Conformational adaptation of UNCG loops upon crowding

Mélanie Meyer, Hélène Walbott, Vincent Oliéric, Jiro Kondo, Maria Costa, Benoît Masquida

Conformational adaptation of UNCG loops upon crowding

MÉLANIE MEYER,[1] HÉLÈNE WALBOTT,[2] VINCENT OLIÉRIC,[3] JIRO KONDO,[4] MARIA COSTA,[2] and BENOÎT MASQUIDA[5]

[1]PPRS, 68000 Colmar, France
[2]Institute for Integrative Biology of the Cell (I2BC), CEA, CNRS, Université Paris-Sud, Université Paris-Saclay, 91198, Gif-sur-Yvette cedex, France
[3]Paul Scherrer Institute, Swiss Light Source, 5232 Villigen PSI, Switzerland
[4]Department of Materials and Life Sciences, Sophia University, 7-1 Kioi-cho, Chiyoda-ku, 102-8554 Tokyo, Japan
[5]UMR7156 GMGM Université de Strasbourg – CNRS, 67084 Strasbourg, France

ABSTRACT

If the A-form helix is the major structural motif found in RNA, the loops that cap them constitute the second most important family of motifs. Among those, two are overrepresented, GNRA and UNCG tetraloops. Recent surveys of RNA structures deposited in the PDB show that GNRA and UNCG tetraloops can adopt tertiary folds that are very different from their canonical conformations, characterized by the presence of a U-turn of a Z-turn, respectively. Crystallographic data from both a lariat-capping (LC) ribozyme and a group II intron ribozyme reveal that a given UUCG tetraloop can adopt a distinct fold depending on its structural environment. Specifically, when the crystal packing applies relaxed constraints on the loop, the canonical Z-turn conformation is observed. In contrast, a highly packed environment induces "squashing" of the tetraloop by distorting its sugar-phosphate backbone in a specific way that expels the first and fourth nucleobases out of the loop, and falls in van der Waals distance of the last base pair of the helix, taking the place of the pair formed between the first and fourth residues in Z-turn loops. The biological relevance of our observations is supported by the presence of similarly deformed loops in the highly packed environment of the ribosome and in a complex between a dsRNA and a RNase III. The finding that Z-turn loops change conformation under higher molecular packing suggests that, in addition to their demonstrated role in stabilizing RNA folding, they may contribute to the three-dimensional structure of RNA by mediating tertiary interactions with distal residues.

Keywords: UNCG tetraloop; Z-turn loop; lariat-capping ribozyme; group II intron

INTRODUCTION

To fulfill their biological roles, RNA molecules adopt specific structures able to interact with their molecular partners, change conformations and carry out catalytic reactions. Their structures result from folding, a process of condensation of the RNA chain in which concatenation of the Watson–Crick base pairs builds up A-form anti-parallel helices interspersed by single-stranded regions. Nevertheless, the experimentally determined RNA structures stored in the PDB indicate that single-stranded regions also adopt specific structures, which allow the RNA to acquire diverse and complex three-dimensional architectures (Masquida et al. 2010; Westhof et al. 2011; Koculi et al. 2012). One of the most abundant types of motifs found in the RNA structural repertoire is represented by apical loops, which cap helices and allow the outgoing RNA strand to be anti-parallel to the ingoing one. Among the loop motifs, tetraloops of the GNRA and UNCG families are overrepresented.

GNRA tetraloops have long been known to weave long-range tertiary contacts (Michel and Westhof 1990; Pley et al. 1994; Cate et al. 1996). The orientation of the adenosine nucleobases in the loop favor interactions involving their sugar or Watson–Crick edges, called A-minor interactions (Doherty et al. 2001; Nissen et al. 2001). The receptors of these loops have been the subject of many studies (for review, see Fiore and Nesbitt 2013), including by in vitro selection methods (Costa and Michel 1997; Robertson et al. 1999; Barrick et al. 2001; Geary et al. 2008). However, in sharp contrast with GNRA loops, UNCG tetraloops have been mostly considered as nucleating RNA folding due to their exceptional thermodynamic stability (Varani et al.

Corresponding author: b.masquida@unistra.fr
Article is online at http://www.rnajournal.org/cgi/doi/10.1261/rna.072694.119.

1991; Antao and Tinoco 1992). Interestingly, the first attempts to crystallize tetraloops embedded in 4 bp hairpins led to solve structures of extended double helix dodecamers incorporating a set of four central mismatches (for review, see Masquida and Westhof 1999). This situation resulted from oligonucleotide dimerization at the high concentration required for crystal growth. It is only in the context of more complex RNA structures, like in three-way junctions from ribozymes, riboswitches, and ribosomes that the actual tetraloop structures could be finally captured by crystallography (Pley et al. 1994; Ban et al. 2000; Ennifar et al. 2000; Wimberly et al. 2000). These data are consistent with the notion that the collapse of the RNA chain into an organized architecture is dominated by the propagation of double-stranded helices and by the formation of tertiary interactions, and not by loop conformations (Woodson 2010).

The mention of the GNRA or UNCG consensus sequence usually points to a characteristic structure thought to be canonical. However, evidence from recent tetraloop surveys indicates that a given sequence can adopt more than one conformation and, conversely, that sequences departing from GNRA or UNCG consensus can also adopt GNRA or UNCG-like loop structures (Bottaro and Lindorff-Larsen 2017; D'Ascenzo et al. 2017). Thus, naming a loop structure after its consensus sequence could be misleading in some cases. To rule out this situation, D'Ascenzo et al. (2017) suggest to name tetraloops after their characteristic turn, that is, U-turn or Z-turn loops, since those turns are characteristic of GNRA or UNCG tetraloops, respectively. The U-turn intervenes between the first and second residue of the tetraloop, while the Z-turn takes place between the third and fourth nucleotide. U-turn-based loops are generally locked by a base pair between the sugar and Hoogsteen edges of the first and fourth residues, respectively. Z-turn loops present a head-to-tail orientation of ribose rings from the third and fourth residues. The fourth residue generally presents a *syn* conformation reminiscent of Z-RNA (Hall et al. 1984; Davis et al. 1986), and donates its Watson–Crick edge to the sugar edge of the first residue. Few exceptions with an *anti* conformation of the fourth residue have been observed. It is only recently that rare but specific Z-turn loop receptors (only three examples up to now) have been identified by another survey of RNA structures (D'Ascenzo et al. 2018). This survey also points to a rare conformational change of a GNRA loop, which binds a Z-turn receptor. These results lead to the update of tetraloop semantics, which now breaks into U-turn and Z-turn loops with their own set of receptors.

UNCG tetraloops adopt the Z-turn conformation far more frequently than any other conformation, as it has been concluded from NMR studies (Allain and Varani 1995; Nozinovic et al. 2010). However in this study, we report rare cases where the conformation of Z-turn tetraloops (D'Ascenzo et al. 2017) is altered due to the steric hindrance applied by crystal packing, which forces the backbone of the second and third residues to take the place of the U1-G4 pair, expelling those residues out of the helix. This arrangement can be seen as "squashing" of Z-turn loops following local densification of macromolecules. The first observation arose from the different conformations adopted by the DP2 loop in two related crystal structures of the *Didymium iridis* lariat-capping ribozyme (LC) (Johansen and Vogt 1994; Meyer et al. 2014) differing in the length of the DP2 helix (Fig. 1). The second observation arose from the close inspection of two crystal structures of a chimeric group II ribozyme derived from the *Oceanobacillus iheyensis* intron. In this case, the intron ribozyme was crystallized both in the presence and in the absence of its 5′ exon substrate, which led to different crystal packing interactions (Costa et al. 2016). In support to our findings based on crystal packing variability, an inspection of RNA structures in the PDB also permitted to find various examples in ribosome structures of loops adopting a squashed conformation. Importantly, the observations made on ribosome structures indicate that the squashed conformations do not result from packing, but from biologically relevant highly packed environments such as those provided by the ribosomal subunits. Interestingly, RNA and/or protein contribute to interactions with the squashed loops. A squashed conformation is also observed in the tetraloop of the double-stranded RNA substrate from a yeast RNase III (Rnt1p) (Song et al. 2017). Altogether, our findings add a property to those of Z-turn loops, the conformation of which can adapt according to the degree of compaction of their immediate surroundings. In this conformational adaptation resulting from the increase of the local molecular density, the sugar-phosphate backbone expels nucleobases out of the loop, which become free to make stacking and H bond interactions with distal residues.

RESULTS

The circularly permutated (CP) form of the LC ribozyme presents a distorted UUCG loop

In vivo, the LC ribozyme (Fig. 1) catalyzes two reactions. The branching reaction results in the formation of a 3 nt lariat upon the nucleophilic attack of the phosphate group of C230 by the 2′ hydroxyl of U232. Due to the chemical reversibility of the transesterification reaction, the ligation is also observed, although the equilibrium is in favor of the branching reaction. In addition, in vitro only, hydrolysis at C230 is also noticeable (Nielsen et al. 2005). In vitro, the coexistence of three different reactions generates a variety of RNA products, which do not favor crystallization. In order to select a unique conformation and improve crystallization of the LC ribozyme, we engineered a circular

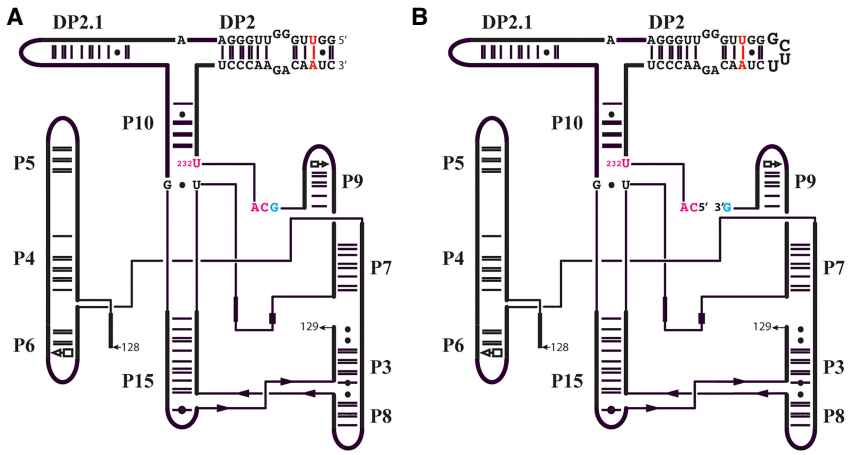

FIGURE 1. Secondary structure schemes of the circularly permutated (CP) LC ribozyme and CP DP2ΔUA derived from the *D. iridis* ribozyme. (*A*) The wild-type form of the LC ribozyme presents both the 5′ and 3′ ends in the DP2 stem. The 2′ hydroxyl group of the catalytic nucleotide, U232 cleaves the phosphate bond between the last nucleotide of the ribozyme G229 (cyan) and the downstream C230 (purple) leading to the formation of a 3 nt lariat. (*B*) In the CP form, the 5′ and 3′ ends reside at the scissile bond while DP2 has been circularized using a UUCG tetraloop. The U–A pair depleted to allow the loop to change conformation is shown in red.

permutation so that the natural 5′ and 3′ ends are tethered by the UUCG loop, and the nucleotides at the catalytic cleavage site, C230 and G229, become the new 5′ and 3′ ends, respectively. An optimized hammerhead construct was added upstream of the unfavorable transcriptional sequence 5′ CAU 3′ corresponding to the lariat sequence (Meyer and Masquida 2014). No interference of the engineered UUCG loop with the overall architecture was expected.

The crystal structure of the CP LC ribozyme (Meyer et al. 2014) shows that the DP2 UUCG loop adopts a squashed conformation where the backbone of residues 2 and 3 takes the place of the U–G pair formed between residues 1 and 4 characteristics of Z-turn loops, expelling nucleotides 1 and 4 into the solvent (Fig. 2A). Close inspection of the crystal packing indicates that the DP2 loop contacts two symmetry-related LC ribozyme molecules. One symmetry-related molecule (sym1) makes van der Waals contacts with residues 2 and 3 and the second symmetry-related molecule (sym2) provides stacking and H-bonding interactions to the expelled G4 (Fig. 3A,C).

Removal of one base pair from the DP2 stem restores the canonical UNCG loop structure

Following this observation, we engineered a CP LC ribozyme construct (CP LC DP2ΔUA) in which one base pair was removed from the DP2 stem (Fig. 1). This RNA crystallized under the same conditions and in identical space group $P2_12_12_1$ with each parameter expanded by <4% with respect to the parent CP LC ribozyme construct (Table 1; PDB: 6gyv). The crystal structure of the CP LC DP2ΔUA construct reveals that the canonical structure of the UUCG loop capping the DP2 stem has been restored (Fig. 2B) and displays the typical Z-turn loop structural characteristics reported in D'Ascenzo et al. (2017). It appears that the removal of one base pair from the DP2 stem restores enough space so that residues U1 and G4 from the loop can stack within the helix and form the *trans* base pair between their Watson–Crick and sugar edges (Leontis and Westhof 2001), respectively. Although the crystal packing remains unchanged, the conformational change of the loop results in the loss of packing interactions due to the different spatial orientation of the fourth residue of the loop. While sym1 still provides van der Waals distance contacts, the increased distance between G4 and sym2 does not allow anymore any type of interactions (Fig. 3C).

Crystal packing induces a conformational change on a UUCG loop engineered at a peripheral section of a group II intron RNA

Group II introns are large self-splicing RNAs and mobile genetic elements of bacterial origin. They are usually regarded as the ancestors of spliceosomal introns and the spliceosome in eukaryotes. Similarly to nuclear pre-mRNA splicing, group II self-splicing proceeds through two consecutive transesterification reactions and the intron is released in a typical branched conformation called the "lariat." The reversibility of the transesterification reactions allows the excised intron lariat to insert back into RNA or DNA targets through the reverse-splicing pathway. Reverse splicing into DNA is at the basis of the genomic mobility of group II introns. Recently, an engineered group II ribozyme derived from the *O. iheyensis* intron was used to obtain two crystal structures of a group II intron lariat primed for reverse splicing (Costa et al. 2016). The lariat form of this chimeric *O. iheyensis* intron was crystallized either alone or in the presence of a nonreactive 5′-exon analog RNA. Interestingly, each condition gave rise to a different crystal form having a distinct set of packing contacts. In the lariat-alone crystal structure, one of the packing interactions involves a UUCG tetraloop engineered at the tip of intron domain III. Remarkably, this crystal contact forces the UUCG loop to adopt a squashed conformation similar to the one observed in the DP2 UUCG loop from the CP LC ribozyme (Fig. 4A). Conversely, however, in

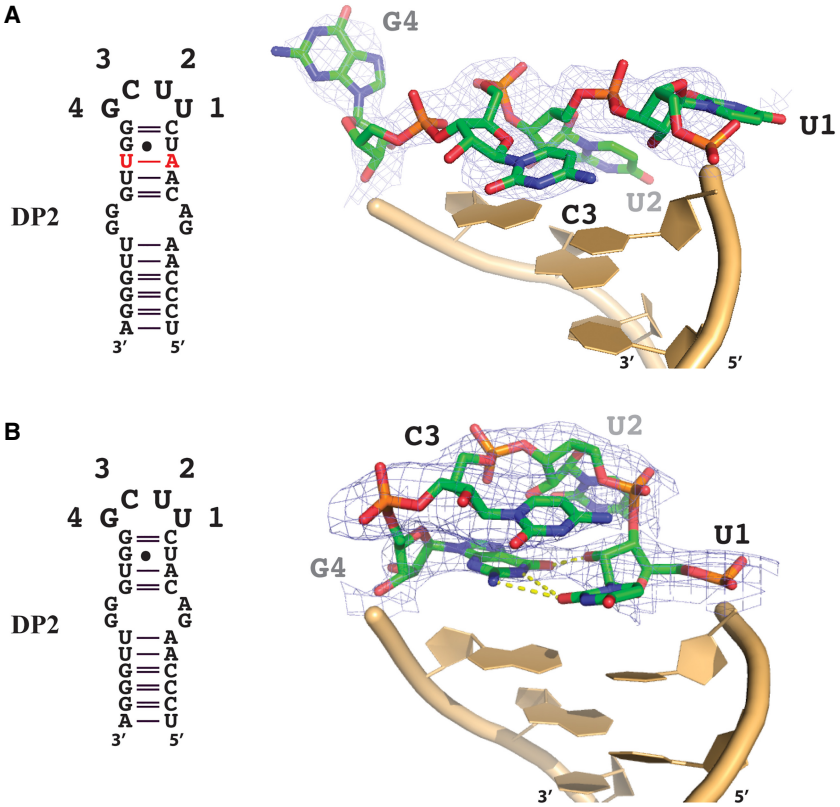

FIGURE 2. Secondary structures and conformations of the DP2 stem–loops. (*A*) The UUCG tetraloop of the CP LC ribozyme presents a squashed conformation (PDB: 6gyv). Squashing results from the proximity between the backbone of residues U2 and C3 and the last base pair of the stem (nucleotides with filled rings), which expels the residues from the loop toward the solvent. The helical conformation is interrupted at the level of U1 and restored only after G4. (*B*) The loop from the CP DP2ΔUA ribozyme adopts the canonical UUCG conformation with a maintained helical continuity at U1 and G4 (PDB: 6g7z). The Z-turn allows reorientation of the RNA chain toward the second strand of the stem. On each panel, the tetraloop is surrounded by a maximum likelihood map $2mF_{obs} - DF_{calc}$ contoured at 1.5 σ.

the context of the 5′-exon-bound lariat structure, the UUCG loop of domain III points toward the solvent and is seen to adopt the canonical conformation expected for the Z-turn tetraloop family (Fig. 4B). Noteworthy, the crystal packing variations do not influence the canonical conformation of the other UUCG loop engineered in intron domain IV.

Occurrences of "squashed" loop conformations within other RNAs

To find biologically relevant examples of our observations, we then investigated whether other squashed tetraloops of any sequence could be found in the PDB. We took advantage of the complete and recent data set gathered by Bottaro and Lindorff-Larsen (2017). In this study, a clustering approach identifies all the structurally distinct RNA tetraloops. Among the 44 clusters reported, four gather loops with structures based on Z-turns typical of UNCG loops (clusters 2, 5, 37, and 44). Three clusters gather conformations close to the squashed loop observed in the LC ribozyme crystal structure (clusters 16, 19, and 41). The contents of the clusters encompassing the Z-turn and the squashed loops are biased at two levels. First, redundancy of crystal or cryo-EM structures of ribosomes contributes most of the yet independent observations. Second, NMR structures usually bring >10 models, which cannot be considered as independent since they result from calculations obtained from a given set of restraints. Consequently, after pruning NMR structures, the final data set contains 1504 Z-turn and 380 squashed loops, respectively. Squashed loops represent ~25% of the Z-turn loops, pointing to the significance of this RNA motif.

Considering squashed loops, the first example extracted from cluster 41 with a UACG loop sequence obeys the UNCG consensus (Fig. 5 in green). This loop belongs to the 1450 region of the *Thermus thermophilus* 16S rRNA, located at the periphery of the 70S ribosome (Maehigashi et al. 2014; Rozov et al. 2015, 2016a,b). The main structural features of this loop are due to its first (U1450) and last (G1453) residues, which bulge out. G1453 interacts with ribosomal protein RPS20. Despite its peripheral location, this loop is not involved in packing. Strikingly, it adopts a canonical UNCG structure in the original crystal structure of the 30S particle alone (1fjg, Wimberly et al. 2000), providing an additional potential biological role for the switch between canonical and squashed conformations. Interestingly, a 3 nt bulge is located exactly 3 bp away from the 1450 loop. This bulge bridges the 23S rRNA (2850 region) and also interacts with RPL19. Inspection of PDB files point to a squashed loop structure during elongation, while a Z-turn loop is preferred during other phases of translation, including initiation, termination, and stress-dependent ribosome stalling. These observations are nonetheless restricted to the *T. thermophilus* ribosome, and require additional data to be confirmed.

The NGNN tetraloop from the double-stranded RNA substrate of the yeast Rnt1p (RNase III family homologous to human drosha and dicer) also adopts the squashed conformation from cluster 41 in the crystal structure of the complex (Liang et al. 2014; Song et al. 2017). This tetraloop (AGUC in pdb file 1k6g) adopts a regular

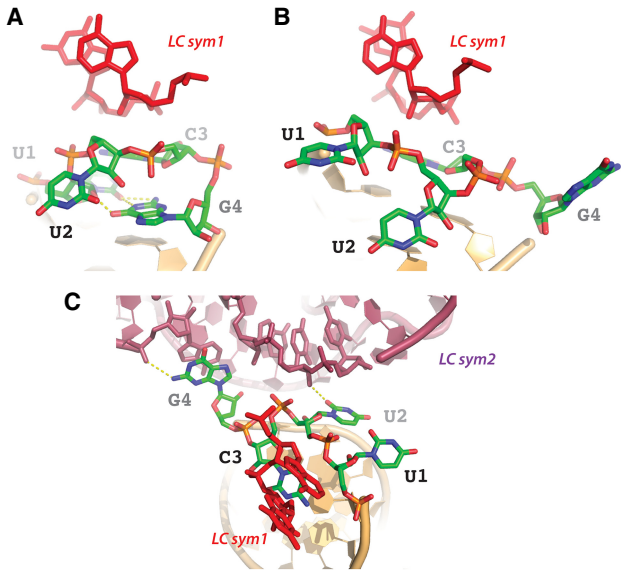

FIGURE 3. Contacts between the two conformations of the DP2 stem–loops of CP LC ribozyme and the symmetry-related molecules. (*A*) When the loop adopts a canonical conformation, the backbone of U2 and C3 is in close contact with the sugar edge of two residues embedded in an A-form helix of a symmetry-related ribozyme (LC sym1). (*B*) Squashing of the loop results in spreading its residues that can contact residues at a much higher distance. LC sym1 remains contacted by residues U2 and C3. (*C*) However, G4 now stacks and H-bonds (N2 group of G4) with an adenine (Phosphate group of A115) from a second symmetry-related molecule (LC sym2) already involved in A-minor interactions (Doherty et al. 2001).

conformation in the free form of the RNA (Lebars et al. 2001), in the sense that the first and fourth residues from the loop form a base pair, which stacks upon the last base pair of the stem. Nevertheless, the kink in the backbone is mediated at the phosphate group between residues 2 and 3, a situation neither typical of U-turn nor Z-turn loops. Mapping the NMR distance constraints from Lebars et al. (2001) onto a Z-turn loop points to only two which cannot be accommodated between the 2′ and 3′ hydrogen atoms from residue L4 and the H6 atom from L3. Since the L4 residue is a purine in the Z-turn loop and not a pyrimidine as in the 1k6g NMR structure, these two constraints may be specific from this sequence. The torsional angles around the phosphate group (α and ζ) remained unconstrained, indicating that most of the torsion angles result from the distance constraints. Moreover, this loop conformation is not represented by any cluster from Bottaro and Lindorff-Larsen (2017), indicating that it does not represent a characteristic fold.

Loops from clusters 16 and 19 also adopt squashed conformations, even though their sequences depart from the UNCG consensus (Table 2; Fig. 5 in cyan and purple, respectively). Cluster 16 characterizes GGAU loops in the 23S rRNA from *T. thermophilus*, and cluster 19, CGAA, UGAG and UUAG loops in the same region of the 23S rRNA from three different organisms, *Haloarcula Marismortui* (Ban et al. 2000), *Deinococcus radiodurans* (Schlüenzen et al. 2001), and *T. thermophilus* (Maehigashi et al. 2014; Rozov et al. 2015). In the case of cluster 16, the loop is situated at the solvent interface of the rRNA, close to ribosomal protein L28. The expelled first residue (G2210) interacts by stacking with U1493 (*T. thermophilus* numbering, pdb 1vvj) and hydrogen-bonds with a non-bridging atom of the phosphate group from G1492. The fourth residue of the loop, also ejected into the solvent, interacts with an arginine residue (R52) from L28. No crystal packing interaction is observed, but the RNA is not fully modeled in this region opening the possibility that the conformation of the loop may be restrained by unobserved contacts.

In contrast, the loops from cluster 19 are well buried in the 50S subunit and interact in the same way in the three considered organisms. The first and fourth residues of the UUAG loop stack on A2430 (*T. thermophilus* numbering) and G2448, respectively. In addition, the WC edge from the fourth residue interacts with the sugar edge of G2445. The second and third residues make shallow groove contacts with nucleotides 2246 to 2248. The very same arrangement is observed for the other loops CGAA and UGAG. Only RNA contacts are observed in this case, reminding of the different situations observed in the crystal packing that drove our study.

The structural characteristics of these loops can thus be summarized as follows. L1 and L4 are bulged out with a strong clustering for the latter. L2 and L3 are mostly located on the shallow groove and deep groove sides, respectively. The conformations for L3 residues are more

TABLE 1. Crystallographic data for CP LC ribozyme DP2ΔUA

PDB Id	6g7z
Resolution range	41.17 – 3.337 (3.456 – 3.337)
Space group	P 2_1 2_1 2_1
Unit cell	59.945 88.791 110.038 90 90 90
Unique reflections	8338 (299)
Completeness (%)	92.64 (33.98)
Wilson B-factor	88.10
Reflections used in refinement	8336 (299)
Reflections used for R-free	417 (15)
R-work	0.2351 (0.3748)
R-free	0.2933 (0.5942)
Number of nonhydrogen atoms	3995
Macromolecules	3970
Ligands	25
RMS (bonds)	0.002
RMS (angles)	0.55
Clashscore	8.91

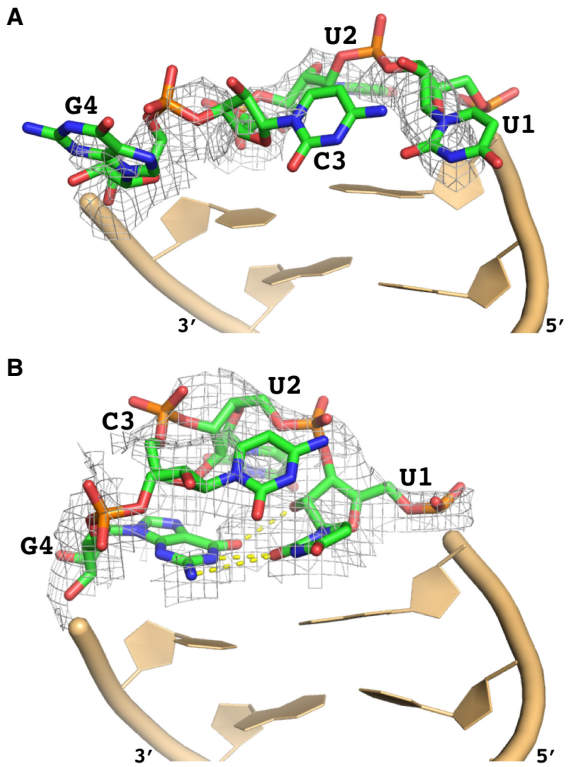

FIGURE 4. Structural flexibility of a UUCG tetraloop present in an engineered group II intron ribozyme. (*A*) The UUCG tetraloop at the tip of intron domain III is involved in a crystal packing contact and adopts a squashed conformation in which residues U1 and G4 are ejected into the solvent (PDB entry: 5j01). (*B*) The same UUCG tetraloop acquires the canonical conformation in a different crystal form (corresponding to the group II intron ribozyme bound to its 5′-exon) in which the loop is not involved in any interaction (PDB entry: 5j02). On each panel, the tetraloop is surrounded by a maximum likelihood map $2mF_{obs} - DF_{calc}$ contoured at 1.5 σ.

constrained than for L2 residues. L4 residues are often engaged in tertiary interactions with neighboring residues, which defines a set of clustered conformations.

DISCUSSION

It is usually accepted that UNCG tetraloops adopt a unique fold with characteristics accurately described elsewhere (D'Ascenzo et al. 2017), in brief, a Z-turn favoring the formation of a *trans* sugar edge-Watson–Crick pair between residues U1 and G4 of the tetraloop, provided that the G residue adopts a *syn* conformation. This conformation of the loop also presents a head-to-tail orientation of the third and fourth ribosomes reminiscent of Z-RNA (Hall et al. 1984; Davis et al. 1986). In another study, Bottaro and Lindorff-Larsen (2017) have shown that tetraloops with sequences distinct from UNCG could also adopt the canonical UNCG loop geometry and that, on rare occasions, UNCG sequences could acquire other conformations. A UUAG loop sequence compatible with the presence of a Z-turn between residues 3 and 4 has also been reported (D'Ascenzo et al. 2018), indicating that under relaxed constraints, loops UGAG and UUAG from cluster 19 may adopt a Z-turn-based conformation.

In the present study, we report observations from crystal structures of two different ribozymes, where a given UUCG loop adopts two distinct conformations according to the crystal packing variations resulting from slight structural changes of the RNA in the asymmetric unit. In each case, the UUCG loops were engineered in order to stabilize the underlying stem and thus, facilitate crystallization of the RNA molecule. Although the aim of these strategies was not to promote inter-molecular interactions, the engineered UUCG loops become involved in crystal packing contacts, which resulted in squashing these loops. The conformations induced by these interactions are very similar, and reveal that the squashed conformation is specifically obtained in response to the backbone–backbone interaction through contacts between phosphate and 2′ hydroxyl groups. In the two occurrences of this interaction observed in our crystal structures, a helix of a symmetry-related molecule contacts the tip of the loop formed by the second and third residues (Fig. 3). However, in the crystal structure of the group II ribozyme, the guanine (L4) of the squashed loop does not interact with a symmetry-related RNA like it does in the CP LC ribozyme crystal structure. This indicates that G-mediated tertiary interaction is not a prerequisite to the folding of the squashed conformation.

The relevance of the distorted UUCG loop conformation we observe is supported by the behavior of specific loops in ribosomal RNAs. Interestingly, the UACG loop from cluster 41 in the study of Bottaro and Lindorff-Larsen (2017) adopts the squashed conformation only in the context of the 70S ribosome since the same loop adopts a canonical conformation in the original structure of the 30S ribosomal subunit (Wimberly et al. 2000). Subunit association is well known to induce RNA conformational changes, suggesting that the one we describe in the context of the ribosome may have a biological significance. Inspection of the other clusters reveals that loops with other sequences can indeed adopt conformations very much related to the squashed loop from the LC ribozyme. The L2 nucleotide presents the most variable position. Residues at positions 3 and 4 occupy very identical positions in spite of belonging to different clusters (Table 3). It appears that the centroid approach developed by Bottaro and Lindorff-Larsen may be too sensitive, and that some clusters could actually be merged under more general structural features.

The resolution of the structures can also be questioned. Usually, structural data mining is performed on structures with resolution better than 2.5–3 Å. In this case, we also took into account structures with resolution worse than 3 Å. Although the positions of loop residues may result from the geometric restraints applied on the models during refinement, the generally crowded surroundings of

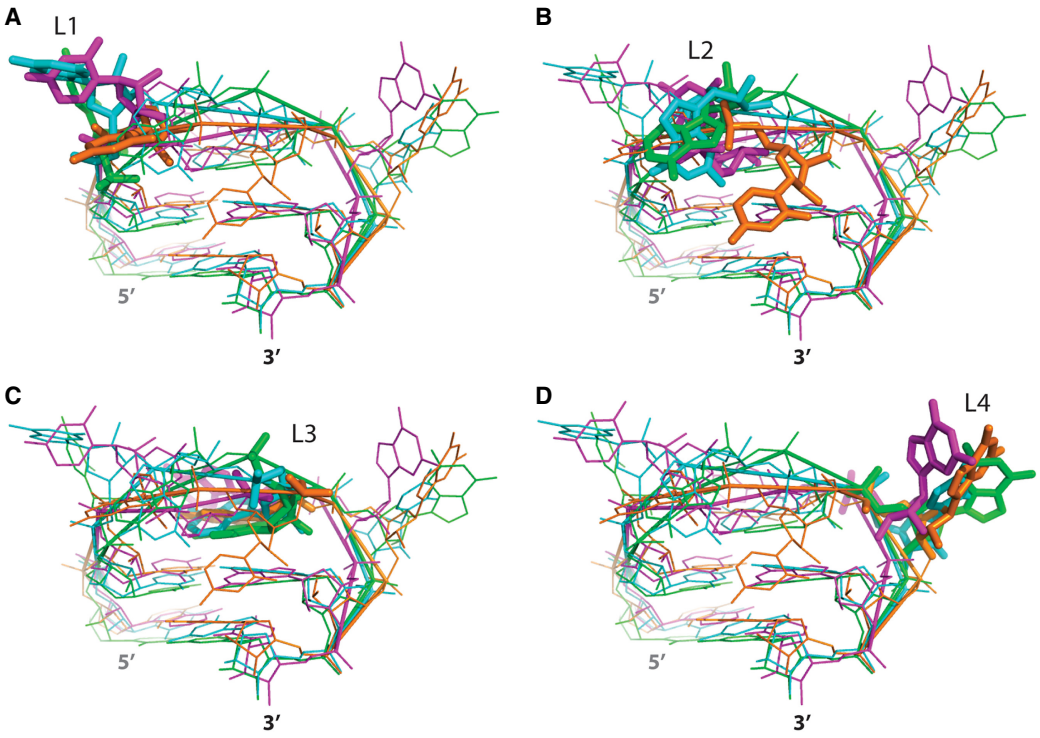

FIGURE 5. Comparison between the conformations of the squashed loops from clusters 41 (green, PDB: 4v6f), 16 (cyan, PDB: 1vvj), and 19 (purple, PDB: 1j5a) from the Bottaro and Lindorff-Larsen study (Bottaro and Lindorff-Larsen 2017) and the squashed loop from the LC ribozyme (orange, PDB: 6gyv). Superimposition of the different types of loops seen from the shallow groove side shows that nucleotides cluster individually in spite of the large differences observed for the torsional angles. Emphasis is given for nucleotides from the loops at first (*A*, L1), second (*B*, L2), third (*C*, L3), and fourth position (*D*, L4).

TABLE 2. Contacts between squashed loops in ribosome structures

PDB	Location	RNA contacts	Protein contacts
C41: UACG[a]			
4v8b UACG	Loop 1448 (16S)	L2 Hbond with 16S-G147GA149	RPS20. No packing interaction
C16: GGAU (23S)[b]			
1vvj GGAU	G2210GAU2214	G1 stack with U1493	Shallow groove contact W70 of RPL2 loop close to RPL28
C19: CGAA, UGAG, UUAG (23S)			
1ffk CGAA	C920GAA923 (well buried into the structure)	L2–L3 interact with RNA backbone (2279–2280) L1 stack with A2467 Adenine L4 WC edge with S edge of G2480	None
1j5a UGAG	U840GAG843 (23S) (well buried into the structure) Same position regardless of nucleotide numbering	L2–L3 interact with RNA backbone (2225–2227) L1 stacks with A2409 Guanine L4 wc edge with s edge of G2424 (opposite strand from Adenine L4 from CGAA loop)	None
1vvj UUAG	U827UAG830 (23S) Same location as CGAA and UGAG (Same file as for C16)	L2–L3 interact with RNA backbone (2246–2248) L1 stacks on A2430 L4 stacks on A2448 Guanine L4 wc edge with s edge of G2445 (same strand as for CGAA)	None

[a]No deformation of the loop in the original *T. thermophilus* crystal structure of 30S subunit (4kvb).
[b]No crystal packing interactions.

TABLE 3. Root-mean-square deviations (Å) between individual residues in squashed loops using the squashed loop from the LC ribozyme as a reference

	L1	L2	L3	L4
4v8b UACG	10.06	7.51	2.25	5.01
1vvj UUAG	6.30	6.35	2.72	3.68
1j5a UGAG	6.44	10.24	4.67	3.61
5j01 Group II UUCG	12.27	6.32	4.29	10.67
Average	8.77	7.60	3.48	5.74

the squashed loops indicate that these conformations more likely result from molecular adaptation. Consequently, squashed loops gather similar structural features, which are (i) residues L1 and L4 bulging out, (ii) the backbone from residues L2 and L3 oriented right above the last helical base pairs, and (iii) the base from residue L3 pointing toward the deep groove side. These features allow residues 1 and 4 to interact with neighboring molecules, either RNAs or proteins, and the backbone of residues 2 and 3 to interact with the shallow groove of helices in the vicinity. Most importantly, the squashed conformation is reversible at least in four occurrences, including one in the ribosome, and one in the dsRNA substrate from Rnt1p.

In summary, our study presents for the first time evidence that loops belonging to the UNCG family can adopt more than one conformation according to the variations of their structural context. Considering that RNA molecules are highly dynamic and that they fulfill their biological functions through complex folding pathways and structural rearrangements, conformational UNCG loop flexibility may play an important role in RNA biology by allowing them to fulfill particular functional needs through their engagement in specific tertiary interactions with neighboring residues. The fact that the loops from cluster 19 are embedded in the same kind of structural environment than the crystal packing from the LC and group II ribozymes corroborates our conclusions.

Interestingly, a parallel can also be drawn between UNCG tetraloops and those belonging to the GANC consensus, which specifically caps the catalytic domain V hairpin in subgroup IIC self-splicing introns. Comparison of the crystal structures of the chimeric *O. iheyensis* intron revealed that the GANC tetraloops are flexible and adopt two different conformations according to the presence or absence of the 5′-exon substrate (Costa et al. 2016). In the presence of the 5′-exon, the GAAC loop of domain V acquires an alternate conformation that allows it to contact a specific intron receptor sequence through a single base stack interaction. Thus, in this case, the flexibility of the GAAC loop plays a crucial biological role by participating in the proper folding of the catalytic center of the intron. It is also interesting to note that this GAAC-receptor interaction is for the moment restricted to group IIC introns since it is specifically adapted to function in the particular structural context of domain V found in these introns.

Finally, the recent description of Z-turn specific receptors (D'Ascenzo et al. 2018) also supports the idea that the different conformations of tetraloops of a given sequence may have distinct receptors. The squashed tetraloops can still interact with macromolecules in the vicinity, especially with the shallow groove of RNA helices, while expelled nucleobases 1 and 4 can stack or H bond with residues in the vicinity.

The present work can thus be seen as the description of a new kind of adaptive tetraloop interaction for which more occurrences may be found in RNA structures that will be solved in the future.

MATERIALS AND METHODS

In vitro transcription, crystallization, and structure resolution of the CP LC DP2ΔUA construct

The CP LC DP2ΔUA was cloned, transcribed and purified as described in Beckert and Masquida (2011) and Meyer and Masquida (2014, 2016). The purified RNA was crystallized under conditions obtained from the crystallization of the CP LC ribozyme (Meyer et al. 2014). One volume of 100 µM of RNA was mixed to one volume of crystallization solution containing 200 mM NaCl, 100 mM HEPES pH 7.5 and 5%–25% PEG 3350. The structure was solved by the molecular replacement method and refined using Refmac (Murshudov et al. 2011) and Phenix (Table 1; Adams et al. 2010).

Data analysis

Representative loops from the various clusters presented in the Bottaro and Lindorff-Larsen study (Bottaro and Lindorff-Larsen 2017) were superimposed to the squashed loop of the LC ribozyme to pick up clusters with similar folds. Torsion angles were determined with x3DNA-DSSR (Lu and Olson 2003). Figures, superimpositions and *RMSD* calculations were made in PyMol (Schrodinger 2010).

ACKNOWLEDGMENTS

B.M. is a USIAS recipient for the exchange program between University of Strasbourg (France) and Sophia University (Tokyo, Japan). This research is supported by the Centre National de la Recherche Scientifique (LABEX ANR-11-LABX-0057_MITOCROSS), the University of Strasbourg. M.C. acknowledges the French agency ANR (grant ANR-10-BLAN-1502) and the BIG Lidex program for funding.

Received July 22, 2019; accepted August 1, 2019.

REFERENCES

Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, et al. 2010. *PHENIX*: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66:** 213–221. doi:10.1107/S0907444909052925

Allain FH, Varani G. 1995. Structure of the P1 helix from group I self-splicing introns. *J Mol Biol* **250:** 333–353. doi:10.1006/jmbi.1995.0381

Antao VP, Tinoco IJ. 1992. Thermodynamic parameters for loop formation in RNA and DNA hairpin tetraloops. *Nucleic Acids Res* **20:** 819–824. doi:10.1093/nar/20.4.819

Ban N, Nissen P, Hansen J, Moore PB, Steitz TA. 2000. The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science* **289:** 905–920. doi:10.1126/science.289.5481.905

Barrick JE, Takahashi TT, Ren J, Xia T, Roberts RW. 2001. Large libraries reveal diverse solutions to an RNA recognition problem. *Proc Natl Acad Sci* **98:** 12374–12378. doi:10.1073/pnas.221467798

Beckert B, Masquida B. 2011. Synthesis of RNA by in vitro transcription. *Methods Mol Biol* **703:** 29–41. doi:10.1007/978-1-59745-248-9_3

Bottaro S, Lindorff-Larsen K. 2017. Mapping the universe of RNA tetraloop folds. *Biophys J* **113:** 257–267. doi:10.1016/j.bpj.2017.06.011

CateJH, Gooding AR, Podell E, Zhou K, Golden BL, Kundrot CE, Cech TR, Doudna JA. 1996. Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science* **273:** 1678–1685. doi:10.1126/science.273.5282.1678

Costa M, Michel F. 1997. Rules for RNA recognition of GNRA tetraloops deduced by *in vitro* selection: comparison with *in vivo* evolution. *EMBO J* **16:** 3289–3302. doi:10.1093/emboj/16.11.3289

Costa M, Walbott H, Monachello D, Westhof E, Michel F. 2016. Crystal structures of a group II intron lariat primed for reverse splicing. *Science* **354:** aaf9258. doi:10.1126/science.aaf9258

D'Ascenzo L, Leonarski F, Vicens Q, Auffinger P. 2017. Revisiting GNRA and UNCG folds: U-turns versus Z-turns in RNA hairpin loops. *RNA* **23:** 259–269. doi:10.1261/rna.059097.116

D'Ascenzo L, Vicens Q, Auffinger P. 2018. Identification of receptors for UNCG and GNRA Z-turns and their occurrence in rRNA. *Nucleic Acids Res* **46:** 7989–7997. doi:10.1093/nar/gky578

Davis PW, Hall K, Cruz P, Tinoco I Jr., Neilson T. 1986. The tetraribonucleotide rCpGpCpG forms a left-handed Z-RNA double-helix. *Nucleic Acids Res* **14:** 1279–1291. doi:10.1093/nar/14.3.1279

Doherty EA, Batey RT, Masquida B, Doudna JA. 2001. A universal mode of helix packing in RNA. *Nat Struct Biol* **8:** 339–343. doi:10.1038/86221

Ennifar E, Nikulin A, Tishchenko S, Serganov A, Nevskaya N, Garber M, Ehresmann B, Ehresmann C, Nikonov S, Dumas P. 2000. The crystal structure of UUCG tetraloop. *J Mol Biol* **304:** 35–42. doi:10.1006/jmbi.2000.4204

Fiore JL, Nesbitt DJ. 2013. An RNA folding motif: GNRA tetraloop-receptor interactions. *Q Rev Biophys* **46:** 223–264. doi:10.1017/S0033583513000048

Geary C, Baudrey S, Jaeger L. 2008. Comprehensive features of natural and in vitro selected GNRA tetraloop-binding receptors. *Nucleic Acids Res* **36:** 1138–1152. doi:10.1093/nar/gkm1048

Hall K, Cruz P, Tinoco I Jr., Jovin TM, van de Sande JH. 1984. 'Z-RNA'—a left-handed RNA double helix. *Nature* **311:** 584–586. doi:10.1038/311584a0

Johansen S, Vogt VM. 1994. An intron in the nuclear ribosomal DNA of *Didymium iridis* codes for a group I ribozyme and a novel ribozyme that cooperate in self-splicing. *Cell* **76:** 725–734. doi:10.1016/0092-8674(94)90511-8

Koculi E, Cho SS, Desai R, Thirumalai D, Woodson SA. 2012. Folding path of P5abc RNA involves direct coupling of secondary and tertiary structures. *Nucleic Acids Res* **40:** 8011–8020. doi:10.1093/nar/gks468

Lebars I, Lamontagne B, Yoshizawa S, Aboul-Elela S, Fourmy D. 2001. Solution structure of conserved AGNN tetraloops: insights into Rnt1p RNA processing. *EMBO J* **20:** 7250–7258. doi:10.1093/emboj/20.24.7250

Leontis NB, Westhof E. 2001. Geometric nomenclature and classification of RNA base pairs. *RNA* **7:** 499–512. doi:10.1017/S1355838201002515

Liang YH, Lavoie M, Comeau MA, Abou Elela S, Ji X. 2014. Structure of a eukaryotic RNase III postcleavage complex reveals a double-ruler mechanism for substrate selection. *Mol Cell* **54:** 431–444. doi:10.1016/j.molcel.2014.03.006

Lu XJ, Olson WK. 2003. 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res* **31:** 5108–5121. doi:10.1093/nar/gkg680

Maehigashi T, Dunkle JA, Miles SJ, Dunham CM. 2014. Structural insights into +1 frameshifting promoted by expanded or modification-deficient anticodon stem loops. *Proc Natl Acad Sci* **111:** 12740–12745. doi:10.1073/pnas.1409436111

Masquida B, Westhof E. 1999. Crystallographic structures of RNA oligonucleotides and ribozymes. In *Oxford handbook of nucleic acid structures* (ed. Neidle S), pp. 533–565. Oxford University Press, Oxford, UK.

Masquida B, Beckert B, Jossinet F. 2010. Exploring RNA structure by integrative molecular modelling. *N Biotechnol* **27:** 170–183. doi:10.1016/j.nbt.2010.02.022

Meyer M, Masquida B. 2014. *Cis*-Acting 5′ hammerhead ribozyme optimization for in vitro transcription of highly structured RNAs. *Methods Mol Biol* **1086:** 21–40. doi:10.1007/978-1-62703-667-2_2

Meyer M, Masquida B. 2016. Polyacrylamide gel electrophoresis for purification of large amounts of RNA. *Methods Mol Biol* **1320:** 59–65. doi:10.1007/978-1-4939-2763-0_5

Meyer M, Nielsen H, Olieric V, Roblin P, Johansen SD, Westhof E, Masquida B. 2014. Speciation of a group I intron into a lariat capping ribozyme. *Proc Natl Acad Sci* **111:** 7659–7664. doi:10.1073/pnas.1322248111

Michel F, Westhof E. 1990. Modelling of the three-dimensional architecture of group-I catalytic introns based on comparative sequence analysis. *J Mol Biol* **216:** 585–610. doi:10.1016/0022-2836(90)90386-Z

Murshudov GN, Skubák P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, Winn MD, Long F, Vagin AA. 2011. *REFMAC5* for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr* **67:** 355–367. doi:10.1107/S0907444911001314

Nielsen H, Westhof E, Johansen S. 2005. An mRNA is capped by a 2′, 5′ lariat catalyzed by a group I-like ribozyme. *Science* **309:** 1584–1587. doi:10.1126/science.1113645

Nissen P, Ippolito JA, Ban N, Moore PB, Steitz TA. 2001. RNA tertiary interactions in the large ribosomal subunit: the A-minor motif. *Proc Natl Acad Sci* **98:** 4899–4903. doi:10.1073/pnas.081082398

Nozinovic S, Fürtig B, Jonker HR, Richter C, Schwalbe H. 2010. High-resolution NMR structure of an RNA model system: the 14-mer cUUCGg tetraloop hairpin RNA. *Nucleic Acids Res* **38:** 683–694. doi:10.1093/nar/gkp956

Pley HW, Flaherty KM, McKay DB. 1994. Model for an RNA tertiary interaction from the structure of an intermolecular complex between a GAAA tetraloop and an RNA helix. *Nature* **372:** 111–113. doi:10.1038/372111a0

Robertson ME, Seamons RA, Belsham GJ. 1999. A selection system for functional internal ribosome entry site (IRES) elements: analysis of the requirement for a conserved GNRA tetraloop in the

encephalomyocarditis virus IRES. *RNA* **5:** 1167–1179. doi:10.1017/S1355838299990301

Rozov A, Demeshkina N, Westhof E, Yusupov M, Yusupova G. 2015. Structural insights into the translational infidelity mechanism. *Nat Commun* **6:** 7251. doi:10.1038/ncomms8251

Rozov A, Demeshkina N, Khusainov I, Westhof E, Yusupov M, Yusupova G. 2016a. Novel base-pairing interactions at the tRNA wobble position crucial for accurate reading of the genetic code. *Nat Commun* **7:** 10457. doi:10.1038/ncomms10457

Rozov A, Westhof E, Yusupov M, Yusupova G. 2016b. The ribosome prohibits the G*U wobble geometry at the first position of the codon–anticodon helix. *Nucleic Acids Res* **44:** 6434–6441. doi:10.1093/nar/gkw431

Schlüenzen F, Zarivach R, Harms J, Bashan A, Tocilj A, Albrecht R, Yonath A, Franceschi F. 2001. Structural basis for the interaction of antibiotics with the peptidyl transferase centre in eubacteria. *Nature* **413:** 814–821. doi:10.1038/35101544

Schrodinger LLC. 2010. *The PyMOL Molecular Graphics System, Version 1.3r1*.

Song H, Fang X, Jin L, Shaw GX, Wang YX, Ji X. 2017. The functional cycle of Rnt1p: five consecutive steps of double-stranded RNA processing by a eukaryotic RNase III. *Structure* **25:** 353–363. doi:10.1016/j.str.2016.12.013

Varani G, Cheong C, Tinoco I Jr. 1991. Structure of an unusually stable RNA hairpin. *Biochemistry* **30:** 3280–3289. doi:10.1021/bi00227a016

Westhof E, Masquida B, Jossinet F. 2011. Predicting and modeling RNA architecture. *Cold Spring Harb Perspect Biol* **3:** a003632. doi:10.1101/cshperspect.a003632

Wimberly BT, Brodersen DE, Clemons WM Jr., Morgan-Warren RJ, Carter AP, Vonrhein C, Hartsch T, Ramakrishnan V. 2000. Structure of the 30S ribosomal subunit. *Nature* **407:** 327–339. doi:10.1038/35030006

Woodson SA. 2010. Compact intermediates in RNA folding. *Annu Rev Biophys* **39:** 61–77. doi:10.1146/annurev.biophys.093008.131334

Conformational adaptation of UNCG loops upon crowding

Mélanie Meyer, Hélène Walbott, Vincent Oliéric, et al.

RNA 2019 25: 1522-1531 originally published online August 19, 2019
Access the most recent version at doi:10.1261/rna.072694.119