JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

Energies and 2′-Hydroxyl Group Orientations of RNA Backbone Conformations. Benchmark CCSD(T)/CBS Database, Electronic Analysis, and Assessment of DFT Methods and MD Simulations

Arnošt Mládek,*[,†] Pavel Banáš,[‡] Petr Jurečka,[‡] Michal Otyepka,[‡] Marie Zgarbová,[‡] and Jiří Šponer*[,†,§]

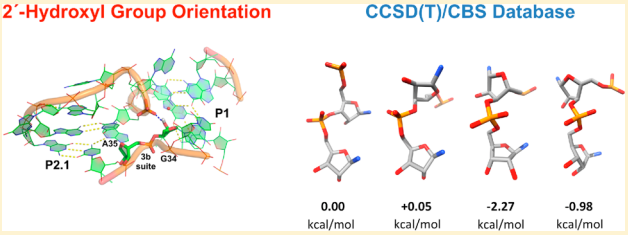
[†]Institute of Biophysics, Academy of Sciences of the Czech Republic, Královopolská 135, 612 65 Brno, Czech Republic

[‡]Regional Centre of Advanced Technologies and Materials, Department of Physical Chemistry, Faculty of Science, Palacky University, tr. 17. listopadu 12, 771 46, Olomouc, Czech Republic

[§]CEITEC, Central European Institute of Technology, Campus Bohunice, Kamenice 5, 625 00 Brno, Czech Republic

Supporting Information

ABSTRACT: Sugar−phosphate backbone is an electronically complex molecular segment imparting RNA molecules high flexibility and architectonic heterogeneity necessary for their biological functions. The structural variability of RNA molecules is amplified by the presence of the 2′-hydroxyl group, capable of forming multitude of intra- and intermolecular interactions. Bioinformatics studies based on X-ray structure database revealed that RNA backbone samples at least 46 substates known as rotameric families. The present study provides a comprehensive analysis of RNA backbone conformational preferences and 2′-hydroxyl group orientations. First, we create a benchmark database of estimated CCSD(T)/CBS relative energies of all rotameric families and test performance of dispersion-corrected DFT-D3 methods and molecular mechanics in vacuum and in continuum solvent. The performance of the DFT-D3 methods is in general quite satisfactory. The B-LYP-D3 method provides the best trade-off between accuracy and computational demands. B3-LYP-D3 slightly outperforms the new PW6B95-D3 and MPW1B95-D3 and is the second most accurate density functional of the study. The best agreement with CCSD(T)/CBS is provided by DSD-B-LYP-D3 double-hybrid functional, although its large-scale applications may be limited by high computational costs. Molecular mechanics does not reproduce the fine energy differences between the RNA backbone substates. We also demonstrate that the differences in the magnitude of the hyperconjugation effect do not correlate with the energy ranking of the backbone conformations. Further, we investigated the 2′-hydroxyl group orientation preferences. For all families, we conducted a QM and MM hydroxyl group rigid scan in gas phase and solvent. We then carried out set of explicit solvent MD simulations of folded RNAs and analyze 2′-hydroxyl group orientations of different backbone families in MD. The solvent energy profiles determined primarily by the sugar pucker match well with the distribution data derived from the simulations. The QM and MM energy profiles predict the same 2′-hydroxyl group orientation preferences. Finally, we demonstrate that the high energy of unfavorable and rarely sampled 2′-hydroxyl group orientations can be attributed to clashes between occupied orbitals.



INTRODUCTION

Nucleic acids are essential biomacromolecules that perform myriad of cellular tasks. While the dominant function of DNA is to preserve genetic information the pool of functions for which RNA is accountable is much larger and still not fully comprehended. A single polynucleotide chain is formed of two elements, a chemically monotonous and highly flexible sugar−phosphate backbone and an ordered succession of nucleobases attached to sugar rings. The ability of RNA molecules to execute miscellaneous functions has its origin in the tremendous variability of complexly organized structures, which are made possible by the 2′-hydroxyl group (O2′H group) attached to the sugar C2′ atom (Figure 1, left). The O2′H group features hydrogen bonding (H-bond) acceptor and donor capabilities, which allow it to interact and stabilize complex tertiary structures and modular motifs indispensable to RNA organization and inherently inaccessible to DNA molecules. RNA molecules create a wide variety of base pairs by systematically combining the three edges—Sugar edge, Watson−Crick edge, and Hoogsteen edge—of the constituent nucleotides.[1−3] Such intermolecular interactions define the shape and conservation patterns of folded RNAs.[4] The complexity of RNA interactions, however, is even greater than suggested by the standard RNA base-pairing classifications.[1−4] For example, many H-bonds of various kinds occur in concert with highly conserved base-phosphate H-bonds.[5,6] The O2′H group can also stabilize the local conformation by a H-bond with its own nucleotide or a subsequent same-strand residue. Regarding intramolecular interactions the preference of the O2′H group for particular H-

Received: September 23, 2013
Published: November 12, 2013

   dx.doi.org/10.1021/ct400837p | J. Chem. Theory Comput. 2014, 10, 463−480

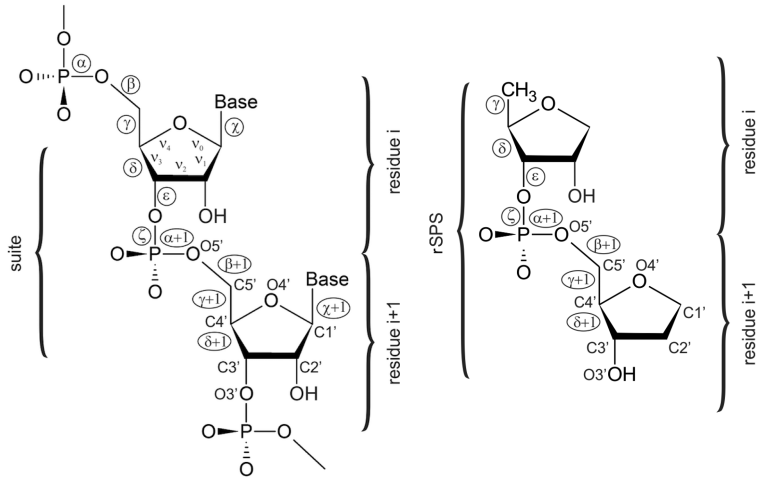


Figure 1. Dinucleotide unit showing the atom numbering and naming of backbone torsion angles. One nucleotide is defined from phosphate to phosphate. Conformation of each residue is given by six backbone torsion angles α...ζ and by the glycosidic torsion angle χ. Suite conformational unit (left) is centered around the phosphate group and is defined as the set of angles going from δ to $\delta + 1$, excluding χ torsion. Conformation of the sugar ring is defined by five internal torsions ν_0...ν_4. The rSPS model system used in the present work (right).

bond acceptors, such as the phosphodiester bridging oxygens (O3′(i) and O5′(i+1), where "i" denotes the residue number in the 5′ → 3′ direction), the anionic phosphate oxygens (O1P and O2P of the (i+1) residue), or the adjacent sugar ring oxygen (O4′(i+1)), is strongly modulated by the sugar−phosphate backbone conformation and vice versa. The fact that the presence of the O2′H group has an impact on both the backbone conformation and the sugar pucker makes description of the RNA backbone potential energy surface (PES) a more challenging task than description of the DNA's one. Orientation of 2′-hydroxyl and the local backbone geometry are tightly coupled and cannot be investigated separately.

Whereas QM studies of base stacking and base pairing are abundant[7−36] there are few QM studies that deal with the conformational properties of the sugar−phosphate backbone,[37−53] mostly considering the more simple DNA backbone. However, QM studies of the sugar−phosphate backbone are needed, since its adequate description is a notorious weakness of contemporary molecular simulation force fields.[54−59] Subtle imbalances in the description of the backbone can lead to an entire degradation of the simulated nucleic acids structure.[54,55] We also suggest that proper description of nucleic acids backbone will be a decisive task for success of future polarization force fields and approximate QM methods designed for QM MD studies of nucleic acids. Description of the backbone is also challenging experimentally. The sugar atoms are much less visible than the nucleobases and phosphates in crystallographically derived electron densities while flexibility of the backbone often precludes its unambiguous refinement.[52,60,61]

We investigate intrinsic stabilities of 46 experimentally determined RNA backbone conformations via high-level QM methods both in gas phase and solvent environment, complementing similar DNA study.[53] We first create a benchmark database of relative stabilities of the backbone rotamers using estimated complete basis set CCSD(T) computations (dubbed as CBS(T)). These calculations are then compared with a wide spectrum of state-of-the-art density functionals (DF) and with AMBER MM force field. We demonstrate that the conformation-dependent hyperconjugation degree of backbone substates is not decisive as its stabilizing effect does not correlate with calculated relative energies. We also derive the O2′H group rotation energy profiles and compare them with the force field data. Potential intramolecular H-bond acceptors are determined and the strength of the most stable H-bonds is assessed by an electronic structure analysis. The O2′H group rotation profiles correlate well with the 2′-hydroxyl group orientation distributions from molecular dynamics (MD) simulations, indicating good sampling of the O2′H group in the simulations.

RNA Structure Overview. The conformational space of RNA backbone is defined by six torsions α, β, γ, δ, ε, and ζ (Figure 1, left).[62] The α torsion is defined as O3′(i−1)-P-O5′-C5′, β as P-O5′-C5′-C4′, γ as O5′-C5′-C4′-C3′, δ as C5′-C4′-C3′-O3′, ε as C4′-C3′-O3′-P(i+1), and ζ as C3′-O3′-P(i+1)-O5′(i+1), where (i+1) denotes atoms of the subsequent nucleotide. It is customary to describe the backbone torsional angles of ∼60° as *plus* (p), of ∼300° (−60°) as *minus* (m), and of ∼180° as *trans* (t). The RNA conformation is further affected by ribose puckering conveniently described by the phase (pseudorotation) angle P, which is defined in terms of five internal ribose torsions ν_0...ν_4.[62,63] The sugar conformations are commonly classified using suffixes X′-endo and -exo, where X′ is one of the sugar ring embedded atoms. The former indicates that the given atom is displaced on the same side of the ring as the C5′ atom, while the latter means that the given atom is displaced on the opposite side. C3′-endo (P ∼ 18°) is the canonical RNA pucker while C2′-endo (P ∼ 162°) pucker occurs in some noncanonical RNA conformations. RNA molecules are precisely structured and combine short canonical helices with endlessly variable noncanonical features starting from single noncanonical base pairs or single bulges up to extremely complex and hierarchically organized molecular building blocks.[3,64]

METHODS

Derivation of RNA Backbone Conformations. RNA backbone adopts multiple conformations (Figure 2). The earlier attempts to classify local conformation space of RNA backbone[65−68] were unified by the current RNA backbone nomenclature.[61] This classification utilizes the concept of **suite** (Figure 1, left) and categorizes RNA backbone conformations into 46 families,[61] each one being tagged by a two-character label. A suite is a structural subset of a dinucleotide going from sugar to

dx.doi.org/10.1021/ct400837p | J. Chem. Theory Comput. 2014, 10, 463−480

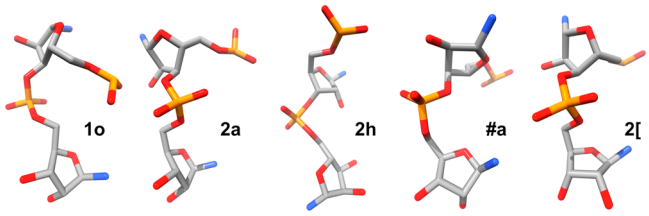


Figure 2. Examples of RNA backbone structural variability, according to Table 1 of ref 61. The N9/N1 nucleobase atoms are blue.

sugar, containing thus only seven backbone torsion angles (δ...δ + 1) and consists of two heminucleotides, the 5′ one is described by δ, ε, ζ torsions, while the 3′ heminucleotide is defined by α + 1, β + 1, γ + 1, and δ + 1 torsion angles. Each family is thus characterized by a seven-dimensional vector consisting of average values of its backbone torsions and a unique two-character cluster name, see Table 1 and ref 61.

The complex two-character label nomenclature reflects complexity of the RNA backbone conformational space.[61] The first character of the suite label describes the 5′ heminucleotide, the second character the 3′ one. For example if the 5′ sugar adopts C3′-endo ($\delta \sim 80°$) conformation the first character would be an odd number or "&", whereas if the pucker is close to C2′-endo ($\delta \sim 150°$) the first character would be "#" or an even number. The number itself refers to the values of remaining ε and ζ torsions. Similar approach is used for defining the 3′ heminucleotide character.[61] Since we are interested in the pucker-type dependent orientation of the 2′-hydroxyl group of the 5′ sugar, we divided the 46 families into four groups according to the 5′/ 3′-sugar conformations (given atoms are endo), namely group I, C3′/C3′; group II, C3′/C2′; group III, C2′/C3′; and group IV, C2′/C2′. For each conformational family an exemplar dinucleotide is quoted by the respective structure NDB code and the "i+1" nucleotide ID number, see Table 1 of ref 61.

RNA Backbone Model System. In our earlier DNA studies,[37,38] we used a model system abbreviated as SPSOM, that is, sugar−phosphate−sugar capped on both ends with methoxy groups. Despite being electronically complete, the SPSOM model suffers from spurious CH···O interactions biasing intrinsic energies. Therefore, we suggested a smaller sugar-to-sugar model (SPS) by replacing the terminal methoxy groups with hydrogen atom and a hydroxyl group at the 5′ and 3′ ends, respectively.[53] The SPS model system is simple enough to be treatable by high-level QM techniques and still sufficiently complete to mimic DNA backbone we slightly modified the SPS model by appending O2′H group to the C2′ of the 5′ sugar to obtain RNA backbone model system. This new model dubbed as rSPS (Figure 1, right), however, is not a fully complete segment of RNA backbone as it lacks the 2′-hydroxyl at the 3′ end. We have nevertheless decided to exclude the second hydroxyl from the model system, as it would insurmountably complicate the computations. The stabilization due to the O2′H···O3′ H-bond at the 3′ terminus, which cannot be easily separated from the studied conformational energetics, is modulated by the sugar conformation which differs for different families. This would bias the energy ranking. Besides if there had been two hydroxyl groups at the 3′ end, O3′ might have served as a H-bond donor, which would be unnatural. It can be argued that ribose and 2′-deoxyribose have different conformational preferences, but the δ + 1 torsion coupled with sugar pucker was constrained during geometry minimizations (vide infra) at the tabulated values (Table 1) and so the respective RNA pucker was used in all computations. Evaluation of the effect of neglecting the 3′-end 2′-hydroxyl group for selected systems with C2′- and C3′-endo puckers shows that the error is rather negligible.

For each class, an exemplar dinucleotide with values of backbone torsion angles as close as possible to the average values was identified.[61] Subsequently, these 46 representative dinucleotides have been pruned to obtain our rSPS model systems and all backbone torsions were adjusted to the values given in Table 1. The coordinates were then used for consecutive constrained geometry optimizations.

Geometrical Optimizations and O2′H Group Scans. Geometrical optimizations of the structures were done separately in gas and solvent environments using TPSS[69] functional complemented with D3 dispersion correction with Becke-Johnson damping,[70,71] dubbed hereafter as TPSS-D3, and the large def2-TZVPP basis set[72,73] augmented with diffuse functions[74] (def2-TZVPPD) important for weakly bound electrons in negatively charged molecules. Large flexible basis sets with diffuse functions, especially in combination with nonhybrid DFs, might result in positive HOMO orbital energies, which would imply unbound electrons.[75−80] Thus, converged wave functions were manually validated. We employed the resolution of the identity (RI) approximation.[81,82] The exchange-correlation contribution was integrated using dense m4 numerical quadrature grid. Convergence threshold of SCF has been tightened to 10^{-7} a.u. while the remaining parameters were kept at their defaults. To take advantage of the robust Berny optimization algorithm[83] implemented in Gaussian03[84] package and the superior scalability of the Turbomole 6.3 code,[85] we used a scheme in which electronic energies and gradients calculated by Turbomole are passed to Gaussian to execute geometry modification. The new coordinates are then passed back and serve as a new input for the next SCF cycle. COSMO continuum solvation model[86,87] was used with a dielectric constant of 78.5. The atomic van der Waals radii were taken as their defaults in Turbomole 6.3. Note, that there are many other sophisticated solvent models and cavity construction methods (including parametrized COSMO-RS), which might provide more accurate solvent response, solvent/solute boundary and solvation energies. We wanted, however, to be consistent with our previous DNA backbone article,[53] and thus, we utilized the same scheme for optimization,that is, TPSS-D3/def2-TZVPPD + COSMO solvent. Nevertheless, since we are primarily interested in gas phase properties of the studied systems, we assume that the COSMO model suffices our purposes.

To preserve the experimentally determined backbone substates, constraints on all backbone torsion angles (δ...δ + 1) was imposed, which is necessary to maintain the desired rotameric family.[37,38,52] Another issue is the initially unknown orientation of the hydroxyl group. Therefore we proceeded as follows. At the first stage, we preoptimized experimental structures with backbone torsions fixed at their tabulated values (Table 1) and the O2′H and O3′H(i+1) orientations, defined hereafter by C3′-C2′-O2′-O2′H and C4′(i+1)-C3′(i+1)-O3′(i+1)-O3′H(i+1) torsions, respectively, were set to 180° and kept frozen. Then, we performed a rigid PES scan in which the C3′-C2′-O2′-O2′H dihedral was rotated by 360° with an increment of 5°. Since we are interested in rough profile shape rather than exact magnitudes of energy barriers and minima depths, and since we want to avoid geometry changes along the scan, the rigid scan is justified. The conformation corresponding to the global O2′H minimum was then reoptimized without O2′H and O3′H

dx.doi.org/10.1021/ct400837p | J. Chem. Theory Comput. 2014, 10, 463−480

Table 1. 46 RNA Backbone Conformations Grouped (I–IV) According to Sugar Conformations Coupled to the δ and δ+1 Torsions[a]

$\delta, \delta+1, \gamma+1$[b]	group	label	δ	ε	ζ	$\alpha+1$	$\beta+1$	$\gamma+1$	$\delta+1$
3, 3, p	I	1a	81	−148	−71	−65	174	54	81
		1m	84	−142	−68	−68	−138	58	86
		1L	86	−115	−92	−56	138	62	79
		&a	82	−169	−95	−64	−178	51	82
		7a	83	−143	−138	−57	161	49	82
		3a	85	−144	173	−71	164	46	85
		9a	83	−150	121	−71	157	49	81
		1g	81	−141	−69	167	160	51	85
		7d	84	−121	−103	70	170	53	85
		3d	85	−116	−156	66	−179	55	86
		5d	80	−158	63	68	143	50	83
3, 3, t		1e	81	−159	−79	−111	83	168	86
		1c	80	−163	−69	153	−166	179	84
		1f	81	−157	−66	172	139	176	84
		5j	87	−136	80	67	109	176	84
3, 2, p	II	1b	84	−145	−71	−60	177	58	145
		1[	83	−140	−71	−63	−138	54	144
		3b	85	−134	168	−67	178	49	148
		1z	83	−154	−82	−164	162	51	145
		5z	83	−154	53	164	148	50	148
		7p	84	−123	−140	68	−160	54	146
3, 2, t		1t	81	−161	−71	180	−165	178	147
		5q	82	−155	69	63	115	176	146
3, 2, m		1o	84	−143	−73	−63	−135	−66	151
		7r	85	−127	−112	63	−178	−64	150
2, 3, p	III	2a	145	−100	−71	−72	−167	53	84
		4a	146	−100	170	−62	170	51	84
		0a	149	−137	139	−75	158	48	84
		#a	148	−168	146	−71	151	42	85
		4g	148	−103	165	−155	165	49	83
		6g	145	−97	80	−156	−170	58	85
		8d	149	−89	−119	62	176	54	87
		4d	150	−110	−172	80	−162	61	89
		6d	147	−119	89	59	161	52	83
2, 3, t		2h	148	−99	−70	−64	177	176	87
		4n	144	−133	−156	74	−143	−166	81
		0i	149	−85	100	81	−112	−178	83
		6n	150	−92	85	64	−169	177	86
		6j	142	−116	66	72	122	−178	84
2, 2, p	IV	2[	146	−101	−69	−68	−150	54	148
		4b	145	−115	163	−66	172	46	146
		0b	148	−112	112	−85	165	57	146
		4p	150	−100	−146	72	−152	57	148
		6p	146	−102	90	68	173	56	148
2, 2, t		4s	150	−112	170	−82	84	176	148
2, 2, m		2o	147	−104	−64	−73	−165	−66	150

[a]Average values of the individual backbone torsions (deg) along with symbolic names of the classes (Label)[61] are given. The conformations are ordered according to ref 61. [b]The two numbers in the first column indicate a pucker type of the 5′ and 3′ sugar moieties, respectively. The third character of the first column refers to γ + 1 torsion, m signifies −60° (*minus*); t, 180° (*trans*); and p, +60° (*plus*).

orientation constraints, so that each structure is derived with its optimal hydroxyl group orientations. The final set of 46 conformations optimized in gas and condensed phases served as input structures for subsequent single point calculations and electronic structure analyses.

Single Point Calculations. Energies of the optimized backbone conformers were calculated using several QM methods and sets of basis functions, both in gas phase and COSMO solvent with the exception of ΔCCSD(T) corrections, which were evaluated in gas phase only.

MP2/CBS calculations were carried out using the Halkier et al. extrapolation scheme[88,89] with augmented Dunning's basis sets[90,91] (aug-cc-pVXZ, where X = D/T/Q) to obtain MP2 energies at the complete basis set limit (CBS). Even though extrapolation to CBS is an effective way how to eliminate both intramolecular basis set superposition (BSSE) and incompleteness (BSIE) errors, still some residual error depreciating the

dx.doi.org/10.1021/ct400837p | J. Chem. Theory Comput. 2014, 10, 463−480

results remain.[92,93] The Hartree−Fock (HF) and the MP2 correlation components are evaluated separately solving the following equations:

$$E_X^{\mathrm{HF}} = E_{\mathrm{CBS}}^{\mathrm{HF}} + A\cdot \mathrm{e}^{-\delta X}$$

$$E_X^{\mathrm{Corr}} = E_{\mathrm{CBS}}^{\mathrm{Corr}} + B\cdot X^{-3}$$

where X is the cardinal angular momentum quantum number of the respective basis set ($X = 2$ for D, $X = 3$ for T, and $X = 4$ for Q) and δ, A and B are extrapolation parameters. While $\delta = 1.43$ and 1.54 for (D,T) and (T,Q) extrapolations, respectively, was taken from the literature,[89] A and B are unique coefficients to be determined by solving the given equations. $E_{\mathrm{CBS}}^{\mathrm{Corr}}$ and $E_{\mathrm{CBS}}^{\mathrm{HF}}$ are the correlation and HF components of the electronic energy at the CBS limit. Analogously, E_X^{Corr} and E_X^{HF} are corresponding terms obtained using the aug-cc-pVXZ basis set.

Our earlier studies on DNA backbone suggests that the ΔCCSD(T) corrections might be omitted for backbone calculations.[38,53] However, the RNA backbone is much more conformationally diverse and includes various intramolecular H-bonds from 2′-hydroxyl group. Thus, in order to provide reliable benchmark data, we augmented the extrapolated MP2(T,Q) energies by ΔCCSD(T)/aug-cc-pVDZ correction for all points. We abbreviate this approximate CCSD(T)/CBS calculation as CBS(T) to indicate that CBS extrapolation has been carried out at the MP2 level of theory only. Since solvation does not significantly alter CCSD(T) amplitudes, solvent CBS(T) energies were estimated using MP2(T,Q) calculated in COSMO and gas phase ΔCCSD(T) corrections.

Performance of three GGA, two meta-GGA, three hybrid, and two double-hybrid DF augmented by D3(BJ) term were evaluated. The GGA level of theory comprises BLYP,[94,95] PBE,[96] and its reparameterized variant revPBE.[97] From the meta-GGA group of DFs, TPSS[69] along with its recalibrated version oTPSS[98] were tested. Hybrid DFs, which take distinct portion of HF exchange into account, are represented by B3LYP[95,99] and two thermochemistry oriented functionals, PW6B95[100] and MPW1B95.[101] The rated double-hybrids are PWPB95[102] and DSD-B-LYP.[103] The former DF[102] contains a mixture of reparameterized Perdew−Wang and Fock (50%) exchange, Becke95 correlation, and 26.9% of spin-opposite scaled perturbative correlation (SOS-PT2). The later[103] is identical to B2PLYP[104] and adds a spin-component-scaled perturbative contribution (SCS-PT2). Both SOS- and SCS-PT2 corrections were calculated without frozen core approximation. Except for PBE, TPSS, and B3LYP considered primarily due to their widespread usage, the listed DFs were selected due to their outstanding score at the respective Jacob's ladder rung in a thorough benchmark study.[105] With the exception of double-hybrids where aug-cc-pVQZ basis set was used for both SCF and PT2 runs, all single point calculations were carried out with the def2-TZVPPD basis set. RI approximation with an appropriate set of auxiliary basis functions[81,82] was employed for all (meta-)GGAs. While exchange-correlation term of (meta-)GGAs was integrated on m4 grid, for hybrids and double-hybrids a denser m5 grid has been used. SCF convergence criteria were set to 10^{-7} a.u., and the remaining parameters were kept at recommended values. All calculations were carried out either with a modified version of Turbomole 5.9 or with Turbomole 6.3.[85]

AIM Analysis of O2′H H-Bonding. Wave functions of the optimized conformers were investigated with an atoms-in-molecules (AIM) analysis[106−108] to reveal and compare the stabilizing effect of the conformationally specific H-bonds involving O2′H group. AIM analyzes the local electron density curvatures and finds critical points (CPs), which can provide information on the intramolecular H-bond network. Thus, an existence of a CP between O2′H and an acceptor oxygen gives a proof of the respective H-bond, while its local characteristics, the electron density (ρ) and its Laplacian ($\nabla^2\rho$), measure the interaction strength. The topologies of the charge densities were computed using the converged B3-LYP/6-31+G** wave functions (note that the empirical dispersion correction does not affect the wave function). The Cartesian 6d and 10f basis functions were substituted for the standard 5d and 7f functions. AIM analysis was done using AIMPAC code.[109,110] For recent discussion of the limitation of the electron topology analyses, see ref 111.

Natural Bond Orbital Analysis. The natural bond orbital (NBO) analysis[112−116] is an electronic structure method, which aims to deepen understanding of the physical principles driving (de)stabilization of the conformers. A great advantage of NBO analysis is that it projects the multicentered orbitals into localized bond orbitals, which are more compatible with the chemical view. In addition, it provides a quantitative estimate of the delocalization effects between localized orbitals. The NBO analysis was applied to compare the extent and stabilizing effect of the electron delocalization in the studied systems. The off-diagonal Fock matrix elements of the optimized conformers were analyzed in gas phase at the B3-LYP/6-311++G(3df, 3pd) level. Further, in order to link the high-energy O2′H group orientations detected in rotation profiles with repulsive orbital interactions, optimized ribose NBO orbitals in C3′- and C2′-endo conformations and their overlaps for different O2′H group directions were calculated and visually inspected at the B3-LYP/6-31+G** level.

Molecular Mechanics Calculations. MM calculations were performed with the aim to compare the benchmark QM data with common simulation force field, including both stabilities of different backbone families and correctness of positioning of the O2′H group. The QM optimized model systems representing all 46 RNA backbone conformations were reoptimized by MM. The minimizations were performed with default ff99bsc0[55,117] parameters (note that our model systems do not include the χ angles, so the χ_{OL3} correction[57] is not applicable to this particular set of calculations, vide infra). Point charges were taken from the Cornell et al. force field;[117] charges on deoxyribose are taken from a deoxyribonucleotide, and the remaining charges are from the ribonucleotide. Charges on hydrogen atoms of the capping methyl group and hydrogen atoms on the C1′ atoms were chosen so as to fix integral charge of the molecule. See Supporting Information (SI) for the list of used point charges. The Gaussian03 program was used as optimization driver with the "external" function and an in-house script linking Gaussian to the sander module of AMBER 11.[118] Optimizations were performed in the gas phase and Poisson−Boltzmann (PB)[119,120] continuum solvent. The grid spacing was set to 0.1 Å, default radii[119,120] and dielectric constant of 78.5 were used. Constraints on seven consecutive backbone torsion angles were applied to keep the desired suite conformation. One additional dihedral angle of the sugar pucker (C4′-O4′-C1′-C2′) was constrained to maintain value of the pseudorotation angle. Orientations of both hydroxyl groups (C3′-C2′-O2′-O2′H and C4′-C3′-O3′-O3′H) were constrained at the QM optimal values so that MM and QM single point energies are comparable. Optimized geometries were subsequently used for rigid scans of

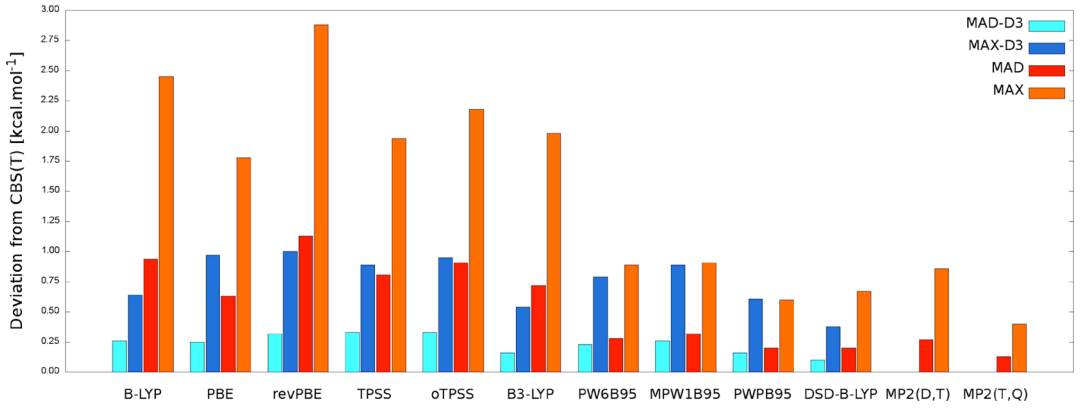


Figure 3. Deviations of the investigated QM methods compared to the reference CBS(T) in gas phase. MAD (cyan/red) and MAX (blue/orange) represent the mean absolute deviation and the maximum absolute deviation, respectively. Performance of density functionals is outlined both with (cyan and blue) and without (red and orange) D3 correction. All tested methods have MAX below 1.0 kcal·mol^{-1}, which can be considered as quite excellent performance.

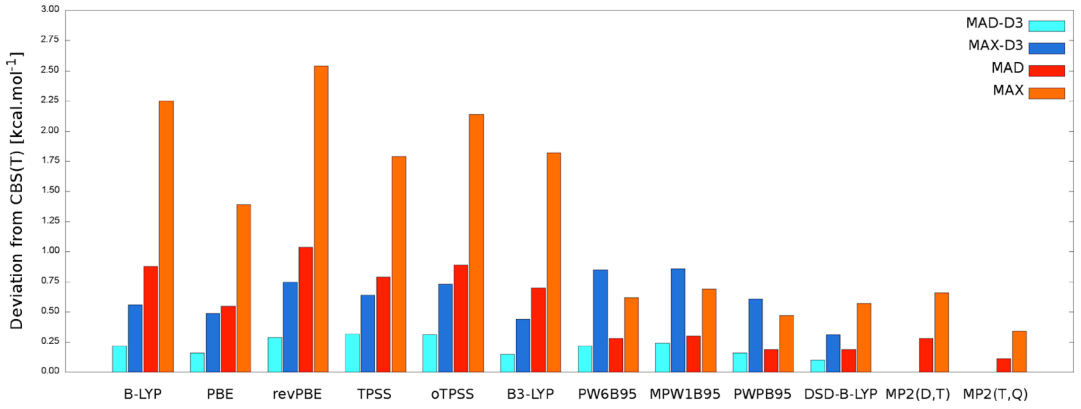


Figure 4. Deviations of the investigated QM methods compared to the reference CBS(T) in COSMO solvent. MAD (cyan/red) and MAX (blue/orange) represent the mean absolute deviation and the maximum absolute deviation, respectively. Performance of density functionals is outlined both with (cyan and blue) and without (red and orange) D3 correction.

the C3′-C2′-O2′-O2′H angle with 5° increment. At each step both vacuum and solvation single point energies were calculated.

Molecular Dynamics Simulations. Molecular simulations of folded RNAs were used to evaluate positioning of the O2′H groups. We have monitored occurrence of different RNA backbone suites in the course of the simulations and evaluated distribution of O2′H groups for each conformational family that is captured by the simulations with statistically significant occurrence.

The classical MD simulations of sarcin-ricin loop domain (SRL) and *glmS* riboswitch were carried out using the AMBER 11[118] package with the ff99bsc0χ_{OL3}[55,57,117,121] force field. The ff99bsc0χ_{OL3} force field is based on AMBER ff99[117] corrected by Barcelona[55] α/γ and Olomouc[57] χ_{OL3} reparameterizations. The ff99bsc0χ_{OL3} force field is currently RNA part of the standard AMBER force field suites (ff10 and ff12SB) and in contrast to the earlier ff94, ff99, and ff99bsc0 versions, the χ_{OL3} correction substantially stabilizes the RNA simulations by preventing formation of spurious ladder-like structures.[54,57] The crystal structure of SRL (Sarcin Ricin Loop) domain from *E. coli* 23S rRNA (PDB ID 3DVZ, resolution of 1.0 Å) and native structure of *glmS* riboswitch precursor from *T. Tengcongensis* (2HO7, 2.9 Å) were used as starting geometries. The native glucosamine-6-phosphate ligand was modeled on the basis of inhibited structure of *glmS* riboswitch from *B. Antrancis* (2NZ4, 2.5 Å; see ref 122 for more details). Systems were immersed into a rectangular box with at least 10 Å between the solute, the box wall and solvated using TIP3P water model[123] and neutralized by net-neutral sodium ions (SRL: r = 1.8680 Å, ε = 0.0028 kcal·mol^{-1}. *glmS*: r = 1.369 Å, ε = 0.0874 kcal·mol^{-1})[124,125] yielding to cation concentration of ∼0.18 M and ∼0.42 M for SRL and *glmS*, respectively. The systems were simulated for 1 μs using the [NpT] ensamble at 298.16 K (Berendsen thermostat[126]). For more details see Supporting Information.

The analysis of the population of C3′-C2′-O2′-O2′H orientation as a function of suite name was performed using an in-house program. The orientation of C3′-C2′-O2′-O2′H torsion was measured for each potential suite in each snapshot of the simulations (snapshots were stored every 10 ps; thus, in total, we analyzed 100 000 snapshots per simulation per suite). Simultaneously, the suite name of a given suite was identified by suitename program.[61] Subsequently, the data were sorted on the basis of suite names and when more than 10 000 of data from both SRL and *glmS* simulations were collected for a given suite, its normalized histogram of O2′H group orientation was calculated. The analysis utilizes ∼89% of the individual suite points; the remaining ones could not be classified.

RESULTS AND DISCUSSION

Benchmark Database of Relative Energies of RNA Sugar−Phosphate Backbone. We have first created a benchmark database of relative energies of 46 backbone

dx.doi.org/10.1021/ct400837p | J. Chem. Theory Comput. 2014, 10, 463−480

Table 2. List of Gas Phase Relative Energies (kcal·mol^{-1}) Referenced Against the 1a Conformation[a]

	GGA			meta-GGA		hybrid			double-hybrid		MP2/CBS			
label	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	CBS(T)	MM
1a	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**0.00**	0.00
1m	−4.02	−4.07	−4.31	−4.18	−4.30	−4.07	−4.13	−4.23	−4.19	−4.04	−4.01	−4.08	**−4.09**	−1.31
1L	2.24	2.34	2.48	2.48	2.51	2.17	2.61	2.70	2.69	2.47	2.71	2.79	**2.60**	3.05
&a	−1.49	−1.25	−1.21	−1.14	−1.01	−1.79	−1.77	−1.73	−1.89	−2.03	−1.85	−1.86	**−1.95**	−2.01
7a	1.03	1.18	1.14	1.17	1.14	0.90	1.21	1.31	1.27	1.07	1.51	1.28	**1.20**	2.94
3a	0.42	0.50	0.39	0.43	0.41	0.40	0.74	0.80	0.75	0.57	1.11	0.74	**0.64**	1.69
9a	0.32	0.27	0.28	0.33	0.32	0.26	0.50	0.53	0.56	0.46	0.81	0.65	**0.56**	2.38
1g	0.51	0.48	0.37	0.58	0.57	0.65	0.81	0.86	0.85	0.84	1.11	1.04	**0.89**	3.57
7d	−0.77	−1.14	−1.04	−1.06	−1.06	−0.39	−0.44	−0.55	−0.32	−0.09	−0.12	−0.13	**−0.17**	3.25
3d	0.46	0.33	0.42	0.31	0.33	0.68	0.99	0.99	1.09	1.05	1.41	1.18	**1.04**	2.85
5d	0.46	0.44	0.25	0.51	0.46	0.54	0.69	0.73	0.65	0.62	0.95	0.71	**0.60**	5.25
1e	1.77	2.09	2.36	2.21	2.61	1.97	2.71	2.82	2.54	1.95	2.19	2.11	**1.92**	6.22
1c	1.05	1.10	1.45	1.44	1.85	1.36	1.83	1.92	1.72	1.39	1.74	1.52	**1.22**	4.39
1f	2.01	1.97	2.07	2.29	2.57	2.32	2.85	2.92	2.70	2.39	3.01	2.55	**2.15**	6.29
5j	4.10	3.97	3.91	4.07	4.24	4.69	5.09	5.22	5.08	5.00	5.49	5.10	**4.75**	11.05
1b	−1.04	−0.69	−0.68	−1.03	−0.87	−0.84	−0.78	−0.75	−0.86	−0.93	−1.16	−1.04	**−0.89**	2.13
1[	−3.39	−3.06	−3.30	−3.44	−3.36	−3.21	−3.24	−3.26	−3.35	−3.24	−3.27	−3.39	**−3.30**	1.06
3b	−0.98	−0.67	−0.63	−0.97	−0.81	−0.83	−0.58	−0.51	−0.60	−0.76	−0.47	−0.76	**−0.71**	3.19
1z	−0.76	−0.60	−0.39	−0.61	−0.36	−0.49	−0.24	−0.17	−0.29	−0.50	−0.49	−0.52	**−0.43**	3.71
5z	−0.94	−0.64	−0.57	−0.82	−0.41	−0.79	−0.71	−0.77	−0.87	−0.84	−0.93	−0.62	**−0.90**	6.30
7p	−0.93	−0.68	−0.70	−1.06	−0.93	−0.60	−0.49	−0.44	−0.48	−0.40	−0.38	−0.45	**−0.41**	4.73
1t	2.33	2.47	2.59	2.60	2.93	2.56	3.02	3.10	2.87	2.50	2.82	2.54	**2.45**	7.64
5q	3.04	2.99	2.67	2.93	3.10	3.44	3.52	3.60	3.39	3.46	3.65	3.32	**3.19**	10.55
1o	−0.04	0.09	−0.07	−0.21	−0.06	0.32	0.37	0.44	0.25	0.25	0.23	0.08	**0.11**	3.82
7r	0.42	0.54	0.60	0.17	0.36	0.76	0.88	0.91	0.89	0.89	1.00	0.79	**0.77**	2.76
2a	−0.80	−0.79	−1.01	−1.05	−1.03	−0.84	−0.68	−0.71	−0.76	−0.84	−0.62	−0.84	**−0.98**	−0.01
4a	1.19	1.27	1.01	1.04	1.03	1.07	1.24	1.23	1.17	1.07	1.36	1.11	**1.06**	3.29
0a	0.72	0.53	0.48	0.27	0.30	0.49	0.42	0.34	0.44	0.53	0.87	0.65	**0.52**	6.57
#a	0.28	−0.02	0.07	−0.32	−0.16	0.00	0.30	0.10	0.19	0.04	0.43	0.23	**0.05**	5.51
4g	3.57	3.59	3.34	3.45	3.43	3.46	3.68	3.68	3.62	3.54	4.00	3.62	**3.62**	5.43
6g	−0.29	−0.22	−0.42	−0.51	−0.52	−0.33	−0.07	−0.06	−0.06	−0.21	0.09	−0.15	**−0.13**	1.32
8d	1.53	1.51	1.33	1.16	1.24	1.67	1.84	1.80	1.79	1.81	1.94	1.78	**1.69**	4.41
4d	1.49	1.47	1.07	1.27	1.25	1.60	1.68	1.72	1.62	1.75	1.92	1.81	**1.63**	5.20
6d	1.46	1.30	1.10	1.23	1.18	1.71	1.95	1.95	2.00	1.97	2.26	2.10	**1.99**	9.47
2h	2.98	2.76	2.56	2.68	2.90	3.30	3.47	3.49	3.38	3.43	3.65	3.42	**3.10**	6.70
4n	6.95	6.40	6.41	6.06	6.16	6.71	6.16	6.01	6.16	6.67	6.86	6.84	**6.51**	15.70
0i	5.84	5.93	5.92	5.80	6.00	5.92	6.14	6.23	6.15	6.10	6.27	6.32	**6.10**	12.14
6n	3.25	3.30	3.13	3.11	3.21	3.47	3.80	3.83	3.73	3.52	3.53	3.57	**3.52**	7.37
6j	1.45	1.13	0.76	1.20	1.26	1.98	2.18	2.16	2.06	2.14	2.40	2.05	**1.76**	8.45
2[	−2.13	−1.73	−2.00	−2.28	−2.08	−1.99	−1.73	−1.72	−1.95	−2.12	−2.12	−2.28	**−2.27**	2.70
4b	−0.08	0.27	0.07	−0.16	0.00	−0.06	0.10	0.15	−0.02	−0.17	−0.07	−0.22	**−0.17**	5.82
0b	0.47	0.84	0.71	0.43	0.60	0.64	1.05	1.09	1.02	0.82	1.03	0.84	**0.88**	8.14
4p	0.52	0.94	0.67	0.41	0.62	0.90	1.13	1.17	1.04	1.07	1.04	1.01	**1.00**	7.46
6p	−1.48	−1.18	−1.29	−1.62	−1.47	−1.20	−1.01	−0.98	−1.04	−1.15	−1.16	−1.15	**−1.04**	6.95
4s	5.29	5.40	4.87	5.14	5.26	5.66	6.00	6.03	5.83	5.73	5.75	5.65	**5.56**	12.89
2o	−0.41	−0.26	−0.42	−0.73	−0.44	−0.12	0.03	0.11	−0.22	−0.35	−0.51	−0.62	**−0.66**	2.86
MAD-D3	**0.26**	**0.25**	**0.32**	**0.33**	**0.33**	**0.16**	**0.23**	**0.26**	**0.16**	**0.10**	**N/A**		**N/A**	**N/A**
MAX-D3	**0.64**	**0.97**	**1.00**	**0.89**	**0.95**	**0.54**	**0.79**	**0.89**	**0.61**	**0.38**				
MAD	**0.94**	**0.63**	**1.13**	**0.81**	**0.91**	**0.72**	**0.28**	**0.32**	**0.20**	**0.20**	**0.27**	**0.13**	**0.00**	**4.20**
MAX	**2.45**	**1.78**	**2.88**	**1.94**	**2.18**	**1.98**	**0.89**	**0.91**	**0.60**	**0.67**	**0.86**	**0.40**		**9.19**

[a]Applied methods are tagged as follows: M1, B-LYP; M2, PBE; M3, revPBE; M4, TPSS; M5, oTPSS; M6, B3-LYP; M7, PW6B95; M8, MPW1B95; M9, PWPB95; M10, DSD-B-LYP; M11, MP2(D,T); and M12, MP2(T,Q). For further details see Methods. MAD and MAX (both in kcal·mol^{-1}) represent the mean absolute deviation and the maximum absolute deviation from the reference CBS(T) energies. MM labels force field calculations. Listed DFT energies include D3 correction; MAD and MAX values are quoted for DFs with and without D3 term.

conformational families. The energies are derived at CBS(T) level, well-established standard for benchmark calculations.[52,53] Relative energies are calculated with respect to the canonical 1a conformation (canonical A-RNA backbone substate). As the backbone torsions are fixed at their experimental values (Table 1), there are only marginal structural differences between the

dx.doi.org/10.1021/ct400837p | J. Chem. Theory Comput. 2014, 10, 463−480

experimental[61] and the QM-optimized geometries. The single point calculations are done on optimized geometries with the key O2′H group oriented into the optimal region for a given conformational family (see Methods). Since no unnatural interactions originating from artificial rSPS terminations occur in our reference structures, the calculated energies reflect the intrinsic backbone rotameric preferences. The benchmark database of optimized rSPS geometries with relative CBS(T) energies is available in Supporting Information.

Benchmarking the DFT Functionals. We have tested 10 DFs (Figures 3 and 4, Table 2 and SI Table S1). They provide quite satisfactory description of the relative energies of the RNA backbone with maximum absolute deviation (MAX) below 1.0 kcal·mol^{-1} in both environments. The range of the relative energies of the 46 conformations is 10.6 kcal·mol^{-1}, and the accuracy of the DFT-D3 methods is in the range 5−14%, considering the structures with the most negative and positive difference. The MAD (mean absolute deviation) of the tested DFs spanning the range ∼0.1−0.3 kcal·mol^{-1} appears slightly lower compared to common benchmark studies of DFT-D for molecular clusters.[127] This accuracy is considerably better compared to simulation force fields (vide infra and ref 53) and should be sufficient for reliable qualitative analysis of nucleic acids. Nevertheless, it should be noted that, due to the fine energy scale of the backbone, even errors of few tenths of kcal·mol^{-1} might sometimes substantially bias computational analyses of nucleic acids, especially when the errors accumulate over multiple nucleotides.[53,54,57]

(*meta*-)GGA Functionals. The tested (*meta*-)GGA group consists of five DFs: B-LYP, PBE, revPBE, TPSS, and oTPSS. All five methods provide good performance. Contrary to the Jacob's ladder expectations numerically simpler GGAs yield in general energies closer to the reference CBS(T) method in both vacuum and COSMO solvent (Figures 3 and 4, Table 2, and SI Table S1). While in gas phase B-LYP outperforms the remaining DFs with MAD and MAX of 0.26 and 0.64 kcal·mol^{-1}, respectively, in COSMO the most accurate DF is PBE with MAD and MAX being as low as 0.16 and 0.49 kcal·mol^{-1}. These results are also in contrast with conclusions in benchmark study on DNA backbone where TPSS was the clear winner at the (meta-)GGA level of theory.[53] Since the RNA rSPS and DNA SPS models differ only in the presence of the 2′-hydroxyl group, the reordering of DFs performance could be due to different electronic description of the intramolecular O2′H···X interactions. On the other hand, the available set of RNA conformations is more diverse than the set of reference DNA backbone geometries,[53] and thus, the present statistics is more robust. The recently reparameterized DFs, revPBE and oTPSS, represent in general no improvement over their original versions, which is in contrary to the recent benchmark study.[105] The most problematic systems (conformations &a, 3d, 5d, 6d, and 5j) common to all (*meta*-)GGAs and well described by hybrid DFs likely indicate requirement of HF exchange (Table 2 and SI Table S1). It is thus obvious that the electronic structure of the sugar−phosphate backbone is a tough test even for modern DFs. Despite that the accuracy of the tested (meta-)GGAs methods is slightly inferior to more elaborate DFs (Figures 3 and 4), their simpler exchange-correlation evaluation and the possibility to take advantage of the RI approximation render them computationally noticeably more feasible. For that reason PBE-D3 and B-LYP-D3 functionals with def2-TZVPPD basis set represent a viable compromise between accuracy and computational demandingness. These methods can be recommended especially for preliminary backbone conformational searches as well as for large-scale computations on complete nucleic acids building blocks.[128] Regarding the latter application, it should be stressed, however, that more testing on systems including nucleobases whose conformational preferences are influenced by stacking interactions need to be done. It should be noted that the currently available tests are done for small model systems, which typically include only limited number of contributions simultaneously.

Hybrid Functionals. The performance of tested hybrid DFs, B3-LYP, PW6B95, and MPW1B95 is shown in Figures 3 and 4. While both Zhao's DFs achieve similar accuracy as PW6B95 yielding slightly better results with gas phase MAD and MAX of 0.23 and 0.79 kcal·mol^{-1}, the far best agreement with reference data is provided by B3-LYP (Figures 3 and 4, Table 2, and SI Table S1). Unlike our previous study[53] dealing with the DNA backbone and also the study of Goerigk[105] where B3-LYP was inferior in comparison with PW6B95 and MPW1B95, it is the second best DF in the present work with gas phase and COSMO MAD 0.16 and 0.15 kcal·mol^{-1}, respectively. This might be attributed to the fact that LYP correlation functional[95] is known to work well in combination with BJ-damped D3 term.[129] Still, the reason why B3-LYP is less satisfactory for DNA backbone[53] is unclear. Even though mixing of HF exchange potential rectifies electronic structures description of several systems and thus represents improvement over (*meta*-)GGAs, there are six systems (1c, 1e, 1f, 2o, 2[, 1t) for which PW6B95 and MPW1B95 (and also PWPB95; vide infra) DFs are jointly less accurate in both environments (Table 2 and SI Table S1). No common structural features of these conformers possibly responsible for the lower accuracy of the listed DFs were found. Since both hybrids and PWPB95 double hybrid contain Perdew−Wang's exchange[130] (PW) and Becke's correlation functional[131] (B95) the origin of the error cannot be unambiguously identified without additional computations. Besides, PW6B95 and MPW1B95 have the highest MAX (0.85 and 0.86 kcal·mol^{-1}, 1t) of all investigated DFs in solvent. Taking into account higher computational requirements of hybrids, PW6B95 and MPW1B95 do not provide any advantage for description of RNA backbone. Although B3-LYP outperforms all mentioned DFs and is the second best functional in the present study, it is less accurate in describing electronic structure of DNA backbone[53] and many systems of other benchmark sets.[105] The present computations confirm that it is relatively straightforward to achieve a good qualitative accuracy with dispersion-corrected DFs; however, fully quantitative agreement with reference calculations is tricky and different tests may favor different DFs.

Double-Hybrid Functionals. Both PWPB95 and DSD-B-LYP provide better results than simpler hybrids and show high correlation with CBS(T) gas phase energies with MAX of 0.61 and 0.38 kcal·mol^{-1} (Table 2 and SI Table S1). Also, gas phase MAD of 0.16 and 0.10 kcal·mol^{-1} ranks them among the most accurate methods within the current study (Figures 3 and 4). Especially the DSD-B-LYP functional shows an astonishing agreement with CBS(T) and, in line with our DNA study,[53] is the clear winner of this benchmark. Even though high computational costs prohibit its practical and widespread usage for larger systems, DSD-B-LYP with quadruple-ζ basis set is the method of choice when highly accurate energies are required.

MP2/CBS and ΔCCSD(T). The MP2/CBS method with (D,T) extrapolation reaches MAX below 0.9 kcal·mol^{-1}. Its performance is slightly inferior to majority of substantially faster DFT methods and is approximately of meta-GGA quality. Part of the error of the MP2/CBS procedure with (D,T) extrapolation is

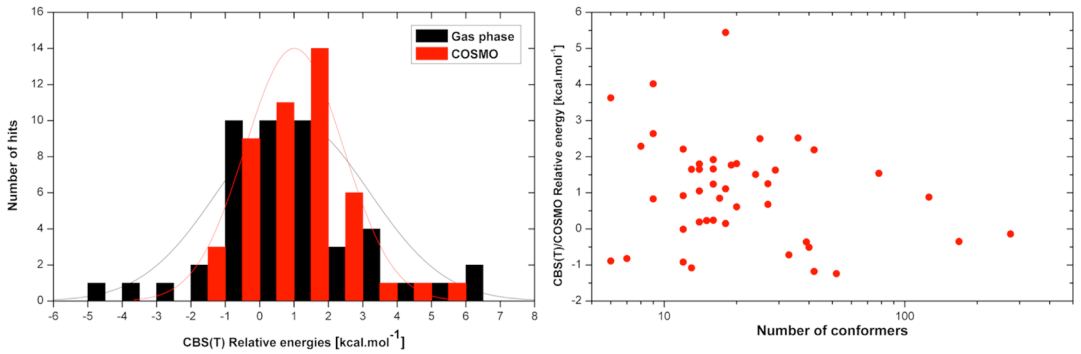


Figure 5. Relative CBS(T) energies (kcal·mol^{-1}) histogram for gas phase (black) and COSMO solvent (red) with fitted normal distributions, width of a bin equals 1.0 kcal·mol^{-1} (left). Correlation of solvent CBS(T) relative energies (kcal·mol^{-1}) of backbone conformations with the logarithm of number of their frequency of occurrence in the database of structures (logarithmic scale) used in the bioinformatics classification of RNA backbone.[61]

due to the smaller basis sets used in the extrapolation and part due to neglect of higher-order electron correlation.

The ΔCCSD(T) corrections calculated using the aug-cc-pVDZ basis set show only small conformational dependence, similarly to data obtained for DNA backbone,[38,53] with the mean and maximum value of ΔΔCCSD(T) being 0.13 and 0.40 kcal·mol^{-1}. This correction also obviously determines the error of the MP2/CBS computation with (T,Q) extrapolation. Note that the present large set of sufficiently diverse RNA backbone conformations covers the whole presently known biologically sampled space, so existence of RNA conformations for which the higher-order correlation effects would be more important is unlikely. Since the average difference between DNA backbone ΔΔCCSD(T) corrections calculated using 6-31+G* and aug-cc-pVDZ basis sets[53] is below 0.05 kcal·mol^{-1}, it is also improbable that further basis set expansion would lead to a dramatic revision of the results. The insensitivity of the backbone energetics to the higher-order correlation effects contrasts studies of molecular clusters.[132] Based on the preceding and current conclusions for DNA/RNA backbone, it can be stated that the MP2 level of theory clearly suffices for qualitative description of backbone electronic structure and that costly CCSD(T) calculations can be avoided unless highly accurate results are demanded.

Molecular Mechanics. MM calculations were performed with the ff99bsc0[55,117] force field in gas phase and PB continuum model. In line with DNA backbone studies[38,53] MM provides by 1 order of magnitude less accurate results compared to QM methods. While gas phase MM results deviate from CBS(T) with MAD and MAX being 4.20 and 9.19 kcal·mol^{-1}, MM-PBSA energies have mean and maximum error of 3.26 and 6.74 kcal·mol^{-1}, respectively (Table 2 and SI Table S1). Taking into account that relative energies of ∼75% systems lie in the interval −1 to +2 kcal·mol^{-1} (according to solvent CBS(T) results, Figure 5) MM-PBSA method resolution of ∼3 kcal·mol^{-1} is insufficient to capture fine energy differences between distinct RNA backbone rotamers. Obviously, one can suggest that our results can be partially influenced by the simple approach of the adjustment of the partial atomic charges after truncation of the model system. However, in our preceding study on DNA backbone, we utilized four different sets of charges with no visible improvement of the results.[53] Thus, we suggest that the differences between the QM and MM data stem rather from the simplicity of the force field model than from utilization of a specific charge set. We expect that a considerable part of the discrepancy can be attributed to polarization effects neglected in the fixed-charge MM calculations. Nevertheless, we still do not want to rule out that some improvement might be reached by a further tuning of the force field torsional parameters.[55,57−59] However, we suggest that it is unlikely that the MM based on constant atomic point charges can provide description of the DNA and RNA backbone that would be comparable in quality with modern DFT methods, since the ability of the torsional terms to tune the force field performance is certainly not limitless.[133]

Distribution of Relative Energies. Despite the enormous progress in quality of QM computations, such computations were so far only rarely considered in bioinformatics studies of nucleic acids.[5] Contemporary bioinformatics is primarily based on experimental structural data, which provide no insights into the energies of the studied local interactions and conformations. Further, biomolecular architectures are primarily shaped by geometrical properties of their components and building blocks, which is known as principle of isostericity.[1−5] Thus, solid bioinformatics insights can be built up without any inclusion of the energies of molecular interactions and conformations. Finally, gas phase energetics is not always representative for solvated nucleic acids. Nevertheless, the recent classification of conserved sugar−phosphate H-bonds in folded RNAs demonstrated that on average the evolution prefers those interactions that are intrinsically more stable.[5] This suggests that intrinsic energetics as captured by QM computations on small model systems is a factor that affects how often are given interactions and conformations utilized in evolution, that is, their frequency of occurrence.[1−5]

We therefore assessed correlation between intrinsic energetics of the RNA backbone and the frequency of occurrence of the different backbone arrangements in the course of the evolution. This issue is, however, quite complex. First, it is important to highlight that different RNA backbone families support nucleic acids bases in creation of different local geometries. This appears to be the primary factor affecting the frequency of occurrence of different backbone substates during the evolution. Further, it should be pointed out that many experimentally determined backbone conformations might be inaccurate or even entirely incorrect since their X-ray geometries are affected by various kinds of data and refinement errors.[61] However, the structural bioinformatics methods using elaborate statistical approaches are capable to infer populated conformational classes/families in the multidimensional conformational space and to screen out outlying structures, which do not pertain to any of those and are probably misleading. It is therefore obvious that sparsely populated RNA backbone classes are prone to be deceptive. On

dx.doi.org/10.1021/ct400837p | J. Chem. Theory Comput. 2014, 10, 463−480

Table 3. Conversion between Torsion Angles Used to Measure the O2′H Group Orientation[a]

torsion	approximate domain					
C3′-C2′-O2′-O2′H[b]	*eclipsed*	*plus* (p)	+e	*trans*	−e	*minus*
C1′-C2′-O2′-O2′H	−e	*minus* (m)	*eclipsed*	*plus*	+e	*trans*
C2′H-C2′-O2′-O2′H	+e	*trans* (t)	−e	*minus*	*eclipsed*	*plus*

[a] *eclipsed* signifies 0°; *plus*, +60° (p); *minus*, −60° (m); *trans*, 180° (t), +e, 120°± 25°; and −e, −120° ± 25°. [b] The present study. Our data can be compared with those by Denning et al.[50,51] by adding −120° to our values.

the other hand there is always a risk to neglect real structures that are infrequently adopted by RNA molecules, as they can be functionally very important on specific occasions (such as catalytic centers of RNA enzymes). Further, an unknown number of stable backbone conformations might be missed by the bioinformatics because of low frequency of occurrence or simply because structures with a given backbone conformation are difficult to obtain, for example, due to crystallization problems or tendency to be disordered. Therefore, the set of 46 backbone conformations is still likely incomplete to certain extent, and less frequent conformations should be taken with more caution. However, since the current consensual backbone database[61] was established on agreement between different methods and clustering approaches,[65−68] we do not anticipate it to contain a lot of erroneous conformational classes of extremely low stability, that is, of suspiciously high energies. Since biologically active structures of macromolecules in general sample geometries on a fine free energy scale of few kcal·mol^{-1}, the existence of significantly populated conformations with high intrinsic energies is doubtful. It is thus expected that the relative energies of backbone substates fall into a rather narrow range.

The distribution of CBS(T) relative energies calculated in gas and condensed phase is plotted in Figure 5 (left) and confirms the above considerations. The CBS(T)/COSMO distribution is appreciably narrower than the gas phase distribution and likely better reflects the real conformational preferences. While in vacuum approximately 33% of substates are more stable than the canonical A-RNA backbone, when the model systems are put in the solvent the respective fraction decreases to ∼27%. The effect of polarizable continuum mimicking water environment is significant and is responsible for relative energy changes as high as ∼4 kcal·mol^{-1} (Table 2 and SI Table S1). Thus, inclusion of solvation effects is essential when predictions about stability and conformational preferences of RNA backbone are made.

We have also analyzed the relation between the intrinsic energy of the different conformations of the isolated RNA backbone model and the frequency of occurrence of the backbone conformations in experimental structures (Figure 5, right). All conformations identified more than 50× (#) (i.e., 1a, 1c, 1b, 1g, 1[, and 2a) have their CBS(T)/COSMO relative energies below 2.1 kcal·mol^{-1}. So, there is no high-energy conformation among those frequently occurring. On the contrary, the three least stable conformations as measured by CBS(T)/COSMO energies, 4g (5.44 kcal·mol^{-1}, 18×), 4n (4.02 kcal·mol^{-1}, 9×), and 0i (3.63 kcal·mol^{-1}, 6×) are rarely observed. Nevertheless, stability of the given backbone rotamer inside a particular RNA context may be decisively affected by its compatibility with the overall architecture and interactions with RNA parts not included in the computations. Still, on average, more populated clusters tend also to be more stable intrinsically. Therefore, we suggest that there exist an evolutionary pressure to preferably use conformations, which are shown as intrinsically stable by QM calculations even on small model systems.

Differences in Hyperconjugation Effect Do Not Contribute to Relative Stability of Different RNA Conformers. Electronic structure of nucleic acid backbone is known to be delocalized and stabilized by a conformational-dependent hyperconjugation effect.[37,38] It has been suggested that hyperconjugation may contribute to difficulties in MM description of nucleic acids.[52] A possible way how to estimate extent of a hyperconjugation network is by summing up the off-diagonal Fock matrix elements, which refer to interactions between bonding (σ) and adjacent antibonding (σ^*) orbitals or lone pairs.

We found that a delocalization degree of a conformation does not correlate with its intrinsic stability. The highest delocalization effect was calculated for 4n, 6n, and 1z conformations with ∼2% stronger hyperconjugation compared to the reference 1a structure. However, both rare group III conformations are of lower stability with relative energies of 3.27 (6n) and 6.02 (4n) kcal·mol^{-1}. On the contrary the less delocalized electronic structures of 0b, 4p, and 6p group IV substates are stabilized by approximately 2−3% weaker hyperconjugation effect compared to 1a but do not show markedly lower stabilities with relative energies lower than 1.4 kcal·mol^{-1}. Even though delocalization of valence electrons contributes certainly to stabilization of RNA backbone its variation among RNA backbone substates is below ∼3%. Moreover, no visible correlation between hyperconjugation extent and total electronic energies was found. Therefore, we suggest that hyperconjugation effect despite stabilizing complex electronic structure of RNA backbone likely contributes rather marginally to the conformational preferences.

Selected Properties of the Individual Backbone Conformers. Full analysis of all geometrical aspects of the 46 very diverse RNA backbone families would be very complex and beyond the scope of this study. The reader can make detailed assessments using our published coordinates and profiles of the 2′-hydroxyl group rotations that are given in SI. In the subsequent text we nevertheless provide some guidelines, especially considering the preferred orientation of the O2′H group, utilizing also information from our MD simulations.

There have been some earlier MD-based attempts to uncover preferred O2′H group orientations.[50,51,134,135] Other studies dealing with placement of 2′-hydroxyl utilized experimental techniques,[136] especially NMR calculations of 2′-hydroxyl group proton exchange rates[137] or measurement of scalar coupling constants.[138,139] Influence of O2′H group position on calculated ribonucleotide chemical shifts NMR parameters has also been investigated.[140]

There are several different, yet equivalent schemes of defining the O2′H group orientation. While in this study we use C3′-C2′-O2′-O2′H torsion angle, Denning et al.[50,51] utilized C1′-C2′-O2′-O2′H torsion instead. In other studies[134,138] 2′-hydroxyl orientation is evaluated by measuring H2′-C2′-O2′-O2′H torsion. However, since the schemes differ in the first atom

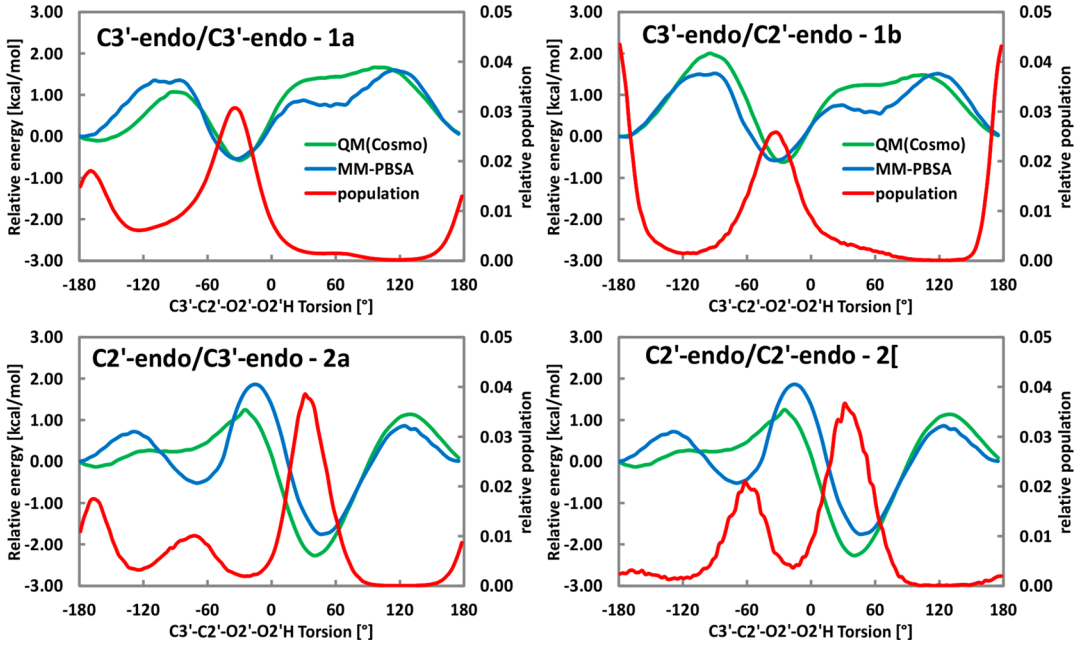


Figure 6. QM/COSMO (green) and MM-PBSA (blue) rigid scan profiles (kcal·mol^{-1}) of 2′-hydroxyl group position supplemented with orientation distributions (red) from MD simulations for the most populated RNA backbone conformations of groups I–IV, namely, 1a, 1b, 2a, and 2[.

only (C3′, C1′, H2′), the torsions are simply shifted by a phase angle (Table 3).

Description of Group I RNA Backbone Conformations. The Group I consist of 17 RNA backbone substates with C3′-endo conformation of both sugar residues (labeled also as 33) and is thus the most frequent (Table 1). The group I can be further divided based on the $\gamma + 1$ torsion into two subgroups labeled as *plus* ($\gamma + 1 \sim -60°$, subgroup 33p) and *trans* ($\gamma + 1 \sim -180°$, 33t). The most frequent conformational class 1a (33p) is adopted by over 75% of RNA nucleotides[61] including the A-form double helix. Even though *γ-trans* DNA conformations are generally assumed to be by 3–6 kcal·mol^{-1} less stable,[37,38,55] it is not the case of the 33t RNA backbone substates. This indicates that penalization of *γ-trans* substates by dihedral terms of force fields that is important to stabilize simulation behavior of canonical DNA and RNA helices[55] may destabilize native *γ-trans* states in noncanonical regions.[53,141] On the contrary, classes 7a (2.52 kcal·mol^{-1}) and 3a (2.50 kcal·mol^{-1}), both pertaining to the 33p subgroup, represent the least stable backbone rotamers as measured by CBS(T) energies in the COSMO environment. While the former conformation is usually encountered in stack switch and inward-dented backbone within double helices,[142] the latter one represents a local deviation of 1a typically ending an A-form strand.[61] Note, however, that the slightly lower stabilities of the conformations might be in RNA outbalanced by non-canonical base-backbone interactions or some tertiary contacts inherently missed in our simple model.

Intrinsically, the most common 2′-hydroxyl group acceptor for group I conformations is the adjacent O3′, as revealed by our O2′H group scans (SI Table S2). With the exception of &a rotamer (a satellite cluster of 1a) whose backbone arrangement enables its O2′H group to establish a bifurcated H-bond with O4′ and O5′ of the *i*+1 nucleotide, the remaining 33p subgroup conformations adopt the most common O2′H···O3′ interaction (SI Table S2). The presence of two H-bonds in &a family is reflected by a somewhat higher stability of ∼−2.0 kcal·mol^{-1} in gas phase (Table 2). However, according to the AIM analysis, both coexisting O2′H···O4′(*i*+1) and O2′H···O5′(*i*+1) H-bonds are approximately of half strength compared to the average O2′H···O3′ interaction and contribute to conformation stabilization equally. There is also an indication of a weak O2′H···O4′(*i*+1) contact in 7a conformation which, however, is not listed since both AIM parameters (ρ, $\nabla^2\rho$) of the respective critical points are below the threshold value of 0.01 a.u., albeit the AIM analysis is not always unambiguous (SI Table S2). In line with the study of Auffinger,[134] except for &a and 7a, we did not detected any O2′H···O4′(*i*+1) H-bond in any of the group I systems. However, a slight modification of backbone torsion angles might bring O4′(*i*+1) closer to O2′H group and thus renders it accessible for H-bond interaction.

The scans revealed two minima for the O2′H group orientation in group I as measured by the C3′-C2′-O2′-O2′H dihedral (see Figure 6 for 1a conformation and SI Figure S1 for group I systems). The global minimum lying approximately in the (−70°, −30°) interval (i.e., the m region) includes the O2′H···O3′ H-bond. The only exception is &a conformation, as its global minimum is slightly shifted to lower values of ∼−80° due to a different H-bonding pattern (SI Figure S1). The second local minimum more pronounced for the 33p subgroup (and insignificant or missing for 33t) is approximately in the t region (∼ 180°) where the O2′H group points outward and does not interact with the backbone H-bond acceptors. Note that the *trans* minimum is missed also in gas phase scans of several 33p conformers probably due to overestimated electrostatic interaction partially screened out by continuum solvent models.

Figure 6 shows that the preferred orientation of the O2′H group is quite well reproduced by the AMBER force field in calculations on the model system. Although the agreement between the QM and MM curves is not perfect, we suggest that it does not indicate any need of a reparameterization of the force field. Recent reparameterization of O2′H group dihedral potential[50] markedly improved the performance of the RNA CHARMM force field.[143] However, in that particular case, the older version of the force field was imbalanced. Obviously, the

dx.doi.org/10.1021/ct400837p | J. Chem. Theory Comput. 2014, 10, 463–480

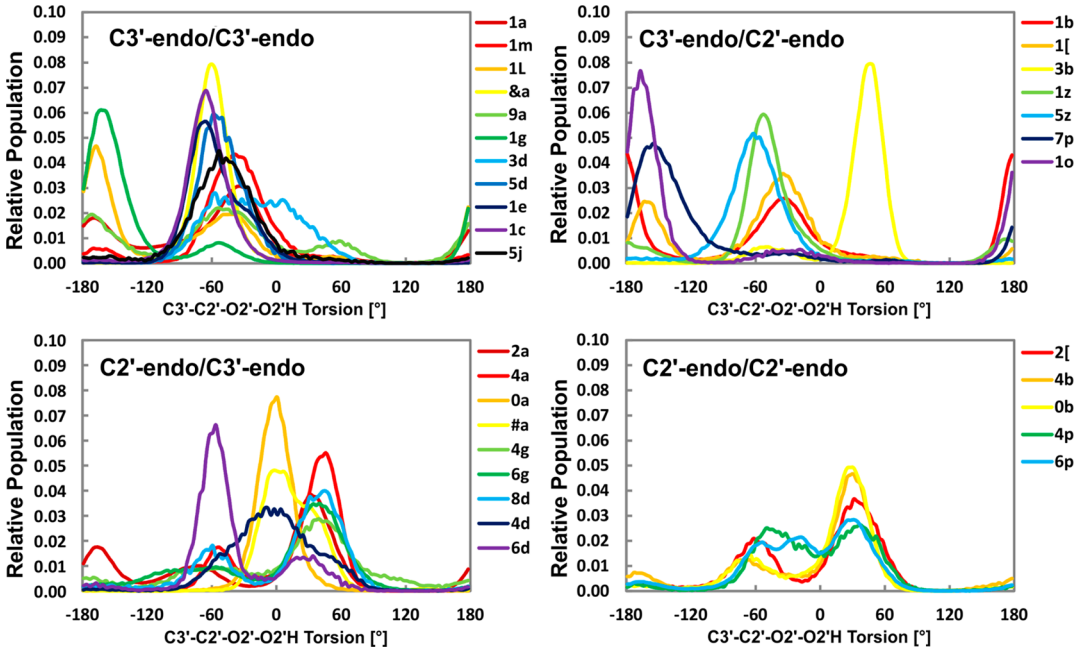


Figure 7. Histograms showing relative populations of the orientations of the C3′-C2′-O2′-O2′H torsions in different suites in simulations of *glmS* riboswitch and SRL using ff99bsc0χ_{OL3} force field. The histograms were calculated only for suites, for which we obtained sufficient sampling of orientations of the C3′-C2′-O2′-O2′H torsion, that is, more than 10 000 of data points. Note that our analysis is still based only on simulations of two RNA molecules. It is likely that the actual picture of the O2′H orientations in some families is highly context-dependent, mainly in those cases when the O2′H group participates in key sugar-edge tertiary interactions as H-bond donor. Such analysis is, however, beyond the scope of this work. Note also that many nucleotides sample different families in the course of the simulation and also not all sampled points can be unambiguously assigned to the RNA families; that is, ∼11% points had to be excluded. Thus, the distributions should be primarily taken as representative examples and do not provide fully comprehensive analysis of the O2′H torsions in RNA simulations.

behavior of the O2′H group in RNA simulations may differ from the prediction made by small backbone model systems, since the O2′H group can be reoriented by RNA tertiary interactions and also by explicitly hydrating water molecules. Nevertheless, both O2′H group minima located by the scan for the group I conformations clearly match the most probable orientations of the 2′-hydroxyl from MD simulations (Figure 7).

Even though the ∼−60° minimum obviously relates to the O2′H···O3′ H-bond in both rSPS model and MD simulations, there are two reasons why the 2′-hydroxyl group of group I conformers adopts also the *trans* orientation (Figure 7). MD reveals that the *trans* position enables the O2′H group to contact explicit water molecules residing approximately below a nucleobase in *anti* orientation[50] or to establish a tertiary interaction with available H-bond acceptors (various sugar-edge interactions, such as the A-minor interactions).[1−3,144] The sugar-edge interactions are crucial to build up specific RNA 3D structures and were extensively studied by QM methods.[7,8,144,145] In the present study, we monitored only the O2′H group orientations in MD simulations, and thus, we do not distinguish the different H-bonding scenarios (H-bond type). We plan to do it in some future studies for specific tertiary contexts. However, the local *trans* minimum from the COSMO O2′H group scans in absence of any water molecules can be also rationalized by NBO analysis of the C3′-endo sugar (vide infra). Interestingly, the intrinsic *trans* substate is well reproduced by the force field. Therefore, both the orbital interaction aspect (minor effect) and the possibility to create a H-bond with surrounding water molecules/H-bond acceptors (major effect) give synergically rise to a non-negligible population of *trans* O2′H orientation in MD simulations, especially for the 33p conformations. These finding are generally in line with other computational studies.[134,51] However, except for the O2′H···O3′ H-bond (m region) and tertiary interaction (t region), Auffinger[134] suggested that the O2′H group interacts also with O4′ of the same sugar. Even though there is a shallow minimum at ∼ *plus* region (Figure 6) associated with the O2′H···O4′ interaction, relative populations of this orientation are rather low (Figures 6 and 7). Nevertheless, it cannot be ruled out that the *plus* orientation enables the O2′H group to act as a H-bond acceptor as the lone pairs of O2′ become accessible for outside H-bond donors. In other words, the *plus* orientation of 2′-hydroxyl seems to be stabilized by some kind of tertiary interaction rather than by the weak O2′H···O4′ H-bond. Similar results were obtained also by NMR scalar coupling measurements[138] and QM calculations on small model systems accompanied with crystallographic survey data.[51] The later study suggested that the 2′-hydroxyl group modulates intrinsic backbone flexibility, which enables RNA backbone to adopt a noncanonical conformation by interaction with nearby O3′. On the contrary, the base plane (i.e., *trans*) orientation of O2′H is typical for the canonical A-RNA conformation. Although results of the current study are in agreement with conclusions of ref 51 in that there are two populated regions, *minus* and *trans*, relative distributions (Figure 6) of the O2′H group of 1a conformation do not show any predominance in the *trans* domain as anticipated by Denning et al.[51] This might be partially attributed to the fact that the 1a suite is not exclusively involved in A-RNA helices but occurs also in noncanonical regions, which in turn might result in increased sampling of the *minus* domain. Let us, however, reiterate that full analysis of the O2′H group orientations in folded RNAs would require a careful consideration of all their specific tertiary contexts,[52] which will be attempted in future studies.

dx.doi.org/10.1021/ct400837p | J. Chem. Theory Comput. 2014, 10, 463−480

Group II Conformations. The 10 membered C3′-endo/C2′-endo group II is further divided according to $\gamma + 1$ torsion into three subgroups: 32p (∼60°, 6 members), 32t (∼180°, 2), and 32m (∼−60°, 2) (Table 1). In line with the 33t subgroup neither *γ-trans* nor *γ-minus* noncanonical conformations show lower intrinsic stabilities compared to the reference 1a family. Except 5z (32p) and 1t (32t) conformations, which form characteristic O2′H···O2P(i+1) and O2′H···O5′(i+1) H-bonds, respectively, all group II systems are stabilized by the common O2′H···O3′ interaction. Although H-bond with anionic O2P is the strongest one as indicated by AIM (SI Table S2), the 5z rotamer found in RNA S-turn[60,146] is the least stable conformation of the group with CBS(T)/COSMO relative energy of 2.19 kcal·mol^{-1}. Contrary to that conformation 1[(32p) separating consecutive bases and allowing intercalation of a drug or another base between is the most stable group II suite with −3.30 and −1.24 kcal·mol^{-1} gas phase and COSMO CBS(T) energies, respectively.

MD simulations identify three preferred 2′-hydroxyl group orientations. Members of the first class, such as 1b and 1[, are stabilized by either O2′H···O3′ interaction indicated by an orientation distribution peak at ∼−40°, or form a H-bond with explicit water molecules or a tertiary contact, as evidenced by populated *trans* orientation (Figure 7). Both probability peaks at m/t regions derived from MD correspond to stationary points on O2′H group rotation energy profile done in COSMO environment (see Figure 6 for 1b and SI Figure S1 for all group II systems). Note that the deeper *minus*-domain minimum of 1[compared to 1b reflects a slightly higher probability of the O2′H···O3′ interaction (Figure 7). Conformers of the second class have only one significantly populated O2′H group position, which relates to either of two located minima on the rotation energetic profile. While 5z and 1z systems have their 2′-hydroxyl exclusively in the *minus* domain and are stabilized by an intramolecular sugar-backbone H-bond, 7p and 1o predominantly form H-bonds with solvating water molecules (Figure 7). The last class consisting of 3b conformation shows a single distribution peak at ∼50° (*plus* region) (Figure 7). MD simulation of *glmS* riboswitch reveals that O2′H group of 3b suite (between G34 and A35 residues) forms a tertiary H-bond with a phosphate group of U59 residue (Figure 8). It is thus obvious that O2′H group orientation is often affected by its specific structural context. A given RNA backbone family can sample various O2′H orientations depending on its structural context.

Group III Conformations. Group III embraces a set of 14 conformations with C2′-endo and C3′-endo conformations of 5′ and 3′ sugars, respectively. While nine suites (2a, 4a, 0a, #a, 4g, 6g, 8d, 4d, and 6d) have their $\gamma + 1$ torsion in canonical *plus* domain (23p), remaining 5 systems (2h, 4n, 0i, 6n, 6j) sample *γ-trans* conformational subspace (23t) (Table 1). Unlike the two previous groups, both 23p and 23t backbone conformations are generally of lower stability compared to reference A-RNA. The most frequent 2a substate (23p) is one the most stable conformations of the group with CBS(T)/COSMO energy being 0.88 kcal·mol^{-1}. The three least energetically favorable conformations 4g, 4n, and 0i with CBS(T)/COSMO energy as high as 5.44, 4.02, and 3.63 kcal·mol^{-1} are very rare.[61] Similarly as for previous group conformations, except for 0a, #a, and 4n, all group III model systems tend to form intramolecular O2′H···O3′ H-bonds of comparable strength (SI Table S2). For the 0a conformation (typical for a cross-stacked A-helix start conformation), a bifurcated H-bond to O3′ and O1P(i+1) acceptors was found by AIM analysis. The strongest H-bond between O2′H and O1P(i+1) forming a so-called backbone-sugar edge[146] stabilizes the #a conformation,[60] which is native to intrastrand GpU dinucleotide platform, a crucial part of SRL domain S-motif. A slightly weaker O2′H···O1P(i+1) H-bond exists also in the rare 4n backbone substate.

In MD, the most populated 2′-hydroxyl orientation of the group III conformers lies in the (−60°, 60°) region, while the *trans* domain is sampled rather negligibly as opposed to groups I and II (Figure 7). Note that whereas the O2′H···O3′ interaction for C3′-endo sugars for group I and II conformations is connected with a *minus* domain distribution maximum, for C2′-endo sugar pucker (groups III and IV) the respective maximum is found in *plus* region (vide infra). The 2′-hydroxyl group rotation scan of 2a suite revealed three minima at ∼ *plus* (40°, global minimum with O2′H···O3′ H-bond), *minus* (−70°, indication of O2′H···O4′(i+1) H-bond), and *trans* (180°, intermolecular interaction with base plane water molecules or tertiary interaction) (Figure 6). This nicely coincides with the respective distribution peaks from MD simulations (Figure 7). Note that 2a is the only backbone substate showing a trimodal distribution of O2′H group orientation. Hydroxyl group rotation energetic profiles for all group III systems have a similar shape (see Figure 6 for 2a and SI Figure S1 for all group III systems). Two maxima at ±130° are ∼1.5 kcal·mol^{-1} above the reference *trans* orientation and likely result from orbital interactions (vide infra). The third maximum around −15° can also be explained by orbital interaction analysis, but due to a conformational dependency its height is variable, for example, of ∼2.5 kcal·mol^{-1} for 2h and nearly negligible in the cases of 4d, #a, 0a, 6n, 6j. The maximum at −15° represents a "transition state" connecting two minima of different depths at ∼ −60° and 60°. Suites with rather high energy barrier around −15° (e.g., 2a, 2h, 8d) show a bimodal (or trimodal for 2a) hydroxyl group orientation distribution with relative population peaks at favorable *minus* and *plus* domains. On the other hand, conformations with low-lying maximum in the ∼−15° region display a broad MD distribution centered around 0° and covering the whole (−60°, 60°) interval (Figure 7). In principle, group III (and also group IV) conformations could establish a sugar···base H-bond (e.g. O2′H···N3(G/A) or O2′H···O2(C/U)). This interaction is, however, possible only for a rather narrow range of glycosidic torsion angles and is not populated in MD simulations.

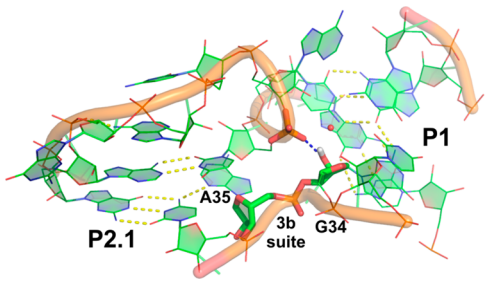


Figure 8. Tertiary interaction between O2′H group of 3b suite (G34 and A35 residues) and a phosphate group of a nearby U59 residue of *glmS* riboswitch. Orientation of the 2′-hydroxyl group corresponds to ∼50° (*plus* region).

Group IV Conformations. Conformations of the least populated group are characterized by C2′-endo pucker of both sugar residues. It can be categorized based on $\gamma + 1$ torsion into three subgroups: 22p (2[, 4b, 0b, 4p, and 6p), 22t (4s), and 22m

dx.doi.org/10.1021/ct400837p | J. Chem. Theory Comput. 2014, 10, 463−480

(2o) (Table 1). The most common 2[suite structurally close to B-form DNA backbone is one of the most stable conformations of the group with relative CBS(T)/COSMO energy of −0.51 kcal·mol^{-1}. To the contrary 4s, the rarest backbone substate of the group, is of lowest stability with CBS(T)/COSMO energy being 2.29 kcal·mol^{-1} above 1a. All group IV model systems form O2′H···O3′ H-bond associated with O2′H group orientation global minimum at ∼30° (SI Table S2). In line with group III systems having 5′ sugar also in C2′-endo conformation, 2′-hydroxyl groups of group IV backbone substates sample a broad region of (−60°, 60°) (Figure 7). An energy barrier of variable height at ∼−15°, identical to that of group III model systems, is clearly responsible for the relative population depression in this region (see Figure 6 for 2[and SI Figure S1 for group IV systems). Also similarly to group III suites (except for 2a) the *trans* domain associated with a shallow local minimum is sampled rather marginally.

Distribution Peaks at O2′H Group *Trans* Orientation. Our Md simulations (Figure 7) show that the o2′h group frequently samples the *trans* region, where it points outward from the sugar−phosphate backbone. However, as noticed above the *trans* orientation can be associated with interaction either with explicit waters or with various H-bond acceptors, such as a phosphate group from distinct segment of the RNA molecule, an O2′H group of a nearby sugar moiety, an adjacent nucleobase, etc. Our present MD analysis (Figure 7) does not differentiate between these interactions and all *trans* domain related H-bonds are averaged in the statistics.

Effect of Solvent and MM vs QM Profiles. QM- and MM-derived energetic profiles of 2′-hydroxyl group rotation were evaluated in both gas phase and solvent environment (SI Figure S1). While the gas phase profiles have in general one deep minimum clearly corresponding to an intramolecular sugar-backbone H-bond, energetic profiles calculated in solvent contain several shallow minima. The higher number of minima in solvent profiles can be ascribed to interaction between the O2′H group and the polarized continuum. Even though there is no explicit H-bond acceptor (e.g., in the *trans* domain), a continuum solvent model effectively mimics presence of surrounding polar molecules. Moreover, the reactive field of implicit solvent affecting solute wave function attenuates electrostatic interactions. Since the H-bonds are predominantly of electrostatic nature, they are destabilized when the system is immersed in a polarizable medium resulting in more shallow PES minima. The inherent inability to encompass effect of environment and the overestimation of electrostatics (compared to solvated nucleic acids) render gas phase energy profile calculations unsuitable for correlation with MD simulation distributions or experimental data.

The differences between QM/COSMO and MM-PBSA profiles are not dramatic. Except for the 4p substate where QM and MM differ in global minimum prediction (QM−*plus*, MM−*minus*; SI Figure S1), there are no profound differences between QM/COSMO and MM-PBSA profiles. Note, however, that our calculations are based on rigid scans. In order to quantitatively compare QM and MM profiles (e.g., for MM reparameterization purposes), a constrained optimization of geometry for each O2′H group would be needed.

Orbital Interaction Analysis. In order to better understand the nature of the energy minima and maxima for the 2′-hydroxyl orientations, we evaluated QM energy profiles of 2′-hydroxyl orientation for C3′- and C2′-endo ribose (Figure 9, see also Methods). The structures near maxima and minima energy

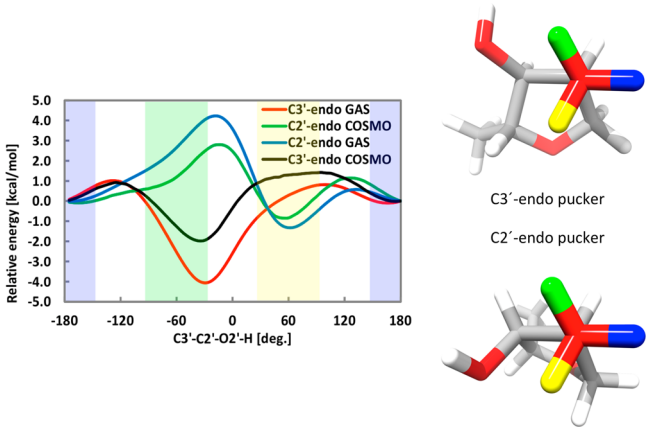


Figure 9. QM- and MM-derived 2′-hydroxyl group rotation scan profiles (kcal·mol^{-1}) for C3′- and C2′-endo puckered ribose and possible orientations of the O2′H group described by the C3′-C2′-O2′-O2′H torsion angle. The *trans* (∼180°), *plus* (∼60°), and *minus* (∼−60°) domain orientations are indicated by blue, yellow, and green colors, respectively.

orientations were further investigated by NBO analysis. For the C3′-endo pucker two minima and two maxima were localized. The global minimum associated with the O2′H···O3′ H-bond lies within the *minus* domain at ∼−30°. The second low-energy orientation in *trans* region is enclosed by two maxima at ∼100° and −130°. Both high-energy orientations approximately maximize the overlap between the O2′ and O3′ nonbonding electron orbitals (Figure 10, first row, left and right). Moreover, both orientations represent eclipsed conformations, which are known to be sterically unstable. Thus, if the C3′-C2′-O2′-O2′H angle equals ∼100° then C1′-C2′-O2′-O2′H is close to 0°, and similarly, if the former torsion approaches −130°, the coupled C2′H−C2′-O2′-O2′H torsion angle goes near 0°. However, when the C3′-C2′-O2′-O2′H angle is approximately −30° (Figure 10, first row, center), then the 2′-hydroxyl group is in optimal orientation to establish the O2′H···O3′ H-bond and also the overlap of nonbonding orbitals of the O2′, O3′, and O4′ atoms is minimal. While the *trans* orientation eliminates the steric clash penalizing the ∼100° and −130° orientations, similarly to the −30° torsion, no intramolecular H-bond is formed. Because of that the *trans* orientation is only a local minimum supporting formation of intermolecular H-bonds (e.g., with water molecules or in sugar-edge interactions).

Two minima and two maxima were found also for the C2′-endo ribose (Figure 9). The global minimum related to the O2′H···O3′ interaction is around 60°. It is surrounded by two nearly symmetrically positioned maxima at approximately −20° and 130°. Both high-energy hydroxyl orientations equally maximize overlap of a C2′-C2′H σ-bonding orbital and a lone electron pair of O2′ (Figure 10, second row, left and right). For the ∼ −20° orientation the O2′H and O3′H hydrogen atoms are as close as ∼2.1 Å and thus repel each other electrostatically. Therefore, the maximum at −20° is slightly higher compared to that at 130° where no such atomic clash is present. Note, however, that the magnitude of the maximum is partially due to a constraint imposed on the C4′-C3′-O3′-O3′H angle (i.e., pseudo ε torsion), which prohibits the 3′-hydroxyl group adjustment in order to release the strain.

The properties of isolated sugar visibly influence the actual 2′-hydroxyl orientation preferences for RNA backbone suites groups I and II with C3′-endo pucker at the 5′ end have two

dx.doi.org/10.1021/ct400837p | J. Chem. Theory Comput. 2014, 10, 463−480

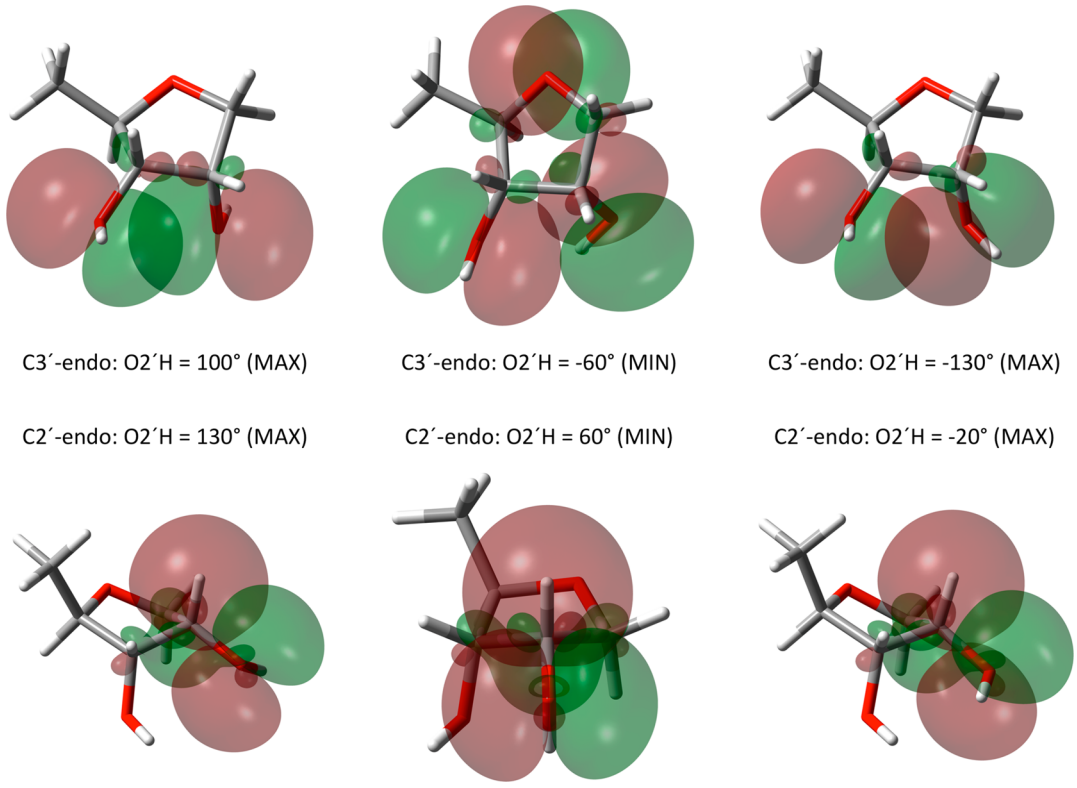


Figure 10. NBO orbitals of maximum (MAX) and minimum (MIN) energy conformations of C3′- and C2′-endo ribose. The listed O2′H group orientations are measured by the C3′-C2′-O2′-O2′H torsion angle (deg).

distribution maxima in ∼ *minus* and *trans* domains. Groups III and IV with C2′-endo pucker at the 5′ end have the respective maxima approximately in *plus* domain with non-negligible population also in *minus* and *trans* regions (Figures 6 and 9). With the exception of secondary distribution maximum in *minus* region for the 5′ C2′-endo suites, all populated domains as well as rarely sampled regions match the respective minima and maxima revealed for the isolated sugar moiety.

CONCLUSIONS

Comprehension of physical principles underlying structural diversity of RNA backbone and C2′ hydroxyl group orientation is of key importance to understand the RNA structural dynamics and to perform theoretical studies on RNA molecules.

We have created a benchmark QM database of relative energies of RNA backbone substates using structurally highly diverse complete set of 46 established backbone conformations (distinct combinations of the $\alpha-\zeta$ backbone dihedral angles). The calculations are carried out at the estimated CCSD(T)/CBS) level of theory both in vacuum and solvent environment.

Using this reference data, we assess performance of the state-of-the-art DFT methods and molecular mechanics to describe the RNA backbone. While all tested dispersion-corrected density functionals provide good agreement with QM benchmark with maximum deviations being not larger than 1 kcal·mol^{-1}, the AMBER ff99bsc0 force field does not reproduce the fine energy differences between the RNA backbone rotamers, in line with our previous DNA backbone study.[53] Our benchmark database of RNA backbone conformations can be readily used for testing of performance of other computational methods intended to study RNA molecules.

From the less-expensive density functionals, B-LYP supplemented with D3-BJ term provides the best trade-off between computational costs and accuracy. Interestingly, B3-LYP with D3 correction outperforms the new PW6B95 and MPW1B95 hybrids and is the second most accurate density functional of the study. Although DSD-B-LYP functional is the most accurate method of the study, high computational demands prevent its large-scale usage. We further confirm that ΔCCSD(T) corrections do not affect significantly the relative energies of different RNA backbone conformations.

We further show that differences in the hyperconjugation effect magnitude do not correlate with total electronic energy ranking and thus represent rather marginal contribution to structural diversity of RNA backbone.

Next, we address the issue of C2′ hydroxyl group orientation preferences. For the 46 backbone conformations, we conducted 2′-hydroxyl group rigid scans and compared the energy profiles with 2′-hydroxyl orientation distributions data obtained from explicit solvent MD simulations of folded RNAs. Energy profiles from rigid scans carried out in continuum solvent correlate with the MD 2′-hydroxyl orientation distributions. The C3′-endo/C3′-endo and C3′-endo/C2′-endo conformational families have two energy minima at ∼−60° and 180° associated in simulations with the O2′H···O3′ H-bond and interaction with base-plane residing water or tertiary interaction, respectively. Contrary to that the less frequent C2′-endo/C3′-endo and C2′-endo/C2′-endo conformations prefer primarily ∼+60° 2′-hydroxyl group orientation associated with the O2′H···O3′ H-bond and −60° and 180° regions associated with water/tertiary interactions.

AIM analysis shows that majority of the 46 conformations prefer generally the accessible O2′H···O3′ H-bond. Approximately 20% of them form alternative H-bond with O1P, O2P, O5′, or O4′ of the adjacent 3′-end nucleotide. It was found that

dx.doi.org/10.1021/ct400837p | J. Chem. Theory Comput. 2014, 10, 463−480

the high energy of the unfavorable and unoccupied orientations of the 2′-hydroxyl group might be elucidated by natural orbital analysis of the sugar moiety alone. While ∼100° and −130° high-energy O2′H orientations of the C3′-endo ribose result in clash between O2′ and O3′ lone electron pairs, the favorable −60° alignment releasing this strain is optimal for establishing the O2′H···O3′ H-bond. Similar clash between O2′ lone pair and C2′-C2′H bonding orbital at ∼130° and −20° renders both orientations to be of high energy.

Even though the force field is incapable to capture the backbone rotameric preferences, it reproduces the QM 2′-hydroxyl orientation energy profiles in our model system rather well. Note that since we obtained energy profiles by rigid scans, quantitative analysis and comparison of QM and MM-derived profiles cannot be commented on. Nevertheless our MM and QM data regarding O2′H group orientation preferences do not suggest any need to reparameterize the respective terms in the Cornell et al. RNA force field and its recent variants.

■ ASSOCIATED CONTENT

Supporting Information

Benchmark database of gas phase and COSMO optimized rSPS geometries with respective CBS(T) relative energies, list of COSMO relative energies, AIM-revealed O2′H group H-bonds, detailed MD simulation protocol, hydroxyl group rotation energy profiles. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Authors

*E-mail: arnost.mladek@gmail.com.
*E-mail: sponer@ncbr.muni.cz.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by research grant from the Grant Agency of the Czech Republic P208/12/1878. Institutional funding was provided by the Ministry of Education, Youth, and Sports of the Czech Republic: projects "CEITEC, Central European Institute of Technology" (CZ.1.05/1.1.00/02.0068), by "RCPTM, Regional Centre of Advanced Technologies and Materials" (CZ.1.05/2.1.00/03.0058) from European Regional Development Fund, and "RCPTM-TEAM", (CZ.1.07/2.3.00/20.0017, M.O., P.B., M.Z., P.J.) from the Operational Program Education for Competitiveness, European Social Fund. The present study was also financially supported by the South Moravian Centre for International Mobility within the framework of the "Brno Ph.D. Talent" scholarship program. Authors of the manuscript are thankful to Dr. Daniel Svozil for contributive cooperation.

■ REFERENCES

(1) Leontis, N. B.; Westhof, E. *RNA* **2001**, *7*, 499−512.
(2) Leontis, N. B.; Stombaugh, J.; Westhof, E. *Nucleic Acids Res.* **2002**, *30*, 3497−3531.
(3) Sponer, J.; Sponer, J. E.; Petrov, A. I.; Leontis, N. B. *J. Phys. Chem. B* **2010**, *114*, 15723−15741.
(4) Leontis, N. B.; Westhof, E. *Curr. Opin. Struct. Biol.* **2003**, *13*, 300−308.
(5) Zirbel, C. L.; Sponer, J. E.; Sponer, J.; Stombaugh, J.; Leontis, N. B. *Nucleic Acids Res.* **2009**, *37*, 4898−4918.
(6) Zgarbova, M.; Jurecka, P.; Banas, P.; Otyepka, M.; Sponer, J. E.; Leontis, N. B.; Zirbel, C. L.; Sponer, J. *J. Phys. Chem. A* **2011**, *115*, 11277−11292.
(7) Sponer, J. E.; Leszczynski, J.; Sychrovsky, V.; Sponer, J. *J. Phys. Chem. B* **2005**, *109*, 18680−18689.
(8) Sponer, J. E.; Spackova, N.; Kulhanek, P.; Leszczynski, J.; Sponer, J. *J. Phys. Chem. A* **2005**, *109*, 2292−2301.
(9) Sharma, P.; Sponer, J. E.; Sponer, J.; Sharma, S.; Bhattacharyya, D.; Mitra, A. *J. Phys. Chem. B* **2010**, *114*, 3307−3320.
(10) Mladek, A.; Sharma, P.; Mitra, A.; Bhattacharyya, D.; Sponer, J.; Sponer, J. E. *J. Phys. Chem. B* **2009**, *113*, 1743−1755.
(11) Vokacova, Z.; Sponer, J.; Sponer, J. E.; Sychrovsky, V. *J. Phys. Chem. B* **2007**, *111*, 10813−10824.
(12) Oliva, R.; Cavallo, L.; Tramontano, A. *Nucleic Acids Res.* **2006**, *34*, 865−879.
(13) Oliva, R.; Cavallo, L. *J. Phys. Chem. B* **2009**, *113*, 15670−15678.
(14) Parker, T. M.; Hohenstein, E. G.; Parrish, R. M.; Hud, N. V.; Sherrill, C. D. *J. Am. Chem. Soc.* **2013**, *135*, 1306−1316.
(15) Guerra, C. F.; Bickelhaupt, F. M.; Snijders, J. G.; Baerends, E. J. *J. Am. Chem. Soc.* **2000**, *122*, 4117−4128.
(16) Hobza, P.; Sponer, J. *Chem. Rev.* **1999**, *99*, 3247−3276.
(17) Sponer, J.; Leszczynski, J.; Hobza, P. *J. Phys. Chem.* **1996**, *100*, 1965−1974.
(18) Sponer, J.; Leszczynski, J.; Hobza, P. *Biopolymers* **2001**, *61*, 3−31.
(19) Sponer, J.; Leszczynski, J.; Hobza, P. *J. Phys. Chem.* **1996**, *100*, 5590−5596.
(20) Luo, R.; Gilson, H. S. R.; Potter, M. J.; Gilson, M. K. *Biophys. J.* **2001**, *80*, 140−148.
(21) Florian, J.; Sponer, J.; Warshel, A. *J. Phys. Chem. B* **1999**, *103*, 884−892.
(22) Lech, C. J.; Heddi, B.; Phan, A. T. *Nucleic Acids Res.* **2013**, *41*, 2034−2046.
(23) Sponer, J.; Sponer, J. E.; Mladek, A.; Banas, P.; Jurecka, P.; Otyepka, M. *Methods* **2013**, DOI: 10.1016/j.ymeth.2013.05.025.
(24) Sponer, J.; Sponer, J. E.; Mladek, A.; Banas, P.; Jurecka, P.; Otyepka, M. *Biopolymers* **2013**, DOI: 10.1002/bip.22322.
(25) Florian, J.; Leszczynski, J. *J. Am. Chem. Soc.* **1996**, *118*, 3010−3017.
(26) Bertran, J.; Oliva, A.; Rodriguez-Santiago, L.; Sodupe, M. *J. Am. Chem. Soc.* **1998**, *120*, 8159−8167.
(27) Gorb, L.; Podolyan, Y.; Dziekonski, P.; Sokalski, W. A.; Leszczynski, J. *J. Am. Chem. Soc.* **2004**, *126*, 10119−10129.
(28) Kelly, R. E. A.; Kantorovich, L. N. *J. Mat. Chem.* **2006**, *16*, 1894−1905.
(29) Sharma, P.; Lait, L. A.; Wetmore, S. D. *Phys. Chem. Chem. Phys.* **2013**, *15*, 2435−2448.
(30) Brandl, M.; Meyer, M.; Suhnel, J. *J. Phys. Chem. A* **2000**, *104*, 11177−11187.
(31) Cooper, V. R.; Thonhauser, T.; Puzder, A.; Schroder, E.; Lundqvist, B. I.; Langreth, D. C. *J. Am. Chem. Soc.* **2008**, *130*, 1304−1308.
(32) Czyznikowska, Z.; Gora, R. W.; Zalesny, R.; Lipkowski, P.; Jarzembska, K. N.; Dominiak, P. M.; Leszczynski, J. *J. Phys. Chem. B* **2010**, *114*, 9629−9644.
(33) Cysewski, P.; Czyznikowska, Z.; Zalesny, R.; Czelen, P. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2665−2672.
(34) Swart, M.; van der Wijst, T.; Guerra, C. F.; Bickelhaupt, F. M. *J. Mol. Model.* **2007**, *13*, 1245−1257.
(35) Langner, K. M.; Sokalski, W. A.; Leszczynski, J. *J. Chem. Phys.* **2007**, *127*, 111102−111105.
(36) Noguera, M.; Rios-Font, R.; Rodriguez-Santiago, L.; Solans-Monfort, X.; Oliva, A.; Bertran, J.; Sodupe, M. *Theor. Chem. Acc.* **2009**, *123*, 105−111.
(37) Svozil, D.; Sponer, J. E.; Marchan, I.; Perez, A.; Cheatham, T. E.; Forti, F.; Luque, F. J.; Orozco, M.; Sponer, J. *J. Phys. Chem. B* **2008**, *112*, 8188−8197.
(38) Mladek, A.; Sponer, J. E.; Jurecka, P.; Banas, P.; Otyepka, M.; Svozil, D.; Sponer, J. *J. Chem. Theory Comput.* **2010**, *6*, 3817−3835.

(39) Foloppe, N.; MacKerell, A. D. *J. Phys. Chem. B* **1999**, *103*, 10955−10964.
(40) MacKerell, A. D. *J. Phys. Chem. B* **2009**, *113*, 3235−3244.
(41) Bosch, D.; Foloppe, N.; Pastor, N.; Pardo, L.; Campillo, M. *J. Mol. Struct.-THEOCHEM* **2001**, *537*, 283−305.
(42) Wang, F. F.; Gong, L. D.; Zhao, D. X. *J. Mol. Struct.-THEOCHEM* **2009**, *909*, 49−56.
(43) Leulliot, N.; Ghomi, M.; Scalmani, G.; Berthier, G. *J. Phys. Chem. A* **1999**, *103*, 8716−8724.
(44) Shishkin, O. V.; Gorb, L.; Zhikol, O. A.; Leszczynski, J. *J. Biomol. Struct. Dyn.* **2004**, *22*, 227−243.
(45) Shishkin, O. V.; Gorb, L.; Zhikol, O. A.; Leszczynski, J. *J. Biomol. Struct. Dyn.* **2004**, *21*, 537−553.
(46) Shishkin, O. V.; Palamarchuk, G. V.; Gorb, L.; Leszczynski, J. *J. Phys. Chem. B* **2006**, *110*, 4413−4422.
(47) Millen, A. L.; Manderville, R. A.; Wetmore, S. D. *J. Phys. Chem. B* **2010**, *114*, 4373−4382.
(48) Churchill, C. D. M.; Wetmore, S. D. *Phys. Chem. Chem. Phys.* **2011**, *13*, 16373−16383.
(49) Millen, A. L.; Kamenz, B. L.; Leavens, F. M. V.; Manderville, R. A.; Wetmore, S. D. *J. Phys. Chem. B* **2011**, *115*, 12993−13002.
(50) Denning, E. J.; Priyakumar, U. D.; Nilsson, L.; Mackerell, A. D. *J. Comput. Chem.* **2011**, *32*, 1929−1943.
(51) Denning, E. J.; MacKerell, A. D. *J. Am. Chem. Soc.* **2012**, *134*, 2800−2806.
(52) Sponer, J.; Mladek, A.; Sponer, J. E.; Svozil, D.; Zgarbova, M.; Banas, P.; Jurecka, P.; Otyepka, M. *Phys. Chem. Chem. Phys.* **2012**, *14*, 15257−15277.
(53) Mladek, A.; Krepl, M.; Svozil, D.; Cech, P.; Otyepka, M.; Banas, P.; Zgarbova, M.; Jurecka, P.; Sponer, J. *Phys. Chem. Chem. Phys.* **2013**, *15*, 7295−7310.
(54) Banas, P.; Hollas, D.; Zgarbova, M.; Jurecka, P.; Orozco, M.; Cheatham, T. E.; Sponer, J.; Otyepka, M. *J. Chem. Theory Comput.* **2010**, *6*, 3836−3849.
(55) Perez, A.; Marchan, I.; Svozil, D.; Sponer, J.; Cheatham, T. E.; Laughton, C. A.; Orozco, M. *Biophys. J.* **2007**, *92*, 3817−3829.
(56) Zgarbova, M.; Luque, F. J.; Sponer, J.; Cheatham, T. E.; Otyepka, M.; Jurecka, P. *J. Chem. Theory Comput.* **2013**, *9*, 2339−2354.
(57) Zgarbova, M.; Otyepka, M.; Sponer, J.; Mladek, A.; Banas, P.; Cheatham, T. E.; Jurecka, P. *J. Chem. Theory Comput.* **2011**, *7*, 2886−2902.
(58) Yildirim, I.; Stern, H. A.; Kennedy, S. D.; Tubbs, J. D.; Turner, D. H. *J. Chem. Theory Comput.* **2010**, *6*, 1520−1531.
(59) Ode, H.; Matsuo, Y.; Neya, S.; Hoshino, T. *J. Comput. Chem.* **2008**, *29*, 2531−2542.
(60) Mladek, A.; Sponer, J. E.; Kulhanek, P.; Lu, X. J.; Olson, W. K.; Sponer, J. *J. Chem. Theory Comput.* **2012**, *8*, 335−347.
(61) Richardson, J. S.; Schneider, B.; Murray, L. W.; Kapral, G. J.; Immormino, R. M.; Headd, J. J.; Richardson, D. C.; Ham, D.; Hershkovits, E.; Williams, L. D.; Keating, K. S.; Pyle, A. M.; Micallef, D.; Westbrook, J.; Berman, H. M. *RNA* **2008**, *14*, 465−481.
(62) Neidle, S. *Principles of Nucleic Acid Structure*; Academia Press, Boston, USA, 2008, 20−33.
(63) Altona, C.; Sundaralingam, M. *J. Am. Chem. Soc.* **1972**, *94*, 8205−8212.
(64) Jaeger, L.; Verzemnieks, E. J.; Geary, C. *Nucleic Acids Res.* **2009**, *37*, 215−230.
(65) Hershkovitz, E.; Tannenbaum, E.; Howerton, S. B.; Sheth, A.; Tannenbaum, A.; Williams, L. D. *Nucleic Acids Res.* **2003**, *31*, 6249−6257.
(66) Hsiao, C.; Mohan, S.; Hershkovitz, E.; Tannenbaum, A.; Williams, L. D. *Nucleic Acids Res.* **2006**, *34*, 1481−1491.
(67) Murray, L. J. W.; Arendall, W. B.; Richardson, D. C.; Richardson, J. S. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 13904−13909.
(68) Schneider, B.; Moravek, Z.; Berman, H. M. *Nucleic Acids Res.* **2004**, *32*, 1666−1677.
(69) Staroverov, V. N.; Scuseria, G. E.; Tao, J. M.; Perdew, J. P. *J. Chem. Phys.* **2003**, *119*, 12129−12137.
(70) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. *J. Chem. Phys.* **2010**, *132*, 154104−154122.
(71) Grimme, S.; Ehrlich, S.; Goerigk, L. *J. Comput. Chem.* **2011**, *32*, 1456−1465.
(72) Schafer, A.; Huber, C.; Ahlrichs, R. *J. Chem. Phys.* **1994**, *100*, 5829−5835.
(73) Weigend, F.; Ahlrichs, R. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297−3305.
(74) Rappoport, D.; Furche, F. *J. Chem. Phys.* **2010**, *133*, 134105−134115.
(75) Jensen, F. *J. Chem. Theory Comput.* **2010**, *6*, 2726−2735.
(76) Sen, K. D. *Chem. Phys. Lett.* **1980**, *74*, 201−202.
(77) Shore, H. B.; Rose, J. H.; Zaremba, E. *Phys. Rev. B* **1977**, *15*, 2858−2861.
(78) Vydrov, O. A.; Scuseria, G. E. *J. Chem. Phys.* **2005**, *122*, 184107−184113.
(79) Cole, L. A.; Perdew, J. P. *Phys. Rev. A* **1982**, *25*, 1265−1271.
(80) Guo, Y. F.; Whitehead, M. A. *Phys. Rev. A* **1989**, *40*, 28−34.
(81) Eichkorn, K.; Treutler, O.; Ohm, H.; Haser, M.; Ahlrichs, R. *Chem. Phys. Lett.* **1995**, *240*, 283−289.
(82) Eichkorn, K.; Weigend, F.; Treutler, O.; Ahlrichs, R. *Theor. Chem. Acc.* **1997**, *97*, 119−124.
(83) Peng, C. Y.; Ayala, P. Y.; Schlegel, H. B.; Frisch, M. J. *J. Comput. Chem.* **1996**, *17*, 49−56.
(84) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Rob, M. A.; Cheeseman, J. R.; Jr., J., A. M.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A.; *Gaussian 03*; Gaussian, Inc.: Wallingford, CT, 2003.
(85) Ahlrichs, R.; Bar, M.; Haser, M.; Horn, H.; Kolmel, C. *Chem. Phys. Lett.* **1989**, *162*, 165−169.
(86) Klamt, A. *WIRES-Comput. Mol. Sci.* **2011**, *1*, 699−709.
(87) Klamt, A.; Schuurmann, G. *J. Chem. Soc.-Perkin Trans. 2* **1993**, 799−805.
(88) Halkier, A.; Helgaker, T.; Jorgensen, P.; Klopper, W.; Koch, H.; Olsen, J.; Wilson, A. K. *Chem. Phys. Lett.* **1998**, *286*, 243−252.
(89) Halkier, A.; Helgaker, T.; Jorgensen, P.; Klopper, W.; Olsen, J. *Chem. Phys. Lett.* **1999**, *302*, 437−446.
(90) Dunning, T. H. *J. Chem. Phys.* **1989**, *90*, 1007−1023.
(91) Woon, D. E.; Dunning, T. H. *J. Chem. Phys.* **1993**, *98*, 1358−1371.
(92) Sponer, J.; Jurecka, P.; Hobza, P. *J. Am. Chem. Soc.* **2004**, *126*, 10142−10151.
(93) Sponer, J.; Riley, K. E.; Hobza, P. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2595−2610.
(94) Becke, A. D. *Phys Rev A* **1988**, *38*, 3098−3100.
(95) Lee, C. T.; Yang, W. T.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785−789.
(96) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865−3868.
(97) Zhang, Y. K.; Yang, W. T. *Phys. Rev. Lett.* **1998**, *80*, 890−890.
(98) Goerigk, L.; Grimme, S. *J. Chem. Theory Comput.* **2010**, *6*, 107−126.
(99) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648−5652.
(100) Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2005**, *109*, 5656−5667.
(101) Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2004**, *108*, 6908−6918.
(102) Goerigk, L.; Grimme, S. *J. Chem. Theory Comput.* **2011**, *7*, 291−309.

(103) Kozuch, S.; Gruzman, D.; Martin, J. M. L. *J. Phys. Chem. C* **2010**, *114*, 20801−20808.
(104) Grimme, S. *J. Chem. Phys.* **2006**, *124*, 034108−034123.
(105) Goerigk, L.; Grimme, S. *Phys. Chem. Chem. Phys.* **2011**, *13*, 6670−6688.
(106) Fradera, X.; Austen, M. A.; Bader, R. F. W. *J. Phys. Chem. A* **1999**, *103*, 304−314.
(107) Bader, R. F. W. *Chem. Rev.* **1991**, *91*, 893−928.
(108) Bader, R. F. W. *Atoms in Molecules. A Quantum Theory*; Oxford University Press: Oxford, U.K., 1990; 13−343.
(109) Biegler-Konig, F.; Schonbohm, J.; Bayles, D. *J. Comput. Chem.* **2001**, *22*, 545−559.
(110) Biegler-Konig, F.; Schonbohm, J. *J. Comput. Chem.* **2002**, *23*, 1489−1494.
(111) Lane, J. R.; Contreras-Garcia, J.; Piquemal, J.-P.; Miller, B. J.; Kjaergaard, H. G. *J. Chem. Theory Comput.* **2013**, *9*, 3263−3266.
(112) Foster, J. P.; Weinhold, F. *J. Am. Chem. Soc.* **1980**, *102*, 7211−7218.
(113) Reed, A. E.; Weinhold, F. *J. Chem. Phys.* **1983**, *78*, 4066−4073.
(114) Reed, A. E.; Weinstock, R. B.; Weinhold, F. *J. Chem. Phys.* **1985**, *83*, 735−746.
(115) Reed, A. E.; Weinhold, F. *J. Chem. Phys.* **1985**, *83*, 1736−1740.
(116) Reed, A. E.; Curtiss, L. A.; Weinhold, F. *Chem. Rev.* **1988**, *88*, 899−926.
(117) 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.
(118) Case, D. A.; Darden, T. A.; Chetham, T. E., III; Simmerling, C. E.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B. P.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossvary, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M. J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, V.; Kovalenko, A.; Kollmann, P. A. *AMBER 11*University of California: San Francisco, 2010.
(119) Luo, R.; David, L.; Gilson, M. K. *J. Comput. Chem.* **2002**, *23*, 1244−1253.
(120) Lu, Q.; Luo, R. *J. Chem. Phys.* **2003**, *119*, 11035−11047.
(121) Perez, A.; Luque, F. J.; Orozco, M. *J. Am. Chem. Soc.* **2007**, *129*, 14739−14745.
(122) Banas, P.; Walter, N. G.; Sponer, J.; Otyepka, M. *J. Phys. Chem. B* **2010**, *114*, 8701−8712.
(123) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.
(124) Aqvist, J. *J. Phys. Chem.* **1990**, *94*, 8021−8024.
(125) Joung, I. S.; Cheatham, T. E., III *J. Phys. Chem. B* **2008**, *112*, 9020−9041.
(126) Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684−3690.
(127) Burns, L. A.; Vazquez-Mayagoitia, A.; Sumpter, B. G.; Sherrill, C. D. *J. Chem. Phys.* **2011**, *134*, 084107.
(128) Sponer, J.; Mladek, A.; Spackova, N.; Cang, X. H.; Cheatham, T. E., III; Grimme, S. *J. Am. Chem. Soc.* **2013**, *135*, 9785−9796.
(129) Goerigk, L.; Kruse, H.; Grimme, S. *Chem. Phys. Chem* **2011**, *12*, 3421−3433.
(130) Perdew, J. P. *Proc. 21st Ann. Int. Symp. Electronic Struct. Sol.*; Akademie Verlag: Berlin, 1991, 11−20.
(131) Becke, A. D. *J. Chem. Phys.* **1996**, *104*, 1040−1046.
(132) Jurecka, P.; Sponer, J.; Cerny, J.; Hobza, P. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1985−1993.
(133) Sponer, J.; Cang, X.; Cheatham, T. E., III *Methods* **2012**, *57*, 25−39.
(134) Auffinger, P.; Westhof, E. *J. Mol. Biol.* **1997**, *274*, 54−63.
(135) Singh, S. B.; Kollman, P. A. *Biophys. J.* **1996**, *70*, 1940−1948.
(136) Bolton, P. H.; Kearns, D. R. *Biochim. Biophys. Acta* **1978**, *517*, 329−337.
(137) Gyi, J. I.; Lane, A. N.; Conn, G. L.; Brown, T. *Nucleic Acids Res.* **1998**, *26*, 3104−3110.
(138) Fohrer, J.; Hennig, M.; Carlomagno, T. *J. Mol. Biol.* **2006**, *356*, 280−287.
(139) Ying, J. F.; Bax, A. *J. Am. Chem. Soc.* **2006**, *128*, 8372−8373.
(140) Fonville, J. M.; Swart, M.; Vokáčová, Z.; Sychrovský, V.; Šponer, J. E.; Šponer, J.; Hilbers, C. W.; Bickelhaupt, F. M.; Wijmenga, S. S. *Chem.—Eur. J.* **2012**, *18*, 12372−12387.
(141) Fadrna, E.; Spackova, N.; Sarzynska, J.; Koca, J.; Orozco, M.; Cheatham, T. E., III; Kulinski, T.; Sponer, J. *J. Chem. Theory Comput.* **2009**, *5*, 2514−2530.
(142) Jang, S. B.; Baeyens, K.; Jeong, M. S.; SantaLucia, J.; Turner, D.; Holbrook, S. R. *Acta Crystallogr. D* **2004**, *60*, 829−835.
(143) Mackerell, A. D.; Wiorkiewiczkuczera, J.; Karplus, M. *J. Am. Chem. Soc.* **1995**, *117*, 11946−11975.
(144) Sponer, J. E.; Reblova, K.; Mokdad, A.; Sychrovsky, V.; Leszczynski, J.; Sponer, J. *J. Phys. Chem. B* **2007**, *111*, 9153−9164.
(145) Sponer, J. E.; Spackova, N.; Leszczynski, J.; Sponer, J. *J. Phys. Chem. B* **2005**, *109*, 11399−11410.
(146) Lu, X. J.; Olson, W. K.; Bussemaker, H. J. *Nucleic Acids Res.* **2010**, *38*, 4868−4876.