Non-Watson−Crick Base Pairing in RNA. Quantum Chemical Analysis of the cis Watson−Crick/Sugar Edge Base Pair Family

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

Institute of Biophysics, Academy of Sciences of the Czech Republic, Královopolská 135, 612 65 Brno, Czech Republic, National Center for Biomolecular Research, Faculty of Science, Masaryk University, Kotlářská 2, 611 37 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: January 8, 2005

Large RNA molecules exhibit an astonishing variability of base-pairing patterns, while many of the RNA base-pairing families have no counterparts in DNA. The cis Watson−Crick/sugar edge (cis WC/SE) RNA base pairing is investigated by ab initio quantum chemical calculations. A detailed structural and energetic characterization of all 13 crystallographically detected members of this family is provided by means of B3LYP/6-31G** and RIMP2/aug-cc-pVDZ calculations. Further, a prediction is made for the remaining 3 cis WC/SE base pairs which are yet to be seen in the experiments. The interaction energy calculations point at the key role of the 2′-OH group in stabilizing the sugar−base contact and predict all 16 cis WC/SE base-pairing patterns to be nearly isoenergetic. The perfect correlation of the main geometrical parameters in the gas-phase optimized and X-ray structures shows that the principle of isosteric substitutions in RNA is rooted from the intrinsic structural similarity of the isolated base pairs. The present quantum chemical calculations for the first time analyze base pairs involving the ribose 2′-OH group and unambiguously correlate the structural information known from experiments with the energetics of interactions. The calculations further show that the relative importance and absolute value of the dispersion energy in the cis WC/SE base pairs are enhanced compared to the standard base pairs. This may by an important factor contributing to the strength of such interactions when RNA folds in its polar environment. The calculations further demonstrate that the Cornell et al. force field commonly used in molecular modeling and simulations provides satisfactory performance for this type of RNA interactions.

Introduction

The structure and dynamics of nucleic acid molecules are influenced by a variety of contributions. Among those, the interactions involving nucleic acid bases are of particular importance. The bases are involved in two qualitatively different mutual interaction types: hydrogen bonding and aromatic base stacking. The canonical and noncanonical H-bonded base pair geometries found in the atomic resolution crystal structures of DNA correspond to local minima on the potential energy surfaces of isolated DNA base pairs.[1−3] The interaction patterns of nucleic acid bases in RNA are considerably more complex. Standard Watson−Crick (WC) AU and GC base pairs account only for ca. 50% of base pairs in large RNAs. The GU wobble base pair is the third most common base pair occurring in RNA and thus represents the third standard RNA base pair. In a sharp contrast to DNA, a large fraction of RNA H-bonding interactions is represented by highly variable noncanonical (non-Watson−Crick, non-WC) base pairs. They are absolutely essential for building up the complex three-dimensional architectures of large RNAs.

The RNA non-WC base pairs show an astonishing variability of base-pairing combinations. They are involved in the formation of internal RNA loops and segments consisting of several consecutive non-WC base pairs (i.e., the RNA motifs). The ribosomal RNAs are a combination of short WC helices and non-WC RNA motifs. Non-WC base pairs also stabilize distant tertiary interactions needed to fold the RNA molecules into their biologically active structure.[4−6]

Each nucleobase possesses three edges (Figure 1). The WC edge, the Hoogsteen (H) or "C−H" edge, and the sugar edge (SE). The SE involves the ribose where the 2′-hydroxyl group is capable of forming efficient H-bonds, in contrast to deoxyribose in DNA. Two nucleobases can interact with each other through any of the three edges.[4] In addition, the nucleobase can be either in cis or trans orientation with respect to its sugar. According to Leontis and Westhof, this leads to a total of 12 distinct families with 168 possible base-pairing patterns.[4] Actually, 110 of them were already identified in the atomic resolution experiments, where the RNA base-pairing repertoire is further extended (e.g., by the cis WC bifurcated base pairs).

The non-WC base pairing in RNA has been extensively characterized by X-ray crystallography,[4−6] NMR,[7] database analyses,[4] and by molecular dynamics simulations.[8] Surprisingly,

* Corresponding authors. judit@ncbr.chemi.muni.cz (J.E.Š.) or sponer@ncbr.chemi.muni.cz (J.Š.).
† Institute of Biophysics, Academy of Sciences of the Czech Republic.
‡ Masaryk University.
§ Jackson State University.
|| Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic.

10.1021/jp050132k CCC: $30.25
Published on Web 02/22/2005

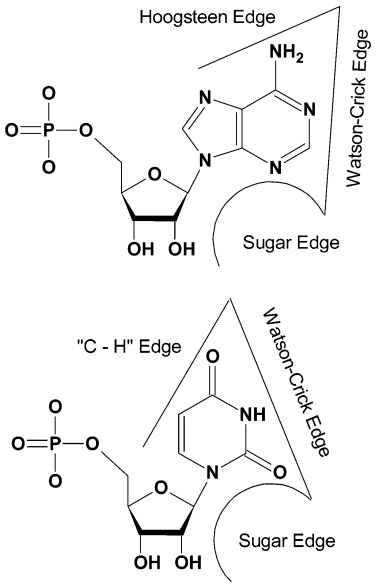

Figure 1. Classification of the interaction sites in purine and pyrimidine nucleobases.

very few quantum chemical (QM) studies have been so far devoted to the non-WC RNA base pairing,[9] despite the fact that quantum chemistry was instrumental in revealing key aspects of DNA base pairing (structures, energies, electronic properties, cation binding, base modifications, hydration, etc.).[1−3,10] For example, a recent characterization of water-mediated RNA base pairs is worth mentioning.[9c] Obviously, the properties of some RNA base-pairing patterns can be evaluated on the basis of studies of the corresponding DNA mismatches. However, the majority of the key RNA non-WC base-pairing families were not yet characterized using QM methods, including all families involving SE interactions. In the absence of such studies, we, for example, do not have any information about the rules governing the relation between the observed non-WC RNA interactions and their intrinsic stabilization energies (i.e., while the structures are known, the corresponding strength of the pairing is unclear). Because the function of biomolecular systems is ultimately determined by the free energies, a proper understanding of the structure−energy relationship in RNA non-WC interactions is of interest. In addition, while molecular dynamics (MD) simulations are nowadays routinely used to study complex RNA molecules, the ability of the force fields to properly describe the non-WC interactions has never been tested.

Visual inspection of the crystallographically found RNA base-pairing patterns shows that many non-WC base pairs are well-paired with two or more hydrogen bonds. On the other hand, there are also pairing patterns with just a single H-bond, highly non-coplanar nucleobase rings, and so forth.[4] It is very likely that some of them are intrinsically very weak and do not correspond even to local minima on the base−base potential energy surface, strikingly contrasting the DNA base-pairing principles. Many of the apparently weak non-WC base pairs still substantially contribute to the RNA architecture, as seen from their conservation patterns. Structural and phylogenetic studies suggest that the shapes of non-WC base pairs are very important for their biological functions, because a given non-WC base pair can often be replaced by other base pairs that have similar shapes (i.e., are isosteric) while mutations leading to nonisosteric base pairs are often not viable.[4]

In the present study, we fill the gap presently existing in the literature on RNA base pairing by examining the molecular interactions of one of the key families of RNA base pairing, the cis Watson−Crick/sugar edge (cis WC/SE) base pairs. In the SE structures, the base (B) is involved in the base pairing via its WC edge, while the nucleoside (N) participates through its sugar edge with the 2′-OH hydroxyl group H-bonded to base B. This paper is thus the first QM study where base pairing is characterized with an involvement of the 2′-OH group. The calculations are performed at the medium level of ab initio theory; however, the recent reference calculations on base pairing show that the methods used in this paper are entirely sufficient for its purpose (see method).[11,12] There are 16 possible base pair members of the cis WC/SE family, while 13 of them have already been seen in RNAs. This makes the cis WC/SE base pair family one of the most biologically relevant RNA interaction patterns with no counterpart in DNA. The main goal of this work is to unambiguously relate the known X-ray base-pairing patterns to their base-pairing energies, thus providing the until-now missing structure−energy relation.

Computational Methods

Geometry Optimizations. Initial structures were built up on the basis of crystal geometries, using structures listed in ref 4b, Figure 6. 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. Wherever possible, we carried out unconstrained geometry optimizations with all parameters relaxed. In a few cases, geometrical constraints had to be imposed to preserve the coordination pattern observed in the crystal structures. Because the choice of constraints was specific for each individual system, they will be described below where relevant.

Geometry optimizations were carried out at the DFT (density functional theory) level of theory using the *Gaussian 98* program package.[13] The density functional was built up by Becke's three-parameter exchange[14] and Lee−Yang−Parr's correlation functional (abbreviated as B3LYP).[15] The 6-31G** basis set was used for geometry optimizations. The B3LYP/6-31G** base pair geometries agree very well with those obtained from reference RI-MP2 (resolution of identity second-order Møller−Plesset perturbational method) calculations executed with an extended cc-pVTZ basis set of atomic orbitals; the B3LYP/6-31G** method overestimates the H-bond distances on average by ca. 0.01−0.05 Å.[11] Note also that the calculated interaction energies are rather insensitive to the base pair geometries, provided the geometries are reasonably close to the minimum.[1,11] We did not use the PW91 DFT method, because although the PW91/6-31G** method gives better absolute values of base-pairing energies compared with the B3LYP/6-31G** technique, the PW91/6-31G** relative base-pairing energies are slightly less accurate and the deformation of the monomers caused by dimerization is exaggerated.[11]

Interaction Energies. Interaction energies were computed on the B3LYP optimized structures using the RIMP2 approach combined with a large aug-cc-pVDZ basis set of atomic orbitals internally stored in the Turbomole code.[16] The RIMP2 interaction energies are close to being identical to true MP2 interaction energies (within 0.03 kcal/mol), while the RIMP2 method is much faster.[12]

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 mainly includes 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. The latter term is mostly repulsive, because the electron correlation reduces the dipole moments of the monomers.[1,2]

All interaction energies were corrected for the basis set superposition error using the standard counterpoise procedure.[17] In contrast to most studies on H-bonding, the interaction energies in the present paper do not include the deformation energies (E^{def}) of monomers upon formation of the base pair, albeit we list this repulsive contribution separately for selected systems. To obtain gas-phase binding energies, E^{def} is to be added to the interaction energies calculated according to eq 1. The main reason for omitting deformation energies in this paper is that in the case of many cis WC/SE base pairs the sugar−base segment undergoes a rather substantial reorientation upon base pairing compared to the isolated nucleoside in the gas phase. Such rearrangements involve the replacement of intramolecular H-bonds by intermolecular base−base and sugar−base contacts, and thus, they are not directly relevant to the strength of the WC/SE interactions. Further, the nucleotide conformation in RNA is also affected by the overall RNA architecture. Thus, while the inclusion of monomer deformation energies is straightforward for small H-bonded systems, the mechanical inclusion of deformation energies for the present base pairs could bias the calculations substantially. We strongly suggest that the deformation energy of monomers is not formally included in the BSSE correction. For a detailed discussion regarding the role of the deformation energies in base-pairing calculations, see ref 11.

Separation of the Base−Base and Base−Sugar Contributions. Besides calculating the interaction energies for the genuine nucleobase−nucleoside dimers, we also 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 an H-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.

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

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

where $E^{BB'}$ is the electronic energy of a system consisting of B and B′ while E^{B} and $E^{B'}$ are the electronic energies of the isolated subsystems B and B′.

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

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

where E^{BS} is the electronic energy of a system consisting of B and S, and E^{B} and E^{S} are the electronic energies of the isolated subsystems B and S.

Reference Calculations. To validate the applied theoretical approaches, we have carried out two sets of preliminary RIMP2 reference calculations for selected systems: geometry optimizations with the aug-cc-pVDZ basis set of atomic orbitals and subsequent interaction energy calculations with the aug-cc-pVTZ basis set. This reference interaction energy value is almost identical to that obtained with the medium-level method utilized in our study. Note that benchmark calculations (such as complete basis set MP2 data with higher-order correlation corrections)[11] were not the purpose of this paper and this task will be accomplished for selected non-WC base pairs from different base pair families soon.

Molecular Mechanics. Molecular mechanics calculations have been performed using the *AMBER7*[18] program package in combination with the Cornell et al. force field parameters (parm99).[19,20] 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.[21] 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 a correction for the deformation energies. Anyway, the AMBER deformation energies are rather negligible compared to the QM ones.[11]

Results and Discussion

Geometrical Characteristics of the cis WC/SE Base Pairs Obtained by Unconstrained Optimizations. Table 1 summarizes the main geometrical parameters of 10 optimized cis WC/SE structures obtained by unconstrained optimizations, corresponding to (at least) local minima on the potential energy surface (see also Figure 2; more details can be subtracted from the coordinate files listed in the Supporting Information). Except for the U.rG base pair, in all optimized structures there are two H-bonds, while the SE nucleotide always acts as an H-bond donor via its 2′-OH group and as an H-bond acceptor via one of its base positions. Note that none of these base pairs contains guanine in the WC position. Three additional base pairs were unstable in the X-ray geometries if optimized in isolation (vide infra), while the remaining three combinations have not yet been detected in RNA structures (vide infra). We also optimized the structures of isolated subsystems, and the Δ(X−H) value in Table 1 shows the elongation of the X−H covalent bond in the complexes compared with isolated optimized monomers. The stretching of the X−H bond is usually a good qualitative indicator of the strength of the H-bonds.[11] Interestingly, the largest Δ(X−H) elongations are observed for the uracil N3−H3 group when interacting with the sugar edge of adenosine and guanosine (0.019 and 0.026 Å, respectively). This suggests a salient H-bonding affinity for uracil among the studied bases. For adenine and cytosine in WC positions, the elongations are much smaller, amounting to 0.010−0.018 Å and 0.012−0.018

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

base pair	donor atom (X)	acceptor atom (Y)	distance X−Y	distance X−H	distance Δ(X−H)[b]
A.rU	N6(A)	O2(U)	2.93	1.017	0.010
	O2′	N1(A)	2.76	0.996	0.023
	O3′			0.978	0.006
A.rA	N6(A)	N3(A′)[c]	3.01	1.025	0.018
	O2′	N1(A)	2.77	0.995	*d*
	O3′			0.978	0.006
A.rG	N6(A)	N3(G)	3.04	1.021	0.014
	O2′	N1(A)	2.78	0.988	*d*
	O3′			0.973	0.001
A.rC	N6(A)	O2(C)	2.83	1.022	0.015
	O2′	N1(A)	2.76	0.994	0.016
	O3′			0.975	0.002
C.rA	N4(C)	N3(A)	3.01	1.024	0.018
	O2′	N3(C)	2.75	0.991	*d*
	O3′	O2(C)	3.95	0.976	0.004
C.rC	N4(C)	O2(C′)[c]	2.83	1.024	0.018
	O2′	N3(C)	2.77	0.988	0.010
	O3′	O2(C)	3.50	0.977	0.004
C.rG	N4(C)	N3(G)	3.05	1.022	0.016
	O2′	N3(C)	2.79	0.986	*d*
	O3′	O2(C)	4.57	0.974	0.002
C.rU	N4(C)	O2(U)	2.93	1.018	0.012
	O2′	N3(C)	2.74	0.990	0.017
	O3′	O2(C)	3.47	0.976	0.004
U.rA	N3(U)	N3(A)	2.99	1.039	0.026
	O2′	O2(U)	2.75	0.977	*d*
	O3′			0.971	0.001
U.rG	N3(U)	N3(G)	3.01	1.032	0.019
	O2′	O4(U)	3.43	0.968	*d*
	O3′			0.972	0.000
	N2(G)	O4(U)	2.92	1.022	0.012

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

Å, respectively. The situation is slightly different in the U.rG complex, where the ribose−uracil H-bond is suppressed by a competing H-bond donated by the N2 amino group of guanine.

Lengths of the O2′−H bonds of the ribose unit involved in sugar−base H-bonding unequivocally reveal the presence of hydrogen-bonding interactions with variable strength. In some cases, we could not determine the O2′−H elongation, because optimization of the isolated nucleoside resulted into an intramolecular H-bond formation involving O2′−H.

For the sake of completeness, we also list the O3′−H bond lengths and distances to nearby acceptors. It is to be noted that the formation of an O3′−H···X H-bond would be an artifact of the model system, because O3′ in RNA is blocked by the attached phosphate group. Fortunately, the data evidently rules out such an H-bond in all our structures and justifies our model. To reassure ourselves of this, we have optimized the structure of the three most stable dimers by replacing the hydrogen atom at O3′ by a methyl group. The total interaction energies as well as the pairwise contributions show (vide infra) that the methylation of O3′ does not influence the strength of the base−nucleoside interaction and the terminal 3′-OH group has no effect on the base pairing in the studied base pairs. Note also that the calculations have to be done without considering the phosphate group, because of its formal charge of −1. This charge would have a profound long-range ionic electrostatic effect in studies of isolated base pairs, while such charge effects are known to be entirely screened in relevant environments.[22] This, together with the very good agreement between the calculated and X-ray structures, justifies our model (vide infra).

All four structures having adenine in the WC position exhibit a similar H-bonding pattern. There is a universal O2′−H···N1(A) sugar−base H-bond complemented by an interbase H-bond of the N6(A) amino group with the N3 endocyclic groups of purines and O2 exocyclic groups of pyrimidines. Among them, the A(N3) site is the strongest H-bond acceptor, as revealed by the elongation of the N6−H bond of 0.018 Å (vide supra). Although the N1(A)−O2′ interatomic distances are very similar in the four structures, steric conditions for the H-bond formation between the ribose and the adenine are much better in the pyrimidine containing nucleosides than in their purine counterparts for the following reasons: While for cytidine and uridine the O2′−H2′···N1(A) H-bond perfectly fits the plane of the interacting adenine (the O2′−H2′−N7(A)−N3(A) dihedral angle being less than 4°), a substantial deviation from coplanarity is observed when having adenosine (−22.1°) and guanosine (−39.4°) in the SE position.

We made an attempt to reoptimize the above 10 base pairs with the Cornell et al. force field starting from the B3LYP/6-31G** optimized geometries. Except for the A.rA, A.rG, and U.rA base pairs, the optimized structures are strikingly different from the initial ab initio geometries and preserved neither the original H-bonding pattern nor the mutual base−base orientation. For example, T-shaped positions of the bases are seen. In addition, in 4 out of the 10 cases, a conformational switch from C3′ endo to C2′ endo has been observed in the ribose part. This result, however, does not necessarily indicate any major inaccuracy of the force field for several reasons. First, the force field is not calibrated to accurately study gas-phase base pairs. The difference may also be partly caused by different optimization criteria of the molecular mechanical code, possibly allowing the local minima to escape. Further, the force field, in contrast to the DFT procedure, does include the dispersion attraction. Therefore, the DFT procedure is considerably less likely to lead to a transition from H-bonded structures to T-shaped or stacked ones, as the stability of T-shaped and stacked structures is severely underestimated by the DFT quantum chemical method.[2] It is to be noted that a very similar result was reported previously in studies of isolated nucleobase triads, where the AMBER optimizations typically resulted in T-shaped or stacked structures while Hartree−Fock optimizations preserved the H-bonded planar structures.[23]

Constrained Optimizations. Three optimized cis WC/SE structures with guanine in the WC position did not retain the hydrogen-bonding patterns determined in their crystal structures. In these cases, geometrical constraints were applied to keep the bases in a crystal-like orientation. Figure 3 presents a summary of the structures obtained from full as well as constrained optimizations.

The crystal structure of the G.rC cis WC/SE pair exhibits two short intermolecular contacts: (i) the closely spaced N2(G) and O2′ enable the formation of either an N2(G)−H···O2′ or an O2′−H···N2(G) H-bond, and (ii) there is an N1(G)−H···O2(C) H-bond. Full optimization, however, resulted in the standard WC/WC base pair. Thus, the C2(C)−N1(G)−C2(G) valence and the C2(C)−N1(G)−C2(G)−N3(G) torsion angles were fixed at the crystal value to prevent displacement of the guanine ring toward the WC edge of cytosine. The resultant gas-phase geometry, in line with the crystal structure, reveals O2′−H···N2(G) and N1(G)−H···O2(C) H-bonds.

There are three potential H-bonding contacts in the crystal geometry of the G.rG cis WC/SE base pair: (i) N2(G)−

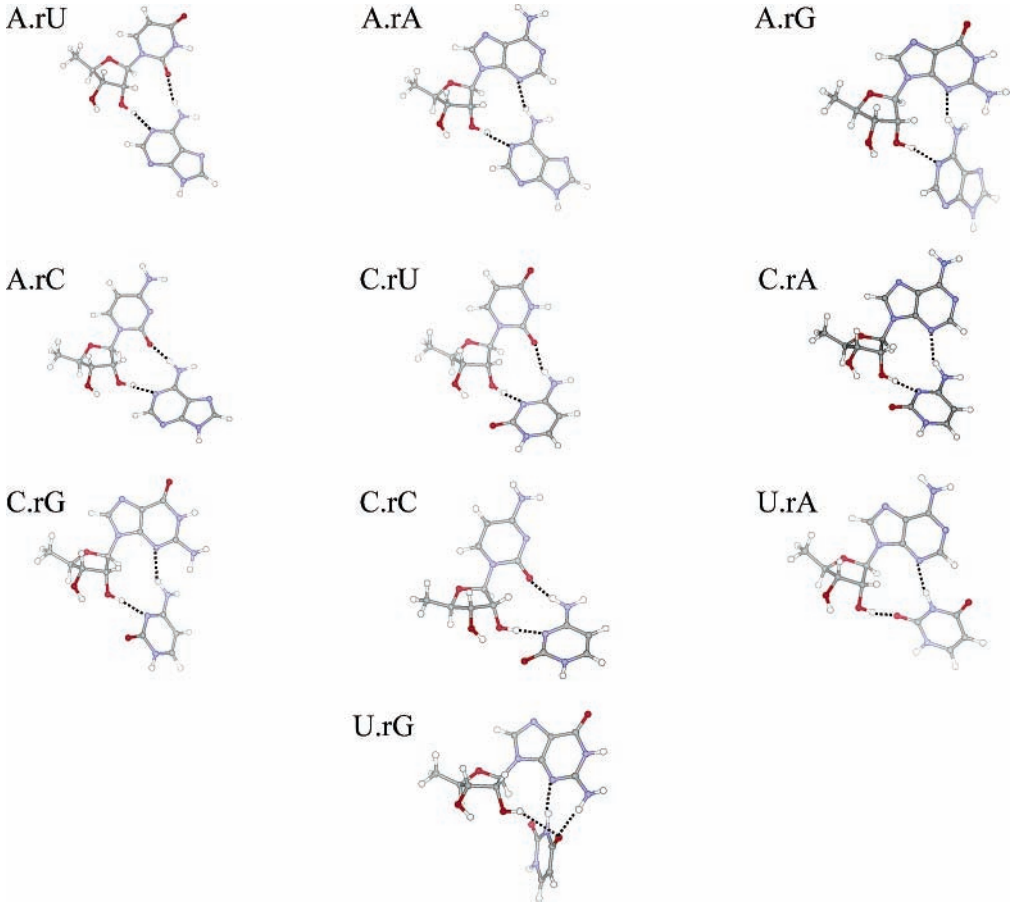

Figure 2. Gas-phase geometries of ten cis WC/SE base pairs obtained from full optimization at B3LYP/6-31G** level of theory. Dashed lines indicate H-bonding contacts.

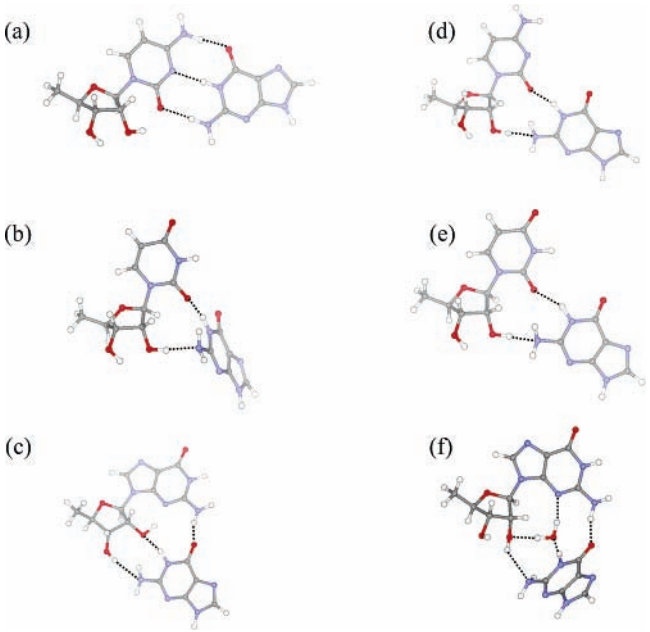

Figure 3. Gas-phase optimized geometries (B3LYP/6-31G** level of theory) of the G.rC, G.rG, and G.rU cis WC/SE base pairs. Structures listed in the left column were obtained with full relaxation of all geometrical parameters: (a) G.rC, (b) G.rU, (c) G.rG. Right column: (d) constrained optimized geometry for G.rC; (e) constrained optimized geometry for G.rU; (f) fully optimized geometry of the water-inserted G.rG cis WC/SE base pair. Dashed lines indicate H-bonding contacts. For description of the geometrical constraints imposed, see text.

H···O2′ or O2′−H···N2(G), (ii) N1(G)−H···N3(G), and (iii) O6(G)···H−N2(G′) (G′ indicates guanine in the SE position). Upon full optimization, the C3′ endo ribose switched to a C2′ endo conformation with a concomitant change in the mutual base−ribose positions inside the guanosine part. The resultant H-bonding pattern significantly differed from the desired one: There was an internal H-bond between 2′-OH and N3(G′), while the interbase interaction was represented by an atypical N2(G′)−H···O6(G) H-bond. Constraining the N9(G)−C1′−O4′−C4′ angle at the crystal value did not result in any relevant geometry, and even subsequent freezing of N3(G′)−C2(G′)−C6(G)−N1(G), N3(G′)−C2(G′)−C6(G), and C2(G′)−C6(G)−N1(G) angles was not useful. In this case, the potential energy surface was so flat near the minimum that optimizations failed. Thus, we conclude that the G.rG cis WC/SE structure is intrinsically unstable. In addition, the X-ray structures suggest that a water bridge mediates the N1(G)−H···N3(G′) interaction. Including the water molecule, indeed, we were able to allocate a minimum reflecting the crystallographically suggested H-bonding pattern. In the optimized structure, the water molecule not only mediates the interaction between guanines but also stabilizes the 2′-OH group in a further H-bond. This cooperative network of H-bonding interactions leads to an overall stabilization of the G···rG pair (see Figure 3f).

In the G.rU complex, the guanine is displaced during the full optimization and forms an amino acceptor O2′−H···N2(G) H-bond.[24] This is accompanied by a rotation of the aromatic rings about the N1(G)−C2(U) axis by ca. 90°, which is not consistent with the intended cis WC/SE arrangement. Thus, we constrained the N1(G)−C2(U)−N1(U), C2(G)−N1(G)−C2(U), N1(G)−C2(U)−N1(U)−C6(U), and C2(G)−N1(G)−C2(U)−N1(U) angles, and this was sufficient to preserve the nearly coplanar position of the two interacting bases. The major

TABLE 2: Interaction Energies and Deformation Energies (kcal/mol) for the Optimized cis WC/SE and Standard WC/WC 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[a]**

cis WC/SE base pair	interaction energy			deformation energy, QM	pairwise terms		nonadditivity term[d]
	RIMP2	HF	AMBER		B···S[b]	B···B′[c]	
A.rU	−16.5	−10.7	−14.9	4.3	−9.4	−4.4	−2.7
	(−17.4)[e]	(−10.1)[e]			(−9.9)[e]	(−5.4)[e]	(−2.1)
A.rA	−17.1	−9.7	−14.3		−6.2	−9.0	−1.9
			(−14.0)[f]				
A.rG	−16.0	−7.9	−14.8	5.9	−8.5	−6.6	−0.9
	(−17.0)[e]	(−7.4)[e]	(−15.4)[f]		(−9.0)[e]	(−7.2)[e]	(−0.8)
A.rC	−17.9	−11.8	−15.7		−9.2	−6.8	−1.9
C.rA	−19.5	−12.0	−17.0		−9.8	−7.0	−2.7
C.rC	−21.8	−16.0	−21.0		−10.4	−8.4	−3.0
C.rG	−17.0	−9.1	−16.4		−9.2	−6.5	−1.3
C.rU	−20.1	−14.4	−19.5	5.9	−10.6	−6.5	−3.0
U.rA	−16.9	−10.8	−15.9		−5.7	−9.9	−1.3
			(−16.4)[f]				
U.rG	−16.0	−8.5	−18.1	6.1	−3.4	−12.5	−0.1
GC WC	−29.4	−25.5	−28.5	4.0			
AU WC	−15.3	−10.4	−13.5	1.7			
GU Wobble	−16.8	−13.5	−15.5	2.3			
C.rMeA[g]	−19.3	−11.6			−9.8	−7.1	−2.4
C.rMeC[g]	−21.0	−15.4			−10.2	−8.1	−2.7
C.rMeU[g]	−19.1	−13.5			−10.1	−6.2	−2.8

[a] For further details, see Methods section. [b] B = nucleobase, interacting via its WC edge; S = sugar. [c] B′= nucleobase interacting via its S edge. [d] The nonadditivity term is approximated as $\Delta E^{BN} - \Delta E^{bb} - \Delta E^{bs}$. Note that its value may be affected by the fact that a covalent bond is cut to derive the individual terms. [e] Reference values from RIMP2/aug-cc-pVTZ//RIMP2/aug-cc-pVDZ calculations (which should be within 1 kcal/mol of the basis set limit[11]) are in parentheses. [f] Values in parentheses have been computed for optimized geometries obtained from AMBER calculations. [g] In these structures, the ribose is methylated at O3′, with no apparent effect on the interaction.

difference between the two optimized structures is in the geometry of the secondary hydrogen bond formed between N1(G) and O2(U). It seems that this contact is optimized at the expense of the above-mentioned ring rotation. The N1(G)−O2(U) distance is by 0.09 Å shorter in the fully relaxed (2.94 Å) geometry than in the constrained (3.03 Å) geometry.

The apparent instability of several base pairs is not surprising. The compact RNAs can stabilize and utilize interbase geometries that are not local minima for isolated base pairs. The likely reason the G.rC pair adopts a WC/WC arrangement upon full optimization is the very steep slope of the potential energy surface around this minimum, which completely masks the shallow minimum corresponding to the WC/SE geometry. The rotation of the aromatic rings in the G.rU structure is due to the great flexibility of the isolated pair and does not change the hydrogen bonding pattern observed in the crystal. This is also reflected by the relatively low energy difference (6.3 kcal/mol, vide infra) between the fully optimized and constrained geometries. The gas-phase instability of the G.rG WC/SE pair, in accord with X-ray studies, indicates that water insertion is essential for this pairing pattern.[9c]

Interaction Energies. Table 2 shows interaction energies of the 10 stable unconstrained cis WC/SE base pairs and compares them with standard RNA base pairs.

The first column in Table 2 presents the total (RIMP2) interaction energy (ΔE). With the aug-cc-pVDZ basis set, we expect that these interaction energies are underestimated (in absolute values) by ca. 2 kcal/mol compared to the basis set limit values, while the relative order of stability is entirely correct.[11] The next column gives the HF component of the interaction energy (ΔE^{HF}, see method). Although these values are markedly reduced compared to the corresponding RIMP2 data, undoubtedly, the classical electrostatic components, such as polarization and charge transfer terms cannot be disregarded when elucidating the stabilizing forces in the cis WC/SE base pairs.[10i]

The difference between the RIMP2 and HF values is mainly due to the attractive dispersion energy, partly reduced by repulsive electron correlation correction to the electrostatic energy. The $\Delta E - \Delta E^{HF} = \Delta E^{cor}$ difference for the cis WS/SE base pairs is in the range −5.7 to −8.1 kcal/mol. These values are by ca. 3 kcal/mol more negative comparing to the three standard base pairs (GC WC, AU WC, and GU Wobble; $\Delta E - \Delta E^{HF} = -3.5$ to −4.9 kcal/mol, see Table 2). Therefore, the nature of stabilization in the cis WC/SE base pairs is similar to base pairs not involving the sugar 2′-OH group, with dominating electrostatic attraction followed by the dispersion term.[1,11] The weight and the absolute value of the dispersion energy are, however, enhanced compared with base pairs involving only the base−base interactions.[11] This reflects the larger contact area of the cis WC/SE base pairs. The increased dispersion contribution relative to standard base pairs may represent a considerable advantage for folding of the RNA molecules in polar environments. Note that the present $\Delta E - \Delta E^{HF}$ term represents the lower estimate of the true dispersion stabilization. As stated above, the basis set size leads to the underestimation of the dispersion attraction by ca. 2 kcal/mol. In addition, the $\Delta E - \Delta E^{HF}$ term contains the repulsive correction to the electrostatic interaction energy (note that the dispersion and electrostatic electron correlation terms cannot be separated).

The third column gives the force-field values calculated by the Cornell et al. force field for the B3LYP-optimized geometry.[19] The force-field values are in very good overall agreement with QM data, with the strength of most base pairs being underestimated by 1−2 kcal/mol. The only exception is the U.rG base pair, where the force-field calculation gives 2 kcal/mol higher interaction energy than the RIMP2 method. The likely reason for this anomaly is an imbalance of the force-field calculations in describing the bifurcated H-bonds formed at the O4 acceptor site of uracil. We have also computed interaction energies for those three structures where the geometry optimization with AMBER has not changed the H-bonding pattern. These

data (listed in column 3 of Table 2 in parentheses next to the corresponding nonoptimized value) illustrate that the interaction energies from AMBER calculations are practically insensitive to minor geometrical changes which do not alter the H-bonding pattern. Thus, despite the small incoherency observed in connection with the U.rG base pair, the force field is capable of effectively reproducing the strength of all H-bonds, including those with O2′−H binding. This is a very encouraging result regarding RNA modeling.

The total strength of the cis WC/SE base pairs is in the range −16 to −22 kcal/mol (i.e., they are rather close in energy). Stabilities of GC WC, AU WC, and GU Wobble base pairs evaluated with exactly the same method are −29.4, −15.3, and −16.8 kcal/mol, respectively, thus spanning a much wider range of energies. As explained in the Methods section, the numbers do not include a correction for the monomer deformation energies, while these are for certain base pairs listed separately in the fourth column of the Table 2. The deformation energies are quite large because of sugar rearrangements (vide supra), and thus, we prefer to separate them from the "net" interaction energies, which we assume to be more relevant to the understanding of base-pairing energies in RNA.

The first column of Table 2 shows that the interaction of cytosine with the nucleosides is noticeably stronger than that of adenine and uracil. The leading term of the cytosine−nucleoside interaction is the sugar−base contribution. On the contrary, the interaction of uracil with nucleosides is prevailed by the base−base contribution. If adenine interacts with a pyrimidine base containing nucleoside, the base−sugar term becomes dominant, while for the purine bases, the base−sugar component is more pronounced.

With one exception, all pairwise interaction energy contributions listed in Table 2 range from −4.4 to −10.6 kcal/mol, showing a mutual compensation between the sugar−base and base−base terms, leading to the overall similarity in the energetics of the base pairs. Except for the two U···nucleoside cis WC/SE pairs, the interbase hydrogen bond always involves an exocyclic amino group of either base, while the sugar−base contact is formed between the O2′ of the ribose and an endocyclic nitrogenous site. Surprisingly, in the U.rG complex, both pairwise contributions fall out of the above range. The extremely weak sugar-base interaction (−3.4 kcal/mol) is counterbalanced with an unusually strong base−base term (−12.5 kcal/mol) supplying a total stability of −16 kcal/mol. The reason for this anomaly is in the formation of competing bifurcated H-bonds involving O4 of uracil. Because of an N2(G)−H···O4(U) H-bonding contact, O4 exhibits strongly deteriorated acceptor activity toward the 2′-OH group of the ribose. Nonadditivity of the pairwise interactions, estimated as $\Delta E^{BN} - \Delta E^{bb} - \Delta E^{bs}$ and listed in the last column of Table 2, are noticeable but not significant.

As noted above, a few X-ray structures required constrained optimizations, and their interaction energies are as follows (not shown in Table 2). The interaction energy for the G.rC complex is −15.2 kcal/mol (i.e., still inside the range of the other base pairs). The dominant contribution comes from the base−base term (−11.5 kcal/mol). The sugar−nucleobase interaction is exceptionally weak (−2.9 kcal/mol), and its HF component is even repulsive (+1 kcal/mol). In contrast, the intermolecular stabilization in the optimized (constrained) structure of G.rU is very low, −9.4 kcal/mol. The interbase and base−sugar terms are weakly attractive (−5.5 and −3.3 kcal/mol, respectively; the latter one is +0.8 kcal/mol at HF level). Evidently, both base pairs with guanine in the WC position have poor sugar−base interactions. Unfortunately, because of the lack of appropriate optimized gas-phase geometry, we could not estimate the interaction energy in the G.rG base pair, and it is quite evident that this base pair is not very stable. However, we have computed the total interaction energy for the analogous water-mediated pair[25] dividing the system into three subsystems, represented by the base in the WC position, guanosine, and water. The resultant value (−29.5 and −18.8 kcal/mol at MP2 and HF levels of theory, respectively) clearly reveals a high-level cooperativity of the H-bonds in this structure, albeit the number cannot be directly compared with the other base pairs. Cooperativity of the H-bonding network in other water-mediated RNA base pairs has already been noticed prior to our work by Brandl et al. Comparing the stability of direct and through-water base pairs, they have concluded that the cooperativity of the H-bonds significantly stabilizes these structures.[9c]

Isostericity. RNA folding depends to a significant extent on steric fit of the molecular interactions. Often, mutations conserving the three-dimensional structure are allowed in RNA. This opens up the way for a number of isosteric substitutions, described in detail by Leontis et al.[4] They consider the C1′−C1′ distance as a measure of the isostericity and classify all known isosteric base pairs with the use of a simple matrix formalism.[4b]

On the basis of the X-ray structures, the 13 known cis WC/SE base pairs form 5 isosteric subfamilies: (i) adenine in the WC position with all 4 possible nucleosides in the SE position, with a characteristic C1′−C1′ distance ranging from 9.0 to 9.5 Å; (ii) cytosine in the WC position with all 4 possible nucleosides in the SE position, with a characteristic C1′−C1′ distance ranging from 7.7 to 8.0 Å; (iii) guanine in the WC position with adenine, cytosine, and uracil in the SE position (characteristic C1′−C1′ distance 8.5−9.1 Å); (iv) uracil with all 4 possible nucleobases in the SE position, representing the shortest C1′−C1′ distance of 5.6−6.5 Å; (v) finally, the guanine−guanosine cis WC/SE self pair itself forms a one-member isosteric subfamily with the longest C1′−C1′ distance of 10.3 Å.

Because of simplifications imposed in the gas-phase model structures (vide supra), instead of the C1′−C1′ distance, we use the corresponding N1−C1′ and N9−C1′ distances (abbreviated as N−C1′) for pyrimidine and purine bases, respectively. Table 3 compares their crystallographically determined and computed values. In Figure 4a, we plotted the experimentally available N−C1′ distances against the corresponding C1′−C1′ values. The excellent correlation clearly verifies that the variation of the C1′−C1′ distances is well-reflected by the N−C1′ values. Similarly, except for G.rG,[26] there is a fair correlation between the data presented in columns 2 and 3 of Table 3, with just a minor systematic discrepancy showing the computed distances slightly longer compared to the crystal ones (see also the correlation diagram in Figure 4b). Thus, the crystallographically suggested classification of the cis WC/SE base pairs into isosteric subfamilies is fully supported by the gas-phase calculations. In other words, the isostericity is determined by the intrinsic properties of the base pairs, including a water insertion into some base pairs that do not form a satisfactory cis WC/SE interaction in the complete absence of solvent molecules.

As noticed above, the interaction energies of the three standard RNA WC/WC base pairs (−15 to −29 kcal/mol) span a much broader interval than those of the cis WC/SE family (−16 to −22 kcal/mol, except for the constrained G···rU). Thus, one may conclude that the substitution principles of the cis WC/

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

		distance	
base pair	measure	crystal[a]	gas-phase QM
A.rA	N9−C1′	8.17	8.35
A.rC	N9−C1′	8.29	8.56
A.rG	N9−C1′	7.86	8.19
A.rU	N9−C1′	8.31	8.51
C.rA	N1−C1′	6.80	6.82
C.rC	N1−C1′	6.56	6.80
C.rG	N1−C1′	6.52	6.76
C.rU	N1−C1′	6.84	6.80
G.rA	N9−C1′	7.9[b]	8.13
G.rC	N9−C1′	8.37	8.57
G.rG	N9−C1′	9.27	8.70
G.rU	N9−C1′	8.39	8.53
U.rA	N1−C1′	5.57	5.65
U.rC	N1−C1′	6.1[b]	6.53
U.rG	N1−C1′	5.19	5.27
U.rU	N1−C1′	6.1[b]	6.44

[a] Crystal data; see ref 4b for further details. [b] Estimated value from model structures (for details, see text).

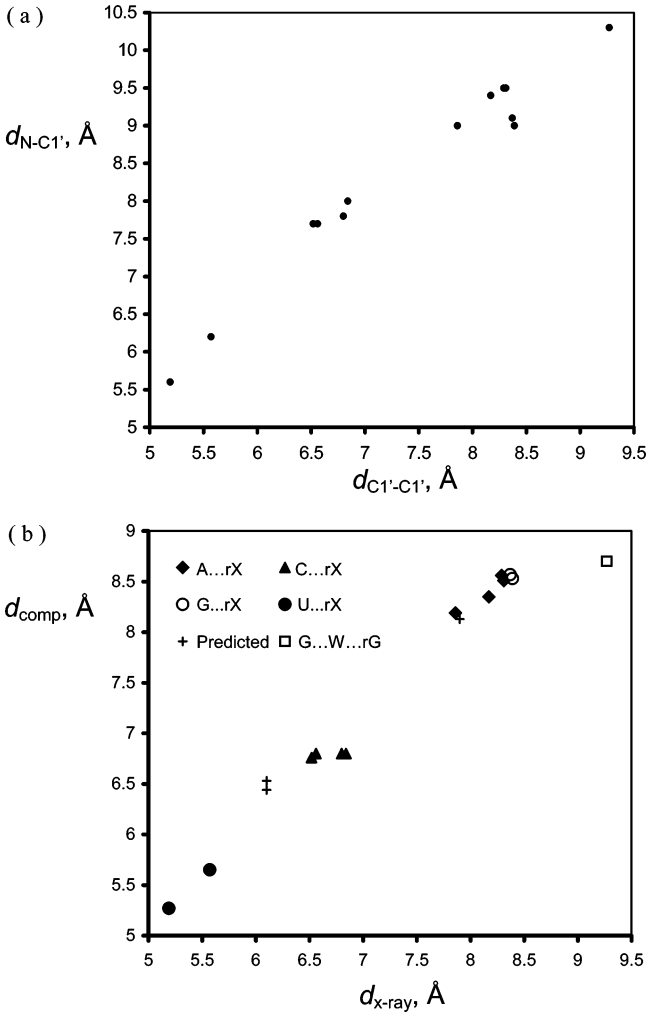

Figure 4. (a) Plot of the experimental N1−C1′ and N9−C1′ distances vs the experimental C1′−C1′ distances for all cis WC/SE base pairs with known crystal structure. (b) Plot of the computed vs crystallographically determined N1−C1′ and N9−C1′ distances.

SE family should indeed be largely controlled by the geometrical factors with likely only marginal effects stemming from the base-pairing energy differences. (Note that GC and AU standard base pairs can often substitute for each other in RNA, despite their pronounced interaction energy difference.) This obviously makes the cis WC/SE family a prominent class of non-WC interactions utilized in building up the complex and modular RNA architectures.

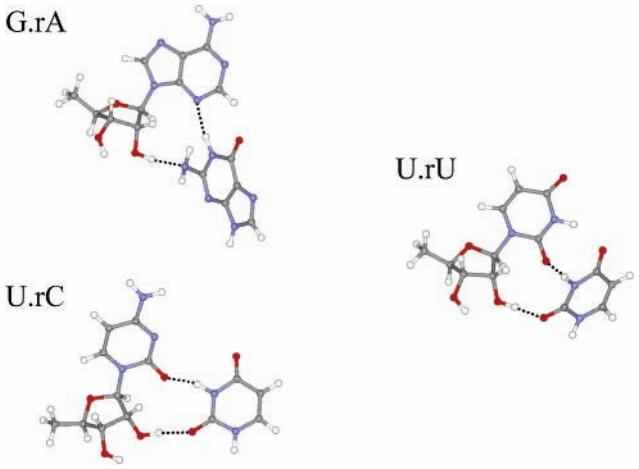

Figure 5. Predicted structures of the U.rC, U.rU, and G.rA cis WC/SE base pairs from B3LYP/6-31G** optimizations.

Predicted Structures. On the basis of Figure 6 of ref 4b, we have built up models for the remaining three cis WC/SE base pairs where the X-ray examples were not yet reported. Their fully optimized gas-phase geometries are shown in Figure 5 (Cartesian coordinates of the structures depicted in the figure can be found in the Supporting Information). The predicted structures are also included in the correlation diagram presented in Figure 4b.

The gas-phase optimized N1−C1′ distances in the U.rC and U.rU structures (6.53 and 6.44 Å, respectively) are somewhat longer than the N9−C1′ distances of the U.rA and U.rG cis WC/SE pairs (5.65 and 5.27 Å, respectively). These findings are in sound agreement with the model building predicting an 0.6−0.9 Å elongation of the C1′−C1′ distances in the U.rPy complexes with respect to their U.rPu counterparts.[4b] Thus, in accord with ref 4b, our gas-phase computations also rank all four cis WC/SE structures with uracil in the WC position into a single isosteric subfamily.

Similarly, the optimized structural parameters of the G.rA cis WC/SE base pair are fully consistent with those of the analogous G.rC and G.rU pairs. Thus, the 0.5 Å reduction of the optimized N−C1′ distance in G.rA with respect to G.rC and G.rU fully confirms the trend proposed on the basis of simple isosteric considerations.[4b]

The computed total interaction energies (−14.6, −15.3, and −17.3 kcal/mol for G.rA, U.rC, and U.rU, respectively) of the three predicted cis WC/SE base pairs overlap the lower edge of the interval found for the crystallographically known structures (−16 to −22 kcal/mol). Thus, the intrinsic properties of these base pairs suggest that they are entirely suitable for building up 3D RNA structures.

Conclusions

We have carried out the first quantum chemical study on one of the key RNA base-pairing families involving sugar−base H-bonding, the cis WC/SE family, consisting of 16 base pairs forming 5 distinct isosteric subfamilies. The calculations complement the structural experimental data by accurate evaluation of the intrinsic stabilities and the nature of interactions in these RNA base pairs. Such structure−energy data were missing until now. A proper understanding of the relation between the

observed structures and the energetics of molecular interactions in compactly folded RNAs is of similar importance to that in proteins.[27]

Twelve out of the thirteen crystallographically known cis WC/SE binding patterns have been found to be intrinsically stable (albeit two of them required a constrained optimization). The only exception is the G.rG base pair, which apparently requires water insertion to be stabilized. It does not rule out that other cis WC/SE base pairs may adopt alternative water-bridged geometries when dictated by the overall RNA fold.

Base-pairing energy in the cis WC/SE structures is surprisingly good and spans a rather narrow range (−16 to −22 kcal/mol), considerably smaller compared with the difference between standard AU and GC base pairs of ca. 14 kcal/mol. The only exception is the G.rU base pair, which is weak (−9.4 kcal/mol) and comparable, for example, to common UU base pairs.[11] There is a substantial degree of mutual compensation in the strength of the sugar−base and base−base contributions. Thus, the cis WC/SE family and subfamilies not only do follow the principles of isostericity, but they also are approximately isoenergetic. This makes them very attractive and versatile building blocks of complex RNA structures. They offer several distinct very well-defined shapes and, further, substantial stabilization energies.

Our calculated structures are in all aspects in quantitative agreement with isostericity matrices by Leontis et al.[4] They perfectly fit the suggested classification of the cis WC/SE structures into isosteric subfamilies made on the basis of crystal data. This, on the other hand, shows that the geometrical similarity of the isosteric base pairs is in fact an intrinsic property of these systems, which is transferable from the complete isolation to the RNA macromolecule. Model building and quantum chemical optimizations also provided stable cis WC/SE structures for the remaining three base pairs not evident in X-ray structures so far. These base pairs also have very good stabilization energies, hinting at the possibility that these interactions will likely be found in future structural studies of RNAs.

The cis WC/SE interactions are primarily of electrostatic nature, similar to standard base pairing. Nevertheless, the electron correlation component (and thus the dispersion attraction) of the interaction energy manifests itself to a modestly increased extent (by ca. 3 kcal/mol) in these systems, due to larger intermonomer contact area. Enhanced electron correlation stabilization should improve the stability of these interactions in a polar environment. The calculations reveal very good performance of the Cornell et al. empirical potential for the base−base and sugar−base interactions involving the hydroxyl group, which justifies performing large-scale MD simulations of complex RNA molecules with extended non-WC regions and interactions.[8]

Acknowledgment. This study was supported by the Wellcome Trust International Senior Research Fellowship in Biomedical Science in Central Europe GR067507 (J.Š., N.Š.) and grant 203/05/0009, Grant Agency of the Czech Republic (J.Š., J.E.Š., N.Š.). J.L., J.Š., and J.E.Š. acknowledge the financial support from NIH grant 3-S06 GM008047 31S1, NSF-CREST grant HRD-0318519, and ONR grant N00034-03-1-0116. Institute of Biophysics acknowledges support by the grant AVOZ50040507, Ministry of Education of the Czech Republic.

Supporting Information Available: Cartesian coordinates of all structures depicted in Figures 2, 3, and 5, together with the atomic charges used in molecular mechanical calculations. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Šponer, J.; Leszczynski, J.; Hobza, P. *J. Phys. Chem.* **1996**, *100*, 1965−1974.

(2) Hobza, P.; Šponer, J. *Chem. Rev.* **1999**, *99*, 3247−3276.

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

(4) (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.

(5) (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 (London)* **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.

(6) (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 (London)* **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.

(7) (a) Vallurupalli, P.; Moore, P. B. *J. Mol. Biol.* **2003**, *325*, 843−856. (b) Finger, L. D.; Trantírek, L.; Johansson, C.; Feigon, J. *Nucleic Acids Res.* **2003**, *31*, 6461−6472. (c) Nixon, P. L.; Rangan, A.; Kim, Y. G.; Rich, A.; Hoffman, D. W.; Hennig, M.; Giedroc, D. P. *J. Mol. Biol.* **2002**, *322*, 621−633. (d) Kolk, M. H.; van der Graaf, M.; Wijmenga, S. S.; Pleij, C. W. A.; Heus, H. A.; Hilbers, C. W. *Science* **1998**, *280*, 434−438.

(8) (a) Reblová, K.; Špačková, N.; Stefl, R.; Csaszar, K.; Koča, J.; Leontis, N. B.; Šponer, J. *Biophys. J.* **2003**, *84*, 3564−3582. (b) Auffinger, P.; Bielecki, L.; Westhof, E. *J. Mol. Biol.* **2004**, *335*, 555−571. (c) Csaszar, K.; Špačková, N.; Stefl, R.; Šponer, J.; Leontis, N. B. *J. Mol. Biol.* **2001**, *313*, 1073−1091. (d) Schneider, C.; Brandl, M.; Suhnel, J. *J. Mol. Biol.* **2001**, *305*, 659−667. (e) Zacharias, M. *Curr. Opin. Struct. Biol.* **2000**, *10*, 311−317. (f) Rážga, F.; Špačková, N.; Reblová, K.; Koča, J.; Leontis, N. B.; Šponer, J. *J. Biomol. Struct. Dyn.* **2004**, *22*, 183−194.

(9) (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.

(10) (a) Šponer, J.; Leszczynski, J.; Hobza, P. *J. Phys. Chem. A* **1997**, *101*, 9489−9495. (b) Meyer, M.; Suhnel, J. *J. Biomol. Struct. Dyn.* **1997**, *15*, 619−624. (c) Barsky, D.; Kool, E. T.; Colvin, M. E. *J. Biomol. Struct. Dyn.* **1999**, *16*, 1119−1134. (d) Podolyan, Y.; Rubin, Y. V.; Leszczynski, J. *J. Phys. Chem. A* **2000**, *104*, 9964−9970. (e) Šponer, J.; Leszczynski, J.; Vetterl, V.; Hobza, P. *J. Biomol. Struct. Dyn.* **1996**, *13*, 695−706. (f) Florian, J.; Šponer, J.; Warshel, A. *J. Phys. Chem. B* **1999**, *103*, 884−892. (g) Kratochvíl, M.; Šponer, J.; Hobza, P. *J. Am. Chem. Soc.* **2000**, *122*, 3495−3499. (i) Guerra, C. F.; Bickelhaupt, F. M.; Snijders, J. G.; Baerends, E. J. *Chem.−Eur. J.* **1999**, *5*, 3581−3594. (k) Brameld, K.; Dasgupta, S.; Goddard, W. A., III *J. Phys. Chem. B* **1997**, *101*, 4851−4859. (l) Sherer, E. C.; York, D. M.; Cramer, C. J. *J. Comput. Chem.* **2003**, *24*, 57−67. (m) Tsuzuki, S.; Uchimaru, T.; Matsumura, K.; Mikami, M.; Tanabe, K. *J. Chem. Phys.* **1999**, *110*, 11906−11910. (n) Florian, J.; Leszczynski, J. *J. Am. Chem. Soc.* **1996**, *118*, 3010−3017. (o) Bertran, J.; Oliva, A.; Rodríguez-Santiago, L.; Sodupe M. *J. Am. Chem. Soc.* **1998**, *120*, 8159−8167. (p) Colson, A.-O.; Sevilla, M. D. *Int. J. Radiat. Biol.* **1995**, *67*, 627−645. (q) Richardson, N. A.; Wesolowski, S. S.; Schaefer, H. F. *J. Phys. Chem. B* **2003**, *107*, 848−853. (r) Šponer, J.; Burda, J. V.; Sabat, M.; Leszczynski, J.; Hobza, P. *J. Phys. Chem. A* **1998**, *102*, 5951−5957. (s) Muñoz, J.; Gelpí, J. L.; Soler-López, M.; Subirana, J. A.; Orozco, M.; Luque, F. J. *J. Phys. Chem. B* **2002**, *106*, 8849−8857. (t) Moroni, F.; Famulari, A.; Raimondi, M.; Sabat, M. *J. Phys. Chem. B* **2003**, *107*, 4196−4202. (u) Guerra, C. F.; Bickelhaupt, F. M., Snijders, J. G.; Baerends, E. J. *J. Am. Chem. Soc.* **2000**, *122*, 4117−4128. (v) Gresh, N.; Šponer, J. *J. Phys. Chem. B* **1999**, *103*, 11415−11427. (x) Muñoz, J.; Šponer, J.; Hobza, P.; Orozco, M.; Luque, F. J. *J. Phys. Chem. B* **2001**, *105*, 6051−6060.

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

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

(13) 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. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.;

Ortiz, J. V.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B. G.; Chen, W.; Wong, M. W.; Andres, J. L.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98*; Gaussian, Inc.: Pittsburgh, PA, 1998.

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

(15) (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.

(16) (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.

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

(18) 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, 2002.

(19) 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.

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

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

(22) Šponer, J. E.; Leszczynski, J.; Glahe, F.; Lippert, B.; Šponer, J. *Inorg. Chem.* **2001**, *40*, 3269−3278.

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

(24) Luisi, B.; Orozco, M.; Šponer, J.; Luque, F. J.; Shakked, Z. *J. Mol. Biol.* **1998**, *279*, 1123−1136.

(25) For the computational algorithm, see, for example, ref 10r.

(26) In the only water-mediated member of the cis WC/SE family, the computed N9−C1′ distance is shorter by 0.57 Å than the crystallographically found value. This deviation is, in our opinion, not significant compared to the absolute value of the parameter (9.27 and 8.70 Å in crystal and gas-phase, respectively). It may stem from interaction with additional water molecules (not refined in the experiment), reflect data and refinement errors in the X-ray structure, or be a result of some other factors not included in the computations.

(27) (a) Biot, C.; Wintjens, R.; Rooman, M. *J. Am. Chem. Soc.* **2004**, *126*, 6220−6221. (b) Biot, C.; Buisine, E.; Kwasigroch, J. M.; Wintjens, R.; Rooman, M. *J. Biol. Chem.* **2002**, *277*, 40816−40822.