Single nucleotide RNA choreography

Chiaolong Hsiao, Srividya Mohan, Eli Hershkovitz[1,2], Allen Tannenbaum[1,2] and Loren Dean Williams*

School of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta, GA 30332-0400, USA, [1]Departments of Electrical and Computer Engineering and [2]Department of Biomedical Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0400, USA

ABSTRACT

New structural analysis methods, and a tree formalism re-define and expand the RNA motif concept, unifying what previously appeared to be disparate groups of structures. We find RNA tetraloops at high frequencies, in new contexts, with unexpected lengths, and in novel topologies. The results, with broad implications for RNA structure in general, show that even at this most elementary level of organization, RNA tolerates astounding variation in conformation, length, sequence and context. However the variation is not random; it is well-described by four distinct modes, which are 3-2 switches (backbone topology variations), insertions, deletions and strand clips.

INTRODUCTION

Prediction and design of three-dimensional structures of large RNAs are best approached using small structural motifs, with modular and hierarchical characteristics (1,2). The simplest, smallest and most frequent RNA motif is known as the tetraloop. Tetraloops are terminal loops, with characteristic four-residue sequences first observed in early phylogenetic comparisons of RNAs (3–5). Tetraloops were seen to connect two anti-parallel chains of double-helical RNA, and so cap A-form stems (2). Isolated stem/tetraloops show well-defined structure, and exceptional thermodynamic stabilities (5–8). Tetraloops are thought to (i) initiate folding of complex RNA molecules (5), (ii) stabilize helical stems (5,9) and (iii) provide recognition elements for tertiary interactions and protein binding (10–13). Tetraloops have been broadly grouped by sequence into three classes (4), which are GNRA (11,14–17), UNCG (5,6,18–21) and CUUG (22,23), (where N can be any nucleotide and R is either G or A.)

Here we re-define and expand the RNA motif concept, unifying what previously appeared to be disparate groups of structures. We focus on the tetraloop motif, and demonstrate increased frequencies, new contexts, unexpected lengths and novel topologies. The results, with broad implications for RNA structure in general, show that even at this most elementary level of organization, RNA tolerates variation in conformation, topology and molecular interactions. However the variation is not random; it is well-described by four distinct modes, which are insertions, deletions, strand clips and 3-2 switches. Collectively we call these four modes DevLS (pronounced Devils, Deviations of Local Structure). The four DevLS are shown in Figure 1.

RNA structure is commonly understood by analysis of base–base interactions and proximities, which led to the concept of isosteric base pairs (24–30). RNA analysis in torsional space can be simplified and reduced in dimensionality with pseudo-bonds, which are vectors between non-bonded atoms (31). Pattern recognition methods have been applied successfully, by Pyle and coworkers (32,33), to geometric relationships between pseudo-bonds. Finally, phylogenetic covariation allows one to decipher RNA secondary and tertiary structure, and thereby infer three-dimensional structure (3,34–36). In an example that is relevant to the results described here, Gutell and coworkers (36) have observed the Lonepair Triloop (LPTL).

Multi-resolution analysis of RNA structure

We look at RNA at various resolutions (or scales) from the finest to coarsest. Note that we are using the term 'resolution' in the sense of signal processing (27) and it should not be confused with 'crystallographic resolution'. Resolution is varied by reducing natural groups of RNA atoms (bases/riboses/phosphates/residues/groups of residues, motifs, etc.) to pseudo-objects, with locations and orientations. Larger numbers of atoms in pseudo-objects correspond to lower resolutions. The basic idea is that important structural features become readily observable only in certain resolution ranges. Therefore resolution is a variable parameter like the tunable magnification of an optical microscope. Analysis of spatial relationships and interactions between RNA pseudo-objects

*To whom correspondence should be addressed. Tel: +1 404 894 9752; Fax: +1 404 894 7452; Email: loren.williams@chemistry.gatech.edu
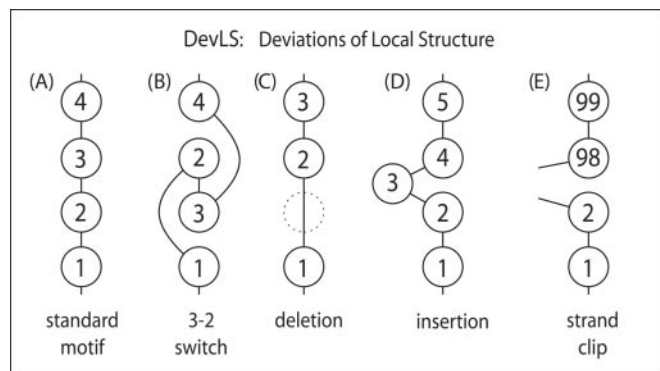
Figure 1. RNA DevLS. (**A**) A generic RNA motif is represented schematically by four circles, which symbolize four residues. (**B**) In a motif with 3-2 switch, the positions of two bases, of residues 3 and 2 in the figure, are interchanged. The backbone linkage is maintained. (**C**) In a deleted motif, a residue is omitted (dashed line). (**D**) In an inserted motif, a residue is added. (**E**) In a strand clipped motif one or more residues is contributed from a remote region of the primary sequence. An insertion, if extensive enough can be equivalent to a strand clip. The numbers indicate the covalent ordering of the residues along the polynucleotide strand. These four DevLS arise from common enabling factors, which operate at the single nucleotide level. These factors are the high RNA backbone length per residue (six bonds separate adjacent residues) and numerous torsional degrees of freedom of RNA nucleotides.

can reveal fundamental RNA architecture that is often obscure at a single resolution. Multi-resolution techniques have been very successful in protein simulations (37,38) and signal, and data processing (39). We use the multi-resolution analysis in combination with molecular interactions in an iterative process to develop empirical motif descriptions. Interactions that become evident in multi-resolution analysis are appended to a search model, leading to empirical motif definitions.

The HM 23S rRNA (1JJ2) is our test 'database'. The crystal structure of the large ribosomal subunit from Haloarcula marismortui has been determined to high resolution by Steitz and Moore (40,41). At 2.4 Å resolution, the atomic positions of the vast majority of the 23S rRNA of HM LSU are well-characterized, and, as of this writing, are more acutely determined than any other large RNA complex [although error and noise cannot be ignored, (42)]. The HM 23S rRNA, with over 2500 residues, constitutes a large database with a rich omnibus of RNA conformation and interactions.

MATERIALS AND METHODS

Detection of RNA tetraloops

To decrease the resolution of RNA, groups of atoms (bases/riboses/phosphates/residues/groups of residues, motifs, etc.) are reduced to pseudo-objects, with locations and in some cases, orientations. Larger numbers of atoms in pseudo-objects correspond to lower resolutions. A very useful space that we have developed, called PBR space (P indicates Phosphate, B indicates Base and R indicates Ribose) is shown in Figure 2. We have defined the center of mass (cm) and orientation of bases, riboses and phosphates. The relative orientations of adjacent bases are given by the angle Θ_{bpn} which is the angle between the two base plane normals. Information on relative positions of riboses is provided by

Scheme 1. Tetraloop fingerprint in PBR Space.

Θ_{rcm}. Information on relative positions of phosphates is given by Θ_{ppp}. Information on relative positions of bases is given by r_{bcm}. RNA motifs are detectable by fingerprints in PBR space (Figure 2B).

PBR space has reduced ability to distinguish among standard tetraloops and those that have undergone deletions, insertions, strand clips or 3-2 switches: at this scale they have certain equivalencies. This blurring is the point of multi-resolution analysis: successively simplify the search space to find patterns that persist from the finer to coarser scales. If a pattern indeed remains at the coarser resolution it will be much easier to discover.

Molecular interaction space, 1st iteration

The 25 tetraloops identified by torsional analysis (43) were used to devise a minimal molecular interaction definition of a tetraloop. Each of the 25 torsionally-derived tetraloops shows an interaction between the O2′ atom of the residue $j - 1$ and the N7 atom of residue $j + 1$. No other hydrogen bonding interaction is conserved. Therefore a search of all $j - 1(O2')$ to $j + 1(N7)$ interactions was conducted, giving 44 hits. Eleven of those are false positives, 33 are valid tetraloops.

Multi-scale-spaces

A variety of scale-spaces from fine to coarse grain are in preliminary use in our lab. We followed this path:

 (i) A tentative scale-space was defined.
 (ii) A preliminary tetraloop fingerprint in that scale-space was established empirically, using the 33 tetraloops identified in torsional spaces and molecular interaction spaces.
(iii) The scale-space was refined, uninformative parameters were discarded, sets of parameters yielding redundant information were consolidated. New parameters were added.
(iv) The new empirical fingerprint for a tetraloop was determined (Scheme 1), which in combination with the molecular interaction definition, gave 41 putative tetraloops.
 (v) The observed tetraloops are inspected and validated. Two tetraloops were determined to be false positives, leaving 39 tetraloops. Therefore the PBR scale-space revealed 6 tetraloops that had eluded us in torsional and our minimal interaction spaces.

Molecular interaction spaces, 2nd iteration

The additional tetraloops found in the scale-space search allowed us to re-evaluate the molecular interaction definition. The revised tetraloop definition allows either $j - 1(O2')$ to $j + 1(N7)$ or $j - 1(O2')$ to $j + 2(N7)$ hydrogen bonds (distance cutoff of 3.5 Å). This definition gives 36 tetraloops, one of
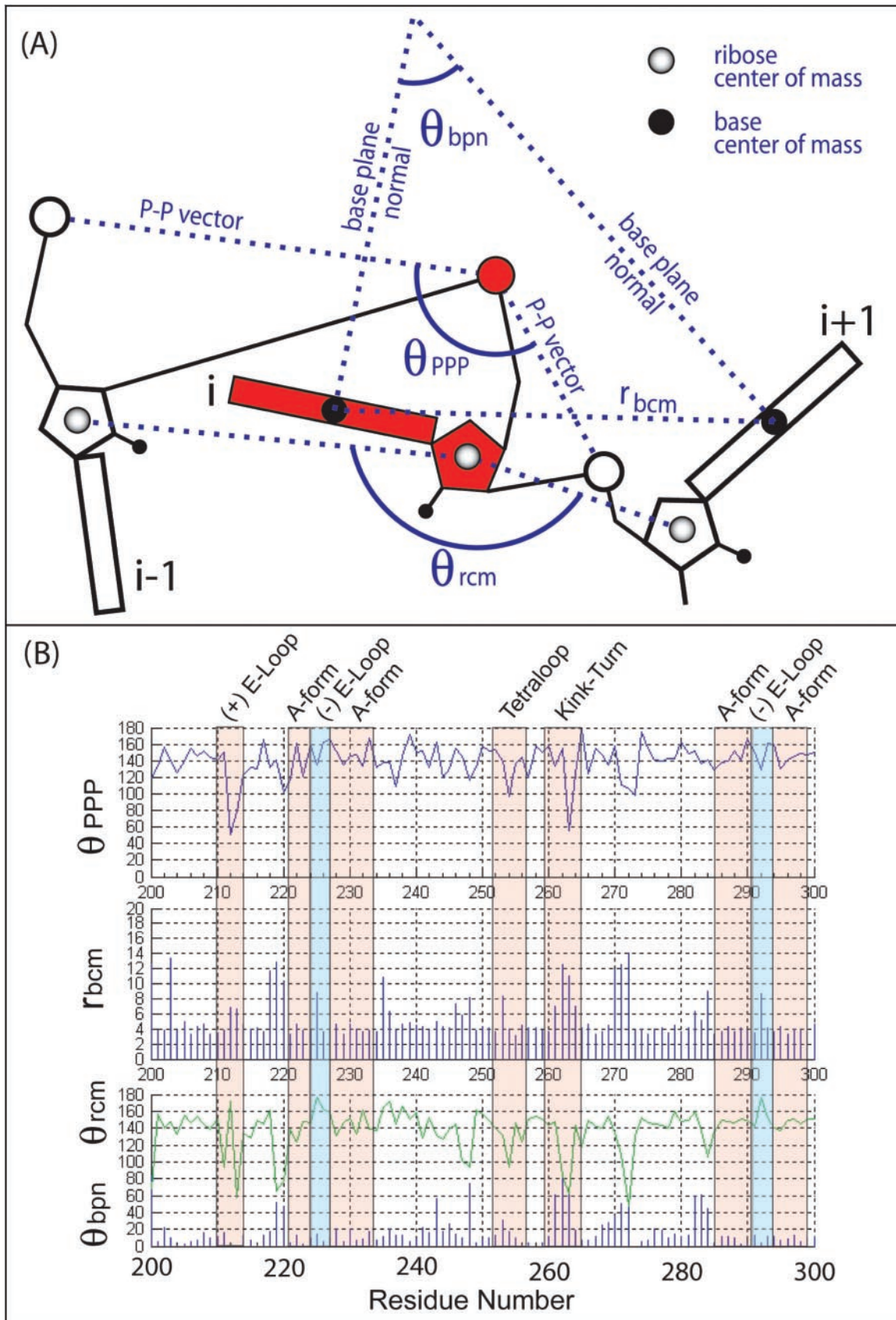
Figure 2. (**A**) PBR space. A decreased resolution view of RNA, where atoms are combined to make pseudo-objects, and special relationships between pseudo-objects are described. (**B**) Residues 200–300 of 1JJ2 in PBR space. Note that A-helices, E-loop Motifs, Kink-Turns, etc give distinctive fingerprints in PBR space.

which was not found in PBR space, plus 33 false positives. None of the false positives are common to the molecular interaction and PBR space. In combination, PBR space and molecular interaction space reveal 40 tetraloops and exclude all false positives.

Molecular interaction spaces, final description

A second class of interaction, $j - 1$(base HB donor) to $j + 2$ (O2P) [or less commonly $j + 1$ (O2P)], is observed in 33 of the 40 observed tetraloops, with only one false positive.

Cartesian spaces

It is necessary that a general, rigorous, objective and transparent statistical definition of similarity be used to validate that the RNA fragments postulated to be similar are indeed similar, and to define false positive and false negative. For these purposes we use RMSDs of atomic positions.

RESULTS

We use multi-resolution approaches and molecular interactions to identify motifs in three-dimensional structures of large RNAs. The results show that tetraloops are commonly adorned with four types of DevLS (Figure 3). DevLS occur in 17 of the 40 observed tetraloops.

Tetraloop family tree

The incorporation of DevLS into the tetraloop definition allows us to build a tetraloop family tree (Figure 4). Tetraloops fall naturally into eight groups, partitioned by the types and sites of DevLS. We have developed a nomenclature to describe tetraloop groups (Figure 3: Tl indicates tetraloop, s indicates standard, d indicates deletion, i indicates insertion, x indicates residue switch and subscripts indicate positions.) The most populated groups are the s-Tl tetraloops (21 members) and d_2-Tl tetraloops (10 members).

Intra-loop interactions

A set of consensus molecular interactions characterize tetraloops throughout the family tree, summarized for the 21 s-Tl tetraloops and the 10 d_2-Tl tetraloops in Table 1. Observed hydrogen bonding interactions are consistent with expectations for 'GNRA' tetraloops [e.g. see (14)] and U-turns. Hydrogen bonding interactions of O2′ of residue $(j - 1)$ with cross-loop base atoms are the most enduring throughout the tetraloop family tree. Twenty of 21 s-Tl tetraloops and 9 of 10 d_2-Tl tetraloops form these hydrogen bonds. s-Tl (1238) is one exception. d_2-Tl (1500), the other exception, has an O2′ $(j - 1)$ to N7 $(j + 1)$ distance of 3.5 Å, which falls nominally outside our hydrogen bonding cut-off.

Although residues $j - 1$ and $j + 2$ appear to be poised to do so, a sheared G-A base pair involving them is infrequent. In s-Tl tetraloops where residue $j - 1$ is G and residue $j + 2$ is A, only a single hydrogen bond links them [also see (17,44)]; the average N3 $(j - 1)$ to N6 $(j + 2)$ distance for s-Tl tetraloops is 4.7 Å. However for a small subset of tetraloops with DevLS, the distance is considerably shorter [3.4 Å (506), 3.5 Å (1707), 3.5 Å (482)], consistent with a true sheared G-A base pair.

As can be seen from Table 1, G and U at position $j - 1$ are interchangeable in terms of cross-loop hydrogen bonding interactions. The hydrogen bond donors N1 and N2 of G are roughly replaceable by donor N3 of U in interactions with the O2P of residue $j + 2$. G is preferred over U at $j - 1$ in s-Tl tetraloops and U is preferred over G in d_2-Tl tetraloops (see Sequence Logo: Figure 4).

DevLS influence helical capping function

Seven tetraloops are flanked by strand clips, which are observed adjacent to but not within tetraloops. All observed strand clipped tetraloops, by definition, cap pseudo-helices, where bases are stacked, and assume a helical form, but are not covalently linked by the backbone. d_2-Tl tetraloops are most frequently associated with clipping (30%). Three d_2-Tl tetraloops are strand clipped directly on the 3′ side of the tetraloop, between residues $j + 2$ and $j + 3$ (d_2-Tl tetraloops 1187, 1809, 2598). One tetraloop is strand clipped between $j - 2$ and $j - 3$ ($x_{3,2}$-Tl 482; Figure 3C). s-Tl 1629 is clipped between residues $j + 3$ and $j + 4$. $x_{3,2}$-Tl 506 is clipped between $j + 4$ and $j + 5$. s-Tl 1238 is clipped between residues $j - 1$ and $j - 2$.

Observed tetraloops are mapped onto the secondary structure, and coded by group in Figure 5. It can be observed that nineteen of 21 s-Tl tetraloops cap helices (45) (not 1238 or 1629, which are clipped). All seven standard topology i-Tl tetraloops (tetraloops with insertions but not 3-2 switches) cap helices. None of the d_2-Tl tetraloops cap unperturbed A-form stems. An unperturbed A-form stem exhibits well-defined molecular interactions such as base pairing and base stacking, with no insertions or strand clipping. Six of 10 d_2-Tl tetraloops cap helices (not 1187, 1749, 1809 or 2598). All non-clipped d_2-Tl associated helices are perturbed by unpaired bases. One d_2-Tl tetraloop (1749) caps neither a helix nor a pseudo-helix. This tetraloop is 'unhinged' in that both terminal residues $(j - 2$ and $j + 1)$ crown a cavity, and are not stacked on adjacent helical regions. Neither of the 3-2 switched tetraloops cap helices.

### Group validation and similarity statistics

We believe that 40 out of 43 entries in the Tetraloop Family Tree are structurally related, and should be described as members of a common motif. This conclusion is supported by Intra-Group and Inter-Group similarity statistics, and by conservation of molecular interactions. Intra-group similarity is characterized by RMSD of atomic positions (RMSD-AP) for atoms that are common within a group, generally backbone atoms. Inter-group similarity is characterized by RMSD-AP of specified backbone atoms that are common between two groups. RMSD-AP is determined after superimposition.

s-Tl tetraloops. The 21 s-Tl tetraloops fit the previous GNRA tetraloop definition. Intra-Group Similarity: the RMSD-AP for all backbone atoms [four residues, $(j - 1)$, (j), $(j + 1)$, $(j + 2)$] is 0.65 Å, giving a natural metric for tetraloop rigidity, and an RMSD-AP norm for evaluating degree of similarity between and within tetraloop groups. The atoms of residue $j + 2$ show the greatest deviations (Figure 6).

d_2-Tl tetraloops. In the 10 members of this group, residue $(j + 2)$ of s-Tl is absent. Residue $j + 3$ of s-Tl becomes
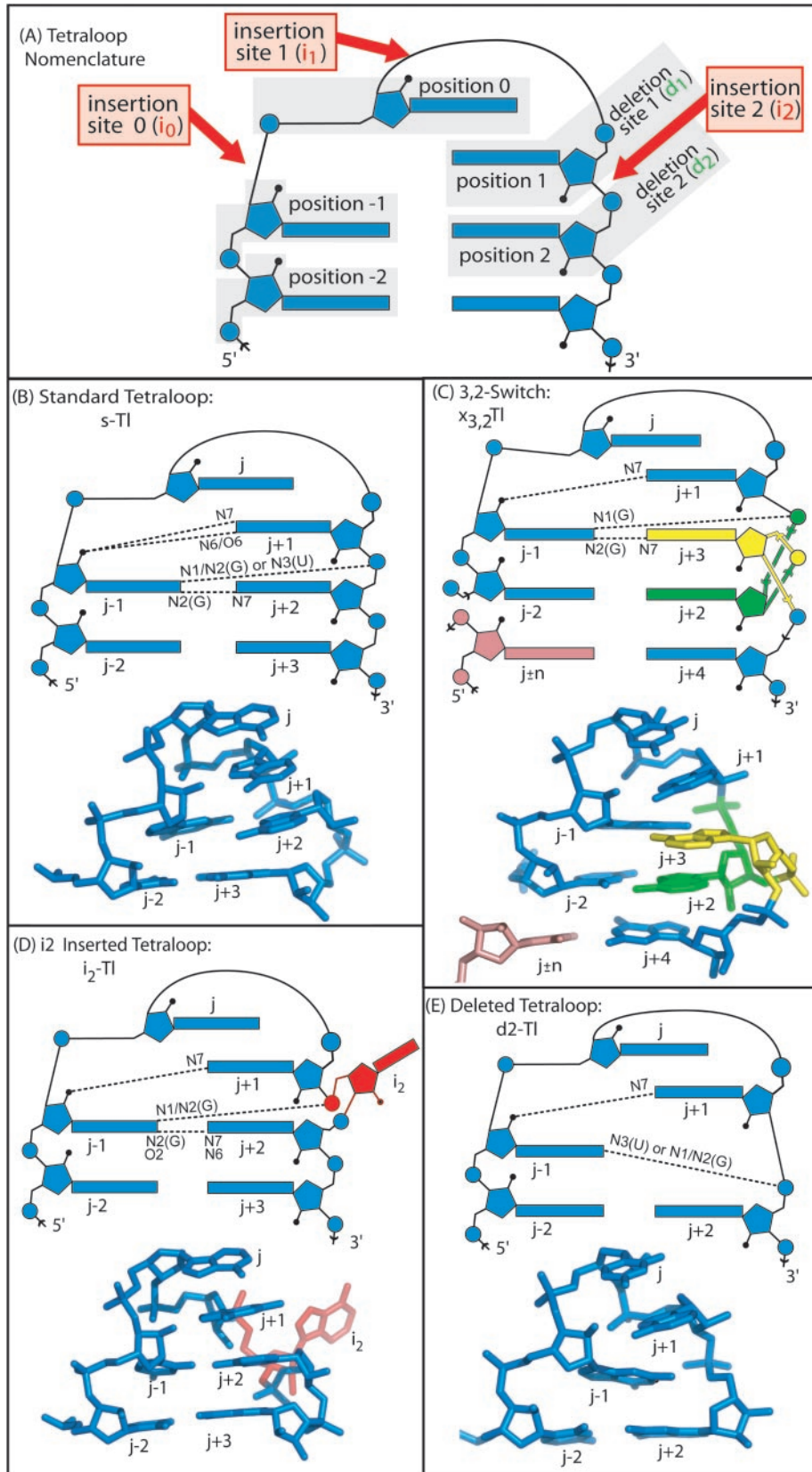
Figure 3. Tetraloops in 1JJ2 adorned with DevLS. (**A**) Observed sites of insertions (red arrows) and deletions (green text) in tetraloops. (**B**) A standard tetraloop (s-Tl tetraloop 805). (**C**) A tetraloop with a 3-2 switch ($x_{3,2}$-Tl tetraloop 482). (**D**) A tetraloop with a residue inserted at the 2 position (i_2-Tl tetraloop 494). (**E**) A tetraloop with a residue deleted at the 2 position. (d_2-Tl tetraloop 1809). Dashed lines represent consensus hydrogen bonds. Hydrogen bond donors and acceptors are indicated. The top of each panel in B–E shows a consensus schematic representation. The bottom of each panel shows a representative 3D structure from 1JJ2.
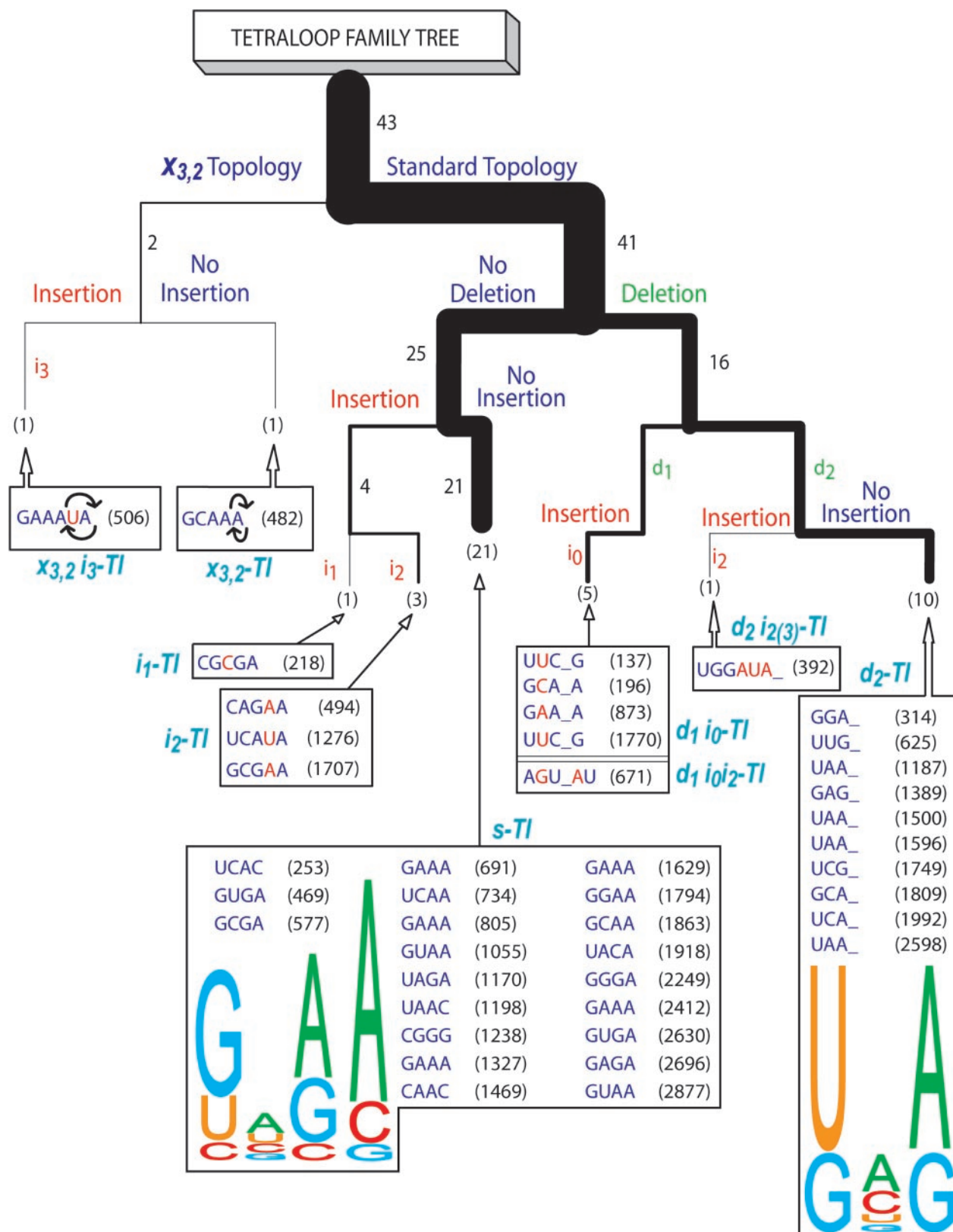
Figure 4. Tetraloop Family Tree. Forty three tetraloops of 1JJ2 are distributed by type of position of DevLS. Insertion positions are indicated in red text. Deletion positions are indicated in green text. The positions of deleted residues are marked by underscores. Number of occurrences is indicated in black, with line widths proportional to frequency. There are eight groups (boxed). The residue number of the first residue and the sequence is given for each tetraloop. The consensus sequence for the s-Tl and d_2-Tl tetraloops are indicated by a sequence Logo representation (53). Entries 196, 671, 873 were described by Huang (46). These were not detected by our methods and are outliers in conformation and molecular interactions.

j + 2 of d_2-Tl. Intra-Group Similarity: the RMSD-AP is 0.30 Å for all backbone atoms [three residues, (j − 1), (j) and (j + 1)] of this group. Thus d_2-Tl tetraloops are more restrained in conformation than s-Tl tetraloops. Inter-Group Similarity: the RMSD-AP is 0.49 Å for the backbone atoms of the ten d_2-Tl tetraloops and those of the corresponding residues [(j − 1), (j) and (j + 1)] of the 21 s-Tl tetraloops. This super-imposition is shown in Figure 6. It can be seen that deletion of residue j + 2 does not appreciably change the positions of the remaining backbone atoms of these tetraloops. However

Table 1. Consensus hydrogen bonding interactions[a] in s-Tl and d_2-Tl tetraloops

s-Tl	j + 1	Frequency	j + 2	Frequency
j − 1	O2′-N7(R)[b]	(19/20)[c]	N1/N2 (G)–O2P[d]	(14/14)
	O2′-N6/O6(R)	(12/20)	N3(U)–O2P	(3/5)
			N2(G)–N7(A)	(13/14)

d_2-Tl	j + 1	Frequency	j + 2	Frequency
j − 1	O2′-N7 (R)	(8/10)	N1 (G)–O2P	(3/3)
	O2′-N6/O6 (R)[e]	(0/10)	N2 (G)–O5′	(3/3)
			N3 (U)–O2P	(7/7)

[a]Hydrogen bonds are determined by geometry (3.4 Å cut-off and reasonable angles).
[b]This field indicates hydrogen bonding interactions between O2′ atoms of residue j − 1 (right column) and N7 atoms of purines at residue j + 1 (top row).
[c]Twenty of 21 s-Tl tetraloops have G or A at position j + 1. Nineteen of these show a hydrogen bond from the O2′ of residue j − 1 to the N7 of residue j + 1.
[d]Fourteen of 14 s-Tl tetraloops with G at j − 1 show a hydrogen bond from either the N1 or the N2 of G(j − 1) to the O2P of residue j + 2, or both.
[e]The hydrogen bond from the O2′ of residue j − 1 to N6/O6 (R), frequency is 13/20, in s-Tl tetraloops is not observed in d_2-Tl tetraloops.

deletion at the j + 2 position is correlated with adjacent helical distortions such as insertions at position 3 (314, 625, 1387, 1992), clipping at position 2 (1187, 1809, 2598), base pair disruption in the stem (1500, 1596) and unhinging (1749).

i_2-Tl. In the three members of this group, a residue (i_2) is inserted at position 2, between residues (j + 1) and (j + 2). It should be noted that insertions in tetraloops are evident in the results of Huang *et al.* (46). Intra-Group Similarity: the RMSD-AP is 0.42 Å for the three i_2-Tl tetraloop backbone atoms [residues (j − 1), (j), (j + 1) and (j + 2), omitting the inserted residues, which show variable positions]. Inter-Group Similarity: the RMSD-AP is 0.76 Å for the common backbone atoms of three i_2-Tl tetraloops and the 21 s-Tl tetraloops. All three members of the i_2-Tl group show the consensus j − 1 O2′ to j + 1 N7 hydrogen bond. One of them (1707) shows hydrogen bonds of j − 1 N1(G) to the O2P of residue i_2 and j − 1 N2(G) to j + 2 N7(A). A second (1276) shows a contact distance just slightly greater than our hydrogen bond cut-off between (j − 1) N3 and i_2 O1P. Two, with pyrimidines at the j − 1 position, show hydrogen bonds of O2 (j − 1) to N6 (j + 2). Therefore, insertion of a residue at position 2 does not appreciably change the atomic positions or significantly alter the nature of the interactions.

$d_2i_2(3)$-Tl. In this tetraloop, as in the d_2-Tl group, residue (j + 2) is deleted. In addition, three residues are also inserted at position 2 [indicated by $i_2(3)$]. This tetraloop demonstr-ates deletion simultaneously with multi-residue insertion. Inter-Group Similarity: the RMSD-AP is 0.30 Å for common backbone atoms [residues (j − 1), (j) and (j + 1)] of the
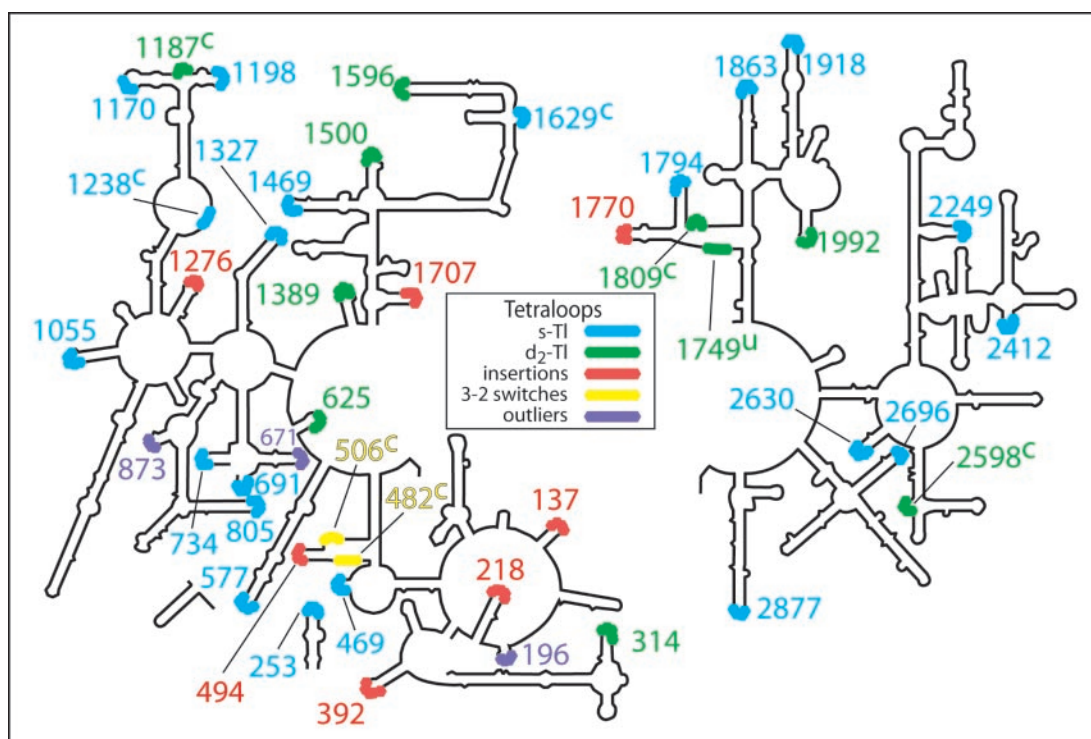

Figure 5. Secondary structure of the HM 23S rRNA (1JJ2). Tetraloop locations and type are indicated by color. Superscripted c's indicate strand clipped tetraloops. A superscripted u indicates the unhinged tetraloop. The strand clipped tetraloops are in contexts in which they do not cap helical stems, as can be inferred from the secondary structure, but do cap pseudohelical stems. The unhinged tetraloop crowns a cavity. Entries 196, 671, 873 were described by Huang (46). These were not detected by our methods and are outliers in conformation and molecular interactions.

$d_2i_2(3)$-Tl tetraloop and the d_2-Tl group. The $j − 1$ O2′ of $d_2i_2(3)$-Tl interacts with the N7 of $j + 1$. The (U) N3 of $j − 1$ interacts with the O1P of $j + 2$. Therefore, the three residue insertion at site 2 does not appreciably change the atomic positions or interactions of the d_2-Tl tetraloop.

i_1-Tl. In this tetraloop, there is a residue inserted at site 1, between residues (j) and (j + 1). Inter-Group Similarity: the RMSD-AP is 0.87 Å for the backbone atoms of this tetraloop and the i_2-Tl tetraloops (omitting the inserted residues). The RMSD-AP is 0.82 Å for the superimposition of common backbone atoms of i_1-Tl(218) and s-Tl(805). In this tetraloop the consensus O2′ $j − 1$ to N7 and O6 $j + 1$ interactions are observed. Therefore, insertion at position 1 does not appreciably change the atomic positions or molecular interactions of this tetraloop.

$x_{3,2}$-Tl. In this tetraloop the positions of the bases of residues $j + 2$ and $j + 3$ are exchanged. Inter-Group Similarity: the RMSD-AP is 0.27 Å for the bases of the $x_{3,2}$-Tl(482) and the bases of s-Tl(1863). Since single residue topology variation is one of the most unexpected discoveries of the multi-resolution method, we provide an illustration of this superimposition (Figure 7). For the superimposition and the RMSD-AP calculation, the ordering of the residues is switched such that $(j − 1)$, (j), (j + 1), (j + 3) of $x_{3,2}$-Tl(482) were superimposed on $(j − 1)$, (j), (j + 1), (j + 2) of s-Tl(1863). We chose tetraloop s-Tl(1863) for this superimposition because it is the only standard tetraloop with the appropriate sequence. In $x_{3,2}$-Tl(482) the consensus O2′ $j − 1$ to N7 and N6 $j + 1$ interactions are observed. In addition the N1 $(j − 1)$ to O2P $(j + 2$, which has replaced $j + 3)$ interaction is maintained. Finally, the N2 $(j − 1)$ to N7 of $j + 3$ (which has replaced $j + 2)$ interaction is conserved. In sum, the positions of the bases and the interactions between them and with the backbone are highly conserved even though the connections linking them differ.

$x_{3,2}i_3$-Tl. In this tetraloop the positions of the bases of residues $j + 2$ and $j + 3$ are exchanged, and in addition, a residue is inserted at position 3. Inter-Group Similarity: the RMSD-AP is 0.45 Å for the common bases of $x_{3,2}i_3$-Tl and s-Tl(691,805,1327,1629), with the base ordering switched
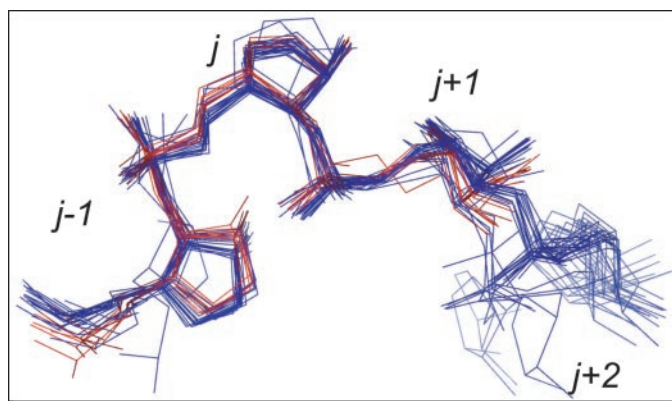
as described above, and the inserted residue omitted. In this group, the topology is the same as $x_{3,2}$-Tl group, and a residue is inserted at position 3, between $(j + 2)$ and $(j + 3)$. In $x_{3,2}i_3$-Tl(506) the consensus O2′ $j − 1$ to N7 $j + 1$ interaction is observed. In addition the N1 $(j − 1)$ to O2P $(j + 2)$ interaction is maintained. Finally, the N2 $(j − 1)$ to N7 of $j + 3$ (which has replaced $j + 2$) interaction is conserved. Therefore the 3-2 switch can accommodate insertions.

d_1i_0-Tl. In these two tetraloops, residue i_0 is inserted between residues $(j − 1)$ and (j) and residue $(j + 1)$ is deleted [Entries 196, 671, 873 were described by Huang (46). These were not detected by our methods and are outliers in conformation and molecular interactions]. This group is equivalent to the previously described UNCG tetraloop (5,6,18–21). The 'looped out' N-residue of UNCG is equivalent to i_0. Intra-Group Similarity: the RMSD-AP is 0.41. Inter-Group Similarity: RMSD-AP is 1.11 Å, for common backbone atoms of the two d_1i_0-Tl tetraloops and s-Tl(805) (which is an average s-Tl tetraloop). It is not clear where this group fits in the family tree (Figure 4) because the cross-loop hydrogen bonding pattern is slightly different from the consensus of other tetraloops. The hydrogen bond from the O2′ of residue $j − 1$ is with the O6 of $j + 2$, not the N7 of $j + 1$, which is deleted from this group of tetraloops. In addition the O2 of $j − 1$ forms a hydrogen bond with the N1 of $j + 2$, which is a G in both members. It is conceivable that further analysis will lead to reassignment of this group to a new position in the family tree or its removal altogether.
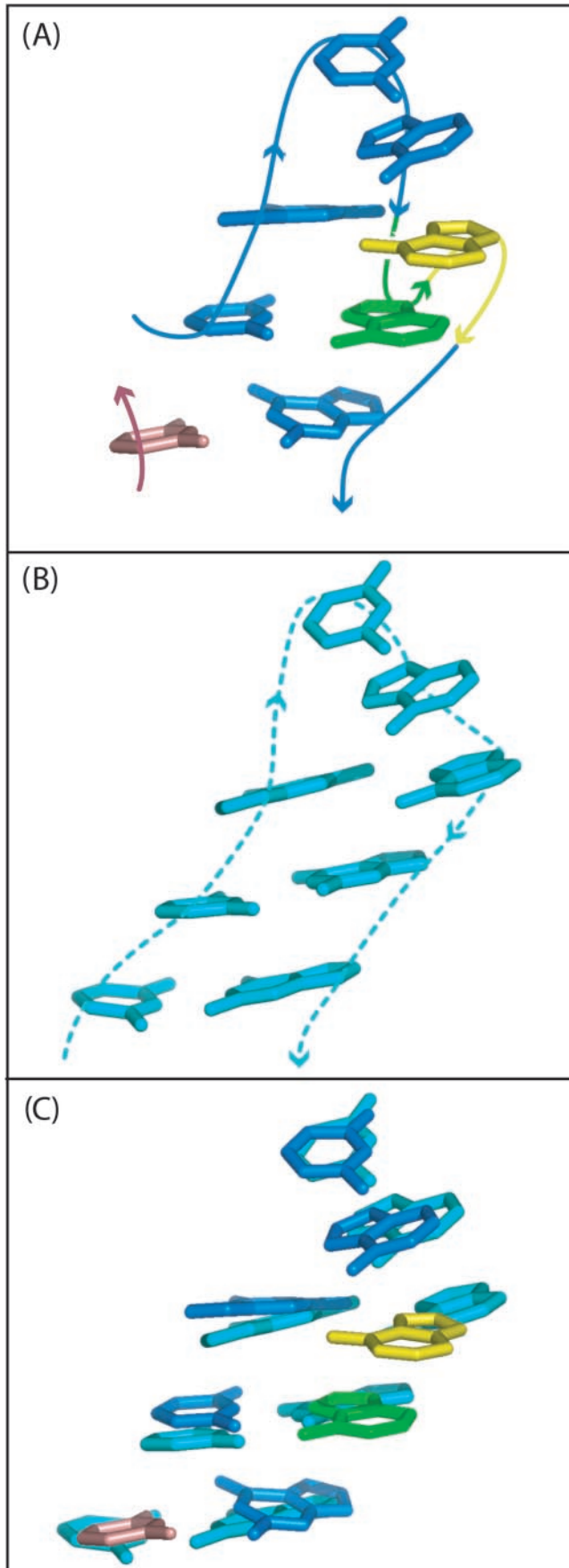
DISCUSSION

On one level the results here correspond well with expectations, confirming that tetraloops have well-defined conformation (given by atomic positions and torsion angles) and molecular interactions (hydrogen bonding and stacking), and sequence constraints. However we arrive at several conclusions that extend or even contradict previous work.

DevLS

We propose a classification scheme where all tetraloops, U-turns and many triloops, pentaloops, etc. are members of a common class (motif) that are elaborated with DevLS— insertions, deletions, strand clips and 3-2 switches. This simplifying scheme can be applied generally to RNA motifs (kink-turns, E-loops, etc.). In fact we observe an E-loop motif in 1JJ2 with two strand clips (residues 911–914, 1045, 1069–1072 and 1293–1294). The commonality of the DevLS between various motifs provides a powerful analytical handle for RNA analysis. One can precisely decompose and describe both polymorphism and the underlying elemental motifs. Approximately one third of the tetraloops in HM 23S rRNA contain DevLS. This significant fraction of tetraloops was not detected in our prior work (43) where DevLS masked tetraloops. The 3-2 switch is, to our knowledge, a previously unrecognized conformational element of RNA. We are not, however, the first to observe insertions, deletions and strand clips. Insertions in tetraloops are evident in the results of Huang *et al.* (46). Deletions in tetraloops give the U-turn motif, and some members of LPTL motif of Lee (36).

Figure 6. Superimposition of 31 tetraloops. The backbone atoms of the first three residues of all s-Tl and d_2-Tl were superimposed. Bases are omitted for clarity.

Insertions, deletions and strand clips in kink-turns and C-like motifs have been noted (40,47).

The 3-2 switch

The 3-2 switch (Figures 1, 3C and 7) re-orders bases such that the effective sequence differs to the primary sequence. Bases with an ordering of 1,2,3,4 in the primary sequence can rearrange, without breaking or altering bonds, to establish a three-dimensional ordering of 1,3,2,4. In a 3-2 switch the RNA backbone skips over one base, then returns to it, then proceeds on in the original direction. We observe three 3-2 switches in the HM 23S rRNA. Two are associated with tetraloops. One is associated with a clipped kink-turn (residues 42–50, 111–115 and 148–149). In addition there are several partial 3-2 switches in which bases 1,3,2 but not 4 are aligned. In sum, RNA accommodates topology variations on the dinucleotide level, whereby the positions and interactions of a series of bases can remain essentially unaltered while the backbone connection linking them varies. We believe 3-2 switches, by partially decoupling covalent sequence from effective sequence, may have significant implications in structure, reactivity and mechanism of evolutionary change.

Tetraloop triplets

The 3-2 switch appears to facilitate tetraloop–tetraloop interactions. We observe that three tetraloops in 1JJ2 associate to form a tetraloop triplet. This tetraloop triplet consists of tetraloops $x_{3,2}$-Tl(482), $x_{3,2}i_3$-Tl(506) and d_2-Tl(314). Tetraloops 482 and 506, which both contain 3-2 switches, associate via an intimate face-to-face interface, which includes base pairing interactions of A(486) with A(511)–each of these is a component of a 3-2 switch. s-Tl(314) stacks on the other two, such that all three j-residues interact. We observe a similar tetraloop triplet in the 23S rRNA of *Deinococcus radiodurans* [1NKW, ref. (48)]. In that structure, tetraloop $x_{3,2}$-Tl (487) and $x_{3,2}i_3$-Tl(510) associate via an intimate face-to-face dimer, which stacks on d_2-Tl(318). We hypothesize that tetraloop triplets play important roles in rRNA folding and stability.

The tetraloop family tree

This tree provides a general, accurate and accessible description of tetraloops and of the relationships among them. The structure-based tree assumes that all tetraloops are members of a single motif class that varies by elaboration with DevLS. To form the tree, the forty observed tetraloops are split first in standard topology and 3-2 switch groups, and are further split by deletions and insertions, according to DevLS positions. Alternative trees with different branching schemes are possible. The tree allows one to readily observe frequencies, relationships between DevLS type and sequence, etc. There

Figure 7. Base positions are conserved in standard and 3-2 switched tetraloops. (**A**) A tetraloop with a 3-2 Switch ($x_{3,2}$-Tl tetraloop 482). The backbone connectivity is indicated by the arrows. The positions of residue j+2 (green) and j+3 (yellow) are switched relative to standard tetraloop. This tetraloop is clipped between residues j − 2 and j − 3. (**B**) A standard tetraloop (s-Tl tetraloop1863), with standard backbone connectivity. (**C**) Superimposition of the bases of the 3-2 Switch and the standard tetraloops. Backbone atoms were not used in the superimposition and are omitted from the diagram for clarity. All bases shown were used for the superimposition.

Table 2. Comparison of LPTL (36) and d_2-Tl tetraloops

	Class[a]	Number[b]	Number[c]	Group[d]
In common	IA	313:317	314	d_2-Tl
	IA	624:628	625	d_2-Tl
	IA	1388:1392	1389	d_2-Tl
	IB	1186:1190	1187	d_2-Tl
	IB	1808:1912	1809	d_2-Tl
	IB	2597:2601	2598	d_2-Tl
	IIA	505:509	506	$x_{3,2}i_3$-Tl
	IA	481:485	482	$x_{3,2}$-Tl
	IIB	482:486	482	$x_{3,2}$-Tl
Identified in present work, not in Lee *et al.* (36)	Three residues inserted at insertion site 2		392	$d_2i_2(3)$-Tl
	No LP; no residue $j − 2$ and $j + 3$ unpaired[e]		1500	d_2-Tl
	No LP; no residue $j − 2$ and $j + 3$ unpaired		1596	d_2-Tl
	No LP; no residue $j − 2$ and $j + 3$ unpaired		1749u[f]	d_2-Tl
	No LP; no residue $j − 2$ and $j + 3$ unpaired		1992	d_2-Tl
Identified in Lee *et al.* (36) not in present work	IB	125:129		Variant conformation[g]
	IB	335:339		Variant conformation[g]
	IIB	326:330		Variant conformation[g]
	IB	1651:1655		Variant conformation[g]
	IB	1966:1970		Variant conformation[g]
	IB	2482:2486		Variant conformation[g]

[a]Classification scheme of LPTLs (36).
[b]Initial and final residue numbers from PDB entry 1JJ2.
[c]Residue number of position $j − 1$ (Figure 3) from PDB entry 1JJ2.
[d]Tetraloop group (Figures 3 and 4), indicating DevLS position and type.
[e]LP indicates lonepair.
[f]The letter 'u' indicates an unhinged d_2-Tl tetraloop.
[g]The conformational states and molecular interaction of these loops are not similar to those of tetraloops.

are many possible family trees. In fact we believe that it may be appropriate, if one were to ignore history, to recast the 10 d_2-Tl tetraloops, which have the greatest conservation of sequence and atomic positions, as the parent motif. In this scheme the current s-Tl group would contain an insertion after residue $j + 1$. With additional data, a more statistically meaningful tree may allow one to infer evolutionary relationships and mechanisms.

Deleted tetraloops, U-turns and LPTLs

The consensus hydrogen bonding interactions and sequence of s-Tl and d_2-Tl tetraloops are consistent with the U-turn motif (16,49–51). The d_2-Tl tetraloop appears to be essentially identical to the original U-turn of Quigley and Rich (51). Gutell and coworkers (36) have used sequence covariation approaches along with visual inspection to detect and describe a motif they refer to as the LPTL. There is considerable overlap of the LPTL motif of Gutell with the d_2-Tl group described here (Table 2). However important distinctions distinguish the two motifs. The d_2-Tl group is characterized by conserved conformation (torsion angles and atomic positions) and molecular interactions, which are also common to the s-Tl group and other tetraloops. In contrast some members of the LPTL group are conformationally distinct from others, and from standard tetraloops. Some d_2-Tl tetraloops lack closing base pairs altogether, and so are not consistent with the LPTL definition.

Variation in the helix capping function of tetraloops

Here, seven of forty tetraloops are strand clipped (Figure 5). Strand clipping allows RNA segments that are remote in the primary sequence to join to form a motif (36,47,52). Strand clipped tetraloops cap pseudo-helices, which commonly do not appear as stems in secondary structure representations. One observed tetraloop caps neither a helix nor a pseudo-helix, but by all other criteria is an average d_2-Tl tetraloop. This tetraloop is 'unhinged' from any helical regions. None of the d_2-Tl tetraloops cap a clean unperturbed helix. In sum, a 'tetraloop' is not necessarily a terminal loop, which by classical definition allows a strand of RNA to fold back on itself to form a helical stem (2).

REFERENCES

1. Chworos,A., Severcan,I., Koyfman,A.Y., Weinkam,P., Oroudjev,E., Hansma,H.G. and Jaeger,L. (2004) Building programmable jigsaw puzzles with RNA. *Science*, **306**, 2068–2072.
2. Moore,P.B. (1999) Structural motifs in RNA. *Annu. Rev. Biochem.*, **68**, 287–300.
3. Woese,C.R., Gutell,R., Gupta,R. and Noller,H.F. (1983) Detailed analysis of the higher-order structure of 16S-like ribosomal ribonucleic acids. *Microbiol. Rev.*, **47**, 621–669.
4. Woese,C.R., Winker,S. and Gutell,R.R. (1990) Architecture of ribosomal-RNA: constraints on the sequence of tetra-loops. *Proc. Natl Acad. Sci. USA*, **87**, 8467–8471.
5. Tuerk,C., Gauss,P., Thermes,C., Groebe,D.R., Gayle,M., Guild,N., Stormo,G., d'Aubenton-Carafa,Y., Uhlenbeck,O.C., Tinoco,I.Jr *et al.* (1988) CUUCGG hairpins: extraordinarily stable RNA secondary structures associated with various biochemical processes. *Proc. Natl Acad. Sci. USA*, **85**, 1364–1368.
6. Cheong,C., Varani,G. and Tinoco,I.Jr (1990) Solution structure of an unusually stable RNA hairpin, 5′ GGAC(UUCG)GUCC. *Nature*, **346**, 680–682.
7. Varani,G., Cheong,C., Tinoco,I.Jr and , Wimberly,B. (1991) Structure of an unusually stable RNA hairpin: conformation and dynamics of an RNA internal loop. *Biochem.*, **30**, 3280–3289.
8. Antao,V.P. and Tinoco,I.Jr (1992) Thermodynamic parameters for loop formation in RNA and DNA hairpin tetraloops. *Nucleic Acids Res.*, **20**, 819–824.
9. Selinger,D., Liao,X. and Wise,J.A. (1993) Functional interchangeability of the structurally similar tetranucleotide loops GAAA and UUCG in fission yeast signal recognition particle RNA. *Proc. Natl Acad. Sci. USA*, **90**, 5409–5413.
10. Michel,F. and 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.
11. Jaeger,L., Michel,F. and Westhof,E. (1994) Involvement of a GNRA tetraloop in long-range tertiary interactions. *J. Mol. Biol.*, **236**, 1271–1276.
12. Puglisi,J.D., Tan,R., Calnan,B.J., Frankel,A.D. and Williamson,J.R. (1992) Conformation of the tar RNA-arginine complex by NMR spectroscopy. *Science*, **257**, 76–80.
13. Cate,J.H., Gooding,A.R., Podell,E., Zhou,K., Golden,B.L., Kundrot,C.E., Cech,T.R. and Doudna,J.A. (1996) Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science*, **273**, 1678–1685.
14. Jucker,F.M., Heus,H.A., Yip,P.F., Moors,E.H. and Pardi,A. (1996) A network of heterogeneous hydrogen bonds in GNRA tetraloops. *J. Mol. Biol.*, **264**, 968–980.

15. Butcher,S.E., Dieckmann,T. and Feigon,J. (1997) Solution structure of a GAAA tetraloop receptor RNA. *EMBO J.*, **16**, 7490–7499.
16. Jucker,F.M. and Pardi,A. (1995) GNRA tetraloops make a U-turn. *RNA*, **1**, 219–222.
17. Correll,C.C. and 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.
18. Akke,M., Fiala,R., Jiang,F., Patel,D. and Palmer,A.G.,IIIrd (1997) Base dynamics in a UUCG tetraloop RNA hairpin characterized by 15N spin relaxation: correlations with structure and stability. *RNA*, **3**, 702–709.
19. Ennifar,E., Nikulin,A., Tishchenko,S., Serganov,A., Nevskaya,N., Garber,M., Ehresmann,B., Ehresmann,C., Nikonov,S. and Dumas,P. (2000) The crystal structure of UUCG tetraloop. *J. Mol. Biol.*, **304**, 35–42.
20. Williams,D.J. and Hall,K.B. (1999) Unrestrained stochastic dynamics simulations of the UUCG tetraloop using an implicit solvation model. *Biophys. J.*, **76**, 3192–3205.
21. Allain,F.H. and Varani,G. (1995) Structure of the p1 helix from group I self-splicing introns. *J. Mol. Biol.*, **250**, 333–353.
22. Jucker,F.M. and Pardi,A. (1995) Solution structure of the CUUG hairpin loop: a novel RNA tetraloop motif. *Biochem.*, **34**, 14416–14427.
23. Baumruk,V., Gouyette,C., Huynh-Dinh,T., Sun,J.S. and Ghomi,M. (2001) Comparison between CUUG and UUCG tetraloops: thermodynamic stability and structural features analyzed by UV absorption and vibrational spectroscopy. *Nucleic Acids Res.*, **29**, 4089–4096.
24. Yang,H., Jossinet,F., Leontis,N., Chen,L., Westbrook,J., Berman,H. and Westhof,E. (2003) Tools for the automatic identification and classification of RNA base pairs. *Nucleic Acids Res.*, **31**, 3450–3460.
25. Leontis,N.B. and Westhof,E. (2003) Analysis of RNA motifs. *Curr. Opin. Struct. Biol.*, **13**, 300–308.
26. Waugh,A., Gendron,P., Altman,R., Brown,J.W., Case,D., Gautheret,D., Harvey,S.C., Leontis,N., Westbrook,J., Westhof,E. *et al.* (2002) RNAML: a standard syntax for exchanging RNA information. *RNA*, **8**, 707–717.
27. Leontis,N.B., Stombaugh,J. and Westhof,E. (2002) Motif prediction in ribosomal RNAs lessons and prospects for automated motif prediction in homologous RNA molecules. *Biochimie*, **84**, 961–973.
28. Leontis,N.B. and Westhof,E. (2001) Geometric nomenclature and classification of RNA base pairs. *RNA*, **7**, 499–512.
29. Lemieux,S. and Major,F. (2002) RNA canonical and non-canonical base pairing types: a recognition method and complete repertoire. *Nucleic Acids Res.*, **30**, 4250–4263.
30. Lee,J.C. and Gutell,R.R. (2004) Diversity of base-pair conformations and their occurrence in rRNA structure and RNA structural motifs. *J. Mol. Biol.*, **344**, 1225–1249.
31. Olson,W.K. (1975) Configuration statistics of polynucleotide chains. A single virtual bond treatment. *Macromolecules*, **8**, 272–275.
32. Duarte,C.M. and Pyle,A.M. (1998) Stepping through an RNA structure: a novel approach to conformational analysis. *J. Mol. Biol.*, **284**, 1465–1478.
33. Duarte,C.M., Wadley,L.M. and Pyle,A.M. (2003) RNA structure comparison, motif search and discovery using a reduced representation of RNA conformational space. *Nucleic Acids Res.*, **31**, 4755–4761.
34. Gutell,R.R., Noller,H.F. and Woese,C.R. (1986) Higher order structure in ribosomal RNA. *EMBO J.*, **5**, 1111–1113.
35. Levitt,M. (1969) Detailed molecular model for transfer ribonucleic acid. *Nature*, **224**, 759–763.
36. Lee,J.C., Cannone,J.J. and Gutell,R.R. (2003) The lonepair triloop: a new motif in RNA structure. *J. Mol. Biol.*, **325**, 65–83.
37. Monge,A., Lathrop,E.J., Gunn,J.R., Shenkin,P.S. and Friesner,R.A. (1995) Computer modeling of protein folding: conformational and energetic analysis of reduced and detailed protein models. *J. Mol. Biol.*, **247**, 995–1012.
38. Betancourt,M.R. (2003) A reduced protein model with accurate native-structure identification ability. *Proteins*, **53**, 889–907.
39. Mallat,S. (1998) *A Wavelet Tour of Signal Processing*, *2nd edn.* Academic Press, NY.
40. Klein,D.J., Schmeing,T.M., Moore,P.B. and Steitz,T.A. (2001) The kink-turn: a new RNA secondary structure motif. *EMBO J.*, **20**, 4214–4221.
41. Ban,N., Nissen,P., Hansen,J., Moore,P.B. and Steitz,T.A. (2000) The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science*, **289**, 905–920.
42. Murray,L.J., Arendall,W.B.,IIIrd, Richardson,D.C. and Richardson,J.S. (2003) RNA backbone is rotameric. *Proc. Natl Acad. Sci. USA*, **100**, 13904–13909.
43. Hershkovitz,E., Tannenbaum,E., Howerton,S.B., Sheth,A., Tannenbaum,A. and Williams,L.D. (2003) Automated identification of RNA conformational motifs: theory and application to the HM LSU 23S rRNA. *Nucleic Acids Res.*, **31**, 6249–6257.
44. Pley,H.W., Flaherty,K.M. and McKay,D.B. (1994) Three-dimensional structure of a hammerhead ribozyme. *Nature*, **372**, 68–74.
45. Elgavish,T., Cannone,J.J., Lee,J.C., Harvey,S.C. and Gutell,R.R. (2001) Aa.Ag@helix.Ends: A:A and a:G base-pairs at the ends of 16S and 23S rRNA helices. *J. Mol. Biol.*, **310**, 735–753.
46. Huang,H.C., Nagaswamy,U. and Fox,G.E. (2005) The application of cluster analysis in the intercomparison of loop structures in RNA. *RNA*, **11**, 412–423.
47. Lescoute,A., Leontis,N.B., Massire,C. and Westhof,E. (2005) Recurrent structural RNA motifs, isostericity matrices and sequence alignments. *Nucleic Acids Res.*, **33**, 2395–2409.
48. Harms,J., Schluenzen,F., Zarivach,R., Bashan,A., Gat,S., Agmon,I., Bartels,H., Franceschi,F. and Yonath,A. (2001) High resolution structure of the large ribosomal subunit from a mesophilic eubacterium. *Cell.*, **107**, 679–688.
49. Auffinger,P. and Westhof,E. (1999) Singly and bifurcated hydrogen-bonded base-pairs in tRNA anticodon hairpins and ribozymes. *J. Mol. Biol.*, **292**, 467–483.
50. Gutell,R.R., Cannone,J.J., Konings,D. and Gautheret,D. (2000) Predicting U-turns in ribosomal RNA with comparative sequence analysis. *J. Mol. Biol.*, **300**, 791–803.
51. Quigley,G.J. and Rich,A. (1976) Structural domains of transfer RNA molecules. *Science*, **194**, 796–806.
52. Nagaswamy,U. and Fox,G.E. (2002) Frequent occurrence of the T-loop RNA folding motif in ribosomal RNAs. *RNA*, **8**, 1112–1119.
53. Schneider,T.D. and Stephens,R.M. (1990) Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res.*, **18**, 6097–6100.