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

Leonardo Darré[1,2], Ivan Ivani[1,2], Pablo D. Dans[1,2], Hansel Gómez[1,2], Adam Hospital[1,2] and Modesto Orozco[1,2,3*]

[1]Institute for Research in Biomedicine (IRB Barcelona). The Barcelona Institute of Science and Technology, 08028 Barcelona, Spain

[2]Joint BSC-IRB Program in Computational Biology, Institute for Research in Biomedicine, 08028 Barcelona, Spain

[3]Department of Biochemistry and Biomedicine, Faculty of Biology, University of Barcelona, 08028 Barcelona, Spain

KEYWORDS: RNA, 2'OH, pucker, QM/MM, data mining, molecular dynamics.

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 conformation of the RNA is still rather limited. In the A-form (sugar in north), the

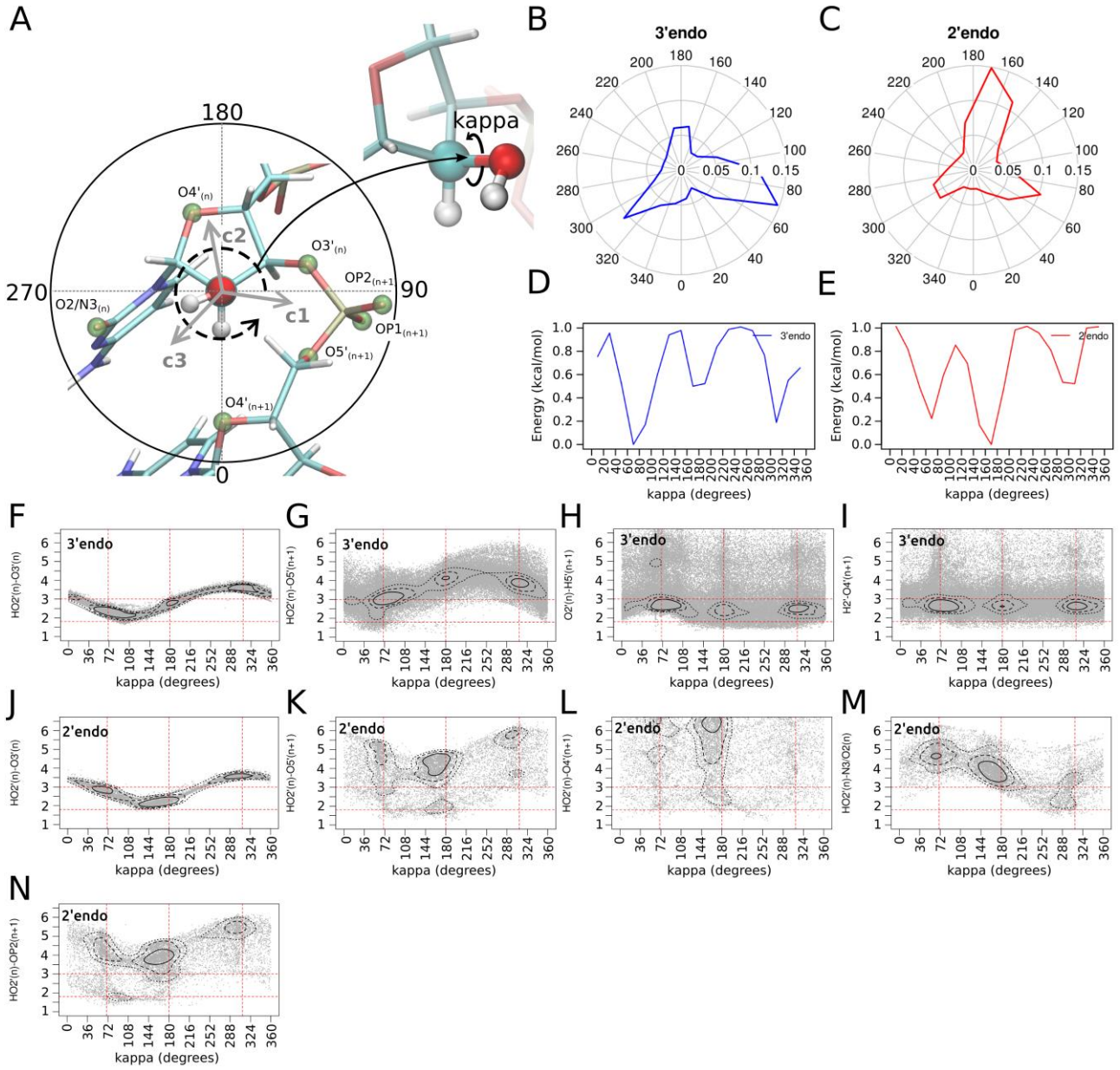


Figure 1. Kappa torsion preferred orientations from the Protein Data Bank. (A) Nucleotide in a RNA strand indicating the possible orientations of the 2'OH and the location of neighboring hydrogen bond acceptors/donors. (B) Probability distribution of the torsion angle between the atoms H2'-C2'-O2'-HO2' 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.

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.

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 Protein Data Bank (PDB) were analyzed (see Supplementary Methods 1), accounting for a total of 174511 2'OH groups, 115513 with the ribose in *north* puckering (0≤pucker phase≤36) and 11626 with the ribose in the unusual *south* puckering (144≤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 26760 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 χ_N=190° and χ_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°, 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 O···H-X (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 forcefield for RNA. They include the following: (i) standard simulations in hairpin and kissing loop RNA motives ; (ii) potentials of mean force (PMF) 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 bsco[23] and chiOL3[24,25] modifications for RNA; some control simulations were performed with a local experimental RNA version of the parmbsc1 forcefield.[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*

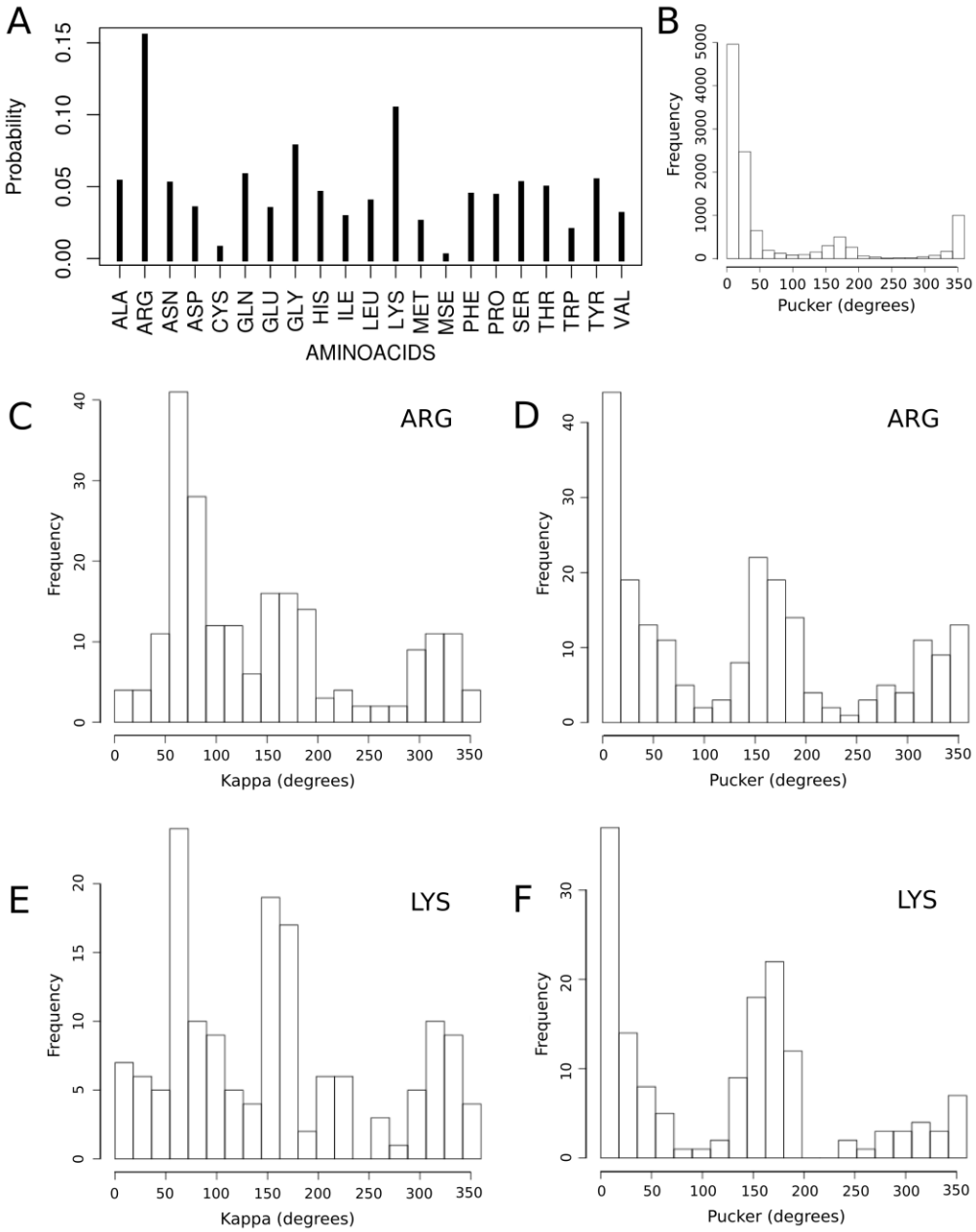


Figure 2. Protein-RNA contacts. (A) Probability of contact between a given amino acid and the 2'OH, given a protein-RNA contact occurs, calculated by counting all contacts (distance ≤ 4 Å) between any protein atom and the oxygen of 2'OH, and splitting the counts per amino acid identity. Multiple atoms 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.

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 $C5'_{(n+1)}$-$H5'_{(n+1)}\cdots O2'$ hydrogen bond and the non-canonical C2'-$H2'\cdots O4'_{(n+1)}$ 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 in those 2'OH interacting with protein residues. As for the κ 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 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 (<3.5Å) 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

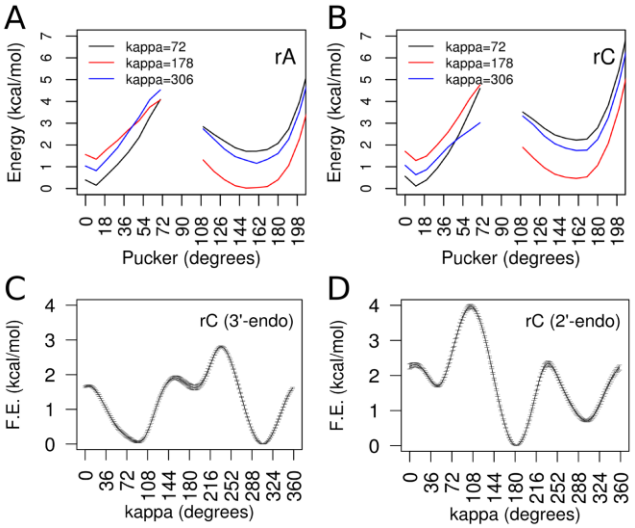


Figure 3. 2'OH κ torsion and sugar pucker phase preferred conformers at the nucleoside level. (A,B) QM potential energy scans of the sugar pucker phase (for adenosine and cytosine, respectively) restraining the κ torsion at the main observed minima in the κ free energy profile (72°, 178° and 306°). The 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.

~1.2 kcal/mol (Figure 3A,B). However, as 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

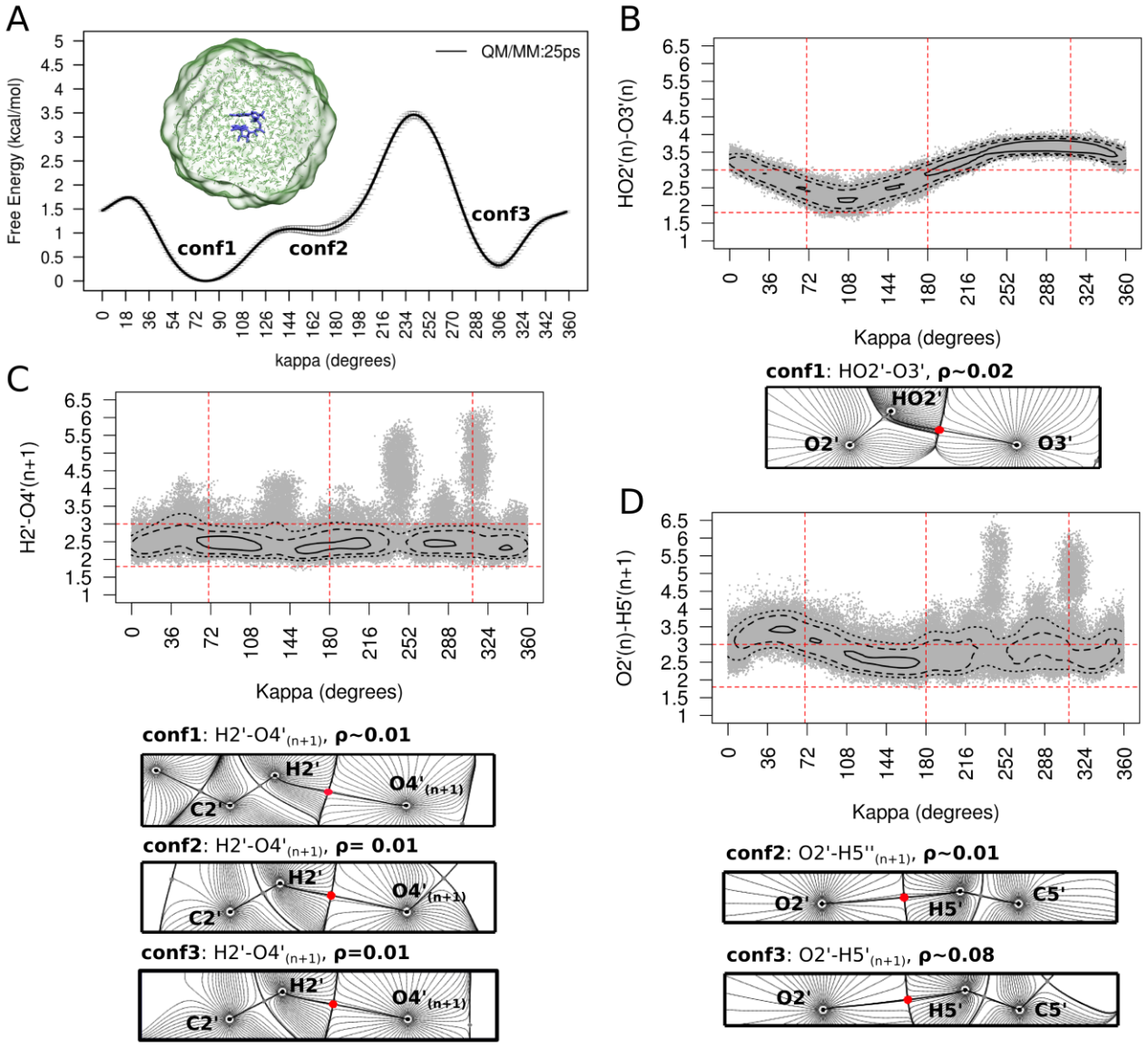


Figure 4. 2'OH κ torsion preferred orientations at the dinucleotide level. (A) US QM/MM free energy profile for the κ torsion of a rCC dinucleotide indicating three main orientations ("conf1", "conf2", and "conf3"). The continuous line and error bars correspond to the average and standard deviation of 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 from 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.

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 Methods and Supplementary Methods 2) dramatically stabilizes the south conformation (see Supplementary Figure S9). 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

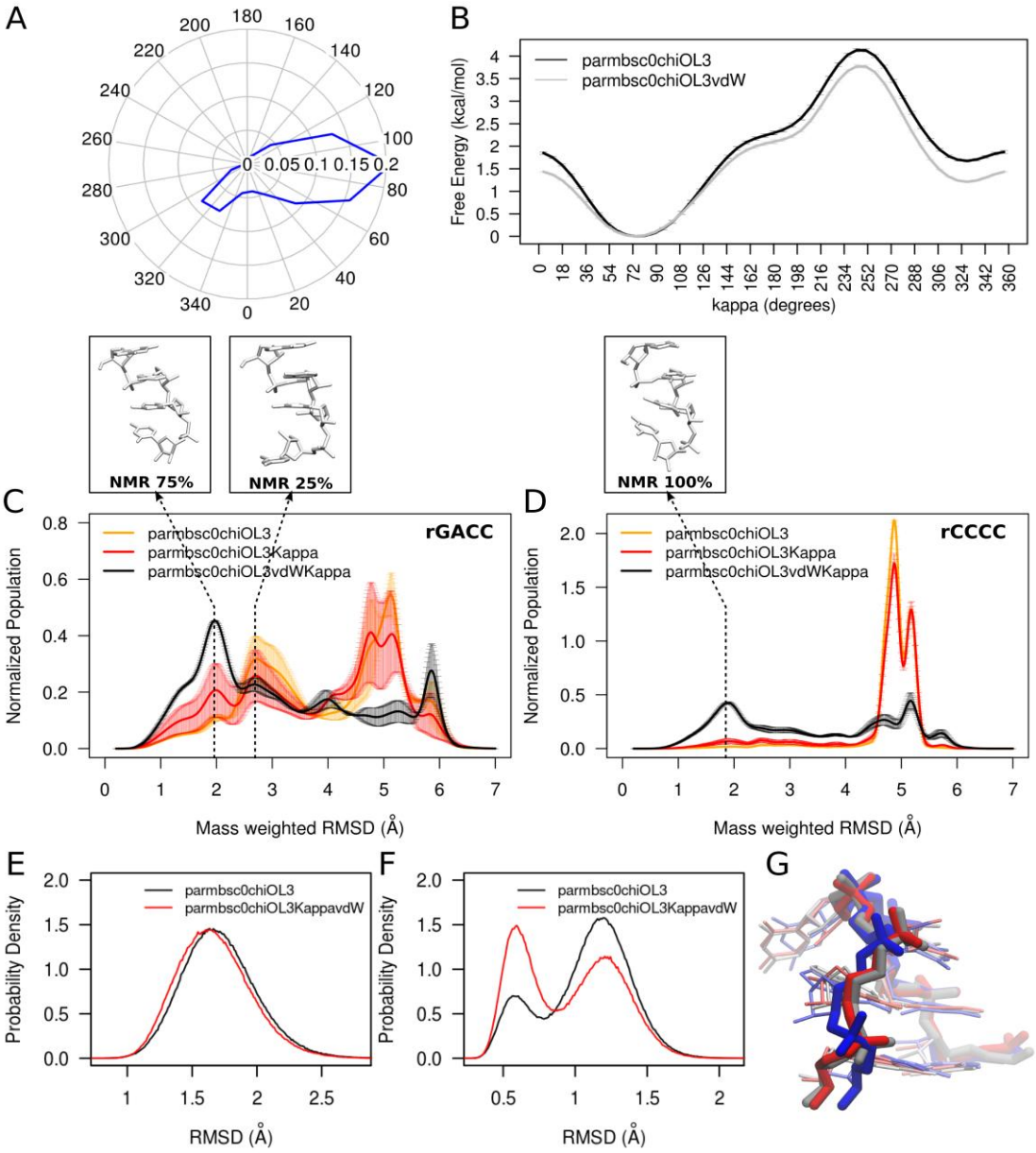


Figure 5. Kappa behavior in RNA MD simulations using parmbscochiOL3. (A) Probability distribution of the κ torsion from unbiased MD simulations of three hairpins and three kissing loops (see Supplementary Methods 2). (B) US MM free energy profile for the κ torsion of a rCC dinucleotide with (gray line) 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 parmbscochiOL3 (orange line), parmbscochiOL3Kappa (red line), and parmbscochiOL3KappavdW (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 (PDBID: 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,[20] 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.[19] (E) RMSD distribution calculated using the backbone atoms of all residues in a RNA hairpin (PDBID: 2KOC) from two unbiased 1μs long MD simulations using parmbscochiOL3 (black line) or parmbscochiOL3KappavdW (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).

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 ($O2'–HO2'\cdots O3$: $\rho \approx 0.020$ au) and non-canonical ($C5'–H5'\cdots O2'$: $\rho \approx 0.009$ au and $C2'–H2'\cdots O4'$: $\rho \approx 0.010$ 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 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 forcefield 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 forcefield 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 error in the κ distribution is related to an incorrect representation of the C2'O2' torsion in current state-of-the-art forcefields, 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 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(GGCACUUCGGUGCC) hairpin with PDB ID 2KOC, containing the UUCG tetra-loop; see Figure 5E-F-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 forcefield.

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

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; 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 the 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, 11, arginine localization near the 2'OH in south-puckering RNA nucleotides.

AUTHOR INFORMATION

Corresponding Author

* modesto.orozco@irbbarcelona.org

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENT

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 parmbscochiOL3 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.* **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.* **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.
(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**, 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* **2016**.
(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.

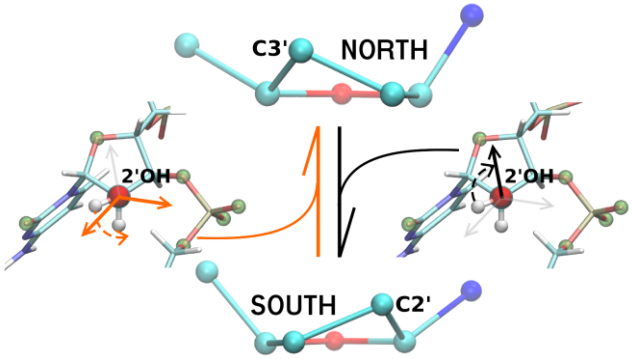


Table of Contents artwork

Supplementary Information

Small Details Matter: the 2'-Hydroxyl as a Conformational Switch in RNA

Leonardo Darré[1,2], Ivan Ivani[1,2], Pablo D. Dans[1,2], Hansel Gómez[1,2], Adam Hospital[1,2] and Modesto Orozco[1,2,3]*

[1]Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology, 08028 Barcelona, Spain

[2]Joint BSC-IRB Program in Computational Biology, Institute for Research in Biomedicine, 08028 Barcelona, Spain

[3]Department of Biochemistry, Faculty of Biology, University of Barcelona, 08028 Barcelona, Spain

* Send correspondence to M.Orozco: modesto.orozco@irbbarcelona.org

Contents

Supplementary Methods 1. Database Analysis.

All the analysis of NMR or X-ray structures was done using local R scripts using the bio3D[1] libraries.

Kappa Torsion Distribution. Two datasets were used to build the kappa torsion empirical distribution: i- the "Full Dataset" which contains the current state of the PDB up to June 2016 for NMR-solved structures containing RNA (610 entries), and ii- the "Non-Redundat Dataset" which contains NMR-solved RNA structures (476 entries) proposed by Leontis et al.[2] to avoid structural redundancy available from the BGSU Structural Bioinformatics Group web page (http://rna.bgsu.edu/rna3dhub/nrlist/), see Supplementary Table 1 for further details. For the Full Dataset, all NMR models in every PDB entry were split into RNA continuous segments (two or more residues), and the kappa torsion angle was measured for every ribonucleotide within a given segment. The canonical hydrogen bond local interactions of the 2'OH group were analyzed by measuring the distance between the 2'OH hydrogen atom and the atoms: O3', O4' and O2 (pyrimidines)/N3 (purines) from the same ribonucleotide, or O5', OP1, OP2 and O4' of the ribonucleotide in 3'. In addition, non-canonical hydrogen bonds were assessed by measuring the distance between the H2' or O2' of a given ribonucleotide and O4' or H5'/H5'' of the ribonucleotide in 3', respectively. To capture the effect of the sugar conformation on the kappa torsion angle, the pucker phase was also measured using Westhof & Sundaralingam definition[3] and obtaining kappa/pucker phase pairs for each analyzed ribonucleotide. Kappa probability distributions were calculated using angle windows of 20 degrees and plotted for 3'endo and 2'endo pucker phases separately, for all bases together or split by base type. The correlation between the kappa torsion angle and the distances to local hydrogen bond acceptors/donors are shown by means of scatter plots and three density contours corresponding to points in the distance-kappa space with density equal to the average density plus one, two or four standard deviations. Finally, kappa distributions were converted to empirical free energies from the relative populations of kappa values between 0 and 360 degrees, considering windows of 20, 15, 10 and 5 degrees, using the relation: $\Delta G_{i/0}=R*T*Ln(P_i/P_0)$, where P_i and P_0 are the population of kappa values for windows i and 0, respectively. The windows [0,20], [0,15], [0,10], and [0,5] were used as reference (window 0) for each of the four striding options mentioned above. The measurement of kappa and pucker and the calculation of the empirical free energy were repeated for the Non-Redundant Dataset although in this case only specific chains and NMR models were used as suggested in the BGSU Structural Bioinformatics Group web page.

Kappa Torsion and Pucker Phase Distributions in 2'OH-ARG/LYS Contacts. Both dataset mentioned in the previous section were filtered keeping only PDB entries corresponding to protein-RNA complexes (see Supplementary Table 2). Kappa and pucker distributions were obtained for ribonucleotides with the 2'OH group in contact with the aminoacids ARG and LYS (distance between any ARG or LYS atom and the oxygen atom of the 2'OH moiety lower or equal to 4 Å). When multiple atoms from the same ARG or LYS residue were in contact with a given ribonucleotide 2'OH, the corresponding kappa/pucker pair was counted only once.

Probability of Contacts Between a Given Aminoacid and the 2'OH Group. The Full and the Non-Redundant Datasets filtered to keep only protein-RNA complexes, which contain only NMR-solved structures, were supplemented with X-ray solved protein-RNA complexes obtained from the current state of the PDB (up to June 2016) or the Leontis et al. non-redundant database, respectively, for resolutions below 2.5 Å. For both NMR/X-ray datasets, the number of contacts (distance <= 4 Å) between any amino acid atom and the 2'OH oxygen atom was counted. When multiple atoms from the same amino acid were in contact with a given ribonucleotide 2'OH, the contact was counted only once to eliminate repeated counts per amino acid. The contacts frequency per amino acid was divided by the total number of observed contacts, thus obtaining the aminoacid-2'OH interaction probability given that a contact exists.

Cluster analysis of Lys and Arg residues close to south puckering. The principal components of the Cartesian coordinates of the NZ atom (Lys) or CZ atom (Arg) were calculated for all occurrences of Lys or Arg residues within 4 Å of the O2’ atom in south puckering nucleotides in the non-redundant database. The first two principal components were used as coordinates to hierarchically cluster the position of the cationic protein side chains in space. Distance histograms between all hydrogen bond donor nitrogen atoms in Lys or Arg and the O2’, O3’, OP1, OP2 and O5’ atoms in RNA were constructed for: (i) all the considered structures containing Lys residues, (ii) all the considered structures containing Arg residues, and (iii) the two main PC-based clusters.

End to end distance measurement. To account for the difference in the conformational space of RNA compared to DNA, the end to end distance was measured for all RNA and DNA fragments in the non-redundant RNA database and all available structures in the Protein Data Bank (up to 22nd Nov 2016), respectively. This was done cutting all nucleic acid fragments into dodecamer strands (removing shorter segments) and measuring the distance between the C1’ atom of the 5’ and 3’ terminal atoms.

Supplementary Methods 2. MD Additional Details.

All classical MD simulations were run using AMBER-14 suite. TLEAP code was used for systems preparation, CPPTRAJ for post-processing and analyzing trajectories and ParmEd to modify and check topologies when needed (e.g. scale torsion angles force constants for HREMD calculations). Restraints were imposed using native AMBER algorithms or by means of the PLUMED 2.2 patch to AMBER-14. Generation of free energy profiles from umbrella sampling simulations was achieved using vFEP.[4]

Unbiased Molecular Dynamics Simulations. Microsecond long MD simulations of six RNA structures corresponding to three hairpins (PDBIDs: 1JJ2, 1Q9A and 2KOC) and three kissing loops (PDBIDs: 1BAU, 2BJ2 and 2RN1) were run using parm99 forcefield[5,6] supplemented with the bsc0[7] and chiOL3[8,9] corrections (here in called "parmbsc0chiOL3") to model the RNA. To take into account solvent model effects, two of the most widely used water models were employed, TIP3P[10] for the hairpin structures and SPC/E[11] for the kissing loops structures. In all cases a 150mM ionic environment was represented using Dang parameters[12–14] for K+ and Cl-. MD simulations were performed in the NPT ensemble using Berendsen thermostat[15] with a time constant of 5 ps^{-1} and the Berendsen barostat with a time constant of 5 ps^{-1}. Equations of motion were integrated using a time step of 2fs with the pmemd.cuda code.[16] Each system was subject to 2000 steps of energy minimization with position restraints in the solute of 25 kcal/mol, followed by 1 ns of position restrained (5 kcal/mol) thermalization in the NVT ensemble and 10 ns unrestrained equilibration in the NPT ensemble. Production MD simulations were run for 1 μs. Non-bonded direct cut-off was set to 9 Å and particle mesh Ewald[17] was used for reciprocal space calculations. All bonds involving hydrogen atoms were constrained by means of SHAKE algorithm.[18]

Hamiltonian Replica Exchange Molecular Dynamics Simulations. The conformational landscape of two tetranucleotides, rGACC and rCCCC, were explored enhancing the sampling by allowing coordinates exchange between eight replicas where all torsion angle force constants are scaled by: 1, 0.9 , 0.8, 0.7, 0.6, 0.5, 0.4, and 0.3, achieving an exchange acceptance in the range of 25-60%. rGACC initial structure was taken from an A-form portion of the *H. marismortui* ribosome crystal structure (PDBID: 3G6E, residues 2623-2626), following the same approach as Henriksen et al.[19] rCCCC initial structure was generated in a random conformation using NAB. The RNA molecule in each system was modelled using parmbsc0chiOL3, solvated using the TIP3P model[10] and neutralized with

three K+ ions using Dang parameters.[12–14] Preparation of both systems for the first set of HREMD involved 2000 steps of position restrained (25 kcal/mol) minimization, and heated during 2 ns of MD from 10-150 K (NVT and 25 kcal/mol position restraints) and from 150-300 K (NPT and 5 kcal/mol position restraints), using a time step of 1 fs. System density at 300 K and 1 Bar was relaxed in 5 ns of 2 fs time step MD in the NPT ensemble with soft position restraints (0.5 kcal/mol) further extended by 500 ps of unrestrained equilibration in NVT. Production HREMD simulations were run in the NVT ensemble at 300 K using the Langevin thermostat with a collision frequency of 2 ps^{-1} and resetting the random seed at each restart to avoid synchronization effects. A 2 fs time step was used with an exchange attempt every 1 ps. Non-bonded direct cut-off was set to 8 Å and particle mesh Ewald[17] was used for reciprocal space calculations. All bonds involving hydrogen atoms were constrained by means of SHAKE algorithm.[18] The independent second run of HREMD simulations were started from the restart structures of the first run after 500 ns, assigning new velocities and equilibrating for 1 ns in the NVT ensemble. Total simulated time for both independent runs was 1.2 μs per replica. Equations of motion were integrated using the pmemd.cuda.MPI code.

Umbrella Sampling Molecular Dynamics Simulations. Classical mechanics umbrella sampling simulations were run for the rCpC dinucleotide to obtain the kappa torsion potential of mean force in order to compare with the corresponding profiles at QM/MM level. For both systems, the solute was modelled using parmbsc0chiOL3 forcefield, solvated using TIP3P water model[10] and neutralized (rCpC) with one K+ ions using Dang parameters[12–14]. The rotation of the kappa torsion was sampled in twenty windows of 18 degrees applying a restraining potential on kappa of 35 kcal/mol. Each window initial configuration was extracted from an exploratory well tempered metadynamics[20] simulation (50 ns; initial Gaussian high of 1.2 kJ/mol; deposition period of 1ps; sigma=0.35 radians; BIASFACTOR=4, T=300 K) of the rCpC dinucleotide, and further equilibrated for 500 ps in the NPT ensemble at 300K and 1 Barr. Production data was collected for 2.5 ns of NPT molecular dynamics for each window. Restraints on beta and gamma backbone torsions, as well as on the sugar pucker were used as in the QM/MM simulations detailed below. Umbrella sampling was also used to obtain the puckering PMF for the cytosine 6 residue of a RNA fragment in complex with protein MIWI PAZ domain (PDB ID: 2XFM; model 6) for the wild type (presence of Lys 316 close to the 2'OH group of cytosine 6) and for a mutant (Lys316Ala). The RNA-protein complex was modeled using parmbsc0chiOL3KappavdW forcefield modification (RNA) and FF14SB (protein), solvated

using TIP3P water model[10] and neutralized with K+ ions using Dang parameters. [12–14] The system was subject to 2000 steps of energy minimization with position restraints on both RNA and protein of 25 kcal/mol, followed by 500 ps of position restrained (5 kcal/mol) thermalization in the NVT ensemble and 500 ps restrained (2.5 kcal/mol) equilibration in the NPT ensemble. Production MD simulations were run for 2.2 ns in the NPT ensemble, keeping the last 1.2 ns for PMF calculation. Position restraints (2.5 kcal/mol) on the RNA (except for residue 6 and atoms C5', H5', H5'', O5', P', OP1 and OP2 of residue 7) where applied to avoid gross changes in the RNA structure during the puckering transition. In the case of the wild type, potential energy walls were placed at 3 Å from O2' and OP2 atoms to keep the contact with Lys 316 during the puckering transition. Non-bonded direct cut-off was set to 8 Å and particle mesh Ewald[17] was used for reciprocal space calculations. All bonds involving hydrogen atoms were constrained by means of SHAKE algorithm.[18] The pucker transition was sampled in twelve windows of 18 degrees applying a restraining potential on the pucker phase (as defined in PLUMED 2.2[22]) of 35 kcal/mol. Calculation of the free energy profile was achieved by means of the vFEP program.[4]

Supplementary Methods 3. QM/MM Additional Details.

All QM/MM dynamics simulation were run using the interface between TERACHEM[23–26] (QM) and AMBER (MM) as implemented in AMBER-14, with a time step for the integration of the equations of motion of 1 fs. Potential energy walls (when required) and/or restraints were enforced by means of PLUMED 2.2[22] patch to AMBER-14. Calculation of the free energy profile from the umbrella sampling trajectories was achieved using vFEP.[4]

Kappa Torsion Potential of Mean Force. Umbrella sampling QM/MM simulations were run to obtain the free energy profile of the C2'O2' (kappa) torsion rotation for a cytosine nucleoside (rC) and for a cytosine dinucleotide (rCpC) in aqueous solution. The system setup was the same as per the classical umbrella sampling calculations (see previous section). In both cases the nucleic acid was treated at the quantum level BLYP/6-31G(d) while the aqueous environment (water or water plus one K^+ ion) was treated at the classical level (TIP3P[10] and Dang parameters[12–14] for ions). The rotation of the kappa torsion was sampled in twenty windows of 18 degrees applying a restraining potential on kappa of 35 kcal/mol. Each window was first equilibrated fully classically ("parmbsc0chiOL3") for 500 ps in the NPT ensemble (300 K and 1 Barr). The restart classical configurations were relaxed at the QM/MM level for 5 ps and production simulations were carried out for 40 and 25 ps for rC and rCpC, respectively. Wavefunction

SCF calculations were done in mixed precision including DFTD3 dispersion corrections.[27] In the case of the rC nucleoside, sugar pucker transitions were frequently observed affecting the sampling of the kappa rotation. Consequently, a potential energy wall as implemented in PLUMED 2.2[22] was applied to the Zx Cartesian coordinate of the ring puckering[21] (a lower wall at Zx=0.3 to maintain the 3'endo conformation or an upper wall at Zx=-0.3 to maintain the 2'endo conformation). The dinucleotide simulation maintained the 3'endo initial pucker, thus the use of walls was not required (that was not the case for the MM simulations where pucker phase restraints were needed). For both rC and rCpC, 5kcal/mol restraints on the beta and gamma backbone torsions were applied to avoid interactions with the phosphate oxygen atoms. For rC additional restraints (5kcal/mol) were also applied on epsilon backbone torsion to keep it at the standard value.

Supplementary Methods 4. Kappa Parametrization. In parm99bsc0chiOL3 the C2'O2' torsion rotation is controlled by three dihedral angles: C1'-C2'-O2'-HO2' (dihedral type: CT-CT-OH-HO), C3'-C2'-O2'-HO2' (dihedral type: CT-CT-OH-HO) and H2'-C2'-O2'-HO2' (dihedral type: H1-CT-OH-HO). To avoid affecting non-RNA OH moieties described using the current AMBER forcefield distributions, a new atom type for the O2' atom was introduced (OK) for refitting the Kappa torsion angle. The dihedral type H1-CT-OH-HO was substituted by H1-CT-OK-HO with a new set of parameters, while the dihedral type CT-CT-OH-HO was renamed CT-CT-OK-HO but keeping the original set of parameters. As in the parmbsc0 and parmbsc1 parametrization procedure, a flexible Metropolis Monte Carlo algorithm was used to fit a truncated third order Fourier series to the difference between: i- QM/MM pmf of the Kappa rotation for the rCpC dinucleotide, and ii- the corresponding pmf obtained at MM level ($parmbsc0chiOL3_{H1\text{-}CT\text{-}OK\text{-}HO=0}$). Both QM/MM and MM potentials of mean force were obtained from umbrella sampling calculations for the sugar in North conformation as described in Supplementary Methods 2 and 3 (see Supplementary Figure 7A). The obtained new parameters (see Supplementary Table 4) were tested on two tetranucleotide systems (rGACC and rCCCC) exhaustively exploring their conformational landscapes by means of Hamiltonian Replica Exchange simulations (see Supplementary Methods 2 for simulation details). In addition to the previous parametrization, a second fitting was performed considering a specific modification of the Lennard-Jones potential (increase in the sigma parameter) between the phosphate oxygen atoms and : i- the ribose O2', O3' atoms, and ii- the amine nitrogen of the base (N6 in A, N2 in G and N4 in C), herein called “parmbsc0chiOL3vdW”. This correction to the Lennard Jones potential is

based on the AMBER parameters revision for organic phosphates proposed by Steinbrecher et al,[28] which was recently shown to improve the description of RNA tetranucleotides.[29] In the present work, instead of including a general Lennard-Jones correction affecting the interaction between the phosphate oxygen atoms and all other atoms in the system, the specific terms affecting only the atoms mentioned above were corrected (see Supplementary Methods 5 section for the parmed.py script). The Kappa torsion parameters were fitted as before but using the parmbsc0chiOL3vdW$_{\text{H1-CT-OK-HO=0}}$ pmf for the MM level reference (see Supplementary Figure 7). The obtained parameters (see Supplementary Table 5) were tested again on the tetranucleotide systems (rGACC and rCCCC) and on microsecond-long unbiased MD simulations of a RNA hairpin (PDB ID: 2KOC; see Supplementary Methods 2 for simulation details).

Supplementary Methods 5. Parmed.py commands for the Lennard-Jones specific interactions modification.

```
changeLJPair @%OS @%N2 3.5958 0.17
changeLJPair @%O2 @%N2 3.5733 0.188944436
changeLJPair @%O2 @%OH 3.4703 0.210199905
changeLJPair @%OS @%OH 3.4928 0.189124298
changeLJPair @%O2 @%OK 3.4703 0.210199905
changeLJPair @%OS @%OK 3.4928 0.189124298
addLJType @O4' radius 1.6837 epsilon 0.1700
parmout OUTFILE
go
```

Supplementary Methods 6: QM/SCRF potential energy surface calculations. RNA Guanine residue 65 and protein Lys residue 32 from the RNA-protein complex structure with PDB ID 4BY9 was used as starting point to build the atomistic models used in Quantum Mechanics (QM) Potential Energy Surface (PES) calculations. This nucleotide-amino acid pair belongs to the most populated cluster (cluster A) of Lys residues within 4 Å of the O2’ atom of south puckering nucleotides (see Supplementary Figure 10). From such structure, the Guanine mono-phosphate (keeping the C5’, H5’, and H5’’ of the nucleotide at 3’, and completing the C5’ valence with a third H atom) and the methyl-ammonium group were kept for the QM calculation while the rest of the atoms where removed. Initially, a first round of QM geometry optimizations was performed restraining the sugar ring torsions ν1 and ν3 to scan the pucker North<->East<->South transition (0 to 190 in 10 degrees steps) and kappa at 216 degrees. Starting from these structures (restrained to the corresponding puckering value) further optimizations were performed restraining kappa to values ranging from 18 to 216 in steps of 18 degrees, giving a total of 240 calculations for

every PES. Additional restraints were applied to β, γ, ε, ζ and α torsions to maintain the experimental conformation of the backbone, and on χ at 190 or 230 degrees for North or South puckering values, respectively, to take into account the correlation between the puckering and the glycosidic torsion. Geometry optimizations in the presence of the methyl ammonium were initially done applying distance restraints between the nitrogen in methyl ammonium and both OP2 and O2' atoms in the nucleoside mono-phosphate. Such restraints were subsequently removed to let the position of the methyl ammonium to relax to the nearest potential energy minimum. This procedure ensured the presence of the hydrogen bonds with the phosphate and 2'OH in the complete PES scan. Geometry optimizations were done with the DL-FIND optimiser[30] implemented in the modular package ChemShell[31,32]. Turbomole 6.6[33] was used to compute energies and gradients at the QM(blyp[34,35]/def2-SVP[36,37]) level of theory and taking advantage of the Resolution-of-the-Identity (RI) approximation[36,38]. Geometry optimizations were performed using the continuum solvation model named Direct Conductor-like Screening Model for Real Solvents (DCOSMO-RS[39], as implemented in Turbomole), with a permittivity ε=78.

Supplementary Table 1. Kappa Analysis (only NMR structures).[a]

	Full Dataset	Non-redundant Dataset
Number of entries	**610 (*7518*)**[b]	**476 (*531*)**[b]
Number of analysed entries	**584 (*7256*)**[b]	**459 (*503*)**[b]
Number of analysed nucleotides	**174511**	**11212**

[a]All available NMR models were used in the PDB (10/06/2016) set analysis, while only specific models were used for the non-redundant dataset (see Supplementary Methods 1).
[b] Number of NMR models for the given set of PDB entries.

Supplementary Table 2. Kappa and pucker analysis for ribonucleotides with 2'OH in

contact (distance <=4 Å) with ARG or LYS (only NMR structures).

		Full Dataset	Non-redundant Dataset
Number of protein-RNA PDB entries available		**107 (*1709*)[a]**	**89 (*135*)[a]**
Number of protein-RNA PDB entries analysed		**107 (*1709*)[a]**	**89 (*135*)[a]**
Number of analysed nucleotides with the 2'OH in contact with:	ARG	**1756[b]**	**212[b]**
	LYS	**1647[b]**	**152[b]**

[a] Number of NMR models for the given set of PDB entries.

[b] Removing repeated kappa/pucker values due to contacts with different atoms of the same aminoacid in a given contact.

Supplementary Table 3. Protein-RNA contacts analysis.

	Full Dataset	Non-redundant Dataset
Number of available PDB entries	**514**	**319**
Number of available X-ray entries	**407**	**230 (*238*)[b]**
Number of available NMR entries	**107 (*1709*)[c]**	**89 (*135*)[c]**
Total number of available models (X-RAY+NMR)	**2116**	**373**
Number of analysed PDB entries	**500**	**307**
Total Number of analysed models (X-RAY+NMR)	**2102**	**361**
Number of analysed contacts (distance<=4Å)	**26760[a]**	**5309[a]**

[a] Removing repeated counts from different atoms of the same aminoacid in a given contact.

[b] Number of X-RAY models for the given set of PDB entries.

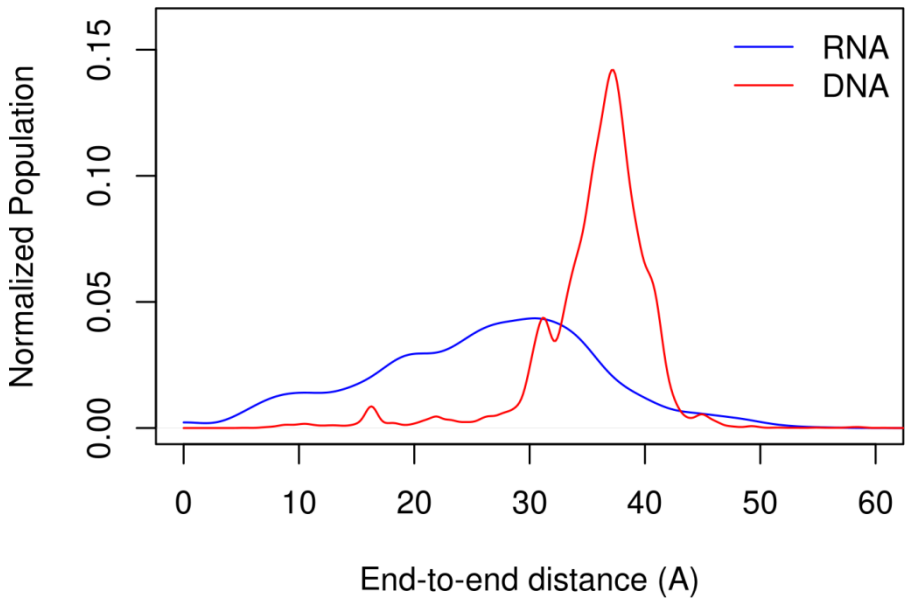
[c] Number of NMR models for the given set of PDB entries.

Supplementary Table 4. H1-CT-OK-HO parameters.

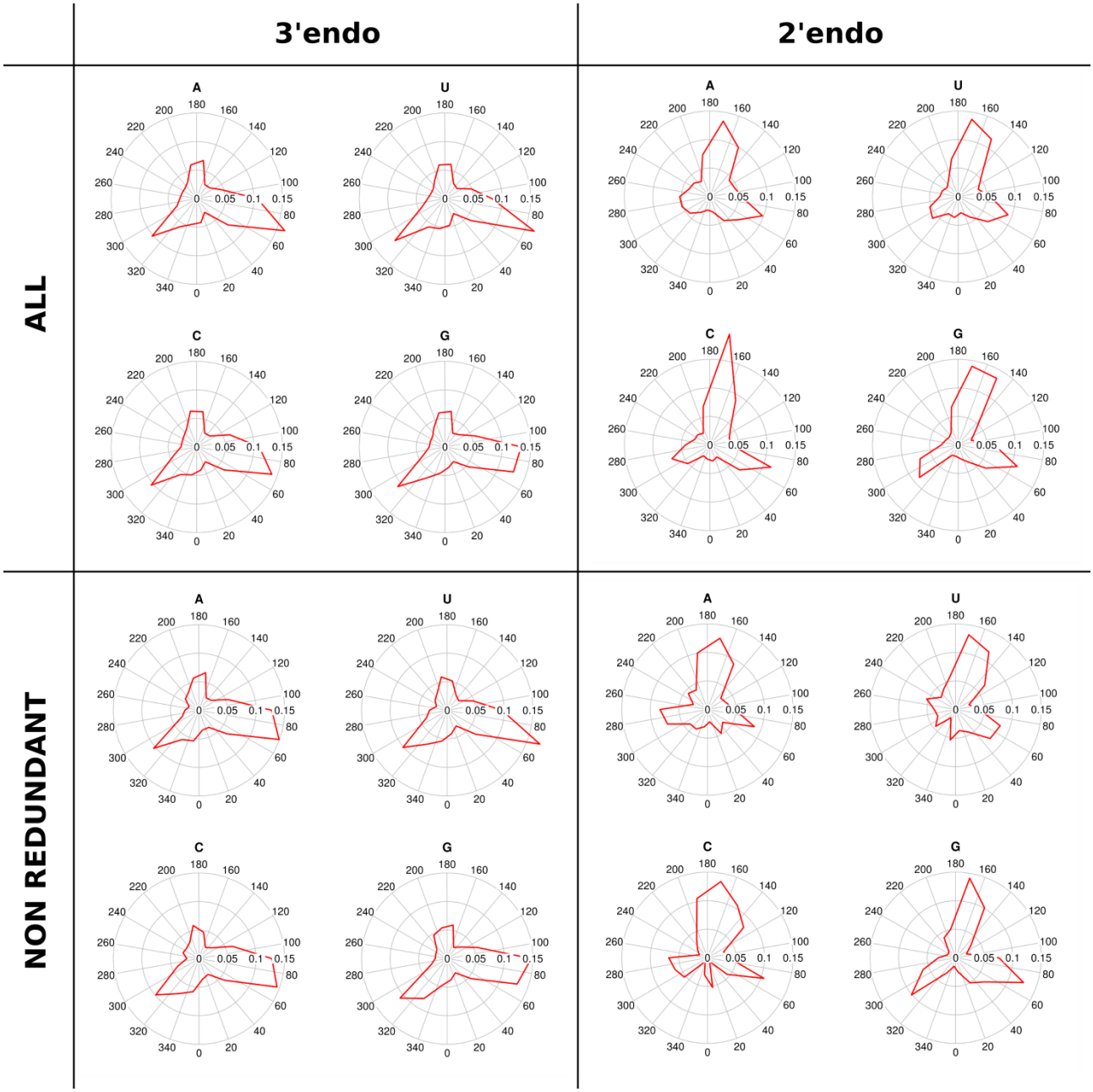
Torsion	$V_n/2$	Phase	Periodicity
H1-CT-OK-HO	0.482	18.8	-3
H1-CT-OK-HO	0.336	59.4	2
H1-CT-OK-HO	0.549	96.9	1

Supplementary Table 5. H1-CT-OK-HO parameters considering vdW specific corrections.

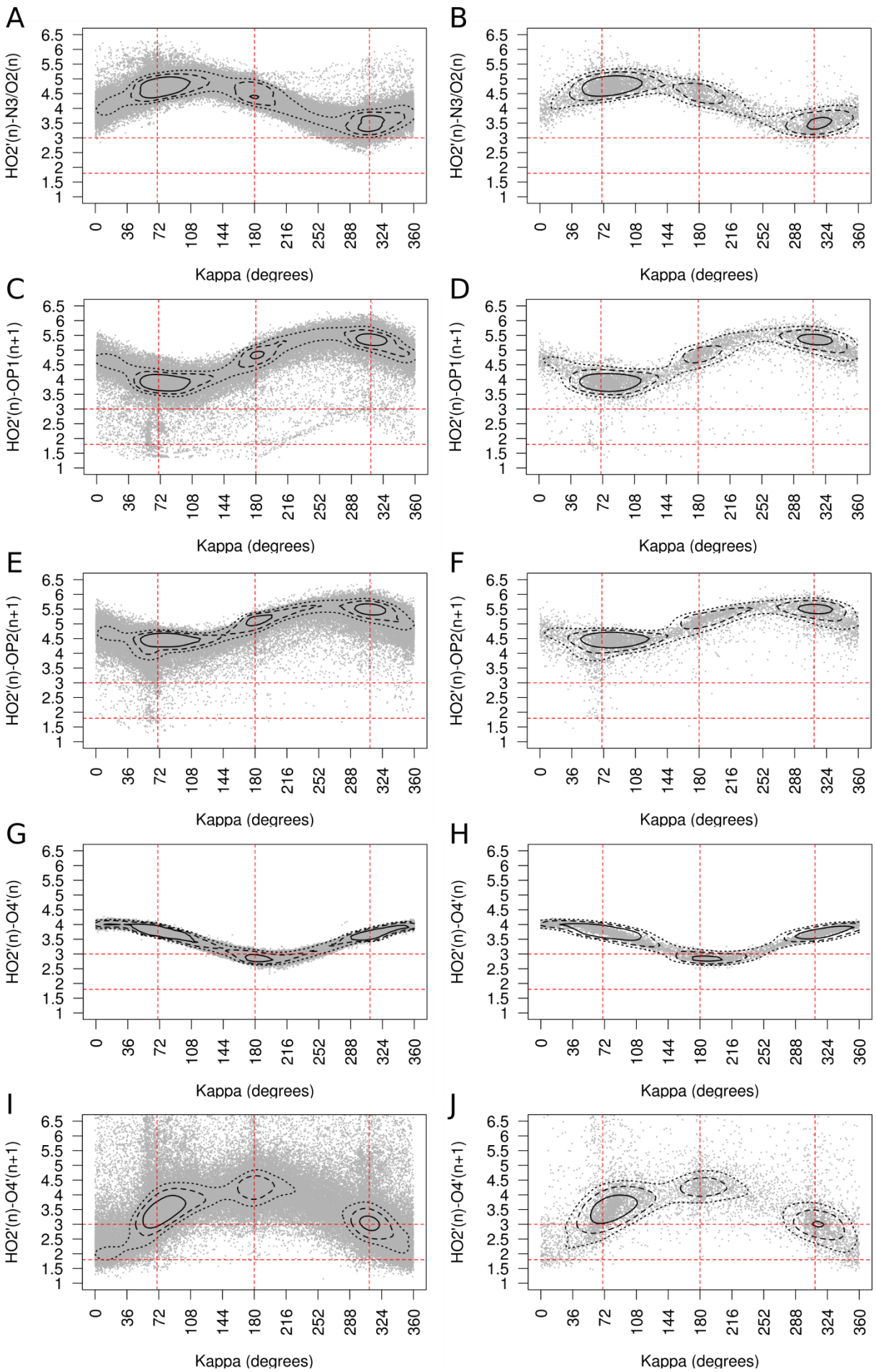
Torsion	$V_n/2$	Phase	Periodicity
H1-CT-OK-HO	0.501	0.0	-3
H1-CT-OK-HO	0.287	74.3	2
H1-CT-OK-HO	0.519	60.7	1



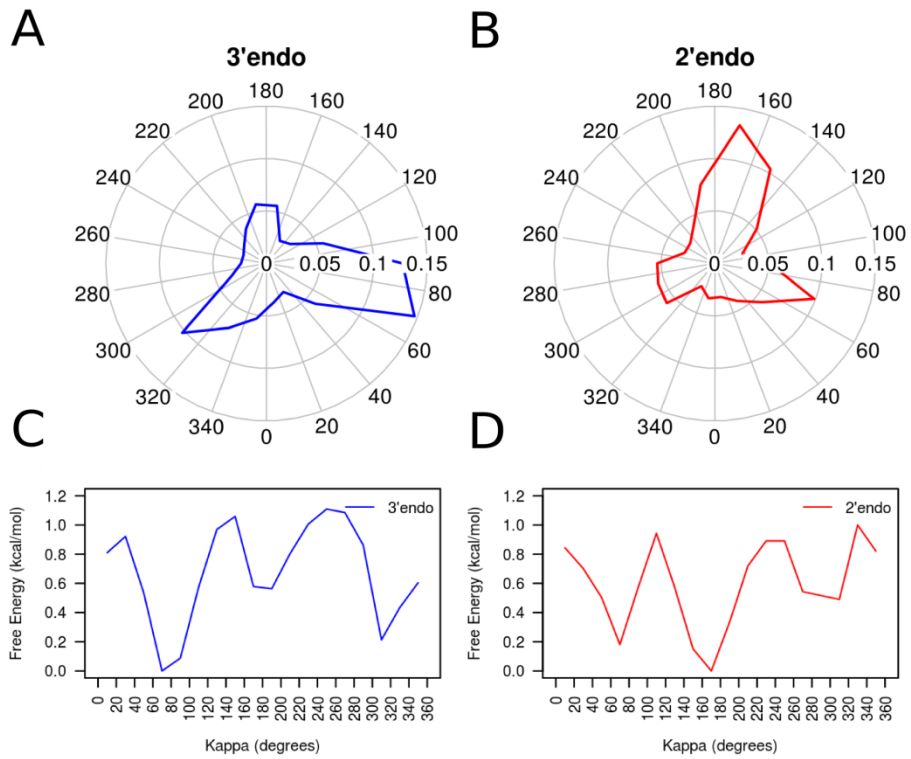
Supplementary Figure 1. End to end distance for all RNA and DNA fragments available in the non-redundant RNA database or in the Protein Data Bank (up to 22nd Nov 2016, with a filter for DNA or DNA-protein entries). Contiguous fragments of twelve residues are considered for the distance measurement discarding those shorter. The distance is defined between the C1' atoms in the 5' and 3' terminal residues.



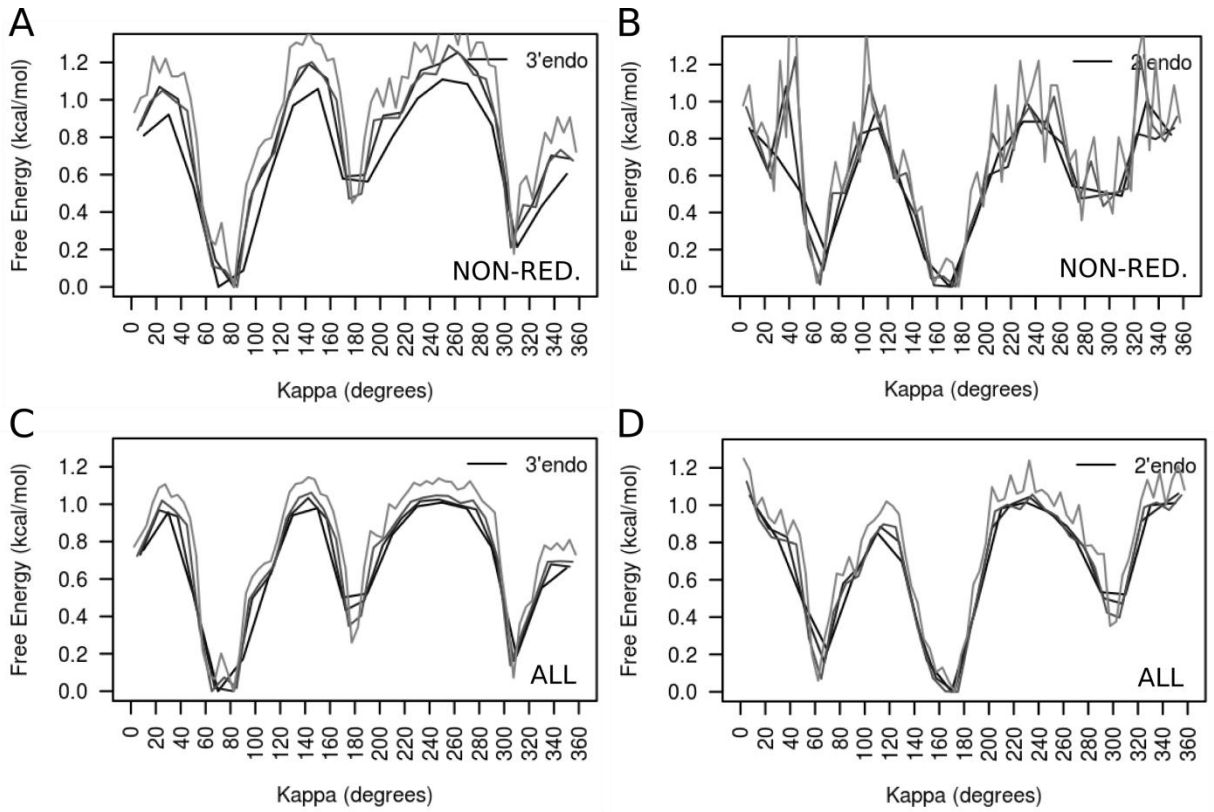
Supplementary Figure 2. Preferred orientations of the kappa torsion per base type. The plots show the probability distribution of the torsion angle between the atoms H2'-C2'-O2'-HO2' for 3'endo or 2'endo ribonucleotides and for the current state of the PDB or a non-redundant database (see Supplementary Methods 1), split by base type.


A
HO2'(n)-N3/O2(n)
B
HO2'(n)-N3/O2(n)
C
HO2'(n)-OP1(n+1)
D
HO2'(n)-OP1(n+1)
E
HO2'(n)-OP2(n+1)
F
HO2'(n)-OP2(n+1)
G
HO2'(n)-O4'(n)
H
HO2'(n)-O4'(n)
I
HO2'(n)-O4'(n+1)
J
HO2'(n)-O4'(n+1)
Kappa (degrees)
0 36 72 108 144 180 216 252 288 324 360
1 1.5 2 2.5 3 3.5 4 4.5 5 5.5 6 6.5

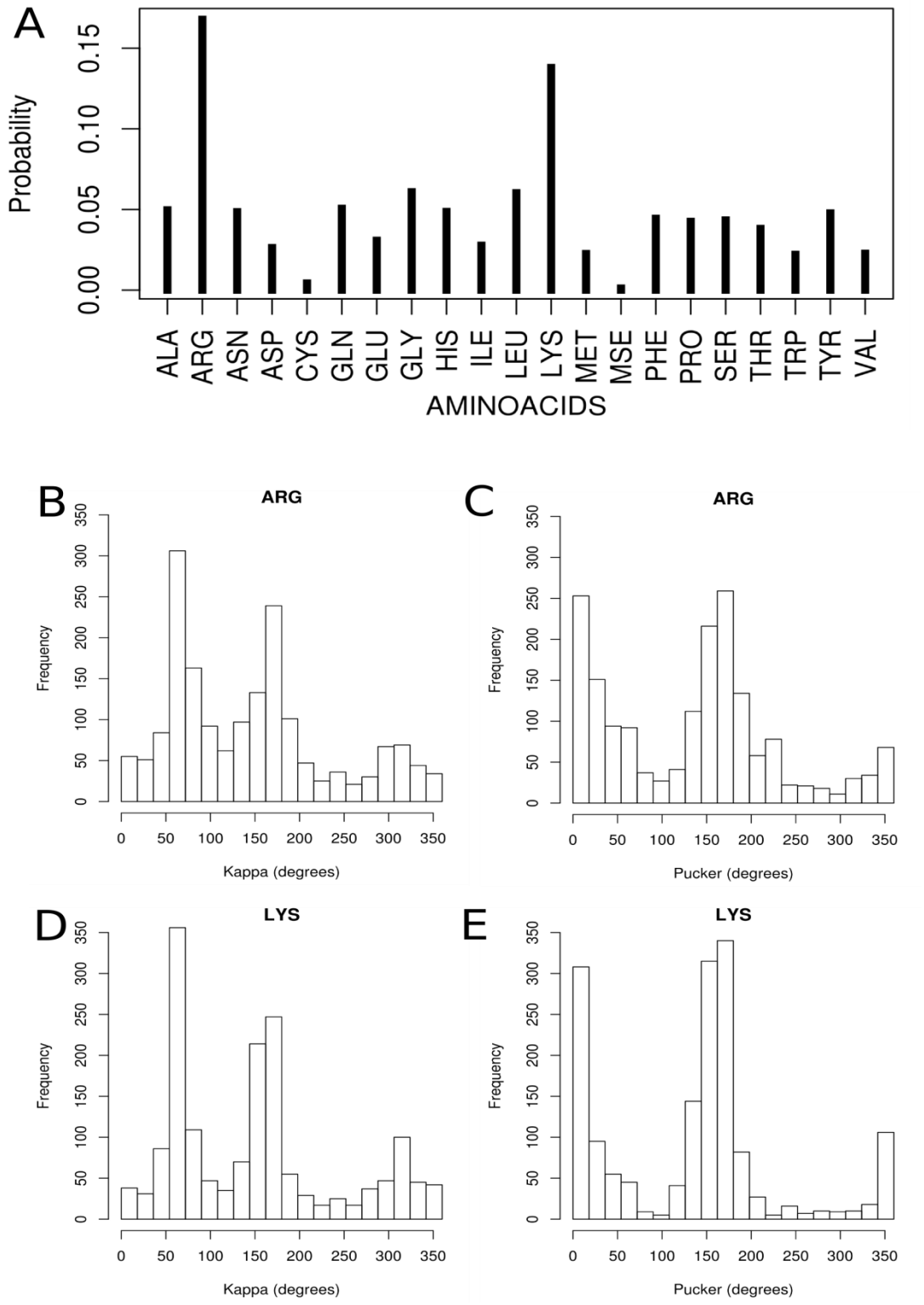
Supplementary Figure 3. Possible hydrogen bonds acceptor/donors nearby the 2'OH group. Scatter plots of kappa torsion vs distance between HO2' and local acceptors of hydrogen bonds, are shown for nucleotides with pucker phase in North for both Full Dataset (A, C, E, G and I) and Non-Redundant Dataset (B, D, F, H and J). Red dotted lines indicate optimal and maximum hydrogen bond distances (horizontal), and kappa 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.



Supplementary Figure 4. Preferred orientations of the kappa torsion from a non-redundant database. (A) 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 a non-redundant database (see Supplementary Methods 1). (B) Same as in (A) but for 2'endo ribonucleotides. (C,D) Empirical free energy calculated from the experimental kappa distributions in (A) and (B), respectively.

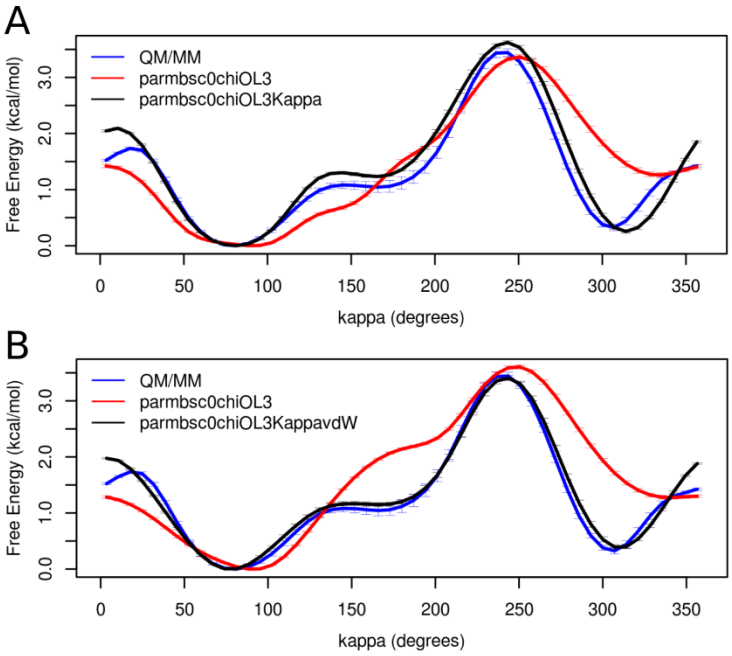


Supplementary Figure 5. Kappa energy profile for different window sizes. (A) Empirical free energy calculated from the kappa distribution of 3'endo ribonucleotides of the non-redundant RNA dataset (see Supplementary Methods 1), splitting the data using four different window sizes: 20 degress (black), 15 degrees (dark gray), 10 degrees (gray), and 5 degrees (light gray). (B) Same as in (A) but for 2'endo ribonucleotides. (C) Same as in (A) but using all current RNA entries in the PDB. (D) Same as (C) but for 2'endo ribonucleotides.

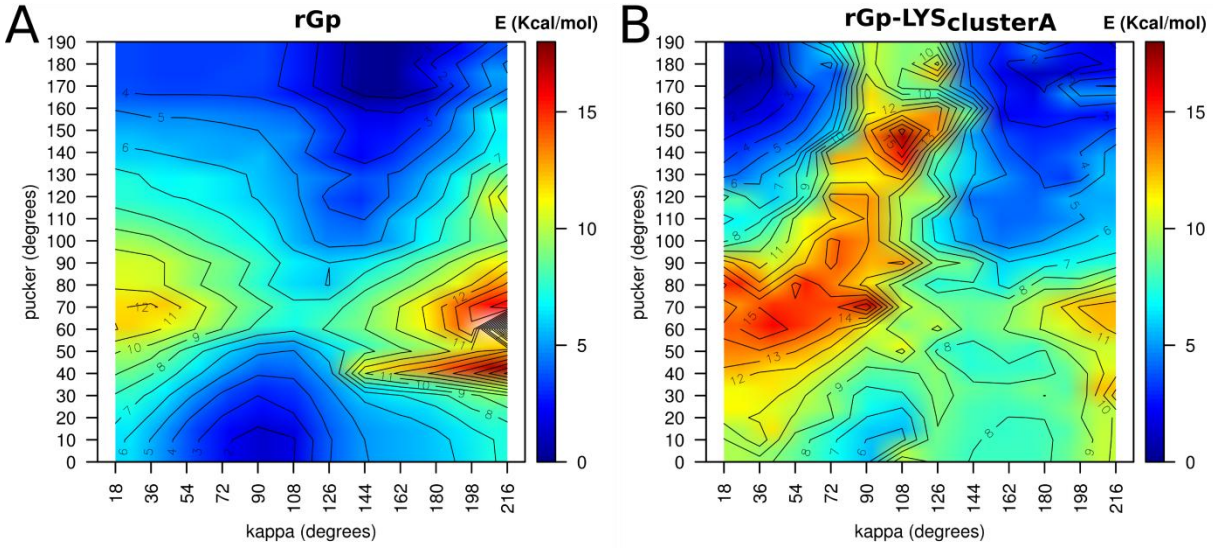


Supplementary Figure 6. Protein-RNA contacts for the Full Dataset. (A) Probability of contact between a given aminoacid and the 2'OH given a protein-RNA contact occur, calculated from counting all contacts (distance<=4 Å) between any protein atom and the oxygen of 2'OH, and splitting the counts per amino acid identity. Multiple atoms of a given aminoacid within the distance cutoff were counted as one contact. All X-ray and NMR

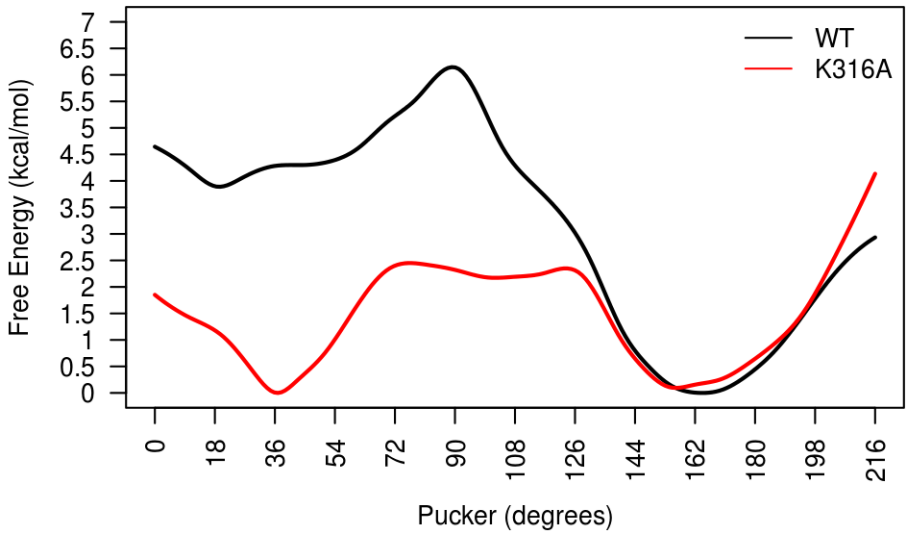
(multiple models) from the Full Dataset (see Supplementary Methods 1) were used. B) Frequency of kappa values for RNA nucleotides in contact with ARG atoms (distance <=4 Å) obtained from NMR (multiple models) structures in the Full dataset. C) Frequency of pucker phase values for RNA nucleotides in contact (distance <=4 Å) with ARG atoms obtained from NMR (multiple models) structures in the Full Dataset. D,E) Same as B and C, respectively, but for LYS amino acid.



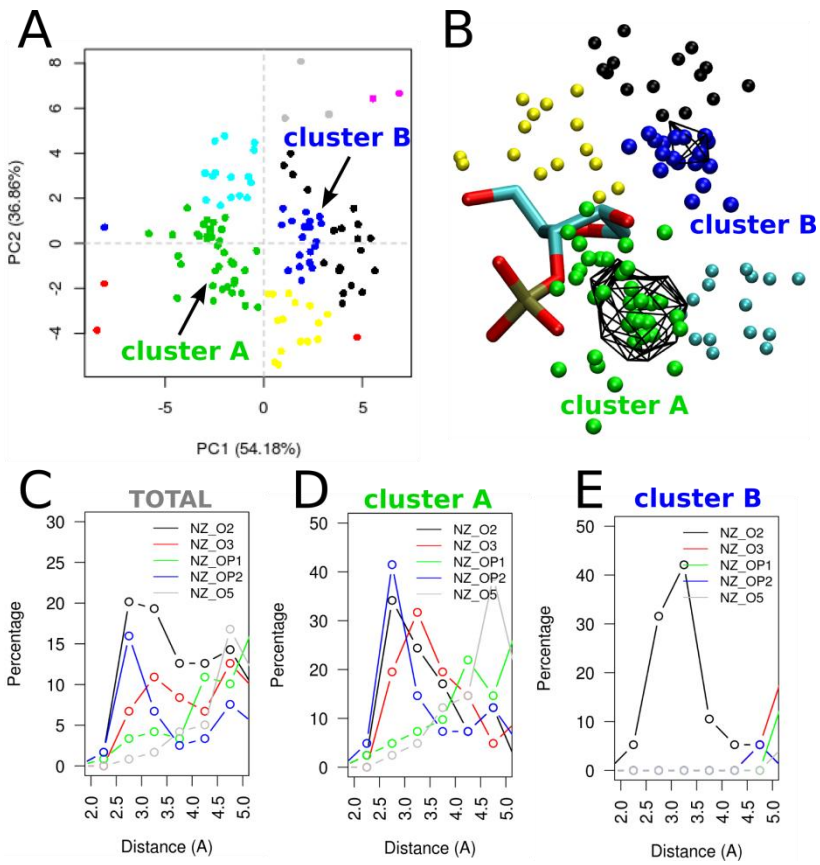
Supplementary Figure 7. Kappa fitting to reproduce QM/MM potential of mean force. (A) US QM/MM (blue), MM parmbsc0chiOL3$_{H1\text{-}CT\text{-}OK\text{-}HO=0}$ (red) and MM parmbsc0chiOL3 with the correction on the kappa torsion (parmbsc0chiOL3Kappa, black) free energy profiles for the kappa torsion of a rCC dinucleotide. The profile and error bars correspond to the average and standard deviation from five energy profiles obtained after 20-25ps every 1 ps (QM/MM) and 2-2.5ns every 100ps (parmbsc0chiOL3$_{H1\text{-}CT\text{-}OK\text{-}HO=0}$ and parmbsc0chiOL3Kappa). (B) Same as in (A) but including the Lennard-Jones modification (see Supplementary Methods 4) on parmbsc0chiOL3$_{H1\text{-}CT\text{-}OK\text{-}HO=0}$ (red) and on parmbsc0chiOL3Kappa (black).



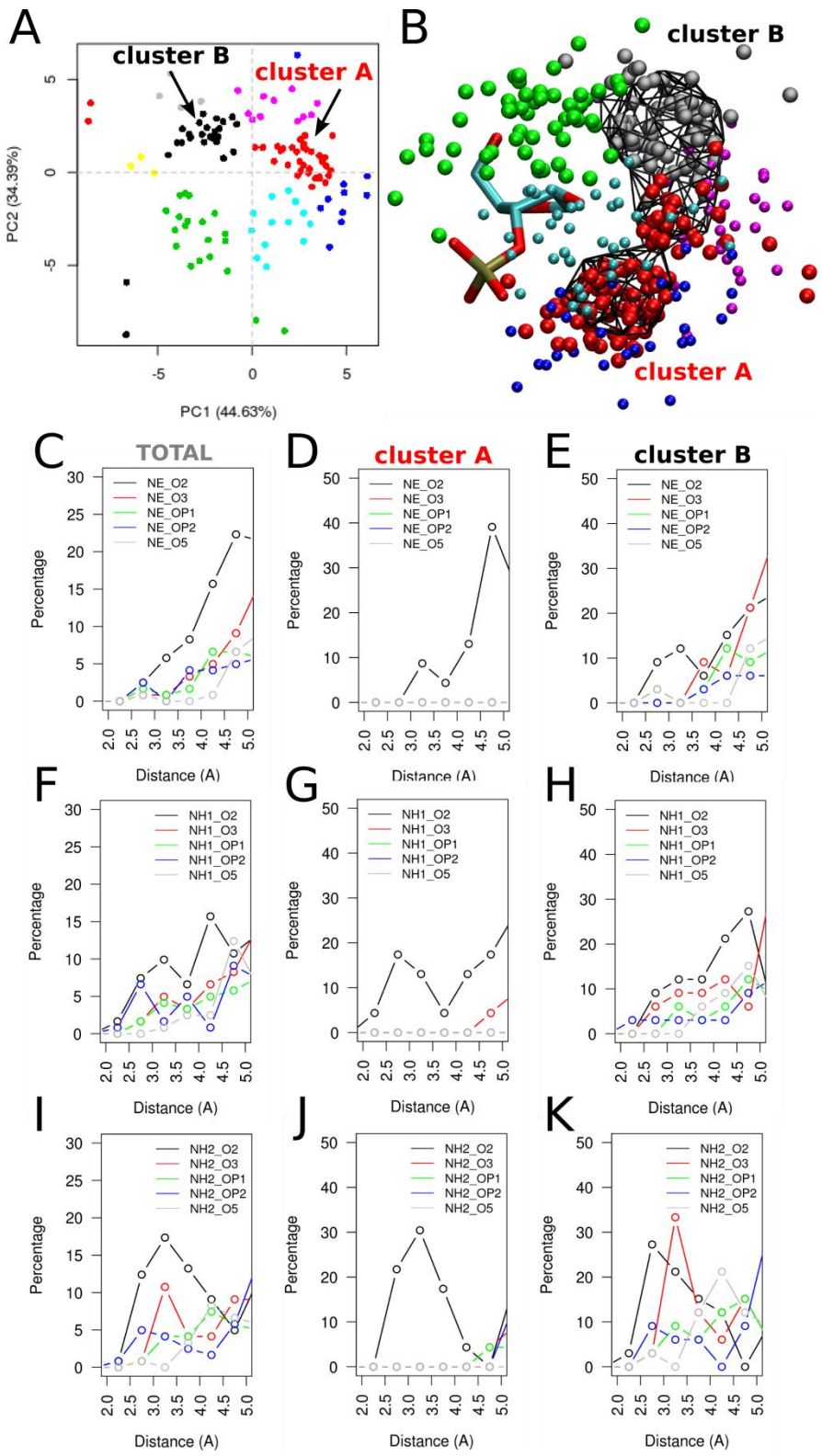
Supplementary Figure 8: κ vs puckering QM/SCRF potential energy surfaces for guanine mono-phosphate in the absence (A) and presence (B) of a lysine analogue (methyl-ammonium) placed at the most populated position (cluster A, see Supplementary Figure 10) for cationic residues nearby the 2'OH of south puckering nucleotides in the non-redundant Database.



Supplementary Figure 9: Puckering PMF of cytosine 6 in the MIWI PAZ domain-RNA complex (PDB ID: 2XFM; model 6) for the wild type (black trace) where Lys 316 is located at cluster A (see Supplementary Figure 10) and for the Lys316Ala mutant.



Supplementary Figure 10. Lysine localization nearby the 2'OH atom in south puckering RNA nucleotides. (A) PCA-based clustering of LYS NZ atom (projection on principal components 1 and 2) for all occurrences of LYS residues within 4 Å of the O2' of south puckering nucleotides in protein-RNA complexes of the non-redundant database. The two most populated clusters are labeled. (B) Position of the LYS ammonium atom around a nucleotide (base omitted for clarity), for the most populated clusters considered in (A), coloured by cluster identity. Occupancy isosurfaces correponding to 60% of maximum occupancy are shown as a black wireframe. (C) Histogram of the distance between NZ atom in LYS and O2', O3', OP1, OP2 and O5' in RNA for all LYS residues considered in (A). (D,E) Same as in (C) but for the two most populated clusters.



Supplementary Figure 11. Arginine localization nearby the 2'OH atom in south puckering RNA nucleotides. (A) PCA-based clustering of ARG CZ atom (projection on principal components 1 and 2) for all occurrences of ARG residues within 4 Å of the O2' of south puckering nucleotides in protein-RNA complexes of the non-redundant database. The two

most populated clusters are labeled. (B) Position of the ARG NE, NH1 and NH2 atoms around a nucleotide (base omitted for clarity), for the most populated clusters considered in (A), coloured by cluster identity. Occupancy isosurfaces correponding to 60% of maximum occupancy are shown as a black wireframe. (C) Histogram of the distance between NE atom in ARG and O2', O3', OP1, OP2 and O5' in RNA for all ARG residues considered in (A). (D,E) Same as in (C) but for the two most populated clusters. (F,I) Same as in (C) but for NH1 and NH2 atom in ARG. (G,H) Same as (F) but for the two most populated clusters. (J,K) Same as in (I) but for the two most populated clusters.

Supplementary References

1. Grant, B. J., Rodrigues, A. P. C., ElSawy, K. M., McCammon, J. A. & Caves, L. S. D. Bio3d: an R package for the comparative analysis of protein structures. *Bioinformatics* **22,** 2695–2696 (2006).
2. Leontis, N. B. & Zirbel, C. L. in *RNA 3D Structure Analysis and Prediction* (eds. Leontis, N. & Westhof, E.) **27,** 281–298 (Springer Berlin Heidelberg, 2012).
3. Westhof, E. & Sundaralingam, M. A method for the analysis of puckering disorder in five-membered rings: the relative mobilities of furanose and proline rings and their effects on polynucleotide and polypeptide backbone flexibility. *J. Am. Chem. Soc.* **105,** 970–976 (1983).
4. Lee, T.-S., Radak, B. K., Pabis, A. & York, D. M. A New Maximum Likelihood Approach for Free Energy Profile Construction from Molecular Simulations. *J. Chem. Theory Comput.* **9,** 153–164 (2013).
5. Cheatham, T. E., Cieplak, P. & Kollman, P. A. A Modified Version of the Cornell *et al.* Force Field with Improved Sugar Pucker Phases and Helical Repeat. *J. Biomol. Struct. Dyn.* **16,** 845–862 (1999).
6. Cornell, W. D. *et al.* A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **117,** 5179–5197 (1995).
7. Pérez, A. *et al.* Refinement of the AMBER Force Field for Nucleic Acids: Improving the Description of α/γ Conformers. *Biophys. J.* **92,** 3817–3829 (2007).

8. Zgarbová, M. *et al.* Refinement of the Cornell et al. Nucleic Acids Force Field Based on Reference Quantum Chemical Calculations of Glycosidic Torsion Profiles. *J. Chem. Theory Comput.* **7,** 2886–2902 (2011).
9. Banáš, P. *et al.* Performance of Molecular Mechanics Force Fields for RNA Simulations: Stability of UUCG and GNRA Hairpins. *J. Chem. Theory Comput.* **6,** 3836–3849 (2010).
10. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79,** 926 (1983).
11. Berendsen, H. J. C., Grigera, J. R. & Straatsma, T. P. The missing term in effective pair potentials. *J. Phys. Chem.* **91,** 6269–6271 (1987).
12. Dang, L. X. & Kollman, P. A. Free Energy of Association of the K+:18-Crown-6 Complex in Water: A New Molecular Dynamics Study. *J. Phys. Chem.* **99,** 55–58 (1995).
13. Dang, L. X. Mechanism and Thermodynamics of Ion Selectivity in Aqueous Solutions of 18-Crown-6 Ether: A Molecular Dynamics Study. *J. Am. Chem. Soc.* **117,** 6954–6960 (1995).
14. Smith, D. E. & Dang, L. X. Computer simulations of NaCl association in polarizable water. *J. Chem. Phys.* **100,** 3757 (1994).
15. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. & Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81,** 3684 (1984).
16. Salomon-Ferrer, R., Götz, A. W., Poole, D., Le Grand, S. & Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J. Chem. Theory Comput.* **9,** 3878–3888 (2013).
17. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An N· log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **98,** 10089 (1993).
18. Ryckaert, J.-P., Ciccotti, G. & Berendsen, H. J. . Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **23,** 327–341 (1977).

19. Henriksen, N. M., Roe, D. R. & Cheatham, T. E. Reliable Oligonucleotide Conformational Ensemble Generation in Explicit Solvent for Force Field Assessment Using Reservoir Replica Exchange Molecular Dynamics Simulations. *J. Phys. Chem. B* **117,** 4014–4027 (2013).
20. Barducci, A., Bussi, G. & Parrinello, M. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Phys. Rev. Lett.* **100,** (2008).
21. Huang, M., Giese, T. J., Lee, T.-S. & York, D. M. Improvement of DNA and RNA Sugar Pucker Profiles from Semiempirical Quantum Methods. *J. Chem. Theory Comput.* **10,** 1538–1545 (2014).
22. Tribello, G. A., Bonomi, M., Branduardi, D., Camilloni, C. & Bussi, G. PLUMED 2: New feathers for an old bird. *Comput. Phys. Commun.* **185,** 604–613 (2014).
23. Ufimtsev, I. S. & Martinez, T. J. Quantum Chemistry on Graphical Processing Units. 3. Analytical Energy Gradients, Geometry Optimization, and First Principles Molecular Dynamics. *J. Chem. Theory Comput.* **5,** 2619–2628 (2009).
24. Titov, A. V., Ufimtsev, I. S., Luehr, N. & Martinez, T. J. Generating Efficient Quantum Chemistry Codes for Novel Architectures. *J. Chem. Theory Comput.* **9,** 213–221 (2013).
25. Götz, A. W., Clark, M. A. & Walker, R. C. An extensible interface for QM/MM molecular dynamics simulations with AMBER. *J. Comput. Chem.* **35,** 95–108 (2014).
26. Isborn, C. M., Götz, A. W., Clark, M. A., Walker, R. C. & Martínez, T. J. Electronic Absorption Spectra from MM and *ab Initio* QM/MM Molecular Dynamics: Environmental Effects on the Absorption Spectrum of Photoactive Yellow Protein. *J. Chem. Theory Comput.* **8,** 5092–5106 (2012).
27. Grimme, S., Antony, J., Ehrlich, S. & Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **132,** 154104 (2010).
28. Steinbrecher, T., Latzer, J. & Case, D. A. Revised AMBER Parameters for Bioorganic Phosphates. *J. Chem. Theory Comput.* **8,** 4405–4412 (2012).
29. Bergonzo, C. & Cheatham, T. E. Improved Force Field Parameters Lead to a Better

Description of RNA Structure. *J. Chem. Theory Comput.* **11,** 3969–3972 (2015).

30. Kästner J, Carr JM, Keal TW, Thiel W, Wander A, Sherwood P: DL-FIND: An Open-Source Geometry Optimizer for Atomistic Simulations. *J. Phys. Chem. A*, **113**, 11856–11865 (2009).
31. ChemShell, a Computational Chemistry Shell, see www.chemshell.org
32. Sherwood P, de Vries AH, Guest MF, Schreckenbach G, Catlow CRA, French SA, Sokol AA, Bromley ST, Thiel W, Turner AJ, et al.: QUASI: A general purpose implementation of the QM/MM approach and its application to problems in catalysis. *J. Mol. Struct. THEOCHEM*, **632**,1–28 (2003).
33. Ahlrichs R, Bär M, Häser M, Horn H, Kölmel C: Electronic structure calculations on workstation computers: The program system turbomole. *Chem. Phys. Lett.*, **162**,165–169 (1989).
34. Becke AD: Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A*, **38**, 3098–3100 (1988).
35. Lee, Yang, Parr: Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B. Condens. Matter*, **37**,785–789 (1988).
35. Eichkorn K, Weigend F, Treutler O, Ahlrichs R: Auxiliary basis sets for main row atoms and transition metals and their use to approximate Coulomb potentials. *Theor. Chem. Accounts Theory, Comput. Model. (Theoretica Chim. Acta)*, **97**,119–124 (1997).
37. Weigend F, Ahlrichs R, Weigend F, Furche F, Ahlrichs R, Leininger T, Nicklass A, Küchle W, Stoll H, Dolg M, et al.: Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.*, **7**,3297 (2005).
38. Eichkorn K, Treutler O, Öhm H, Häser M, Ahlrichs R: Auxiliary basis sets to approximate Coulomb potentials. *Chem. Phys. Lett.*, **240**,283–290 (1995).
39. Sinnecker S, Rajendran A, Klamt A, Michael Diedenhofen § and, Frank Neese: Calculation of Solvent Shifts on Electronic g-Tensors with the Conductor-Like Screening Model (COSMO) and Its Self-Consistent Generalization to Real Solvents (Direct COSMO-RS). J. Phys. Chem. A, 110,2235–2245 (2006).