A critical base pair in k-turns that confers folding characteristics and correlates with biological function

Scott A. McPhee1, Lin Huang1 & David M.J. Lilley1

Kink turns (k-turns) are widespread elements in RNA that mediate tertiary contacts by kinking the helical axis. We have found that the ability of k-turns to undergo ion-induced folding is conferred by a single base pair that follows the conserved A\textbullet{}G pairs, that is, the 3b\textbullet{}3n position. A Watson–Crick pair leads to an inability to fold in metal ions alone, while 3n = G or 3b = C (but not both) permits folding. Crystallographic study reveals two hydrated metal ions coordinated to O6 of G3n and G2n of Kt-7. Removal of either atom impairs Mg2+ induced folding in solution. While SAM-I riboswitches have 3b\textbullet{}3n sequences that would predispose them to ion-induced folding, U4 snRNA are strongly biased to an inability to such folding. Thus riboswitch sequences allow folding to occur independently of protein binding, while U4 should remain unfolded until bound by protein. The empirical rules deduced for k-turn folding have strong predictive value.
kin turns (k-turns) are ubiquitous sequences that generate a tight kink within an RNA helix, mediating tertiary interactions in the folding of large assemblies such as the ribosome, and often serving as the target for specific binding proteins. Because of this, k-turns have a key role in the assembly of ribosomes, the spliceosome and box C/D snoRNPs, as well as seven distinct riboswitch species. The standard k-turn comprises a duplex interrupted by a three-nucleotide bulge followed by G-A and A-G base pairs (Fig. 1), and the adenine nucleobases make key cross-strand hydrogen bonds that stabilize the kinked conformation.

K-turn-containing RNA exists in a two-state equilibrium between the folded conformation and a relatively extended structure, and is strongly biased towards the extended structure in the absence of some specific process promoting folding. Several factors can drive the equilibrium towards the kinked structure of the k-turn. These include tertiary contacts and protein factors can drive the equilibrium towards the kinked structure of the sequences of the k-turns of ribosomes, the spliceosome and box C/D snoRNPs, as well as seven distinct riboswitch species. The standard k-turn comprises a duplex interrupted by a three-nucleotide bulge followed by G-A and A-G base pairs (Fig. 1), and the adenine nucleobases make key cross-strand hydrogen bonds that stabilize the kinked conformation.

Results
Ion-induced folding determined by a key sequence element. The extensively-studied Kt-7 of the Haloarcula marismortui ribosome, and the k-turn of the SAM-I riboswitch, both fold into the characteristic kinked structure on addition of metal ions alone. However, in marked contrast the k-turns of the archaeal box C/D (Supplementary Fig. 3) and the human U4 snRNA (Supplementary Fig. 4) do not fold upon addition of metal ions. Both box C/D and U4 k-turns fold on binding of L7Ae protein, and indeed co-crystal structures of both show that these k-turns are folded, so each is intrinsically capable of adopting the k-turn conformation, yet metal ions alone fail to achieve folding. These biologically important k-turns divide into two classes on the basis of their ability to be folded by metal ions, evidently a result of their sequence. As the G-A and A-G pairs are strongly conserved at the 1b-2n/1n positions (the nomenclature is shown in Fig. 1), the important difference must lie elsewhere, and our suspicion turned to the 3b-3n position that follows the conserved G-A pairs.

We took a short RNA duplex with a central Kt-7 sequence and fluorophores at both 5’-termini, enabling us to follow folding into the kinked conformation by the increase in FRET efficiency (EFRET) as the end-to-end distance shortens. The experiment was performed in a background of 90 mM Tris-borate (pH 8.3), with Mg2+ titrated as the only cation present. For the natural Kt-7 sequence EFRET increased from ~0.2 to 0.56 on addition of Mg2+ ions, with a value of [Mg2+]1/2 = 70 µM. The analysis was repeated for the 15 species in which the 3b-3n position was replaced by each combination of the four nucleotides (Fig. 2a, Supplementary Table 1). A range of folding abilities were found from full folding (for example, natural Kt-7; 3b-3n = G-A-G) to those exhibiting a complete inability to fold under these conditions (for example, 3b-3n = G-C). Yet even the 3b-3n = G-C sequence underwent folding upon addition of L7Ae protein, so it is not intrinsically unable to adopt the k-turn structure.

3b-3n sequences correlate with biological function. Thus the 3b-3n sequence acts as a key discriminator, conferring ion-dependent folding properties. Is this reflected in the distribution of k-turn sequences as a function of biological role? We examined the distribution of 3b-3n sequences in two important functional RNA species, comparing several thousand SAM-I and U4 k-turn sequences downloaded from the Rfam database. These two were chosen because of their contrasting environments. The SAM-I k-turn mediate a key tertiary contact6-11 to create a ligand-binding pocket in a riboswitch not known to bind a protein, while the U4 snRNA k-turn binds the 15.5 kDa protein during spliceosome assembly17. The results are plotted as a histogram in Fig. 2b, showing the occurrence of the 3b-3n sequences for the two species ranked horizontally by the folding ability of Kt-7 with the same 3b-3n sequence. It is apparent that the two species cluster at opposite ends of the folding spectrum. The SAM-I k-turn sequences are strongly biased towards an ability to fold in Mg2+ ions, with 60% having 3b-3n = A-G, which is the best folding sequence, and just 0.1% being C-G or G-C. Contrast, 97% of the U4 k-turn sequences are predicted to be unable to fold in Mg2+ ions, with a very strong bias to 3b-3n = G-C or G-U, and less than 0.03% being A-G. Interestingly, modification of the human U4 k-turn sequence by conversion of 3b-3n from G-C to A-G conferred an ability to fold in response to addition of Mg2+ ions (Supplementary Fig. 4).

Empirical sequence rules for ion-induced folding. Examination of the 3b-3n sequences displayed in array form and scored by folding ability (Fig. 2c) reveals some interesting patterns. First, Watson–Crick pairs (Fig. 2c—ascending diagonal) plus GU are all poor folders, with G-C and C-G especially bad. By contrast, guanine directly coordinates metal ions. Analysis of many natural RNA sequences such as those of the ribosome.

Figure 1 | The sequences of some standard k-turns. The top sequence shows a standard k-turn with the standard nomenclature for nucleotide positions indicated and the 3b-3n position highlighted. Below are shown the sequences of the k-turns of H. marismortui ribosomal Kt-7, the SAM-I riboswitch, human U4 snRNA and box C/D snoRNA.
the presence of $3n = G$ (Fig. 2c—third column) or $3b = C$ (Fig. 2c—second row) associates with ability to fold in Mg$^{2+}$ ions. However, since $3b / C \neq 3n / G$ is unfolded in metal ions, the first rule takes precedent over the second. The two best-folding k-turns both have G at the 3n position, and 95% of SAM-I k-turns have G at 3n. Moreover, the k-turns of the glycine$^{21}$, lysine$^{22}$ and cobalamine$^{23}$ riboswitches also have $3n = G$.

A structural explanation of the $3n = G$ rule. Systematic investigation of the ion-induced folding of Kt-7 shows that the most readily folded sequences are those with either $3n = G$ or $3b = C$. There are no high-resolution crystal structures available for $3b = C$ k-turns, and at the present time we cannot rationalize this effect. However, we can provide a molecular explanation for the $3n = G$ behaviour.

We have previously presented a crystal structure of H. marismortui Kt-7 as a protein-free duplex at 2.3 Å resolution$^{24}$. We subsequently obtained crystals diffracting to 2.0 Å (Table 1), whereupon we observed two hydrated metal ions bound in the major groove of the NC helix adjacent to G2n and G3n (Fig. 3a). The electron density for the inner coordination sphere of water molecules is very clear, and both metal ions have octahedral symmetry. Thus they are most probably Mg$^{2+}$ ions, although we cannot exclude the possibility that they are Na$^{+}$ ions on the basis of the metal-O distances. Ion M1 has exchanged two adjacent inner-sphere water molecules with G2n and G3n O6...
atoms, while G3n O6 makes an inner-sphere contact with both ions (Fig. 3b, Supplementary Fig. 5).

Removal of O6 from G2n or G3n impairs ion-induced folding. Having observed the two ions bound to G3n and G2n in the crystal, we then sought to test the importance of these interactions in the folding of Kt-7 in solution. This was examined by atomic mutagenesis whereby the participating O6 atoms were selectively removed by individual substitution of guanine by 2-aminopurine. Folding was analysed using a gel electrophoretic method10. A 65 bp RNA duplex with a central k-turn-containing RNA section was electrophoresed in 15% polyacrylamide in the presence of 90 mM Tris-borate (pH 8.3), 2 mM Mg2+ ions. The folded structure of the unmodified k-turn results in pronounced electrophoretic retardation (Fig. 4). However, removal of either G2n or G3n O6 atoms significantly impaired the ability to fold on addition of Mg2+ ions, that is, resulted in less retarded electrophoretic mobility, whereas the corresponding modification of G1b had a minor effect. This provides a direct connection between the metal ions observed to bind to G2n and G3n O6 atoms by crystallography, and the ability of the k-turn to fold in response to the presence of Mg2+ ions. Thus binding of the divalent metal ions to guanine O6 at the 2n and 3n positions is the key determinant allowing the Kt-7 k-turn to fold unassisted by protein binding.

Discussion
While all k-turns can be folded by protein binding and/or the formation of tertiary contacts, not all will fold spontaneously in the presence of metal ions, and we have found that a major determinant of this behaviour resides in the 3b•3n sequence. From a systematic analysis of 3b•3n sequence variants of Kt-7, we have formulated a
set of rules that have predictive value; application of these can convert the U4 k-turn from non-ion folding into one that is fully folded, in the presence of Mg$^{2+}$ ions for example. One of the rules is that 3n = G, and a high-resolution structure of Kt-7 as a free duplex RNA provides an explanation. Two hydrated, octahedrally-coordinated metal ions are directly bound to the O6 atoms of G2n and 3n. These are probably Mg$^{2+}$ ions, although they could conceivably be Na$^+$ ions, but since both ions can induce folding of Kt-7 then probably either can coordinate at this position. The amidite chemistry 26,27. Fluorescein (Link Technologies) and Cy3 (GE Healthcare) folding properties conferred by the 3b is bound by L7Ae.

RNA synthesis. Oligoribonucleotides were purified by gel electrophoresis in polyacrylamide in the presence of 7 M urea. The full-length RNA product was visualized by ultraviolet shadowing. The band was excised and electroeluted using an Elutrap (Whatman) into 40% aqueous methylamine (Sigma-Aldrich) for 30 min. at 65 °C deprotected using a 1:1 solution of 35% aqueous ammonia (Fisher Scientific) and incubated at 65 °C for 30 min. at 65 °C. All oligoribonucleotides were introduced into the RNA sections as required.

Expression and purification of A. fulgidus L7Ae. The gene encoding full-length Archaeoglobus fulgidus L7Ae was cloned into a modified pET-Duet1 plasmid (Novagen)28 using the HindIII and EcoRI sites. The L7Ae gene was fused upstream of a hexahistidine-encoding sequence with a PreScission-cleavable linker. The hexahistidine-L7Ae fusion protein was expressed in Escherichia coli BL21-Gold (DE3) pLysS cells (Stratagene) induced with 0.2 mM IPTG at 20 °C for 12 h. Harvested cells were resuspended in 20 mM Tris–HCl (pH 8.0), 10 mM NaCl, 10 mM imidazole, 1 M phenylmethylsulfonyl fluoride (buffer A) and lysed by sonication. The protein suspension was heated at 85 °C for 20 min in the presence of 10 mM MgCl$_2$ to denature endogenous protein and this was removed by centrifugation at 18,000 x g for 30 min at 4 °C. L7Ae was loaded onto a HiTrap column (GE Healthcare), washed with 25 mM imidazole in buffer A, and the protein was eluted with 500 mM imidazole in buffer A. The six-His tag was cleaved from L7Ae by PreScission protease in 20 mM HEPES-Na (pH 7.6), 100 mM NaCl, 0.5 mM EDTA (buffer C) at 4 °C for 16 h. L7Ae was applied to a heparin column (GE Healthcare) and eluted at 250 mM NaCl in a gradient from 50 to 2,000 mM NaCl in 20 mM HEPES-Na (pH 7.6). The protein was further purified using a Superdex 200 gel filtration column in a buffer containing 5 mM Tris-HCl (pH 8.0), 100 mM NaCl.

The protein concentration was measured by absorbance at 280 nm using a molar extinction coefficient of 5,240 M$^{-1}$ cm$^{-1}$ for L7Ae. The protein was concentrated to 20 mg ml$^{-1}$ in buffer containing 5 mM Tris-HCl (pH 8.0), 100 mM NaCl and stored at −20 °C as aliquots.

FRET analysis of k-turn folding. FRET efficiency was measured from a series RNA duplexes terminally 5’ labelled with fluorescein and Cy3, containing central k-turn sequences and variants. Absorption spectra were measured in 90 mM Tris-borate (pH 8.3) in 2 μl volumes using a Thermo Scientific NanoDrop 2000c spectrophotometer. Spectra were deconvoluted using a corresponding RNA species labelled only with Cy3, and fluorescence ratios calculated using a two-state model for ion-induced folding, that is, $E_{FRET}$ = constant for metal ion binding and $E_{FRET}$ = 1 for L7Ae. The protein was further purified using a Superdex 200 gel filtration column in a buffer containing 5 mM Tris-HCl (pH 8.0), 100 mM NaCl and stored at −20 °C as aliquots.

Gel electrophoretic analysis of k-turn folding. RNA species were electrophoresed in 13% polyacrylamide (29:1, acrylamide: bis) gels in 90 mM Tris-borate (pH 8.3) plus 2 mM Mg$^{2+}$ ions. Electrophoresis was performed at 120 V at 4 °C for at least 72 h, with recirculation of the buffer at >1 litre h$^{-1}$. Gels were stained using SYBR Gold (Life Technologies), washed in MilliQ water and visualized on a Typhoon FLA 9500 (GE Healthcare).

The sequences used for the electrophoretic experiments were (written 5’ to 3’): Kt-7 upper strand: 5’-CGCAAGAAGGACAGAAGGAGCAGUCAGUGGGAAGGGAACAU GUCGGCCGGAGCGATGGAAGGAGG-3’ Kt-7 lower strand: 5’-CCCTGTCACCCGATGCACCGUAGGGAAGGGAGCAGUG CUGAGGGGCAGCAGGTAAGAGG-3’

The DNA sections of these oligonucleotides are shown underlined. Modified nucleotides were introduced into the RNA sections as required.

Crystal structure determination and refinement. The crystallized construct had the sequence (written 5’ to 3’): 5’-GGCGAGAAGACGGGGAGCCG-3’.
Self-complementary sequence forms the structure shown in Supplementary Fig. 6, containing two Kt-7 motifs. A solution of 1 mM RNA in 5 mM Tris–HCl (pH 8.0) and 100 mM NaCl was heated to 95 °C for 1 min. The solution was then cooled to 20 °C and MgCl₂ was added to a final concentration of 10 mM. The hanging-drop vapour diffusion method was used for crystallization. A volume of 1 µl of RNA was mixed 1:1 with well solution comprising 3.5 M Na formate, 0.1 M Na acetate (pH 4.6) at 20 °C. Crystals (approximate dimensions 150 × 20 × 20 µm³) with space group P6₁22 were grown in a few days. Crystals were briefly washed in well solution supplemented with 30% glycerol. The crystals were flash frozen by mounting in nylon loops and plunging into liquid nitrogen. A 2.0 Å resolution data set was collected on beamline I03 at the Diamond Light Source (Harwell, UK). The resolution cutoff for the data was determined by examining both CC1/2 and difference map of the magnions, as described previously. The structure was determined by molecular replacement. H. marismortui Kt-7 (PDB 4C40) was used as the search model using the program Phaser. The remaining ligands and waters were added to the model on the basis of inspection of electron density difference maps. Structural models were built in Coot and RCrane. The structure was refined with Refmac5 (ref. 35) from the CCP4 suite of programs and Phenix refine. Model geometry and the fit to electron-density maps were monitored with MOLPROBITY.

References
1. Klein, D. J., Schmeing, T. M., Moore, P. B. & Steitz, T. A. The kink-turn: a new RNA secondary structure motif. EMBO J. 20, 4214–4221 (2001).
2. Vidovic, I., Nottrott, S., Hartmuth, K., Luhrmann, R. & Ficner, R. Crystal structure of the spliceosomai 15.5 kD protein bound to a U4 snRNA fragment. Molec. Cell. 6, 1331–1342 (2000).
3. Moore, T., Zhang, Y., Fenley, M. O. & Li, H. Molecular basis of box C/D RNA-protein Interactions: Cocrystal structure of archaeal L7Ae and a box C/D RNA. Structure 12, 807–818 (2004).
4. Hanna, T. & Ferré-D’Amare, A. R. Structure of protein L7Ae bound to a K-turn derived from an archaeal box H/A RNA at 1.8 Å resolution. Structure 12, 893–903 (2004).
5. Li, L. & Ye, K. Crystal structure of an H/A box ribonuclease protein. Nature 443, 302–307 (2006).
6. Montange, R. K. & Batey, R. T. Structure of the S-adenosylmethionine riboswitch regulatory mRNA element. Nature 441, 1172–1175 (2006).
7. Lescoute, A., Leontis, N. B., Massire, C. & Westhof, E. Recurrent structural RNA motifs, isoelectricty matrices and sequence alignments. Nucleic Acids Res. 33, 2395–2409 (2005).
8. Liu, J. & Lilley, D. M. J. The role of specific 2'-hydroxyl groups in the stabilization of the folded conformation of kink-turn RNA. RNA 13, 200–210 (2007).
9. Daldrop, P. & Lilley, D. M. J. The plasticity of a structural motif in RNA: structural polymorphism of a kink turn as a function of its environment. RNA 19, 357–364 (2013).
10. Goetz, T. A., Melcher, S. E., Norman, D. G. & Lilley, D. M. J. The kink-turn motif in RNA is dimorphic, and metal ion dependent. RNA 10, 254–264 (2004).
11. Schroeder, K. T., Daldrop, P. & Lilley, D. M. J. RNA tertiary interactions in a riboswitch stabilize the structure of a kink turn. Structure 19, 1233–1240 (2011).
12. Turner, B., Melcher, S. E., Wilson, T. J., Norman, D. G. & Lilley, D. M. J. Induced fit of RNA on binding the L7Ae protein to the kink-turn motif. RNA 11, 1192–1200 (2005).
13. Turner, B. & Lilley, D. M. J. The importance of G: A hydrogen bonding in the metal ion- and protein-induced folding of a kink turn RNA. J. Mol. Biol. 381, 431–442 (2008).
14. Wang, J. et al. Single-molecule observation of the induction of k-turn RNA structure on binding L7Ae protein. Biophys. J. 103, 2541–2548 (2012).
15. Koonin, E. V., Bork, P. & Sander, C. A novel RNA-binding motif in omnipotent suppressors of translation termination, ribosomal proteins and a ribosome modification enzyme? Nucleic Acids Res. 22, 2166–2167 (1994).
16. Ban, N., Nissen, P., Hansen, J., Moore, P. B. & Steitz, T. A. The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. Science 290, 905–920 (2000).
17. Nottrott, S. et al. Functional interaction of a novel 15.5kD [U4/U6/U5] tri-snRNA protein with the S'-stem loop of U4 snRNA. EMBO J. 18, 6119–6133 (1999).
18. Kuhn, J. F., Tran, E. J. & Maxwell, E. S. Archaeal ribosomal protein L7 is a functional homolog of the eukaryotic 15.5kD Snu13p snoRNP core protein. Nucleic Acids Res. 30, 931–941 (2002).
19. Watkins, N. J., Dickmanns, A. & Lüthi, R. Conserved stem II of the box C/D motif is essential for nucleolar localization and is required, along with the 15.5k protein, for the hierarchical assembly of the box C/D snoRNP. Mol. Cell. Biol. 22, 8342–8352 (2002).
20. Burge, S. W. et al. Rfam 11.0: 10 years of RNA families. Nucleic Acids Res. 41, D226–D232 (2013).