Revisiting GNRA and UNCG folds: U-turns versus Z-turns in RNA hairpin loops

LUIGI D’ASCENZO,1 FILIP LEONARSKI,1,2 QUENTIN VICENS,1 and PASCAL AUFFINGER1
1Université de Strasbourg, CNRS, Architecture et Réactivité de l’ARN, UPR 9002, F-67000 Strasbourg, France
2Faculty of Chemistry, University of Warsaw, 02-093 Warsaw, Poland

ABSTRACT
When thinking about RNA three-dimensional structures, coming across GNRA and UNCG tetraloops is perceived as a boon since their folds have been extensively described. Nevertheless, analyzing loop conformations within RNA and RNP structures led us to uncover several instances of GNRA and UNCG loops that do not fold as expected. We noticed that when a GNRA does not assume its “natural” fold, it adopts the one we typically associate with a UNCG sequence. The same folding interconversion may occur for loops with UNCG sequences, for instance within tRNA anticodon loops. Hence, we show that some structured tetranucleotide sequences starting with G or U can adopt either of these folds. The underlying structural basis that defines these two fold types is the mutually exclusive stacking of a backbone oxygen on either the first (in GNRA) or the last nucleobase (in UNCG), generating an oxygen–π contact. We thereby propose to refrain from using sequences to distinguish between loop conformations. Instead, we suggest using descriptors such as U-turn (for “GNRA-type” folds) and a newly described Z-turn (for “UNCG-type” folds). Because tetraloops adopt for the largest part only two (inter)convertible turns, we are better able to interpret from a structural perspective loop interchangeability occurring in ribosomes and viral RNA. In this respect, we propose a general view on the inclination for a given sequence to adopt (or not) a specific fold. We also suggest how long-noncoding RNAs may adopt discrete but transient structures, which are therefore hard to predict.

Keywords: RNA folding; RNA motif; structure prediction; tetraloop; tRNA anticodon

INTRODUCTION
RNA architecture is modular and hierarchical, which implies that secondary structural elements such as double stranded helices, hairpins, and single-stranded loops are linked by tertiary interactions that guide the assembly process (Hendrix et al. 2005; Cruz and Westhof 2009; Butcher and Pyle 2011). The majority of hairpin stems are capped by GNRA or UNCG tetranucleotide sequences—where N is any base and R is a purine (Cheong et al. 2015; Hall 2015). These tetranucleotide loops adopt distinctive folds that involve extensive and well-described networks of hydrogen bonds and stacking interactions (Cheong et al. 1990; Heus and Pardi 1991; Allain and Varani 1995; Jucker and Pardi 1995a; Jucker et al. 1996; Ennifar et al. 2000; Correll and Swinger 2003; Nozinovic et al. 2010). For GNRA and UNCG loops, it is generally assumed that the sequence commands a unique fold. Hence, upon considering sequence alignments and secondary structures of RNA families for which no 3D structures are available, we presume that we understand how these tetraloops fold.

Here, we present structural evidence that challenges these expectations by identifying GNRA sequences that adopt a UNCG fold and vice versa, both in tetraloops closed by a Watson–Crick base pair and in tetraloop-like motifs embedded in larger ribosomal and tRNA loops (Auffinger and Westhof 2001). Although this loop dimorphism remains rare within the pool of RNAs for which we currently possess 3D data, it led us to question some basic assumptions we make about RNA folding and structure prediction.

To better characterize these interconversions, we propose a more general structure-based tetraloop and tetraloop-like identification scheme that involves on one side the classical and well-described U-turn (Gutell et al. 2000) and, on the other, a newly defined “Z-turn,” which is based on the UNCG tetraloop fold and the Z-RNA CpG step it encompasses (D’Ascenzo et al. 2016). We establish that these two turns and variants thereof are key to the tetraloop and tetra-loop-like folding landscape, but also to most turns in RNAs. A typical and infrequent tetranucleotide fold that does not conform to these rules will be described in more detail elsewhere. Here, before pursuing, we need first to (re)define

© 2017 D’Ascenzo et al. This article, published in RNA, is available under a Creative Commons License (Attribution 4.0 International), as described at http://creativecommons.org/licenses/by/4.0/.
U-turns and Z-turns as they appear in structured tetranucleotide folds within hairpins (see also Materials and Methods).

**U-turn and U_{SH}-turn signatures**

A U-turn is a tetranucleotide motif that was first identified in tRNA anticodon and T-loops (Quigley and Rich 1976; Gutell et al. 2000; Auffinger and Westhof 2001; Klosterman et al. 2004) and has since been characterized in a large variety of structural motifs starting with a uridine or a pseudouridine. In that respect, U-turns were sometimes called uridine-turns or π-turns (Kim and Sussman 1976; Jucker and Pardi 1995a).

U-turns were also associated with “G-starting” motifs such as GNRA tetraloops (Fig. 1A), or more recently in tetranucleotide motifs involving a protonated cytosine like a uC’UAu loop (Gottstein-Schmidtke et al. 2014). In short, a U-turn involves a hydrogen bond between the first nucleobase— with a U/G/C’ imino or amino nitrogen atom—and an OP atom of the fourth nucleotide. This base—phosphate hydrogen bond is of the “5/4/3BPh” type according to a recent classification (Zirbel et al. 2009). It ensues that the 1-4 G·A trans-Sugar/Watson–Crick pair (t-SW) occurring in GNRA loops should not be considered as a U-turn determinant although it is essential for interactions with GNRA receptors (Fiore and Nesbitt 2013).

As an important outcome, the characteristic 1-4 nucleobase—phosphate (or nucleobase–OP) hydrogen bond imposes the formation of an oxygen–π or phosphate–π stacking contact between the first nucleobase and an OP atom of the third nucleotide. A PDB survey led to an average OP–π stacking distance of 3.0 ± 0.2 Å, with a maximum distance of 3.5 Å. This oxygen–π contact, which is a further characteristic of U-turns, has rarely been described (Egli and Sarkhel 2007).

It emerges that these two features, namely the 1-4 nucleobase–OP hydrogen bond and the OP–π stacking contacts, are sufficient to unambiguously characterize a U-turn. The latter criterion allows us also to distinguish between regular and partially degenerated or unfolded U-turns, which correspond to loops with no oxygen–π stacking contact and are most often found at RNA–protein interfaces. However, such occurrences are rare (see the following section).

A U-turn variant has been identified for UNAC sequences (Fig. 1B). These loops were found to mimic GNRA tetraloops since their backbone conformations are similar (Zhao et al. 2012). The 1-4 interaction involves a U+C trans-Sugar/ Hoogsteen (t-SH) pair instead of a hydrogen bond involving the OP atom of the fourth nucleotide as in more typical U-turns. Yet, in the examples we collected, the OP–π contact between the first nucleobase and an OP atom of the third nucleotide is conserved. In the following, we call this U-turn variant a “U_{SH}-turn” because of the consistent presence of a 1-4 t-SH pair.

Note that the cGANCgc tetraloop in group IIC introns has a backbone that is similar to that of a U-turn and a 1-4 G·A t-SW pair (Keating et al. 2008). Although rare, these GANC loops are examples of structured tetraloops with no oxygen–π contact. For all U-turns, it is important to note that the last three nucleobases are stacked in a manner that their exposed Watson–Crick edges can establish specific tertiary contacts such as, for example, within anticodon–codon associations or with cognate receptors (Fiore and Nesbitt 2013; Tanaka et al. 2013).

**Z-turn and Z_{anti}-turn signatures**

UNCG tetraloops are not based on a U-turn but on a newly defined “Z-turn”: they embed a trans-Sugar/Watson–Crick (t-SW) interaction between the first and fourth nucleobase, associated with a C2’-endo pucker of the third residue, and a syn conformation of the fourth residue. In addition, the third and fourth ribose rings adopt an uncommon head-to-tail orientation (Fig. 1C). This particular combination of rare structural features is characteristic of Z–DNA/RNA motifs and implies an O4′–π stacking contact (Egli and Sarkhel 2007; D’Ascenzo et al. 2016). The 3-4 O4′–π stacking contact in Z-turns is comparable with the 1-3 OP–π stacking contact in U-turns. Furthermore, the average stacking distance (3.1 ± 0.2 Å) and the maximum distance (3.5 Å) are similar in both turns. Thus, we can assume that to define a Z-turn as found in UNCG loops, we can rely on both the 1-4 base pair essentially of the t-SW type as described below, and the O4′–π stacking contact.

Such a definition is not based on the syn conformation of the fourth nucleotide and therefore allows us to consider rare motifs where the O4′ stacking involves bases in anti, such as found in some CUUG folds (Fig. 1D; Jucker and Pardi 1995b). Hence, as for U-turns, we can define two Z-turn subcategories: the main Z-turn or Z_{syn}-turn— with the fourth nucleobase in syn—and the less frequent “Z_{anti}-turn” variant—with the fourth nucleobase in anti. Most Z_{anti}-turns are not associated with a t-SW 1-4 pair but with a cis-Watson–Crick/ Watson–Crick (c-WW) pair. As such, these Z_{anti}-turns are also known as di-loops. Interestingly, the characteristic C2’-endo sugar pucker of UNCG tetraloops seems to be conserved in all Z-turn types.

**U-turns and Z-turns dominate the tetranucleotide folding landscape in RNA hairpins**

In our unified definition of U-turns and Z-turns in RNA hairpins, each turn is distinguished by the presence of either a 1-3 or 3-4 oxygen–π contact (Egli and Sarkhel 2007). With the above-defined criteria, we searched the PDB for occurrences of these two turns and their variants in crystal and NMR structures, among tetranucleotide sequences embedded in RNA hairpin loops (Table 1). As expected, U-turns in tetranucleotide sequences starting with G, U, or C’ are the most frequent, followed by Z-turns in UNCG tetraloops. U_{SH} turns are less frequent and are
associated with UNAC sequences. $Z_{\text{anti}}$-turns are slightly more frequent and diverse and comprise essentially CNNG sequences. The “Uncategorized” motifs are mostly of the partially unfolded U-turn type—where the 1–4 interaction is present, but not the OP–π stacking contact. They correspond also to folds that are too rare and/or disordered to allow for their assignment to any clearly defined category, or to partially unfolded conformations induced by proteins. The rare GANC tetranucleotide loop has only been identified in group IIC introns based on structural and phylogenetic evidence and has only been reported when bound to its cognate receptor (Keating et al. 2008). Thus, our early assumption that the

**FIGURE 1.** Examples of a GNRA “U-turn” and a UNCG “Z-turn” along with their $U_{\text{anti}}$-turn and $Z_{\text{anti}}$-turn variants (1–4 bp and relevant nucleobase–phosphate hydrogen bonds are shown in the insets). In all panels, the cyan dashed lines mark contact distances between the OP/O4′ atoms—emphasized as yellow spheres—and the stacked nucleobase that are associated with oxygen–π contacts ≤3.5 Å (see Materials and Methods and insets of panels A and C). For clarity, all nonrelevant OP atoms were hidden. The C = G closing base pairs are shown in white. For all secondary structures, symbols according to the Leontis and Westhof nomenclature were used (Leontis and Westhof 2001; Nasalean et al. 2009). (A) $G_{\text{anti}}$AAA tetraloop (chain A) adopting a classical U-turn (symbolized by a circled “U”). The first G and the phosphate of the third nucleotide involved in an OP–π contact are marked in red as well as the oxygen atoms of the phosphate involved in the 1–4 base–phosphate hydrogen bond. The three stacked A nucleotides are colored in wheat. (B) U$_{\text{anti}}$CAC tetraloop (chain 0) adopting the rare $U_{\text{anti}}$-turn variant (symbolized by a circled “U”). The first U and the phosphate of the third nucleotide are marked in red. The three stacked CAC nucleotides and part of their backbone are colored in wheat. (C) $U_{\text{anti}}$CCG tetraloop (chain B) adopting a Z-turn (symbolized by a circled “Z”). The CpG step forming a Z-RNA motif is shown in red. The two ribose O4′ atoms of the CpG step are shown in yellow to mark the characteristic head-to-tail orientation of the sugars. The fourth nucleotide adopts a syn conformation. The UpC step is colored in wheat. (D) $C_{\text{anti}}$UUGu pentaloop (chain 1) adopting a rare $Z_{\text{anti}}$-turn variant (symbolized by a circled “Z”). The UpG step forming a Z-RNA motif, with the G adopting an anti instead of a syn conformation, is shown in red. The two ribose O4′ atoms of the CpG step are shown in yellow to mark the characteristic head-to-tail orientation of the sugars. The CpU step is colored in wheat and the bulged “u” in blue.
The largest part of tetranucleotide folds in hairpins is based on a U-turn or a Z-turn comprising an oxygen–π stacking contact, which is supported by this survey. Consequently, we can assume that most GNRA and UNCG tetranucleotide fold predictions based on sequence alignments are correct (Table 1). However, these data also indicate that some sequences expected to form a U-turn are associated with a Z-turn and vice versa. Thus, the sequence of a tetraloop does not systematically dictate its fold. For instance, we identified a GCAAu sequence that adopts a Zanti-turn (Fig. 2). Further, one GUGA sequence of the GNRA type adopting a Z-turn was observed in an RNA–protein complex (Fig. 3A). NMR structures of anticodon loops containing the U33NCG sequence were found to adopt a Z-turn under specific conditions, in agreement with their sequence but not with the expected anticodon–codon binding scheme (see below). These examples are more thoroughly described in the following sections. A detailed report describing the structural features of tetranucleotide folds will be provided elsewhere, the main purpose of this account being to establish the interchangeability between U-turns and Z-turns.

**GNRA and GNYA dimorphism**

Loop dimorphism came upon us serendipitously. We found that it deserved special attention, as we realized that it

---

**Table 1.** Number of U-turns, Z-turns and their variants (as defined in the text and in Fig. 1) associated with tetranucleotide sequences involving a 1–4 nucleobase-nucleotide contact and occurring in RNA hairpin loops not longer than eight residues.

| Sequence | U-turn | U$_{anti}$-turn | Z-turn | Z$_{anti}$-turn | Uncategorized |
|----------|--------|-----------------|-------|----------------|--------------|
| GNRA     | 1353 [416] | — | 2 [2] | — | 38 [21] $^a$ |
| In tetraloops | 106 [29] | — | — | — | 68 [29] $^a$ |
| In larger hairpins | 106 [29] | — | — | — | 68 [29] $^a$ |
| GNRC     | 47 [17] | — | — | — | 5 [4] $^a$ |
| In tetraloops | 1 [1] | — | — | — | 1 [1] $^b$ |
| In larger hairpins | 18 [11] | — | — | — | 3 [3] $^c$ |
| GNYY     | 12 [7] | — | — | — | 1 [1] $^c$ |
| In tetraloops | 6 [4] | — | — | — | 5 [4] $^a$ |
| In larger hairpins | 706 [252] | — | — | — | 55 [24] $^d$ |
| UNNN (not UNCG) | — | — | — | — | — |
| In tetraloops | 46 [13] | NMR (2MQT/V) | — | — |
| In larger hairpins | 706 [252] | — | — | 2 [1] 55 [24] $^d$ |
| C(+)/AAC | — | — | — | — |
| In tetraloops | 57 [2] | — | — | — |
| In larger hairpins | 57 [2] | — | — | — |
| CNNG     | — | — | 6 [4] NMR (1RNG) | — |
| In tetraloops | — | — | — | 74 [52] |
| In larger hairpins | — | — | — | 74 [52] |
| CNNG(notG) | — | — | — | 3 [2] $^a$ |
| In tetraloops | — | — | — | 24 [9] $^d$ |
| In larger hairpins | — | — | — | 24 [9] $^d$ |
| ANN      | — | — | — | 12 [6] $^d$ |
| In tetraloops | — | — | — | 111 [51] $^d$ |
| In larger hairpins | — | — | — | 111 [51] $^d$ |
| Total | 2822 [891] | 46 [13] | 248 [69] | 148 [86] | 379 [184] |

$^a$Mostly U/Z-turn-like, but with nonstandard geometry (oxygen–π stacking or hydrogen bond distances above 3.5 Å).

$^b$GANC loops in group IIC introns.

$^c$Mostly tetraloop folds in hairpins that are not inducing turns (will be discussed elsewhere).

$^d$Mostly unstructured.

These data were derived from a survey of X-ray structures from the PDB (October 2016; resolution ≤3.0 Å). The estimated number of nonredundant occurrences is given in brackets. Tetranucleotide sequences having at least one atom with a B-factor >79 Å$^2$ were excluded. "NMR" in the table refers to folds for which only NMR structures are available; the corresponding PDB codes are given in parenthesis. These structures are not included in the total.
impacted our ability to derive three-dimensional structures from secondary structures. Upon looking at GNRA and GNYA loops, we noted that the phylogenetically conserved cGUGAg loop that caps helix 93 in domain V of all large ribosomal subunits adopts the expected U-turn. However, the same cGUGAg loop located within a 21-nt-long ribosomal fragment in a complex with a pseudouridine synthase adopts an unexpected Z-turn, which is made possible through the formation of a 1–4 G•A t-SW pair (Fig. 3A; Czudnochowski et al. 2014). Whether the Z-turn is induced by the pseudouridine synthase or by crystal constraints is unclear. However, it is tempting to speculate that some RNA binding proteins and modification enzymes could recognize and/or induce Z-turns in GNRA sequences.

Loop dimorphism was also observed in larger motifs containing GNRA sequences, such as the phylogenetically conserved 7-nt uGAAAgg loop that caps helix 35a in domain II of large ribosomal subunits (Hsiao et al. 2006; Nasalean et al. 2009; D’Ascenzo et al. 2016). In every X-ray and cryo-EM structure of a ribosome available to date (including mitochondrial ribosomes), this uGAAAgg—or uGACAgg in Homo sapiens mitochondrial ribosome, only one X-ray occurrence of a uGACAg in Deinococcus radiodurans (Fig. 3C) has been reported, where the tetranucleotide sequence adopts a U-turn (Table 1). Yet, NMR experiments illustrated that a cGUUAg loop (Ihle et al. 2005) and a uGCUAg loop (Melchers et al. 2006) can adopt a Z-turn rather than the anticipated U-turn (PDB codes: 1Z30 and 2EVY).

Overall, although such dimorphism is not frequent among structured RNAs (Table 1), it might be relevant when deriving the structures of noncoding RNA that may adopt several transient folds in order to achieve their functions within a large diversity of environments (Cech and Steitz 2014). It would therefore be interesting to explore how such conformational changes occur in vivo, especially since an anti to syn conversion could not easily be fathomed without stem unwinding.

**UNCG dimorphism: U-turns or Z-turns in tRNA anticodon loops?**

It is generally well appreciated that longer loops—from pentaloops to larger motifs—can embed tetranucleotide sequences that adopt U-turns (Hsiao et al. 2006). One of the most biologically relevant systems to incorporate this fold is the 7-nt-long tRNA anticodon loop. In the context of protein synthesis, any U_{3,NNN} sequence will adopt a U-turn (Auffinger and Westhof 2001) so that the three anticodon bases are able to associate with the three complementary...
FIGURE 3. (Legend on next page)
bases of the codon on the messenger RNA (mRNA). However, would a U_{33}NCG anticodon sequence naturally adopt that classical U-turn conformation required for translation instead of the more cogent Z-turn? Do such anticodon loops manage to switch from U-turns to Z-turns and, if yes, which environmental context would direct such a structural transition or impose one over the other fold?

In that respect, it could be envisaged that nucleotide modifications play a role in facilitating or preventing U_{33}NCG anticodon loops from adopting a Z-turn. NMR experiments were performed on four variants of tRNA^{Arg}_{1,2} stem-loops possessing a U_{33}ACG sequence and containing diverse combinations of RNA modifications such as A_{34}/I and C_{32}/S_{2C}—PDB codes: 2KRP/Q/V/W (Cantara et al. 2012). This study revealed that all modified and nonmodified anticodon loops adopt a Z-turn, although the absence of a natural m^{2}A_{33} post-transcriptional modification could have biased the outcome. In any case, it seems fair to state that the extent of nucleotide modifications modulates the conformational plasticity of the tRNA^{Arg}_{1,2} anticodon loop in order to secure the essential U-turn conformation (Sundaram et al. 2000). However, in its unmodified state, the loop could also adopt a Z-turn and be recognized by specific proteins, as in the above-mentioned 4LGT pseudouridine synthase complex (Fig. 3A).

To summarize, these U_{33}ACG anticodon sequences can successively adopt at least three distinct folds. They journey from a Z-turn in their free state, through a “degenerated” fold when bound to their cognate tRNA synthetases—see for example, tRNA^{Arg}_{1} with a U_{33}JCG anticodon; PDB code: 1F7U (Delagoutte et al. 2000)—to end with a classical U-turn when interacting with mRNA codons. RNA modifications—or their absence—may determine how anticodon loops fold, thereby altering or suppressing the tRNA codon-reading capacity.

Could Z-turns of U_{33}NCG anticodon loop sequences be associated with a specific biological function? Would a Z-turn be necessary for the recognition of modification sites by tRNA synthetases? In that case, could Z-turns within anticodon loops also occur when other NpG steps replace CpG within the U_{33}NCG sequence? After all, it has been established that almost all dinucleotide sequences can adopt Z- RNA conformations (see Fig. 3A,B for GpA and ApA Z-steps) and therefore be part of Z-turns (D’Ascenzo et al. 2016). Indeed, a NMR structure of a UCAGu pentaloop with an ApG Z-step has been reported—PDB code: 1Q7S (Theimer et al. 2003). If that hypothesis holds true, 16 out of the 64 anticodon sequences ending with a G—thereby comprising the four U_{33}NCG sequences—could potentially adopt a Z-turn. Our understanding of translation regulation, of decoding rules and of the role of modified bases in tRNAs could be expanded by these findings (Grosjean and Westhof 2016).

Are other folds possible for U_{33}NNN sequences? A different UGAA fold has been reported in the NMR structure of an RNA hairpin—PDB code: 1AFX (Butcher et al. 1997). However, we did not consider this fold since no 1–4 interaction was present and since this loop has not been reported elsewhere. We already described UNAC sequences (Zhao et al. 2012) that can adopt the alternative U_{511}–t–SH pair (Fig. 1B). We also identified a UUUAa pentanucleotide sequence in a ribosome structure that adopts the Z_{anti}-turn variant and that is closed by a 1–4 U–A c-WW pair (Fig. 3D). Thus, U_{33}NNN anticodon loops can theoretically adopt any of the four folds we described, depending on the nature of nucleotide 36 and the associated structural context. Although most of these folds are rarely found in experimental structures, they can transiently appear in the folding pathways of these loops depending on sequence and modification levels.

**Which turns for CNNN and ANNN sequences?**

Similarly, we wondered whether CNNN sequences adopt a unique fold specific to their sequence or multiple conformations. When the C nucleotide is protonated, typical U-turns can be formed as shown by NMR and in ribosomes—see C_{4871}AACu in Halocarcula marismortui (Gottstein-Schmidtke et al. 2014). It was inferred from NMR and thermodynamic measurements (Proctor et al. 2002) as well as X-ray crystallography (Fig. 3E) that CNNG sequences can form either Z-turns—PDB code: 1ROQ—(Du et al. 2003; Oberstrass et al. 2006; Schwalbe et al. 2008), or Z_{anti}-turns.
For the latter, the 1–4 \( C = G \) c-WW pair was significantly buckled, probably due to constraints imposed by the "di-loop" fold—PDB code: 1RNG (Jucker and Pardi 1995b). Interestingly, the cCAAGg loop that caps helix 14 of the small subunits of eukaryotic ribosomes (Fig. 3E) takes the place of a UACG loop in bacterial ribosomes, both forming a Z-turn. Besides UNNC, CNNC sequences could potentially form USH-turns, although the latter have not yet been observed (Fig. 3F). Again, these loops starting with a C residue display an unanticipated plasticity, suggesting that the fold they adopt is largely context dependent.

Tetranucleotide sequences starting with an adenine are almost nonexistent, at least in crystallographic structures (Table 1). If they exist, they do not seem to display a significant and/or stable 1–4 contact as reported for the other loops described here. Hence, especially when the loop interacts with a protein, it is difficult to refer to these tetranucleotides as being "structured." However, we do not exclude the possibility that additional motifs might emerge in newly deposited crystal or NMR structures. For instance, since a UUUaA pentaloop with a \( Z_{\text{anti}} \)-turn implying a 1–4 U–A c-WW pair was observed, an ANNUn pentaloop with a similar turn and a 1–4 A–U pair cannot be dismissed. Such possibilities have been reported by NMR for uGUUC and CUUGu pentaloops—PDB code: 2L6I (Lee et al. 2011).

**Phylogenetic considerations on tetranucleotide loops in RNA**

Phylogenetic data on 16S rRNA suggested early on that helix 6 (positions 83–86 in *Escherichia coli* 16S rRNA) is capped either by a CUUG (45%), a UUCG (36%), or a GCAA (13%) tetraloop (Woese et al. 1990; Konings and Gutell 1995).
Thus, it could be concluded that this stem can be capped either by a Z–turn or by a U-turn. According to our present study, these three sequences can also adopt a Z-turn. Such loop polymorphism might complicate the interpretation of biochemical data, for example, when highly conserved GAAA tetraloops in 16S rRNA are substituted by a UACG sequence (Sahu et al. 2012). In addition, the fact that this loop is unstructured in the 4YBB *Escherichia coli* crystal structure (resolution: 2.1 Å) might interrogate classical phylogenetic data interpretations. Indeed, in the seven UNCG tetranucleotide sequences deduced from the 16S *Escherichia coli* 2D structure, only three adopt a canonical Z-turn and the other sequences appear in disordered regions with, however, a G nucleotide in *syn* for four of them. The reasons as to why these loops appear as disordered are not yet understood.

Thus, sequence interchangeability might be hiding structural similarity. As noted above, the Z-turn GAAA loop capping helix 35a in the 50S of *Halobaculur marismortui* could exchange with YNMG sequences. Further, convincing evidence of sequence exchange that leads to similar folds has been reported in studies of viral RNA hairpins (Melchers et al. 2006; Liu et al. 2009; Zoll et al. 2011; Clabbers et al. 2014; Prostova et al. 2015).

**Sequence–structure relationships**

It is our hope that the data we gathered (summarized in Fig. 4) will help to interpret tetranucleotide sequence variations from a structural perspective, as they inform on the prevalence of a sequence to adopt (or not) a given fold. For example, GNNA sequences with a 1–4 G+A base pair can adopt a classical GNRA U-turn fold but also a Z-turn and even a Zanti-turn, but not a Us+1-turn. Similarly, UNNG sequences can adopt U-turns and Z-turns, but not the two other less frequent variants. Finally, the GNNG and GNNU sequences are only found in the U-turn category. This classification reflects our current understanding of tetranucleotide turns and might be completed or refined with the advent of new non-coding RNA structures.

**Final thoughts about folds and structure prediction**

We report that tetraloop and tetranucleotide folds are not systematically determined by their sequence, possibly because of subtle changes in their environment and in the sequence of connected residues. A logical implication of this observation is that, for any given RNA sequence for which the 3D structure is not available, we are unable to ascertain with 100% confidence how the hairpins it contains will fold. With prior knowledge acquired on ribozymes (Schultes and Bartel 2000; Woodson 2015) and riboswitches (Garst et al. 2011; Batey 2015), we became aware that the same RNA sequence can adopt distinct folds in order to carry out specific functions. The structural analysis we present here reveals that only two folds dominate the tetranucleotide landscape. Consequently, predicting whether GNRA, UNCG, or related sequences within any noncoding RNA will adopt a U-turn involving a phosphate–π stacking contact or a Z-turn with a O4′–π stacking ceases to be a straightforward exercise. Without additional stereochemical rules, the structure adopted by such tetranucleotide sequences might remain complex to predict and more structural information on these essential folds needs to be accumulated. It could therefore be informative to see how current 3D structure prediction methods would perform when confronted with such noncompliant pieces of the RNA puzzle (Miao et al. 2015).

Efforts to fold these tetranucleotide sequences by molecular dynamics simulations are currently only partially successful, although significant progress has been made in that direction (Kührova et al. 2013; Haldar et al. 2015; Miner et al. 2016). Such modeling attempts have now to face new challenges: finding not only one, but two or more folds, while grasping their relationship with the environment. Recently, some simple procedures based on diffusion maps and Markov models found the alternative Z-turn fold of a GAAA loop (Bottaro et al. 2016). Such methods are however currently limited to small fragments—4 nt and no closing base pair in that instance. Although this represents an essential first step in assessing folding pathways, it will certainly be much more challenging to predict the occurrence of such folds or turns embedded in the core of complex RNP particles like ribosomes.

Tetraloop fold variability probably only makes for the tip of the iceberg in the folding adaptability that characterizes regulatory RNAs. Regardless of how daunting they may seem, scenarios of folding plasticity at the local level are both attractive and relevant for molecules that comprise several thousands of nucleotides and that are thought to be mostly devoid of well-defined 3D structures (Gardini and Shiekhattar 2015; Rivas et al. 2017). We could envision how this plasticity of the most basic RNA folds would be well suited to regulatory RNAs that are obligatory opportunists, by *nature*. The race is on toward “overturning more rules” about RNA structure and folding (Cech and Steitz 2014).

**MATERIALS AND METHODS**

We searched the PDB (October 2016; X-ray data; resolution ≤3.0 Å) for tetranucleotide sequences in RNA hairpins that involve a 1–4 nucleobase–nucleotide interaction and an oxygen–π contact as defined below. For that purpose, we used the DSSR program (Lu et al. 2015). DSSR was also used to isolate tetranucleotide sequences embedded in loops comprising not more than eight residues. For characterizing 1–3 and 3–4 oxygen–π contacts, we specified in DSSR a 3.5 Å cutoff between the OP/O4′ oxygen atom and the nucleobase plane. In addition, the projection of the OP/O4′ oxygen on the base plane had to lie within the surface of the nucleobase aromatic cycles. A polygon-offset of 0.5 Å was used to take into

*U-turn versus Z-turn in RNA hairpin loops*
ACKNOWLEDGMENTS

P.A. and Q.V. wish to thank Professor Eric Westhof for ongoing support and helpful discussions, as well as Professors Neocles Leonaris and Richard Giegé for useful comments on the manuscript. Q.V. also acknowledges Dr. Yaser Hashem for support. The authors acknowledge the support of the French “Ministère de la recherche et de l’enseignement” (to L.D.); Polish Ministry of Higher Education and Science (Mobility Plus programme) (1103/MOB/2013/0 to F.L.); LabEx: ANR-10-LABX-0036_NETRNA (to Q.V.); and CNRS (funding for open access charge).

Received September 8, 2016; accepted November 29, 2016.

REFERENCES

Allain FHT, Varani G. 1995. Structure of the P1 helix from group I self-splicing introns. J Mol Biol 250: 333–353.

Amunts A, Brown A, Toots J, Scheres SH, Ramakrishnan V. 2015. Ribosome. The structure of the human mitochondrial ribosome. Science 348: 95–98.

Auffinger P, Westhof E. 2001. An extended structural signature for the tRNA anticodon loop. RNA 7: 334–341.

Batey RT. 2015. Riboswitches: still a lot of undiscovered country. RNA 21: 560–563.

Bottaro S, Gil-Ley A, Bussi G. 2016. RNA folding pathways in stop motion. Nucleic Acids Res 44: 5883–5891.

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

Butcher SE, Dieckmann T, Feigon J. 1997. Solution structure of the conserved 16S-like ribosomal RNA UGAA tetraloop. J Mol Biol 268: 348–358.

Cantara WA, Bilbille Y, Kim J, Kaiser R, Leszczyńska G, Malkiewicz A, Agris PF. 2012. Modifications modulate anticodon loop dynamics and codon recognition of E. coli tRNAAsp-A1. J Mol Biol 416: 579–597.

Cech TR, Steitz JA. 2014. The noncoding RNA revolution—trashing old rules to forge new ones. Cell 157: 77–94.

Cheong C, Varani G, Tinoco I. 1990. Solution structure of an unusually stable RNA hairpin, 5’GGAC(UUCG)GUCC. Nature 346: 680–681.

Cheong H, Kim N, Cheong C. 2015. RNA structure: tetraloops. In ELS. Wiley, Chichester, UK.

Clabers MTB, Olslohorn RCL, Gultyaev AP. 2014. Tospovirus ambisense genomic RNA segments use almost complete repertoire of stable tetraloops in the intergenic region. Bioinformatics 30: 1800–1804.

Correll CC, Swinger K. 2003. Common and distinctive features of GNRA tetraloops based on a GUAA tetraloop structure at 1.4 Å resolution. RNA 9: 355–363.

Cruz JA, Westhof E. 2009. The dynamic landscapes of RNA architecture. Cell 136: 604–609.

Czudnochowski N, Ashley GW, Santi DV, Alian A, Finer-Moore J, Stroud RM. 2014. The mechanism of pseudouridine synthases from a covalent complex with RNA, and alternate specificity for U2605 versus U2604 between close homologs. Nucleic Acids Res 42: 2037–2048.

D’Ascenzo L, Leonarsi F, Vicens Q, Auffinger P. 2016. ‘Z-DNA like’ fragments in RNA: a recurring structural motif with implications for folding, RNA/protein recognition and immune response. Nucleic Acids Res 44: 5944–5956.

Delagoutte B, Moras D, Cavarelli J. 2000. Transfer-RNA aminocoylation by arginyl-transfer-RNA synthase: induced conformations during substrates binding. EMBO J 19: 5599–5610.

Du Z, Yu J, Andino R, James TL. 2003. Extending the family of UNCG-like tetraloop motifs: NMR structure of a CACG tetraloop from coxsackievirus B3. Biochemistry 42: 4373–4383.

Egli M, Sarkhel S. 2007. Lone pair- aromatic interactions: to stabilize or not to stabilize. Acc Chem Res 40: 197–205.

Ennifar E, Nikulin A, Tischchenko S, Seganov A, Nevskaya N, Garber M, Ehresmann B, Ehresmann C, Nikonorov S, Dumas P. 2000. The crystal structure of UUCG tetraloop. J Mol Biol 304: 35–42.

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

Gardini A, Shiekhattar R. 2015. The many faces of long noncoding RNAs. FEBS J 282: 1647–1657.

Garst AD, Edwards AL, Batey RT. 2011. Riboswitches: structures and mechanisms. Cold Spring Harb Perspect Biol 3: a003533.

Gottstein-Schmidtke SR, Duchscheidt-Ferner E, Groher F, Weigand JE, Gottstein D, Suess B, Wohnert J. 2014. Building a stable RNA U-turn with a protonated cytidine. RNA 20: 1163–1172.

Grosjean H, Westhof E. 2016. An integrated, structure- and energy-based view of the genetic code. Nucleic Acids Res 44: 8020–8040.

Gutell RR, Cannone JJ, Konings D, Gautheret D. 2000. Predicting U-turns in ribosomal RNA with comparative sequence analysis. J Mol Biol 300: 791–803.

Haldar S, Kührova P, Banas P, Spivok W, Sponer J, Hobza P, Otyepka M. 2015. Insights into stability and folding of GNRA and UNCG tetra-loops revealed by microsecond molecular dynamics and well-tempered metadynamics. J Chem Theory Comput 11: 3866–3877.

Hall KB. 2015. Mighty tiny. RNA 21: 630–631.

Hendrix DK, Brenner SE, Holbrook SR. 2005. RNA structural motifs: building blocks of a modular biomolecule. Q Rev Biophys 38: 221–243.

Heus HA, Pardi A. 1991. Structural features that give rise to the unusual stability of RNA hairpins containing GNRA loops. Science 253: 191–194.
Hsiao C, Mohan S, Hershkovitz E, Tannenbaum A, Williams LD. 2006. Single nucleotide RNA choreography. *Nucleic Acids Res* **34**: 1481–1491.

Ihle Y, Ohlenschlager O, Hafner S, Duchardt E, Zacharias M, Seitz S, Zell R, Ramachandran B, Gorlach M. 2005. A novel cGUGAg tetraloop structure with a conserved yYNMGG-type backbone conformation from cloverleaf I of bovine enterovirus 1 RNA. *Nucleic Acids Res* **33**: 2003–2011.

Jucker FM, Pardi A. 1995a. GNRA tetraloops make a U-turn. *RNA* **1**: 219–222.

Jucker FM, Pardi A. 1995b. Solution structure of the CUUG hairpin loop: a novel RNA tetraloop motif. *Biochemistry* **34**: 14414–14427.

Jucker FM, Heus HA, Yop PF, Moors HHM, Pardi A. 1996. A network of heterogeneous hydrogen bonds in GNRA tetraloops. *J Mol Biol* **264**: 968–980.

Keating KS, Toor N, Pyle AM. 2008. The GANC tetraloop: a novel motif in the group IIC intron structure. *J Mol Biol* **383**: 475–481.

Kim SH, Sussman JL. 1976. π-turn is a conformational pattern in RNA loops and bends. *Nature* **260**: 645–646.

Klosterman PS, Hendrix DK, Tamura M, Holbrook SR, Brenner SE. 2004. Three-dimensional motifs from the SCOR, structural classification of RNA database: extruded strands, base triples, tetraloops and U-turns. *Nucleic Acids Res* **32**: 2342–2352.

Konings DAM, Guttell R. 1995. A comparison of thermodynamic foldings with comparatively derived structures of 16S and 16S like RNAs. *RNA* **1**: 559–574.

Kührova P, Banas P, Best RB, Sponer J, Otyepka M. 2013. Computer folding of RNA tetraloops? Are we there yet? *J. Chem Theory Comput* **9**: 2115–2125.

Lee CW, Li L, Giedroc DP. 2011. The solution structure of the hepatitis B virus post-transcriptional regulatory element. *Nat Struct Mol Biol* **18**: 1130–1136.

Lee CW, Li L, Giedroc DP. 2011. The solution structure of the hepatitis B virus post-transcriptional regulatory element. *Nat Struct Mol Biol* **18**: 1130–1136.

Lee CW, Li L, Giedroc DP. 2011. The solution structure of the hepatitis B virus post-transcriptional regulatory element. *Nat Struct Mol Biol* **18**: 1130–1136.

Lee CW, Li L, Giedroc DP. 2011. The solution structure of the hepatitis B virus post-transcriptional regulatory element. *Nat Struct Mol Biol* **18**: 1130–1136.

Lee CW, Li L, Giedroc DP. 2011. The solution structure of the hepatitis B virus post-transcriptional regulatory element. *Nat Struct Mol Biol* **18**: 1130–1136.