Article

Lone Pair . . . π Contacts and Structure Signatures of r(UNCG) Tetraloops, Z-Turns, and Z-Steps: A WebFR3D Survey

Craig L. Zirbel [1] and **Pascal Auffinger** [2,*]

[1] Department of Mathematics and Statistics, Bowling Green State University, Bowling Green, OH 43403, USA; zirbel@bgsu.edu

[2] Architecture et Réactivité de l'ARN, UPR 9002, Institut de Biologie Moléculaire et Cellulaire du CNRS, Université de Strasbourg, 67084 Strasbourg, France

[*] Correspondence: p.auffinger@ibmc-cnrs.unistra.fr; Tel.: +33-3-8841-7049; Fax: +33-3-8860-2218

Abstract: Z-DNA and Z-RNA have long appeared as oddities to nucleic acid scientists. However, their Z-step constituents are recurrently observed in all types of nucleic acid systems including ribosomes. Z-steps are NpN steps that are isostructural to Z-DNA CpG steps. Among their structural features, Z-steps are characterized by the presence of a lone pair . . . π contact that involves the stacking of the ribose O4′ atom of the first nucleotide with the 3′-face of the second nucleotide. Recently, it has been documented that the CpG step of the ubiquitous r(UNCG) tetraloops is a Z-step. Accordingly, such r(UNCG) conformations were called Z-turns. It has also been recognized that an r(GAAA) tetraloop in appropriate conditions can shapeshift to an unusual Z-turn conformation embedding an ApA Z-step. In this report, we explore the multiplicity of RNA motifs based on Z-steps by using the WebFR3D tool to which we added functionalities to be able to retrieve motifs containing lone pair . . . π contacts. Many examples that underscore the diversity and universality of these motifs are provided as well as tutorial guidance on using WebFR3D. In addition, this study provides an extensive survey of crystallographic, cryo-EM, NMR, and molecular dynamics studies on r(UNCG) tetraloops with a critical view on how to conduct database searches and exploit their results.

Keywords: nucleic acid; X-ray; cryo-EM; NMR; molecular dynamics simulations; structure mining; isosteric; isostructural

1. Introduction

Defining and naming structural motifs is key to better comprehending and manipulating nucleic acid systems and is a root of structural ontology [1,2]. A significant part of nucleic acid motifs has already been categorized with various levels of precision. Efficient tools to search these motifs in PDB structures are currently accessible [3–10]. However, some motifs have not yet been defined with the finest possible levels of detail. Henceforth, the structural features of a few of the best-known and most essential of these motifs can still surprise us. We illustrate this by focusing on r(UNCG) tetraloops that are ubiquitous in RNA systems [11].

A Background section summarizes the main characteristics of these tetraloops and how their structural signature evolved and was recently augmented by the discovery of an embedded Z-step that is isostructural to CpG steps in Z-DNA [12,13]. In the same section, we discuss the importance of lone pair . . . π or lp . . . π contacts in the context of Z-steps and r(UNCG) tetraloops and address their structural and energetic features. In nucleic acids, lp . . . π contacts are defined as involving the stacking of an oxygen atom with a nucleobase [12,14–17]. To conclude the Background section, we survey molecular dynamics (MD) simulations related to r(UNCG) and Z-DNA systems that contain Z-steps and discuss current challenges that are related to obtaining stable dynamical models.

In the Results section, we describe a novel WebFR3D [5] implementation of symbolic constraints allowing one to search lp . . . π contacts in RNA. To the best of our knowledge,

WebFR3D is currently the only publicly available structural search tool that embeds such functionality. We illustrate how these new WebFR3D constraints can be used to search PDB structures for Z-steps, for exploring Z-turns, and for defining effective structural signatures for r(UNCG) tetraloops and their variants. Obtaining such signatures is of great interest for providing sound structural references that can be used to better tune and validate MD simulations. We conclude by discussing future evolutions of the service that will incorporate the ability to search for Z-steps in DNA and to search for motifs that include modified nucleotides.

2. Background: r(UNCG) Tetraloops and Lone Pair . . . π Contacts

2.1. The Structural Signature of r(UNCG) Tetraloops Has Evolved

The first r(UUCG) tetraloop NMR structure was characterized some 30 years ago [18,19]. This 1990 structure established that the $r(U_5UCG_8)$ loop embeds a reverse-wobble $U_5 \bullet G_8$ pair with the guanine adopting a *syn* conformation, with a cytosine . . . phosphate contact and with extensive base stacking (Figure 1). The structural model also involved a *C2′-endo* conformation for U6 and C7 associated with a specific backbone dihedral angle sequence [18–20]. As a further r(UNCG) characteristic, the Watson-Crick edge of the U_6 nucleobase is usually fully exposed to the solvent.

The $r(UUU_7G)$ NMR structure looks like that of the parent $r(UUC_7G)$ tetraloop. Along with r(GNRA) tetraloops, r(UNCG) loops are among the most stable RNA tetraloops while $r(UUU_7G)$ loops were found to be less stable [18,21]. The C7/U7 replacement, which leads to an apparent loss of the cytosine . . . phosphate contact (Figure 1d), was considered to be at the origin of the lower thermodynamic stability of the $r(UUU_7G)$ loop.
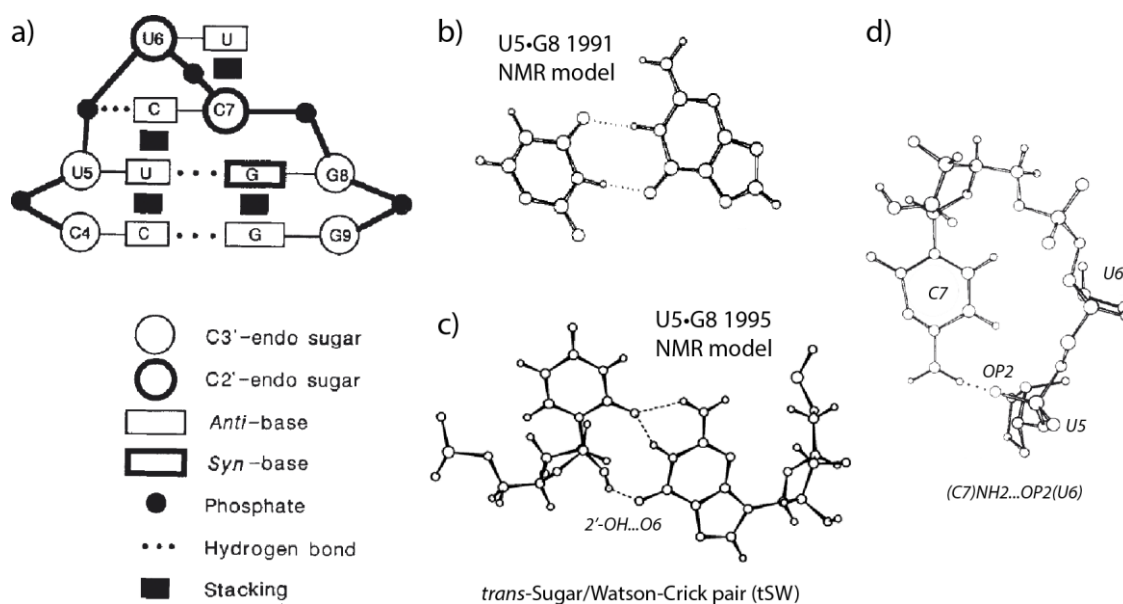

Figure 1. Early and revised $r(U_5UCG_8)$ NMR model structures. (**a**) First NMR structural signature (adapted with permission from [18]). (**b**) First model of the U•G pair in an r(UUCG) loop (adapted with permission from [19]). (**c**) The preceding model was revised in 1995 to lead to the correct trans-Sugar/Watson-Crick (tSW) base pair arrangement (adapted with permission from [22]). Note that a C•G tSW pair would be perfectly isosteric to the U•G tSW pair. (**d**) A C7 . . . U5 base . . . phosphate interaction, annotated as 8BPh, is part of the r(UUCG) tetraloop signature (adapted with permission from [18]).

Some parts of the original r(UUCG) structure were revised in 1995 through improved NMR protocols [22]. The new model mainly differs in the hydrogen bonding pattern of the $U_5 \bullet G_8$ pair that was found to involve a 2′-OH group (Figure 1). The authors noted that the resonance rate of exchange of the 2′-OH hydrogen is like that of NH groups involved in

Watson-Crick pairs. Thus, this 2′-OH forms a stable hydrogen bond. This contrasts with the fact that a majority of the 2′-OH resonances are unobservable given that the hydroxyl proton is usually in fast exchange with those of the solvent.

In a subsequent 2010 "high-resolution" NMR structure (PDBid: 2KOC), the tetraloop U•G pair was annotated as *trans*-wobble [23]. Additionally, based on NOESY water-RNA cross peaks, it was suggested that a water molecule was firmly associated with this base pair. Later, this base pair was annotated as *trans*-Sugar/Watson-Crick (tSW) based on the Leontis and Westhof base pair nomenclature [5,24,25].

The first X-ray structure of a r(UUCG) tetraloop was obtained in 2000 (PDBid: 1F7Y; resolution: 2.8 Å; [24]). The authors described failures related to early attempts to crystallize standalone r(UUCG) hairpins and identified a successful strategy that consisted in embedding the tetraloop in a larger RNA system, namely a 16S rRNA fragment complexed with a *Thermus thermophilus* S15 r-protein. The two r(UUCG) tetraloops embedded in the system were similar to those of the 1995 revised NMR structure [22,24].

More recently, NMR studies reported r(UNCG) conformational heterogeneity [25,26]. A study using exact nuclear Overhauser enhancement data (eNOEs) led to the characterization of a two-state r(UUCG) structural ensemble [26]. This two-state dynamic was explored through the integration of molecular dynamics (MD) simulations with eNOEs [27]. The dominant state (state A; 90%) corresponds to the consensus $r(U_6UCG_9)$ structure, while the low-populated state B (~10%) is characterized by the absence of the $U_6•G_9$ tSW pair with C_8 and G_9 partially exposed to the solvent. However, the possibility that conformations differing from states A and B can be present was suggested (Figure 2). The advantage of combining NMR data with structure prediction and MD simulations is to be able to circumvent current MD simulation shortcomings and imprecisions to construct more reliable dynamic conformational ensembles [28,29]. MD simulation shortcomings are addressed in Section 2.3.
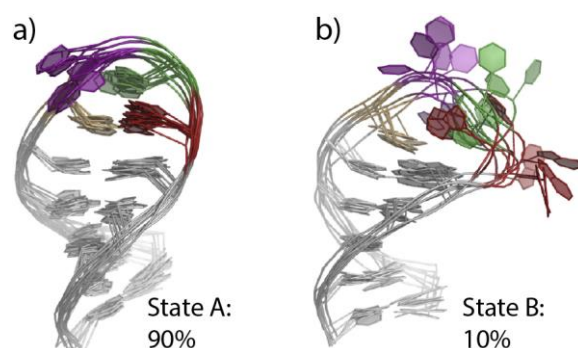

Figure 2. Two NMR ensembles of an r(UUCG) tetraloop derived from eNOE experiments. (**a**) Representative (random) conformations sampled from state A corresponding to a canonical fold. (**b**) Representative (random) conformations sampled from state B corresponding to less structured conformations. Note the fraying out of the guanine of the U•G pair. The PDBids of these NMR ensembles are 6BY4 and 6BY5. Figure adapted from [27].

2.2. r(UNCG) Tetraloops, Although Called Z-Turns, Comprise a CpG Z-Step with a Lone Pair . . . π Contact

Up to recently, it has escaped the attention of most structural biologists except Richardson et al. [20], that the CpG step of r(UNCG) tetraloops adopt a conformation that is also found in CpG containing double-helical Z-DNA structures. Therefore, it is important to describe some of the main structural characteristics of these CpG steps.

Because they are characteristic of the Z-DNA zig-zag left-handed structure, Z-DNA CpG steps were named Z-steps [12,13]. By extension, any DNA or RNA NpN dinucleotide that adopts a conformation close to that of the Z-DNA CpG step is called a Z-step. The fact that almost all dinucleotide NpN sequences, when placed in the proper environment, can form Z-steps stresses the structural importance of this recurring nucleic acid motif [12,30].

Further, UNCG-like tetraloops and larger loops containing Z-steps are called Z-turns [13,31]. Rare examples of Z-turns with an r(GAAA) sequence featuring an ApA Z-step were identified in the core of ribosomes and in a few other RNAs [13,31]. They are discussed in Section 3.5.

The Z-DNA CpG Z-step conformation (Figure 3) involves a 3′-nucleotide in a *syn* conformation, a 5′-nucleotide (deoxy)ribose with a C2′-*endo* pucker, and an lp . . . π contact with an oxygen (O4′) to nucleobase (G) contact distance below 3.5 Å and sometimes close to 2.8 Å. Z-steps also embed a characteristic ribose head-to-head (*The head-to-tail versus head-to-head terms sometimes lead to confusion. Here we chose to name head-to-tail the ribose orientation as it occurs in a regular helical structure by analogy with elephants walking in a line and holding the tail ahead with their trunk; we named head-to-head the ribose orientation shown in Figure 3b that occurs in Z DNA and r(UNCG) CpG steps. Note that we wrongly used "head-to-tail" in [13]*) orientation. The 3.5 Å boundary was set by examining the oxygen to nucleobase plane distance histograms derived from various NpG steps in nucleic acids [12,13,17] although other authors preferred 4.0 Å boundaries [15].
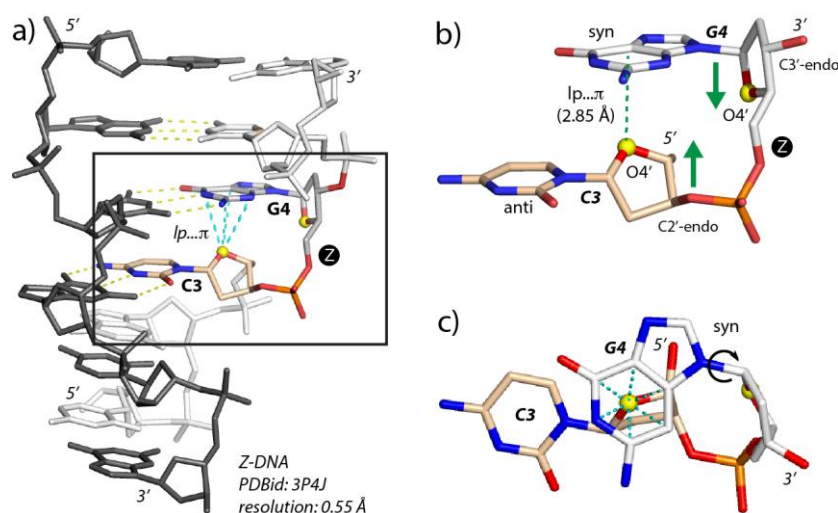

Figure 3. A representative Z-DNA CpG step and associated lp . . . π contact. (**a**) View of a 0.55 Å X-ray structure of a *d*(CpGpCpGpCpG)₂ duplex [32]. A representative CpG segment is boxed and colored as in (**b**); the remaining nucleotides of this strand are in white. The second strand is in dark grey. The lp . . . π contacts and the base pair hydrogen bonds associated with the CpG segment are marked by cyan and yellow dashed lines, respectively. This structure contains six lp . . . π contacts, one for each CpG step. (**b**) Close-up of the CpG motif boxed in (**a**). The O4′ atoms are yellow; the cytosine (C) and guanine (G) carbons are wheat and white, respectively. The O4′ atom of C3 stacks with the 3′ face of G4. Note the characteristic Z-DNA alternation of *anti* and *syn* conformations [12]. The head-to-head orientation of the two sugar rings, common for Z-steps, is marked by green arrows. (**c**) Orthogonal view of b). In (**a**,**b**), the "Z" letter (circled with a black background) marks the occurrence of a Z-step.

The *syn* conformation of the CpG Z-step guanine implies a χ torsion angle value around the glycosidic bond of about 60° instead of 120° for the usual *anti* conformation [33,34]. One of the rarely noted peculiarities of a Z-step is that the Watson-Crick sites of these base pairs are aligned, resulting in a modest average 2° rotation for a CpG step in Z-DNA compared to a 60° rotation for a Z-DNA GpC step and a ≈ 35° rotation for any step in B-DNA.

The energetic contribution of the lp . . . π contact is not precisely known, although it is appreciated that the interaction is of a weak non-covalent type [14,17]. We established earlier that the origins of the short ≈2.8 Å contacts observed in X-ray structures cannot be explained by orbital effects as implied by the "lp . . . π interaction" terminology. Thus, we proposed that these "lp . . . π interactions" could be named "oxygen . . . π contacts" to avoid

interpretation issues [35–37]. Given the weak non-covalent character of these interactions, it has been hypothesized that short lp . . . π contacts occur primarily in structurally strained motifs such as those found in Z-DNA and r(UNCG) tetraloops [17,37].

The dihedral angle variations of a CpG Z-step seem relatively limited when considering Z-DNA structures. The left-handed double-helical Z-DNA is constructed from alternating pyrimidine–purine (YpR) and purine–pyrimidine (RpY) steps. The YpR steps are characterized by a single backbone conformation whereas the RpY steps may adopt two distinct conformations known as Z_I and Z_{II} [38–41]. In the Z_I form the phosphate groups are shifted deeper inside the helix towards the groove and in the Z_{II} form the phosphate groups are rotated away from the groove (Figure 4). Sometimes, alternate conformations of phosphate groups are observed in high-resolution X-ray structures [34]. The consensus r(UNCG) dihedral backbone sequence is N**1a**U**1z**N2[C**6n**G**1a**N following the nomenclature established by Richardson et al. [20].
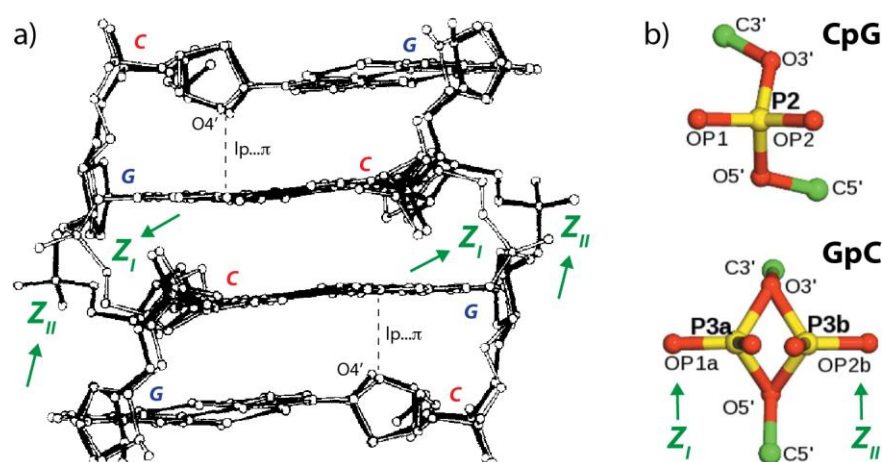

Figure 4. Z_I and Z_{II} conformations. (**a**) Superimposition of two Z-DNA *d*(CGCG)$_2$ structures (PDBid: 1ZNA; resolution 1.5 Å) showing Z_I (light bonds) and Z_{II} (dark bonds) conformations (adapted with permission from [38]). Short lp . . . π contacts are marked by a dotted line. (**b**) Two characteristic conformations of a CpG (top) and a GpC (bottom) step extracted from a d(CGCGCGCGCGCG)2 dodecamer (PDBid: 4OCB; resolution: 0.75 Å). The CpG step displays mostly a single conformation while four of the five GpC steps of the dodecamer were modeled as alternate Z_I and Z_{II} conformations thanks to the use of phosphorous anomalous signals (adapted from [34]).

Two other local Z-DNA conformations were reported. These are the Z and Z′ forms that adopt a C3′-*endo* and an infrequent C4′-*exo* guanine sugar pucker, respectively [38,42]. Overall, the base pair orientation and stacking configuration seem not affected by changes in the guanine sugar pucker (Z, Z′) or the GpC backbone conformations (Z_I, Z_{II}).

2.3. r(UNCG) and Z-DNA Molecular Dynamics Simulation Challenges

Despite the apparent simplicity of these motifs, MD simulations of r(UNCG) tetraloops and Z-DNA fragments present many unresolved challenges that are related to an incomplete understanding of the forces at play in these systems [43–53]. Therefore, we provide next a survey of current MD simulation issues.

A recent report summarized some of these challenges for RNA tetraloops that could not all be resolved by recent parameterization efforts [54–59]. The characteristic r(UUCG) native structure is lost in ns to µs standard MD simulations or does not fold correctly because of at least two different effects. The first of those is excessive stabilization of unfolded single-stranded RNA structures by intramolecular base . . . phosphate and sugar . . . phosphate interactions. The second relates to the destabilization of the native folded state by underestimation of the native hydrogen bonds including the stem base pairing. The drift from the native structure was described as a progressive and undesired loss of key signature interactions.

Simulations using the newly developed DESRES force field documented several r(UUCG) unfolding-refolding events on a 20 µs time scale [55] but these were found to be sensitive to tiny changes in parameterization. For instance, changing the monovalent ion model suffices to completely lose the tetraloop native structure [54,60]. Other issues related to the use of the DESRES force field are described in references [54,55,57,60]. The authors of these studies conclude that they are not currently aware of any existing RNA force field that would accurately represent these tetraloop systems on µs time scales.

Thus, if the NMR data suggest the existence of a dominant and several minor states for the tetraloop, MD simulations must be able to sample these various populations and their variations related to changes in temperature and environmental conditions [61]. These issues will certainly push MD force fields and simulation protocols into further challenges and necessitate a finer understanding of the physico-chemistry underlying these phenomena.

Some attention was recently drawn to the parameterization of the lp ... π contact constitutive of Z-steps, an issue that has never been fully addressed [37]. To this point, it is unknown to what extent a potentially imbalanced description of lp ... π contacts causes MD simulations of Z-step-containing systems to ill-behave. However, two factors are worthy of consideration. The first is related to a misrepresentation of the Lennard-Jones or vdW parameters of the nucleobase atoms. It has been established that the vdW parameters of sp^2 carbon atoms attached to electron-withdrawing groups are largely overestimated by the Bondi tabulation established in 1964, a tabulation still in use [17,37]. A smaller effective sp^2 carbon vdW radius could explain the short lp ... π contacts observed in Z-DNA and r(UNCG) loops. It can be noted that the current AMBER force-field versions use Lenard-Jones parameters for the nucleobase carbon atoms that are identical to those of phenyl ring carbons as noted in [37]. The second effect of importance for MD simulations is an overestimation of the repulsive part of the Lennard-Jones potential of current force fields when compared to potentials derived from high-level quantum-mechanics calculations. Although adapting the Lennard-Jones parameters is a difficult task needing significant reparameterization efforts, progress in this direction would certainly help to improve the quality of MD simulations of systems containing lp ... π contacts and of biopolymers in general [37].

Besides r(UNCG) tetraloops, Z-DNA structures also represent a significant challenge for MD simulations [53,62–65]. This can be linked to an inappropriate representation of Z_I/Z_{II} equilibrium states and has been recently addressed through backbone dihedral angle reparameterizations. Interestingly, these reports also mention a sensitivity of the explored substates to the monovalent ion parameters and water models suggesting that the entire molecular ecosystem must be modeled with the greatest possible accuracy to achieve a precise balance of interatomic forces. Given the presence of Z-steps in both Z-DNA and r(UNCG) tetraloops, progress in the simulation of these systems is definitely linked.

3. Results

3.1. Defining lp ... π Contacts

As mentioned in [51], the native state definition of a structural model can have a dramatic impact on reported populations of folded states derived from MD simulations, NMR studies, or X-ray surveys. Therefore, the following aims at defining an operational structural signature for characterizing r(UNCG) tetraloops, associated Z-steps, and more generally Z-turns. For that, we start this section by defining lp ... π contacts that are common to all these motifs.

Since a significant portion of dinucleotide fragments involving lp ... π contacts are not of the Z-step type, it is important to first define an lp ... π contact signature. The characteristic of these contacts is that the involved oxygen atom stacks on a nucleobase face with a contact distance to the nucleobase plane ≤ 3.5 Å [12,13,17]. We refer to the two faces as the 3' and the 5'-face; for a definition of 3'-and 5'-nucleobase faces, see the Method Section.

The lp . . . π distance has also to be >2.0 Å to exclude rare "in-plane" contacts. Issues related to alternate conformations in database searches are described in [66]. Such conformations (with occupancies < 1.0) will be ignored in the following. Note that for generic lp . . . π contacts, the two nucleotides do not need to be consecutive or to belong to the same strand and the nucleobases do not need to be stacked.

3.2. Finding lp . . . π Contacts with WebFR3D

The WebFR3D server can search for lp . . . π interactions using text strings such as s3O4′ to indicate that the 3′-face of one nucleotide stacks with the O4′ atom of a second nucleotide. The first two characters, s3 or s5, mark a stacking with the nucleobase 3′ or 5′-face. The second part of the text string marks which backbone oxygen atom is involved in the stacking, for instance: OP1, OP2, O2′, O3′, O4′, or O5′ (nucleobase O2, O4, O6, and non-hydrogenated nitrogens are not considered). Note that the search string is directional. One may write that the 3′-face of nucleotide 1 (nt1) forms an s3O4′ interaction with the O4′ atom of nt2, or that nt2 forms a sO4′3 interaction with the 3′-face of nt1.

When searching for nucleotides forming lp . . . π contacts, various abbreviations can be used. For instance, "O" is generic for OP1, OP2, O2′, O3′, O4′, and O5′ backbone oxygen atoms. Thus, sO3 indicates stacking involving a backbone oxygen of nt1 and the 3′-face of nt2 while sO indicates stacking of a backbone oxygen atom on either the 3′ or 5′-face of nt2. Figure S1 (Supplementary Materials) shows the results of an sO search that identifies 246 lp . . . π contacts in the 7K00 ribosomal structure, 143 of which involve an O4′ atom (sO4′) and 62 involve an OP atom (sOP or sOP1/sOP2).

3.3. Z-Steps and Z_{anti}-Steps Identified by WebFR3D

An ideal Z-step signature, as observed in Z-DNA and r(UNCG) tetraloops [12], involves two consecutive nucleotides where the 5′-nucleobase is in *anti*; the 5′-sugar is in *C2′-endo*; the 3′-nt is in *syn*, and the O4′ oxygen of the 5′-ribose stacks with the 3′-face of the 3′-nt to form an lp . . . π contact. Finally, in Z-DNA and r(UNCG) tetraloops, a ribose head-to-head orientation occurs (Figure 3b). As noted above, the consensus r(UNCG) dihedral backbone sequence for the CpG step is N**1a**U**1z**N2[C**6n**G**1a**N following the nomenclature established by Richardson et al. [20]. We note also that the head-to-head orientation of the two ribose is a consequence of the formation of an lp . . . π contact.

In the 7K00 ribosome structure at 1.98 Å resolution, eleven instances of Z-step motifs are identified that comprise six different sequences (Figure S2). In a larger search, most of the 16 possible r(NpN) sequences can be identified by WebFR3D when searching current representative sets of structures at different resolution thresholds. For instance: ApA, ApG, GpA, GpG, CpA, CpG, UpA, and UpG Z-steps are identified in structures with resolution ≤ 2.0 Å; ApC, ApU, CpU, UpC, and UpU Z-steps are identified in structures with resolutions between 2.0 Å and 2.5 Å, and GpC, GpU Z-steps are identified in structures with resolutions between 2.5 Å and 3.0 Å. This confirms earlier conjectures stating that Z-steps can involve any of the 16 NpN sequences except for CpC Z-steps in RNA structures [12,13]. Overall, we found that NpR are more frequent than NpY steps and without surprise that CpG is the most frequent Z-step in RNA as it is in DNA.

Z-steps with a 3′-nucleobase in *anti* that we called Z_{anti}-steps [13]. They correspond to a Z-step subcategory that, along with many other variants, is not discussed here. All these variants can be identified by WebFR3D. Note that the *syn/anti* constraint is sometimes not very effective for the 3′-nt in a Z-step since the χ angles are in borderline regions. Therefore, it is suggested to use as an alternative an nt1 . . . nt2 s53 stacking constraint when the *syn* constraint leads to questionable results. More precisely, s53 implies that the second base is turned and adopts a *syn* conformation. For searching Z_{anti}-steps, where the second base is in *anti*, an s55 constraint can be used.

3.4. r(UNNG) Z-Turn Signatures Derived from X-ray and Cryo-EM Structures

To explore the X-ray and cryo-EM structures of $r(U_1NCG_4)$ tetraloops, we used the WebFR3D server to extract a set of instances of r(UNNG) motifs from structures with resolutions $\leq$ 2.0 Å [5]. The search involves an $r(nU_2NNG_5n)$ sequence with a tSW $U_2 \bullet G_5$ base pair (Figure 5). The "3.230" representative set of structures (March 2022) with a resolution $\leq$ 2.0 Å was used [67]. The search led to the 14 hits described below. This structural set comprises eight r(UNCG) X-ray structures, the six other tetraloop structures originating from the 7K00 *Escherichia coli* ribosome cryo-EM structure at 1.98 Å resolution [68]. This ensemble of 14 structures consists of ten r(UUCG), two r(UACG), and two r(UCCG) motifs. With resolutions < 3.0 Å, we found only a single occurrence of an r(UGCG) loop in a *Trypanosoma cruzi* cryo-EM structure (PDBid: 5T5H; res.: 2.54 Å) with a reasonably good electron density map [69] suggesting that this sequence seldom folds as a Z-turn. The third position of r(UNNG) Z-turns can also be a U as discussed in Section 2.1 (Figure 1d) and observed in the 6CK5 riboswitch structure at 2.49 Å resolution [70] or an A as in the 2.7 Å resolution 6YWS ribosomal structure [71].

A closer examination of the eight high-resolution X-ray structures ($\leq$2.0 Å) reveals that the electron densities for the 6DCB and to a lesser extent the 5OB3 tetraloops are of poor quality as also implied by some backbone irregularities, the absence of modeled water molecules, and by the fact that the bottom cytosine residue of the 5OB3 adopts an irregular *syn* conformation. Therefore, these two structures were discarded from the ensemble and the focus was placed on the six X-ray structures 5Y85, 3U4M, 7KKV, 4ARC, 7EOG, and 7P0V with modeled water molecules. The 3U4M structure presents the best defined experimental electron densities, as visualized with Coot [72]. There, as well as in 7KKV and 7P0V, a water molecule in the U$\bullet$G cleft of the tSW pair could be accurately modeled (Figure 6a). This water molecule is certainly at the origin of the NOESY spectra cross peak signal observed between water and the first uridine of the loop as detailed in Section 2.1. Although the 7EOG model has the best resolution (1.50 Å) of the set, the electron density maps are less well defined, and no water molecule was modeled in the density that is visible in the U$\bullet$G cleft of the tSW pair. Thus, we discarded 7EOG from our validated structural ensemble.

Figure 5. *Cont.*

FR3D output

Query name: r(UNNG) tetraloop motif query (resolution better or equal to 2.0 Angs.). Found 14 candidates from 461 of 461 files in 3 seconds. See and modify query | See CSV output
Columns of the table show candidate number in similarity order, checkbox to display coordinates or not, structure resolution, discrepancy from query in geometric or mixed searches, units matching each position in the query, sequence of the units and backbone connectivity, glycosidic
bond conformation if requested, pair and stack interactions present, base-phosphate interactions, base-ribose interactions, oxygen stacking interactions, and the number of nested AU, GC, GU Watson-Crick pairs crossed by each annotated interaction.

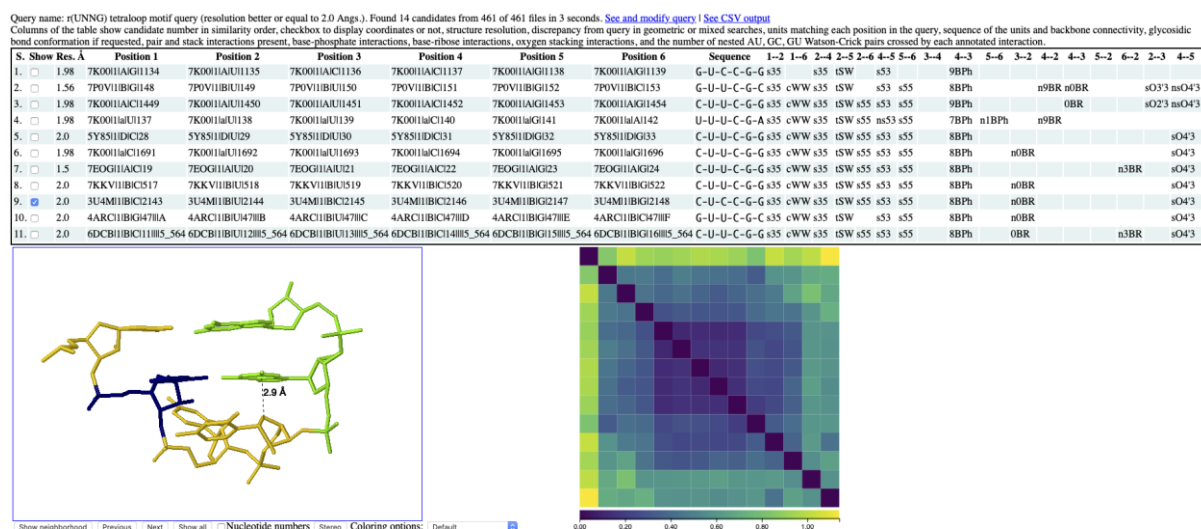| S.	Show	Res. Å	Position 1	Position 2	Position 3	Position 4	Position 5	Position 6	Sequence	1→2	1→6	2→4	2→5	2→6	4→5	5→6	3→4	4→3	5→6	3→2	4→2	4→3	5→2	6→2	2→3	4→5
1.	☐	1.98	7K00|1|A|G|1134	7K00|1|A|U|1135	7K00|1|A|C|1136	7K00|1|A|C|1137	7K00|1|A|G|1138	7K00|1|A|G|1139	G-U-C-C-G-G	s35		s35	tSW		s53		9BPh									
2.	☐	1.56	7P0V|1|B|G|148	7P0V|1|B|U|149	7P0V|1|B|U|150	7P0V|1|B|C|151	7P0V|1|B|G|152	7P0V|1|B|C|153	G-U-U-C-G-C	s35	cWW	s35	tSW		s53	s55	8BPh			n9BR	n0BR				sO3'3	nsO4'3
3.	☐	1.98	7K00|1|A|C|1449	7K00|1|A|U|1450	7K00|1|A|U|1451	7K00|1|A|C|1452	7K00|1|A|G|1453	7K00|1|A|G|1454	C-U-U-C-G-G	s35	cWW	s35	tSW	s55	s53	s55	9BPh				0BR				sO2'3	nsO4'3
4.	☐	1.98	7K00|1|a|U|1137	7K00|1|a|U|1138	7K00|1|a|U|1139	7K00|1|a|C|1140	7K00|1|a|G|141	7K00|1|a|A|142	U-U-U-C-G-A	cWW	s35	tSW	ns53	s55			7BPh	n1BPh		n9BR						
5.	☐	2.0	5Y8S|1|D|C|28	5Y8S|1|D|U|29	5Y8S|1|D|U|30	5Y8S|1|D|C|31	5Y8S|1|D|G|32	5Y8S|1|D|G|33	C-U-U-C-G-G	s35	cWW	s35	tSW	s55	s53	s55	8BPh									sO4'3
6.	☐	1.98	7K00|1|a|C|1691	7K00|1|a|U|1692	7K00|1|a|U|1693	7K00|1|a|C|1694	7K00|1|a|G|1695	7K00|1|a|G|1696	C-U-U-C-G-G	s35	cWW	s35	tSW	s55	s53	s55	8BPh		n0BR							sO4'3
7.	☐	1.5	7EOG|1|A|C|19	7EOG|1|A|U|20	7EOG|1|A|U|21	7EOG|1|A|C|22	7EOG|1|A|G|23	7EOG|1|A|G|24	C-U-U-C-G-G	s35	cWW	s35	tSW	s55	s53	s55	8BPh						n3BR			sO4'3
8.	☐	2.0	7KKV|1|B|C|517	7KKV|1|B|U|518	7KKV|1|B|U|519	7KKV|1|B|C|520	7KKV|1|B|G|521	7KKV|1|B|G|522	C-U-U-C-G-G	s35	cWW	s35	tSW	s55	s53	s55	8BPh		n0BR							sO4'3
9.	☑	2.0	3U4M|1|B|C|2143	3U4M|1|B|U|2144	3U4M|1|B|C|2145	3U4M|1|B|C|2146	3U4M|1|B|G|2147	3U4M|1|B|G|2148	C-U-C-C-G-G	s35	cWW	s35	tSW	s55	s53	s55	8BPh		n0BR							sO4'3
10.	☐	2.0	4ARC|1|B|G|47|I|A	4ARC|1|B|U|47|II|B	4ARC|1|B|U|47|II|C	4ARC|1|B|C|47|II|D	4ARC|1|B|G|47|II|E	4ARC|1|B|C|47|II|F	G-U-U-C-G-C	s35	cWW	s35	tSW		s53	s55	8BPh		n0BR							sO4'3
11.	☐	2.0	6DCB|1|B|C|11|III|5_564	6DCB|1|B|U|12|III|5_564	6DCB|1|B|U|13|III|5_564	6DCB|1|B|C|14|III|5_564	6DCB|1|B|G|15|III|5_564	6DCB|1|B|G|16|III|5_564	C-U-U-C-G-G	s35	cWW	s35	tSW	s55	s53	s55	8BPh		0BR				n3BR			sO4'3

Figure 5. A WebFR3D symbolic query for r(nU$_2$NNGn) Z-turn motifs. (**Top**) Panel showing the query constraints. (**Bottom**) Partial view of the WebFR3D result panel showing the first 12 instances, the tetraloop of the 3U4M structure, and a heat map showing the 14 by 14 matrix of mutual geometric discrepancies between the instances. The 2.9 Å lp … π contact between the ribose O4′ and the guanine C2 atom is visualized by a dashed line. This search shows that imposing a trans-Sugar … Watson-Crick (tSW) U$_2$•G$_5$ pair (yellow box) and nucleotides that are consecutive in the (n, n + 1) or (5′ to 3′) direction ("next" in blue boxes) is sufficient to retrieve fourteen r(nUNNGn) folds with the Z-turn characteristics discussed in Section 3.4 and present in the 3.230 representative set of structures [67]. However, a more stringent query is needed to discard problematic structures from this and lower resolution ensembles (Figure S3). Use the following link to retrieve the results of this search: http://rna.bgsu.edu/webfr3d/Results/62c6809102051/62c6809102051.html (accessed on 1 May 2022).

Among the six r(UNCG) loops found in the 1.98 Å resolution 7K00 cryo-EM structure, the best-defined densities are those of r(U$_{1692:a}$UCG) (Figure 6b) and to a lesser extent of the r(U$_{343:A}$ACG) loop. The former tetraloop is the only match to the 3U4M X-ray (2.0 Å) structure in terms of the quality of both the model and the experimental data. The U•G bridging water molecule is visible in these two tetraloops. On the other hand, the r(U$_{1135:A}$CCG), r(U$_{1450:A}$CCG), and r(U$_{138:a}$UCG) loops feature poor densities as shown by the PDB deposited map visualized with Coot [72]. Finally, the r(U$_{420:A}$UCG) loop has a (C$_{423:A}$)O2 … N1(G$_{424:A}$) distance of 2.28 Å because C$_{423:A}$ is unduly modeled in *syn* (Figure 6c).

At this stage, a word of caution is needed. Since these structures were deposited by experienced researchers, it is reasonable to trust the data within the limits of the claimed resolution. However, numerous studies reported that blindly trusting deposited PDB structures can lead to serious data interpretation flaws [66,74–78]. Occasionally, as mentioned above, it is required to inspect the experimental density maps to correct inconsistencies that were not perceived by the authors of the structures. For instance, without careful inspection of electron densities, one could be inclined to consider that the *syn* conformation of the 3rd residue of the 5OB3 tetraloop and other minor structural deviations are the result of natural hairpin dynamics. However, they more likely result from poor local modeling due to an incomplete refinement process combined or not with insufficient experimental data [74,75,78–81].
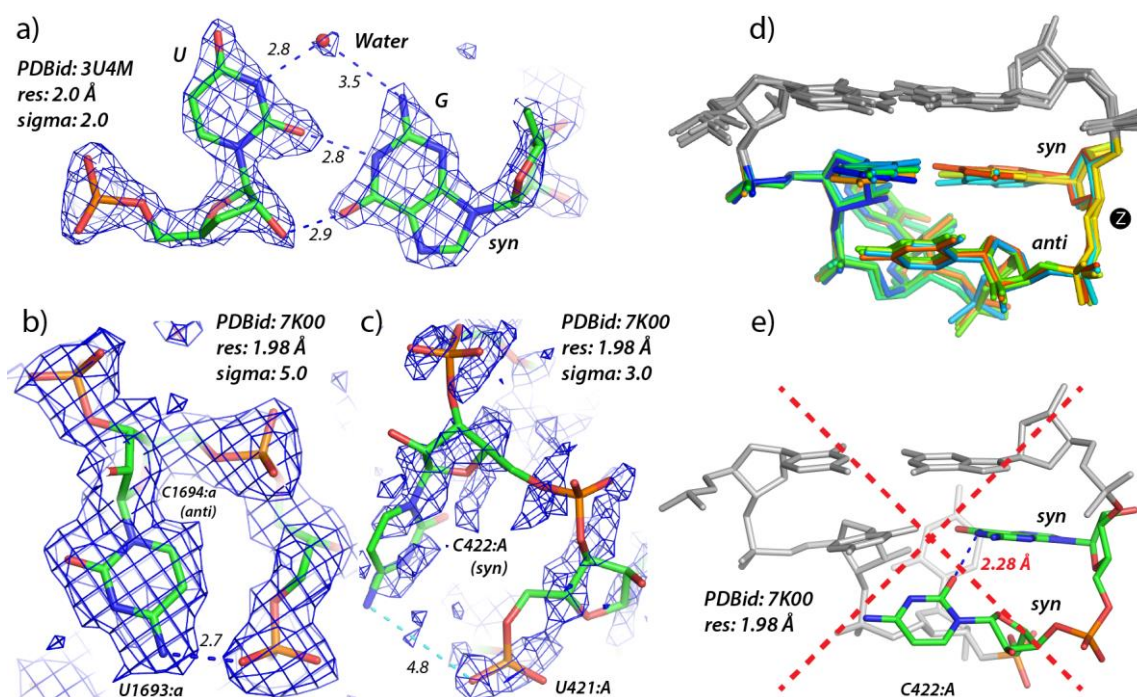
Figure 6. **r(UUCG) structural details derived from X-ray and cryo-EM.** (**a**) A tSW $U_{2144:B} \bullet G_{2147:B}$ pair that is part of an r(UUCG) tetraloop showing a water molecule bridging an amino and an imino Watson-Crick site (PDBid: 3U4M). (**b**,**c**) Views of the interaction are shown in Figure 1d for two tetraloops extracted from the 7K00 cryo-EM structure. In (**c**), the density is poor at a 3.0 sigma level and the C422:A residue has been modelled in *syn* leading to a loss of an important interaction shown in (**b**) where the density at a higher 5.0 sigma level is unambiguous. Note that X-ray and cryo-EM density sigma levels are not equivalent. (**d**) Backbone superimposition of the five high-quality 5Y85, 3U4M, 7KKV, 4ARC, and 7K00 r($U_{1692:a}$UCG) Z-turns mentioned in the text. The superimposition was made by using Chimera [73]. (**e**) View of the 7K00 r($U_{420:A}$UCG) tetraloop with the cytosine in *syn* and an unrealistic 2.28 Å hydrogen-bond contact [66]. This short contact reveals major modelling issues in regions of a 1.98 Å resolution cryo-EM structure and explains why this tetraloop structure must be discarded.

Finding structural inconsistencies in a structural ensemble is tedious work. It can be realized, as we detailed above, on small structural ensembles through careful examination of experimental data to validate or invalidate the model. However, for the larger structural ensembles obtained for searches with resolutions ≤ 3.0 Å, this is an impossible task. Thus, more constraints need to be added to filter out inappropriate structures. A search involving resolutions ≤ 2.0 Å with the Figure 5 criteria generates 14 hits, while the same search using the Figure S3 criteria that add an nt4-nt5 sO4′3 constraint and an *anti* constraint on nt3 and nt4 generates 7 hits closely matching our manually curated ensemble. Similarly, a search involving resolutions ≤ 3.0 Å with the Figure 5 criteria generates 82 hits, while the same search using the Figure S3 criteria generates 49 hits.

The water molecule in the U●G cleft might also be considered as part of the r(UNCG) signature (Figure 6a). However, WebFR3D does not currently allow solvent molecule searches. Additionally, including a solvent constraint would result in a low number of hits given that water molecules are only present in a subset of high-resolution structures. Yet, the presence of this water molecule should be considered a hallmark of high-quality experimental data associated with accurate modeling.

3.5. Finding Unusual r(GNNA) Z-Turns

WebFR3D can be used to search regular r(UNCG) Z-turns but can also be used to explore the sequence variability of these structural motifs. In preceding studies [13,31], we

identified loops with r(GAAA) sequences that adopt a Z-turn conformation with a tSW G•A pair (Figure 7). This was surprising since most of the r(GAAA) or r(GNRA) tetraloops are known to adopt a U-turn conformation. Such conformations are retrieved by WebFR3D when the r(nGNNAn) sequence, tSW G•A pair, and nt4-nt5 sO4'3 constraints are imposed (Figure S4). WebFR3D also isolated the r(GACA) sequence (a non-GNRA sequence; Figure S4) that has the ability to fold as a Z-turn in the 2.2 Å resolution 7OF0 human mitochondrial ribosome [82] and an r(GUGA) Z-turn in the 1.3 Å resolution 4LGT structure [83]. These r(GAAA), r(GUGA), and r(GACA) Z-turns occur all at the same location in the core of the large ribosomal sub-unit [13,31]. It is important to note that to find these motifs, no tetraloop closing base pair constraint must be imposed given that most if not all of them are pentaloops with a 5th bulging residue. We call this U-to-Z tetraloop motif transition a shapeshifting process that is similar in spirit to the flipons described by A. Herbert. Flipons define a class of sequences capable of forming either left or right-handed helical structures [84,85].
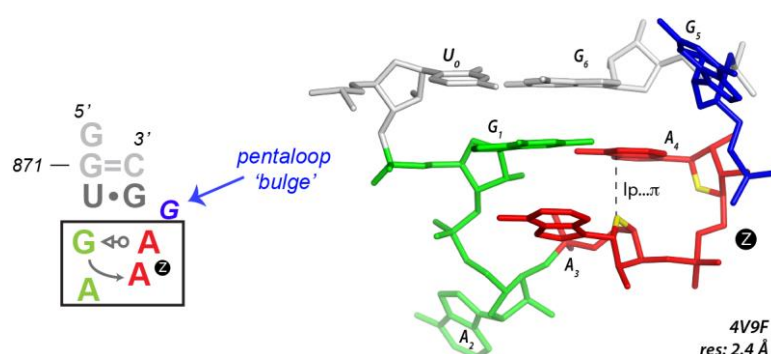
Figure 7. An r(G$_1$AAAg) pentaloop that adopts a Z-turn rather than an expected U-turn. The Z-step is marked by a circled "Z" in both the secondary and tertiary structure of the ribosomal H35a tetraloop of the *Haloarcula marismortui* large ribosomal subunit (PDBid: 4V9F). This motif is conserved in all known ribosome structures. It can be considered as a "shapeshifting" motif exemplifying the ability of an r(GAAA) loop to switch from U-turns to Z-turns in specific environments [13,31]. Note that a U-turn is characterized by an s3OP2 phosphate . . . π contact between the first and third residue (not shown), while Z-turns are characterized by sO4'3 between the 3rd and 4th residue and Z$_{anti}$-turns by sO4'5 between the 3rd and 4th residue (Figure S5).

3.6. UNNG versus CNNG Z-Turns: Isosteric or Not?

It is appreciated that U•G and C•G tSW pairs that comprise a Y•G pair O2 . . . N1 hydrogen bond (see Figure 1c) are isosteric [86]. As such it seems reasonable to assume that r(UNNG) and r(CNNG) sequences would be equally favored. However, it is observed that r(UNNG) are more represented than r(CNNG) tetraloops [13]. This is reflected by WebFR3D searches that identify 56 instances of r(UNNG) versus 14 instances of r(CNNG) Z-turns at resolutions $\leq$ 3.0 Å.

Although the observed Z-turn structures are isostructural, we propose a hypothesis that could explain the underrepresentation of r(CNNG) tetraloops. When cytosine and guanines are close in space, they tend to form C=G pairs and this results in the formation of a Z$_{anti}$-turn di-loop [13] with the 4th guanine in *anti* rather than *syn* while the ribose head-to-head conformation is preserved. However, in Z$_{anti}$-turns, the sO4'5 lp . . . π contact prevails given the *anti* conformation of the 5'-nucleotide. This Z$_{anti}$-turn structure seems more constrained and competes with the formation of an r(CNNG) Z-turn. We found 23 r(CNNG) Z$_{anti}$-turns at resolutions $\leq$ 3.0 Å. It is interesting that these Z$_{anti}$-turns are still isosteric with Z-turns. They keep the head-to-head ribose orientation in the NpG step, the sO4'3 is replaced by a sO4'5 contact, and the second base is bulged out. It could be argued that Z$_{anti}$-turns with a U$_1$•G$_4$ cWW pair could also compete with the canonical r(UNNG) Z-turns. Only one such r(UUCG) Z$_{anti}$-turn has been found, in the 6ERI chloroplast ribosome structure at 3.0 Å resolution [87].

The Figure S5 search expands on the number of sequences that can form Z_{anti}-turns characterized previously [13]. At resolutions ≤ 3.0 Å, we found 32 examples of Z_{anti}-turns that involved the r(CNNG) sequence described above, but also r(UNNA), r(UUCG), and r(GNNC) sequences. Only two of them with an r(CUUG) and an r(UUCG) is closed by a Watson-Crick pair and can be considered as tetraloop hairpins (PDBid: 6AZ3; res. 2.5 Å and PDBid: 7P7Q; res. 2.4 Å). However, the fact that these structures are few, and most of them are located in ribosomes of medium resolution, calls for caution and needs confirmation.

4. Discussion

4.1. Outliers: Validating, Correcting, or Discarding

An exploratory data analysis of the WebFR3D results is a good way to spot outliers for a given set of structures. A rough identification of outliers may be based on the heatmaps provided by the WebFR3D searches (Figures 5 and S1–S5) that allow rapid identification of structural variations, which can then be individually inspected. These outliers may correspond to rare but real conformations of a given motif or to a locally deficient model based on poorly interpreted experimental data as detailed in Section 3.4. When outliers are identified, it is advised to verify if the model agrees with experimental data and with current knowledge to decide if the explored structural fragment should be categorized as a new conformation, excluded from the dataset, or corrected.

While correcting the structures seems the best option, choosing this process depends on the quality of the available electron density maps. When experimental data are of poor quality, it is not advised to attempt such a correction, and discarding the structure seems a better choice. However, when one is interested in using a structure for initiating MD simulations, it is advised to check the structures and eventually correct any visible flaws [78].

4.2. Use of Structural Signatures for MD Simulations

If one is concerned by the integrity of an r(UNCG) structural model derived from MD simulations [51], the WebFR3D constraints defined in Figure S3 should be monitored. Moreover, additional structural features should be examined like the Figure 1d cytosine . . . ribose interaction that is also part of the structural signature of an r(UNCG) loop. The variations in ribose puckers and dihedral angles also need to be scrutinized. Moreover, the presence of a water molecule in the cleft of the tSW U•G pair (Figure 6a) could be monitored since this water might be considered as part of the r(UNCG) loop signature and should display long residency times as suggested by NMR data. The question remains whether MD simulations should reproduce this solvation feature. The answer is probably yes. However, these solvation features may be less stable than the tSW base pairing.

Recently, several of these contacts were monitored in MD simulations including a labile 0BPh contact involving the G of the first stem base pair [54,59]. Such C-H . . . O contact [88] are very subtle and difficult to model although recent parameterization efforts involving a modification of vdW radii improved the sampling of the loops. Unfortunately, the conclusions of these studies reveal that we are currently far from mastering all interactions that are essential for modeling the apparently "simple" tetraloop systems [54,59].

4.3. WebFR3D Limitations and Strategy

Additional features are regularly being added to the WebFR3D search interface but, understandably, not all conceivable search constraints are available. For example, one cannot use the presence or absence of a solvent molecule as a search constraint, nor can one constrain a search by sugar pucker or backbone conformation. Instead, these features must be evaluated on a case-by-case basis in the relatively small set of instances returned by searches targeting high-resolution structures.

A recurrent issue in searches of biological 3D structures is over-constraining the search and missing instances of interest because they happen to lack one or more of the specified criteria. Sometimes, one should use as few symbolic constraints as possible to get the

instances of interest; a search that insists on all idealized features of a motif may return no or very few results! For avoiding such issues, one can widen the search criteria by allowing "near" annotations. For example, when searching for an r(UNNG) turn and using a U•G tSW base pair constraint, one may wish to allow both "near" and "true" tSW base pairs by typing "tSW ntSW UG" or "n+tSW UG" in the appropriate WebFR3D yellow search box. This will return a wider range of instances, which can be evaluated manually. Finally, if one already has an instance of a desired motif, one can search for geometrically similar instances up to a user specified geometric discrepancy value, while still imposing a minimal set of symbolic constraints, and thus avoid over-constraining a search based on incorrect guesses about how instances of interest will vary.

5. Conclusions

Our WebFR3D lp ... π constraint implementation was demonstrated to be effective in retrieving instances of all categories of lp ... π contacts between consecutive and non-consecutive nucleotides in RNA. As such, we confirmed preceding findings that assessed that basically all 16 NpN combinations can form Z-steps with the current exception of CpC steps and showed without surprise that the most represented ones are CpG Z-steps. The current implementation of WebFR3D is also able to retrieve all kinds of motifs including a phosphate ... π contact.

We were also able to successfully retrieve all types of Z-turns described elsewhere [13]. In addition, through the fine tuning of structural constraints, we could eliminate from an ensemble of r(UNCG) structures with resolutions < 2.0 Å, locally poorly resolved structures and refine the r(UNCG) signature to provide better comparison points for structural and molecular dynamics (MD) simulation studies.

WebFR3D proved also effective in being able to search for r(NNNN) sequences that fold as Z-turns, that is to search for all the sequences able to adopt a given 3D shape. For instance, we retrieved r(GAAA) motifs that are isostructural to r(UNCG) Z-turns. Another highlighted example is related to r(CNNG) sequences that can fold as Z-turns and are therefore almost completely isostructural to r(UNCG) loops. Alternatively, they can fold as Z_{anti}-turns and adopt a structure with a $C_1=G_4$ Watson-Crick base pair and a G_4 nucleotide in *anti* rather than in *syn*. These loops are also isostructural with r(UNCG) loops.

Such examples demonstrate that searching by shape is complementary to searching by sequence and that when studying an RNA for which 3D structures are not available, one should consider the multiple shapes a given sequence can adopt by exploring their diversity with tools such as WebFR3D.

6. Methods

WebFR3D search tool: WebFR3D is the online implementation of FR3D, a general-purpose RNA motif search tool [4,5]. WebFR3D makes it possible to search for RNA motifs of one to over a dozen nucleotides in different search modes. In a geometric search, the user specifies a list of RNA nucleotides from a 3D structure from the Protein Data Bank and a set of structures to search, and WebFR3D is guaranteed to find all matches up to a user-specified tolerance called the discrepancy cutoff. Geometric searches can be augmented with symbolic constraints to require certain base identities, pairwise interactions, or chain continuity constraints. Some motifs can be described entirely by symbolic constraints, and WebFR3D makes it possible to search for those based on the constraints alone. WebFR3D has pre-computed annotations of RNA base pairs, base stacking, base-backbone, and other interactions for all RNA-containing 3D structures in PDB. For statistical surveys, one can avoid the redundancy inherent in the 3D structure database by searching a representative set of 3D structures, at a chosen resolution threshold [67].

Attribution of a nucleobase 3′ and 5′-face: For each standard RNA base, the BGSU RNA pipeline [4,5] calculates the geometric center of the heavy (non-hydrogen) atoms of the base, weighting each heavy atom equally. From this point, the normal vector to the base is oriented to point toward the 3′-direction in a regular RNA helix. We refer to that

nucleobase face as the 3′-face; the opposite face becomes the 5′-face. This convention was introduced earlier and is also used to annotate base-base stacking by FR3D [4].

lp . . . π search: For each nucleotide pair, the relative locations of the base of the first nucleotide and the backbone oxygen atoms of the second are computed in the following way. For each of the oxygen atoms in the phosphate backbone of the second nucleotide, the vertical coordinate along the normal vector of the first base, called z, is determined. For $|z| \leq 3.5$ Å, the oxygen atom is projected onto the plane of the first base; for an lp . . . π contact, it must lie inside a nucleobase ring. If more than one oxygen atom projects inside a ring, we keep the oxygen atom with the smallest $|z|$ value. We annotate an oxygen stacking interaction according to the oxygen atom and the face of the first base, producing a string such as sO4′3 when the O4′ atom stacks with the 3′-face. See Figure S6 for plots of the projected oxygen atoms having minimal $|z|$ values.

When the criteria for an lp . . . π contact are not met, we check for "near" stacking interactions to soften the boundaries between contacts that have all the desired properties and contacts that are nowhere close to sO stacking.

Case 1: An oxygen has 3.5 Å $< |z| \leq 3.6$ Å and projects inside a ring; we keep the oxygen atom with the smallest $|z|$ value.

Case 2: An oxygen has $|z| \leq 3.5$ Å and projects outside the ring(s) but within an ellipse around a ring which we describe now.

For each standard RNA base ring, the ellipse that comes closest to the five or six atoms on a ring is drawn (note that the eccentricities of the ellipses are small but not zero). Then we expand the ellipse to 0.3 Å beyond the corners of the ring, as shown in green in the bottom panel of Figure S6. Among such oxygen atoms, we annotate the one that is closest to the geometric center of the base, to produce an annotation such as nsO4′3 where "n" stands for "near". For searching both regular and near instances, one may use a text string such as n+sO4′3. Figure S6 shows the projected oxygen atoms in sO and near sO (nsO) interactions. This methodology resembles that described in [12].

References

1. Auffinger, P.; Westhof, E. An extended structural signature for the tRNA anticodon loop. *RNA* **2001**, *7*, 334–341. [CrossRef] [PubMed]
2. Leontis, N.B.; Altman, R.B.; Berman, H.M.; Brenner, S.E.; Brown, J.W.; Engelke, D.R.; Harvey, S.C.; Holbrook, S.R.; Jossinet, F.; Lewis, S.E.; et al. The RNA Ontology Consortium: An open invitation to the RNA community. *RNA* **2006**, *12*, 533–541. [CrossRef] [PubMed]
3. Lu, X.J.; Olson, W.K. 3DNA: A versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures. *Nat. Protoc.* **2008**, *3*, 1213–1227. [CrossRef] [PubMed]
4. Sarver, M.; Zirbel, C.L.; Stombaugh, J.; Mokdad, A.; Leontis, N.B. FR3D: Finding local and composite recurrent structural motifs in RNA 3D structures. *J. Math. Biol.* **2008**, *56*, 215–252. [CrossRef] [PubMed]
5. Petrov, A.I.; Zirbel, C.L.; Leontis, N.B. WebFR3D—A server for finding, aligning and analyzing recurrent RNA 3D motifs. *Nucleic Acids Res.* **2011**, *39*, W50–W55. [CrossRef]
6. Lu, X.J.; Bussemaker, H.J.; Olson, W.K. DSSR: An integrated software tool for dissecting the spatial structure of RNA. *Nucleic Acids Res.* **2015**, *43*, e142. [CrossRef]
7. Sweeney, B.A.; Roy, P.; Leontis, N.B. An introduction to recurrent nucleotide interactions in RNA. *Wiley Interdiscip. Rev. RNA* **2015**, *6*, 17–45. [CrossRef]
8. Parlea, L.G.; Sweeney, B.A.; Hosseini-Asanjan, M.; Zirbel, C.L.; Leontis, N.B. The RNA 3D Motif Atlas: Computational methods for extraction, organization and evaluation of RNA motifs. *Methods* **2016**, *103*, 99–119. [CrossRef]
9. Berman, H.M.; Lawson, C.L.; Schneider, B. Developing community resources for nucleic acid structures. *Life* **2022**, *12*, 540. [CrossRef]
10. Roy, P.; Bhattacharyya, D. Contact networks in RNA: A structural bioinformatics study with a new tool. *J. Comput. Aided Mol. Des.* **2022**, *36*, 131–140. [CrossRef]
11. Leontis, N.B.; Lescoute, A.; Westhof, E. The building blocks and motifs of RNA architecture. *Curr. Opin. Struct. Biol.* **2006**, *16*, 279–287. [CrossRef] [PubMed]
12. D'Ascenzo, L.; Leonarski, F.; Vicens, Q.; Auffinger, P. 'Z-DNA like' fragments in RNA: A recurring structural motif with implications for folding, RNA/protein recognition and immune response. *Nucleic Acids Res.* **2016**, *44*, 5944–5956. [CrossRef] [PubMed]
13. D'Ascenzo, L.; Leonarski, F.; Vicens, Q.; Auffinger, P. Revisiting GNRA and UNCG folds: U-turns versus Z-turns in RNA hairpin loops. *RNA* **2017**, *23*, 259–269. [CrossRef] [PubMed]
14. Egli, M.; Sarkhel, S. Lone pair-aromatic interactions: To stabilize or not to stabilize. *Acc. Chem. Res.* **2007**, *40*, 197–205. [CrossRef]
15. Chawla, M.; Chermak, E.; Zhang, Q.; Bujnicki, J.M.; Oliva, R.; Cavallo, L. Occurrence and stability of lone pair-π stacking interactions between ribose and nucleobases in functional RNAs. *Nucleic Acids Res.* **2017**, *45*, 11019–11032. [CrossRef] [PubMed]
16. Kozelka, J. Lone pair-π interactions in biological systems: Occurrence, function, and physical origin. *Eur. Biophys. J.* **2017**, *46*, 729–737. [CrossRef] [PubMed]
17. Kruse, H.; Mrazikova, K.; D'Ascenzo, L.; Sponer, J.; Auffinger, P. Short but weak: The Z-DNA lone-pair . . . π conundrum challenges standard carbon Van der Waals radii. *Angew. Chem. Int. Ed. Engl.* **2020**, *59*, 16553–16560. [CrossRef]
18. Cheong, C.; Varani, G.; Tinoco, I. Solution structure of an unusually stable RNA hairpin, 5'GGAC(UUCG)GUCC. *Nature* **1990**, *346*, 680–681. [CrossRef]
19. Varani, G.; Cheong, C.; Tinoco, I., Jr. Structure of an unusually stable RNA hairpin. *Biochemistry* **1991**, *30*, 3280–3289. [CrossRef]
20. Richardson, J.S.; Schneider, B.; Murray, L.W.; Kapral, G.J.; Immormino, R.M.; Headd, J.J.; Richardson, D.C.; Ham, D.; Hershkovits, E.; Williams, L.D.; et al. RNA backbone: Consensus all-angle conformers and modular string nomenclature (an RNA Ontology Consortium contribution). *RNA* **2008**, *14*, 465–481. [CrossRef]
21. Sheehy, J.P.; Davis, A.R.; Znosko, B.M. Thermodynamic characterization of naturally occurring RNA tetraloops. *RNA* **2010**, *16*, 417–429. [CrossRef] [PubMed]
22. Allain, F.H.-T.; Varani, G. Structure of the P1 helix from group I self-splicing introns. *J. Mol. Biol.* **1995**, *250*, 333–353. [CrossRef] [PubMed]
23. Nozinovic, S.; Furtig, B.; Jonker, H.R.; Richter, C.; Schwalbe, H. High-resolution NMR structure of an RNA model system: The 14-mer cUUCGg tetraloop hairpin RNA. *Nucleic Acids Res.* **2010**, *38*, 683–694. [CrossRef] [PubMed]
24. Ennifar, E.; Nikulin, A.; Tishchenko, S.; Serganov, A.; Nevskaya, N.; Garber, M.; Ehresmann, B.; Ehresmann, C.; Nikonov, S.; Dumas, P. The crystal structure of UUCG tetraloop. *J. Mol. Biol.* **2000**, *304*, 35–42. [CrossRef] [PubMed]
25. Borkar, A.N.; Vallurupalli, P.; Camilloni, C.; Kay, L.E.; Vendruscolo, M. Simultaneous NMR characterisation of multiple minima in the free energy landscape of an RNA UUCG tetraloop. *Phys. Chem. Chem. Phys.* **2017**, *19*, 2797–2804. [CrossRef]
26. Nichols, P.J.; Henen, M.A.; Born, A.; Strotz, D.; Guntert, P.; Vogeli, B. High-resolution small RNA structures from exact nuclear Overhauser enhancement measurements without additional restraints. *Commun. Biol.* **2018**, *1*, 61. [CrossRef]
27. Bottaro, S.; Nichols, P.J.; Vogeli, B.; Parrinello, M.; Lindorff-Larsen, K. Integrating NMR and simulations reveals motions in the UUCG tetraloop. *Nucleic Acids Res.* **2020**, *48*, 5839–5848. [CrossRef]
28. Orioli, S.; Larsen, A.H.; Bottaro, S.; Lindorff-Larsen, K. How to learn from inconsistencies: Integrating molecular simulations with experimental data. *Prog. Mol. Biol. Transl. Sci.* **2020**, *170*, 123–176.

29. Liu, B.; Shi, H.; Al-Hashimi, H.M. Developments in solution-state NMR yield broader and deeper views of the dynamic ensembles of nucleic acids. *Curr. Opin. Struct. Biol.* **2021**, *70*, 16–25. [CrossRef]

30. Nichols, P.J.; Bevers, S.; Henen, M.; Kieft, J.S.; Vicens, Q.; Vogeli, B. Recognition of non-CpG repeats in Alu and ribosomal RNAs by the Z-RNA binding domain of ADAR1 induces A-Z junctions. *Nat. Commun.* **2021**, *12*, 793. [CrossRef]

31. D'Ascenzo, L.; Vicens, Q.; Auffinger, P. Identification of receptors for UNCG and GNRA Z-turns and their occurrence in rRNA. *Nucleic Acids Res.* **2018**, *46*, 7989–7997. [CrossRef] [PubMed]

32. Brzezinski, K.; Brzuszkiewicz, A.; Dauter, M.; Kubicki, M.; Jaskolski, M.; Dauter, Z. High regularity of Z-DNA revealed by ultra high-resolution crystal structure at 0.55 Å. *Nucleic Acids Res.* **2011**, *39*, 6238–6248. [CrossRef] [PubMed]

33. Sokoloski, J.E.; Godfrey, S.A.; Dombrowski, S.E.; Bevilacqua, P.C. Prevalence of *syn* nucleobases in the active sites of functional RNAs. *RNA* **2011**, *17*, 1775–1787. [CrossRef] [PubMed]

34. Luo, Z.P.; Dauter, M.; Dauter, Z. Phosphates in the Z-DNA dodecamer are flexible, but their P-SAD signal is sufficient for structure solution. *Acta Cryst.* **2014**, *D70*, 1790–1800. [CrossRef]

35. Martinez, C.R.; Iverson, B.L. Rethinking the term "pi-stacking". *Chem. Sci.* **2012**, *3*, 2191–2201. [CrossRef]

36. Carter-Fenk, K.; Herbert, J.M. Reinterpreting π-stacking. *Phys. Chem. Chem. Phys.* **2020**, *22*, 24870–24886. [CrossRef]

37. Mrazikova, K.; Sponer, J.; Mlynsky, V.; Auffinger, P.; Kruse, H. Short-range imbalances in the AMBER Lennard-Jones potential for (deoxy)ribose . . . nucleobase lone-pair . . . π contacts in nucleic acids. *J. Chem. Inf. Model.* **2021**, *61*, 5644–5657. [CrossRef]

38. Drew, H.R.; Dickerson, R.E. Conformation and dynamics in a Z-DNA tetramer. *J. Mol. Biol.* **1981**, *152*, 723–736. [CrossRef]

39. Wang, A.H.J.; Quigley, G.J.; Kolpak, F.J.; Vandermarel, G.; Vanboom, J.H.; Rich, A. Left-handed double helical DNA—Variations in the backbone conformation. *Science* **1981**, *211*, 171–176. [CrossRef]

40. Gessner, R.V.; Frederick, C.A.; Quigley, G.J.; Rich, A.; Wang, A.H.J. The molecular structure of the left-handed Z-DNA double helix at 1.0 Å atomic resolution. Geometry, conformation, and ionic interactions of d(CGCGCG). *J. Biol. Chem.* **1989**, *264*, 7912–7935. [CrossRef]

41. Svozil, D.; Kalina, J.; Omelka, M.; Schneider, B. DNA conformations and their sequence preferences. *Nucleic Acids Res.* **2008**, *36*, 3690–3706. [CrossRef] [PubMed]

42. Saenger, W. *Principles of Nucleic Acid Structure*; Springer: New York, NY, USA, 1984.

43. Banas, P.; Hollas, D.; Zgarbova, M.; Jurecka, P.; Orozco, M.; Cheatham III, T.E.; Sponer, J.; Otyepka, M. Performance of molecular mechanics force fields for RNA simulations: Stability of UUCG and GNRA hairpins. *J. Chem. Theory. Comput.* **2010**, *6*, 3836–3849. [CrossRef] [PubMed]

44. Chen, A.A.; Garcia, A.E. High-resolution reversible folding of hyperstable RNA tetraloops using molecular dynamics simulations. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 16820–16825. [CrossRef]

45. Kührova, P.; Banas, P.; Best, R.B.; Sponer, J.; Otyepka, M. Computer folding of RNA tetraloops? Are we there yet? *J. Chem. Theory Comput.* **2013**, *9*, 2115–2125. [CrossRef] [PubMed]

46. Bergonzo, C.; Henriksen, N.M.; Roe, D.R.; Cheatham, T.E., 3rd. Highly sampled tetranucleotide and tetraloop motifs enable evaluation of common RNA force fields. *RNA* **2015**, *21*, 1578–1590. [CrossRef]

47. Haldar, S.; Kuhrova, P.; Banas, P.; Spiwok, V.; Sponer, J.; Hobza, P.; Otyepka, M. Insights into stability and folding of GNRA and UNCG tetraloops revealed by microsecond molecular dynamics and well-tempered metadynamics. *J. Chem. Theory Comput.* **2015**, *11*, 3866–3877. [CrossRef] [PubMed]

48. Giambasu, G.M.; York, D.M.; Case, D.A. Structural fidelity and NMR relaxation analysis in a prototype RNA hairpin. *RNA* **2015**, *21*, 963–974. [CrossRef]

49. Bottaro, S.; Banas, P.; Sponer, J.; Bussi, G. Free energy landscape of GAGA and UUCG RNA tetraloops. *J. Phys. Chem. Lett.* **2016**, *7*, 4032–4038. [CrossRef]

50. Aytenfisu, A.H.; Spasic, A.; Grossfield, A.; Stern, H.A.; Mathews, D.H. Revised RNA dihedral parameters for the Amber force field improve RNA molecular dynamics. *J. Chem. Theory Comput.* **2017**, *13*, 900–915. [CrossRef]

51. Kuhrova, P.; Best, R.B.; Bottaro, S.; Bussi, G.; Sponer, J.; Otyepka, M.; Banas, P. Computer folding of RNA tetraloops: Identification of key force field deficiencies. *J. Chem. Theory Comput.* **2016**, *12*, 4534–4548. [CrossRef]

52. Sponer, J.; Bussi, G.; Krepl, M.; Banas, P.; Bottaro, S.; Cunha, R.A.; Gil-Ley, A.; Pinamonti, G.; Poblete, S.; Jurecka, P.; et al. RNA structural dynamics as captured by molecular simulations: A comprehensive overview. *Chem. Rev.* **2018**, *118*, 4177–4338. [CrossRef] [PubMed]

53. Zgarbova, M.; Sponer, J.; Jurecka, P. Z-DNA as a touchstone for additive empirical force fields and a refinement of the alpha/gamma DNA torsions for AMBER. *J. Chem. Theory Comput.* **2021**, *17*, 6292–6301. [CrossRef] [PubMed]

54. Mrazikova, K.; Mlynsky, V.; Kuhrova, P.; Pokorna, P.; Kruse, H.; Krepl, M.; Otyepka, M.; Banas, P.; Sponer, J. UUCG RNA tetraloop as a formidable force-field challenge for MD simulations. *J. Chem. Theory. Comput.* **2020**, *16*, 7601–7617. [CrossRef] [PubMed]

55. Tan, D.; Piana, S.; Dirks, R.M.; Shaw, D.E. RNA force field with accuracy comparable to state-of-the-art protein force fields. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, E1346–E1355. [CrossRef]

56. Cesari, A.; Bottaro, S.; Lindorff-Larsen, K.; Banas, P.; Sponer, J.; Bussi, G. Fitting corrections to an RNA force field using experimental data. *J. Chem. Theory Comput.* **2019**, *15*, 3425–3431. [CrossRef]

57. Kuhrova, P.; Mlynsky, V.; Zgarbova, M.; Krepl, M.; Bussi, G.; Best, R.B.; Otyepka, M.; Sponer, J.; Banas, P. Improving the performance of the Amber RNA force field by tuning the hydrogen-bonding interactions. *J. Chem. Theory Comput.* **2019**, *15*, 3288–3305. [CrossRef]

58. Chen, J.; Liu, H.; Cui, X.; Li, Z.; Chen, H.F. RNA-specific force field optimization with CMAP and reweighting. *J. Chem. Inf. Model.* **2022**, *62*, 372–385. [CrossRef]

59. Mlynsky, V.; Janecek, M.; Kuhrova, P.; Frohlking, T.; Otyepka, M.; Bussi, G.; Banas, P.; Sponer, J. Toward Convergence in Folding Simulations of RNA Tetraloops: Comparison of Enhanced Sampling Techniques and Effects of Force Field Modifications. *J. Chem Theory Comput* **2022**, *18*, 2642–2656. [CrossRef]

60. Kuhrova, P.; Mlynsky, V.; Zgarbova, M.; Krepl, M.; Bussi, G.; Best, R.B.; Otyepka, M.; Sponer, J.; Banas, P. Correction to "Improving the performance of the Amber RNA force field by tuning the hydrogen-bonding interactions". *J. Chem. Theory Comput.* **2020**, *16*, 818–819. [CrossRef]

61. Ferner, J.; Villa, A.; Duchardt, E.; Widjajakusuma, E.; Wohnert, J.; Stock, G.; Schwalbe, H. NMR and MD studies of the temperature-dependent dynamics of RNA YNMG-tetraloops. *Nucleic Acids Res.* **2008**, *36*, 1928–1940. [CrossRef]

62. Zgarbova, M.; Luque, F.J.; Sponer, J.; Cheatham, T.E., 3rd; Otyepka, M.; Jurecka, P. Toward improved description of DNA backbone: Revisiting Epsilon and Zeta torsion force field parameters. *J. Chem. Theory Comput.* **2013**, *9*, 2339–2354. [CrossRef] [PubMed]

63. Zgarbova, M.; Sponer, J.; Otyepka, M.; Cheatham, T.E., 3rd; Galindo-Murillo, R.; Jurecka, P. Refinement of the sugar-phosphate backbone torsion beta for AMBER force fields improves the description of Z- and B-DNA. *J. Chem. Theory Comput.* **2015**, *11*, 5723–5736. [CrossRef]

64. Galindo-Murillo, R.; Robertson, J.C.; Zgarbova, M.; Sponer, J.; Otyepka, M.; Jurecka, P.; Cheatham, T.E., 3rd. Assessing the current state of Amber force field modifications for DNA. *J. Chem. Theory Comput.* **2016**, *12*, 4114–4127. [CrossRef] [PubMed]

65. Fakharzadeh, A.; Zhang, J.; Roland, C.; Sagui, C. Novel eGZ-motif formed by regularly extruded guanine bases in a left-handed Z-DNA helix as a major motif behind CGG trinucleotide repeats. *Nucleic Acids Res.* **2022**, *50*, 4860–4876. [CrossRef] [PubMed]

66. Kruse, H.; Sponer, J.; Auffinger, P. Comment on "Evaluating Unexpectedly Short Non-covalent Distances in X-ray Crystal Structures of Proteins with Electronic Structure Analysis". *J. Chem. Inf. Model.* **2019**, *59*, 3605–3608. [CrossRef] [PubMed]

67. Leontis, N.B.; Zirbel, C.L. Nonredundant 3D structure datasets for RNA knowledge extraction and benchmarking. In *RNA 3D Structure Analysis and Prediction*; Leontis, N.B., Westhof, E., Eds.; Springer: Berlin/Heidelberg, Germany, 2012; pp. 281–298.

68. Watson, Z.L.; Ward, F.R.; Meheust, R.; Ad, O.; Schepartz, A.; Banfield, J.F.; Cate, J.H. Structure of the bacterial ribosome at 2 Å resolution. *Elife* **2020**, *9*, e60482. [CrossRef]

69. Liu, Z.; Gutierrez-Vargas, C.; Wei, J.; Grassucci, R.A.; Ramesh, M.; Espina, N.; Sun, M.; Tutuncuoglu, B.; Madison-Antenucci, S.; Woolford, J.L., Jr.; et al. Structure and assembly model for the *Trypanosoma cruzi* 60S ribosomal subunit. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 12174–12179. [CrossRef] [PubMed]

70. Knappenberger, A.J.; Reiss, C.W.; Strobel, S.A. Structures of two aptamers with differing ligand specificity reveal ruggedness in the functional landscape of RNA. *Elife* **2018**, *7*, e36381. [CrossRef] [PubMed]

71. Itoh, Y.; Naschberger, A.; Mortezaei, N.; Herrmann, J.M.; Amunts, A. Analysis of translating mitoribosome reveals functional characteristics of translation in mitochondria of fungi. *Nat. Commun.* **2020**, *11*, 5187. [CrossRef]

72. Casanal, A.; Lohkamp, B.; Emsley, P. Current developments in Coot for macromolecular model building of electron cryo-microscopy and crystallographic data. *Protein Sci.* **2020**, *29*, 1069–1078. [CrossRef]

73. Pettersen, E.F.; Goddard, T.D.; Huang, C.C.; Couch, G.S.; Greenblatt, D.M.; Meng, E.C.; Ferrin, T.E. UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25*, 1605–1612. [CrossRef] [PubMed]

74. Dauter, Z.; Wlodawer, A.; Minor, W.; Jaskolski, M.; Rupp, B. Avoidable errors in deposited macromolecular structures: An impediment to efficient data mining. *IUCrJ* **2014**, *1*, 179–193. [CrossRef] [PubMed]

75. Minor, W.; Dauter, Z.; Helliwell, J.R.; Jaskolski, M.; Wlodawer, A. Safeguarding structural data repositories against bad apples. *Structure* **2016**, *24*, 216–220. [CrossRef] [PubMed]

76. Leonarski, F.; D'Ascenzo, L.; Auffinger, P. Mg^{2+} ions: Do they bind to nucleobase nitrogens? *Nucleic Acids Res.* **2017**, *45*, 987–1004. [CrossRef]

77. Wlodawer, A. Stereochemistry and validation of macromolecular structures. *Methods Mol. Biol.* **2017**, *1607*, 595–610.

78. Wlodawer, A.; Dauter, Z.; Porebski, P.J.; Minor, W.; Stanfield, R.; Jaskolski, M.; Pozharski, E.; Weichenberger, C.X.; Rupp, B. Detect, correct, retract: How to manage incorrect structural models. *FEBS J.* **2018**, *285*, 444–466. [CrossRef]

79. Rupp, B.; Wlodawer, A.; Minor, W.; Helliwell, J.R.; Jaskolski, M. Correcting the record of structural publications requires joint effort of the community and journal editors. *FEBS J.* **2016**, *283*, 4452–4457. [CrossRef]

80. Williams, C.J.; Headd, J.J.; Moriarty, N.W.; Prisant, M.G.; Videau, L.L.; Deis, L.N.; Verma, V.; Keedy, D.A.; Hintze, B.J.; Chen, V.B.; et al. MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci.* **2018**, *27*, 293–315. [CrossRef]

81. Roversi, P.; Tronrud, D.E. Ten things I 'hate' about refinement. *Acta Crystallogr. D Struct. Biol.* **2021**, *77*, 1497–1515. [CrossRef]

82. Hillen, H.S.; Lavdovskaia, E.; Nadler, F.; Hanitsch, E.; Linden, A.; Bohnsack, K.E.; Urlaub, H.; Richter-Dennerlein, R. Structural basis of GTPase-mediated mitochondrial ribosome biogenesis and recycling. *Nat. Commun.* **2021**, *12*, 3672. [CrossRef]

83. Czudnochowski, N.; Ashley, G.W.; Santi, D.V.; Alian, A.; Finer-Moore, J.; Stroud, R.M. The mechanism of pseudouridine synthases from a covalent complex with RNA, and alternate specificity for U2605 versus U2604 between close homologs. *Nucleic Acids Res.* **2014**, *42*, 2037–2048. [CrossRef] [PubMed]

84. Herbert, A. A genetic instruction code based on DNA conformation. *Trends Genet.* **2019**, *35*, 887–890. [CrossRef] [PubMed]

85. Herbert, A.; Karapetyan, S.; Poptsova, M.; Vasquez, K.M.; Vicens, Q.; Vögeli, B. Special Issue: A, B and Z: The structure, function and genetics of Z-DNA and Z-RNA. *Int. J. Mol. Sci.* **2021**, *22*, 7686. [CrossRef] [PubMed]
86. Stombaugh, J.; Zirbel, C.L.; Westhof, E.; Leontis, N.B. Frequency and isostericity of RNA base pairs. *Nucleic Acids Res.* **2009**, *37*, 2294–2312. [CrossRef]
87. Boerema, A.P.; Aibara, S.; Paul, B.; Tobiasson, V.; Kimanius, D.; Forsberg, B.O.; Wallden, K.; Lindahl, E.; Amunts, A. Structure of the chloroplast ribosome with chl-RRF and hibernation-promoting factor. *Nat. Plants* **2018**, *4*, 212–217. [CrossRef]
88. Auffinger, P.; Louise-May, S.; Westhof, E. Molecular dynamics simulations of the anticodon hairpin of tRNA(asp): Structuring effects of C-H . . . O hydrogen bonds and of long-range hydration forces. *J. Am. Chem. Soc.* **1996**, *118*, 1181–1189. [CrossRef]