The pseudotorsional space of RNA

LEANDRO GRILLE,[1,2,5] DIEGO GALLEGO,[3,4,5] LEONARDO DARRÉ,[2,3] GABRIELA DA ROSA,[1,2] FEDERICA BATTISTINI,[3,4] MODESTO OROZCO,[3,4] and PABLO D. DANS[1,2,3]

[1]Computational Biophysics Group, Department of Biological Sciences, CENUR Litoral Norte, Universidad de la República, 50000 Salto, Uruguay

[2]Bioinformatics Unit, Institute Pasteur of Montevideo, 11400 Montevideo, Uruguay

[3]Molecular Modelling and Bioinformatics Group, Institute for Research in Biomedicine, The Barcelona Institute of Science and Technology, 08028 Barcelona, Spain

[4]Department of Biochemistry and Molecular Biology, Faculty of Biology, University of Barcelona, 08028 Barcelona, Spain

ABSTRACT

The characterization of the conformational landscape of the RNA backbone is rather complex due to the ability of RNA to assume a large variety of conformations. These backbone conformations can be depicted by pseudotorsional angles linking RNA backbone atoms, from which Ramachandran-like plots can be built. We explore here different definitions of these pseudotorsional angles, finding that the most accurate ones are the traditional η (eta) and θ (theta) angles, which represent the relative position of RNA backbone atoms P and C4′. We explore the distribution of $\eta - \theta$ in known experimental structures, comparing the pseudotorsional space generated with structures determined exclusively by one experimental technique. We found that the complete picture only appears when combining data from different sources. The maps provide a quite comprehensive representation of the RNA accessible space, which can be used in RNA-structural predictions. Finally, our results highlight that protein interactions lead to significant changes in the population of the $\eta - \theta$ space, pointing toward the role of induced-fit mechanisms in protein–RNA recognition.

Keywords: bioinformatics; computational chemistry; conformations; data mining; structure

INTRODUCTION

RNA molecules participate in almost any genomic process. The repertoire of RNA molecules is growing continuously: We have known for decades that messenger RNAs (mRNA) transport the genomic information which is transformed into proteins with the help of ribosomal and transfer RNAs, but very recently many other functional RNAs have emerged. For example, long noncoding (lncRNA) and small interfering RNA (siRNA) are essential for the regulation of gene expression and gene activation/silencing, and some microRNAs (miRNAs) are known to trigger mRNA degradation (Hombach and Kretz 2016). Small nuclear RNA (snRNA) is critical for the function of the spliceosome, and ribozymes are RNA molecules capable of catalyzing specific biochemical reactions. This astonishing diversity in function is due to the ability of RNA to assume a surprising variety of conformations (Vicens and Kieft 2022), which structurally speaking places the RNA closer to proteins than to DNA.

At physiological conditions, unlike DNA, which is mostly found as a regular double helix, RNA molecules can adopt several secondary structure motifs. These motifs can in turn bind together through intermolecular interactions like in kissing-loops, or fold guided by intramolecular interactions to form highly complex 3D structures like pseudoknots and junctions (Butcher and Pyle 2011). Although only <5% of the structures determined experimentally and deposited in the Protein Data Bank (PDB) contain RNA, they represent a large variety of those secondary structure motifs. Among the most studied, hairpins with three to 11 bases in the loop region can be found. Some of them have particular structural elements, such as U-turns and A-turns typical of tRNA, while others display a particular sequence preference, like the known GNRA and UNCG tetra-loops common in rRNA. Symmetric and asymmetric internal-loops also encompass a significant number of important motifs, like K-turn, C-loop, docking-elbow, bulged-G, and twist-up motifs, which in most of the cases are composed by a

---

[5]Co-first authors.

Corresponding authors: modesto.orozco@irbbarcelona.org, pdans@pasteur.edu.uy

Article is online at http://www.rnajournal.org/cgi/doi/10.1261/rna.079821.123.

combination of important structural elements like S-turns, base triples, cross-strand stacking, and extruded nucleotides. In addition to this rich conformational landscape, RNA molecules are often found in functionally specific protein–RNA complexes, that is, the ribosome or the spliceosome, where the effect of the protein on the RNA conformation is unclear. Monovalent cations (mainly Na^+ and K^+) and divalent cations also play several important roles, usually being found cocrystallized with RNA. For instance, Mg^{2+} is known to stabilize the structure of particular hairpins (Misra and Draper 1998), or to actively participate in the folding of tRNA molecules (Misra and Draper 1998). This presents several questions regarding how these interactions modify the RNA conformational space, or if certain conformations are preferred when bound to proteins or cations.

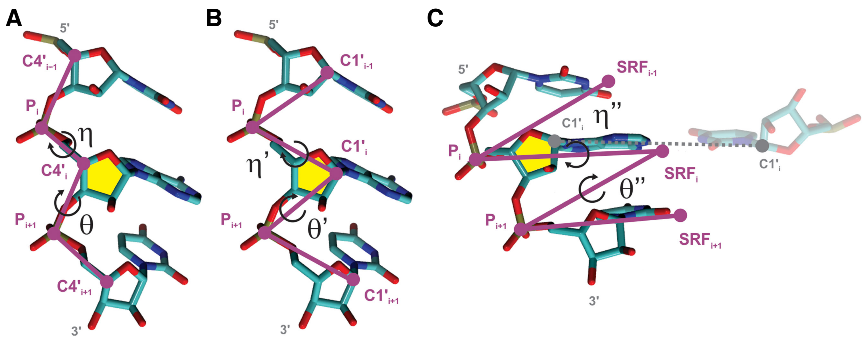

FIGURE 1. Structural representation of the pseudotorsional angles of RNA. (*A*) Definition of the η (eta: $C4'_{i-1}$, P_i, $C4'_i$, P_{i+1}) and the θ (theta: P_i, $C4'_i$, P_{i+1}, $C4'_{i+1}$) angles. (*B*) Definition of the η' (eta': $C1'_{i-1}$, P_i, $C1'_i$, P_{i+1}) and the θ' (theta': P_i, $C1'_i$, P_{i+1}, $C1'_{i+1}$) angles. (*C*) Definition of the η'' (eta'': SRF_{i-1}, P_i, SRF_i, P_{i+1}) and the θ'' (theta'': P_i, SRF_i, P_{i+1}, SRF_{i+1}) angles, where SRF stands for standard reference frame as implemented in DSSR (Lu et al. 2015).

The first attempt to rationalize and classify RNA motifs was based on the nucleobase pairing capabilities, that is, the ability to form hydrogen bonds. Saenger's classification (Saenger 1984) included 29 different base pairs in RNA. Several years later, a new classification defining 12 discrete groups was introduced by Leontis and Westhof (2001), who also introduced a general nomenclature to describe base pairing rules. These groups describe canonical and noncanonical pairs, regardless of the structural motif. However, no information is given about the sugar-backbone conformation. Initial attempts to classify the RNA backbone conformational space led to 37 (Hershkovitz et al. 2003; Murray et al. 2003) and 32 (Schneider et al. 2004) backbone rotamers, highlighting the complexity of the RNA backbone when all the torsions are considered. In 2008, the RNA Ontology Consortium (ROC) updated and refined the clustering of the multidimensional dihedral angle distributions into 46 discrete backbone conformers. More recently, Černý et al. (2020) defined 96 + 1 dinucleotide conformer classes (for RNA and DNA) by analyzing almost six million steps from nonredundant structures found in the PDB. Using a different approach, first Olson (1980), and later Pyle's group (Wadley et al. 2007), proposed the usage of two pseudobonds (traced between the C4′ and P atoms of the backbone) to simplify the backbone conformational space. Considering these pseudobonds, the dihedral space of the phosphodiester linkage can be described by only two pseudotorsional angles: η (eta: $C4'_{i-1}$, P_i, $C4'_i$, P_{i+1}) and θ (theta: P_i, $C4'_i$, P_{i+1}, $C4'_{i+1}$) (Fig. 1A). "Ramachandran-like" plots of $\eta-\theta$ allowed us to identify 11 major backbone conformers, discriminating according to the sugar conformation (North or South). It was demonstrated that nucleotides with similar $\eta-\theta$ values had almost identical backbone conformations, and a relationship was detected between the backbone conformation and the base position, pointing out the ability of $\eta-\theta$ to discriminate between base orientations.

Aiming to increase structural characterization, two other pseudotorsional angles have been defined, namely $\eta'-\theta'$, where the C4′ in $\eta-\theta$ was replaced by C1′ (Keating et al. 2011), and $\eta''-\theta''$, formed between the phosphate atoms and a standardized point in the nucleobase plane (Lu et al. 2015), according with the standard reference frame (SRF) as defined by Olson et al. (2001). $\eta'-\theta'$ was less explored even though in principle it seems more adequate than $\eta-\theta$ to fit the RNA backbone to X-ray electron density maps. The $\eta''-\theta''$ has the additional advantage that it contains information about the nucleobase orientation with respect to the RNA backbone, something not explicitly provided by the other two pairs of pseudotorsions. No comprehensive study has been published comparing the informational load of these pseudotorsional angles.

In recent years, the number of structures determined by X-ray crystallography (X-ray), nuclear magnetic resonance (NMR), or cryogenic electron microscopy (Cryo-EM) techniques, and available in public databases, has grown to the point where there is enough nonredundant data to extract statistical meaningful information. This allows us to compare isolated RNA molecules versus those bound to proteins, finding any potential bias related to the experimental sources of 3D structures, or to start adding sequence context to major conformers. Accordingly, in this work, we used an updated version of our veriNA3d R (Gallego et al. 2019) package to revisit the three pairs of pseudotorsional angles ($\eta-\theta$, $\eta'-\theta'$, and $\eta''-\theta''$) as they emerge from nonredundant experimental structures available to date. We found that $\eta-\theta$ were the best pseudotorsions to describe RNA conformational space. We found 14 major states and detected nonnegligible differences in sampling of the different conformations, depending on the technique used to determine the structure. Finally, the presence of binding can lead to the exploration of typically empty regions of the conformational space, pointing toward the impact of induced fit in RNA–protein recognition and structural stabilization for the case of cations.

RESULTS AND DISCUSSION

The continuous growth in number of experimental RNA structures creates demands for periodic updates of the analyses based on structural data mining. Since the seminal work of Wadley et al. (2007), the number of deposited RNA structures in public databases like the PDB has increased by a factor of 4.5 (from 1333 in 2007 to 5978 in 2022) (Supplemental Fig. S2), providing a more complete picture of the RNA conformational space. We considered a complete data set containing all the structures and subdivisions in different subsets presented in Table 1. For example, structures were divided by the experimental techniques used to describe them. In the same way, the *complete-data set* was filtered in "naked" (*naked-subset*) and "interacting" (*protein-subset, monovalent-subset*, and *divalent-subset*) RNA fragments with the purpose of addressing the possible effects on RNA folding introduced by interacting partners ubiquitously found in biological contexts. Statistics on the number of nucleotides and RNA structures analyzed in this work are reported in Table 1. Finally, we analyzed the sequence dependence on the $\eta-\theta$ space and explored how to relate transitions in $\eta-\theta$ clusters with the rotation of canonical backbone torsions α, β, γ, δ, ε, ζ, and χ.

How much did the global $\eta-\theta$ RNA pseudotorsional space change from its conception to date?

The $\eta-\theta$ pseudotorsion (Fig. 1A), was classified according to the sugar-puckering conformation (North nonhelical and South) and analyzed following the same strategy reported by Pyle and coworkers on our *complete-data set*. Surprisingly, although we dramatically increased the number of analyzed structures (Table 1), our results for $\eta-\theta$ showed roughly the original distribution, except for two unreported HDRs for nonhelical nucleotides with North sugars (herein clusters numbered VII and VIII in Fig. 2A); and one new cluster in the $\eta-\theta$/South space (cluster V, Fig. 2B). These results reveal only moderate changes in the sampled experimental $\eta-\theta$ space, indicating that either the RNA motifs space was already extensively explored by the X-ray structures available at that time, or that certain motifs remain elusive to experimental techniques.

TABLE 1. Summary of data sets used in this work and number of nucleotides and structures included in each of them

	Structures	Nucleotides		
	Total	North	South	Total
Pyle data set (Wadley et al. 2007)	73	6813	811	7624
complete-data set	1756	94,696	15,461	110,157
xray-subset	1109	45,299	6946	52,245
nmr-subset	601	12,194	1471	13,665
em-subset	56	47,156	8774	55,930
naked-subset	1236	42,632	5042	47,674
protein-subset	850	47,997	9761	57,758
monovalent-subset	55	1857	493	2350
divalent-subset	244	6960	1293	8253

The nonhelical North HDR described originally by Pyle and coworkers have either helical η ($145° < \eta < 190°$) or helical θ ($190° < \theta < 245°$) values. Notably, the new clusters found in the $\eta-\theta$ North space (VII and VIII) have the same pattern. Therefore, all HDRs fall within the black dashed lines in Figure 2A, indicating that North nucleotides are usually located in the frontier of helical motifs. In contrast, the RNA nucleotides with sugar moieties in C2′-endo are, as expected, mostly found outside the helical region defined by black dashed lines in Figure 2B.

We investigated the structural composition of the identified clusters using the RMSD values for all trinucleotides of a given HDR, and a representative one was determined as the most similar to the others in the same region (the structure of all representative nucleotides is shown in Fig. 2C for new clusters reported herein and Supplemental Fig. S3). Using Pyle's similarity score (Wadley et al. 2007; see Table 2), we found that the new clusters VII–VIII in $\eta-\theta$/North and cluster V in $\eta-\theta$/South displayed high similarity percentages (98.8, 96.0, and 93.3, respectively), and therefore they are formed by a group of nucleotides with the same backbone conformation.

The stacking was described geometrically using the R-vector from base (i) to base ($i \pm 1$) defined by the Bussi group (Bottaro et al. 2014) and energetically from thousands of QM calculations (see Materials and Methods). In this way, a systematic description of each HDR for both South and North sugar conformations was obtained (see Supplemental Table S1). The results in Figure 2D for clusters I in both North/South sugar conformations, display a range of energy that covers values from slightly repulsive stacking contributions to very stable ones (2 to −14 kcal/mol). Note how stacking interactions of the central nucleotide with its nearest neighbors are key for explaining conformations found in North cluster I (nHDR-I), while one to three interactions are the main players in stabilizing the structures found in cluster I/South (Fig. 2D). In this particular cluster, most trinucleotides come from internal-loops (49%) or junction-loops (20%), which are secondary motifs that contain several unpaired bases, some of them having flipped-out nucleobases as shown in Figure 2I.

The intramolecular hydrogen bond capability of the 2′OH group of the ribose in RNA (Darre et al. 2019), is another

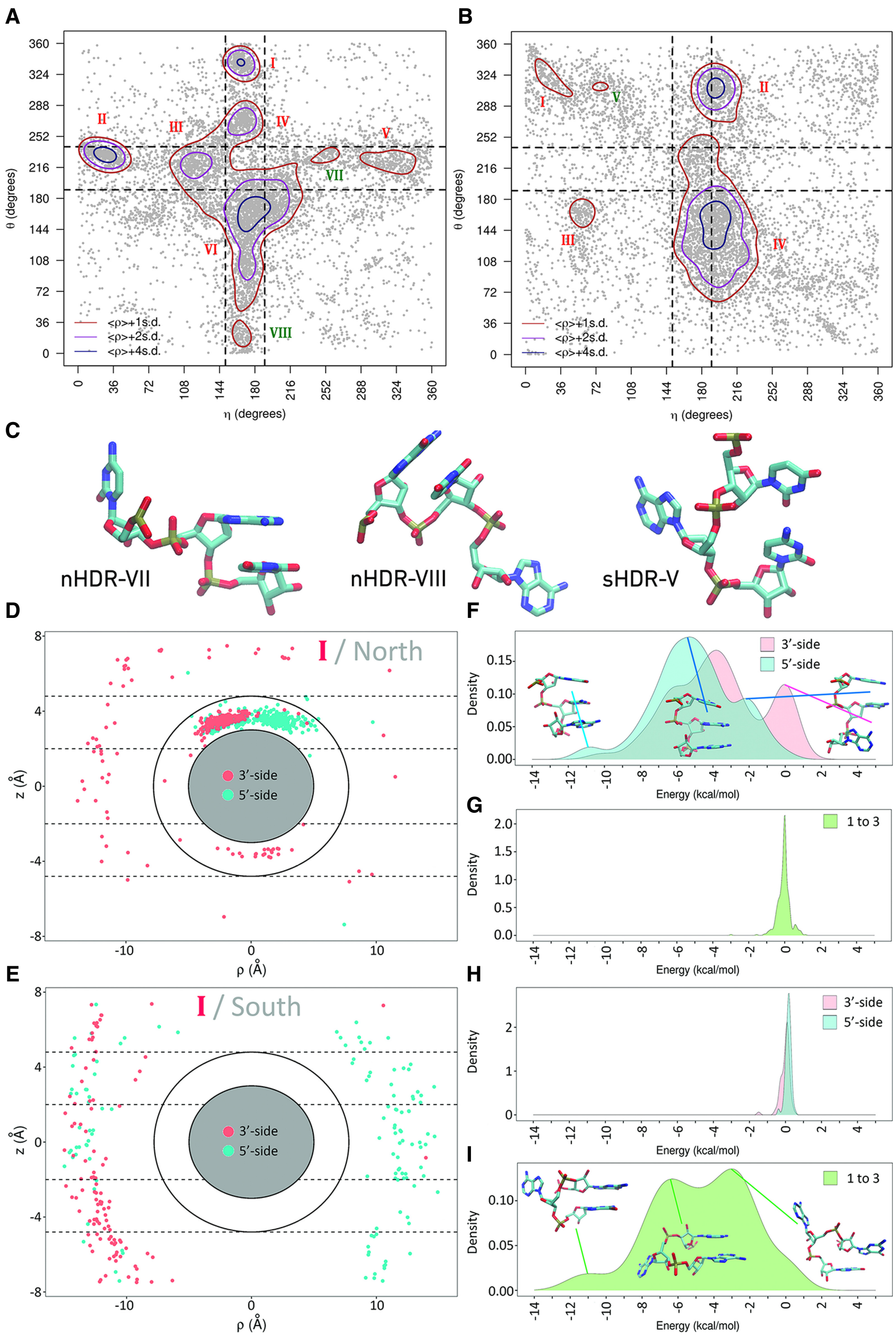

FIGURE 2. (*A*) $\eta - \theta$ from the complete-data set for nonhelical nucleotides with sugar conformation in North. Density contours of $\bar{\rho} + X \cdot \sigma$ ($X = 1$, 2, or 4; dark red, violet, and dark blue, respectively) highlight regions of the plots with a significant population of nucleotides. Cluster or HDRs previously identified by Pyle and coworkers are labeled in red, while new clusters found in this work appear in dark green. Black dashed lines delimit the region occupied by canonical helical A-form. (*B*) Same as (*A*) for nucleotides in South configuration. (*C*) Representative structure for each new cluster found. (*D*) Example of the stacking analysis for cluster I North. ρ and *z* components of the R-vector (Bottaro et al. 2014) between the given nucleobase and its 5′ (green) and 3′ (pink) neighboring nucleobases are shown. The inner space between ellipses and dotted lines shows the region in which stacking occurs. (*E*) Same as (*D*) for cluster I South. (*F*) Interaction energy between the given nucleobase and its 5′ (green) and 3′ (pink) neighboring for cluster I North. (*G*) Direct interaction energy between the 5′ and 3′ nucleobases (one to three interactions) for cluster I North. (*H*) Same as (*F*) for cluster I South. (*I*) Same as (*G*) for cluster I South.

TABLE 2. Summary of clusters' similarities, and the number of PDBs in each cluster for $\eta - \theta$ conformational space

Puckering	Cluster	Similarity	*n*
North	I	97.3	404
North	II	99.2	600
North	III	89.5	327
North	IV	95.2	309
North	V	93.8	242
North	VI	73.2	1622
North	VII	98.8	81
North	VIII	96.0	124
South	I	87.0	138
South	II	94.0	631
South	III	93.2	148
South	IV	68.1	2457
South	V	93.3	15

cornerstone contact among HDRs found far from the helical region. As described in detail in Supplemental Table S1, the 2′OH-phosphate interaction is a contact found prevalently in clusters I, II, and V in $\eta - \theta$/North. This shows the versatile role of this interaction in stabilizing noncanonical motifs while also serving as a primary molecular switch contributing to specific protein–RNA recognition (Darre et al. 2019). Our detailed analysis of each HDR and the correlations of $\eta - \theta$ pseudotorsions with all backbone angles also allowed for a more general observation: all HDRs in the $\eta - \theta$ space are interconnected by a seemingly simple combination of torsional rotations (Fig. 3 for η and Supplemental Fig. S4 for θ). Transitions in the η axis appear mainly driven by changes in α, while transitions in θ depend on changes in ζ. This suggests an interpretation for RNA conformational transitions occurring at two levels: (i) a "coarse" level, which allows RNA to move between HDRs, and (ii) a "fine" level on how these movements translate to the all-atom space. This observation is developed below.

Global $\eta' - \theta'$ and $\eta'' - \theta''$ RNA pseudotorsional space

The pseudotorsions $\eta' - \theta'$ and $\eta'' - \theta''$ in their definition share similarities with the original $\eta - \theta$ torsions (Fig. 1B, C). In $\eta' - \theta'$ the C1′ atom is chosen instead of C4′, while in $\eta'' - \theta''$ the SRF as defined by Olson and coworkers (Olson et al. 2001) is used to account for the relative orientation of the base with respect to the backbone. It should be noted that the location of the SRF depends upon the width of base-pair C1′ – C1′ spacing and the pivoting of complementary bases in the base-pair plane, and hence to analyze single-stranded RNA an idealized base-pair reference is needed.

A first inspection of the $\eta' - \theta'$ space (Supplemental Fig. S5A,B) revealed a very similar distribution to the one provided by $\eta - \theta$, but with lower power to discriminate backbone substates as less clusters are visible in the North/South planes (nine in $\eta' - \theta'$ vs 13 in $\eta - \theta$, compare Fig. 2 with Supplemental Fig. S5). Despite its reduced descriptive power at the backbone level, the $\eta' - \theta'$ pseudotorsional space might be advantageous for mapping X-ray models to their electron density maps, given that C1′ is used as a key reference atom to fit the experimental densities into molecular models (Keating and Pyle 2010; Gruene and Sheldrick 2011; Keating et al. 2011; Li et al. 2019). Coarse-grained models for RNA (Dans et al. 2019) based on $\eta' - \theta'$ could also benefit over $\eta - \theta$ since the C1′ atom is used to define the major/minor groove in curvilinear helicoidal coordinates (Lavery et al. 2014); therefore, the grooves' dimensions would have a direct readout.

To our knowledge, no data mining of PDB has been done to date using the $\eta'' - \theta''$ pseudotorsional space. In our data sets, the $\eta'' - \theta''$ space looks similar to $\eta - \theta$ and $\eta' - \theta'$ (Supplemental Fig. S5C,D), but less defined clusters are observed (eight in total). As a result, the power to discriminate backbone substates is even lower than the $\eta' - \theta'$ one, as North clusters I to IV belong to the same HDR with contour $\bar{\rho} + 1 \cdot \sigma$. Each cluster is then a more heterogenic mixture of backbone conformations, and hence $\eta'' - \theta''$ appears to be not so useful to describe and understand substate diversity as found experimentally.

Additionally, we analyzed the power of each pseudotorsion in explaining the secondary motifs displayed by RNA molecules. This was quantified by analyzing the composition of secondary motifs (stem, stem-end, internal-loop, hairpin-loop, junction-loop, ss-non-loop, and bulge) that are present in the defined clusters. According to Supplemental Figure S6, there are no clear advantages of one pseudotorsion over the others in separating the secondary motif among clusters. Hairpin loops are present in all the clusters but are more abundant in cluster II/North for the three angles. In the same way, internal-loops are more present in South, cluster III in $\eta - \theta$ and cluster XI in $\eta' - \theta' / \eta'' - \theta''$. In summary, even if the $\eta' - \theta'$ definition might have some advantages in terms of modeling, the traditional $\eta - \theta$ space is preferable to describe the conformational space accessible by the RNA backbone.

Experimental techniques describing RNA conformations

Originally, only high-resolution structures obtained by X-ray crystallography were used to analyze the $\eta - \theta$ space (Wadley et al. 2007). However, to date, an important number of RNA structures were determined by means of solution NMR, and more recently by Cryo-EM which has reached a resolution close to that of an X-ray while allowing for the determination of huge RNA complexes (Natchiar et al. 2017). We then divided the *complete-data set* into three subsets, one for each method, in order to analyze

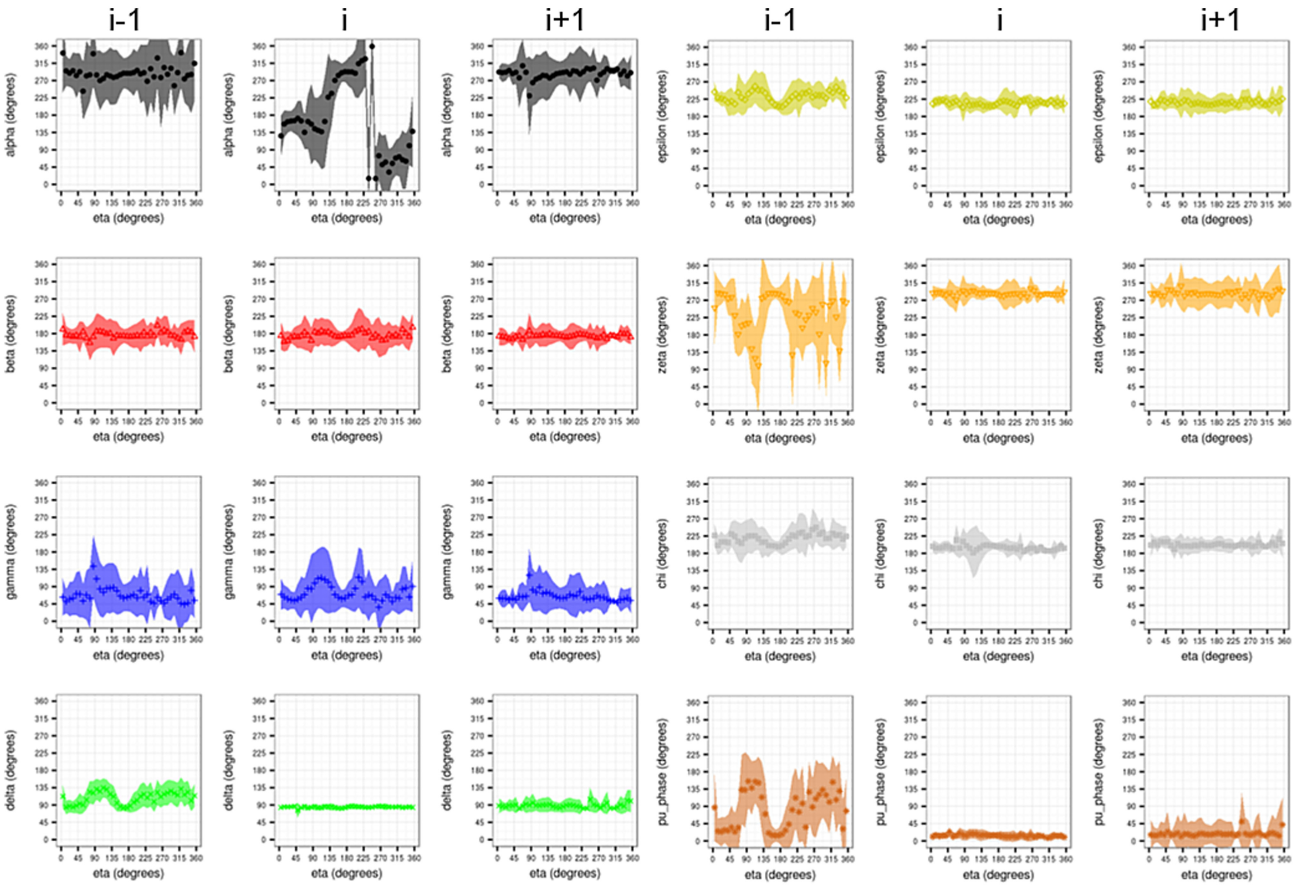

FIGURE 3. Distributions of real dihedral angles in the nucleotides $i-1$ (panels on the *left*), *i* (mid-panels), and $i+1$ (panels on the *right*), as a function of the pseudodihedral angle eta (η) in the central nucleotide *i*. Panels show: α (black), β (red), γ (blue), δ (green), ε (yellow), ζ (orange), ξ (gray), and the puckering phase (brown).

their contributions and limits: *xray-subset*, *nmr-subset*, and *em-subset* (see Materials and Methods). As shown in Table 1, a significant number of nucleotides can still be analyzed despite the division performed.

The *xray-subset* (Fig. 4A,B), although being dramatically enlarged in the number of structures, reproduces well Pyle's $\eta-\theta$ space (Wadley et al. 2007). The *em-subset* (Fig. 4C,D) and *nmr-subset* (Fig. 4E,F) show new clusters VII/VIII in the North plane and cluster V in the South plane. Clusters VII–VIII in $\eta-\theta$/North are mostly exclusively populated by trinucleotides found in large ribosomal complexes that could have been determined with atomic resolution only recently (Nakane et al. 2020). Apart from these two new clusters, trinucleotides in the *em-subset*, extracted from only 32 PDB structures (Table 1), reproduce by themselves all the conformational substates observed in the *complete-data set*. Remarkably, ribosomes seem to exploit geometrically and energetically all the possible backbone substates to achieve their complex 3D folded structures that in turn depend on RNA–protein interactions. Interesting results are obtained when analyzing the pseudotorsional space revealed by solution NMR experiments (*nmr-subset*). As shown in Figure 4E,F, only certain conformations seem to be detected when using this technique (clusters II–III–IV–VI in North and IV–V in South), where large deviations from the canonical $\eta-\theta$ values are not frequently sampled. The new cluster V/South described herein for the *complete-data set* (Fig. 2B) is mainly composed by structures determined using NMR. Furthermore, focusing on the *nmr-subset*, the large and heterogeneous cluster IV/South can be divided into two clearly distinct populations (Fig. 4F).

Being based on completely different approaches and physical observables, which make comparisons rather difficult, the main differences among these three methods might arise from features like the experimental temperature at which data collection is produced. While Cryo-EM typically cools down the sample between 4 K (liquid helium cryogen) and 77 K (nitrogen, propane, or ethane coolers) (Doerr 2017), the structures from the *xray-subset* were collected at 112 ± 44 K and NMR ones at near NTP conditions 293 ± 28 K. Above a given temperature threshold, thermal fluctuations of RNA molecules in solution might be preventing the transient stabilization of some extreme and less stable backbone substates. Consequently, some

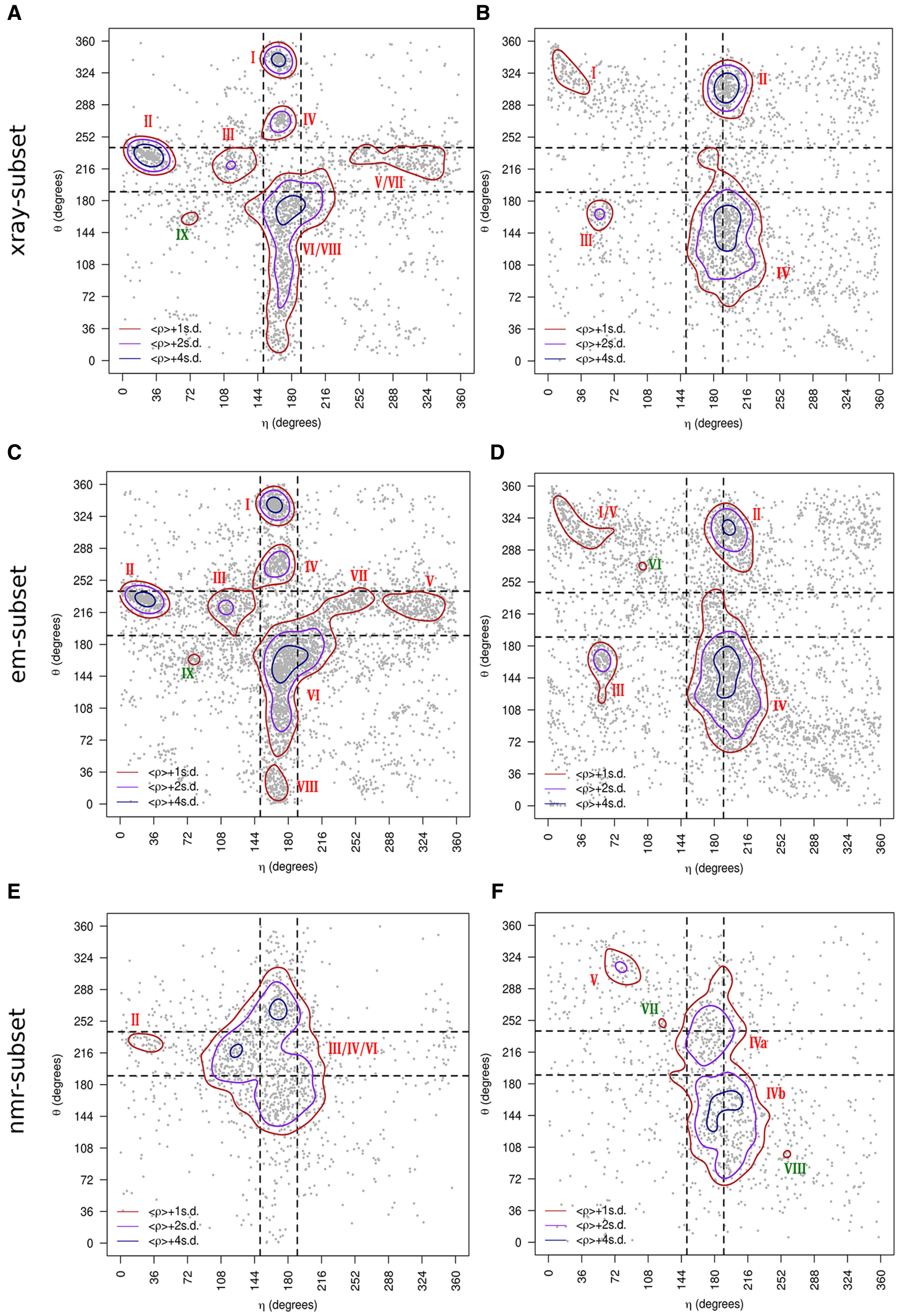

FIGURE 4. $\eta-\theta$ conformational space as detected by each experimental technique. Results are divided into North (panels on the *left*) or South (panels on the *right*) sugar conformations. Density contours of $\bar{\rho}+X\cdot\sigma$ ($X=1$, 2, or 4; dark red, violet, and dark blue, respectively) highlight regions of the plots with a significant population of nucleotides. Black dashed lines delimit the region occupied by canonical helical A-form (those structures are removed from the analyses as described in the Materials and Methods section). Cluster numbers are identified in red or dark green when they were not seen in the *complete-data set* depicted in Figure 2. (*A*) Densities and cluster analysis of the *xray-subset* for central nucleotides with sugars in North. (*B*) Same as (*A*) for South conformations. (*C*,*D*) are the same as (*A*) and (*B*) for the *em-subset*. (*E*,*F*) are the same as (*A*) and (*B*) for the *nmr-subset*.

substates detected by Cryo-EM and X-ray may not be captured when performing NMR experiments. Additionally, the use of smaller X-ray structures to fit large and/or multimeric Cryo-EM densities could be another important reason behind the similarity found between both data sets in terms of eta–theta. Other differences between these three methods might arise from the different size of the RNA molecules in each of the data sets. Apparently, this should not be relevant as the average chain length of the analyzed structures is very similar between NMR and X-ray data sets (26 and 34 residues on average, respectively, while it is 323 for the *em-subset*). However, a closer look into our data sets indicates that both the X-ray and Cryo-EM subsets contain several long RNA chains such as ribosomes with thousands of nucleotides that can sample all the conformational space of the RNA backbone, whereas the longest structure in the NMR set is 155 residues (PDB ID: 2N1Q) and contains mostly A-form RNA. Apparently, the longer an RNA molecule is, the easier it is to explore alternative nonhelical conformations, but this phenomenon could also be influenced by the contacts with proteins, which is explored hereafter.

How do proteins affect the RNA $\eta-\theta$ space?

We divided the *complete*-data set into "naked" RNA and protein-bound RNA. This allowed for the analysis of the intrinsic or "naked" conformations that RNA could adopt (*naked-subset*) and the effect produced by proteins (*protein-subset*). The $\eta-\theta$ plots for North puckering (Fig. 5) reveal that, while the *protein-subset* (RNA with a protein residue closer than 5 Å) roughly resembles the *complete-data set*, the *naked-subset* shows a nucleotide distribution confined to the helical region neighborhood (clusters III, IV, and VI). Clusters I/III/IV show a noticeable reduction in their population and no HDR are found near θ values of ∼0° (new cluster VIII in the *complete-data set*). These differences are an evident consequence of the divergent $\eta-\theta$ density distributions in the absence and presence of proteins. That is, the density of the *naked-subset* biases the sampling of the *complete-data set* toward near-helical regions, while proteins preferentially enrich conformations far from the helical center. Summarizing, these results indicate that proteins can stabilize North pucker RNA segments with η and θ values far from the helical conformation, which are less sampled in "naked" RNA. In the case of South puckering nucleotides, the absence of proteins enriches cluster IV in the helical region, splitting the cluster into two sub-HDR regions.

How do cations affect the RNA $\eta-\theta$ space?

To analyze the effect of the direct contact of monovalent (Na^+ and K^+) and divalent (Ca^{2+}, Mg^{2+}, and Mn^{2+}) cations on $\eta-\theta$, we filtered the *complete-data set*, keeping only those trinucleotides with at least one cation closer than 5 Å. Cation binding to RNA can serve to stabilize a specific 3D structure, perform a catalytic role (Hendrix et al. 2005) and is a requirement for folding (Nguyen et al. 2019). As previously observed (see Table 1), structures containing divalent cations (particular Mg^{2+}) are significantly more abundant than the ones containing Na^+ and K^+ (Stefan et al. 2006) (although the assignment is not always accurate [Auffinger et al. 2011]), while it was proposed that the formation of RNA-structural motifs seems to be Mg^{2+} independent, since this cation binds preferentially to RNA regions others than the well-established secondary conformations (hairpins, internal-loops, junctions, etc.) (Hsiao et al. 2009). Figure 6 and Supplemental Table S2 highlight the main results observed for the *monovalent*- and *divalent-subsets* upon removal of those helical nucleotides. All clusters of the *complete-data set* (Fig. 2A,B) are still present when only the structures with cations are retained (Fig. 6A), several of them being denser (clusters I, III, V, VII, and VII in both North nonhelical and South), and even new minor clusters appeared that were not significant at the *complete-data set* level (Fig. 6B,C). Enhanced clusters involve backbone conformations far from the helical regions, and new ones are located at extreme η values (<72° and >300°). This supports the structural role for a significant portion of the cations determined experimentally that are particularly enriched around backbone conformations highly distorted, stabilizing peculiar RNA motifs. We quantified this observation by computing the number of trinucleotides in helical and nonhelical regions with and without monovalent and divalent cations (Supplemental Table S2). Although RNA is mostly found not to be surrounded by cations (from 89.7% to 97.3% of the time), a slight enrichment of divalent and monovalent cations is seen going from helical to nonhelical regions (2.4% for divalent cations and 1.0% for monovalent).

Sampling of the $\eta-\theta$ space is nucleobase dependent

The large number of nucleotides in each cluster allows for a reasonable analysis on the sequence dependencies in the $\eta-\theta$ space. Notable differences are observed when the *complete-data set* $\eta-\theta$ map for North sugar rings is decomposed into nucleobase identity as evidenced in Figure 7. Despite fluctuations in their relative populations, clusters nearby the helical region (III, IV, and VI), and cluster II are clearly observed for all four nucleobases, as well as South clusters II and III. However, cluster I/North is depopulated of adenine (<12%), cluster V is almost exclusive of guanine (present at that position more than 50% of the time), and uracil is almost absent in clusters V and VII (with an occurrence below 9% and 7%, respectively). Considering that these later clusters are more populated in protein–RNA complexes (see previous sections), this result suggests that the nucleobase-dependent sampling of the $\eta-\theta$ space is, in part, a consequence of sequence-specific protein–RNA contacts.

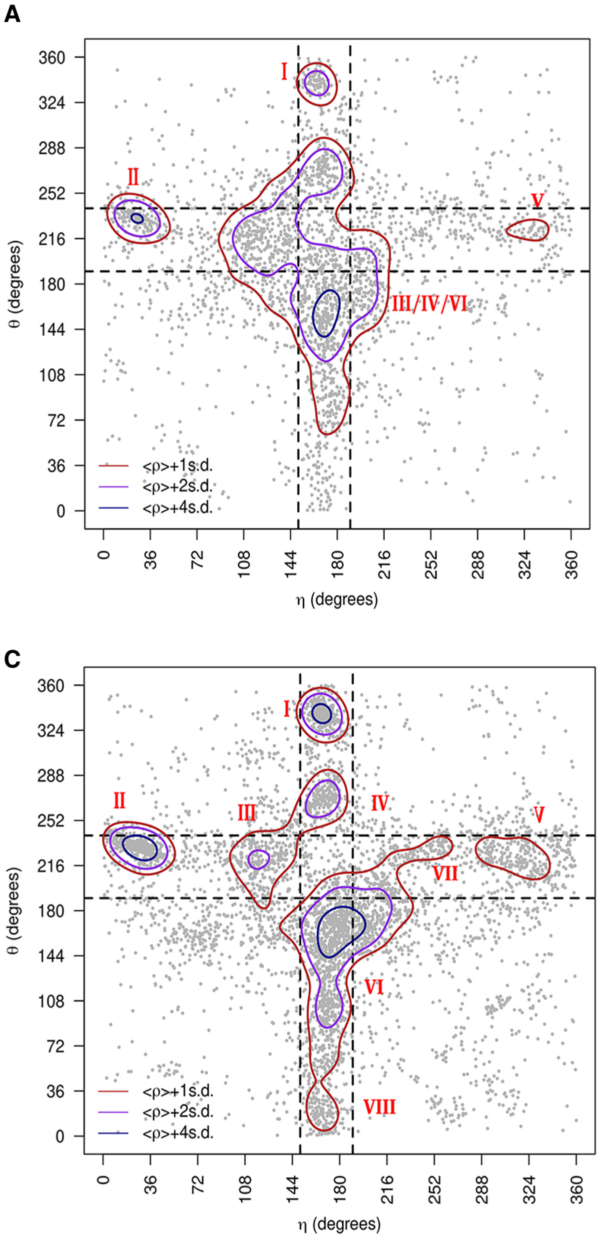

FIGURE 5. $\eta-\theta$ conformational space described by the naked-subset and protein-subset. Results are divided into nonhelical North (panels on the *left*) or South (panels on the *right*) sugar conformations. Density contours of $\bar{\rho}+X\cdot\sigma$ ($X=1$, 2, or 4; dark red, violet, and dark blue, respectively) highlight regions of the plots with a significant population of nucleotides. Black dashed lines delimit the region occupied by canonical helical A-form (those structures are removed from the analyses as described in the Materials and Methods section). Cluster numbers are identified in red or dark green when they were not seen in the *complete-data set* depicted in Figure 2. (*A*) Densities and cluster analysis of the naked-subset for central nucleotides with sugars in North. (*B*) Same as (*A*) for South conformations. (*C*,*D*) are the same as (*A*) and (*B*) for the protein-subset.

Furthermore, the presence of purines in the trinucleotide motif displayed as logo-plots is slightly dominant outside the helical region in North, GAA, AAG, GGA, and GAA being the most common sequences for clusters II, IV, IV, and VIII. In the case of South sugar rings, nucleobase dependence has been also observed to some extent, with cluster V mainly populated by uracil and adenine (52% and 27%, respectively), and adenine being mainly responsible for cluster IV (52% of the time).

A small set of torsional perturbations applied to a helical structure fully explain the RNA conformational diversity

All the North $\eta-\theta$ clusters fall into the space of helical η (~160°) or helical θ (~215°) (see the red lines in all our $\eta-\theta$ maps). Therefore, in most cases the backbone can move from one cluster to another just by rotating η or θ. For example, to go from the HDR-II ($\eta=30°$, $\theta=215°$) to the HDR-III

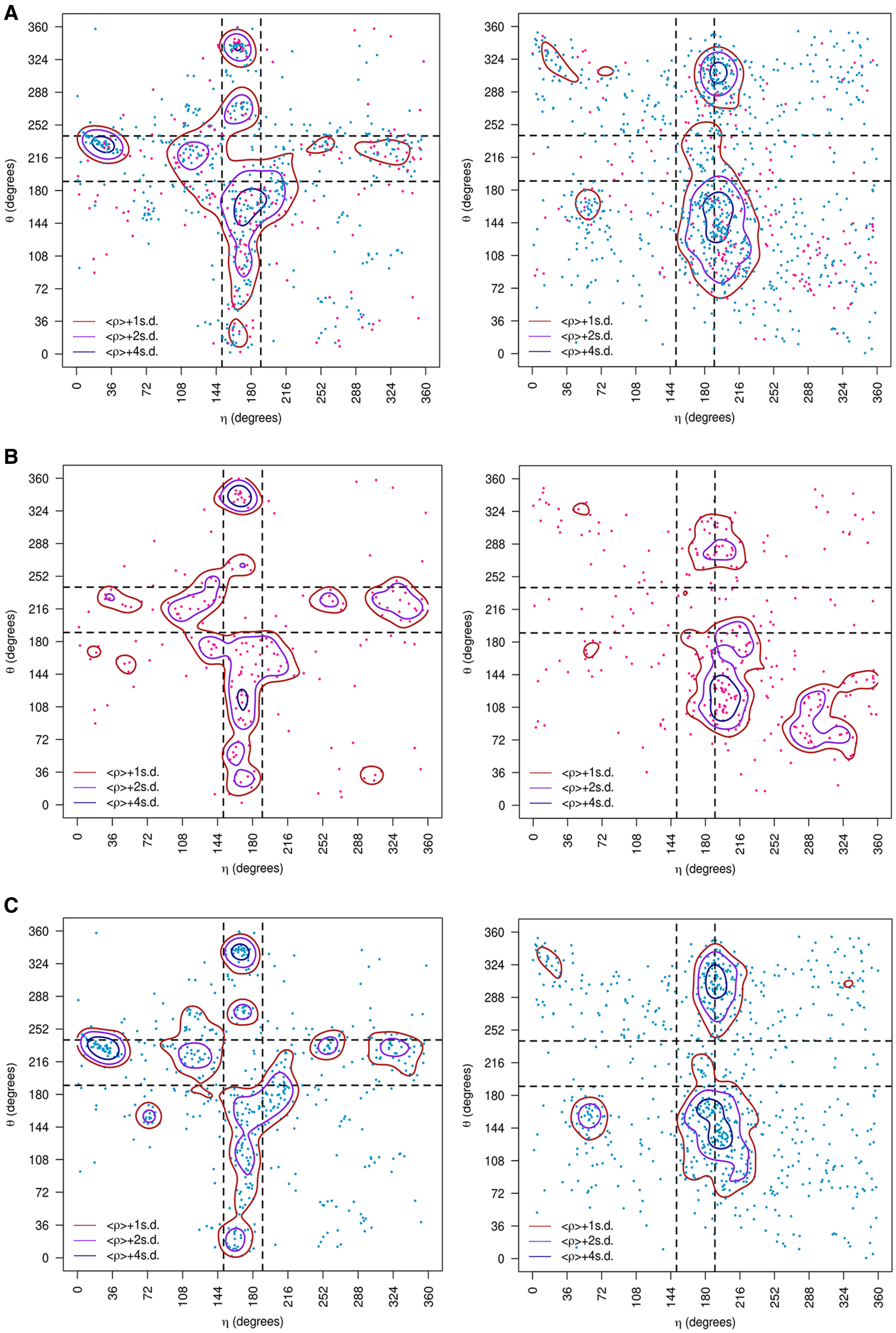

FIGURE 6. $\eta - \theta$ conformational space described by the monovalent-subset and divalent-subset. Results are divided into nonhelical North (panels on the *left*) or South (panels on the *right*) sugar conformations. Density contours of $\bar{\rho} + X \cdot \sigma$ ($X = 1$, 2, or 4; dark red, violet, and dark blue, respectively) highlight regions of the plots with a significant population of nucleotides. Black dashed lines delimit the region occupied by canonical helical A-form (those structures are removed from the analyses as described in the Materials and Methods section). (*A*) Densities and cluster analysis of the both types of cations together (monovalent in pink dots and divalent in turquoise) over the clusters obtained for the complete-data set. (*B*) Clusters obtained by using only the monovalent-subset. (*C*) Idem than (*B*) using the divalent-subset.

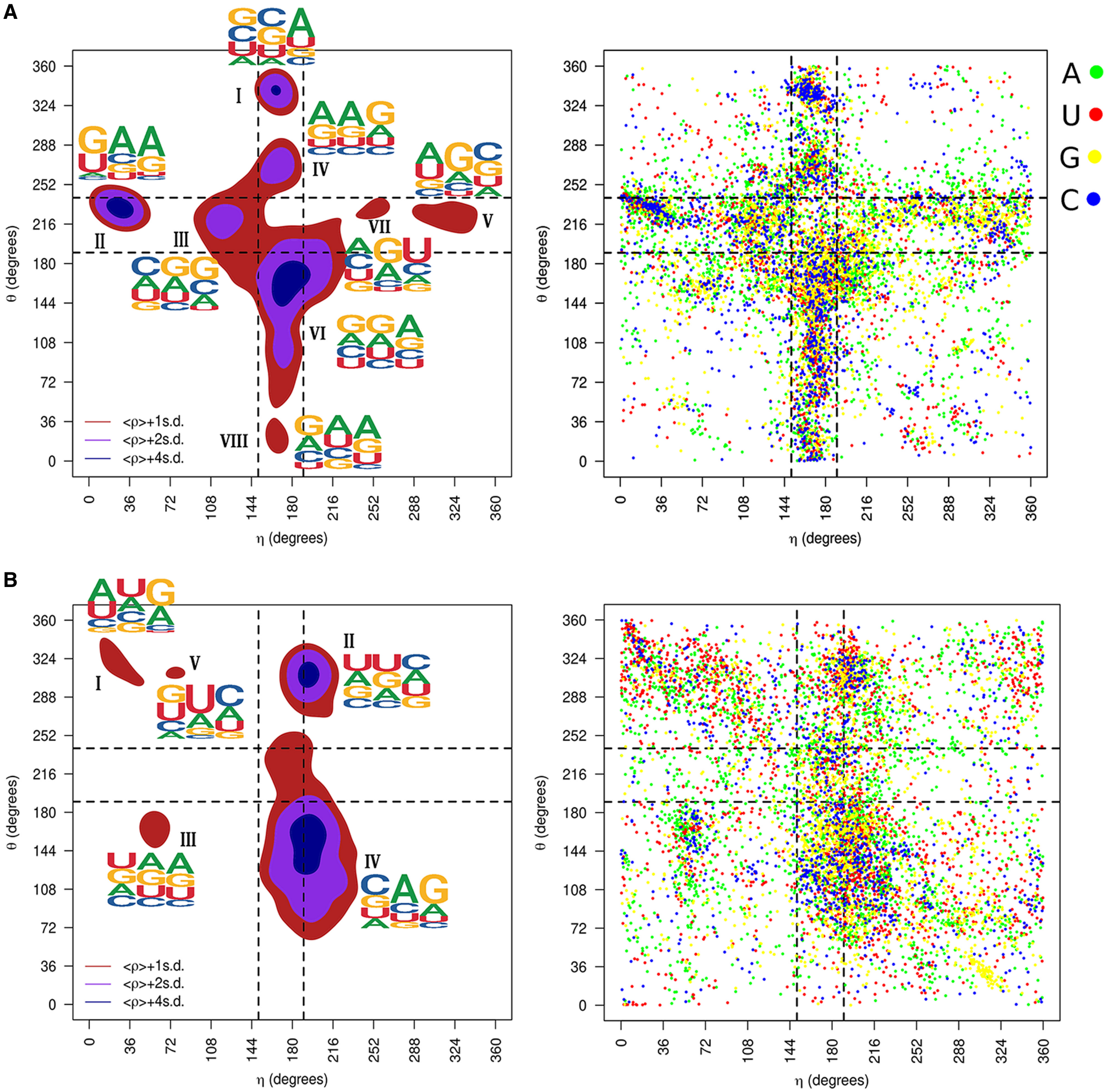

FIGURE 7. $\eta - \theta$ conformational space of the complete-data set. Results are divided into nonhelical North (*top*-panels) or South (*bottom*-panels) sugar conformations. Density contours of $\bar{\rho} + X \cdot \sigma$ ($X = 1$, 2, or 4; dark red, violet, and dark blue, respectively) highlight regions of the plots with a significant population of nucleotides. Black dashed lines delimit the region occupied by canonical helical A-form (those structures are removed from the analyses as described in the Materials and Methods section). Cluster numbers are identified in black in the *left* panels. (*A*) The panel on the *left* shows densities and sequence logo-plots of the complete-data set for sugars in North. The panel on the *right* shows the corresponding dot-plot with adenine (green), uracil (red), guanine (yellow), and cytosine (blue). (*B*) Same as (*A*) for South conformations.

($\eta = 110°$, $\theta = 215°$), the backbone only needs to move along the η axis. To acquire insight on the specific torsion changes coupled to η and θ, average values of α, β, γ, δ, ε, ζ, χ, and puckering were calculated for η and θ values in 5° windows. This was done for the central, and for the first 5′- and 3′-neighbors. Figure 3 for η and Supplemental Figure S4 for θ suggests that simple combinations of specific torsional perturbations encode the transitions between nHDRs. We have shown before that the torsions of η and θ can be traced by the movements of the real dihedral angles $\alpha, \beta, \gamma, \delta, \varepsilon, \zeta, \chi$, and puckering. Considering the η direction (Fig. 3), discrete perturbations in ε, ζ, and puckering in the 5′-neighbor, and in α and γ in the central nucleotide result in the sequential transitions: nHDR-II ↔ nHDR-III ↔ helical ↔ nHDR-VII ↔ nHDR-V.

Considering the θ direction (Supplemental Fig. S4), perturbations in α, γ, and ζ for the central nucleotide, and in α, γ, and puckering for the 3′-neighbor, explain the sequential transitions: nHDR-I ↔ nHDR-IV ↔ helical ↔ nHDR-VI ↔ nHDR-VIII. These results suggest that moving from the helical conformation to any of the nHDRs can be achieved by specific combinations of discrete torsional perturbations. This result implies that the complete North puckering RNA conformational space can be rationalized as a helical structure plus a small set of specific combinations of torsional perturbations.

Conclusions

We explored here the conformational space of the "experimental" RNA backbone using our package veriNA3d (Gallego et al. 2019). The large size of the database allowed dividing the data into different subsets, exploring relationships that were not addressed before, such as the effect of proteins on the RNA backbone, or intrinsic differences due to the source of experimental structures. In addition, alternatives to $\eta-\theta$ to study the backbone conformation have been used, such as the $\eta'-\theta'$ (which uses the C1′ instead of C4′) and the $\eta''-\theta''$ (which uses a standardized point in the base plane instead of the C4′). We explored these alternative maps but found out that they tend to form less clusters and are more heterogeneous, so we focused our analysis on the original $\eta-\theta$ space.

Quite surprisingly, the updated $\eta-\theta$ maps show overall a similar general distribution to that found in Pyle's 2007 study (Wadley et al. 2007). However, we detected the presence of three small new clusters not reported previously. In general, the X-ray and Cryo-EM subsets show almost equivalent $\eta-\theta$ maps, while the NMR-subset is much more limited, covering only the clusters close to the helical A-form. One plausible explanation is that X-ray and Cryo-EM data sets include very large structures with thousands of nucleotides (e.g., ribosomes) and proteins that expand the conformational space that can be sampled by RNA in nonhelical regions. Overall, it is clear that complementary sources of data are needed to fully recover the landscape of RNA backbone conformations.

Comparing samplings in *naked-subset* and *protein-subset*, we find evidence that protein contacts allow the RNA backbone to adopt the most extreme conformations. This agrees with previous findings in which proteins are shown to modulate the RNA conformation through contacts with the 2′OH (Darre et al. 2019). By their side, a significant portion of monovalent and divalent cations also appears, stabilizing backbone conformations far from the helical regions.

We also find that the transitions between clusters can be traced by the rotation of specific dihedral angles, and only a few changes are necessary to go from one cluster to another. Given that the $\eta-\theta$ maps are particularly interesting for coarse-grained applications in modeling RNA (Dans et al. 2019), our finding introduces a new way of moving in the $\eta-\theta$ space by using the real dihedral angles.

MATERIALS AND METHODS

Atomic-resolution structural data and RNA equivalence classes

The primary source of RNA-structural data is the PDB. This database is highly redundant and is also biased toward those structures that are determined more easily experimentally, which means that before starting with any desired analysis, some degree of filtering of structures deposited in the PDB is required to obtain a nonredundant set (Leontis and Zirbel 2012). The Leontis group classified types of redundancy and they generate weekly lists of "Equivalence Classes" of RNA on their website (http://rna.bgsu.edu/rna3dhub/nrlist). Each Equivalence Class contains all the RNA redundant chains and a representative structure. Equivalence Classes contain RNA chains rather than complete PDB IDs. A nonredundant set of structures is provided by the list of Equivalence Class Representatives. They offer different lists of X-ray structures according to a resolution threshold, and an additional list called "All" including all experimental conditions. For our purpose, the "All" list was used here, as defined in the release version 3.134.

Nonredundant data sets

From the Leontis list, different sets of nonredundant structures were obtained: *complete-data set* using all Equivalence Classes, *protein-subset* when the RNA is in contact with a protein, *monovalent/divalent-subsets* when in contact with cations (K^+/Na^+ and Ca^{2+}/Mg^{2+}/Mn^{2+}, respectively) and *naked-subset* which contains free RNA. In addition, the *complete-data set* was also divided into three additional subsets representing the determination techniques: *xray-subset*, *nmr-subset*, and *em-subset* for X-ray, NMR, and Cryo-EM techniques, respectively. For the X-ray and Cryo-EM structures, a resolution threshold of 2.4 Å was used to ensure a correct description of the sugar puckering (Jain et al. 2015). We considered only RNA chains with three or more nucleotides, the minimum number to measure the eta–theta parameters. Next, all the nucleotides were analyzed and divided into two sets according to their puckering state (North or South). North nucleotides (C3′-endo) were selected according to these criteria: $\delta = 84° \pm 30°$; phase between −18° and 54°; and Dp distance (base-phosphate perpendicular distance, defined in Jain et al. 2015) >2.9 Å. South nucleotides (C2′-endo) were defined with: $\delta = 147° \pm 30°$; phase between 126° and 198°; and Dp distance ≤2.9 Å. Nucleotides that lacked atoms or whose bond distances exceed the threshold of 2 Å were considered as "broken" and excluded. Nucleotides obtained using X-ray with the phosphate or the C4′/C1′ atom with B-factor over 60 were also excluded (Keating et al. 2011). The final nucleotides in the data set were restricted to the A, U, G, and C residues that have a complete backbone starting from the sugar of the "$i-1$" nucleotide (5′ neighbor) to the sugar of the "$i+1$" (3′ neighbor).

Definitions and calculations

Once the list of structures was compiled, we computed the geometrical analysis of the RNA backbone. $\eta-\theta$ and other dihedrals were defined as in Wadley et al. (2007). Puckering was measured as in Westhof and Sundaralingam (1985). Stacking was represented geometrically using the R vectors defined by Bottaro et al. (2014) and quantified using a QM approach (see below). In Bottaro's approach, the *z*-axis measures the vertical distance between two nucleobases having their glycosidic bond (angle ξ) in the canonical anticonformation (Bottaro et al. 2014). According to this reference frame, the central nucleobase of a canonical trinucleotide will have positive values of *z* toward the 5′ side and negative values toward the 3′ side. The calculation of the distance between points in the $\eta-\theta$ map and the scoring of $\eta-\theta$ clusters were carried out as described in Wadley et al. (2007). RMSD, εRMSD (Bottaro et al. 2014), torsions, and distances were computed in R using our package veriNA3d (Gallego et al. 2019). RMSD calculations included backbone atoms from $C4'_{i-1}$ to $C4'_{i+1}$ or $C1'_{i-1}$ to $C1'_{i+1}$ in the case of $\eta'-\theta'$. For the other nucleotide descriptors, such as base pairing and $\eta''-\theta''$, DSSR software was used (Lu et al. 2015). See Figure 1 for a structural representation of the pseudotorsional angles used to characterize RNA at the trinucleotide level.

Ramachandran-like plots

The $\eta-\theta$ plot of North nucleotides revealed a large, highly populated region near its center (Supplemental Fig. S1). This region contains nucleotides with a backbone conformation typical of an A-RNA helix, so we refer to it as the *helical* region. It has a density that is almost two orders of magnitude higher than any other region of the $\eta-\theta$ map. To study the surrounding regions with precision, we separated the set of North nucleotides into *helical* and *nonhelical* nucleotides like in Wadley's work (Wadley et al. 2007). Separate $\eta-\theta$ maps were so constructed for the *nonhelical* North nucleotides and all South nucleotides. The same approach was applied to $\eta'-\theta'$ and $\eta''-\theta''$ plots. For each plot, the density was measured using a kernel density estimation (KDE), with 361 × 361 grid points and a bandwidth = 40. These parameters were chosen in order to exactly match the original results obtained by Pyle and coworkers using their data set published in 2007 (Wadley et al. 2007). From the densities at each point of the $\eta-\theta$ map, the mean ($\langle\rho\rangle$) and standard deviation (σ or s.d.) were calculated and contour levels at $\langle\rho\rangle + X \bullet \sigma$ (using $X = 1$, 2, and 4) were added to define high-density regions (HDRs) or clusters. We defined as a cluster or HDR, trinucleotides grouped locally whose estimated density values were above $\langle\rho\rangle + 2$ s.d. (i.e., clusters I, II, III, IV, and VI in Fig. 2A). We also defined as an HDR or cluster, a given isolated group of trinucleotides having a density above $\langle\rho\rangle + 1$ s.d. (clusters V, VII, and VIII in Fig. 2A).

Stacking energy calculation

At the trinucleotide level, three combinations of stacking interactions were computed: the central base (i) with the nucleotide at its 5′-side ($i-1$), the central base with the nucleotide at its 3′-side ($i+1$), and ($i-1$)-to-($i+1$) direct interaction that does not involve the central base. To compute the stacking energy, the backbone of each trinucleotide was removed and replaced by a methyl group. Then the geometry of the individual bases was replaced by that optimized at the MP2/cc-pVTZ level of theory (Woon and Dunning 1993). The optimization was done on isolated nucleobases, defining an "ideal" geometry for each of the canonical nucleobases (A, C, G, and U). At the trinucleotide level, the coordinates of each of the three nucleobases were then replaced by their "ideal" version, based on the overlap of the ring atoms N1, C2, N3, C4, C5, C6 for pyrimidines, and N1, C2, N3, C4, C5, C6, N7, C8, N9 for purines. Quantum mechanical energies were calculated at the dispersion-corrected (Grimme et al. 2010, 2016) DFT level B3LYP-D3(BJ) with the double-polarized def2-TZVPP basis set (Weigend and Ahlrichs 2005), and interaction energies were corrected by the basis-set superposition error (BSSE) following the Boys and Bernadi counterpoise method (Gutowski et al. 1993).

SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

Received September 1, 2023; accepted September 1, 2023.

REFERENCES

Auffinger P, Grover N, Westhof E. 2011. Metal ion binding to RNA. *Met Ions Life Sci* **9:** 1–35. doi:10.1039/9781849732512-00001

Bottaro S, Di Palma F, Bussi G. 2014. The role of nucleobase interactions in RNA structure and dynamics. *Nucleic Acids Res* **42:** 13306–13314. doi:10.1093/nar/gku97

Butcher SE, Pyle AM. 2011. The molecular interactions that stabilize RNA tertiary structure: RNA motifs, patterns, and networks. *Acc Chem Res* **44:** 1302–1311. doi:10.1021/ar200098

Černý J, Božíková P, Svoboda J, Schneider B. 2020. A unified dinucleotide alphabet describing both RNA and DNA structures. *Nucleic Acids Res* **48:** 6367–6381. doi:10.1093/nar/gkaa383

Dans PD, Gallego D, Balaceanu A, Darré L, Gómez H, Orozco M. 2019. Modeling, simulations, and bioinformatics at the service of RNA structure. *Chem* **5:** 51–73. doi:10.1016/j.chempr.2018.09.015

Darre L, Ivani I, Dans PD, Gómez H, Hospital A, Orozco M. 2019. Small details matter: the 2′hydroxyl as a conformational switch in RNA. *J Am Chem Soc* **138:** 16355–16363. doi:10.1021/jacs.6b09471

Doerr A. 2017. Cryo-electron tomography. *Nat Methods* **14:** 34. doi:10.1038/nmeth.4115

Gallego D, Darré L, Dans PD, Orozco M. 2019. VeriNA3d: an R package for nucleic acids data mining. *Bioinformatics* **3:** 5334–5336. doi:10.1093/bioinformatics/btz553

Grimme S, Antony J, Ehrlich S, Krieg H. 2010. 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. doi:10.1063/1.3382344

Grimme S, Hansen A, Brandenburg JG, Bannwarth C. 2016. Dispersion-corrected mean-field electronic structure methods. *Chem Rev* **116:** 5105–5154. doi:10.1021/acs.chemrev.5b00533

Gruene T, Sheldrick GM. 2011. Geometric properties of nucleic acids with potential for autobuilding. *Acta Crystallogr A* **67:** 1–8. doi:10.1107/S0108767310039140

Gutowski M, Van Duijneveldt-Van De Rijdt JGCM, Van Lenthe JH, Van Duijneveldt FB. 1993. Accuracy of the Boys and Bernardi function counterpoise method. *J Chem Phys* **98:** 4728. doi:10.1063/1.465106

Hendrix D, Brenner S, Holbrook S. 2005. RNA structural motifs: building blocks of a modular biomolecule. *Q Rev Biophys* **38:** 221–243. doi:10.1017/S0033583506004215

Hershkovitz E, Tannenbaum E, Howerton SB, Sheth A, Tannenbaum A, Williams LD. 2003. Automated identification of RNA conformational motifs: theory and application to the HM LSU 23S rRNA. *Nucleic Acids Res* **31:** 6249–6257. doi:10.1093/nar/gkg835

Hombach S, Kretz M. 2016. Non-coding RNAs: classification, biology and functioning. *Adv Exp Med Biol* **937:** 3–17. doi:10.1007/978-3-319-42059-2_1

Hsiao C, Tannenbaum E, Vandeusen H, Hershkovitz E, Perng G, Tannenbaum AR, Dean WL. 2009. Complexes of nucleic acids with group I and II cations. In *Nucleic acid-metal ion interactions* (ed. Hud NV), pp. 1–38. The Royal Society of Chemistry, United Kingdom.

Jain S, Richardson DC, Richardson JS. 2015. Computational methods for RNA structure validation and improvement. *Methods Enzymol* **558:** 181–212. doi:10.1016/bs.mie.2015.01.007

Keating KS, Pyle AM. 2010. Semiautomated model building for RNA crystallography using a directed rotameric approach. *Proc Natl Acad Sci* **107:** 8177–8182. doi:10.103/pnas.0911888107

Keating KS, Humphris EL, Pyle AM. 2011. A new way to see RNA. *Q Rev Biophys* **44:** 433–466. doi:10.1017/S0033583511000059

Lavery R, Maddocks JH, Pasi M, Zakrzewska K. 2014. Analyzing ion distributions around DNA. *Nucleic Acids Res* **42:** 8138–8149. doi:10.1093/nar/gku504

Leontis NB, Westhof E. 2001. Geometric nomenclature and classification of RNA base pairs. *RNA* **7:** 499–512. doi:10.1017/s1355838201002515

Leontis NB, Zirbel CL. 2012. Nonredundant 3D structure datasets for RNA knowledge extraction and benchmarking. *RNA 3D Struct Anal Predict* **27:** 281–298. doi:10.1007/978-3-642-25740-7_13

Li S, Olson WK, Lu XJ. 2019. Web 3DNA 2.0 for the analysis, visualization, and modeling of 3D nucleic acid structures. *Nucleic Acids Res* **47:** 26–34. doi:10.1093/nar/gkz394

Lu X, Bussemaker HJ, Olson WK. 2015. DSSR: an integrated software tool for dissecting the spatial structure of RNA. *Nucleic Acids Res* **43:** 142–155. doi:10.1093/nar/gkv716

Misra VK, Draper DE. 1998. On the role of magnesium ions in RNA stability. *Biopolymers* **48:** 113–135. doi:10.1002/(SICI)1097

Murray LJW, Arendall WB, Richardson DC, Richardson JS. 2003. RNA backbone is rotameric. *Proc Natl Acad Sci* **100:** 13904–13909. doi:10.1073/pnas.1835769100

Nakane T, Kotecha A, Sente A, McMullan G, Masiulis S, Brown PMG, Grigoras IT, Malinauskaite L, Malinauskas T, Miehling J, et al. 2020. Single-particle cryo-EM at atomic resolution. *Nature* **587:** 152–156. doi:10.1038/s41586-020-2829-0

Natchiar SK, Myasnikov AG, Kratzat H, Hazemann I, Klaholz BP. 2017. Visualization of chemical modifications in the human 80S ribosome structure. *Nature* **551:** 472–477. doi:10.1038/nature24482

Nguyen HT, Hori N, Thirumalai D. 2019. Theory and simulations for RNA folding in mixtures of monovalent and divalent cations. *Biol Sci* **116:** 21022–21030. doi:10.1073/pnas.1911632116

Olson WK. 1980. Configurational statistics of polynucleotide chains. An updated virtual bond model to treat effects of base stacking. *Macromolecules* **13:** 721–728. doi:10.1021/ma60075a045

Olson WK, Bansal M, Burley SK, Dickerson RE, Gerstein M, Harvey SC, Heinemann U, Lu X, Neidle S, Shakked Z, et al. 2001. A standard reference frame for the description of nucleic acid base-pair geometry. *J Mol Biol* **313:** 229–237. doi:10.1006/jmbi.2001.4987

Saenger W. 1984. *Principles of nucleic acid structure*. Springer, New York.

Schneider B, Morávek Z, Berman HM. 2004. RNA conformational classes. *Nucleic Acids Res* **32:** 1666–1677. doi:10.1093/nar/gkh333

Stefan LR, Zhang R, Levitan AG, Hendrix DK, Brenner SE, Holbrook SR. 2006. MeRNA: a database of metal ion binding sites in RNA structures. *Nucleic Acids Res* **34:** 131–134. doi:10.1093/nar/gkj058

Vicens Q, Kieft JS. 2022. Thoughts on how to think (and talk) about RNA structure. *Proc Natl Acad Sci* **119:** e2112677119. doi:10.1073/pnas.2112677119

Wadley LM, Keating KS, Duarte CM, Pyle AM. 2007. Evaluating and learning from RNA pseudotorsional space: quantitative validation of a reduced representation for RNA structure. *J Mol Biol* **372:** 942–957. doi:10.1016/j.jmb.2007.06.058

Weigend F, Ahlrichs R. 2005. 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:** 3297–3305. doi:10.1039/B508541

Westhof E, Sundaralingam M. 1985. Additions and corrections: 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 polypentide backbone flexibility. *J Am Chem Soc* **107:** 970–976. doi:10.1021/ja00304a600

Woon DE, Dunning TH. 1993. Gaussian basis sets for use in correlated molecular calculations. III. The atoms aluminum through argon. *J Chem Phys* **98:** 1358. doi:10.1063/1.464303

RNA
A PUBLICATION OF THE RNA SOCIETY

The pseudotorsional space of RNA

Leandro Grille, Diego Gallego, Leonardo Darré, et al.

RNA 2023 29: 1896-1909 originally published online October 4, 2023
Access the most recent version at doi:10.1261/rna.079821.123