

## REVIEW

## The kink-turn in the structural biology of RNA

Lin Huang and David M. J. Lilley\*

Cancer Research UK Nucleic Acid Structure Research Group, MSI/WTB Complex, The University of Dundee, Dow Street, Dundee DD1 5EH, UK

Quarterly Reviews of Biophysics (2018), 51, e5, page 1 of 32 doi:10.1017/S0033583518000033

**Abstract.** The kink-turn (k-turn) is a widespread structural motif found in functional RNA species. It typically comprises a three-nucleotide bulge followed by tandem *trans* sugar edge-Hoogsteen G:A base pairs. It introduces a sharp kink into the axis of duplex RNA, juxtaposing the minor grooves. Cross-strand H-bonds form at the interface, accepted by the conserved adenine nucleobases of the G:A basepairs. Alternative acceptors for one of these divides the k-turns into two conformational classes N3 and N1. The base pair that follows the G:A pairs (3b:3n) determines which conformation is adopted by a given k-turn. k-turns often mediate tertiary contacts in folded RNA species and frequently bind proteins. Common k-turn binding proteins include members of the L7Ae family, such as the human 15-5k protein. A recognition helix within these proteins binds in the widened major groove on the outside of the k-turn, that makes specific H-bonds with the conserved guanine nucleobases of the G:A pairs. L7Ae binds with extremely high affinity, and single-molecule data are consistent with folding by conformational selection. The standard, simple k-turn can be elaborated in a variety of ways, that include the complex k-turns and the k-junctions. In free solution in the absence of added metal ions or protein k-turns do not adopt the tightly-kinked conformation. They undergo folding by the binding of proteins, by the formation of tertiary contacts, and some (but not all) will fold on the addition of metal ions. Whether or not folding occurs in the presence of metal ions depends on local sequence, including the 3b:3n position, and the -1b:-1n position (5' to the bulge). In most cases -1b:-1n = C:G, so that the 3b:3n position is critical since it determines both folding properties and conformation. In general, the selection of these sequence matches a given k-turn to its biological requirements. The k-turn structure is now very well understood, to the point at which they can be used as a building block for the formation of RNA nano-objects, including triangles and squares.

1. An introduction to the kink-turn (k-turn) in RNA structure 2
2. The occurrence of k-turns 2
3. The structure of a standard k-turn 4
4. Two conformational classes adopted by k-turns 4
  - 4.1. *HmKt-7* can exist in both N3 and N1 conformations 5
5. The conformational class is determined by the 3b,3n sequence 6
6. Variations on the structural theme of the standard k-turn 8
  - 6.1. Non-standard, simple k-turns 8
  - 6.2. Complex k-turns 10
  - 6.3. k-junctions 11
  - 6.4. To be or not to be a k-turn 12
  - 6.5. Identification of k-turns in RNA sequence and structure 14
7. The folding of k-turns 14
  - 7.1. Folding of k-turns on the addition of metal ions 16
  - 7.2. Folding of k-turns on the binding of proteins 16
  - 7.3. Folding of k-turns induced by tertiary interactions 16

\* Author for correspondence: David M. J. Lilley, Cancer Research UK Nucleic Acid Structure Research Group, MSI/WTB Complex, The University of Dundee, Dow Street, Dundee DD1 5EH, UK. Tel: (+44)-1382-384243; Email: [d.m.j.lilley@dundee.ac.uk](mailto:d.m.j.lilley@dundee.ac.uk)



<b>8. Recognition of k-turns by proteins of the L7Ae family</b>	<b>16</b>
<b>9. The process of binding k-turns by proteins of the L7Ae family</b>	<b>19</b>
<b>10. Sequence dependence of ion-dependent folding</b>	<b>20</b>
10.1. The 3b:3n sequence is a critical determinant of folding properties in metal ions	20
10.2. A molecular explanation for the 3n = G folding rule	21
10.3. The -1b:-1n sequence is another important determinant of folding properties in metal ions	21
<b>11. The distribution of sequences for natural k-turn elements</b>	<b>22</b>
11.1. Application of the folding rules to a set of k-turns of unknown structure	24
<b>12. Effect of N<sup>6</sup>-methyladenine inclusion on k-turn folding of snoRNAs</b>	<b>25</b>
<b>13. The k-turn as a building unit in nano-construction</b>	<b>25</b>
13.1. Association of a two-k-turn unit in crystal lattices	25
13.2. The molecular structure of a six-k-turn nano-object	27
<b>14. In conclusion</b>	<b>28</b>
<b>Acknowledgements</b>	<b>28</b>
<b>References</b>	<b>28</b>

## 1. An introduction to the kink-turn (k-turn) in RNA structure

The k-turn is a common structural motif in many functional RNA species, that introduces a sharp bend into the axis of duplex RNA. In most cases, k-turn structures act as binding sites for specific proteins and frequently mediate short- or long-range tertiary contacts. The secondary structures of medium to large RNA species can be reduced to double-helical sections connected by junction elements that alter the axial trajectory and allow tertiary interactions to occur. The k-turn is probably the most common such element in RNA. The structural and folding characteristics of k-turns and their dependence on local sequence have been studied extensively, such that they are arguably the best understood of the motifs from which RNA is built.

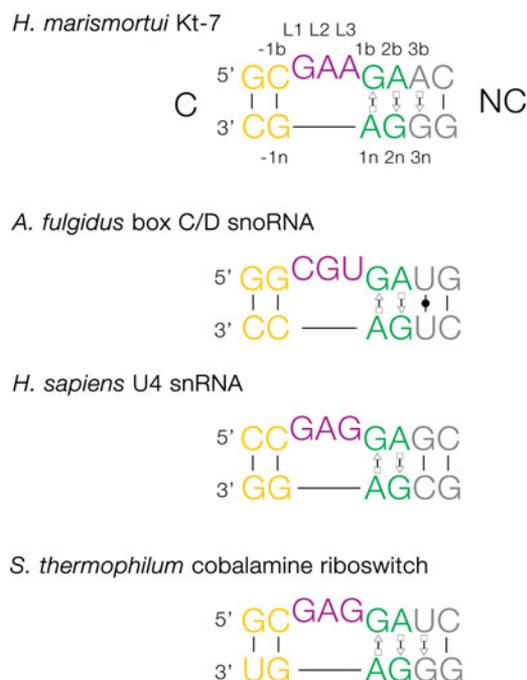
The basic, standard k-turn comprises a (usually) three-nucleotide bulge followed on the 3' side by tandem G:A and A:G pairs (Fig. 1). The helices 5' and 3' to the loop are termed the C- and NC-helices, respectively (names derived from the presence of canonical or non-canonical base pairing). The standard nomenclature (Liu & Lilley, 2007) for the nucleotide positions numbers outwards from the ends of the loop, positively in the 3' direction and negatively in the 5' direction. The nucleotides located on the loop-containing strand (that is normally drawn as the top strand) take the suffix *b*, and those on the other strand take the suffix *n*. The nucleotides of the loop are named L1, L2 and so on numbered from the 5' end. In a subset of k-turns, termed k-loops, the C helix is replaced by a terminal loop with no formal base-pairing.

k-turns kink the helical axis by juxtaposing the minor grooves of the NC and C helices, with the major groove displayed on the outer face of the structure. The helical axes are not exactly coplanar, but there is an included angle of 50° in projection. This angle is quite tightly constrained by the hydrogen bonds that form in the core, and thus the angle between the axes for the known k-turns occupies a narrow range between 45° and 55° (Daldrop & Lilley, 2013).

The structure is stabilized by important A-minor H-bonding interactions between the conserved adenine nucleobases of the tandem G•A pairs and 2-hydroxyl groups of the C-helix and loop. These are discussed further in section 3.

## 2. The occurrence of k-turns

k-turns are extremely widespread, being found in most classes of functional RNA in all domains of life. They were identified as a common sequence element in RNA (Winkler *et al.* 2001), but first recognized as a recurring structural motif when Steitz and coworkers solved the crystal structure of the large ribosomal subunit of the archaeon *Haloarcula marismortui* (Ban *et al.* 2000; Klein *et al.* 2001). k-turns are also found in the small ribosomal subunit (Wimberly *et al.* 2000), and in the ribosomes of bacteria (Clemons *et al.* 1999; Schuwirth *et al.* 2005) and eukaryotes (Ben-Shem *et al.* 2011). Most ribosomal k-turns mediate long- or short-range tertiary contacts and the majority are bound by proteins. Kt-7 of the *H. marismortui* 50S ribosomal



**Fig. 1.** Sequences of some representative simple k-turns and the nomenclature of the nucleotide positions. Four examples of k-turns are shown, drawn from the large ribosomal subunit (Kt-7), box C/D snoRNP, the spliceosomal B complex U4 snRNA and a (cobalamine) riboswitch. The nomenclature is shown for Kt-7. The helix 5' to the loop is called C, and that to the 3' is called NC. The loop nucleotides are designated  $L_n$  numbered from 5' to 3'. For the non-loop nucleotides, those on the loop strand take the suffix  $b$  and those on the other strand take  $n$ . The nucleotides of the NC helix are positively numbered outwards (i.e. 5' to 3' on the loop-containing strand) from the loop, while those of the C helix are negatively numbered also outward from the loop (3' to 5' on the loop-containing strand). When this nomenclature is applied to complex k-turns the nucleotides are numbered according to their positions in the structure rather than their order in the primary sequence. The coloring of the nucleotides is followed throughout this review.

subunit (Klein *et al.* 2001) has a sequence that is close to the consensus for standard k-turns (Fig. 1), and is the best characterized of any k-turn despite the fact that it is atypical in one respect (see section 4.1); we shall refer to this as *HmKt-7* subsequently.

A k-turn plays a role in the assembly of the eukaryotic spliceosome, found in U4 snRNA that is contained within the B-complex that precedes the tri-snRNP complex (Vidovic *et al.* 2000; Wozniak *et al.* 2005). The U4 k-turn binds the 15.5k protein, the eukaryotic ortholog of the L7Ae protein (see section 8). In an ordered assembly process (Nottrott *et al.* 2002) the complex then recruits Prp31 (Liu *et al.* 2007) in a manner closely analogous to the assembly of the box C/D snoRNP discussed below (section 12). This complex can be observed in the recent cryo-EM structure of the spliceosomal B complex of yeast (Plaschka *et al.* 2017). Given the size and complexity of the spliceosome, it seems quite possible there will be further k-turns formed at different stages of the spliceosome cycle.

k-turns are important structural elements in at least seven different riboswitches. The first was discovered in the SAM-I riboswitch (Montange & Batey, 2006), and examples were later found in the structures of the cyclic-diGMP (Smith *et al.* 2009, 2011), cobalamine (Peselis & Serganov, 2012) and T-box (Zhang & Ferre-D'Amare, 2013) riboswitches. Biochemical experiments have also strongly indicated the presence of k-turns in lysine (Blouin & Lafontaine, 2007) and glycine (Baird & Ferre-D'Amare, 2013) riboswitches. Lastly, bacterial and plant thiamine pyrophosphate (TPP) riboswitches have an elaborated form of k-turn called a k-junction (Serganov *et al.* 2006; Thore *et al.* 2006; Wang *et al.* 2014), discussed below in section 6.3. The riboswitches are not generally known to bind proteins, and their role appears to be the mediation of tertiary contacts to create a stable fold (and thereby a ligand binding site) in these relatively small, autonomously folding RNA species.

k-turns play a critical role in the assembly of the box C/D and H/ACA snoRNP structures that carry out guided O<sup>2</sup>'-methylation and pseudouridylation, respectively, of archaeal and eukaryotic RNA (Hamma & Ferré-D'Amaré, 2004; Moore *et al.* 2004; Szewczak *et al.* 2005; Youssef *et al.* 2007). k-turn-forming sequences are also found in U3 snoRNP involved in guided nucleolytic processing of rRNA (Beltrame & Tollervey, 1995; Gerbi *et al.* 2001; Marmier-Gourrier *et al.* 2003). The core of the box C/D snoRNA forms an internal loop, the strands of which provide the 12 nt guide sequences for two target



RNA species. The loop is flanked by the box C and D, and box C' and D' sequences that form k-turns or k-loop structures at each end (Bleichert *et al.* 2009; Kiss-Laszlo *et al.* 1996; Lin *et al.* 2011; Tran *et al.* 2003; Tycowski *et al.* 1996; Watkins *et al.* 2000; Xue *et al.* 2010; Ye *et al.* 2009). The k-turns bind L7Ae-family protein (L7Ae in archaea, 15-5k in eukaryotes) as the first stage of snoRNP assembly (McKeegan *et al.* 2007; Omer *et al.* 2002; Schultz *et al.* 2006; Watkins *et al.* 1998, 2002). This will be discussed further in section 12.

k-turns play a significant role in the control of translation of some mRNA species. A k-turn was found in a stem-loop contained within the pre-mRNA encoding the L30e ribosomal protein of *Saccharomyces cerevisiae* (Mao *et al.* 1999; White *et al.* 2004). L30e is a member of the L7Ae family of proteins, and binds a k-turn within its pre-mRNA, thereby repressing its own translation. It has recently been shown that in archaea L7Ae similarly regulates its own translation by binding to a k-turn within the 5'-UTR of its structural gene *l7ae*, thereby occluding the Shine-Dalgarno sequence (Daume *et al.* 2017). This may be a more general mechanism, as L7Ae has been shown to bind to mRNA encoding the box C/D snoRNP components Nop5 and fibrillarin (Daume *et al.* 2017) (for discussion of box C/D assembly see section 12).

Thus k-turns are key architectural elements involved in some of the most important functions of RNA in the cell. These include translation, various kinds of posttranscriptional modification including splicing, 2'-O-methylation and pseudouridylation, and in the control of gene expression. And the list does not stop there. For example, a k-turn occurs in the ribozyme ribonuclease P that processes the 5'-terminus of tRNA (Lai *et al.* 2017; Reiter *et al.* 2010). Still more k-turns exist of currently unknown function. For example, Breaker and colleagues (Weinberg *et al.* 2017) have recently presented a bioinformatic analysis of 224 candidate structured RNAs found in non-coding regions of bacteria, a number of which contain putative k-turn structures – see section 6.5. Thus it is highly probable that many more are waiting to be discovered, for example in the analysis of long non-coding RNA species.

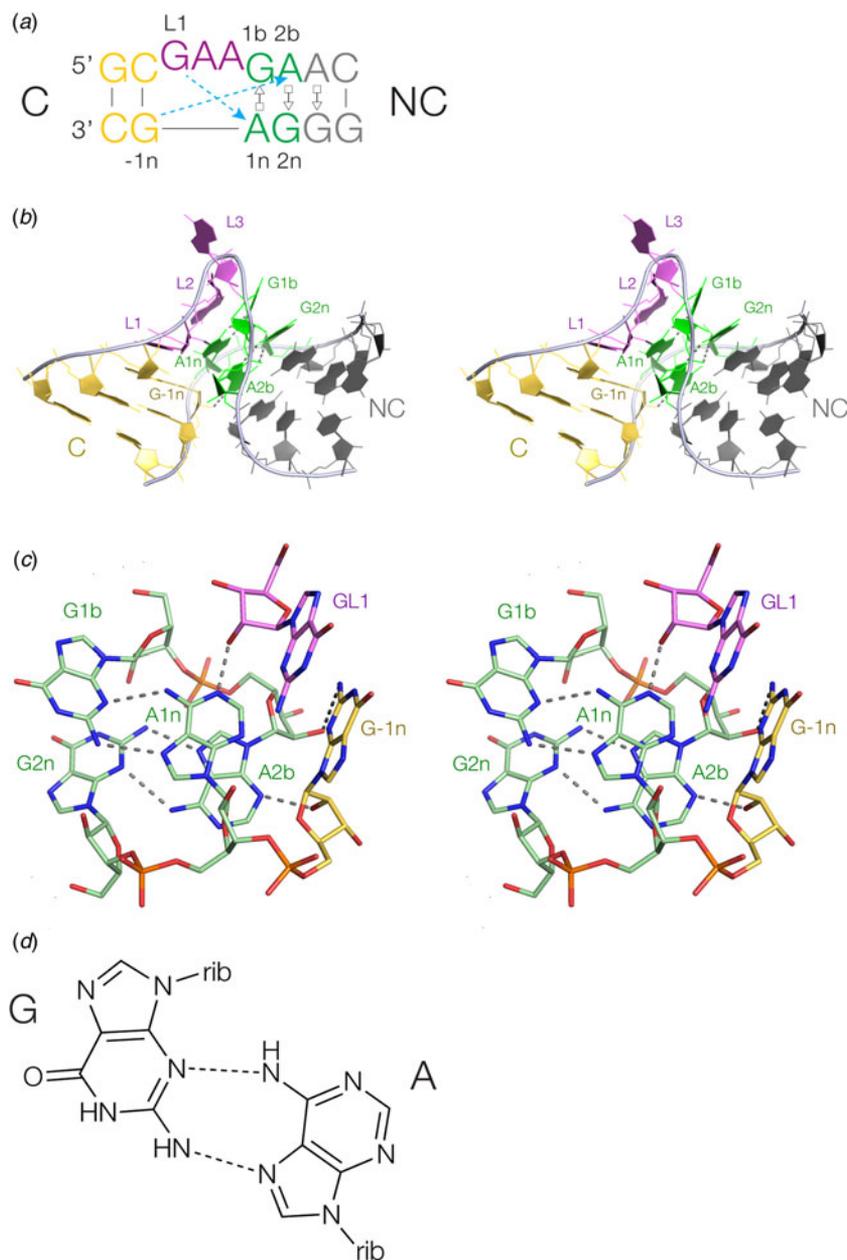
### 3. The structure of a standard k-turn

Tandem G:A, A:G basepairs are the core of the k-turn structure, each forming *trans* G(sugar edge):A(Hoogsteen edge) base pairs (Fig. 2). The 1b:1n pair forms H-bonds from GN2 to AN7, and from AN6 to GN3. The 2b:2n pair is similar, except that in a subset of k-turns called the N1 class (see section 4 below) the AN6 to GN3 is too long to be considered a stable H-bond. The nucleobase of L1 is stacked onto the end of the C-helix, while that of L2 is stacked onto the end of the NC helix. The 1b:1n pair is strongly buckled such that G1b is rotated  $\sim 25^\circ$  out of the plane, and L2 adopts a *syn* conformation about the glycosyl bond in order to maximize stacking on A1n. L3 is directed away from the RNA structure into the solvent, making no interactions within the RNA.

In the folded k-turn structure the minor grooves of the NC and C helices are juxtaposed, with the sugar edges of the conserved adenine nucleobases (1n and 2b) directed towards the minor groove of the C helix. These act as acceptors for critical cross-strand H-bonds donated by 2'-OH groups (Fig. 2). A1n N1 accepts a proton from the O2' of L1 (the 5' nucleotide of the loop) (Lescoute *et al.* 2005; Liu & Lilley, 2007), while A2b N1 or N3 (see the following section) accepts one from the O2' of the nucleotide at  $-1n$  (Daldrop & Lilley, 2013; Reblova *et al.* 2011). These two H-bonds are extremely important to the integrity of the folded structure. Disruption of the L1 to A1n H-bond was shown to prevent the folding of *HmKt-7* in response to the addition of  $Mg^{2+}$  ions (Liu & Lilley, 2007). In k-turn variants that disallow one or other then surrogate equivalents form. For example, *HmKt-58* has a loop of just two nucleotides and thus it is the O2' of the nucleotide in the  $-1b$  position that donates a proton to A1n N1 (Liu & Lilley, 2007). Additional H-bonds form adventitiously in given k-turns. In many k-turns L3 O2' donates an H-bond to the *pro-S* non-bridging O of the L1/L2 phosphate; removal of the L3 2'-OH from *HmKt-7* led to impaired ion-induced folding (Liu & Lilley, 2007).

### 4. Two conformational classes adopted by k-turns

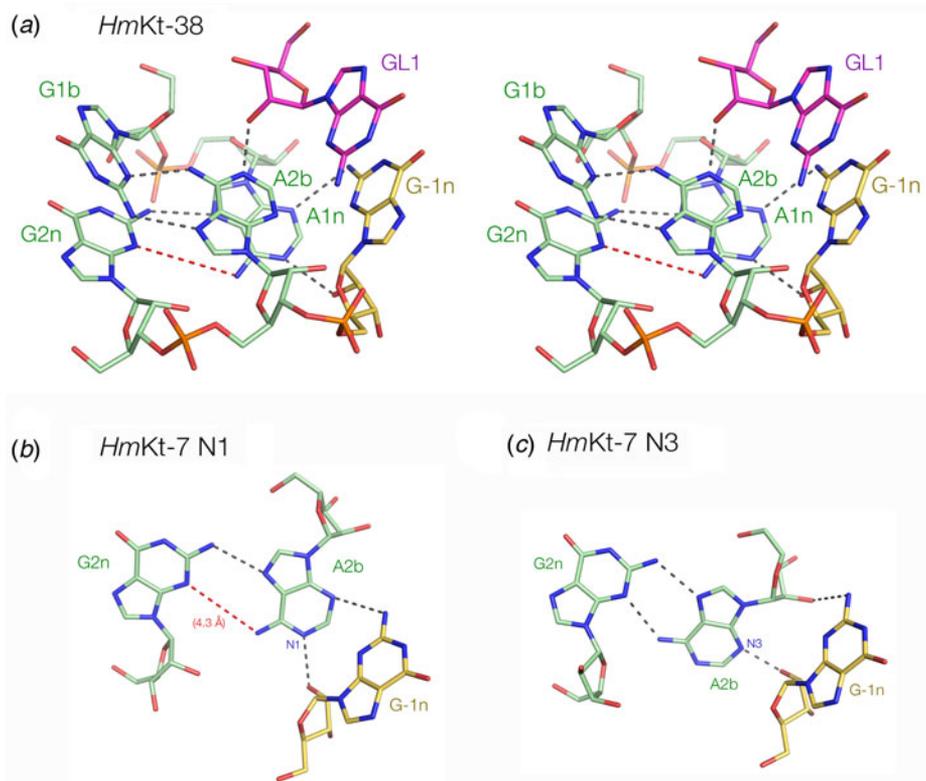
Comparison of the available structures of k-turns reveals that the H-bond donated by  $-1n$  O2' can be accepted by A2b at either N3 or N1 (Fig. 3). This divides the known k-turns into two classes termed N3 or N1, with approximately equal numbers in each group (Table 1). In order to accept a proton at N1 (i.e. adopt the N1 conformation), the nucleobase of A2b must rotate, stretching the A2b N6 to G2n N3 distance typically to  $>4 \text{ \AA}$ , too long to be a stable H-bond. This is compensated in the great majority of N1-class k-turns by the formation of a second H-bond from by G-1n N2 to A2b N3, with good length and geometry. The switch between the N3 and N1 conformations leads to a rotation about the C helix axis (Daldrop & Lilley, 2013), potentially affecting any tertiary interaction that the helix makes.



**Fig. 2.** The structure of a simple k-turn. (a) The sequence of *HmKt-7* with the key cross-strand H-bonds indicated by the broken cyan arrows. These are donated by the 2'-hydroxyl groups of the L1 and  $-1n$  ribose groups, and accepted by the conserved adenine nucleobases at the 1n and 2b positions, respectively. (b) and (c) Parallel-eye stereoscopic image of the structure of *HmKt-7*. The k-turn structure was determined as a duplex at 2.0 Å resolution and is in the N3 conformation (protein data base (PDB) ID 4CS1). The overall structure of the k-turn viewed from the side of the loop is shown in (b), while the G:A basepairs and the cross-strand H-bonds are shown in (c). All H-bonds are shown as broken lines here and throughout the review. (d) The chemical structure of the sheared *trans*-sugar-Hoogsteen G:A basepairs.

#### 4.1 *HmKt-7* can exist in both N3 and N1 conformations

The structural class for a given k-turn is generally invariant, irrespective of its environment. However, there is an interesting exception to this. In the *H. marismortui* large ribosomal subunit Kt-7 adopts the N1 conformation (Ban *et al.* 2000). Yet when taken out of the ribosomal context it invariably adopts the N3 conformation, as free RNA (Daldrop & Lilley, 2013; Huang & Lilley, 2013) or bound by L7Ae protein (Huang & Lilley, 2013). In fact aside from the ribosome, 16 independent structures (some being crystallographically-independent molecules within the same lattice) have been determined for *HmKt-7*, all of which are N3 structures (Huang *et al.* 2016). Thus it seems that the environment of the archaeal ribosome has forced Kt-7 to adopt the N1 conformation even though this is clearly less stable in isolation compared with the N3 structure. In



**Fig. 3.** The N3 and N1 conformations of the k-turn. (a) A parallel-eye stereoscopic image of the structure of the core of the ribosomal k-turn *HmKt-38* core as a representative N1 structure (PDB ID 1FFK). Note that the G1b:A1n basepair has the standard two H-bonds, while the A2b:G2n nucleobases are only connected by a single H-bond from G2nN2 to A2bN7. The distance between A2bN6 and G2nN3 is 4.7 Å (shown by the broken red line) is too long to be stably H-bonded. (b) and (c) A comparison of the H-bonding in the 2n:2b:-1n triple-nucleotide interaction in *HmKt-7* in the N1 (b, in the ribosome; PDB ID 1FFK) and N3 (c, as a duplex PDB ID 4CS1), conformations. Note again the long A2bN6 to G2nN3 distance in the N1 conformation, shown as a broken red line in b.

the ribosome, Kt-7 is contained within a duplex that is organized by a three-way junction and whose terminal loop makes a loop-loop interaction with an adjacent arm of the junction. It is also bound by the L24 protein. Evidently, these interactions increase the relative stability of the N1 conformation in the ribosomal context. There is one other example of conformational switching known. The conformation of the non-standard (see section 6.1) ribosomal k-turn Kt-23 from *Thelohania solenopsae* was determined to be N1 class in the context of the SAM-I riboswitch, but N3 when bound to L7Ae (Huang & Lilley, 2014).

## 5. The conformational class is determined by the 3b:3n sequence

The structural class of most k-turns is intrinsic, unaffected by context. A given k-turn will adopt the same conformation as free RNA or protein bound (Huang & Lilley, 2013), when contained within the ribosome, the SAM-I riboswitch or bound by L7Ae (Huang *et al.* 2016). Thus k-turn conformation is evidently determined by the nucleotide sequence. Given the strong conservation of most nucleotides of the standard k-turn (the tandem G:A, A:G pairs and very frequently  $-1b:-1n = C:G$ ) the possible location of the determinant of conformation is very limited. We found that the key determinant was the 3b:3n position. The known k-turn structures can be plotted in a  $4 \times 4$  array of the 16 possible 3b:3n sequences (Fig. 4) (Huang *et al.* 2016), whereupon it is apparent that a given 3b:3n sequence is associated with either N3 or N1 conformation, except for *HmKt-7* as discussed above. Moreover, this holds true irrespective of the sequence context of the k-turn. For example, three different k-turns are known with 3b:3n = U:G, one contained within the cobalamine riboswitch (Peselis & Serganov, 2012), one within *Thermus thermophilus* ribosome and one bound by L7Ae protein, and all three adopt the N1 conformation. Three k-turns have 3b:3n = G:C, and all three adopt the N3 conformation. Similarly there are three k-turns with 3b:3n = U:U, all of which are in the N3 conformation. All the natural k-turns occupy eight cells of the  $4 \times 4$  array, so a systematic synthetic approach was adopted to determine the preferred conformations of the remaining 3b:3n sequences where possible. *HmKt-7* inserted in place of the natural SAM-I riboswitch k-turn was systematically modified at the 3b:3n position. Each was placed in

**Table 1.** Summary of the sequence and conformation of simple, standard k-turns

Name	PDB ID	Resolution	Space group	−1b:−1n	Loop	3b:3n	3b:3n conf	Conformation
SAMI rsw <i>T. tengongensis</i>	3GX5	2.4	P4 <sub>3</sub> 2 <sub>1</sub> 2	C:G	GAC	A:G	tSH	N3
SAMI rsw + YbxF <i>B. subtilis</i>	3V7E	2.8	C2	C:G	GAU	A:G	tSH	N3
SAMI rsw <i>B. subtilis</i>	4KQY	3.02	P3 <sub>1</sub> 2 <sub>1</sub>	C:G	GAC	A:A	tSH	N3
c-di-GMP-II rsw <i>C. acetobutylicum</i>	3Q3Z	2.51	P2 <sub>1</sub>	C:G	AAU	U:U	cWC	N3
cobalamine rsw <i>S. thermophilum</i>	4GXY	3.05	P3 <sub>1</sub> 2 <sub>1</sub>	C:G	GAG	U:G	tSH	N1
T-box rsw + YbxF <i>O. iheyensis</i>	4LCK	3.2	C222 <sub>1</sub>	C:G	GAU	G:C	cWC	N3
Kt-7 <i>H. marismortui</i>	3CC2	2.4	C222 <sub>1</sub>	C:G	GAA	A:G	tSH	N1
Kt-7 <i>T. thermophilus</i>	4Y4O	2.3	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	C:G	GUG	U:G	tSH	N1
Kt-7 <i>D. radiodurans</i>	2ZJR	2.91	I222	C:G	UUU	C:C	cWC	N1
Kt-46 <i>H. marismortui</i>	3CC2	2.4	C222 <sub>1</sub>	C:G	GAA	A:G	tSH	N3
ScKt-46 <i>S. cerevisiae</i>	4U4R	2.8	P2 <sub>1</sub>	U:A	GAC	A:C	tSH	N3
TtKt-78 <i>T. thermophilus</i>	3UMY	1.9	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	C:G	UGU	A:G	tSH	N3
box C/D + L7Ae <i>A. fulgidus</i>	1RLG	2.7	P23	G:C	CGU	U:U	cWC	N3
box C/D + L7Ae <i>S. solfataricus</i>	3PLA	3.15	P4 <sub>1</sub> 2 <sub>1</sub> 2	U:A	UGU	U:U	cWC	N3
pre-mRNA + L30e <i>S. cerevisiae</i>	1T0 K	3.24	P4 <sub>1</sub> 2 <sub>1</sub> 2	C:G	AGA	U:G	tSH	N1
U4 + 15.5k <i>H. sapiens</i>	1E7 K	2.9	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	C:G	AAU	G:C	cWC	N3
U4 + Prp31 + 15.5k <i>H. sapiens</i>	2OZB	2.6	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	C:G	AAU	G:C	cWC	N3
U4atac + Prp31 + 15.5k <i>H. sapiens</i>	3SIU	2.63	P2 <sub>1</sub>	C:G	AAU	G:U	cWC	N3

The name of each k-turn (and binding protein where relevant) are shown, above the organism of origin.

The PDB ID and resolution for each structure are shown, and the space group in which the structure was solved. The −1b:−1n, loop and 3b:3n sequences are tabulated, together with the conformation of the 3b:3n as either *cis*-Watson-Crick (cWC) or *trans*-sugar-Hoogsteen (tSH). Lastly, the conformation as N3 or N1 type is shown for each k-turn. rsw, riboswitch.

crystallization trials resulting in structures determined at 1.71–3.31 Å resolution for 12 of the possible 16 sequences (Huang *et al.* 2016). These additional structures are included (red) in Fig. 4. Of these six fill previously empty cells of the array, leaving only two unoccupied cells. The six remaining structures fall into cells with pre-existing data from the natural k-turns; in every case the newly-determined structures adopt the same conformation as the other members of the same cell. This study also demonstrates that the SAM-I riboswitch is good vehicle for studying k-turn structure, being flexible enough not to perturb the natural conformation of a given k-turn.

In the *HmKt-7* context five of the 3b:3n variants retain the N3 conformation, while seven switch to the N1 conformation. We conclude that the 3b:3n position is the major sequence determinant of k-turn conformation. The full array (Fig. 4) can be used to predict the conformation of new k-turns. However, there is no clear pattern that reveals itself, nor a set of rules relating 3b:3n sequence to conformation. The 3b:3n nucleotides adopt two kinds of basepair, either *cis*-Watson-Crick or *trans*-Hoogsteen-sugar edge pairs. Yet there is no apparent correlation of the 3b:3n basepair type with k-turn conformation. However, there is a structural correlation based on base pair width. The rotation of the A2b nucleobase required to accept a donated proton at N1 changes the width of the base pair. Taking the ensemble of k-turn structures available, the mean C1'-C1' distance for the 2b:2n base pair for the N3 conformation structures is 8.93 ± 0.22 Å. However, that for the N1 conformation is

		3n				
		A	C	G	U	
3b	A	3 3	3	33 331 3		
	C	1	1	3	1	riboswitches
	G		3 33	1	3	ribosome
	U	1	1	1 1 1 1	3 33 3	bound to L7Ae Kt7 variants

**Fig. 4.** The conformation of k-turns as a function of the 3b:3n sequence, shown as a  $4 \times 4$  array where the rows show the 3b and the columns the 3n nucleotides (Huang *et al.* 2016). The Watson-Crick base pairs are found on the ascending diagonal. The N3 (3) or N1 (1) conformation is shown for each k-turn. These structures have been determined for riboswitches (black), within the ribosome (cyan), as complexes with L7Ae protein (magenta) and as 3b:3n sequence variants of *HmKt-7* (red). Note that where multiple k-turns have the same 3b:3n sequence these, in general, adopt the same conformation. The one exception to this is found in the 3b:3n = A:G cell, where a single case has adopted the N1 conformation. This is *HmKt-7* in the context of the ribosome where the environment evidently forces the k-turn to adopt a less favorable conformation. In all other circumstances, *HmKt-7* adopts the N3 conformation.

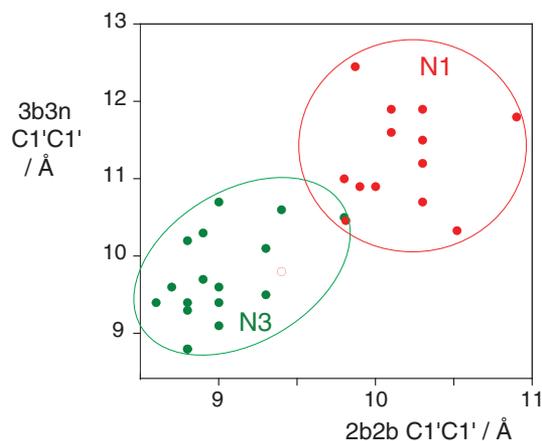
$10.17 \pm 0.40$  Å. Thus the local helical width at the 2b:2n base pair is significantly greater for the N1 conformation k-turns. If we plot the C1'-C1' distances for the 2b:2n base pairs against those for the 3b:3n base pairs for the entire set of k-turn structures we find that they cluster into two distinct groups (Fig. 5); the sets of N3 and N1 structures are very nearly disjoint,  $N3 \cap N1$  is almost empty. The N3 structures have shorter values of both C1'-C1' distances, while the N1 structures have longer values for both. Interestingly, there is one N1 structure that lies with the N3 structures on this plot, that of *HmKt-7* in the ribosome. This reveals its true character because it is intrinsically an N3 k-turn that has been forced to adopt the N1 structure in the ribosome (see section 4). Ribosomal *HmKt-7* is very much the exception that proves the rule.

## 6. Variations on the structural theme of the standard k-turn

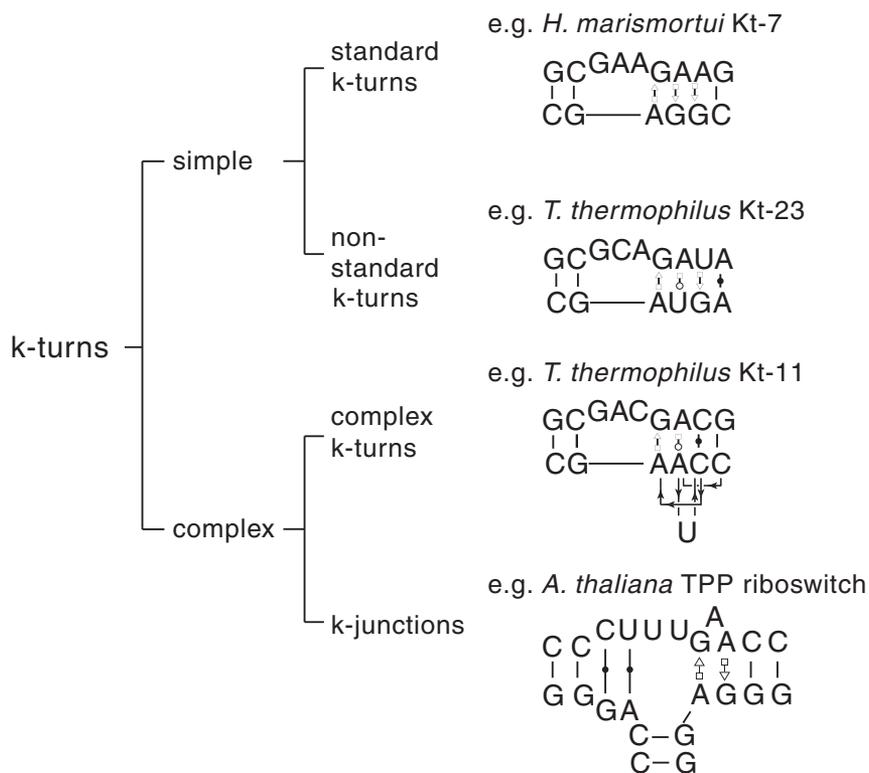
The most common form of the k-turn motif has a three-nucleotide bulge followed by the tandem G:A, A:G pairs, but there is some variation found in natural sequences. We have already noted that in the ribosome some k-turns depart from the standard k-turn motif, with loops that differ from the usual 3 nt, and some have a nucleotide on the lower strand opposing the bulge. But even the G:A, A:G pairs are not completely invariant. Our classification scheme is shown in Fig. 6. Simple k-turns retain the basic framework of the k-turn, whereas in the complex k-turns the nucleotides of the G:A, A:G pairs do not map linearly onto the RNA sequence. Within these groups, non-standard k-turns have a nucleotide substitution within the G:A, A:G pairs. In k-junctions, the k-turn has been elaborated into a helical junction by the presence of an additional helix. While these variations from the standard k-turn reflect significant changes in sequence, we find that in general all the standard structural features of the k-turn are retained.

### 6.1 Non-standard, simple k-turns

In non-standard k-turns, there is a substitution of one of the nucleotides in the otherwise-highly conserved tandem G:A, A:G pairs at the 1b:1n or 2b:2n positions. This occurs most frequently at the 2b:2n position. Kt-23 of the 16S rRNA in the small ribosomal subunit provides a good example. In this k-turn the 2n nucleotide is rather infrequently found to be G, with a frequency of occurrence  $U > C > G > A$ . While a G2nU substitution in *HmKt-7* prevented ion-induced folding *in vitro*, Kt-23 from *Thermus thermophilus* (where 2n = U) folded normally into the kinked conformation on the addition of  $Mg^{2+}$  ions. Moreover, in the *T. thermophilus* ribosome Kt-23 forms a completely normal k-turn structure (Fig. 7), that superimposes with *HmKt-7* with a root mean square deviation (RMSD) = 0.60 Å. The A:G pair at the 2b:2n position is replaced by a *trans* Hoogsteen-Watson Crick A•U pair, that places the adenine nucleobase in essentially the location that it would be found in a standard k-turn. The A1n is also in its normal position, and both adenine nucleobases accept the standard cross-strand H-bonds. The k-turn adopts an N1 conformation structure, with an H-bond from G-1n O2' to A2b N1 plus G-1n N2 to A2b N3, and G1 O2' to A1n N1. The 3b:3n base pair is a *trans* U:G basepair with a C1'-C1' distance of 11.0 Å; the local widening of the helix may be important, see section 5.

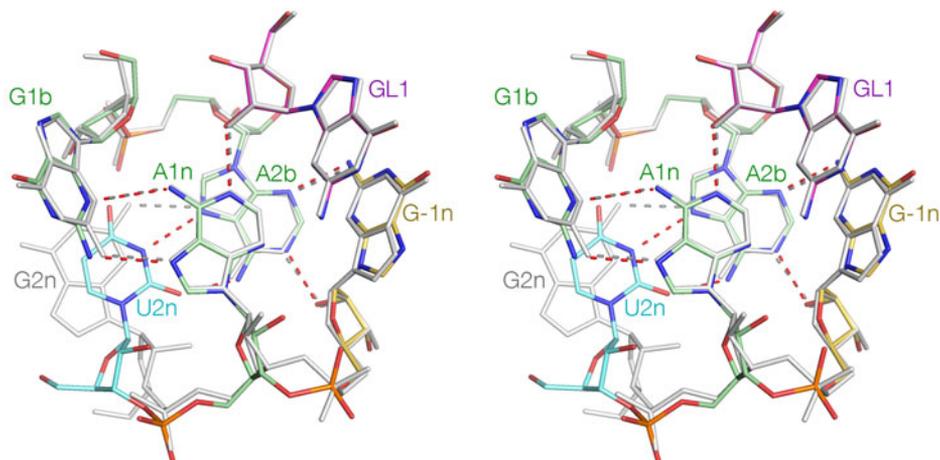


**Fig. 5.** A plot of C1'-C1' distances for the 2b:2n and 3b:3n basepairs (Huang *et al.* 2016). These are colored red for N1 structures and green for N3 structures. Note that these form two distinct clusters with longer distances for both basepairs for the N1 structures compared with the N3 structures. Note however that the ribosomal *HmKt-7* k-turn (shown as a red open circle) falls within the N3 cluster.

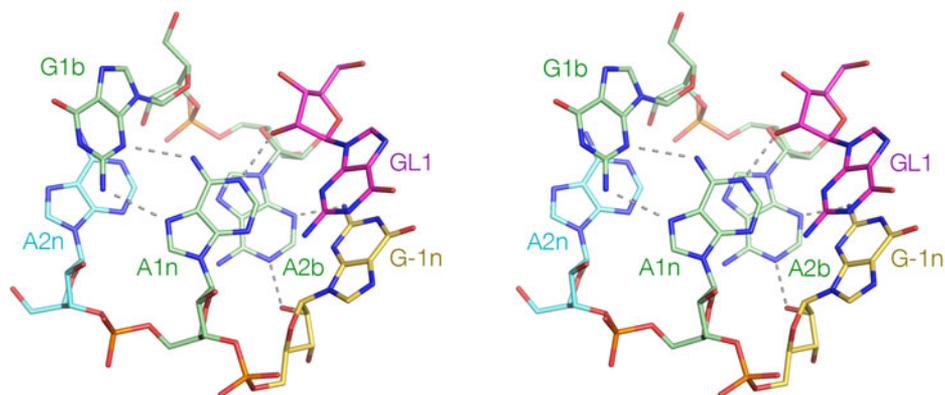


**Fig. 6.** A scheme showing the classification of k-turns, with examples. In the simple, non-standard k-turns one or more nucleotides of the G•A basepairs are substituted by a different nucleotide. In the complex k-turns the nucleotides of the tandem G•A basepairs do not map linearly onto the sequence. k-junctions contain an additional helix on the non-loop-containing strand.

The least-common 2b:2n sequence found in Kt-23 is A:A at <1%. Bioinformatic searching revealed an example that occurs in *Telohania solenopsae* (Schroeder *et al.* 2012), a parasite of the fire ant, *Solenopsis invicta*. It is a simple k-turn that only differs from a standard k-turn in the 2n position. It exhibits weak folding in Mg<sup>2+</sup> ions but folds normally on binding L7Ae protein. It was functional on substitution into the SAM-I riboswitch in place of the normal k-turn, i.e. the modified riboswitch bound SAM normally. The structure of *T. solenopsae* k-turn was solved in the context of the riboswitch by crystallography (Schroeder *et al.* 2012). It adopted the standard k-turn conformation, with all the normal H-bonds (Fig. 8). It could be superimposed with *HmKt-7* with an RMSD of 1.6 Å and an almost perfect superposition of the A1n and A2b nucleobases. These accept the

*T. thermophilus* Kt-23

**Fig. 7.** The structure of *T. thermophilus* Kt-23, a simple, non-standard k-turn with U at the 2n position in place of the normal G. The sequence is shown (upper) with the non-standard U2n nucleotide highlighted in cyan. A superposition of *Tt*Kt-23 (colored; PDB ID 2WH1) with *Hm*Kt-7 (grey) is shown as a parallel-eye stereoscopic image (lower). This shows the close superposition of the key A1n, A2b and G-1n nucleotides between the two structures despite the G2nU substitution, making all the standard cross-strand H-bonds.

*T. solenopsae* Kt-23

**Fig. 8.** The structure of *T. solenopsae* Kt-23, a simple, non-standard k-turn with A at the 2n position in place of the normal G (Schroeder *et al.* 2012). The sequence is shown (upper) with the non-standard A2n nucleotide highlighted in cyan. The core of the *Ts*Kt-23 is shown as a parallel-eye stereoscopic image (lower; PDB ID 4AOB), showing that all the standard cross-strand H-bonds are present.

normal cross-strand H-bonds, forming an N1 structure. The A2b:A2n pairing is disposed like a *trans*-Hoogsteen-sugar edge basepair, although the nucleobases are spaced apart (e.g. A2b N6 - A2n N3 = 5.0 Å) and unconnected by H-bonds.

### 6.2 Complex k-turns

In complex k-turns the nucleotides of the tandem G•A basepairs do not map onto the sequence in a linear way – they appear out of place in the normal secondary structure depiction. However, on folding into the k-turn conformation they nevertheless



find their conventional position within the structure and make the same interactions. We, therefore, use our standard nomenclature to label these nucleotides according to their position in the structure, rather than the sequence.

Kt-11 in the small ribosomal subunit of *T. thermophilus* provides a good example (Fig. 9). In the NC helix, the non-bulged strand has the linear order 5'-C4n-A2n-U-C3n-A1n-3'. The backbone forms an S-shaped turn that places C3n under A2n despite being to its 3' side in the sequence, and the 1n and 2n nucleotides are separated by two nucleotides in the primary sequence. Despite the scrambled connectivity of this sequence, the RNA folds into a normal k-turn structure in the ribosome with all the nucleotides in their correct places. Once again, the A1n and A2b are placed in the normal k-turn positions in the structure, with the near-universal cross-strand H-bonding. A2b N1 accepts a proton from G-1nO2', so the structure is in the N1 conformation. The A2b:A2n basepair is *trans*-Watson-Crick-Hoogsteen, connected by two H-bonds, with a significant rotation of the A2n nucleobase relative to the normal G2n in *HmKt-7* for example. Interestingly, the structure of Kt-11 in the *Escherichia coli* ribosome is virtually identical to that of *T. thermophilus*. The non-bulged strand of the NC helix makes a pronounced S-turn with the linear order 5'-G4n-A2n-U-U3n-A1n-3', and the A2b:A2n basepair is *trans*-Hoogsteen-Watson-Crick. Nevertheless, the RNA again folds into the k-turn geometry, placing A1n and A2b into their standard positions where they make the usual cross-strand H-bonds in the N1 conformation. The core can be superimposed with that of *HmKt-7* (Fig. 9c) with an RMSD = 0.58 Å.

We noticed that a topologically-equivalent k-turn occurs in a structure determined for the Constitutive Transport Element RNA bound to the export factor TAP (Teplova *et al.* 2011). The non-bulged strand has the linear order 5'-U4n-A2n-A-G-A-C3n-A1n-3', and the A2b:A2n basepair is again *trans*-Hoogsteen-Watson-Crick. The S-turn trajectory of the non-bulged strand is even more pronounced than *TtKt-11* as it now has three consecutive nucleotides in the strand section that passes backwards, yet once again A1n and A2b are located in the correct position to make the normal cross-strand H-bonding interactions with GL1 and G-1n.

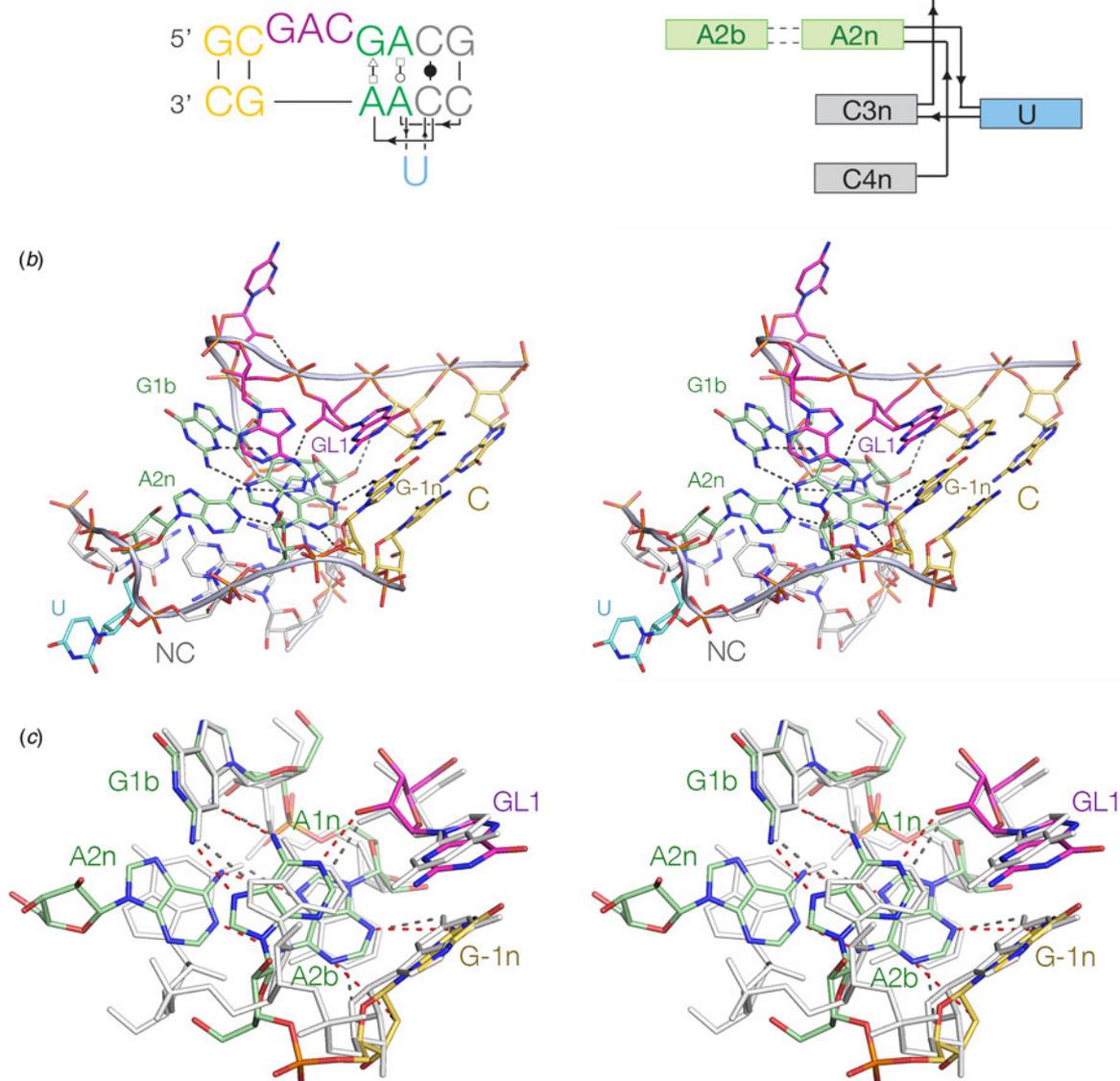
### 6.3 k-junctions

k-junctions represent the most extreme form of the k-turn. These went largely unnoticed for several years but were identified when a computer program was written to search for helices with the relative disposition of those in k-turns within RNA structures downloaded from the PDB (Wang *et al.* 2014). While this exercise identified a number of known k-turns, it surprisingly flagged the thiamine pyrophosphate (TPP) riboswitches which on closer inspection were found to contain a new class of k-turn (Wang *et al.* 2014). These riboswitches contain a three-way junction from which two helices emanate that interact to create the ligand binding site. But it emerged that the junction was, in fact, a k-turn, where the third helix (we call the T helix) is located on the non-bulged strand opposite the loop of the k-turn. This had also been noted in a footnote in an earlier study of A-minor interactions by Jaeger and colleagues (Geary *et al.* 2011).

The *Arabidopsis thaliana* TPP riboswitch (Thore *et al.* 2006) contains a good example of a k-junction (Fig. 10). As ever, the core of the k-junction has tandem G:A, A:G pairs, labeled according to our nomenclature. Both are *trans* sugar(G):Hoogsteen (A) basepairs. G1b and A2b are separated by an additional A that is directed down into the major groove of the NC helix to make a triple base interaction with the U5b:A5n basepair. The C helix terminates at the junction with conventional Watson Crick C:G and U:A basepairs, but for reasons that will shortly become apparent we designate the C and U as formally part of the loop, termed L1 and L1.1. These two basepairs lie between and coaxial with the C and T helices in the folded structure. L1 O2' donates an H-bond to A1n N1 and G-1n donates H-bonds from O2' and N2 to A2b N1 and N3 respectively. Thus the standard cross-strand H-bonds are present, and the k-junction adopts an N1 conformation. The structure can be well superimposed with a standard N1 k-turn such as ribosomal *HmKt-7*. Superposition using the backbone atoms for -1n, L1, 1n, 2b and 2n gives an RMSD of 1.58 Å.

The *E. coli* TPP riboswitch (Serganov *et al.* 2006) also contains a k-junction in an equivalent position. The global shape is closely similar to that of the *A. thaliana* TPP riboswitch, with the coaxial stacking of the C and T helices. The main difference is that the 1b nucleotide is an adenine so that the 1b:1n basepair is a *trans*-Hoogsteen-Hoogsteen A:A basepair. What takes the role of the loop is a CGU sequence, where the CL1 and GL2 are basepaired with their complements on the opposing strand. The A1n and A2b play their usual role, accepting H-bonds L1 O2' to A1n N1, G-1n O2' and N2 to A2b N1 and N3, respectively. Thus this k-junction is also an N1 structure. An A1nC mutant of the riboswitch exhibited impaired folding and no longer bound TPP in calorimetric experiments (Wang *et al.* 2014), demonstrating the importance of the k-junction in the function of the riboswitch.

Analysis of over 11 000 TPP riboswitch sequences indicates that the k-junction is common to all and that the critical adenines at the 1n and 2b positions are conserved in >99.5% of the sequences. In a wider context, the k-junction occurs in the

*T. thermophilus* Kt-11

**Fig. 9.** The structure of the complex k-turn *T. thermophilus* Kt-11. The sequence of the k-turn is shown top left. (a) A cartoon of the structure of *TtKt-11*, showing the connectivity of the nucleotides, that are labeled according to their position in the k-turn structure. (b) The overall structure of *TtKt-11* (PDB 2WH1) viewed from the side of the non-loop strand. (c) The core of *TtKt-11* (colored) superimposed with that of *HmKt-7* (grey). There is good superposition of the key A1n, A2b and G-1n nucleotides, with an RMSD = 0.579 Å. (b) and (c) are shown as parallel-eye stereoscopic images.

ribosome, with two examples in the *H. marismortui* 50S ribosomal subunit. Given that k-junctions are not easy to identify without detailed analysis, it is likely that further examples are waiting to be found.

We note that all the known complex k-turns and k-junctions adopt the N1 conformation. While this is more stable than the N3 conformation in about half the standard k-turns, it clearly confers an advantage in the more diverged structures. It is likely that the N1 conformation provides some structural flexibility that allows the k-turn geometry to form in the more complex sequences.

#### 6.4 To be or not to be a k-turn

The potential of the adenine nucleobases in tandem G•A pairs for participating in A-minor interactions can be exploited more widely. In the human signal recognition particle RNA, tandem *trans* sugar-Hoogsteen G:A basepairs mediate a long-





Conversely, the role of k-turns within a given species may be taken up by a different kind of kink in the helix, acting in a functionally-equivalent manner, but lacking all the characteristic structural features of the k-turn. In the ribonuclease P ribozyme, a tightly-kinked RNA helix (Meyer *et al.* 2012b) facilitates the formation of a loop–receptor interaction that creates the binding site for the substrate. In some species, this is achieved using a k-turn but in others, a different kinked structure termed the pk-turn is used. Globally the pk-turn of the *Thermotoga maritima* ribozyme is very similar to the geometry of a k-turn, but it lacks all the standard features of the k-turn including the G:A pairs. Yet the pk turn and the k-turn can be exchanged with preservation of function. Mutation of *T. maritima* RNase P to replace the pk turn by a k-turn resulted in retention of ribozyme activity, while a SAM-I riboswitch in which the k-turn had been replaced by the pk turn retained the ability to bind SAM ligand (Daldrop *et al.* 2013).

Another example of structural and functional redundancy is found in the lysine riboswitch. A putative k-turn was identified in the *Bacillus subtilis* lysine riboswitch and shown to bind L7Ae protein (Blouin & Lafontaine, 2007). However, a k-turn was not found in the structure of the lysine riboswitch of *T. maritima* (Garst *et al.* 2008; Serganov *et al.* 2008). Rather its function was taken by a different kinked structure lacking G:A pairs, and not stabilized by A-minor H-bonds.

### 6.5 Identification of k-turns in RNA sequence and structure

In the light of experience studying k-turns over the last decade, how easy is it to recognize k-turn motifs within RNA structures, or indeed potential k-turns given RNA sequence information?

Identifying potential k-turns in RNA sequence reliably is not easy, and in principle, the two strands could be widely separated within the overall RNA structure. It is relatively straightforward to predict basepairing and secondary structure by analysis of nucleotide covariation, yet the k-turns can still be difficult to find. In general, we use three stages to do this, performed manually:

1. Focus on junctional sections and try to identify tandem G:A, A:G pairs, since these are the most well-conserved features. Where available, phylogenetic data should indicate strong conservation of the two adenine nucleotides.
2. Seek unpaired nucleotides (normally three) preceding the G:A 5' to the G.
3. Check the –1b:–1n position. In the majority of k-turns, this is C:G. This is not required, but if a given sequence conforms then it adds confidence that it really is a k-turn.

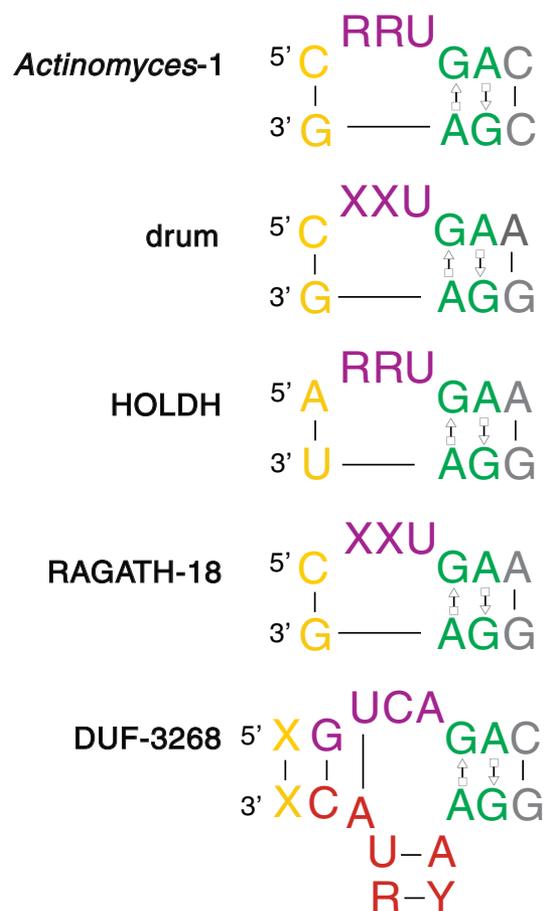
While this works well for standard k-turns, it gets progressively more difficult for non-standard and complex k-turns or k-junctions. There are also procedures based on statistical methods that have been proposed for identifying k-turn motifs (Bayrak *et al.* 2017).

We have applied this approach to the collection of 224 candidate structured RNA sequences recently identified by Breaker and colleagues (Weinberg *et al.* 2017) by comparative analysis of specific subsets of intergenic regions. Rigorous application of the method identified four putative standard k-turns, and one k-junction. These are shown in Fig. 11 and listed in Table 2.

Identifying k-turns within three-dimensional RNA structure is a rather different problem, but once again the task becomes progressively more difficult moving from standard k-turns to the more complex varieties. Two computational tools designed to analyze RNA structure include can search for k-turns. These are the RNA 3D Motif Atlas (Parlea *et al.* 2016) and RMDetect (Cruz & Westhof, 2011). Simple k-turns have generally been identified by authors publishing new structures, but more complex ones have often been missed, especially the k-junctions. We found the k-junctions by focusing on the disposition of the C and NC helices (Wang *et al.* 2014) (see section 6.3). A program written in Python 2.7 downloaded RNA structures from the PDB and then searched for helices with the required relative inclination similar to that of known k-turn structures. Standard Watson-Crick (plus G:U) basepairs were identified, from which four vectors were defined; one for position and three for orientation. This was calibrated against known k-turn structures and then used to search for new structures. A full description of the program was given in the Supplementary Information of Wang *et al.* (2014). Application of the program identified the k-junctions of the TPP riboswitches and the Constitutive Transport Element RNA.

## 7. The folding of k-turns

A k-turn in free solution in a buffered solution lacking added metal ions or proteins is predominantly unfolded (Goody *et al.* 2004), adopting the structure generally formed by an RNA duplex containing a three-nucleotide bulge (Lilley, 1995). The



**Fig. 11.** Putative k-turns and one k-junction identified in a collection of 224 candidate structured RNA sequences from the Breaker laboratory (Weinberg *et al.* 2017). The most common sequence amongst the variations found in nature are shown for each. These k-turns and k-junction are tabulated in Table 2, with their predicted properties.

k-turn conformation must be stabilized by something specific, either the presence of metal ions, protein binding or the formation of tertiary interactions.

Experimentally folding has generally been studied mainly by two methods. One is polyacrylamide gel electrophoresis which is sensitive to the angle included by the two helical arms that have normally been extended to increase sensitivity, usually using

**Table 2.** Four putative k-turns and one k-junction within a series of 224 structured RNA sequences that have recently been presented by Breaker and colleagues (Weinberg *et al.* 2017)

RNA source	Location in RNA	3b:3n	Predicted properties	
			Conformation	Folding in ions
Actinomyces-1	P2	C:C > C:U	N1	Fold
Drum	P1	A:G > >A:A > U:U	N3	Fold
HOLDH	P1	A:G > C:A > G:A	N3/N1	Fold
RAGATH-18	P1	A:G	N3	Fold
DUF3268	J2,3,4	C:G		

The k-turn structures were identified applying the rules written in the text. Where available, the relative frequency of the most common 3b:3n sequences are given, and the most frequent are shown in Fig. 11. The properties of the k-turns are predicted by application of the data summarized in Fig. 17. The conformations of Actinomyces-1, drum and RAGATH-18 are predicted unambiguously, but the distribution of sequences for HOLDH suggest this is more variable. We do not know how reliably the rules can be applied to a k-junction, although we note that all other known k-junctions adopt an N1 conformation.



DNA for synthetic convenience (Lilley, 2008). The other main approach is to use fluorescence resonance energy transfer (FRET) of RNA with terminally-attached donor and acceptor fluorophores. The efficiency of energy transfer ( $E_{\text{FRET}}$ ) depends on the inverse sixth power of the distance between the fluorophores (Förster, 1948), and this distance becomes shorter when the RNA becomes more kinked. Small-angle X-ray scattering interferometry with RNA containing terminally-attached gold nanoparticles has also been used (Shi *et al.* 2016), where Fourier transformation of the processed scattering curves yields the distance between the nanoparticles. EPR has also been used to study the structure of k-turns (Esquiaqui *et al.* 2016).

Fluorescence lifetime distributions of donor-acceptor-labeled k-turn-containing RNA indicated that in the absence of added metal ions and proteins there is a conformational equilibrium between folded (i.e. kinked) and unfolded (extended) RNA that is strongly biased towards the unfolded state (Goody *et al.* 2004). We have found three main mechanisms that affects the position of this equilibrium, driving the population into the kinked conformation.

### 7.1 Folding of k-turns on the addition of metal ions

Some k-turns (although not all, as we discuss below) will adopt the folded conformation on the addition of metal ions (Goody *et al.* 2004; Liu & Lilley, 2007). This is revealed as a retardation of electrophoretic mobility, or an increase in FRET efficiency  $E_{\text{FRET}}$ . Titration curves of  $E_{\text{FRET}}$  as a function of ionic concentration are well fitted by a two-state folding process (Fig. 12a). Folding occurs in either divalent or monovalent ions, with half concentrations of  $[\text{Mg}^{2+}]_{1/2} \sim 100 \mu\text{M}$  and  $[\text{Na}^+]_{1/2} \sim 30 \text{mM}$ . The intracellular concentration of unbound  $\text{Mg}^{2+}$  ions has been measured using fluorescent indicator probes at close to 1 mM (Csernoch *et al.* 1998; Li-Smerin *et al.* 2001), so this would be sufficient to achieve folding.

### 7.2 Folding of k-turns on the binding of proteins

Many k-turns in nature are bound by specific proteins. The archetypal k-turn binding proteins are the L7Ae family (discussed below). The folding of *HmKt-7* induced by the binding of *Archeoglobus fulgidus* L7Ae (*Afl7Ae*) is shown using FRET in Fig. 12b. The affinity of L7Ae for k-turn RNA is extremely high, and the folding observed as the increase in  $E_{\text{FRET}}$  reflects stoichiometric binding as discussed in section 9. Protein-induced folding is not restricted to proteins of the L7Ae family. The majority of k-turns in the ribosome are bound by proteins of greatly differing structure and recognized in very different ways. This can be seen by the examples of *E. coli* L24 and *T. thermophilus* S17 (Fig. 13a, c). FRET analysis shows that Kt-7 adopts the kinked conformation on the binding of either protein (Fig. 13b, d). Thus the folding of k-turn RNA on binding proteins in general, is not a property of a single class of proteins.

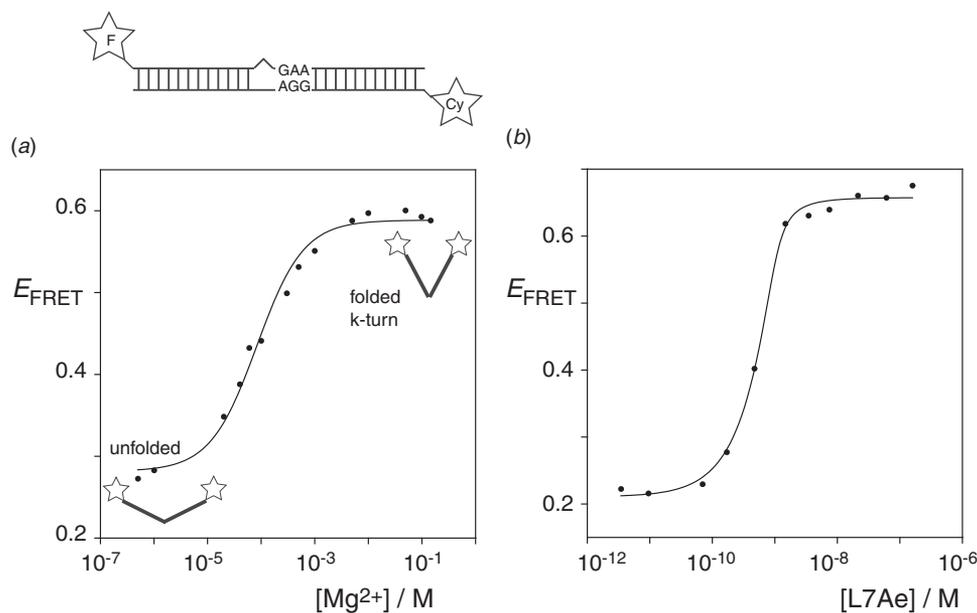
### 7.3 Folding of k-turns induced by tertiary interactions

Tertiary interactions can also stabilize the folded conformation of the k-turn. This was first shown in a G2nA mutant of the SAM-I riboswitch k-turn (Schroeder *et al.* 2011). Although this k-turn was unable to fold in isolation on the addition of metal ions, it did fold in the context of the riboswitch. A crystallographic structure of the RNA showed all the normal interactions of the k-turn despite the presence of the mutant A2b:A2n basepair. In the riboswitch the C-helix mediates a tertiary contact between its terminal loop and a receptor, creating the SAM ligand binding site (demonstrated by calorimetry), and evidently, these tertiary contacts lead to the induced folding of the k-turn.

## 8. Recognition of k-turns by proteins of the L7Ae family

The L7Ae family proteins are found bound to a variety of k-turns in different cellular structures (Koonin *et al.* 1994; Watkins *et al.* 2000). The family includes the archaeal and eukaryotic proteins L7Ae, L30e and S12e, the human 15-5k protein (Nottrott *et al.* 1999) and the yeast Nhp2 and Snu13p proteins. Functional substitutions between some of these proteins have been found to be tolerated (Rozhdestvensky *et al.* 2003). YbxF and YlxQ are bacterial homologs (Baird *et al.* 2012). L7Ae-family proteins are bound to k-turns in the ribosome (Ban *et al.* 2000), the U4-U6.U5 spliceosomal complex (Nottrott *et al.* 1999; Vidovic *et al.* 2000), box C/D (Kuhn *et al.* 2002; Szewczak *et al.* 2002; Watkins *et al.* 2002) and H/ACA snoRNPs (Hamma & Ferré-D'Amaré, 2004; Li & Ye, 2006; Rozhdestvensky *et al.* 2003), U3 snoRNP (Marmier-Gourrier *et al.* 2003), telomerase (Pogacic *et al.* 2000) and archaeal RNaseP (Cho *et al.* 2010). L7Ae-family proteins are thus very much the archetypal k-turn binding proteins, essential in many cellular processes including ribosome structure, spliceosome assembly and guided site-specific modification of RNA. RIP-seq experiments have indicated that translation of L7Ae is autoregulated in archaea (Daume *et al.* 2017).

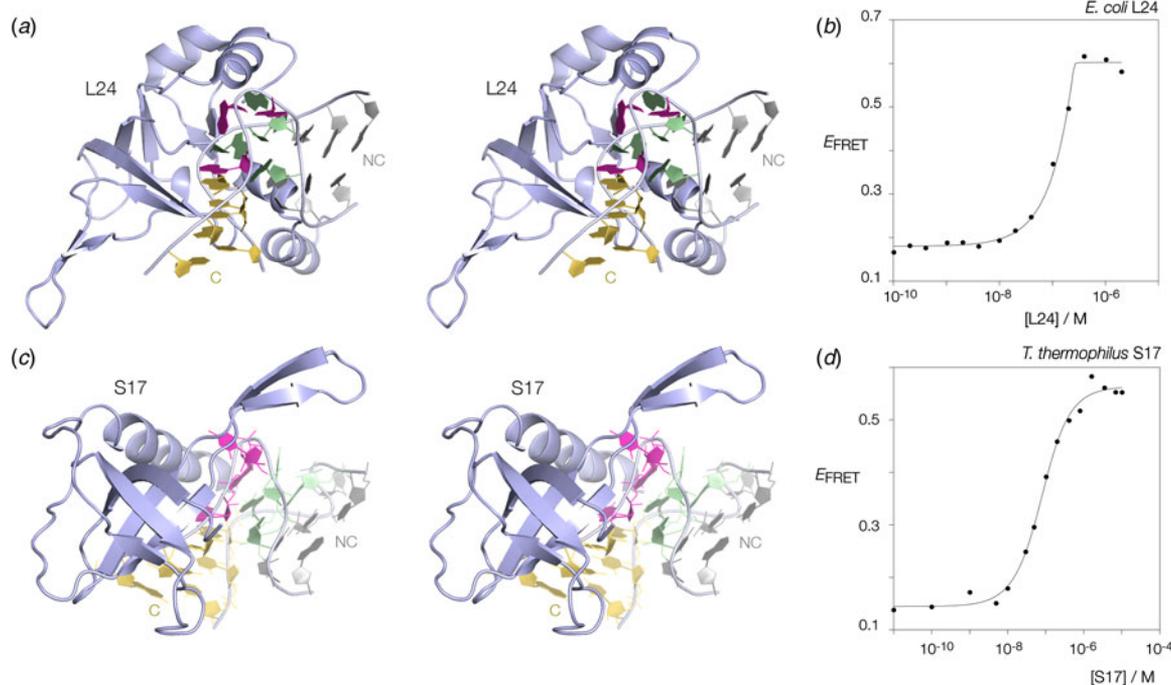
Crystal structures have been determined for a number of L7Ae-family protein-RNA complexes. These include 15-5k bound to the human U4 snRNA k-turn (Vidovic *et al.* 2000), archaeal L7Ae bound to a box C/D k-turn (Moore *et al.* 2004; Suryadi



**Fig. 12.** Folding of *HmKt-7* in response to the addition of metal ions and the binding of protein. In these experiments, the folding of a 27 bp RNA duplex with a centrally-located *HmKt-7* is studied by fluorescence spectroscopy (top). The RNA is 5'-terminally labeled with fluorescein and Cy-3, and the efficiency of energy transfer between the fluorophores ( $E_{\text{FRET}}$ ) measured as a function of the ion or protein concentration. In the absence of these, the structure is relatively extended with a corresponding low value of  $E_{\text{FRET}}$ . On folding into the kinked conformation the distance between the fluorophores shortens, resulting in an increase in  $E_{\text{FRET}}$  (these species are shown schematically in part (a)). As the population of the kinked form increases the mean value of  $E_{\text{FRET}}$  rises, and the data are well fitted by a two-state model, shown by the lines in the plots. (a) Titration with  $\text{Mg}^{2+}$  ions. The fit corresponds to two-state folding induced by the non-cooperative binding of  $\text{Mg}^{2+}$  ions with a  $[\text{Mg}^{2+}]_{1/2} \sim 100 \mu\text{M}$ . (b) Titration with *AfL7Ae* protein. Addition of the protein leads to a folding of the k-turn, with a typically slightly higher end point compared with the ion titration. The affinity of *L7Ae* for k-turn RNA is extremely high (Turner & Lilley, 2008), so that binding is essentially stoichiometric and cannot be used to estimate an apparent  $K_d$ .

*et al.* 2005) and to a box H/ACA k-loop (Hamma & Ferré-D'Amaré, 2004). The structure of *L7Ae* bound to *Kt-15* has also been determined in the context of the complete large ribosomal subunit (Ban *et al.* 2000); however, *Kt-15* is a complex k-turn, so less directly comparable with the simple k-turn complexes. We determined the structure of *A. fulgidus* *L7Ae* protein (*AfL7Ae*) bound to the archetypal k-turn *H. marismortui* *Kt-7*, probably the best-characterized k-turn (Fig. 14). This was solved at a near-atomic resolution of 2.3 Å, giving an electron density map of excellent quality where each side chain was unambiguous. Taking all these structures together we can describe the manner of recognition of and interaction with standard k-turns in general terms:

1. *L7Ae* proteins bind in the major groove that is splayed around the outside of the k-turn (Fig. 14a). While DNA binding proteins frequently place a recognition helix in the major groove to make sequence-specific interactions, this is generally not possible in RNA due to the narrow and deep geometry. However, in the k-turn the major groove is opened by its location on the outside face. Thus the *L7Ae*-family complexes resemble helix-turn-helix and zinc finger proteins in this respect.
2. Two elements of the protein comprise the RNA binding interface. The first is a very basic  $\beta$ -strand : turn :  $\alpha$ -helix element. The second is a short loop of hydrophobic residues that make a 180° turn.
3. The  $\alpha$ -helix is located in the major groove of the NC helix, directing its N-terminal end towards the major groove edges of the guanine nucleobases of the G:A pairs (Fig. 14b).
4. The C-terminal end of the  $\alpha$ -helix lies close to the non-bulge-containing strand of the NC arm of the k-turn, from which basic side chains make non-specific interactions with the RNA backbone. The arginine in the third turn (R41 in *AfL7Ae*) makes a universal H-bond with the 3n/4n phosphate. This interaction is observed in all the complexes studied. The lysine in the second turn (K37 in *AfL7Ae*) is bonded to the 4n/5n phosphate in the *Kt-7* complex but is more variable over the ensemble of structures. These interactions will locate the position of the NC helix, and therefore interrogate the angle of the helices included in the k-turn. At the N-terminal end of the  $\beta$ -sheet two additional basic side chains contribute to the non-specific electrostatic stabilization of the complex.
5. Side chains located at the N-terminal end of the  $\alpha$ -helix make sequence-specific interactions with the guanine nucleobases of the tandem G:A, A:G pairs in the core of the k-turn. A glutamate from the first turn (E34 in *AfL7Ae*) is H-bonded to

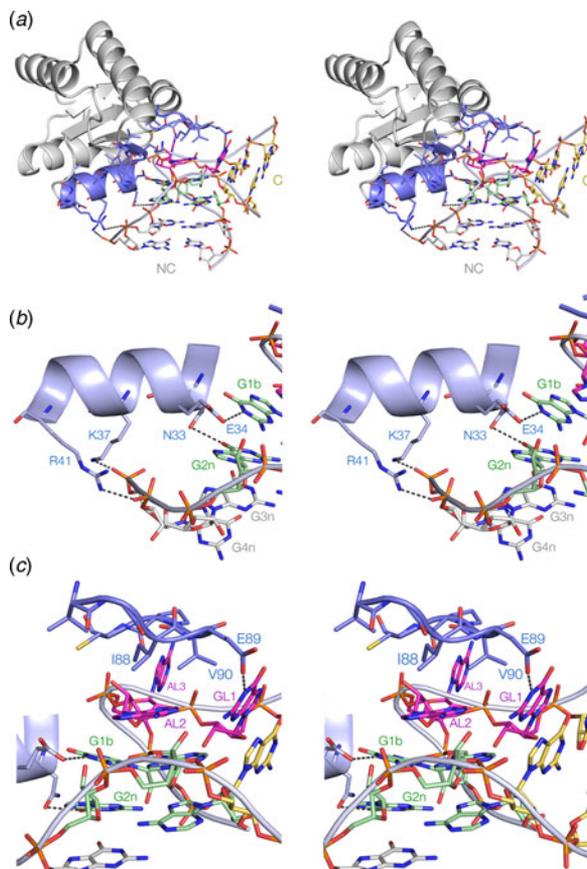


**Fig. 13.** Protein-induced folding of k-turns by ribosomal proteins other than L7Ae. Two examples are shown here, one from the large and one from the small ribosomal subunit. (a) Parallel-eye stereoscopic image of L24 bound to Kt-7 in the *H. marismortui* 50S ribosomal subunit, PDB ID 2QA4 (Kavran & Steitz, 2007). The core of the protein comprises 9 strands of  $\beta$  sheet that organize peripheral  $\alpha$  helices that interact with the RNA. Two  $\alpha$  helices at the N- (17–22) and C- (106–113) terminal ends of the protein define a cleft through which the non-loop strand of the k-turn passes. The N-terminal  $\alpha$  helix is quite basic and lies adjacent to the minor groove, making only non-specific contacts. The C-terminal  $\alpha$  helix is directed into the major groove and is somewhat equivalent to the helix of the L7Ae proteins (see Fig. 14). Asp 105 makes two H-bonds with G1b, and the nucleobase lies at the N-terminal pole of the  $\alpha$  helix dipole. There is no contact with the loop nucleotides of the k-turn. (b) Folding of *HmKt-7* induced by the binding of L24 protein analysed by FRET. The experiment was performed analogously to that in Fig. 12b.  $E_{\text{FRET}}$  rises upon addition of the L24 protein as the RNA population is driven into the folded conformation of the k-turn. The line is a fit to a two-state model of folding induced by protein binding. (c) Parallel-eye stereoscopic image of S17 bound to Kt-11 in the *T. thermophilus* 30S ribosomal subunit, PDB ID 4V5E (Weixlbaumer *et al.* 2008). The structure of S17 is very different from either L24 or L7Ae, comprising 7 strands of  $\beta$  sheet plus coil, and a single C-terminal  $\alpha$  helix. The loop of the k-turn lies between the  $\alpha$  helix and the  $\beta$  domain.  $\beta$  strand 40–45 plus the coil region 12–16 make backbone contacts with the loop strand of the C helix, while the  $\alpha$  helix passes across the apex of the k-turn in a transverse manner, above the L1 and L2 nucleobases. Tyr 95 makes a hydrophobic interaction with these. (d) Folding of *HmKt-7* induced by the binding of S17 protein analysed by FRET. In a similar manner to L24,  $E_{\text{FRET}}$  rises upon addition of the S17 protein as the RNA population is driven into the folded conformation of the k-turn. The line is a fit to a two-state model of folding induced by protein binding.

G1b N1. This seems to be a universal interaction, that is observed in the L7Ae-box C/D and 15-5 k-U4 k-turn interactions. The adjacent asparagine (N33 in *Afl7Ae*) is H-bonded to G2n O6; this also occurs in the box C/D interaction. In an equivalent complex of L7Ae with the non-standard k-turn Kt-23 of *Telohenia solenopsae* (where  $2n = A$ ) N33 is H-bonded to A2n N6 (Huang & Lilley, 2014).

- In all the structures, G1b O6 is located almost on the axis of the  $\alpha$ -helix at the N-terminal end. The exocyclic oxygen atom has a significant partial negative charge, that will be stabilized electrostatically by the positive pole of the helix dipole.
- The hydrophobic loop covers the L1 and L2 nucleobases, burying a surface area of  $732 \text{ \AA}^2$  in the L7Ae-Kt-7 complex and making good Van der Waals contact (Fig. 14c). Contact with the loop is maximized by the *syn* conformation of L2. The glutamate of the loop (E89 in *A. fulgidus* L7Ae) is H-bonded to L1 GN1. A similar interaction was observed in the L7Ae-box C/D complex.

The L7Ae family of proteins is highly specific for binding k-turns in RNA, and the structural analysis shows the basis for this. By making contacts in the groove of the NC helix (the helix) and the loop region (the hydrophobic loop) the protein is sensitive to the angle of the kink in the RNA. At the same time, the N-terminal region makes base-specific contacts with the two guanine nucleobases of the tandem G:A pairs. However, L7Ae makes no contacts with the C helix, so that k-loops (where the C-helix is replaced by a terminal loop) can also form complexes (Hamma & Ferré-D'Amaré, 2004).



**Fig. 14.** The molecular interaction between an archaeal L7Ae protein and the standard k-turn *HmKt-7*. The complex between L7Ae of *A. fulgidus* and *HmKt-7* as a stem-loop structure was solved at 2.3 Å resolution (Huang & Lilley, 2013). Parallel-eye stereoscopic images of the structure (PDB 4BW0) are shown. (a) An overall view of the complex. The L7Ae protein is shown in cartoon form, with the sections making contact with the RNA highlighted in blue. The view is from the non-loop side of the k-turn so that the C helix is directed rightwards. (b) The alpha helix located in the major groove on the outer face of the k-turn. Key amino acid side chains that interact with the RNA are shown in stick form. K37 and R41 contact the backbone while N33 and E34 make specific interactions with the guanine nucleobases of the G:A basepairs G2n and G1b, respectively. (c) The hydrophobic loop that caps the loop region of the k-turn. Note the hydrophobic side chains of I88 and V90 on the face of the protein loop contacting the L1 and L2 nucleobases. In this complex, the carboxylate group of E89 makes a specific contact with the GL1 nucleobase.

## 9. The process of binding k-turns by proteins of the L7Ae family

Binding of L7Ae proteins to k-turns result in the formation of the kinked conformation. That is true for the majority of k-turns, including many of those that will not fold in response to the addition of metal ions. There is, however, some sequence specificity, residing largely in the sequence of the loop of the k-turn, but the rules have not been fully investigated. The folding is readily followed *via* the change of end-to-end distance using FRET (e.g. Fig. 12b). Such data are fitted well by a two-state process. However, the data are characteristic of stoichiometric binding, indicating a high affinity. Indeed, it turns out that the affinity of binding is too high to measure by these means. Any RNA concentration that is detectable by fluorescence is too high to be in equilibrium in the binding process. Instead we studied the binding of *HmKt-7* by *AfL7Ae* by measuring the rates of association ( $k_{\text{on}}$ ) and dissociation ( $k_{\text{off}}$ ) (Turner & Lilley, 2008). The former was close to diffusion-controlled and was measured by stopped-flow. The affinity of binding was then calculated from the ratio ( $K_d = k_{\text{off}}/k_{\text{on}}$ ) and found to be  $K_d = 10$  pM.

We often loosely refer to the folding of k-turns on the addition of protein as ‘induced fit’, but this has a specific meaning that may be misleading here. In principle, the change in conformation in the RNA population could result from either active or passive action by the protein. Induced fit (Koshland, 1958) implies that the protein binds to the RNA in its unfolded state and somehow mechanically forces the structure to fold into the kinked structure. Alternatively, the folding of the population could be driven by conformational selection (Tsai *et al.* 2001). In this mechanism, the protein would bind a pre-existing sub-fraction that is folded (in the case of k-turns we know this exists (Goody *et al.* 2004)) and thus change the position of equilibrium for the population by mass action. Thus the key difference is whether or not the folded conformation exists prior to protein



binding, and this has been extensively discussed for a variety of macromolecular interactions (Csermely *et al.* 2010; Hammes *et al.* 2009; Okazaki & Takada, 2008; Pitici *et al.* 2002; Weikl & von Deuster, 2008; Zhou, 2010). Single-molecule FRET experiments were therefore used to study the conformation of k-turn RNA at the moment of binding (Wang *et al.* 2012). This was achieved by tethering a U1A-Afl7Ae fusion protein to a quartz slide, and then injecting fluorescent donor-acceptor-labeled *HmKt-7*. The fluorescent RNA becomes immobilized on the binding, and the FRET efficiency reports on the RNA conformation at a given moment.

It was observed that the bound molecules existed only in a high- $E_{\text{FRET}}$  state (i.e. the k-turns were folded into the kinked conformation) and remained stable so with no transitions to a low- $E_{\text{FRET}}$  state observed in many molecules analyzed. Real-time analysis of binding was used to try to catch bound RNA in a less-than-fully folded state, but none was detected down to an 8 ms frame rate (Fig. 15) (Wang *et al.* 2012). We cannot exclude the possibility that the structure is induced to form by L7Ae protein on a faster timescale. A prerequisite of the conformational selection mechanism is that the folded structures of the free and bound RNA are the same. We have determined the structure of *HmKt-7* both as free RNA and bound by *Afl7Ae* at high resolution and found that the two states are closely similar, with an all-atom RMSD of 0.83 Å. Thus very little conformational adjustment of the k-turn structure should be required on binding to L7Ae. All the present data are fully consistent with a conformational selection mechanism, although we cannot say this is completely proven at this time.

## 10. Sequence dependence of ion-dependent folding

We have already noted that k-turns such as *HmKt-7* fold into their kinked conformation on the addition of micromolar concentrations of divalent metal ions (see Fig. 12a), or millimolar concentrations of monovalent metal ions. However, we had also found that some other k-turns, exemplified by the *AfboxC/D* and human U4 k-turns remain unfolded in the presence of these ions (McPhee *et al.* 2014). This does not result from a fundamental inability to adopt the k-turn conformation because all of these sequences fold into the standard k-turn structure on binding L7Ae proteins. Yet some k-turns fold on the addition of metal ions and some do not. A difference in some critical sequence must determine the different folding characteristics, so what is this? In fact, the possible location of such a key element is rather limited. Comparing the sequences of *HmKt-7* and human U4 k-turns the tandem G:A pairs are standard, and both have  $-1b:-1n = C:G$ . We, therefore, turned to the next position in the NC helix and examined the effect of varying the 3b:3n sequence, which varies widely between k-turn sequences.

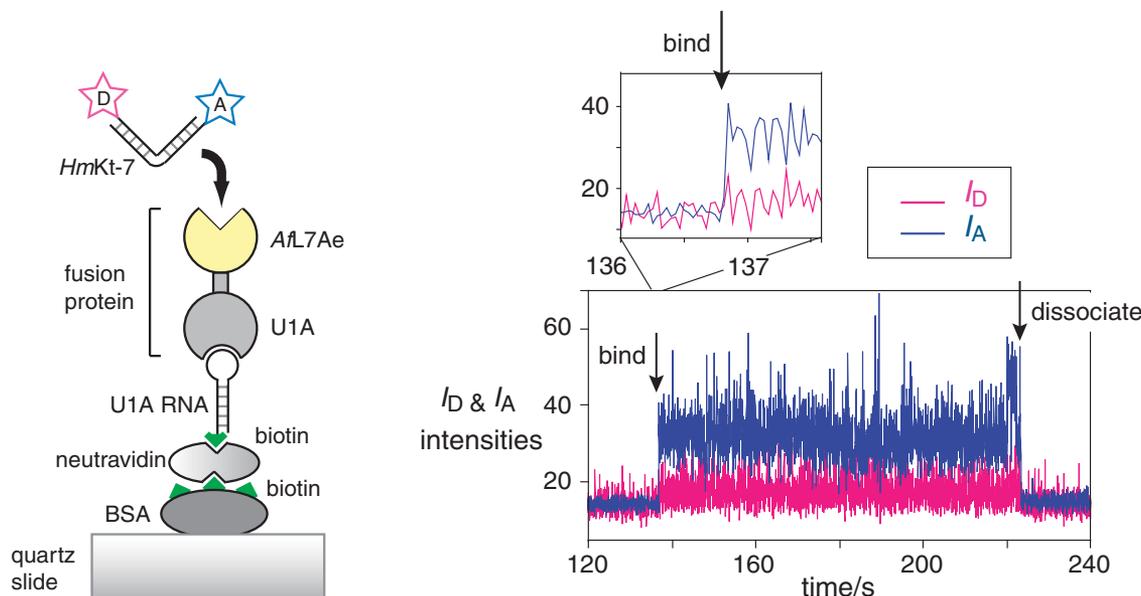
### 10.1 The 3b:3n sequence is a critical determinant of folding properties in metal ions

The role of the 3b:3n sequence was systematically investigated in the context of the *HmKt-7* sequence by construction all possible 16 variants in that position. Their folding in response to the addition of  $Mg^{2+}$  ions was studied by FRET (McPhee *et al.* 2014). The result was unambiguous, changing the 3b:3n sequence determined the folding properties of the *HmKt-7* variants between full folding (e.g. the natural sequence 3b:3n = A:G) and none at all (e.g. 3b:3n = G:C or C:G). Some underwent intermediate folding in response to the addition of metal ions (e.g. 3b:3n = A:C, or U:U) and the whole range of folding end-points was obtained for the set of k-turn variants. Yet any could be folded by the binding of L7Ae protein. All can thus adopt the k-turn conformation, but some will not do so on the addition of metal ions alone.

The dependence of the ion-dependent folding characteristics of the *HmKt-7* variants can be displayed as a  $4 \times 4$  array, where the rows are the 3b variants and the columns are the 3n variants (Fig. 16a). These have been color-coded to reflect folding ability (red = good folding and blue = poor folding in metal ions). From this, a pattern emerges, and we can define a set of ‘rules’ (McPhee *et al.* 2014).

1. 3b:3n = Watson-Crick base pairs (on the ascending diagonal) lead to poor or non-existent folding of *HmKt-7* in metal ions. This is worst for the G:C and C:G pairs. G:U is also a very weakly-folding sequence.
2. k-turns with 3b = C (second row) fold well in response to metal ions.
3. k-turns with 3n = G (third column) also fold well in response to metal ions.
4. Rule 1 takes precedence over 2 and 3 because 3b:3n = C:G fails to fold on the addition of metal ions alone.

Note that these rules are sufficient to define the folding properties of the entire folding space. Although the rules are deduced from the analysis of the folding properties of the *HmKt-7* variants, they are clearly quite general. We were able to switch the folding characteristics of the human U4 k-turn from no folding to full folding in metal ions by changing the 3b:3n sequence from G:C to A:G (McPhee *et al.* 2014), moving from a blue to a red square in Fig. 16a in a kind of ‘snakes and ladders’ move. And we had noted a strong tendency for SAM-I riboswitch k-turns to have 3b:3n sequences that confer ion-dependent folding, while U4 k-turns have a marked propensity to have 3b:3n sequences that disallow folding in metal ions. This suggests that



**Fig. 15.** Single-molecule analysis of L7Ae-induced folding (Wang *et al.* 2012). The scheme (*left*) shows the construct used. Afl7Ae (high-lighted yellow) was fused at the C-terminus of U1A protein. The fusion U1A-L7Ae was bound to a U1A RNA stem-loop biotinylated at its 5' end, that was attached to the surface of a quartz slide via neutravidin-biotin-BSA. The HmKt-7 RNA was 5'-terminally-labeled with Cy3 donor and Cy5 acceptor and becomes immobilized when it binds to the Afl7Ae of the fusion. This only appears as a fluorescent spot on the image of the surface when bound. The plots of the donor ( $I_D$  magenta) and acceptor ( $I_A$  blue) intensity as a function of time (*right*) show the moments at which a single RNA molecule binds and dissociates. Expansion of the region in which RNA binds to the protein (shown as the upper trace) reveals that the bound molecule achieves full FRET efficiency within a single frame (i.e. 16 ms), consistent with it being folded at the moment of binding.

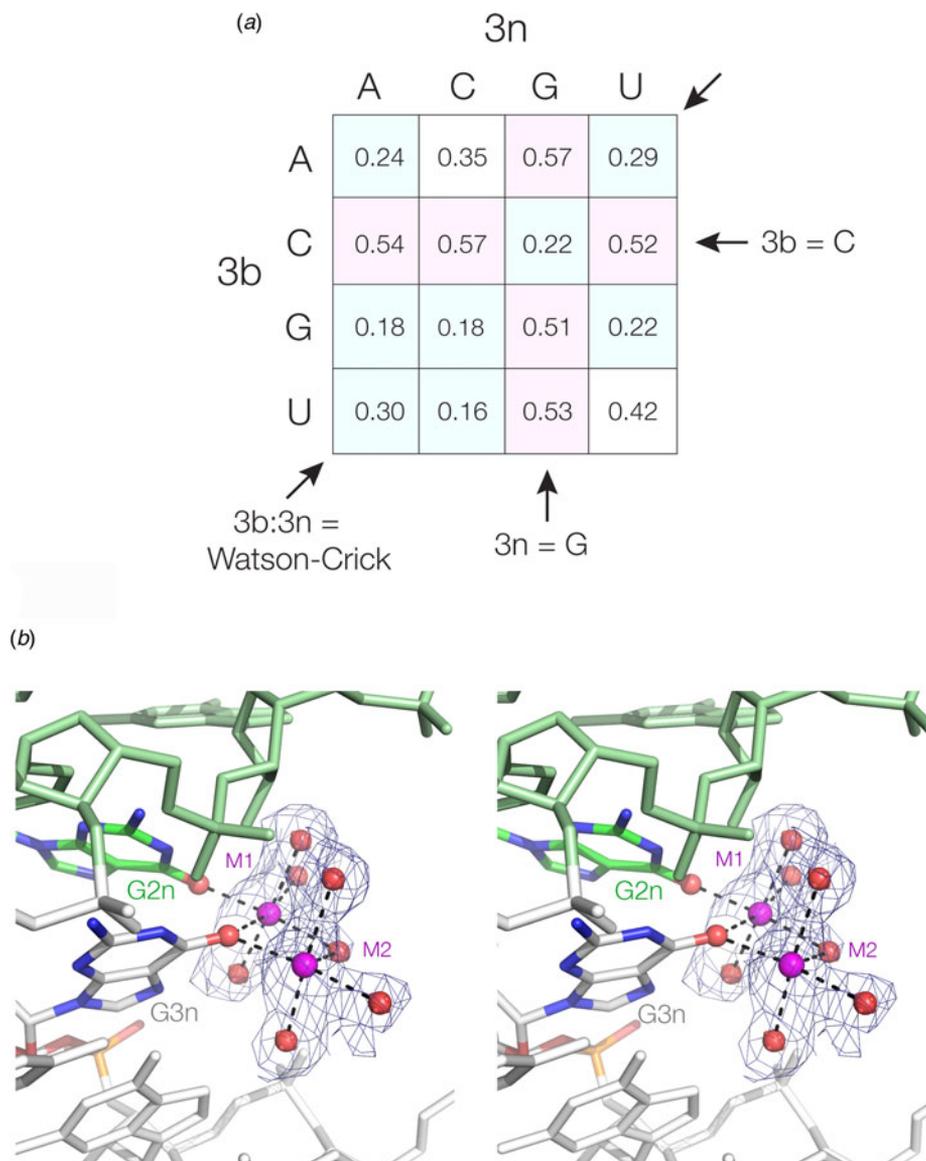
the rules deduced for *HmKt-7* apply widely. As we shall see in section 11 below, these rules, plus those determining k-turn conformation, can rationalize a large body of data and have the considerable predictive ability.

### 10.2 A molecular explanation for the $3n = G$ folding rule

While the rules for the effect of the  $3b:3n$  sequence on folding of k-turns in metal ions are largely empirical, we can explain the  $3n = G$  rule. In the course of the refinement of a 2.0 Å resolution crystal structure of an RNA duplex containing *HmKt-7*, we found electron density in the major groove adjacent to the Hoogsteen edges of the  $G2n$  and  $G3n$  nucleobases (McPhee *et al.* 2014). Further refinement revealed that the density corresponded to two octahedrally-coordinated hydrated metal ions (Fig. 16*b*). Given the symmetry, these were probably  $Mg^{2+}$  ions, although the water O-metal distances were slightly ambiguous. Importantly both ions were directly coordinated to the O6 atoms of  $G2n$  and  $G3n$  as inner-sphere complexes, and indeed one of the metal ions was coordinated to the O6 atoms of both  $G2n$  and  $G3n$ . Atomic mutagenesis experiments studying the ion-dependent folding of *HmKt-7* in which a given  $GO6$  was selectively removed (i.e. guanine to 2-aminopurine substitution) revealed that both  $G2n$  and  $G3n$  O6 atoms were required for full folding in metal ions, but that of  $G1b$  was of little significance (McPhee *et al.* 2014). The crystallographic and atomic mutagenesis data show the importance of  $G3n$  O6 on the coordination of metal ions and k-turn folding, thus explaining the  $3n = G$  rule.

### 10.3 The $-1b:-1n$ sequence is another important determinant of folding properties in metal ions

For the great majority of ribosomal and riboswitch k-turns  $-1b:-1n = C:G$ , and in our original analysis of the sequence dependence of k-turn folding properties we ignored this position, so it remained constant as C:G in all the constructs. However, we were aware that an important k-turn did not conform to the rules discussed in the previous section, the *Afbox C/D* k-turn. This has  $3b:3n = U:U$  that in the *HmKt-7* context confers reasonably good folding in metal ions, yet the *Afbox C/D* k-turn exhibits no folding in metal ions alone (although it folds on binding Afl7Ae). We, therefore, systematically exchanged four elements between the *Afbox C/D* and *HmKt-7* k-turns, i.e. the  $-1b:-1n$ ,  $3b:3n$ ,  $4b:4n$  and loop sequences, and examined the folding properties in metal ions (Ashraf *et al.* 2017). From this analysis, we deduced that the most important difference between this two k-turn is the  $-1b:-1n$  sequence, which is G:C in the *Afbox C/D* k-turn. Just inverting this basepair in *HmKt-7* (i.e. converting to  $-1b:-1n = G:C$ ) leads to very poor folding in the presence of metal ions. The reverse change in the *Afbox C/D* k-turn (generating  $-1b:-1n = C:G$ ) leads to the largest improvement in ion-dependent folding for a single change. While the  $4b:4n$  sequence had a small effect on



**Fig. 16.** The effect of the 3b:3n sequence on the folding of *HmKt-7* in metal ions. (a) 16 *HmKt-7* variants with all possible 3b:3n sequence were synthesized as fluorophore-labeled duplexes and  $E_{\text{FRET}}$  measured as a function of  $\text{Mg}^{2+}$  ion concentration (as in Fig. 12a). The end point of the titration is shown for each sequence, as a measure of the extent of folding in metal ions. These are shown in the form of a  $4 \times 4$  array, where the rows are the 3b and the columns the 3n sequences. Those that fold well (final  $E_{\text{FRET}} \geq 0.5$ ) are colored red, and those that fold poorly or not at all (final  $E_{\text{FRET}} \leq 0.3$ ) are colored blue. (b) Close-up view of metal ions directly bound to G2n and G3n (McPhee *et al.* 2014). Parallel-eye stereoscopic image of a high-resolution crystal structure of *HmKt-7* (PDB ID 4CS1) showing two hydrated metal ions in the major groove bound to G2n and G3n O6 atoms. The metal ions (magenta) and oxygen atoms (red) of water molecules of hydration and guanine O6 are shown as spheres. Direct metal ion–oxygen interactions are shown by broken lines. The electron density is taken from the  $F_o - F_c$  omit map contoured at  $2\sigma$ .

folding characteristics, this was smaller and less systematic than that for  $-1b:-1n$ . Probably, for this reason, the great majority of k-turns have evolved  $-1b:-1n = C:G$ . This includes complex k-turns such as *HmKt-11*, and the TPP riboswitch k-junctions. The biggest exception to this is the box C/D k-turns, that may be in part explained by a requirement to create an adenine  $\text{N}^6$ -methylation target in some of these (see section 12).

## 11. The distribution of sequences for natural k-turn elements

We have now seen that the 3b:3n base pair plays the key role in determining both the conformation of the k-turn (i.e. N3 or N1, Fig. 4), and the folding properties of the k-turn, whether or not it will fold in response to metal ions alone (Fig. 16). These



empirical rules are combined and summarized in Fig. 17a. It is difficult to extract any simple patterns from the distribution, beyond those discussed above for the folding characteristics. One can, however, use the array as a look-up table to predict the properties of new k-turns, or to relate the sequences to probable function. One point that emerges from inspection of the array is that 3b:3n = A:G is the only cell associated with folding in metal ions and an N3 conformation. If these properties are both required, then that is the only sequence that will confer them. We will take four classes of k-turn and analyze them in terms of their 3b:3n sequence. 4 × 4 arrays of the distributions of 3b:3n sequences may be found in the Supplementary Information of Huang and Lilley (Huang *et al.* 2016). For each group of k-turns we present the distribution of 3b:3n sequences according to their folding and conformational properties in the form of bar charts in Fig. 17b.

*Ribosomal Kt-46.* Analysis of 3181 sequences showed that Kt-46 has an extremely strong 3b:3n sequence bias, with only five cells of the 4 × 4 array having non-zero populations. 99.7% are 3b:3n = A:G, giving ion-dependent folding, and >99.9% have 3b:3n conferring the N3 conformation. Clearly, Kt-46 is required to fold in ions alone and adopt the N3 conformation. As noted above, 3b:3n = A:G is the only cell that confers both properties and thus Kt-46 is strongly biased in its 3b:3n sequence.

*Bacterial ribosomal Kt-7.* In contrast to Kt-46, analysis of 2722 Kt-7 sequences showed that >99% have 3b:3n sequences that favor the N1 conformation, of which >92% have 3b:3n = U:G. In addition, 99.7% of the sequences conform to either 3b = C or 3n = G, conferring ion-induced folding. So we deduce that bacterial Kt-7 is required to adopt the N1 conformation and fold in response to ions alone. It is interesting that in the archaeal 50S ribosomal subunit *HmKt-7* (which has the N3-conferring 3b:3n = A:G sequence) is forced to adopt an N1 conformation (section 4.1).

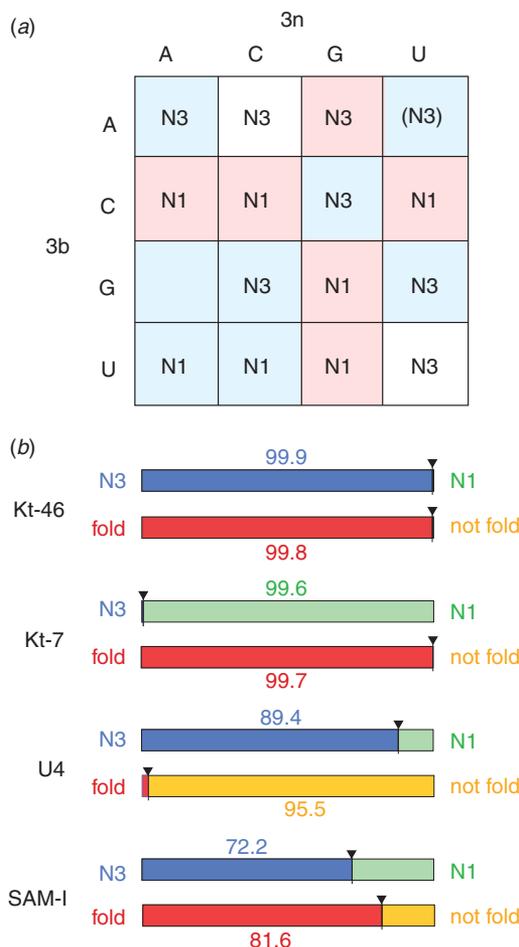
*U4 snRNA k-turns.* Sequence analysis for 9235 U4 snRNA k-turns reveals that all 16 possible 3b:3n are found, although 12 of these are populated to <1.4%. All the red colored cells (i.e. those with 3b:3n sequences that confer ion-induced folding) are avoided, with all but one populated at ≤0.3%. 63.6% have 3b:3n = G:C, conferring no folding in metal ions and an N3 conformation. The next largest (22.7%) is 3b:3n = G:U, with the same folding and conformational properties. 3b:3n = A:C is populated at 2.4%, so that ≥88.6% should exhibit weak or non-existent folding in metal ions and adopt the N3 conformation. Interestingly, the third largest fraction of sequences (7.1%) are in the 3b:3n = A:U cell, another non-ion-folding cell but with an unassigned conformation. It, therefore, seems probable that the 3b:3n = A:U sequence also confers the N3 conformation, so this has been included provisionally in Fig. 17a, although this remains unconfirmed by structural analysis at present. If that is true, however, then the fraction of U4 snRNA k-turns with an N3 conformation and weak or zero folding on the addition of metal ions would rise to a total of >95%.<sup>1</sup>

*The SAM-I riboswitch k-turn.* The SAM-I riboswitch k-turns have the broadest distribution of sequence and properties. Analysis of 4755 sequences showed that 73.4% have 3b:3n sequences that should confer an N3 conformation, while 21.7% should adopt an N1 conformation. We had previously noted that both N3 and N1 conformation k-turns could be accommodated into the SAM-I riboswitch, making it a rather useful experimental vehicle for the structural study of k-turns. ≥86.6% of the sequences should fold in the presence of metal ions, and indeed 81.7% conformed either to 3b = C or 3n = G. The most populated (60%) cell of the 4 × 4 array is 3b:3n = A:G. Thus the majority of the SAM-I riboswitch k-turns should adopt the N3 conformation and undergo folding in metal ions, but there is a significant minority that is likely to adopt the N1 conformation.

Overall we see that the ribosomal and SAM-I riboswitch k-turns have evolved 3b:3n sequences that are consistent with folding in metal ions alone. For the riboswitches, there are no known binding proteins that might assist folding, so they are probably required to fold unaided. So we might expect that k-turns within riboswitches would generally be required to fold in response to metal ions. In the lysine riboswitch the highest frequency 3b:3n = U:G > A:G > C:C. The majority of the ribosomal k-turns are bound by protein, and yet perhaps during the biogenesis of the ribosome they are required to adopt the correct stable fold before protein binding. We see that both Kt-46 and Kt-7 have 3b:3n sequences that predispose the k-turns to fold in metal ions, despite being of different conformations (N3 and N1 respectively). In the *H. marismortui* 50S ribosomal subunit Kt-46, Kt-7, Kt-58, Kt-78 and the k-junction J4,5 all have 3b:3n = A:G. The k-turn of the human U4 snRNA contrasts with the ribosomal and riboswitch k-turns in that it has evolved 3b:3n sequences that prevent ion-induced folding. Instead, they have selected 3b:3n sequences that confer N3 conformation but will not fold in metal ions alone. These will be unable to fold stably until the L7Ae or 15.5k protein binds. We conclude that during the formation of the spliceosomal B-complex the folding is incomplete until protein binding has occurred. The sequence of the RNA ensures that it cannot adopt a stable fold in the absence of the protein. Note that there is a second way to achieve a lack of ion-dependent folding so that a k-turn must await the binding of a protein. Some of the box C/D k-turns (e.g. that of *A. fulgidus*) have selected –1b:–1n = G:C, so fail to fold on the addition of metal ions despite having 3b:3n = U:U (section 10.3).

Another interesting case is that of the archaeal k-turns in the 5'-UTR of the mRNA encoding L7Ae protein (introduced in section 2). The 5'-terminus of the mRNA folds into a hairpin that includes the Shine-Dalgarno sequence that overlaps the

<sup>1</sup> We have recently solved a new k-turn structure that has 3b:3n = A:U, and this does indeed adopt the N3 conformation.



**Fig. 17.** The influence of the 3b:3n sequence on conformation and folding of k-turns. (a) A summary  $4 \times 4$  array where the rows are the 3b and the columns the 3n sequences, combining effect of the identity of the 3b and 3n positions on both N3 or N1 conformation (Huang *et al.* 2016), and whether or not the k-turn will fold in response to addition of metal ions (McPhee *et al.* 2014). The latter is shown by color as in Fig. 16, where red cells denote folding in metal ions and blue ones denote failure to fold on the addition of metal ions. The resulting conformation is shown as the collective result from the analysis of natural k-turns and the HmKt-7 variants as shown in Fig. 4. In the case of 3b:3n = A:G, all but one (i.e. *HmKt-7* within the ribosome) are N3 structures so this cell is designated N3. The 3b:3n = A:U cell is provisionally assigned as N3 in the light of the distribution of U4snRNA sequences. (b) Bar plots showing the occurrence of 3b:3n sequence variation for four k-turns (Huang *et al.* 2016). For each k-turn type two bar plots are shown. The top one shows the fraction of 3b:3n sequences that confer N3 (blue) vs N1 (green) conformation. The lower one shows the fraction of 3b:3n sequences that confer ion-induced folding (red) vs. inability to fold in metal ions (yellow).

bulged strand of a standard simple k-turn. If L7Ae binds to the k-turn it should prevent association with the ribosome, thereby repressing its own translation (Daume *et al.* 2017). A corresponding k-turn-containing stem-loop is found in a wide variety of archaea, including *Crenarchaeota* and *Euryarchaeota*, and it is instructive to examine the predicted folding properties of these. The repression mechanism would suggest a requirement that the k-turn does not fold until the L7Ae protein binds, i.e. they should not undergo ion-induced folding. The k-turn of *Sulfolobus acidocaldarius* has 3b:3b = A:U, and  $-1b:-1n = G:C$ , both of which confer absence of folding in metal ions alone. The corresponding k-turns of *A. fulgidus* and *Methanosarcina acetivorans* also have 3b:3b = A:U, while that of *Thermococcus kodakaraensis* have 3b:3b = G:U, so all are non-ion-folding k-turns. These have the folding properties anticipated for a genetic switch regulated by the binding of the gene product L7Ae. We also note that all the 3b:3b sequences are predicted to confer an N3 conformation.

### 11.1 Application of the folding rules to a set of k-turns of unknown structure

As an exercise, we can apply the empirical set of rules to the putative k-turn sequences (Fig. 11) that we identified in the structured RNA species recently presented by Breaker and co-workers as discussed in section 6.5. The predicted properties are summarized in Table 2. Apart from HOLDH the k-turns have  $-1b:-1n = C:G$  and three of the four also have A:G as



the most frequent 3b:3n sequence. Thus most are expected to fold in response to the addition of metal ions. The k-turn identified in the *Actinomyces-1* sequence has 3b:3n = C:C > C:U indicating that this will also fold in response to the presence of metal ions. The folding properties of the HOLDH k-turn are harder to predict because the -1b:-1n = A:U has not been studied experimentally; however it has the 3b:3n = A:G sequence that is the best for ion-induced folding.

The drum and RAGATH-1 k-turns have 3b:3n sequences indicating a strong tendency to adopt the N3 conformation. By contrast, the *Actinomyces-1* k-turn has 3b:3n sequences indicating an N1 conformation. HOLDH is the least clear case. The most frequent 3b:3n sequence is A:G, conferring N3 conformation, whereas the next highest is C:A that confers N1 conformation.

The DUF-3268 k-junction has 3b:3n = C:G. If this follows the behavior of the simple k-turns then it should not fold in metal ions and adopt an N3 conformation. However, the *A. thaliana* TPP riboswitch k-junction also has 3b:3n = C:G yet adopts an N1 conformation (see section 6.3). That indicates the rules cannot be applied to k-junctions with confidence. It seems that the k-junctions have a strong tendency to adopt the N1 conformation, and thus we expect this one to do so as well.

## 12. Effect of N<sup>6</sup>-methyladenine inclusion on k-turn folding of snoRNAs

Using X-ray crystallography we have recently shown that *trans* sugar edge-Hoogsteen G:A basepairs are disrupted if the adenine nucleobase is methylated at N6 (Huang *et al.* 2017). This is in marked contrast to *cis*-Watson-Crick A:U and A:G basepairs that form normally with N<sup>6</sup>-methyladenine (Fig. 18). Since *trans* sugar edge-Hoogsteen G:A basepairs form the core of the k-turn it is perhaps unsurprising that N<sup>6</sup>-methylation of the 1n adenine prevents folding of *HmKt-7* (S. Ashraf, LH and DMJL, unpublished data). In principle, this might form the basis of a regulatory mechanism controlled by reversible methylation in cells.

The principal methyl transferase in eukaryotes that is responsible for generating N<sup>6</sup>-methyladenine in RNA is METTL3-METTL14 (Liu *et al.* 2014). This has the preferred target DRACH (where D denotes A, G or U, R is A or G and H is A, C or U), with GAC as the most common site of methylation (Dominissini *et al.* 2012; Meyer *et al.* 2012a). The central A is the site of methylation. Thus for the A1n of a k-turn to become methylated would require -1n to be cytosine, yet in the majority of k-turns -1b:-1n = C:G. However, we had noted that in some box C/D snoRNA k-turns the -1b:-1n base pair is inverted, so that -1n = C thus creating the required GAC methylation target. We recently performed a bioinformatic data mining analysis of human box C/D and C'/D' snoRNA sequences, finding that 27 have the required GAC target sequence, of which 14 have been found to be methylated in the cell (Huang *et al.* 2017). Thus a significant sub-set of human box C/D and C'/D' k-turn sequences undergo adenine N<sup>6</sup>-methylation at the critical A1n position.

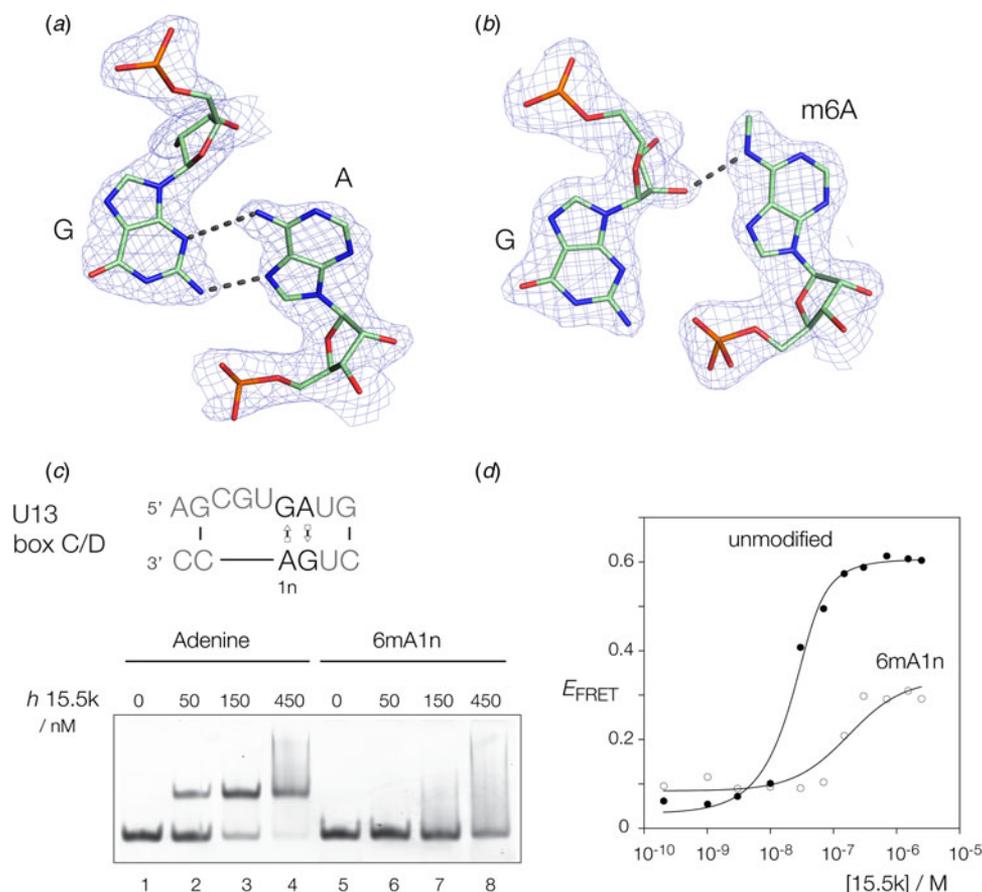
Assembly of the box C/D snoRNP species requires the ordered association of a series of proteins with the RNA (Bleichert *et al.* 2009; Kiss-Laszlo *et al.* 1996; Lin *et al.* 2011; Tran *et al.* 2003; Tycowski *et al.* 1996; Watkins *et al.* 2000; Xue *et al.* 2010; Ye *et al.* 2009). In the first stage, the 15.5k protein binds to the k-turns. Once this has occurred the next proteins (Nop56 and Nop58 in eukaryotes) bind. These are analogous to Prp31 that binds to the complex of 15.5k with the U4 k-turn in the spliceosomal B complex (Liu *et al.* 2007) (see section 2). Lastly, two molecules of the 2'O-methyl transferase fibrillarin associate with the complex to generate the catalytically-active snoRNP. Importantly Watkins *et al.* (2002) found that if the binding of 15.5k to the human box C/D k-turn is prevented by sequence changes known to interfere with k-turn folding, assembly of the box C/D snoRNP does not occur. We showed *in vitro* that inclusion of N<sup>6</sup>-methyladenine at the 1n position of some representative human box C/D k-turns both prevented specific 15.5k protein binding, and the proper folding of the k-turn (Huang *et al.* 2017) (Fig. 18). While this is perhaps a step away from demonstrating an *in vivo* regulatory mechanism based on reversible methylation of k-turn RNA, it is very plausible and further experiments to test this are under way.

## 13. The k-turn as a building unit in nano-construction

The precise geometry of the k-turn suggests a possible use as a building block in the construction of molecular nano-objects (Jasinski *et al.* 2017). RNA has previously been used as a material for nano-scale engineering, and helical bulges have been used to construct a square or triangular molecules (Boerneke *et al.* 2016; Dibrov *et al.* 2011; Ohno & Inoue, 2015). Moreover, this provides a test of our conformational understanding, whether or not we can manipulate the structure at will to design and generate molecular assemblies. Saito and colleagues (Ohno *et al.* 2011) made the first steps on this path, using atomic force microscopy (AFM) to visualize triangular objects comprising three k-turns bound by the L7Ae protein.

### 13.1 Association of a two-k-turn unit in crystal lattices

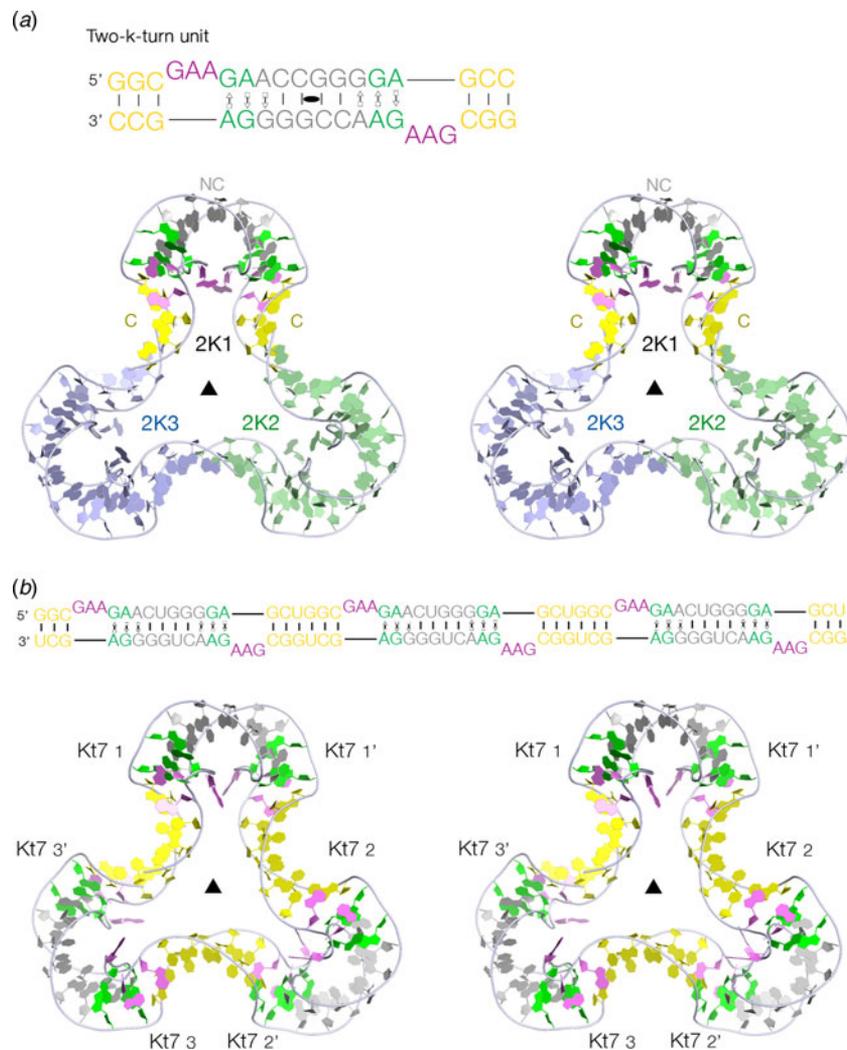
The basic structural element used in our studies comprised a two-fold symmetric duplex with two *HmKt-7* sequences (Fig. 19a) (Huang & Lilley, 2016). Their loops are located on opposite strands and the k-turns are associated through



**Fig. 18.** The effect of N<sup>6</sup>-methyladenine (N<sup>6</sup>mA) inclusion on *trans* sugar-Hoogsteen G:A base pairing, and its potential effect on box C/D snoRNP assembly (Huang *et al.* 2017). (a) and (b) Crystal structures of G:A (PDB ID 5LR3; at 1.65 Å resolution) and G: N<sup>6</sup>mA (PDB ID 5LR4 at 1.72 Å resolution) base pairs in a duplex as tandem G:A, A:G pairs flanked by G:U base pairs. The unmodified G:A base pairs (a) form sheared *trans*-sugar-Hoogsteen G:A base pairs like those found in the k-turns. By contrast the in the G:<sup>6</sup>mA pair (b) there are no H-bonds connecting the nucleobases but rather N<sup>6</sup>MAN6 forms an H-bond with GO2'. (c) Binding of human 15.5k protein (L7Ae ortholog) to the box C/D k-turn of human U13 snRNA. The sequence of the k-turn is shown. The RNA was prepared with either adenine or N<sup>6</sup>mA at the 1n position of the k-turn. Radioactive RNA was incubated with increasing concentrations of 15.5k protein, and free RNA and complexes with bound protein separated by electrophoresis in polyacrylamide. 15.5k-bound complexes migrate as retarded species. RNA with unmodified adenine (tracks 1–4) binds 15.5k protein as a discrete retarded complex, with a smear of non-specific binding at the highest protein concentration (track 4). By contrast, RNA with N<sup>6</sup>mA at the 1n position (tracks 5–8) does not give rise to a specific complex, but only non-specific binding. (d) Analysis of 15.5k protein-induced folding of human U13 snRNA using steady-state FRET. FRET efficiency ( $E_{\text{FRET}}$ ) was measured using RNA terminally labeled with fluorescein donor and Cy-3 acceptor analogous to that in Fig. 12. The RNA was prepared with either adenine (filled circles) or N<sup>6</sup>mA (open circles) at the 1n position of the k-turn.  $E_{\text{FRET}}$  was measured as a function of added 15.5k protein concentration, and the data fitted to a simple binding isotherm. That for the unmodified RNA undergoes a two-state folding process similar to that shown for Afl7Ae binding to HmKt-7 (Fig. 12b). In contrast, the N<sup>6</sup>mA-containing RNA requires a higher concentration of 15.5k protein and achieves a significantly lower end-point. It is likely that this corresponds to the non-specific binding observed in the electrophoretic experiments shown in part b.

their NC helices, with ten basepairs separating the two loops, including the G:A pairs. The structure had been determined by X-ray crystallography at high resolution, revealing it to have an overall horse-shoe shape (McPhee *et al.* 2014). It has a central dyad but is not fully planar. Within the structure, the HmKt-7 elements adopt their standard N3 k-turn structure. The potential of this two-k-turn unit (the 2 K unit) became apparent when intermolecular contacts between units within the crystal lattice were examined. In fact, two crystal forms were obtained. One had tetragonal symmetry (P<sub>4</sub>2<sub>2</sub>), where two 2 K units form a dumbbell shape. The second had hexagonal symmetry (P<sub>6</sub>3<sub>2</sub>2), and in this case, three 2 K units were associated to form a triangular shape (Fig. 19a). In both cases, the interaction between molecules occurred by end-to-end coaxial stacking of the C-helices, with preservation of helical symmetry and stacking (base plane separation of 3.3 Å).

Two further structures were obtained for 2 K units comprising HmKt-7 with altered 3b:3n sequences, crystallized as complexes with bound Afl7Ae protein. That with 3bU:3nU crystallized with orthorhombic symmetry (P<sub>2</sub><sub>1</sub>2<sub>1</sub>), and formed a



**Fig. 19.** The k-turn as a unit for nano-construction. The structure of a triangular molecular object comprising six k-turns (Huang & Lilley, 2016). (a) The two k-turn unit (2 K unit) is a duplex with two *HmKt-7* sequences related by a two-fold rotation so the loops are on opposite strands and they are connected by their common NC helix (PDB ID 4CS1; at 2.0 Å resolution). They pack in two ways to form a crystal lattice. The parallel-eye stereoscopic image shows that in which three 2 K units associated by end-to-end stacking, related by the three-fold rotation axis shown. (b) A single RNA duplex containing three 2 K units, i.e. six *HmKt-7* sequences with alternating polarity. In the crystal, the molecule adopts a triangular conformation with a random rotational setting within the lattice creating a crystallographic three-fold rotation axis as shown (PDB ID 5G4 T; at 2.75 Å resolution). Thus the molecule has a pseudo-symmetry with symmetry point group  $D_3$ .

triangular association in the lattice based on coaxial end-to-end stacking as before. However, the central pore was two-fold larger than that found for the protein-free triangular structure based on unmodified Kt-7. By contrast, the 2 K unit based on *HmKt-7* 3bG:3nC crystallized with monoclinic symmetry ( $C1\ 2\ 1$ ), and in this case, four 2 K units (i.e. 8 k-turns and 8 L7Ae molecules) adopted a square assembly by coaxial end-to-end stacking. Thus by small changes of the sequence the geometry and dimensions of the assemblies can be radically altered.

### 13.2 The molecular structure of a six-k-turn nano-object

With this background and the experience of manipulating the structural properties of the basic building block a single duplex RNA was synthesized containing three consecutive 2 K using (derived from a single strand of 57 nt), i.e. six *HmKt-7* motifs alternating in polarity. This crystallized without requiring the addition of L7Ae, in a hexagonal space group ( $P6_322$ ), diffracting to 2.75 Å resolution (Huang & Lilley, 2016). The molecule folded into a triangular shape with end-on-end coaxial stacking at the open end (Fig. 19b). Remarkably, this was so perfectly formed that the individual molecules crystallized with static disorder, making no distinction between a covalently continuous junction between C helices and the open end. Thus each



phosphate group connecting the two-k-turn units had an occupancy of 2/3. Effectively the molecule has an apparent point group symmetry  $D_3$ , with a shape reminiscent of a cyclohexane molecule in its chair conformation.

These studies demonstrate that the k-turn is an effective and versatile building block in the assembly of nano-scale molecular objects, and demonstrates just why k-turns are widely used in natural RNA molecules to mediate tertiary contacts and organize architecture over a large scale. The k-turn is one of nature's most useful nano-scale structure-building elements in functional RNA molecules.

## 14. In conclusion

The k-turn is an extremely widespread structural motif in RNA architecture, involved in the construction of many key RNA structures of the cell that are important in translation, RNA processing, and genetic control. We have seen that the formation of k-turns is the essential first step in the formation of RNA-protein assemblies such as the box C/D snoRNP and that this can be controlled by methylation of a key adenine nucleotide in the k-turn. The k-turn structure has been the subject of intense study for the last decade or more so that it is arguably the most well-understood RNA motif. We now comprehend these structures in considerable detail, and how folding and conformational properties are determined by local sequence. We see how the sequences of individual k-turns fit their structural properties to the biological requirements of their situation. Moreover, we now understand the structure to the point at which we can use it as the basis for the design of defined nano-structural objects. This perhaps is no surprise, since the k-turn is essentially a major component of nature's nano-technological construction set.

## Acknowledgements

We thank Terry Goody, Jia Liu, Ben Turner, Tim Wilson, Peter Daldrop, Kersten Schroeder, Jia Wang, Jo Ouellet, Scott McPhee and Saira Ashraf for their contributions to studies of k-turn structures in RNA, and all past and present members of the DMJL lab for many discussions. We thank our collaborators, especially Xuesong Shi, Dan Herschlag and Benoit Masquida, and Cancer Research UK, the Wellcome Trust and the HFLP for funding. X-ray crystallographic data have been collected at the Diamond and ESRF synchrotrons.

## References

- AHL, V., KELLER, H., SCHMIDT, S. & WEICHENRIEDER, O. (2015). Retrotransposition and crystal structure of an Alu RNP in the ribosome-stalling conformation. *Molecular Cell* **60**(5), 715–727.
- ASHRAF, S., HUANG, L. & LILLEY, D. M. J. (2017). Sequence determinants of the folding properties of box C/D kink-turns in RNA. *RNA* **23**(12), 1927–1935.
- BAIRD, N. J. & FERRE-D'AMARE, A. R. (2013). Modulation of quaternary structure and enhancement of ligand binding by the K-turn of tandem glycine riboswitches. *RNA* **19**(2), 167–176.
- BAIRD, N. J., ZHANG, J., HAMMA, T. & FERRE-D'AMARE, A. R. (2012). Ybxf and YlxQ are bacterial homologs of L7Ae, and bind K-turns but not K-loops. *RNA* **18**(4), 759–770.
- BAN, N., NISSEN, P., HANSEN, J., MOORE, P. B. & STEITZ, T. A. (2000). The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science* **289**(5481), 905–920.
- BAYRAK, C. S., KIM, N. & SCHLICK, T. (2017). Using sequence signatures and kink-turn motifs in knowledge-based statistical potentials for RNA structure prediction. *Nucleic Acids Research* **45**(9), 5414–5422.
- BELTRAME, M. & TOLLERVEY, D. (1995). Base pairing between U3 and the pre-ribosomal RNA is required for 18S rRNA synthesis. *The EMBO Journal* **14**(17), 4350–4356.
- BEN-SHEM, A., GARREAU, DE, LOUBRESSE, N., MELNIKOV, S., JENNER, L., YUSUPOVA, G. & YUSUPOV, M. (2011). The structure of the eukaryotic ribosome at 3.0 Å resolution. *Science* **334**(6062), 1524–1529.
- BLEICHERT, F., GAGNON, K. T., BROWN, B. A., MAXWELL, E. S., LESCHZINER, A. E., UNGER, V. M. & BASERGA, S. J. (2009). A dimeric structure for archaeal box C/D small ribonucleoproteins. *Science* **325**(5946), 1384–1387.
- BLOUIN, S. & LAFONTAINE, D. A. (2007). A loop-loop interaction and a K-turn motif located in the lysine aptamer domain are important for the riboswitch gene regulation control. *RNA* **13**(8), 1256–12567.
- BOERNEKE, M. A., DIBROV, S. M. & HERMANN, T. (2016). Crystal-structure-guided design of self-assembling RNA nanotriangles. *Angewandte Chemie* **55**(12), 4097–4100.
- CHO, I. M., LAI, L. B., SUSANTI, D., MUKHOPADHYAY, B. & GOPALAN, V. (2010). Ribosomal protein L7Ae is a subunit of archaeal RNase P. *Proceedings of the National Academy of Sciences USA* **107**(33), 14573–14578.
- CLEMONS, W. M., MAY, J. L. C., WIMBERLY, B. T., MCCUTCHEON, J. P., CAPEL, M. S. & RAMAKRISHNAN, V. (1999). Structure of a bacterial 30S ribosomal subunit at 5.5 Å resolution. *Nature* **400**(6747), 833–841.



- CRUZ, J. A. & WESTHOF, E. (2011). Sequence-based identification of 3D structural modules in RNA with RMDetect. *Nature Methods* **8**(6), 513–521.
- CSERMELY, P., PALOTAI, R. & NUSSINOV, R. (2010). Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. *Trends in Biochemical Sciences* **35**(10), 539–546.
- CSERNOCH, L., BERNENGO, J. C., SZENTESI, P. & JACQUEMOND, V. (1998). Measurements of intracellular Mg<sup>2+</sup> concentration in mouse skeletal muscle fibers with the fluorescent indicator mag-indo-1. *Biophysical Journal* **75**(2), 957–967.
- DALDROP, P. & LILLEY, D. M. J. (2013). The plasticity of a structural motif in RNA: structural polymorphism of a kink turn as a function of its environment. *RNA* **19**(0), 357–364.
- DALDROP, P., MASQUIDA, B. & LILLEY, D. M. J. (2013). The functional exchangeability of pk- and k-turns in RNA structure. *RNA Biology* **10**, 445–452.
- DAUME, M., UHL, M., BACKOFEN, R. & RANDAU, L. (2017). RIP-Seq Suggests translational regulation by L7Ae in archaea. *mBio* **8**(4), e00730-17.
- DIBROV, S. M., MCLEAN, J., PARSONS, J. & HERMANN, T. (2011). Self-assembling RNA square. *Proceedings of the National Academy of Sciences USA* **108**(16), 6405–6408.
- DOMINISSINI, D., MOSHITCH-MOSHKOVITZ, S., SCHWARTZ, S., SALMON-DIVON, M., UNGAR, L., OSENBURG, S., CESARKAS, K., JACOB-HIRSCH, J., AMARIGLIO, N., KUPIEC, M., SOREK, R. & RECHAVI, G. (2012). Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature* **485**(7397), 201–206.
- ESQUAQUI, J. M., SHERMAN, E. M., YE, J. D. & FANUCCI, G. E. (2016). Conformational flexibility and dynamics of the internal loop and helical regions of the kink-turn motif in the glycine riboswitch by site-directed spin-labeling. *Biochemistry* **55**(31), 4295–4305.
- FORSTER, T. (1948). Zwischenmolekulare energiewanderung und fluoreszenz. *Annales de Physique* **2**, 55–75.
- GARST, A. D., HEROUX, A., RAMBO, R. P. & BATEY, R. T. (2008). Crystal structure of the lysine riboswitch regulatory mRNA element. *Journal of Biological Chemistry* **283**(33), 22347–22351.
- GEARY, C., CHWOROS, A. & JAEGER, L. (2011). Promoting RNA helical stacking via A-minor junctions. *Nucleic Acids Research* **39**(3), 1066–1080.
- GERBI, S. A., BOROVJAGIN, A. V., EZROKHI, M. & LANGE, T. S. (2001). Ribosome biogenesis: role of small nucleolar RNA in maturation of eukaryotic rRNA. *Cold Spring Harbor Symposia on Quantitative Biology* **66**, 575–590.
- GOODY, T. A., MELCHER, S. E., NORMAN, D. G. & LILLEY, D. M. J. (2004). The kink-turn motif in RNA is dimorphic, and metal ion dependent. *RNA* **10**, 254–264.
- HAMMA, T. & FERRÉ-D'AMARÉ, A. R. (2004). Structure of protein L7Ae bound to a K-turn derived from an archaeal box H/ACA sRNA at 1.8 Å resolution. *Structure* **12**(5), 893–903.
- HAMMES, G. G., CHANG, Y. C. & OAS, T. G. (2009). Conformational selection or induced fit: a flux description of reaction mechanism. *Proceedings of the National Academy of Sciences USA* **106**(33), 13737–13741.
- HUANG, L., ASHRAF, S., WANG, J. & LILLEY, D. M. (2017). Control of box C/D snoRNP assembly by N6-methylation of adenine. *EMBO Reports* **18**, 1631–1645.
- HUANG, L. & LILLEY, D. M. J. (2013). The molecular recognition of kink turn structure by the L7Ae class of proteins. *RNA* **19**, 1703–1710.
- HUANG, L. & LILLEY, D. M. J. (2014). Structure of a rare non-standard sequence k-turn bound by L7Ae protein. *Nucleic Acids Research* **42**(7), 4734–4740.
- HUANG, L. & LILLEY, D. M. J. (2016). A quasi-cyclic RNA nano-scale molecular object constructed using kink turns. *Nanoscale* **8**, 15189–15195.
- HUANG, L., WANG, J. & LILLEY, D. M. J. (2016). A critical base pair in k-turns determines the conformational class adopted, and correlates with biological function. *Nucleic Acids Research* **44**(11), 5390–5398.
- JASINSKI, D., HAQUE, F., BINZEL, D. W. & GUO, P. (2017). Advancement of the emerging field of RNA nanotechnology. *ACS Nano* **11**(2), 1142–1164.
- KAVRAN, J. M. & STEITZ, T. A. (2007). Structure of the base of the L7/L12 stalk of the *Haloarcula marismortui* large ribosomal subunit: analysis of L11 movements. *Journal of Molecular Biology* **371**(4), 1047–1059.
- KISS-LASZLO, Z., HENRY, Y., BACHELLERIE, J. P., CAIZERGUES-FERRER, M. & KISS, T. (1996). Site-specific ribose methylation of preribosomal RNA: a novel function for small nucleolar RNAs. *Cell* **85**(7), 1077–1088.
- KLEIN, D. J., SCHMEING, T. M., MOORE, P. B. & STEITZ, T. A. (2001). The kink-turn: a new RNA secondary structure motif. *The EMBO Journal* **20**(15), 4214–4221.
- KOONIN, E. V., BORK, P. & SANDER, C. (1994). A novel RNA-binding motif in omnipotent suppressors of translation termination, ribosomal proteins and a ribosome modification enzyme? *Nucleic Acids Research* **22**(11), 2166–2167.
- KOSHLAND, D. E. (1958). Application of a theory of enzyme specificity to protein synthesis. *Proceedings of the National Academy of Sciences USA* **44**, 98–104.
- KUHN, J. F., TRAN, E. J. & MAXWELL, E. S. (2002). Archaeal ribosomal protein L7 is a functional homolog of the eukaryotic 15.5kD/Snu13p snoRNP core protein. *Nucleic Acids Research* **30**(4), 931–941.
- LAI, L. B., TANIMOTO, A., LAI, S. M., CHEN, W. Y., MARATHE, I. A., WESTHOF, E., WYSOCKI, V. H. & GOPALAN, V. (2017). A novel double kink-turn module in euryarchaeal RNase P RNAs. *Nucleic Acids Research* **45**(12), 7432–7440.
- LESCOUTE, A., LEONTIS, N. B., MASSIRE, C. & WESTHOF, E. (2005). Recurrent structural RNA motifs, isostericity matrices and sequence alignments. *Nucleic Acids Research* **33**(8), 2395–2409.
- LI, L. & YE, K. (2006). Crystal structure of an H/ACA box ribonucleoprotein particle. *Nature* **443**(7109), 302–307.



- LI-SMERIN, Y., LEVITAN, E. S. & JOHNSON, J. W. (2001). Free intracellular  $Mg^{2+}$  concentration and inhibition of NMDA responses in cultured rat neurons. *The Journal of Physiology* **533**(Pt 3), 729–743.
- LILLEY, D. M. J. (1995). Kinking of DNA and RNA by base bulges. *Proceedings of the National Academy of Sciences USA* **92**(16), 7140–7142.
- LILLEY, D. M. J. (2008). Analysis of branched nucleic acid structure using comparative gel electrophoresis. *Quarterly Reviews of Biophysics* **41**(1), 1–39.
- LIN, J., LAI, S., JIA, R., XU, A., ZHANG, L., LU, J. & YE, K. (2011). Structural basis for site-specific ribose methylation by box C/D RNA protein complexes. *Nature* **469**(7331), 559–563.
- LIU, J. & LILLEY, D. M. J. (2007). The role of specific 2'-hydroxyl groups in the stabilization of the folded conformation of kink-turn RNA. *RNA* **13**(2), 200–210.
- LIU, J., YUE, Y., HAN, D., WANG, X., FU, Y., ZHANG, L., JIA, G., YU, M., LU, Z., DENG, X., DAI, Q., CHEN, W. & HE, C. (2014). A METTL3-METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation. *Nature Chemical Biology* **10**(2), 93–95.
- LIU, S., LI, P., DYBKOV, O., NOTTROT, S., HARTMUTH, K., LUHRMANN, R., CARLOMAGNO, T. & WAHL, M. C. (2007). Binding of the human Prp31 Nop domain to a composite RNA-protein platform in U4 snRNP. *Science* **316**(5821), 115–120.
- MAO, H., WHITE, S. A. & WILLIAMSON, J. R. (1999). A novel loop-loop recognition motif in the yeast ribosomal protein L30 autoregulatory RNA complex. *Nature Structural Biology* **6**, 1139–1147.
- MARMIER-GOURRIER, N., CLERY, A., SENTRY-SEGAULT, V., CHARPENTIER, B., SCHLOTTER, F., LECLERC, F., FOURNIER, R. & BRANLANT, C. (2003). A structural, phylogenetic, and functional study of 15.5-kD/Snu13 protein binding on U3 small nucleolar RNA. *RNA* **9**(7), 821–838.
- MCKEEGAN, K. S., DEBIEUX, C. M., BOULON, S., BERTRAND, E. & WATKINS, N. J. (2007). A dynamic scaffold of pre-snRNP factors facilitates human box C/D snoRNP assembly. *Molecular and Cellular Biology* **27**(19), 6782–6793.
- MCPHEE, S. A., HUANG, L. & LILLEY, D. M. J. (2014). A critical base pair in k-turns that confers folding characteristics and correlates with biological function. *Nature Communications* **5**, 5127.
- MEYER, K. D., SALETORRE, Y., ZUMBO, P., ELEMENTO, O., MASON, C. E. & JAFFREY, S. R. (2012a). Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. *Cell* **149**(7), 1635–1646.
- MEYER, M., WESTHOF, E. & MASQUIDA, B. (2012b). A structural module in RNase P expands the variety of RNA kinks. *RNA Biology* **9**(3), 254–260.
- MONTANGE, R. K. & BATEY, R. T. (2006). Structure of the S-adenosylmethionine riboswitch regulatory mRNA element. *Nature* **441**(7097), 1172–1175.
- MOORE, T., ZHANG, Y., FENLEY, M. O. & LI, H. (2004). Molecular basis of box C/D RNA-protein interactions; cocrystal structure of archaeal L7Ae and a box C/D RNA. *Structure* **12**(5), 807–818.
- NOTTROT, S., HARTMUTH, K., FABRIZIO, P., URLAUB, H., VIDOVIC, I., FICNER, R. & LUHRMANN, R. (1999). Functional interaction of a novel 15.5kD [U4/U6.U5] tri-snRNP protein with the 5' stem-loop of U4 snRNA. *The EMBO Journal* **18**(21), 6119–6133.
- NOTTROT, S., URLAUB, H. & LUHRMANN, R. (2002). Hierarchical, clustered protein interactions with U4/U6 snRNA: a biochemical role for U4/U6 proteins. *The EMBO Journal* **21**(20), 5527–5538.
- OHNO, H. & INOUE, T. (2015). Designed regular tetragon-shaped RNA-protein complexes with ribosomal protein L1 for bionanotechnology and synthetic biology. *ACS Nano* **9**(5), 4950–4956.
- OHNO, H., KOBAYASHI, T., KABATA, R., ENDO, K., IWASA, T., YOSHIMURA, S. H., TAKEYASU, K., INOUE, T. & SAITO, H. (2011). Synthetic RNA-protein complex shaped like an equilateral triangle. *Nature Nanotech* **6**(2), 116–120.
- OKAZAKI, K. & TAKADA, S. (2008). Dynamic energy landscape view of coupled binding and protein conformational change: induced-fit versus population-shift mechanisms. *Proceedings of the National Academy of Sciences USA* **105**(32), 11182–11187.
- OMER, A. D., ZIESCHE, S., EBHARDT, H. & DENNIS, P. P. (2002). In vitro reconstitution and activity of a C/D box methylation guide ribonucleo-protein complex. *Proceedings of the National Academy of Sciences of the USA* **99**(8), 5289–5294.
- PARLEA, L. G., SWEENEY, B. A., HOSSEINI-ASANJAN, M., ZIRBEL, C. L. & LEONTIS, N. B. (2016). The RNA 3D Motif Atlas: computational methods for extraction, organization and evaluation of RNA motifs. *Methods* **103**, 99–119.
- PESELIS, A. & SERGANOV, A. (2012). Structural insights into ligand binding and gene expression control by an adenosylcobalamin riboswitch. *Nature Structural & Molecular Biology* **19**(11), 1182–1184.
- PITTCI, F., BEVERIDGE, D. L. & BARANGER, A. M. (2002). Molecular dynamics simulation studies of induced fit and conformational capture in U1A-RNA binding: do molecular substates code for specificity? *Biopolymers* **65**(6), 424–435.
- PLASCHKA, C., LIN, P. C. & NAGAI, K. (2017). Structure of a pre-catalytic spliceosome. *Nature* **546**(7660), 617–621.
- POGACIC, V., DRAGON, F. & FILIPOWICZ, W. (2000). Human H/ACA small nucleolar RNPs and telomerase share evolutionarily conserved proteins NHP2 and NOP10. *Molecular Cell Biology* **20**(23), 9028–9040.
- REBLOVA, K., SPONER, J. E., SPACKOVA, N., BESSEOVA, I. & SPONER, J. (2011). A-minor tertiary interactions in RNA kink-turns. Molecular dynamics and quantum chemical analysis. *The Journal of Physical Chemistry. B* **115**(47), 13897–13910.
- REITER, N. J., OSTERMAN, A., TORRES-LARIOS, A., SWINGER, K. K., PAN, T. & MONDRAGON, A. (2010). Structure of a bacterial ribonuclease P holoenzyme in complex with tRNA. *Nature* **468**(7325), 784–789.
- ROZHDESTVENSKY, T. S., TANG, T. H., TCHIRKOVA, I. V., BROSIUS, J., BACHELLERE, J.-P. & HÜTTENHOFER, A. (2003). Binding of L7Ae protein to the K-turn of archaeal snoRNAs: a shared RNA binding motif for C/D and H/ACA box snoRNAs in Archaea. *Nucleic Acids Research* **31**(3), 869–877.
- SCHROEDER, K. T., DALDROP, P. & LILLEY, D. M. J. (2011). RNA tertiary interactions in a riboswitch stabilize the structure of a kink turn. *Structure* **19**(9), 1233–1240.



- SCHROEDER, K. T., DALDROP, P., MCPHEE, S. A. & LILLEY, D. M. (2012). Structure and folding of a rare, natural kink turn in RNA with an A•A pair at the 2b•2n position. *RNA* **18**(6), 1257–1266.
- SCHULTZ, A., NOTTROT, S., WATKINS, N. J. & LUHRMANN, R. (2006). Protein-protein and protein-RNA contacts both contribute to the 15-5 K-mediated assembly of the U4/U6 snRNP and the box C/D snoRNPs. *Molecular Cell Biology* **26**(13), 5146–5154.
- SCHUWIRTH, B. S., BOROVINSKAYA, M. A., HAU, C. W., ZHANG, W., VILA-SANJURJO, A., HOLTON, J. M. & CATE, J. H. (2005). Structures of the bacterial ribosome at 3.5 Å resolution. *Science* **310**(5749), 827–834.
- SERGANOV, A., HUANG, L. & PATEL, D. J. (2008). Structural insights into amino acid binding and gene control by a lysine riboswitch. *Nature* **455** (7217), 1263–1267.
- SERGANOV, A., POLONSKAIA, A., PHAN, A. T., BREAKER, R. R. & PATEL, D. J. (2006). Structural basis for gene regulation by a thiamine pyrophosphate-sensing riboswitch. *Nature* **441**(7097), 1167–1171.
- SHI, X., HUANG, L., LILLEY, D. M. J., HARBURY, P. B. & HERSCHLAG, D. (2016). The solution structural ensembles of RNA kink-turn motifs and their protein complexes. *Nature Chemical Biology* **12**(3), 146–152.
- SMITH, K. D., LIPCHOCK, S. V., AMES, T. D., WANG, J., BREAKER, R. R. & STROBEL, S. A. (2009). Structural basis of ligand binding by a c-di-GMP riboswitch. *Nature Structural & Molecular Biology* **16**(12), 1218–1223.
- SMITH, K. D., SHANAHAN, C. A., MOORE, E. L., SIMON, A. C. & STROBEL, S. A. (2011). Structural basis of differential ligand recognition by two classes of bis-(3′–5′)-cyclic dimeric guanosine monophosphate-binding riboswitches. *Proceedings of the National Academy of Sciences USA*, **108**, 7757–7762.
- SURYADI, J., TRAN, E. J., MAXWELL, E. S. & BROWN, B. A. (2005). The crystal structure of the *Methanocaldococcus jannaschii* multifunctional L7Ae RNA-binding protein reveals an induced-fit interaction with the box C/D RNAs. *Biochemistry* **44**(28), 9657–9672.
- SZEWCAK, L. B., DEGRECORIO, S. J., STROBEL, S. A. & STEITZ, J. A. (2002). Exclusive interaction of the 15.5 kD protein with the terminal box C/D motif of a methylation guide snoRNP. *Chemistry & Biology* **9**(10), 1095–1107.
- SZEWCAK, L. B., GABRIELSEN, J. S., DEGRECORIO, S. J., STROBEL, S. A. & STEITZ, J. A. (2005). Molecular basis for RNA kink-turn recognition by the h15.5 K small RNP protein. *RNA* **11**(9), 1407–1419.
- TEPLOVA, M., WOHLBOLD, L., KHIN, N. W., IZAURRALDE, E. & PATEL, D. J. (2011). Structure-function studies of nucleocytoplasmic transport of retroviral genomic RNA by mRNA export factor TAP. *Nature Structural & Molecular Biology* **18**(9), 990–998.
- THORE, S., LEIBUNDGUT, M. & BAN, N. (2006). Structure of the eukaryotic thiamine pyrophosphate riboswitch with its regulatory ligand. *Science* **312**(5777), 1208–1211.
- TRAN, E. J., ZHANG, X. & MAXWELL, E. S. (2003). Efficient RNA 2′-O-methylation requires juxtaposed and symmetrically assembled archaeal box C/D and C′/D′ RNPs. *The EMBO Journal* **22**(15), 3930–3940.
- TSAI, C. J., MA, B., SHAM, Y. Y., KUMAR, S. & NUSSINOV, R. (2001). Structured disorder and conformational selection. *Proteins* **44**(4), 418–427.
- TURNER, B. & LILLEY, D. M. J. (2008). The importance of G.A hydrogen bonding in the metal ion- and protein-induced folding of a kink turn RNA. *The Journal of Molecular Biology* **381**(2), 431–442.
- TYCOWSKI, K. T., SMITH, C. M., SHU, M. D. & STEITZ, J. A. (1996). A small nucleolar RNA requirement for site-specific ribose methylation of rRNA in *Xenopus*. *Proceedings of the National Academy of Sciences USA* **93**(25), 14480–14485.
- VIDOVIC, I., NOTTROT, S., HARTMUTH, K., LUHRMANN, R. & FICNER, R. (2000). Crystal structure of the spliceosomal 15.5 kD protein bound to a U4 snRNA fragment. *Molecular Cell* **6**(6), 1331–1342.
- WANG, J., DALDROP, P., HUANG, L. & LILLEY, D. M. (2014). The k-junction motif in RNA structure. *Nucleic Acids Research* **42**(8), 5322–5331.
- WANG, J., FESSL, T., SCHROEDER, K. T., OUELLET, J., LIU, Y., FREEMAN, A. D. & LILLEY, D. M. J. (2012). Single-molecule observation of the induction of k-turn RNA structure on binding L7Ae protein. *Biophysical Journal* **103**(12), 2541–2548.
- WATKINS, N. J., DICKMANN, A. & LUHRMANN, R. (2002). Conserved stem II of the box C/D motif is essential for nucleolar localization and is required, along with the 15.5 K protein, for the hierarchical assembly of the box C/D snoRNP. *Molecular and Cellular Biology* **22** (23), 8342–8352.
- WATKINS, N. J., NEWMAN, D. R., KUHN, J. F. & MAXWELL, E. S. (1998). In vitro assembly of the mouse U14 snoRNP core complex and identification of a 65-kDa box C/D-binding protein. *RNA* **4**(5), 582–593.
- WATKINS, N. J., SEGALT, V., CHARPENTIER, B., NOTTROT, S., FABRIZIO, P., BACHI, A., WILM, M., ROSBASH, M., BRANLANT, C. & LUHRMANN, R. (2000). A common core RNP structure shared between the small nucleolar box C/D RNPs and the spliceosomal U4 snRNP. *Cell* **103**(3), 457–466.
- WEIKL, T. R. & VON DEUSTER, C. (2008). Selected-fit versus induced-fit protein binding: kinetic differences and mutational analysis. *Proteins* **75**, 104–110.
- WEINBERG, Z., LUNSE, C. E., CORBINO, K. A., AMES, T. D., NELSON, J. W., ROTH, A., PERKINS, K. R., SHERLOCK, M. E. & BREAKER, R. R. (2017). Detection of 224 candidate structured RNAs by comparative analysis of specific subsets of intergenic regions. *Nucleic Acids Research* **45** (18), 10811–10823.
- WEIXLBAUMER, A., JIN, H., NEUBAUER, C., VOORHEES, R. M., PETRY, S., KELLEY, A. C. & RAMAKRISHNAN, V. (2008). Insights into translational termination from the structure of RF2 bound to the ribosome. *Science* **322**(5903), 953–956.
- WHITE, S. A., HOEGER, M., SCHWEPPE, J. J., SHILLINGFORD, A., SHIPILOV, V. & ZARUTSKIE, J. (2004). Internal loop mutations in the ribosomal protein L30 binding site of the yeast L30 RNA transcript. *RNA* **10**(3), 369–377.
- WIMBERLY, B. T., BRODERSEN, D. E., CLEMONS, JR., W. M., MORGAN-WARREN, R. J., CARTER, A. P., VONRHEIN, C., HARTSCH, T. & RAMAKRISHNAN, V. (2000). Structure of the 30S ribosomal subunit. *Nature* **407**(6802), 327–339.
- WINKLER, W. C., GRUNDY, F. J., MURPHY, B. A. & HENKIN, T. M. (2001). The GA motif: an RNA element common to bacterial antitermination systems, rRNA, and eukaryotic RNAs. *RNA* **7**(8), 1165–1172.



- WOZNIAK, A. K., NOTTROT, S., KUHN-HOLSKEN, E., SCHRODER, G. F., GRUBMULLER, H., LUHRMANN, R., SEIDEL, C. A. & OESTERHELT, F. (2005). Detecting protein-induced folding of the U4 snRNA kink-turn by single-molecule multiparameter FRET measurements. *RNA* **11**(10), 1545–1554.
- XUE, S., WANG, R., YANG, F., TERNS, R. M., TERNS, M. P., ZHANG, X., MAXWELL, E. S. & LI, H. (2010). Structural basis for substrate placement by an archaeal box C/D ribonucleoprotein particle. *Molecular Cell* **39**(6), 939–949.
- YE, K., JIA, R., LIN, J., JU, M., PENG, J., XU, A. & ZHANG, L. (2009). Structural organization of box C/D RNA-guided RNA methyltransferase. *Proceedings of the National Academy of Sciences USA* **106**(33), 13808–13813.
- YOUSSEF, O. A., TERNS, R. M. & TERNS, M. P. (2007). Dynamic interactions within sub-complexes of the H/ACA pseudouridylation guide RNP. *Nucleic Acids Research* **35**(18), 6196–6206.
- ZHANG, J. & FERRE-D'AMARE, A. R. (2013). Co-crystal structure of a T-box riboswitch stem I domain in complex with its cognate tRNA. *Nature* **500**(7462), 363–366.
- ZHOU, H. X. (2010). From induced fit to conformational selection: a continuum of binding mechanism controlled by the timescale of conformational transitions. *Biophysical Journal* **98**(6), L1517.