N1−C1′	8.06	8.17
	C/rG	N1−C1′	7.83	7.62
	C/rU[b]	N1−C1′	∼8.06[c]	8.15
I_2	G/rC	N9−C1′	8.05	8.43
	G/rU	N9−C1′	8.29	8.39
I_3	U/rA	N1−C1′	6.52	6.70
	U/rC	N1−C1′	7.23	7.37
	U/rU[b]	N1−C1′	7.23[c]	6.94
I_4	U/rG	N1−C1′	7.74	8.00
predicted as I_2 by us[d]	G/rA[b]	N9−C1′		7.81
	G/rG[b]	N9−C1′		8.05

[a] Crystal data, see ref 1b for further details. [b] No crystallographic data are available for this structure. [c] Estimated value from ref 1b. [d] Not assigned by Leontis et al.[1b]

The H-bonding patterns of the trans WC/SE base pairs lead to several distinct isosteric subfamilies. In the following, we will discuss each isosteric subfamily separately in order to illuminate the geometric and energetical consequences of these differences. The main geometrical parameters of the characteristic interatomic contacts are summarized in Table 1. Table 2

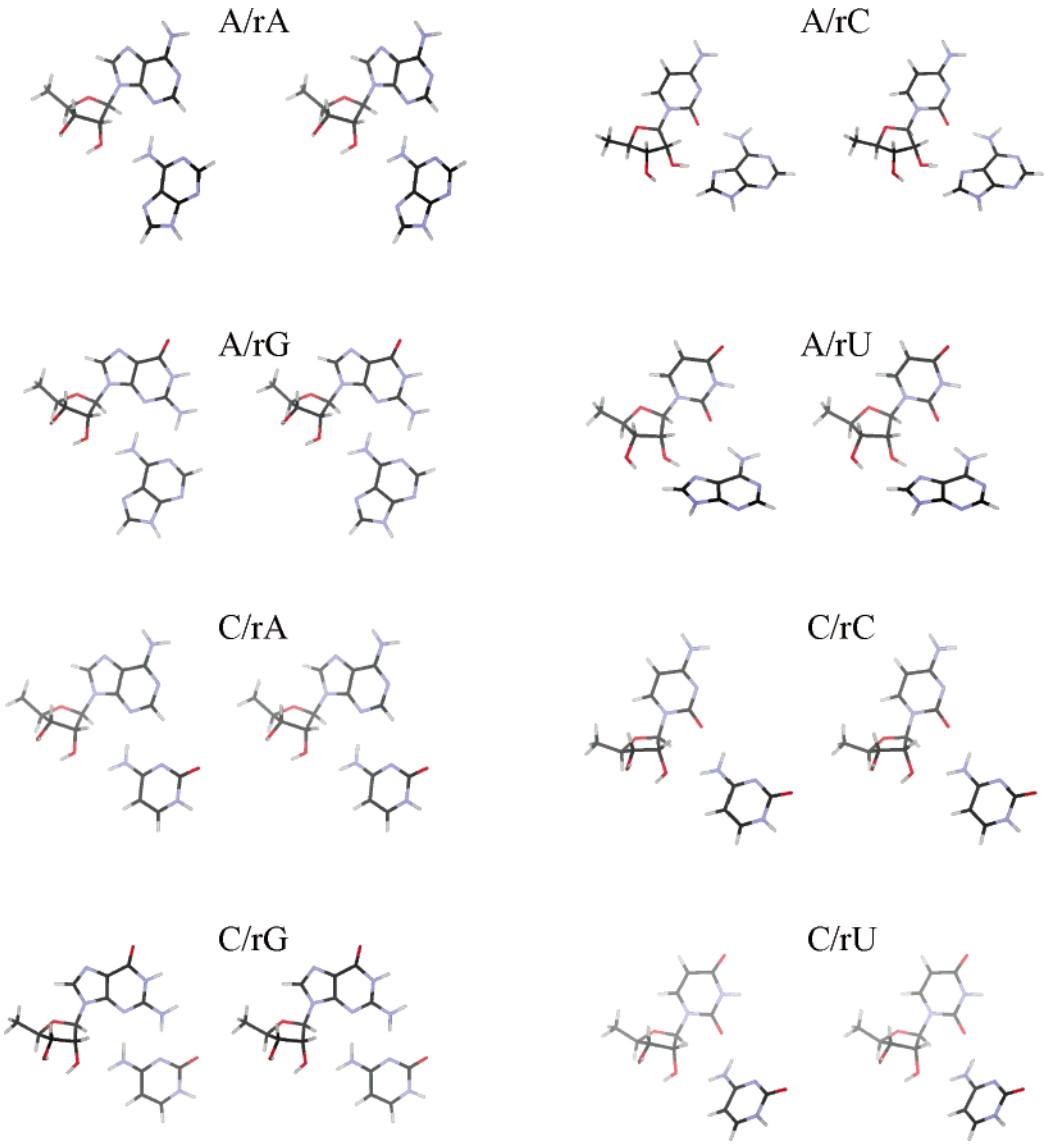

Figure 2. Gas-phase optimized geometries for the I_1 isosteric subfamily of the trans WC/SE binding type computed at the Becke3LYP level of theory (stereoview).

overviews the interaction energy data obtained from RIMP2 calculations and compares them with molecular mechanics calculations using the Cornell et al. force field.[15] Similar to the cis WC/SE base pairs, the AMBER results reproduce the RIMP2 values with an accuracy of ~2 kcal/mol.[7] This means that the AMBER force field appears to be well suited for the simulations of RNA structures containing this class of base pairs.

Prior to discussing the individual trans WC/SE base pair subfamilies, let us call attention to a common geometrical feature of all binding patterns involving a direct ribose−nucleobase contact. Due to the non-coplanarity of the nucleobase and ribose rings in the nucleoside structures, the base pairs are typically somewhat nonplanar. Interestingly, the QM-optimized structures of isolated pairs appear to be mostly more nonplanar compared to the X-ray examples. However, despite some differences, the X-ray and computed structures show the same H-bonding patterns (vide infra) and thus the QM structures are considered as being entirely representative.

I_1 Isosteric Subfamily. The eight trans WC/SE base pairs with A and C in the WC position form the largest isosteric subgroup, with a characteristic C1′−N distance of 7.8−8.7 Å (see Table 3). Five of its members were observed in available RNA X-ray structures. The optimized geometries of the group members are depicted in Figure 2, and for more detailed comparison, the coordinates of all geometries are given in the Supporting Information.

Crystallographic studies assign a central role to the amino group of the WC base in this subfamily. The X-ray data suggest that the amino group donates one of its hydrogens to form a H-bond with the 2′-OH of the sugar, while its second hydrogen is involved in a H-bond with one of the acceptor positions of the SE base. In the majority of the structures, no other contacts contribute to the stabilization of these base pairs.

The picture drawn by the ab initio calculations is somewhat more complicated. Depending on the input orientation of the 2′-OH group, we have found two stable optimized geometries for each base pair, sharply different in their H-bonding patterns. If the 2′-OH group points toward the exocyclic amino group of the WC base with its oxygen, a conventional N−H···O2′ type H-bond is formed (see Figure 2). If the 2′-OH group donates its hydrogen to N6(A) or N4(C) of the WC nucleobase, however, an amino-acceptor O2′−H···N bond is formed (see Figure 3 and Table 4).[19]

When comparing the calculated structures with the experimental ones, the agreement is better for those with conventional H-bonding than for those showing the amino-acceptor interac-

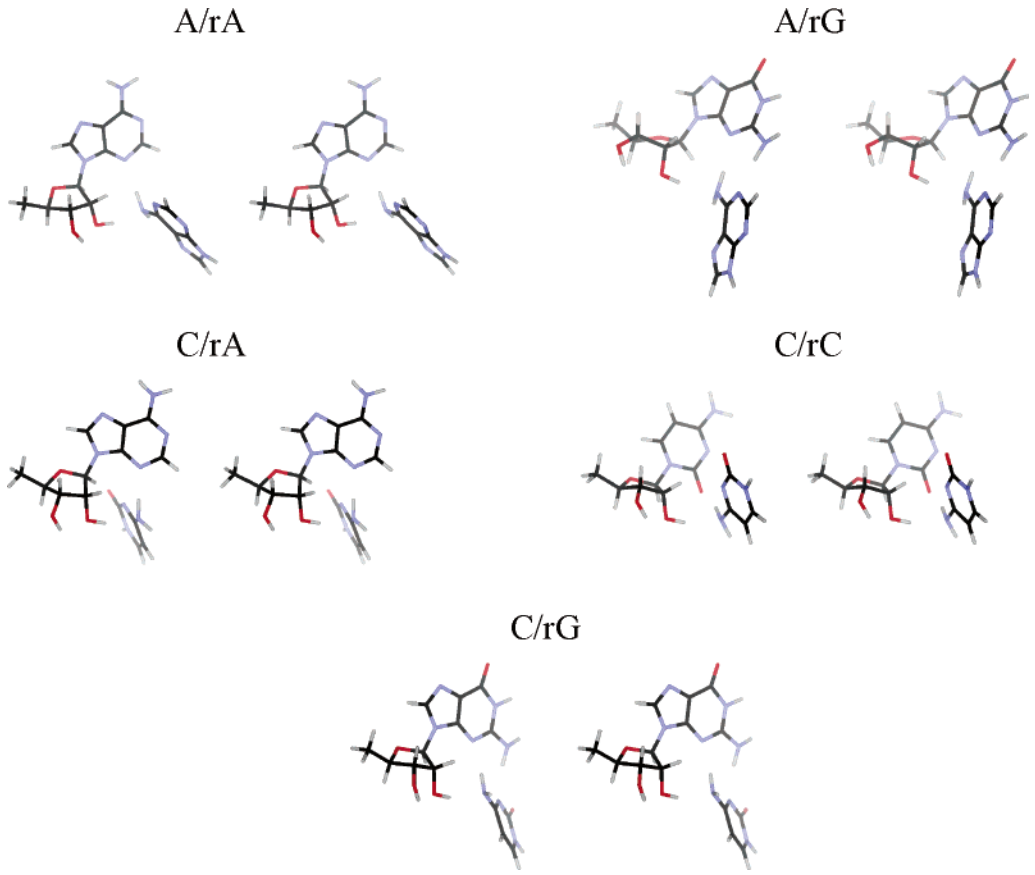

Figure 3. Possible amino-acceptor interactions in the I_1 subfamily (stereoview). Geometries were computed at the Becke3LYP level of theory.

TABLE 4: Main Parameters of the Alternative Amino-Acceptor Optimized Geometries Obtained for the I_1 Isosteric Subfamily[a]

base pair	donor atom (X)	acceptor atom (Y)	distance		
			X−Y	X−H	Δ(X−H)[b]
A/rA	O2′	N6(A)	2.90	0.980	*d*
	N6(A)	N3(A′)[c]	3.01	1.031	0.024
A/rG	O2′	N6(A)	2.95	0.976	*d*
	N6(A)	N3(G)	3.04	1.029	0.022
	N2(G)	N1(A)	3.01	1.026	0.016
C/rA	O2′	N4(C)	2.94	0.978	*d*
	N4(C)	N3(A)	2.99	1.034	0.025
C/rC	O2′	N4(C)	2.89	0.983	*d*
	N4(C)	O2(C′)[c]	2.86	1.021	0.012
C/rG	O2′	N4(C)	2.94	0.975	*d*
	N4(C)	N3(G)	3.03	1.031	0.022
	N2(G)	N3(G)	2.97	1.026	0.016

[a] Interatomic distances are given in angstroms. Cartesian coordinates are listed in the Supporting Information for all structures. [b] Elongation of the X−H covalent bond with respect to the isolated optimized monomers. [c] The prime sign denotes the base connected to the nucleoside. [d] Reference value is not relevant due to internal H-bond formation inside the fully optimized nucleoside structure.

tion. This is illustrated in Figures 4 and 5, which show an overlay of the computed (red) and X-ray (blue) geometries. The calculations (for conventional H-bonding) fairly reproduce the X-ray C1′−N distances (see Table 3). For A/rA and C/rC base pairs, the amino-acceptor structures provide better interaction energies than the conventional H-bonding. Thus, all the data suggest that formation of the amino-acceptor interactions cannot be entirely ruled out in conjunction with appropriate additional (external) interactions and represents a viable option for pairing.

The total interaction energies of the five crystallographically known members of the subfamily range from −9.7 to −20.9 kcal/mol (see Table 2). In the weaker structures (i.e., C/rC, A/rA, and C/rA with interaction energies of −9.7, −9.9, and −10.8

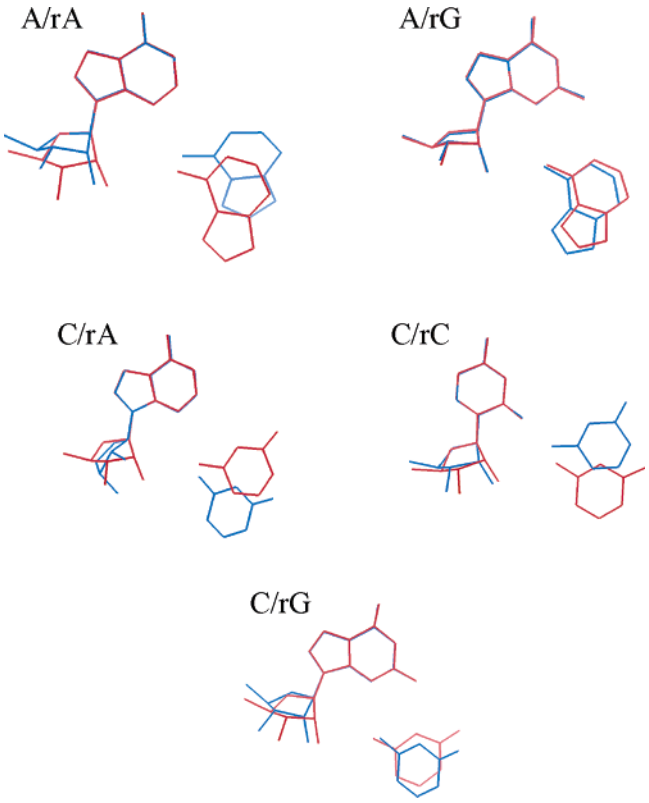

Figure 4. Overlay of the computed and X-ray geometries for the I_1 isosteric subfamily. Computed geometries were obtained assuming conventional N−H· · ·O2′ H-bonding.

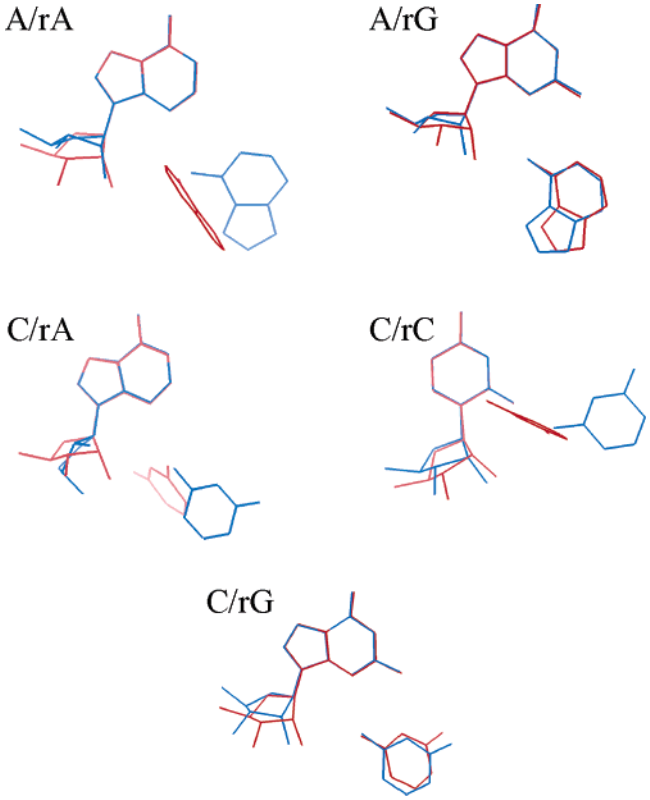

Figure 5. Overlay of the computed and X-ray geometries for the I_1 isosteric subfamily. Computed geometries were obtained assuming amino-acceptor bonds between 2′-OH and the WC base.

kcal/mol, respectively), beyond the featuring N−H···O2′ H-bond, there is only a weak interbase contact with participation of the second amino hydrogen of the WC base. For a summary of the interbase contacts, see Table 1. In the more stable structures (i.e., A/rG and C/rG with interaction energies of −16.5 and −20.9 kcal/mol, respectively), however, additional interbase H-bonds enhance the stability of the pair (see Table 1). This is reflected by the outstandingly large value of the interbase interaction energy term (−12.5 and −15.9 kcal/mol for A/rG and C/rG, respectively). Let us here call attention to the sound agreement of the RIMP2 and AMBER total interaction energies (see columns 1 and 3 of Table 2). Our data demonstrate that the AMBER force field performs well for classical N−H···O2′ as well as O2′−H···N amino-acceptor interactions. Note, however, that the AMBER energies were obtained with QM-optimized structures which, in contrast to the force field, allow amino-group pyramidalization in amino-acceptor interactions.

Out of the four trans WC/SE structures predicted by the isostericity modeling of Leontis et al., three structures (A/rC, A/rU, and C/rU) were assigned to the I_1 isosteric subfamily. Leontis et al. suggested a similar base pairing pattern for these base pairs as seen in the X-ray structure of C/rC. Our computations, however, were not able to identify such structures for A/rC and A/rU, because the N6(A)−H···O2′ contact was replaced with an energetically more favorable O2′−H···N7-(A) H-bond. This yielded ∼6 kcal/mol extra stability to these base pairs (see Figure 2 and Table 2). Interestingly, the loss of the H-bond between N6(A) and the ribose is accompanied with a noticeable non-coplanarity of the bases. Note that rather often even base pairs not involving the sugar unit are intrinsically substantially nonplanar, while the nonplanarity usually brings only a modest energy improvement.[20]

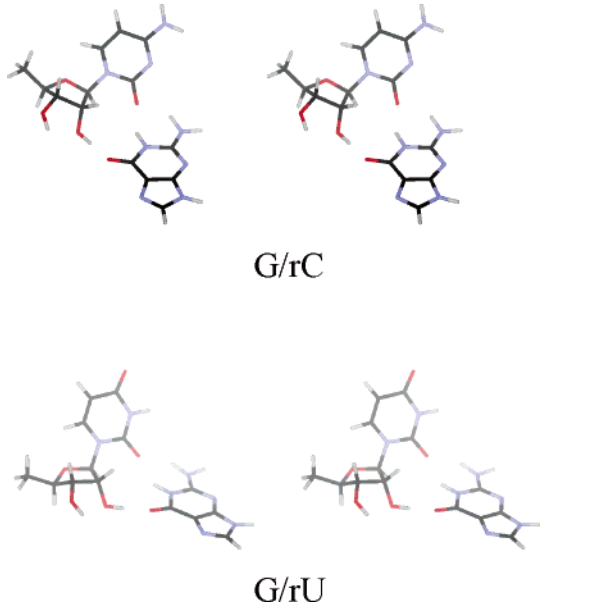

Figure 6. Gas-phase optimized geometries for the I_2 isosteric subfamily of the trans WC/SE binding type computed at the Becke3LYP level of theory (stereoview).

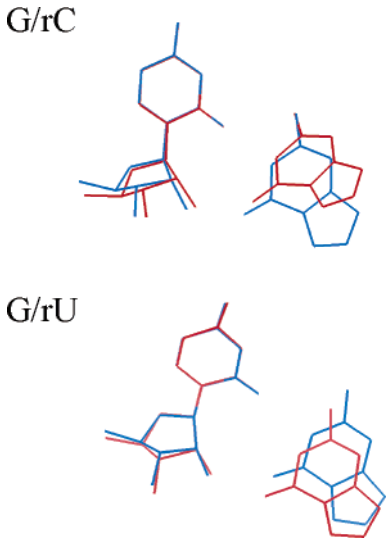

Figure 7. Overlay of the computed and X-ray geometries for the I_2 isosteric subfamily.

The QM-predicted C1′−N distances in the two base pairs (6.63 and 6.55 Å for A/rC and A/rU, respectively) are well below the value originally predicted (8.06 Å) by isostericity analogy, due to the non-coplanarity of the bases. C/rU, on the other hand, strongly resembles C/rC not only in its geometric parameters but also in the energetics of the base pairing (Table 2). The computed C1′−N distance in C/rU (8.15 Å) perfectly fits the proposed one (8.06 Å).

I_2 Isosteric Subfamily. The I_2 isosteric subfamily comprises two members, the G/rC and G/rU pairs, and represents the strongest trans WC/SE interaction subtype (for the DFT-optimized geometries and a comparison of the computed and X-ray geometries, see Figures 6 and 7, respectively). The base pairing (−28.4 and −23.1 kcal/mol for G/rC and G/rU, respectively) is stronger than that in any other cis and trans WC/SE structures. It seems that guanine is the best-suited base to interact with other bases in the trans WC/SE fashion. In addition to the obvious O2′−H···O2(G) sugar−base H-bond guanine is able to donate two H-bonds toward the SE nucleobase

via its N1 and N2 nitrogens. While there is no difference in the strength of the sugar−base contacts, in the G/rU structure, the base−base contribution is slightly reduced as compared to the G/rC contact (see the corresponding data in the last two columns of Table 2). The likely reason for this might be the lower basicity of O2 in uracil with respect to cytosine.[21a,b] Equivalently, this result may be interpreted as being caused by the lower dipole moment (polarity of charge distribution) of uracil[21c] or by different secondary interactions.[21d] All three explanations are actually closely interrelated. This is in agreement with elongations of the interacting intramolecular N−H bonds of guanine that are noticeably higher in G/rC than in G/rU (see the last column of Table 1). On the other hand, similar to the cis WC/SE interactions, the computed C1′−N distances are slightly longer than their crystallographic counterparts (Table 3).

Due to the outstanding gas-phase stability of these two binding patterns, we attempted to reoptimize them with the AMBER program. The optimized geometries and, consequently, the interaction energies (listed in Table 2) were practically identical to those obtained from quantum chemical calculations, providing a justification of force field calculations. Note that in the case of weaker WC/SE base pairs, standard AMBER structure relaxation often leads to different pairing patterns due to a lack of local minima; this is discussed in detail in ref 7.

I_3 Isosteric Subfamily. This base pair family includes three members, all having uracil in the WC position (U/rA, U/rC, and U/rU). The uracil N3 and O4 positions interact with the 2′-OH of the ribose and the SE nucleobase (see Figure 8).

The total interaction strength ranges from −9.9 to −17.7 kcal/mol. While there is a good balance of the sugar−base and base−base interactions in the U/rU and U/rC structures, in the U/rA base pair, the sugar−base term completely diminishes due to internal H-bond formation between rA(2′-OH) and rA(N3). This is, however, not surprising viewing the sugar conformation in the crystal structure. Similar to the C/rA binding pattern, the sugar adopts a C2′-endo conformation and thereby its hydroxyl becomes predisposed to form an internal H-bond with the adjacent adenine, rather than with the WC base. Let us note here that no C2′-endo conformations have been observed in the cis WC/SE base pair family.[7] As will be pointed out below, both trans WC/SE U/rA and C/rA base pairs, being C2′-endo in the database analysis by Leontis et al.,[1b] can be smoothly reoptimized with C3′-endo pucker.

There is a relatively good overall agreement between the computed and X-ray structures of U/rA and U/rC (see Figure 9). Accordingly, the computed C1′−N distances for these two base pairs follow the generally observed trend; that is, they are slightly longer than the crystallographically determined ones. In contrast, for the predicted U/rU structure, computations underestimate the X-ray value by about 0.2−0.3 Å. This contrasts with the general trend established for the cis WC/SE base pair family. Note, however, that these differences are negligible considering the medium resolution of the X-ray structures.

I_4 Isosteric Subfamily. U/rG is the only water-mediated trans WC/SE pair, which itself forms a separate isosteric subfamily. Geometry optimizations of the isolated model system resulted in two isomeric H-bonding patterns. In the structure depicted in Figure 10a, the water molecule donates only one of its hydrogen atoms to the H-bonding, creating a slightly looser contact with a total interaction energy of −26.4 kcal/mol and a C1′−N distance of 8.00 Å. Note that due to water insertion the interaction energy is calculated for the base−nucleoside−water trimer and thus cannot be directly compared with those base pairs that are not water-bridged. In the more stable structure (see Figure 10b), both hydrogen atoms of the water molecule participate in the H-bonding, in a similar manner as that observed in the water-mediated G/rG cis WC/SE base pair (see Figure 3f and the corresponding discussion in ref 7). This leads to a considerable contraction of the C1′−N distance to 7.34 Å and a consequent raise of the intermolecular interaction strength (−30.9 kcal/mol for the trimer). Thus, similar to other water-mediated structures, insertion of a water molecule optimizes the intermolecular contacts in such a way to establish an extensive H-bonding network.[6c] The reason why water insertion is needed in U/rG is obvious: due to the steric hindrance of the N2 amino group of guanine, its N3 nitrogen cannot accept H-bonds from endocyclic H-bond donors, such as N3 of uracil. Let us note, however, that the computed C1′−N distance of the less stable interaction pattern reproduces better the crystal value (7.74 Å). Comparison of the optimized and X-ray geometries in Figure 11 also reveals that the less stable variant is in better overall agreement with that found in crystal. Note that in an RNA system the water molecule, presumably a long-residency one, may rattle between two or more local binding

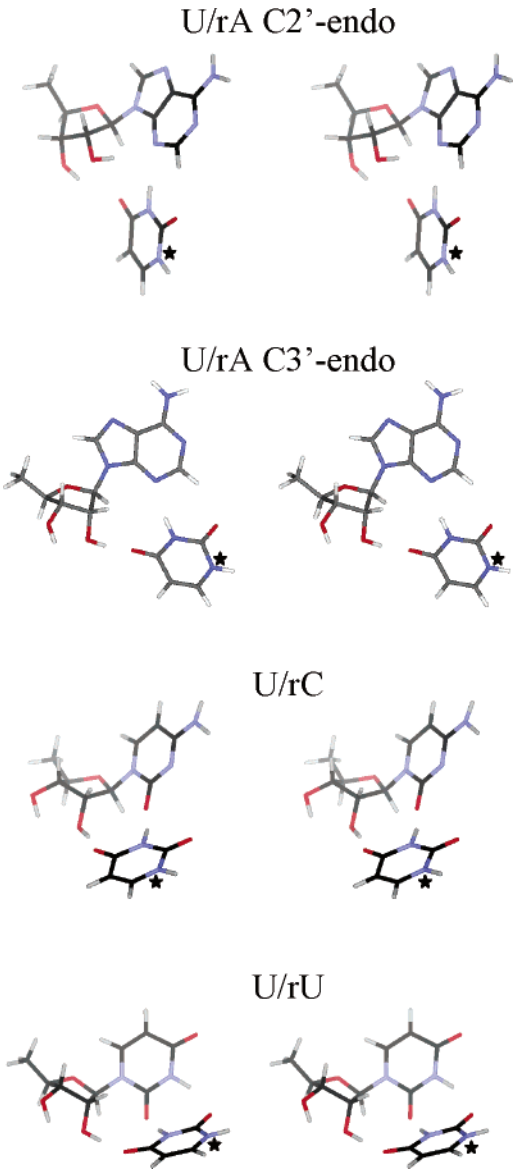

Figure 8. Gas-phase optimized geometries for the I_3 isosteric subfamily of the trans WC/SE binding type computed at the Becke3LYP level of theory (stereoview). The N1 position of uracil is marked with a star.

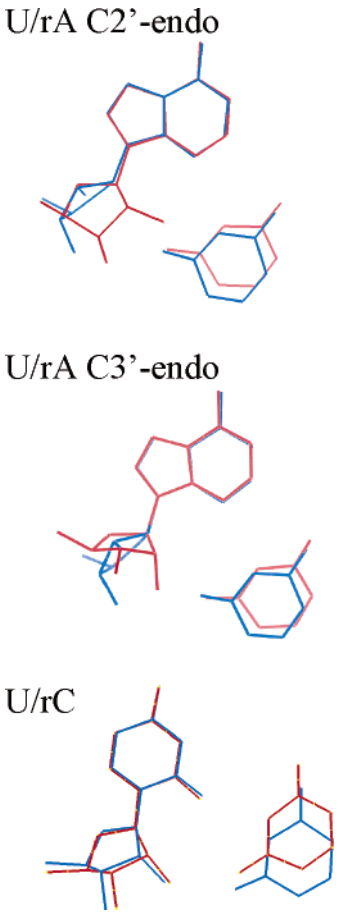

Figure 9. Overlay of the computed and X-ray geometries for the I_3 isosteric subfamily.

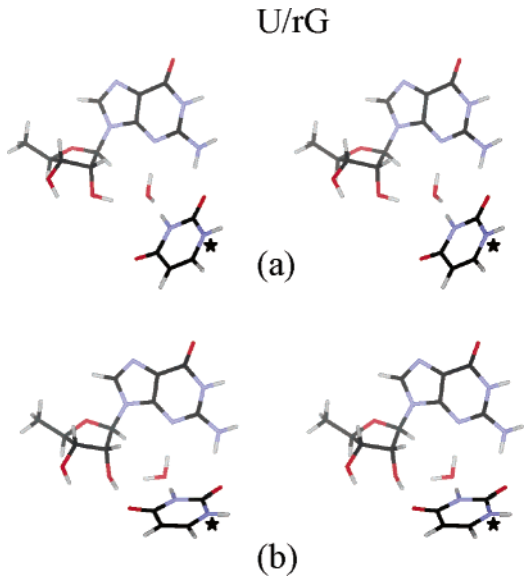

Figure 10. Gas-phase optimized geometries of the U/rG trans WC/SE base pair with different H-bonding patterns (stereoview). Both structures were obtained at the B3LYP level of theory: (a) less stable structure, $\Delta E = -26.4$ kcal/mol; (b) more stable structure, $\Delta E = -30.9$ kcal/mol. The N1 position of uracil is marked with a star.

patterns while its position may be determined by its involvement in a broader network of H-bonds.[22]

Conformational Switch from C3′-endo to C2′-endo. While the cis WC/SE base pairs rigorously maintain the conventional C3′-endo ribose conformation, C2′-endo pucker has been observed twice in the crystal geometry of the trans structures.

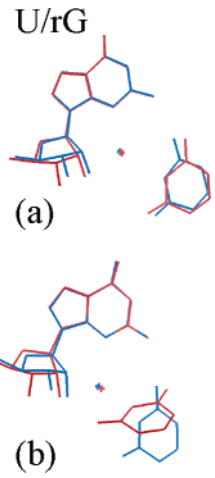

Figure 11. Overlay of the computed and X-ray geometries for the U/rG trans WC/SE base pair with different H-bonding patterns. Both structures were obtained at the B3LYP level of theory: (a) less stable structure, $\Delta E = -26.4$ kcal/mol; (b) more stable structure, $\Delta E = -30.9$ kcal/mol.

Both binding patterns involving the C2′-endo sugar conformation occur in the crystal of *M. marismortui* (NDB code RR0033, resolution 2.3 Å).[3e] To clarify the driving force of this conformational change, we have optimized the structure of the above two C2′-endo base pairs starting from the crystal geometry but inverting the C2 and C3 positions of the ribose into the C3′-endo pucker.

The total electronic energy of the C/rA pair does not depend much on the sugar conformation; the energy difference between the C2′-endo and C3′-endo structures is below 3 kcal/mol. Although crystallographic examples predict a C2′-endo pucker in this base pair, the C1′−N distance in the gas-phase optimized geometry (6.54 Å) strongly underestimates the experimental value (7.88 Å). The most likely reason for the shrinkage of the C1′−N distance in the gas phase is the significant noncoplanarity of the aromatic rings. In contrast, the optimized structure with C3′-endo ribose provides a satisfactory estimate (7.96 Å) of the experimental C1′−N distance and preserves the coplanarity of the nucleobases even in the gas-phase optimized structure. Thus, we recommend accepting the crystallographic assignment of the sugar conformation in this binding pattern with caution. The intramolecular H-bond noticed above for the C2′-endo conformation has been abolished.

On the basis of the total electronic energies, the U/rA structure with the C3′-endo ribose conformation is preferred over its C2′-endo counterpart by −4.3 kcal/mol (for optimized geometries, see Figures 8 and 9). The difference of the total interaction energies is 7.3 kcal/mol in favor of the C3′-endo pucker. Thus, the stability increase can be fully attributed to the improvement of the intermolecular contacts. Rearrangement of the intermolecular H-bonds, however, has not influenced significantly the mutual base−base positions, leaving the C1′−N distance practically unchanged (6.7 Å in both structures.). Thus, on the basis of the relative stabilities and steric parameters, we do not see any obvious reason why the C2′-endo conformation is adopted in the crystal, although it obviously could be due to the overall backbone topology. We can conclude that, in isolation, both base pairs should prefer the C3′-endo arrangement. Further discussion of the occasional occurrence of C2′-endo puckers in RNA crystal structures is beyond the scope of the present paper.[23]

Isostericity. In our previous paper, we suggested that the isostericity of the cis WC/SE base pairs is rooted in the intrinsic

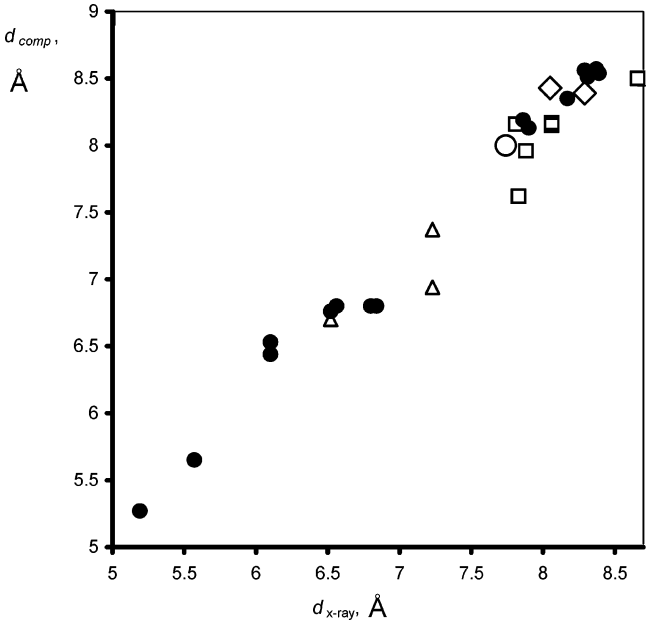

Figure 12. Correlation diagram between the computed (d_{comp}, Å) and crystallographically suggested ($d_{X\text{-}ray}$, Å) C1′−N distances for the cis and trans WC/SE base pair families: (●) cis WC/SE base pairs; (□) I_1 subfamily of the trans WC/SE base pairs; (◊) I_2 subfamily of the trans WC/SE base pairs; (△) I_3 subfamily of the trans WC/SE base pairs; (○) I_4 subfamily of the trans WC/SE base pairs.

properties of the isolated structures. We have shown that the base pairing energy is very similar for the members of the family and established a linear relationship between the computed and crystallographic C1′−N distances.

In Figure 12, we present a diagram which plots the computed versus experimental C1′−N distances (including those predicted by isostericity modeling). In addition to the trans WC/SE structures, the plot also includes data for the cis WC/SE family. Numerical data of the trans structures presented in the figure are listed in Table 3 along with those of the predicted G/rG and G/rA base pairs (vide infra). Except for trans A/rC and A/rU (not shown in the diagram), the newly added data are consistent with the established linear relationship. Note that neither of the above two exceptions have so far been identified in RNA crystals. Therefore, here, instead of X-ray data, we have used estimated values from ref 1b for comparison with the computed C1′−N distances. The estimation was made assuming a complete structural similarity with C/rC; however, the rigorous geometry optimization of the isolated A/rC and C/rC base pairs resulted in a different trans WC/SE H-bonding pattern (vide supra). Thus, we suggest revising the classification of trans WC/SE A/rC and A/rU according to the computed DFT structural parameters, although the subfamily I_1 was correctly predicted by the isostericity modeling.

Predicted Structures. We also attempted to predict the structure and stability of the G/rG and G/rA trans WC/SE base pairs. There were not even suggested structural data available for these two members of the trans WC/SE base pair family. Hence, we have arbitrarily selected initial geometries for optimization in the following way. In part, we projected the H-bonding pattern of the I_2 subfamily to the G/rG and G/rA structures assuming an O2′−H···O6(G) type sugar−base binding. Then, we rotated the purine rings into a nearly perpendicular position to enable interaction with N3 of the SE base. This is a universal interaction type, present in all trans WC/SE base pairs having rA or rG in the SE position.

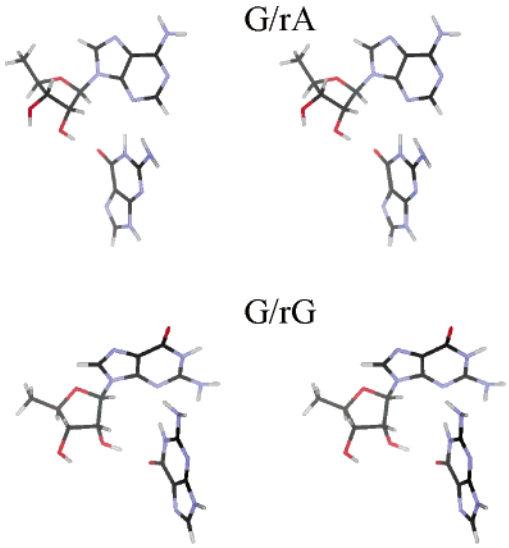

Figure 13. Optimized structures of the hypothetical G/rG and G/rA base pairs computed at the Becke3LYP level of theory (stereoview).

The optimized geometries of G/rG and G/rA are depicted in Figure 13. The characteristic H-bond parameters (see Table 1) and C1′−N distances (listed in Table 3) make the new structures unambiguously similar to the two known trans WC/SE G/rPyr base pairs. Nevertheless, the corresponding H-bonding contacts are slightly longer than those in the I_2 subfamily, leading to a moderate reduction of the interaction energies (see Table 2). Note, however, that, with interaction energies of −20.8 and −19.9 kcal/mol, these base pairs are intrinsically very stable. In summary, we suggest a tentative assignment of these base pairs to the I_2 isosteric subfamily.

Comment on Force Field Accuracy. The energy data show that the force field interaction energies are in a very reasonable agreement with the QM calculations. Note, however, that the structures were not reoptimized by the force field. We nevertheless assume that it is entirely appropriate to use the QM-optimized geometries for the comparison. The potential energy surface of these RNA base pairs is more complex (and flat) compared to base pairs not involving sugars. Thus, it is fair to assume that a gas-phase optimization by the force field may lead to a somewhat different geometry of some base pairs, as shown for the cis WC/SE family in our preceding study.[7] More specifically, the force field often tends to increase the nonplanarity of the base pairs compared to QM HF and DFT optimizations.[7,24] Nevertheless, to the best of our knowledge, this does not indicate any significant force field defect. The AMBER force field is primarily tuned for condensed phase simulations, while RNA simulations specifically aimed to study non-Watson−Crick base pairs suggested a very good correspondence between MD and X-ray data.[5a] The difference may also be partly caused by (i) the truncation of the subsystems for the force field calculations when comparison with QM data is attempted or (ii) perhaps underestimation of the dispersion forces in the course of the DFT QM optimization. The present results thus confirm that the Cornell et al. force field has a very good balance of intermolecular interaction terms for nucleobases. As discussed in detail elsewhere, the agreement for base pairing is partly due to the compensation of errors, since the missing polarization/charge transfer effects are compensated for by a more polar charge distribution (the AMBER charges are derived at the HF level). In addition, the deformation energies of the bases are substantially higher from force field calculations.[11] AMBER charge distribution based on electrostatic potential

fitting appears to be optimal for the electrostatic part of the base−base interactions, while the Lennard-Jones term allows a surprisingly good match with the absolute values of interaction energies, even for base stacking.[25−27] Actually, the absolute AMBER values compare better with the recent reference QM calculations[11,27] than with the medium level QM data reported in 1990s. The older QM data underestimated the intermolecular electron correlation (dispersion) effects. The latest QM calculations are based on a complete basis set extrapolation at the MP2 level followed by higher-order electron correlation correction, which is assumed to guarantee a highly accurate description of the intermolecular electron correlation effects. In conclusion, the AMBER force field provides, within the framework of pair-additive force fields, a close-to-ideal description of a wide range of nucleobase interactions. More artifacts in simulations are likely caused by the sugar−phosphate backbone parametrization, as indicated for example by backbone flips leading to incorrect loop topology in guanine quadruplex simulations[28] and rather unexpected long-living substates in extended B-DNA simulations.[29,30]

Conclusions

Due to the presence of the 2′-OH hydroxyl group of ribose, RNA molecules utilize an astonishing variability of base pairing patterns to build up their structures and perform the biological functions. We have characterized the trans Watson−Crick/sugar edge (trans WC/SE) RNA base pair family using reliable quantum chemical calculations. Our study is the first computational analysis of this type of base pairing.

The complete trans WC/SE base pair family consists of 16 different base pairs. Ten of them are known from RNA X-ray structures, four were suggested earlier on the basis of qualitative isostericity modeling, and two were entirely unknown.

In this study, the gas-phase optimized geometries and interaction energies are reported for all 16 trans WC/SE base pairing patterns. The 10 structures having their X-ray precedents are in very good agreement with the available X-ray crystal data. Where C2′-endo pucker is present in the X-ray structure, QM calculations are done considering both the C2′-endo and C3′-endo arrangements, with the later one being more stable in isolation. Some trans WC/SE base pairs can adopt two isomers, depending on the orientation of the O2′−H group. In one isomer, the hydroxyl group is involved in a conventional H-bond; in the other, it forms an amino-acceptor interaction with C(N4) or A(N6) of the other base. X-ray structures are consistent with the first option; however, energy data clearly show that formation of the amino-acceptor contact is feasible and cannot be ruled out.

Stable structures are predicted for all of the remaining six possible members of this family not seen in RNAs so far. Among these six base pairs, the computations substantially refine two structures suggested earlier on the basis of simple isosteric considerations, while, for two additional trans WC/SE base pairs predicted in this study, no arrangement was suggested before. Thus, our study brings a complete set of trans WC/SE base pairing patterns. All the modeled structures are very stable.

The interaction strength of the trans WC/SE base pairs spans a considerably wider range (from −8 to −28 kcal/mol) than that of their cis counterparts (from −16 to −22 kcal/mol). Thus, many trans WC/SE base pairs are comparably stable as the GC WC base pair (−29.4 kcal/mol when evaluated with the same method), whereas some others are considerably weaker than the AU WC one (−15.3 kcal/mol). On the contrary, the characteristic C1′−N distances adopted in the crystals (from 6.5 to 8.7 Å) as well as in the isolated models (from 6.6 to 8.5 Å) fall within a tighter interval in the trans structures compared to the cis WC/SE base pairs (from 5.2 to 9.3 Å and from 5.3 to 8.7 Å for the crystal and gas-phase structures, respectively). This suggests that, relative to their cis analogues, in the trans WC/SE base pairs, the interaction strength gets an elevated role at distinguishing between the isosteric subgroups. It is well established that the gas-phase strength of base pairs cannot be directly compared to the stabilities of the pairs in nucleic acids. Nevertheless, it is fair to assume that at least part of the base pair energy variability is transferred to nucleic acids, especially for base pairs buried deeply inside large RNAs with reduced solvent screening. In other words, the strength of binding should play a more important role in this family compared to the cis WC/SE one.

In the majority of the studied trans WC/SE structures, the leading interaction energy term is the base−base contribution. Comparing the corresponding sugar−base and base−base interaction energy terms (see the last two columns of Table 2) in those 10 base pairs with a gas-phase stability exceeding 15 kcal/mol, one may recognize that in six cases the base−base interaction exceeds the sugar−base term by at least 2 kcal/mol, while in two additional cases the two contributions are roughly equal. The opposite is true for the cis WC/SE base pairs. As shown in Table 2 of ref 7, here, the sugar−base term is favored over the base−base contribution with a score of 7:3. Thus, cis−trans *isomerism of the WC/SE base pairs determines the distribution of interaction energy over the sugar−base and base−base terms*. This may affect the balance of interactions involving the sugar and bases when other interactions compete with or complement the WC/SE base pair.

The linear relationship found between the computed and crystallographically determined C1′−N (and thus also C1′−C1′) interbase distances of the cis WC/SE base pairs can be extended to all WC/SE interaction patterns. This finding, on one hand, means that the isostericity of the WC/SE base pairs is rooted in the atomic level similarity of the isolated base pairs. In addition, it demonstrates that DFT calculations of isolated model geometries are capable of reproducing base pair geometries occurring in RNA crystals.

Interaction energies obtained from molecular mechanics calculations, in addition to the previously published results on the cis WC/SE base pairs, justify the applicability of the Cornell et al. force field[16,17] for qualitative molecular modeling studies on RNA.

Acknowledgment. This study was supported by Wellcome Trust International Senior Research Fellowship in Biomedical Science in Central Europe GR067507 (J.Š. and N.Š.) and grant 203/05/0009, Grant Agency of the Czech Republic (J.Š., J.E.Š., and N.Š.). J.L., J.Š., and J.E.Š. acknowledge the financial support from NIH grant S06 GM008047, NSF-CREST grant HRD-0318519, and ONR grant N00034-03-1-0116. The Institute of Biophysics acknowledges support by the grant AVOZ50040507, Ministry of Education of the Czech Republic. We are very indebted to our reviewers for their very constructive comments on the manuscript.

Supporting Information Available: Table summarizing the angle of the C1′−N vectors and Cartesian coordinates of all structures depicted in Figures 2−11 and 13. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) (a) Leontis, N. B.; Westhof, E. *Q. Rev. Biophys.* **1998**, *31*, 399−455. (b) Leontis, N. B.; Stombaugh, J.; Westhof, E. *Nucleic Acids Res.* **2002**, *30*, 3497−3531.

(2) (a) Cate, J. H.; Gooding, A. R.; Podell, E.; Zhou, K. H.; Golden, B. L.; Kundrot, C. E.; Cech, T. R.; Doudna, J. A. *Science* **1996**, *273*, 1678−1685. (b) Ferre-D'Amare, A. R.; Zhou, K. H.; Doudna, J. A. *Nature* **1998**, *395*, 567−574. (c) Scott, W. G.; Murray, J. B.; Arnold, J. R. P.; Stoddard, B. L.; Klug, A. *Science* **1996**, *274*, 2065−2069.

(3) (a) Moore, P. B.; Steitz, T. A. *Annu. Rev. Biochem.* **2003**, *72*, 813−850. (b) Wimberly, B. T.; Brodersen, D. E.; Clemons, W. M.; Morgan-Warren, R. J.; Carter, A. P.; Vonrhein, C.; Hartsch, T.; Ramakrishnan, V. *Nature* **2000**, *407*, 327−339. (c) Klein, D. J.; Schmeing, T. M.; Moore, P. B.; Steitz, T. A. *EMBO J.* **2001**, *20*, 4214−4221. (d) Nissen, P.; Ippolito, J. A.; Ban, N.; Moore, P. B.; Steitz, T. A. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 4899−4903. (e) Ban, N.; Nissen, P.; Hansen, J.; Moore, P. B.; Steitz, T. A. *Science* **2000**, *289*, 905−920. (f) Lewis, H. A.; Musunuru, K.; Jensen, K. B.; Edo, C.; Chen, H.; Darnell, R. B.; Burley, S. K. *Cell* **2000**, *100*, 323−332. (g) Su, L.; Chen, L.; Egli, M.; Berger, J. M.; Rich, A. *Nat. Struct. Biol.* **1999**, *6*, 285−292.

(4) Ogle, J. M.; Carter, A. P.; Ramakrishnan, V. *Trends Biochem. Sci.* **2003**, *28*, 259−266.

(5) (a) Reblová, K.; Špačková, N.; Stefl, R.; Csaszar, K.; Koča, J.; Leontis, N. B.; Šponer, J. *Biophys. J.* **2003**, *84*, 3564−3582. (b) Leontis, N. B.; Westhof, E. *RNA* **1998**, *4*, 1134−1153. (c) Vallurupalli, P.; Moore, P. B. *J. Mol. Biol.* **2003**, *325*, 843−856.

(6) (a) Šponer, J.; Mokdad, A.; Šponer, J. E.; Špačková, N.; Leszczynski, J.; Leontis, N. B. *J. Mol. Biol.* **2003**, *330*, 967−978. (b) Brandl, M.; Meyer, M.; Suhnel, J. *J. Biomol. Struct. Dyn.* **2001**, *18*, 545−555. (c) Brandl, M.; Meyer, M.; Suhnel, J. *J. Phys. Chem. A* **2000**, *104*, 11177−11187. (d) Hobza, P.; Šponer, J.; Cubero, E.; Orozco, M.; Luque, F. J. *J. Phys. Chem. B* **2000**, *104*, 6286−6292. (e) Pan, Y.; Priyakumar, D.; MacKerell, A. D., Jr. *Biochemistry* **2005**, *44*, 1433−1443.

(7) Šponer, J. E.; Špačková, N.; Kulhánek, P.; Leszczynski, J.; Šponer, J. *J. Phys. Chem. A* **2005**, *109*, 2292−2301.

(8) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; 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. J.; Cui Q.; Morokuma, K.; Salvador, P.; Dannenberg, J. J.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Baboul, A. G.; 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.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Andres, J. L.; Gonzalez, C.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian98*; Gaussian, Inc.: Pittsburgh, PA, 2001.

(9) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648−5652.

(10) (a) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785−789. (b) Miehlich, B.; Savin, A.; Stoll, H.; Preuss, H. *Chem. Phys. Lett.* **1989**, *157*, 200−206.

(11) Šponer, J.; Jurečka, P.; Hobza, P. *J. Am. Chem. Soc.* **2004**, *126*, 10142−10151.

(12) (a) Eichkorn, K.; Treutler, O.; Oehm, H.; Haeser, M.; Ahlrichs, R. *Chem. Phys. Lett.* **1995**, *242*, 652−660. (b) Weigend, F.; Haeser, M. *Theor. Chem. Acc.* **1997**, *97*, 331−340. (c) Weigend, F.; Haeser, M.; Patzelt, H.; Ahlrichs, R. *Chem. Phys. Lett.* **1998**, *294*, 143−152.

(13) Jurečka, P.; Nachtigall, P.; Hobza, P. *Phys. Chem. Chem. Phys.* **2001**, *3*, 4578−4582.

(14) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553−566.

(15) Case, D. A.; Pearlman, D. A.; Caldwell, J. W.; Cheatham, T. E., III; Wang, J.; Ross, W. S.; Simmerling, C. L.; Darden, T. A.; Merz, K. M.; Stanton, R. V.; Cheng, A. L.; Vincent, J. J.; Crowley, M.; Tsui, V.; Gohlke, H.; Radmer, R. J.; Duan, Y.; Pitera, J.; Massova, I.; Seibel, G. L.; Singh, U. C.; Weiner, P. K.; Kollman, P. A. *AMBER7*; University of California: San Francisco, CA, 2002.

(16) 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.

(17) Cheatham, T. E., III; Cieplak, P.; Kollman, P. A. *J. Biomol. Struct. Dyn.* **1999**, *16*, 845−862.

(18) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269−10280.

(19) (a) Luisi, B.; Orozco, M.; Šponer, J.; Luque, F. J.; Shakked, Z. *J. Mol. Biol.* **1998**, *279*, 1123−1136. (b) Šponer, J.; Hobza, P. *J. Am. Chem. Soc.* **1994**, *116*, 709−714.

(20) Šponer, J.; Florian, J.; Hobza, P.; Leszczynski, J. *J. Biomol. Struct. Dyn.* **1996**, *13*, 827−833.

(21) (a) Russo, N.; Toscano, M.; Grand, A.; Jolibois, F. *J. Comput. Chem.* **1998**, *19*, 989−1000. (b) Wolken, J. K.; Tureček, F. *J. Am. Soc. Mass Spectrom.* **2000**, *11*, 1065−1071. (c) Šponer, J.; Leszczynski, J.; Hobza, P. *J. Phys. Chem.* **1996**, *100*, 1965−1974. (c) Pranata, J.; Wierschke, S. G.; Joergensen, W. L. *J. Am. Chem. Soc*. **1991**, *113*, 2810−2819.

(22) Reblová, K.; Špačková, N.; Koča, J.; Leontis, N. B.; Šponer, J. *Biophys. J.* **2004**, *87*, 3397−3412.

(23) Schneider, B.; Moravek, Z.; Berman, H. M. *Nucleic Acids Res.* **2004**, *32*, 1666−1677.

(24) Šponer, J.; Burda, J. V.; Mejzlik, P.; Leszczynski, J.; Hobza, P. *J. Biomol. Struct. Dyn.* **1997**, *14*, 613−628.

(25) Hobza, P.; Kabeláč, M.; Šponer, J.; Mejzlik, P.; Vondrášek, J. *J. Comput. Chem.* **1997**, *18*, 1136−1150.

(26) Šponer, J.; Gabb, H. A.; Leszczynski, J.; Hobza, P. *Biophys. J.* **1997**, *73*, 76−87.

(27) Jurečka, P.; Šponer, J.; Hobza, P. *J. Phys. Chem. B* **2004**, *108*, 5466−5471.

(28) Fadrná, E.; Špačková, N.; Štefl, R.; Koča, J.; Cheatham, T. E.; Šponer, J. *Biophys. J.* **2004**, *87*, 227−242.

(29) Varnai, P.; Zakrzewska, K. *Nucleic Acids Res.* **2004**, *32*, 4269−4280.

(30) Beveridge, D. L.; Barreiro, G.; Byun, K. S.; Case, D. A.; Cheatham, T. E.; Dixit, S. B.; Giudice, E.; Lankaš, F.; Lavery, R.; Maddocks, J. H.; Osman, R.; Seibert, E.; Sklenar, H.; Stoll, G.; Thayer, K. M.; Varnai, P.; Young, M. A. *Biophys. J.* **2004**, *87*, 3799−3813.