JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Article

pubs.acs.org/JACS

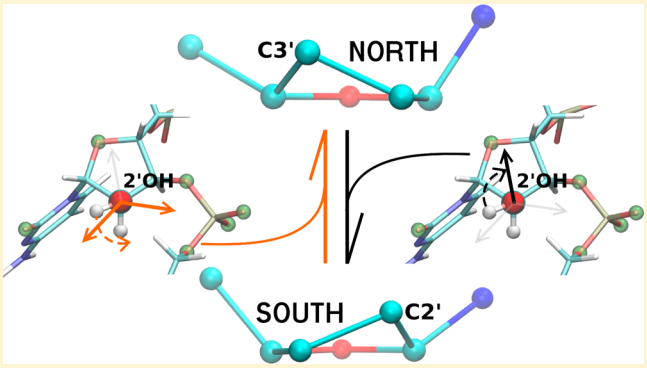
Small Details Matter: The 2′-Hydroxyl as a Conformational Switch in RNA

Leonardo Darré,[†,‡] Ivan Ivani,[†,‡] Pablo D. Dans,[†,‡] Hansel Gómez,[†,‡] Adam Hospital,[†,‡] and Modesto Orozco*[,†,‡,§]

[†]Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology, 08028 Barcelona, Spain
[‡]Joint BSC-IRB Program in Computational Biology, Institute for Research in Biomedicine, 08028 Barcelona, Spain
[§]Department of Biochemistry and Biomedicine, Faculty of Biology, University of Barcelona, 08028 Barcelona, Spain

Supporting Information

ABSTRACT: While DNA is mostly a primary carrier of genetic information and displays a regular duplex structure, RNA can form very complicated and conserved 3D structures displaying a large variety of functions, such as being an intermediary carrier of the genetic information, translating such information into the protein machinery of the cell, or even acting as a chemical catalyst. At the base of such functional diversity is the subtle balance between different backbone, nucleobase, and ribose conformations, finely regulated by the combination of hydrogen bonds and stacking interactions. Although an apparently simple chemical modification, the presence of the 2′OH in RNA has a profound effect in the ribonucleotide conformational balance, adding an extra layer of complexity to the interactions network in RNA. In the present work, we have combined database analysis with extensive molecular dynamics, quantum mechanics, and hybrid QM/MM simulations to provide direct evidence on the dramatic impact of the 2′OH conformation on sugar puckering. Calculations provide evidence that proteins can modulate the 2′OH conformation to drive sugar repuckering, leading then to the formation of bioactive conformations. In summary, the 2′OH group seems to be a primary molecular switch contributing to specific protein−RNA recognition.



INTRODUCTION

There is general consensus that life originated in an RNA-world, as this oligonucleotide is a very versatile entity that is able to self-replicate, transmitting information to descendants, and at the same time adopt complex three-dimensional structures, acting as catalyzers of complex reactions. However, at an early point of evolution, DNA was selected as the primary carrier of genetic information, while RNA maintained a myriad of other functions, the most important ones related to translating DNA information into protein sequence. Although DNA presents several non-canonical structures (such as triplexes, quadruplexes, H-junctions, and others),[1] it is most often found as a self-complementary right-handed double helix. RNA, which in physiological conditions is single-stranded, displays a more complex conformational landscape where double-helix fragments are linked by single-stranded segments and flanked by different kinds of loops, bulges, flipped bases, and non-canonical base pairs. These motifs form modular secondary structure domains that combine in very complex three-dimensional structures, some of them, of large biological impact and having been exquisitely refined by evolution,[2−6] displaying globular structures that are not so common in DNA (see Supplementary Figure 1). This conformational richness is likely to be mandatory for the large variety of biological functions of RNA.

Despite their coexistence in some cellular organelles, nature has completely separated DNA and RNA functional spaces, which is quite surprising considering the minuscule chemical differences between them: the presence/absence of one methyl group at position 5 of uridine, and the presence/absence of a hydroxyl at position 2′ of the sugar. The latter add several distinctive hydrogen bond interactions (2′OH−nucleobase or 2′OH−phosphate contacts) that might contribute to the stabilization of non-helical motifs and that can modify conformational preferences of the nucleotide. In fact, there is a general consensus that the presence of the 2′OH drives the puckering preferences of the sugar from south (S, C2′endo) to north (N, C3′endo) conformations, which is known to drive a global conformational change from the B- to the A-form.[7] However, our understanding of the connection between the rotational state of the C2′−O2′ bond and the local and global

Received: September 8, 2016
Published: November 28, 2016

DOI: 10.1021/jacs.6b09471
J. Am. Chem. Soc. 2016, 138, 16355−16363

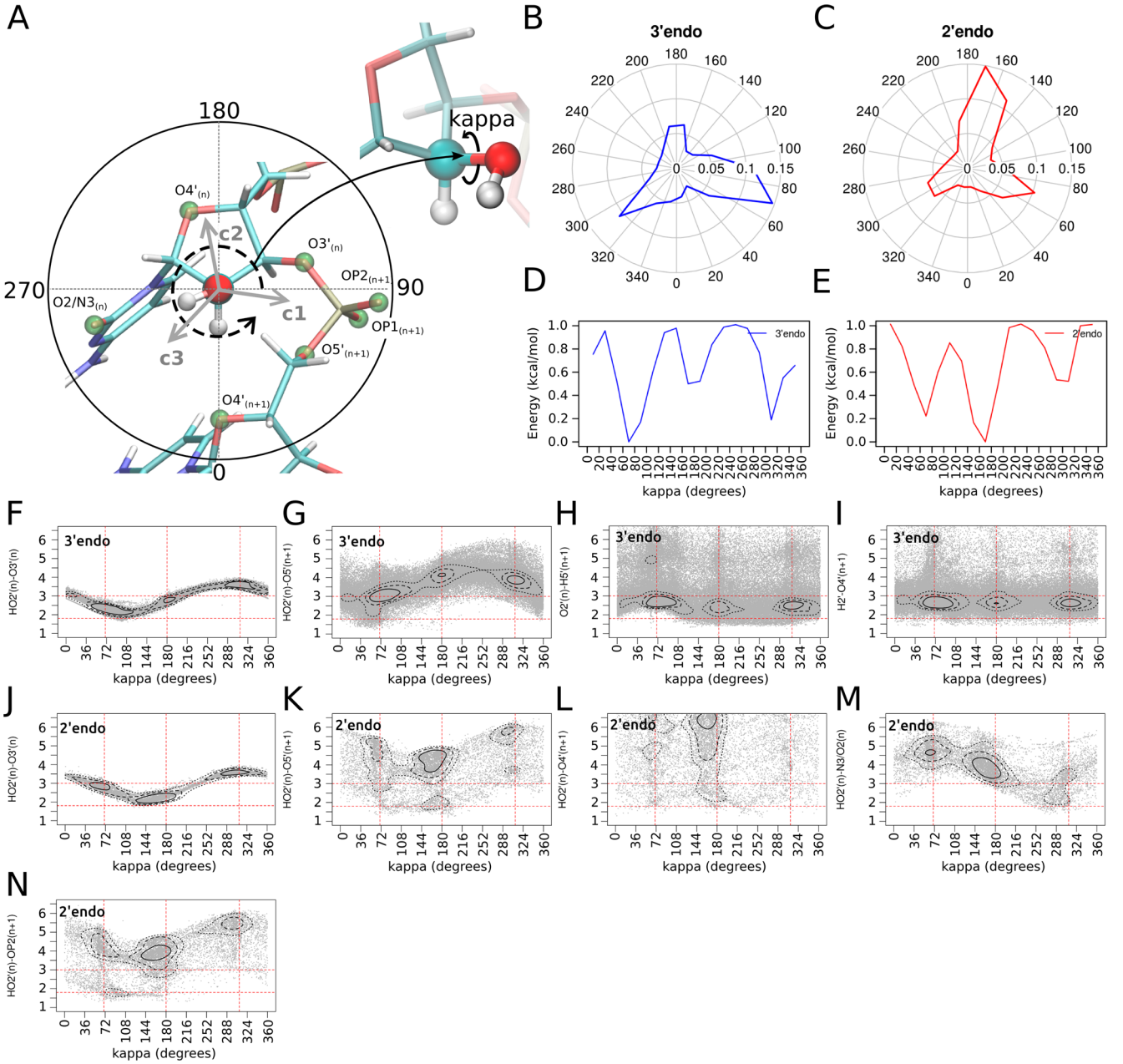


Figure 1. Kappa torsion preferred orientations from the Protein Data Bank. (A) Nucleotide in a RNA strand indicating the possible orientations of the 2′OH and the loction of neighboring hydrogen bond acceptors/donors. (B) Probability distribution of the torsion angle between the atoms H2′−C2′−O2′−HO2′ for all the 3′endo ribonucleotides of the RNA dataset obtained from the current state of the PDB. (C) Same as in (A) but for 2′endo ribonucleotides. (D,E) Empirical free energy calculated from the experimental κ distributions in (B) and (C), respectively. Scatter plots of κ torsion vs distance between HO2′ and local acceptors/donors of hydrogen bonds are shown for nucleotides with pucker phase in 3′endo (F−I) and 2′endo (J−N). Red dotted lines indicate optimal and maximum hydrogen bond distances (horizontal) and κ rotation minimum energy positions (vertical). Contour lines correspond to points with density values equal to the average density plus 1 (dotted line), 2 (dashed line) and 4 (continuous line) standard deviations.

conformation of the RNA is still rather limited. In the A-form (sugar in north), the 2′OH could adopt three preferred orientations, pointing toward the O3′ atom (*gauche*+ measured from H2′), the nucleobase (*gauche*-), or the O4′ atom (*trans*).[8] The first two are the most frequent ones, according to NMR[9] and quantum mechanics (QM) calculations,[10] and subject to orientation-specific hydration. When water interacts with the 2′OH in the base orientation the A-form is stabilized,[9,11] while non-canonical conformations gain stabilization from water molecules interacting with the 2′OH in the O3′ orientation.[12] The less frequent south sugar conformation has received less attention but is believed to favor a 2′OH oriented mainly toward the O3′ atom but with a C2′−O2′ torsion shifted to *trans* orientation.[8,10]

In the present work, a combination of database analysis, atomistic molecular dynamics (MD), high-level QM, and hybrid quantum mechanics/molecular mechanics (QM/MM) calculations was used to explore in detail the conformational preferences of the C2′OH bond and its specific impact on the sugar puckering, which in turn, defines the RNA conformation.

DOI: 10.1021/jacs.6b09471
J. Am. Chem. Soc. 2016, 138, 16355−16363

The 2′OH rotation is found to bias the sugar pucker preference, evidencing its role as a major determinant of RNA conformation and a molecular switch, which can be tuned by proteins and other effectors to induce changes on the RNA structure.

METHODS

Database Mining. All NMR-solved RNA structures deposited in the Protein Data Bank (PDB) were analyzed (see Supplementary Methods 1), accounting for a total of 174 511 2′OH groups, 115 513 with the ribose in north puckering ($0 \leq$ pucker phase ≤ 36) and 11 626 with the ribose in the unusual south puckering ($144 \leq$ pucker phase ≤ 180). The orientation around the C2′−O2′ torsion (herein called kappa, κ) was defined using the atoms H2′−C2′−O2′−HO2′, following Auffinger and Westhof[8] (see Figure 1A). In order to analyze the potential role of the 2′OH group in modulating protein−RNA interactions and the connection with the RNA local conformation, we performed additional analysis using only RNA−protein complexes, solved again by NMR. Additionally we explored heavy-atom contacts involving the 2′OH group considering not only NMR but also X-ray (resolution ≤2.5 Å) protein−RNA complex structures, which means exploration of 26 760 2′OH groups from 500 PDB entries. Additional details of the database analysis can be found in Supplementary Methods 1. To double-check the observations made from the datasets mentioned above, the same analysis was repeated using a non-redundant database[13] containing both NMR and X-ray (resolution ≤ 2.5 Å) solved structures (see Supplementary Tables 1−3 for details).

Quantum Simulations. The pseudo-rotational profile of ribose was first explored along the north ↔ east ↔ south transition path. To avoid discontinuities in the energy profiles, geometry optimizations at each point were performed keeping β, γ, ε, and χ torsions at their standard values in RNAs. In the case of χ, the dependence on the sugar puckering was taken into account, setting $\chi_N = 190°$ and $\chi_S = 230°$. To explore whether pseudo-rotation was dependent on the orientation of the C2′−O2′ bond, profiles were calculated fixing the κ angle at three typical values (72°, 178°, and 306°; the most populated values found in our database analysis). Energy profiles were obtained at the B3LYP/6-31++G(d,p) level, and selected points were refined at the MP2/aug-cc-pVDZ level. All profiles were obtained in water as simulated by the IEFPCM continuum method.[14]

Analysis of electron distribution using Bader's atoms in molecules (AIM) theory[15−17] was performed on reduced clusters representative of the most prevalent orientations of the κ angle (three replicas per relevant κ orientation). Single-point calculations at the MP2(FC)/6-31G(d,p) level were performed at the dinucleotide level, removing the base at 3′ and completing the valence of the C1′, O5′, and O3′ atoms with H atoms. This analysis allowed us to explore the potential formation and intensity of canonical O−H···X (for X = O) or non-canonical X−H···O (for X = C) hydrogen bonds by searching for bond critical points connecting such atoms and quantifying the associated electron density. The AIM-UC package[18] was used for the AIM analysis.

Additional QM/SCRF calculations were performed to determine the impact of the presence of a cationic group in the vicinities of the O2′ group on the κ vs puckering energy bi-dimensional map (see Supplementary Methods 6 for detailed explanation of these QM calculations).

Classical Simulations. A large variety of MD simulations were performed to analyze the connection between RNA and C2′O2′ conformation and assess the reliability of a state-of-the-art force field for RNA. They include the following: (i) standard simulations in hairpin and kissing loop RNA motives; (ii) potentials of mean force (PMFs) of the κ rotation at the dinucleotide (rCpC) level using umbrella sampling (US) with an 18° interval grid of the κ torsional space (500 ps equilibration and 2.5 ns of averaging per window); and (iii) Hamiltonian-replica exchange molecular dynamics (H-REMD) to evaluate the conformational landscape of two small RNA tetranucleotides (rGACC and rCCCC) for which experimental structural data in solution are available.[19,20] All calculations were performed using the parm99 force field[21,22] supplemented with the bsc0[23] and chiOL3[24,25] modifications for RNA; some control simulations were performed with a local experimental RNA version of the parmbsc1 force field.[26] Electroneutrality was achieved by adding K^+ and extra K^+Cl^- to generate a 150 mM concentration (taking Dang's parameters[27−29] to represent ions). US calculations were performed to determine the free energy associated with S ↔ N conformational transition in a protein−RNA (MIWI PAZ domain bound to RNA; PDB ID 2XFM model 6) complex in the wild type, where Lys316 is in the vicinity of O2′, and a mutant protein, where Lys316 is substituted by an alanine (K316A mutant). Details of these calculations can be found in Supplementary Methods 2.

QM/MM Simulations. Hybrid QM/MM simulations were used extensively to analyze the free energy profile of the C2′O2′ rotation for an isolated rC nucleoside and an r(CpC) dinucleotide in aqueous solution. The BLYP/6-31G(d) functional was used to represent the nucleic acid, while the solvent was represented at the classical level. US free energy profiles were computed by scanning in 18° intervals the κ torsional space (5 ps equilibration and 40 or 25 ps of averaging per window for the nucleoside rC or the dinucleotide rCpC, respectively). Extended descriptions of QM/MM simulations can be found in Supplementary Methods 3.

RESULTS AND DISCUSSION

C2′O2′ Torsion Experimental Distribution. The 2′OH group of a ribose in the major north conformation (ratio N:S is around 10:1 in the database) samples three rotational states in NMR-PDB (Figure 1B): (i) the κ region between 40° and 140° (peak at ∼70°; conformer 1), (ii) the κ region between 140° and 240° (peak at ∼180°; conformer 2), and (iii) the κ region between 240° and 40° (peak at ∼310°; conformer 3). Transforming populations into conformational free energies (Figure 1D) points to a nearly barrier-less rotation, with three minima of free energies of 0 (conformer 1), ∼0.5 (conformer 2), and ∼0.2 kcal/mol (conformer 3). Interestingly, no significant differences are found in the κ torsional distribution for the four ribonucleotides (Supplementary Figure 2), suggesting that base−sugar contacts are not crucial to determine the 2′OH group orientation (see below). Contact analysis reveals some interactions that appear in all the conformations of this set of structures, such as a non-canonical C5′$_{(n+1)}$−H5′$_{(n+1)}$···O2′ hydrogen bond and the non-canonical C2′−H2′···O4′ $_{(n+1)}$ hydrogen bond previously reported by Auffinger and Westhof,[8] while others, like the strong O2′−HO2′···O3′ hydrogen bond, appear only in conformer 1 (Figure 1F−I). Close contacts between the 2′OH group and the nucleobase, or the OP1/2 groups, are uncommon in experimental structures of north riboses (Supplementary Figure 3). Conformer 2, which was the least populated orientation for north puckering, becomes dominant for south riboses, probably due to the formation of O2′−HO2′···O3′ hydrogen bonds (Figure 1J−N). Conformer 1 instead becomes the second most populated orientation and conformer 3 the least populated one (see Figure 1C,E). This is reflected on the relative free energy difference between the conformers 1, 2, and 3 (∼0.2, 0, and ∼0.5 kcal/mol, respectively; see Figure 1E). Some variability (∼0.1 kcal/mol) is observed in the relative energy values, depending on partitioning of the κ coordinate, in particular for the south-puckering profile (Supplementary Figure 5C,D); however, the overall trend remains consistent. Furthermore, when the calculation is repeated for the non-redundant dataset, equivalent results are obtained (Supplementary Figures 4 and 5A,B).

To gain additional information, we focus our study on those 2′OH's interacting with protein residues. As for the κ

DOI: 10.1021/jacs.6b09471
J. Am. Chem. Soc. 2016, 138, 16355−16363

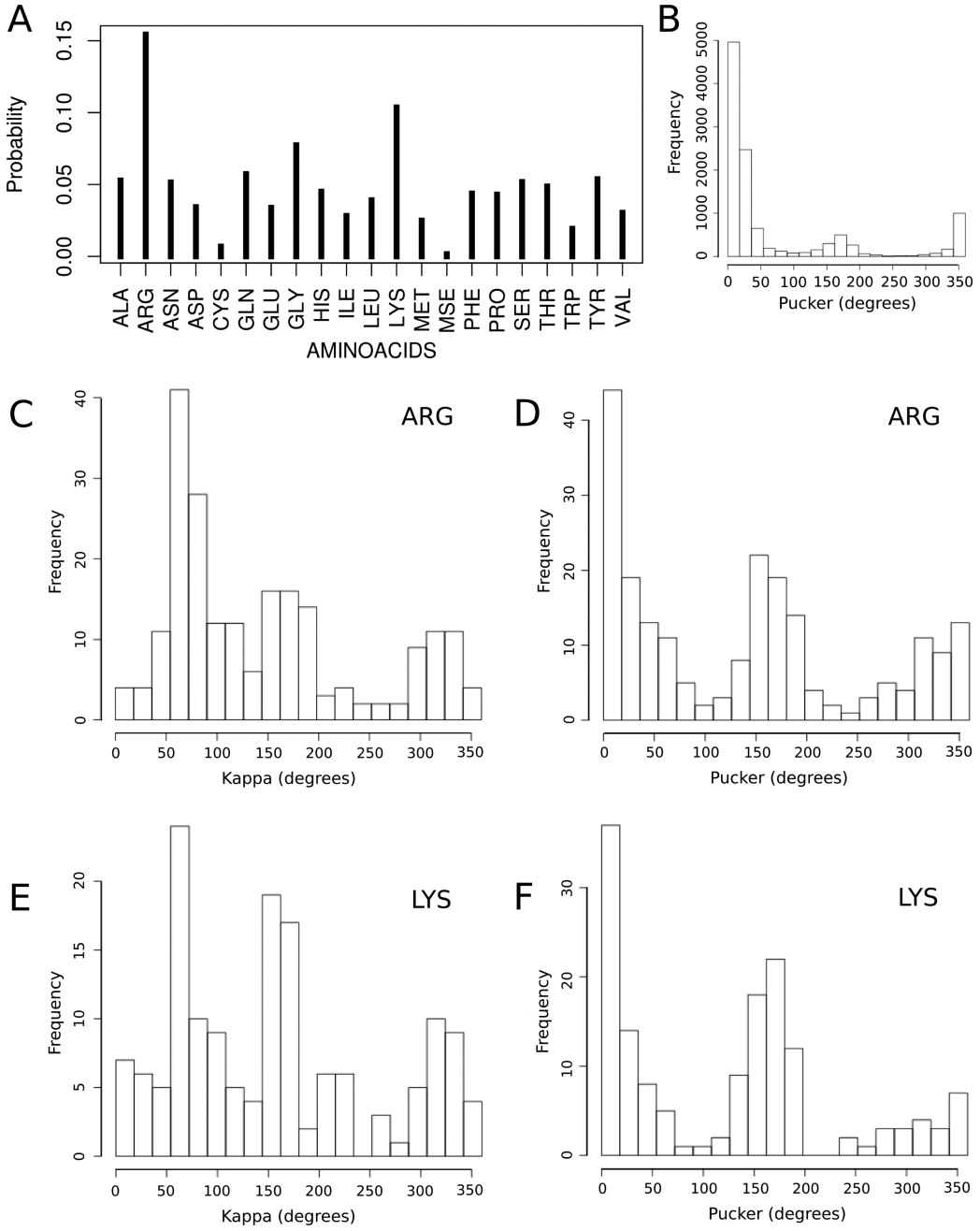


Figure 2. Protein−RNA contacts. (A) Probability of contact between a given amino acid and the 2′OH, given a protein−RNA contact occurs, calculated by counting all contacts (distance ≤4 Å) between any protein atom and the oxygen of 2′OH, and splitting the counts per amino acid identity. Multiple atoms of a given amino acid within the distance cutoff were counted as one contact. X-ray and NMR chains and models specified in the non-redundant dataset were used; see Supplementary Methods 1 for details. (B) Frequency of pucker phase values for all RNA nucleotides obtained from NMR structures in the non-redundant dataset. (C) Frequency of κ values for RNA nucleotides in contact with ARG atoms (distance ≤4 Å) obtained from NMR structures in the non-redundant dataset. (D) Frequency of pucker phase values for RNA nucleotides in contact (distance ≤4 Å) with ARG atoms obtained from NMR structures in the non-redundant dataset. (E,F) Same as (C) and (D), respectively, but for LYS amino acid.

distribution analysis, both the full and the non-redundant datasets were originally used. However, although qualitatively similar trends are observed, differences between the two datasets point toward some bias in the full dataset, which leads us to discuss below only results from the non-redundant dataset (see Figure 2; results from the full dataset can be found in Supplementary Figure 6). Figure 2A shows that Lys and Arg are the preferred interacting partners among all amino acids. Interaction of 2′OH with these protein side chains leads to a stabilization of *conformers 1 and 2* and a parallel enrichment in south puckering (Figure 2B−E). Altogether, analysis of experimental databases strongly suggests that sugar puckering and C2′OH rotational states are coupled, and that proteins interacting with the C2′OH can modulate the sugar puckering by biasing κ torsional preferences, which can lead to global structural changes in RNA. Cluster analysis of the localization of the Lys and Arg hydrogen bond donor nitrogen atoms close to south-puckering nucleotides (see Supplementary Methods

DOI: 10.1021/jacs.6b09471
J. Am. Chem. Soc. 2016, 138, 16355−16363

1) indicates a distribution around the O2′ atom that concentrates mainly on two sites. The first site (cluster A) is localized between the phosphate and 2′OH groups, while the second site (cluster B) is in contact only with the 2′OH (Supplementary Figures 10A,B and 11A,B). In the case of cluster A, around 60% (Lys) and 90% (Arg) of the population show the hydrogen bond donor nitrogen very close (<3.5 Å) to the O2′ atom. For cluster B these values increase to 75% (Lys) and 92% (Arg).

Puckering and C2′O2′ Torsion Are Coupled in Ribonucleosides. The database analysis above can be subject to criticism, since the orientation of the C2′−O2′ bond is not directly observed in the spectra but inferred from indirect restraints. Thus, to support our database analysis, we first performed QM studies of the pseudo-rotation profile of ribose for the three C2′O2′ rotational states in dilute aqueous solution (see Methods). For both adenosine and cytosine, in the north state, conformer 1 is the most stable orientation, and conformer 3 is close in energy (∼0.5 kcal/mol), while conformer 2 is disfavored by ∼1.2 kcal/mol (Figure 3A,B). However, as

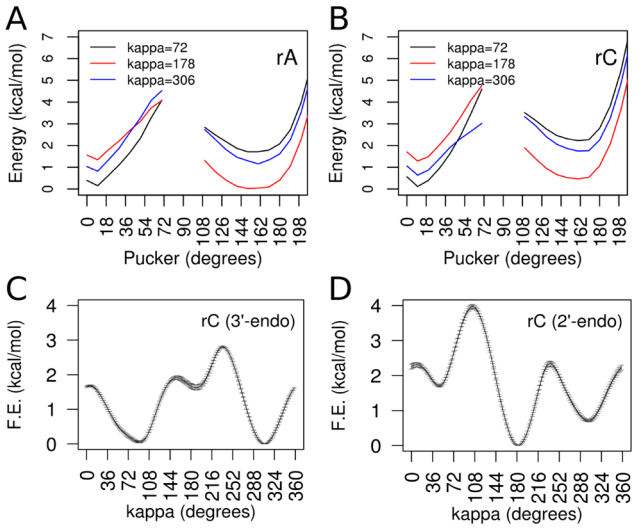


Figure 3. 2′OH κ torsion and sugar pucker phase preferred conformers at the nucleoside level. (A,B) QM potential energy scans of the sugar pucker phase (for adenosine and cytosine, respectively) restraining the κ torsion at the main observed minima in the κ free energy profile (72°, 178°, and 306°). The dependence of χ on the sugar puckering was taken into account by fixing χ = 190° for pucker values in the range 0−70° and χ = 230° for pucker values in the range 110−216°. (C,D) US QM/MM free energy profiles for the κ torsion of a rC nucleoside with restraints on the sugar pucker phase at the 3′endo and 2′endo conformations, respectively. The continuous line and error bars correspond to the average and standard deviation of the free energy, respectively, calculated from the energy profiles obtained after 31, 32, 33, 34, 35, 36, 37, 38, 39, and 40 ps of US simulation.

suggested from database analysis, conformer 2 (poorly populated in the north state) becomes the most stable orientation when the sugar samples the south state. We found it very exciting that, if conformer 2 is forced, north and south relative energies invert, with the latter becoming the most stable sugar puckering state (Figure 3A,B). This suggests that already at the nucleoside level the orientation of the 2′OH group can induce changes in sugar puckering. Very encouragingly, similar results are obtained when flexibility and explicit solvent are considered in QM/MM PMFs of the C2′−O2′ rotation (see Methods and Supplementary Methods 3), with restraints in sugar puckering (see Figure 3C,D). In summary, QM/SCRF and QM/MM calculations provide a picture of the κ torsional space of the nucleoside that qualitatively agrees with the database analysis of RNA motives. Furthermore, it reinforces the idea that C2′O2′ torsion and puckering are coupled and that biasing of the κ torsion can lead to changes in puckering, which in turn dramatically affects the RNA conformation. This effect was also observed when comparing the κ vs puckering QM/SCRF (see Supplementary Methods 6) potential energy surface (PES) of a guanine nucleoside monophosphate in the presence/absence of a Lys analogue placed in the cluster A site (see previous section for the definition of cluster A and Supplementary Figure 8 for the QM/SCRF PES results), supporting the idea that protein cationic side chains act as one of such puckering biasing agents. In fact, the presence of a Lys residue (316) in cluster A of cytosine 6 in the RNA−MIWI PAZ domain complex (PDB ID 2XFM model 6; see Supplementary Methods) dramatically stabilizes the south conformation (see Supplementary Figure 9). This effect is lost when Lys is mutated to Ala (see Supplementary Figure 9), highlighting the importance of the cationic residue in modulating RNA conformation, independent of more global structural effects imposed by the protein.

C2′O2′ Torsion in RNA Oligomers. State-of-the-art simulations discussed above present a major caveat: the neglect of the polynucleotide environment, which can force the approach of different interactors to the 2′OH group, modifying the intrinsic properties of nucleosides described in the previous section. To solve this potential caveat, we computed the QM/MM PMF of the κ rotation in the rCpC dinucleotide in explicit solvent (see Methods and Supplementary Methods 3). Very encouragingly, results in Figure 4A qualitatively agree with the observed preferred orientations for north-puckering riboses in the PDB analysis (Figure 1B,D), with conformer 1 being the global minimum, followed closely by conformer 3, ∼0.3 kcal/mol higher energy, and conformer 2, the least stable, ∼1 kcal/mol above conformer 1. Conformational transitions between conformers 1 and 3 happen through a ∼1.8 kcal/mol free energy barrier localized at the eclipsed $\kappa \approx 0$ value. These values are also consistent with high-level QM calculations in solution for an isolated nucleoside and in astonishing agreement with database analysis. Although these energy barriers are relatively small compared with global RNA−protein interaction free energy, they are comparable with specific RNA−protein residue interactions (e.g., Asn, Gln, and Arg interaction energy with typical nucleotides of −0.8, −1.5, and −0.9 kcal/mol, respectively[30]). Thus, such transitions can affect local structural rearrangements, impacting in the definition of the RNA bioactive conformation. In addition, the 2′OH contacts that were frequent in NMR-refined structures are also frequent in our QM/MM trajectories. Bader's analysis of electron densities in QM/MM snapshots (see Figure 4B−D and Methods) confirms the formation of hydrogen bond interactions, both canonical (O2′−HO2′···O3′: $\rho \approx 0.020$ au) and non-canonical (C5′−H5′···O2′: $\rho \approx 0.009$ au and C2′−H2′···O4′: $\rho \approx 0.010$ au). These electron density values confirm that "non-canonical" O···H−C hydrogen bonds are quite stable (∼2−3 kcal/mol as estimated from the linear relationship between the interaction energy and bond critical point density reported in Cubero et at.[31]), not far from a medium-strength canonical hydrogen bond. This confirms previous claims on the stabilizing role of ribose aliphatic

DOI: 10.1021/jacs.6b09471
J. Am. Chem. Soc. 2016, 138, 16355−16363

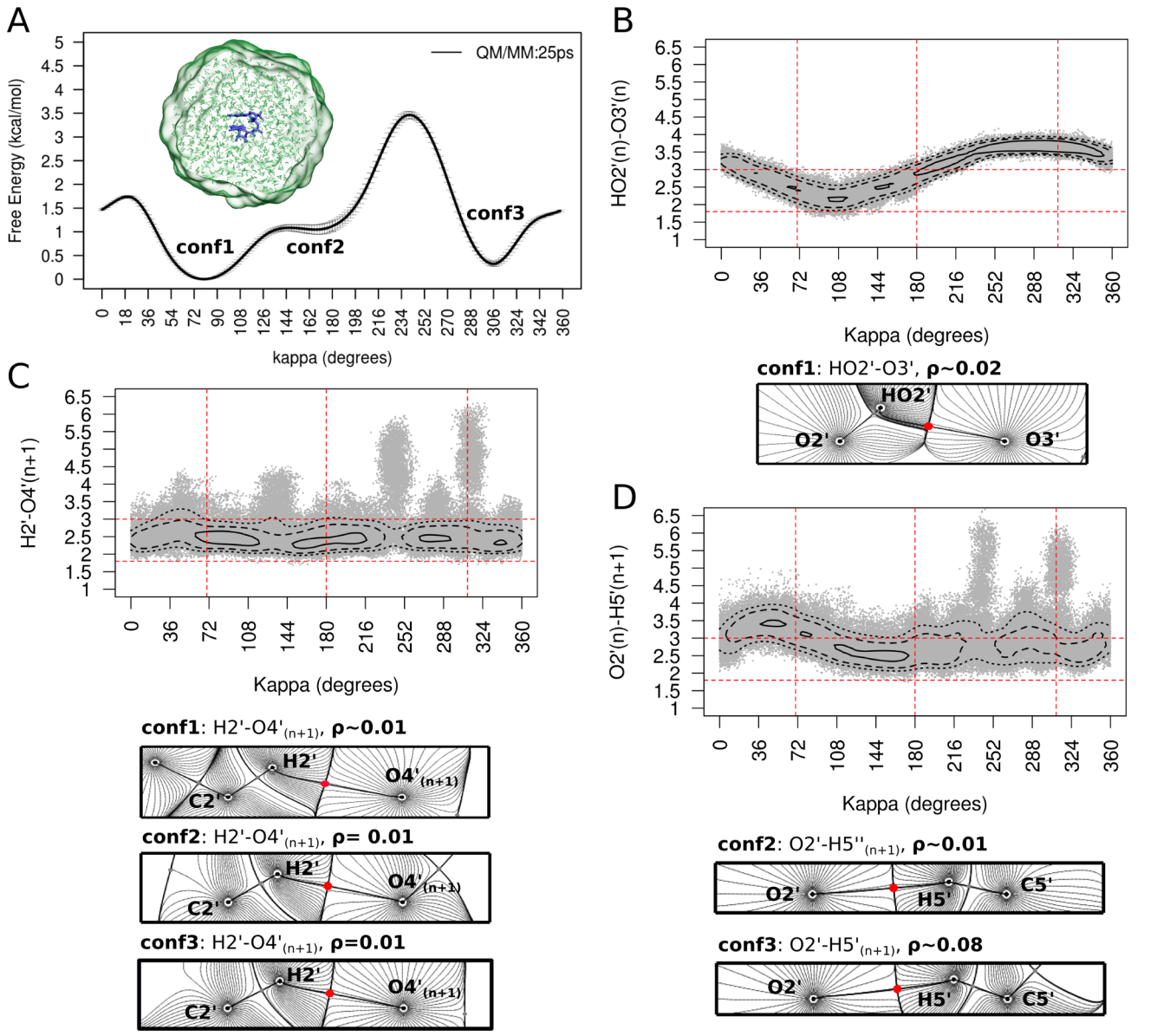


Figure 4. 2′OH κ torsion preferred orientations at the dinucleotide level. (A) US QM/MM free energy profile for the κ torsion of a rCC dinucleotide indicating three main orientations ("conf1", "conf2", and "conf3"). The continuous line and error bars correspond to the average and standard deviation of the free energy, respectively, calculated from the energy profiles obtained after 20, 21, 22, 23, 24, and 25 ps of US simulation. A snapshot of the simulated system is shown, indicating in blue the QM region (rCC dinucleotide) and in green the MM region (water and K^+ ion). (B) κ vs HO2′−O3′ distance scatter plot obtained form the US QM/MM simulation. Red dotted lines indicate optimal and maximum hydrogen bond distances (horizontal), and κ rotation minimum energy positions (vertical). Contour lines correspond to points with density values equal to the average density plus 1, 2, and 4 standard deviations. In addition, AIM projection on the O2′−HO2′···O3′ plane is also shown for a simulation snapshot corresponding to "conf1". The position of the bond critical point, the atomic nuclei involved in the interaction, and gradient field lines are indicated with red and black dots and gray lines, respectively. The density at the bond critical point (average over three simulation snapshots taken from "conf1") is also shown. (C) Same as (B) but for the interaction between C2′−H2′···O4′$_{(n+1)}$. In this case, the AIM analysis is shown for conformers 1−3. (D) Same as in (B) but for the interaction between C5′$_{(n+1)}$−H5′$_{(n+1)}$···O2′. In this case, the AIM analysis is shown for conformers 2 and 3.

hydrogens as "non-canonical" hydrogen bond donors in modified oligonucleotides.[32,33]

Impact of the C2′O2′ Torsion on the Global RNA Structure. We performed MD simulations of a variety of standard RNA motives (see Supplementary Methods 2) to see whether the most accurate RNA force field is able to capture the κ distribution found in database analysis and QM/MM calculations. Results in Figure 5A clearly indicate major errors in the κ distribution, which is dramatically biased toward conformer 1. This artifact is clearly related to a poor description of the C2′O2′ torsion, as highlighted by MM PMF calculations of the κ torsion (Figure 5B), which, compared with the QM/MM reference (Figure 4A), show a serious unbalance in the conformer 1 vs conformer 3 ratio. This behavior is not corrected if a local RNA adaptation of the DNA parmbsc1 force field is used (data not shown; official version of parmbsc1[26] should be used only for DNA), and only slightly improved (Figure 5B) if a correction in the Lennard-Jones specific interaction between the O2′ and the phosphate oxygens is introduced (see Supplementary Methods 4 and 5). Thus, the

DOI: 10.1021/jacs.6b09471
J. Am. Chem. Soc. 2016, 138, 16355−16363

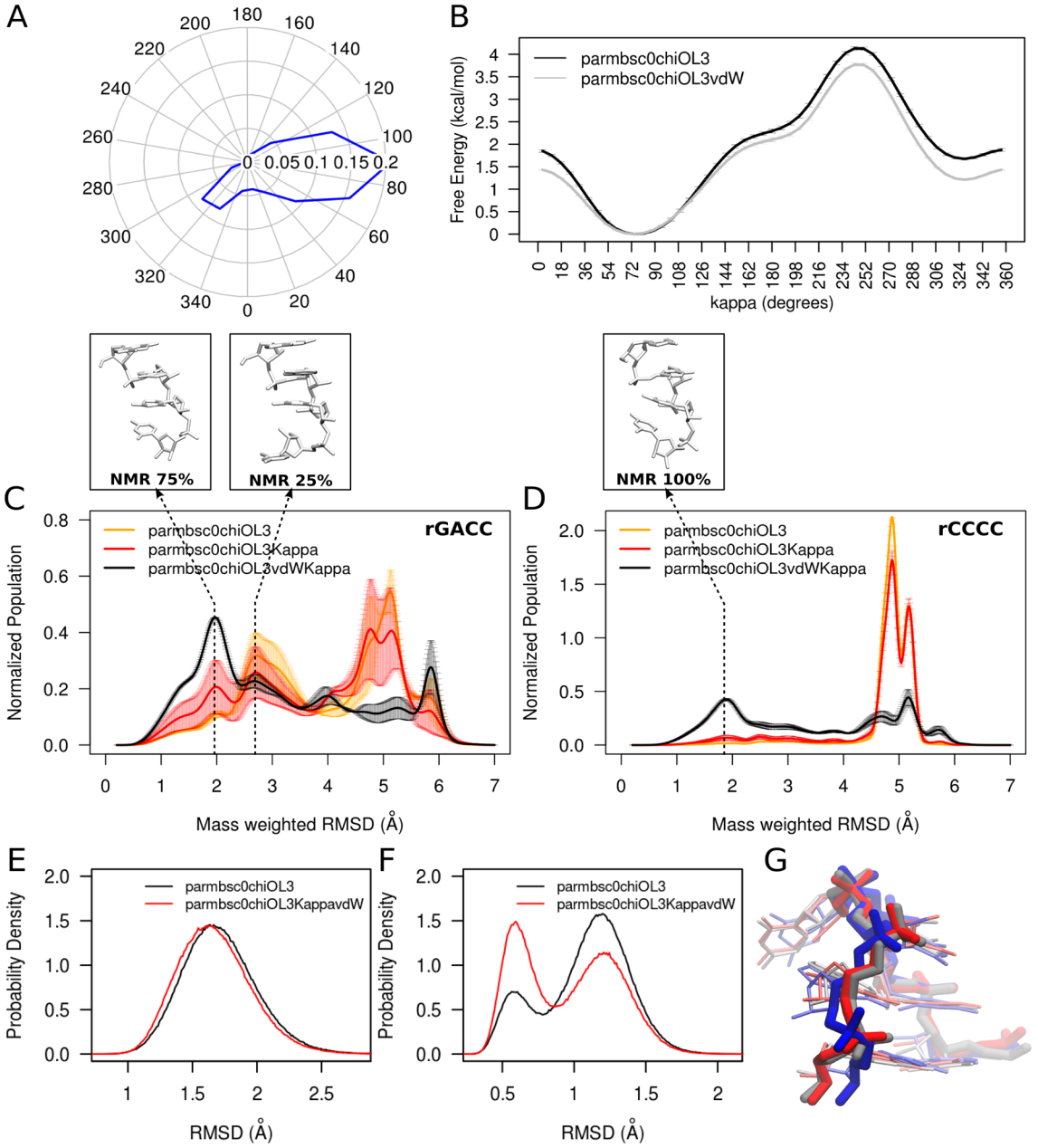


Figure 5. Kappa behavior in RNA MD simulations using parmbsc0chiOL3. (A) Probability distribution of the κ torsion from unbiased MD simulations of three hairpins and three kissing loops (see Supplementary Methods 2). (B) US MM free energy profile for the κ torsion of a rCC dinucleotide with (gray line) and without (black line) a specific correction in the Lennard-Jones potential between the phosphate oxygen atoms and the hydroxyl oxygen atoms (see Supplementary Methods 4). The profile and error bars shown correspond to the average and standard deviation from five energy profiles obtained between 2 and 2.5 ns, every 100 ps. (C) RMSD distribution (calculated using all atoms) from H-REMD simulations of the rGACC tetranucleotide using parmbsc0chiOL3 (orange line), parmbsc0chiOL3Kappa (red line), and parmbsc0chiOL3KappavdW (black line). Error bars correspond to the standard deviation from the average (continuous line) obtained from two duplicates of the H-REMD simulations (see Supplementary Methods 2). The reference structure used for the alignment (prior to RMSD calculation) corresponds to an A-form portion of the *Haloarcula marismortui* ribosome crystal structure (PDB ID: 3G6E, residues 2623−2626). Representative structures of the first two peaks are indicated with dotted arrows and correspond to NMR major and NMR minor structures,[19] respectively. (D) Same as in (C) but for the tetranucleotide rCCCC. In this case, the reference structure used for the alignment corresponds to a canonical A-form generated using NAB. A representative structure of the first peak is indicated with a dotted arrow and corresponds to the unique conformation observed in NMR.[18] (E) RMSD distribution calculated using the backbone atoms of all residues in a RNA hairpin (PDB ID: 2KOC) from two unbiased 1 μs long MD simulations using parmbsc0chiOL3 (black line) or parmbsc0chiOL3KappavdW (red line). (F) Same as in (E) but for the region containing the loop plus the first stem base pair. (G) Three-dimensional representation of the region considered in (F) taken from the experimental structure (gray) and the centroids of the clusters corresponding to peaks at ∼0.6 Å (red) and 1.2 Å (blue) in the RMSD distribution shown in (F).

error in the κ distribution is related to an incorrect representation of the C2′O2′ torsion in current state-of-the-art force fields, which cannot be corrected by other backbone parametrization. The impact of this inaccuracy is maximized in RNAs showing low levels of secondary structure, as is the case with the tetranucleotides r(GACC) and r(CCCC), where the

DOI: 10.1021/jacs.6b09471
J. Am. Chem. Soc. 2016, 138, 16355−16363

incorrect sampling of the κ torsion contributes to the formation of artifactual contacts, stabilizing incorrect structures for the oligo in H-REMD simulations (see Figure 5C,D). Correction of the C2′O2′ torsion to reproduce the QM/MM κ profiles (see Supplementary Figure 7) improves the results (Figure 5C,D), but there is a problem of transferability of the parameters between nucleotides in the middle and termini of the stand, as the presence/absence of neighboring phosphates generate different environments. Adding the specific Lennard-Jones tuning (see before) improves the fitting, guarantees transferability (see Supplementary Methods 4 and 5, and Supplementary Figure 7), and yields a much better representation of the tetranucleotide conformational space (see Figure 5 C,D).

Very encouragingly, the improvement in simulations obtained by treating more accurately the C2′O2′ torsion is also visible in a longer system (the 14 mer r(GGCACUUCG-GUGCC) hairpin with PDB ID 2KOC, containing the UUCG tetra-loop; see Figure 5E−G). These results highlight the importance of the suggested modifications, especially in regions of linkage between single- and double-stranded regions, but patches commented here should not be taken as a new validated force field.

CONCLUSIONS

By combining a variety of complementary techniques (database analysis, high-level QM calculations, QM/MM, and classical simulations), we provide convincing evidence that the C2′O2′ torsion is strongly coupled with sugar puckering while also being involved in a myriad of nonbonded contacts. Some C2′O2′ torsional states favor the transition to unusual puckerings, the presence of which is required in several protein−RNA contacts. Our results demonstrate that protein contacts with 2′OH correlates with an increase of south pucker sugar ring frequency and, furthermore, that a Lys residue placed in the most populated position observed in the database analysis can bias the C2′O2′ torsional state by forming specific hydrogen bonding with both the 2′OH and phosphate groups, leading to a north → south transition. Such transition introduces changes in the structure of the RNA, which is often required for functional RNA−protein complexes. Thus, the results presented herein support the C2′O2′ torsion as a trigger for a general novel induced-fit mechanism of protein−RNA recognition. Finally, our results raise concerns about the current state-of-the-art RNA force fields, but also suggest that recalibration of the C2′O2′ torsion can lead to an improved description of unusual RNA conformations.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.6b09471.

Supplementary Methods 1, database analysis, 2, MD additional details, 3, QM/MM additional details, 4, κ parametrization, 5, parmed.py commands for the Lennard-Jones specific interactions modification, and 6, QM/SCRF PES calculations; Supplementary Tables 1, κ torsion angle analysis, 2, κ and pucker analysis for ribonucleotides with 2′OH in contact with ARG or LYS, 3, protein−RNA contacts analysis; 4, H1-CT-OK-HO parameters, and 5, H1-CT-OK-HO parameters considering vdW specific corrections; and Supplementary Figures 1, end-to-end distance for all RNA and DNA fragments, 2, preferred orientations of the κ torsion per base type, 3, possible hydrogen bonds near the 2′OH group, 4, preferred orientations of the κ torsion angle from a non-redundant database, 5, κ energy profile for different window sizes, 6, protein−RNA contacts from the Full Dataset, 7, κ fitting to reproduce QM/MM PMF, 8, κ vs puckering QM/SCRF PESs for guanine monophosphate in the absence/presence of methylammonium, 9, puckering PMF of cytosine 6 in the MIWI PAZ domain−RNA complex, 10, lysine localization near the 2′OH in south-puckering RNA nucleotides, and 11, arginine localization near the 2′OH in south-puckering RNA nucleotides (PDF)

AUTHOR INFORMATION

Corresponding Author

*modesto.orozco@irbbarcelona.org

ORCID

Leonardo Darré: 0000-0001-5280-8579

Pablo D. Dans: 0000-0002-5927-372X

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work has been supported by the Spanish Ministry of Science (BFU2014-61670-EXP), the Catalan SGR, the Instituto Nacional de Bioinformática, and the European Research Council (ERC SimDNA), the European Union's Horizon 2020 research and innovation program under grant agreement no. 676556 (MuG), the MINECO project BIO2015-64802-R, the Biomolecular and Bioinformatics Resources Platform (ISCIII PT 13/0001/0030) cofunded by the Fondo Europeo de Desarrollo Regional (FEDER), and the MINECO Severo Ochoa Award of Excellence (Government of Spain) (awarded to IRB Barcelona). M.O. is an ICREA academia researcher. L.D. is a SNI (Sistema Nacional de Investigadores; ANII, Uruguay) researcher. P.D.D. is a SNI and PEDECIBA (Programa de Desarrollo de las Ciencias Básicas) researcher. The authors also acknowledge the Barcelona Supercomputing Center for CPU and GPU time on MareNostrum and MinoTauro computers. Federica Battistini, Fernando Romeo, and Adria Ferandez are acknowledged for providing parmbsc0-chiOL3 MD trajectories of hairpin and kissing hairpin systems, and Diego Gallego for contributing to the R-scripting used in the experimental database analysis.

REFERENCES

(1) Neidle, S. *Principles of nucleic acid structure*, 1st ed.; Elsevier/Academic Press: Amsterdam/Boston, 2008.

(2) Caetano-Anollés, G.; Caetano-Anollés, D. *Comput. Struct. Biotechnol. J.* **2015**, *13*, 427.

(3) Petrov, A. S.; Gulen, B.; Norris, A. M.; Kovacs, N. A.; Bernier, C. R.; Lanier, K. A.; Fox, G. E.; Harvey, S. C.; Wartell, R. M.; Hud, N. V.; Williams, L. D. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112* (50), 15396.

(4) Petrov, A. S.; Bernier, C. R.; Hsiao, C.; Norris, A. M.; Kovacs, N. A.; Waterbury, C. C.; Stepanov, V. G.; Harvey, S. C.; Fox, G. E.; Wartell, R. M.; Hud, N. V.; Williams, L. D. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111* (28), 10251.

(5) Saint-Leger, A.; Bello, C.; Dans, P. D.; Torres, A. G.; Novoa, E. M.; Camacho, N.; Orozco, M.; Kondrashov, F. A.; Ribas de Pouplana, L. *Sci. Adv.* **2016**, *2* (4), e1501860.

(6) Zhang, J.; Ferré-D'Amaré, A. *Life* **2016**, *6* (1), 3.

DOI: 10.1021/jacs.6b09471
J. Am. Chem. Soc. 2016, 138, 16355−16363

(7) Soliva, R.; Luque, F. J.; Alhambra, C.; Orozco, M. *J. Biomol. Struct. Dyn.* **1999**, *17* (1), 89.
(8) Auffinger, P.; Westhof, E. *J. Mol. Biol.* **1997**, *274* (1), 54.
(9) Fohrer, J.; Hennig, M.; Carlomagno, T. *J. Mol. Biol.* **2006**, *356* (2), 280.
(10) Mládek, A.; Banáš, P.; Jurečka, P.; Otyepka, M.; Zgarbová, M.; Šponer, J. *J. Chem. Theory Comput.* **2014**, *10* (1), 463.
(11) Egli, M.; Portmann, S.; Usman, N. *Biochemistry* **1996**, *35* (26), 8489.
(12) Denning, E. J.; MacKerell, A. D. *J. Am. Chem. Soc.* **2012**, *134* (5), 2800.
(13) RNA 3D structure analysis and prediction. In *Nucleic acids and molecular biology*; Leontis, N. B., Westhof, E., Eds.; Springer: Heidelberg/New York, 2012.
(14) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **2009**, *113* (18), 6378.
(15) Bader, R. F. W. *J. Phys. Chem. A* **1998**, *102* (37), 7314.
(16) Bader, R. F. W. *Chem. Rev.* **1991**, *91* (5), 893.
(17) Bader, R. F. W. *Atoms in molecules: a quantum theory*; The International series of monographs on chemistry; Clarendon Press/Oxford University Press: Oxford/New York, 1994.
(18) Vega, D.; Almeida, D. *J. Comput. Methods Sci. Eng.* **2014**, *14* (No. 1−3), 131.
(19) Tubbs, J. D.; Condon, D. E.; Kennedy, S. D.; Hauser, M.; Bevilacqua, P. C.; Turner, D. H. *Biochemistry* **2013**, *52* (6), 996.
(20) Yildirim, I.; Stern, H. A.; Tubbs, J. D.; Kennedy, S. D.; Turner, D. H. *J. Phys. Chem. B* **2011**, *115* (29), 9261.
(21) Cheatham, T. E.; Cieplak, P.; Kollman, P. A. *J. Biomol. Struct. Dyn.* **1999**, *16* (4), 845.
(22) 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* (19), 5179.
(23) Pérez, A.; Marchán, I.; Svozil, D.; Sponer, J.; Cheatham, T. E.; Laughton, C. A.; Orozco, M. *Biophys. J.* **2007**, *92* (11), 3817.
(24) Zgarbová, M.; Otyepka, M.; Šponer, J.; Mládek, A.; Banáš, P.; Cheatham, T. E.; Jurečka, P. *J. Chem. Theory Comput.* **2011**, *7* (9), 2886.
(25) Banáš, P.; Hollas, D.; Zgarbová, M.; Jurečka, P.; Orozco, M.; Cheatham, T. E.; Šponer, J.; Otyepka, M. *J. Chem. Theory Comput.* **2010**, *6* (12), 3836.
(26) Ivani, I.; Dans, P. D.; Noy, A.; Pérez, A.; Faustino, I.; Hospital, A.; Walther, J.; Andrio, P.; Goñi, R.; Balaceanu, A.; Portella, G.; Battistini, F.; Gelpí, J. L.; González, C.; Vendruscolo, M.; Laughton, C. A.; Harris, S. A.; Case, D. A.; Orozco, M. *Nat. Methods* **2015**, *13*, 55.
(27) Dang, L. X.; Kollman, P. A. *J. Phys. Chem.* **1995**, *99* (1), 55.
(28) Dang, L. X. *J. Am. Chem. Soc.* **1995**, *117* (26), 6954.
(29) Smith, D. E.; Dang, L. X. *J. Chem. Phys.* **1994**, *100* (5), 3757.
(30) de Ruiter, A.; Zagrovic, B. *Nucleic Acids Res.* **2015**, *43* (2), 708.
(31) Cubero, E.; Orozco, M.; Hobza, P.; Luque, F. J. *J. Phys. Chem. A* **1999**, *103* (32), 6394.
(32) Martin-Pintado, N.; Deleavey, G. F.; Portella, G.; Campos-Olivas, R.; Orozco, M.; Damha, M. J.; González, C. *Angew. Chem., Int. Ed.* **2013**, *52* (46), 12065.
(33) Martín-Pintado, N.; Yahyaee-Anzahaee, M.; Deleavey, G. F.; Portella, G.; Orozco, M.; Damha, M. J.; González, C. *J. Am. Chem. Soc.* **2013**, *135* (14), 5344.

DOI: 10.1021/jacs.6b09471
J. Am. Chem. Soc. 2016, 138, 16355−16363