Control of box C/D snoRNP assembly by N⁶-methylation of adenine

Lin Huang, Saira Ashraf, Jia Wang & David MJ Lilley*

Abstract

N⁶-methyladenine is the most widespread mRNA modification. A subset of human box C/D snoRNA species have target GAC sequences that lead to formation of N⁶-methyladenine at a key trans Hoogsteen-sugar A-G base pair, of which half are methylated in vivo. The GAC target is conserved only in those that are methylated. Methylation prevents binding of the 15.5-kDa protein and the induced folding of the RNA. Thus, the assembly of the box C/D snoRNP could in principle be regulated by RNA methylation at its critical first stage. Crystallography reveals that N⁶-methylation of adenine prevents the formation of trans Hoogsteen-sugar A-G base pairs, explaining why the box C/D RNA cannot adopt its kinked conformation. More generally, our data indicate that sheared A-G base pairs (but not Watson–Crick base pairs) are more susceptible to disruption by N⁶-methylation and are therefore possible regulatory sites. The human signal recognition particle RNA and many related Alu retrotransposon RNA species are also methylated at N6 of an adenine that forms a sheared base pair with guanine and mediates a key tertiary interaction.

Keywords: epigenetics; G-A base pairs; k-turn; RNA methylation; signal recognition particle

Subject Categories: RNA Biology; Structural Biology

DOI: 10.15252/embr.201743967 | Received 20 January 2017 | Revised 21 May 2017 | Accepted 22 May 2017

Introduction

All cellular RNA is subject to dynamic covalent modification, and posttranscriptional modification of RNA is diverse and widespread [1,2]. N⁶-methyladenine is the most common modification in RNA [3–5]. It is found in mRNA (a typical eukaryotic mRNA will have several such methylated adenine nucleotides), as well as in lncRNA species such as Xist [4,5] and MALAT1 [6,7]) and in viral RNA [8]. This epigenetic marker has proposed roles in the modulation of RNA stability [9], control of translation efficiency [10,11], and in gene regulation [12], and N⁶-methylation is most frequently observed in regions of mRNA indicative of control functions [5]. The level of modification is subject to regulation by methyltransferase [13] and demethylase [14,15] enzymes, and they have been linked to human disease [15–18].

In this work, we have uncovered a putative role of epigenetic regulation in a subset of human snoRNP complexes that are involved in the site-specific modification of RNA. In archaea and the eukaryotes, the box C/D snoRNP complexes direct the site-specific 2'-O-methylation of rRNA and tRNA by providing complementary guide RNA for specificity and a SAM-dependent methyl transferase enzyme to modify the target ribose [19–26]. The complex is equivalent to a bulged RNA duplex, where both 12 nucleotide strands of the central region are the guide sequences that hybridize to the target RNA molecules to be methylated (Fig 1A). The conserved box C/D and C’/D’ sequences are located in the duplexes that flank the guide region, and both adopt the k-turn conformation (k-turn structure is reviewed in [27]). A series of proteins assembles on the snoRNA [28–32]. In the first step, a member of the L7Ae family, the 15.5-kDa protein (15.5k) in humans, binds to each of the k-turns. This is followed by a Nop58 binding to box C/D and Nop56 binding to box C’/D’, associating through a coiled-coil domain to form a heterodimer. Finally, two molecules of the methyl transferase fibrillarin associate with the complex to generate the catalytically active snoRNP. Thus, the key event that initiates the assembly of the complex is the association between the 15.5-kDa protein and the k-turn. In particular, Watkins et al [30] showed that if the binding of 15.5k to the human box C/D k-turn is prevented by sequence changes that are known to disrupt k-turn folding, this blocks assembly of the box C/D snoRNP.

Binding of L7Ae-family proteins induces the formation of the tightly kinked k-turn structure [33–35] (Fig 1B). The core of the k-turn structures formed by the box C/D and C’/D’ elements comprises consecutive trans sugar-Hoogsteen G-A base pairs (sometimes termed sheared base pairs) that position the conserved adenine nucleobases to make key cross-strand hydrogen bonds that stabilize the conformation (Fig 1C). In general, k-turns fall into two classes depending on whether or not they will fold in response to the presence of metal ions [36]. We have noted that box C/D k-turns remain unfolded under these conditions, and so require protein-induced folding. Once this has been achieved the assembly of the snoRNP can proceed to form the active methylation complex. However, a process that blocks the folding of the k-turn would likely prevent the assembly of a functional snoRNP complex.

While the majority of k-turn structures have a C-G base pair at the -1b,-1n position, we have noted that in the box C/D k-turn...
sequences this base pair is sometimes inverted, so that the -1n nucleotide is a cytosine. Since the highly conserved core of the k-turn requires 2n = G and 1n = A, this creates a GAC sequence on the non-bulged strand (boxed) that is a potential target for N6-methylation of the conserved adenine at the 1n position. The two helical arms of the k-turn are named C (canonical) and NC (non-canonical) as indicated.

C Structure of a standard box C/D k-turn, together with the chemical structure of the G1b A1n and A2b G2n trans sugar-Hoogsteen G/C base pairs (PDB 1RLG [74]).

Figure 1. Assembly of box C/D snoRNA, and their k-turn structures.
A Scheme depicting the assembly of box C/D snoRNP. The first stage is the binding of the 15.5k protein to the box C/D and C'/D' k-turns.
B The general sequence of a k-turn, with the standard nomenclature of nucleotide positions. Note that when the -1n nucleotide is cytosine, this generates a GAC sequence on the non-bulged strand (boxed) that is a potential target for N6-methylation of the conserved adenine at the 1n position. The two helical arms of the k-turn are named C (canonical) and NC (non-canonical) as indicated.
C Structure of a standard box C/D k-turn, together with the chemical structure of the G1b A1n and A2b G2n trans sugar-Hoogsteen G/A base pairs (PDB 1RLG [74]).

sequences this base pair is sometimes inverted, so that the -1n nucleotide is a cytosine. Since the highly conserved core of the k-turn requires 2n = G and 1n = A, this creates a GAC sequence on the non-bulged strand (boxed in Fig 1B) that is a putative target for methylation of adenine N6 [9]. Given the critical role of the A1n nucleobase in the folded k-turn, it seemed highly probable that N6-methylation would interfere with folding. We have therefore studied the occurrence of -1n = C in human box C/D and C'/D' sequences. We find that a significant subset have -1n = C, about half of which are known to be methylated, and that those that are methylated are strongly conserved. We further show here that N6-methyladenine at the 1n position of box C/D and C'/D' k-turns prevents specific binding of the 15.5k protein, and consequent induced folding, and use X-ray crystallography to provide a structural explanation for this. Indeed, our structural analysis indicates that sheared A-G base pairs are much more sensitive to N6-methylation of adenine compared to Watson–Crick base pairs, and we show that these effects extend to other structures involving such A-G base pairs.

Results

N6-adenine methylation of box C/D sequences

Analysis of the targets for methylation by the METTL3–METTL14 methyl transferase complex in eukaryotes shows the preferred
sequence to be 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 [4,5]. The box C/D and C'/D' sequences adopt the k-turn conformation on binding the 15.5-kDa protein (15.5k), with the unbulged strand having GA at the 2n,1n sequence. In most other k-turns, the -1n nucleotide is G, but in some box C/D-type k-turns it is C, thus creating a GAC sequence with the central adenine as a potential methylation target (Fig 2A). N6-methylation of the A1n is C, thus creating a GAC sequence with the central adenine as a k-turns, the -1n nucleotide is G, but in some box C/D-type k-turns it is C, thus creating a GAC sequence with the central adenine as a potential methylation target (Fig 2A). N6-methylation of the A1n might be anticipated to affect the folding of the k-turn into the kinked conformation. We therefore analyzed the occurrence of -1n = C in human box C/D and C'/D' sequences (although many D' boxes could not be reliably annotated either because of the short length of these snoRNA-like genes, or the lack of evolutionary conservation or sequence motif signals).

Using the snoRNABase [37] and snOPY [38] databases, we identified 27 human snoRNA sequences with 2n = G, 1n = A, and -1n = C, comprising 17 box D and 10 box D'. These are potential methylation targets, so we then examined the RMBase database [39], which contains single-nucleotide or high-resolution data on N6-methyladenine sites [40,41], to see whether any had been experimentally demonstrated to be methylated in vivo. From this, we identified 14 human sequences that were shown to be N6-methylated at A1n (eight from box D and six from box D') (Table 1, Fig 3). Most of these have been identified in more than one independent experiment. The A1n of SNORD13 from chromosome 8 was found to be methylated in 11 different sequencing experiments. Paralogs of SNORD13 and SNORD46 encoded on different chromosomal locations were found to be methylated.

If the N6-methylation of A1n in these k-turn-forming snoRNA species is functionally significant then we might expect that the -1n = C sequence that is required to generate the methylation target should be conserved in other eukaryotes. We therefore examined the conservation of the -1n position in other organisms using the snoRNABase and snOPY databases (Appendix Table S1). We found that the -1n sequence is highly conserved as C for those 14 sequences that are methylated in vivo. This is 100% conserved in SNORD71 and SNORD101. 26 SNORD71 sequences are conserved in vertebrates including human, rat, chicken, frog, and fugu (pufferfish). Outside the conserved C, D, C', D' and guide region, only the -1n C and its complementary paired base (-1b = G) are 100% conserved (Fig 2B). In marked contrast, for 12 of 13 box C/D RNAs that are not methylated -1n is significantly less conserved as C. This is shown graphically in Fig 2C. Evidently, the -1n position of the snoRNA k-turn is critical for function in the methylated sequences, but much less so in those that are not.

N6-methylation of adenine 1n interferes with 15.5k-induced folding of box C/D and C'/D' k-turns

We have seen that in a subset of human box C/D k-turn sequences the -1n position is a cytosine, creating a GAC methylation target on the non-bulged strand. The METTL3-METTL14 methyl transferase complex could therefore convert the A1n of the k-turn to N6-methyladenine. We know that k-turn folding is most sensitive to modification of the G1b-A1n base pair, and we therefore set out to test whether or not N6-methyladenine would impair folding of box C/D-type k-turns. In general, these k-turn have sequences that are not susceptible to ion-induced folding [36] so that they require the binding of protein to undergo folding into the kinked conformation. For the human snoRNA species this is the 15.5k protein. We examined the folding of representative human box C/D and C'/D' k-turns in response to addition of human 15.5k using two approaches. One was by electrophoretic migration in non-denaturing polyacrylamide gels, where protein binding results in retardation of the RNA. The other is a spectroscopic method based upon fluorescence resonance energy transfer (FRET) between fluorescein and Cy3 fluorophores attached to the termini of a short snoRNA duplex with a central k-turn motif; protein-induced folding into the kinked conformation brings the fluorophores closer together, resulting in an increased efficiency of energy transfer (E\textsubscript{FRET}).

Results are shown for one box C/D k-turn from SNORD13 (U13) snoRNA, and one box C'/D' k-turn from SNORD62A (U62A) snoRNA (Fig 4). The unmethylated RNA species both migrate as a discrete band of retarded mobility when incubated with either human 15.5k or Archeoglobus fulgidus L7ae (Af L7ae) protein (Fig 4A and C). At higher 15.5k concentrations, further retarded species are observed as a smear running up the gel, indicative of additional non-specific binding. Over the same range of 15.5k protein concentration used in the electrophoresis, the fluorescence spectroscopy reveals an increase in energy transfer (final E\textsubscript{FRET} = 0.6) indicative of a folding of the k-turns (Fig 4B and D). The affinity of Af L7ae for a standard k-turn has been measured in the pM range [42], so it is expected that we observe stoichiometric binding in these experiments, and consequently, we cannot measure a dissociation constant for this class of protein. However, collectively these data indicate that the 15.5k protein binds to and induces the folding of the box C/D and box C'/D' k-turns.

By contrast, the corresponding RNA species with N6-methyladenine at the 1n position exhibited very different behavior. Methylated versions of both U13 and U62A RNA failed to form discrete retarded complexes observed by gel electrophoresis (Fig 4A and C), rather forming smears up the gels at high 15.5k concentrations, indicative of non-specific binding. The fluorescence data (Fig 4B and D) show evidence of structural transitions at higher 15.5k concentration, with a significantly lower endpoint (final E\textsubscript{FRET} = 0.3) compared to the unmethylated RNA. Moreover, the shape of the transition indicates rather greater cooperativity. These results suggest non-specific binding of multiple protein molecules that fails to fold the methylated RNA into a normal k-turn conformation.

We have explored the binding of 15.5k protein to SNORD62A (U62A) snoRNA with and without N6 methylation of A1n using isothermal titration calorimetry (Appendix Fig S1). Binding results in the evolution of heat in an exothermic reaction for both species, and the fitted thermodynamic data are tabulated in Appendix Table S2. The titration of the unmethylated RNA is consistent with stoichiometric binding as anticipated, with close to 1:1 15.5k:RNA ratio in an enthalpically driven binding reaction. However, the character of the binding to the N6A-methylated RNA is different, having lower values of AH and TAS and weaker apparent binding affinity, with a consequently different 15.5k protein–RNA ratio. This clearly reflects a significantly different protein binding mode between unmethylated and N6A-methylated RNA, consistent with the results of the electrophoretic and FRET analysis, and confirming non-specific binding when the snoRNA is methylated.

Evidently, N6-methylation of A1n disrupts the proper binding and folding of box C/D and C'/D' k-turns by the human 15.5k protein.
A structural basis for the effect of mA

In order to understand the structural basis for the effect of N6-methyladenine on k-turn conformation, we have used X-ray crystallography to investigate the structural effect of this modification on different base pairs involving adenine. Each species was formed by the hybridization of a self-complementary strand that contains a 5-bromocytosine nucleotide used to calculate phases for the structure.

**Figure 2. Bioinformatic analysis of box C/D and C’/D’ methylation in the human genome.**

- **A** WebLogo plot showing the occurrence of box C/D sequences in human snoRNA. The frequency of distribution for the -1b,-1n pair is CG 65.27%; AU 8.78%; UA 4.96%; UC 4.20%; CC 3.82%.

- **B** Sequence alignment for SNORD71. The boxes C, D, C’, and D’ are boxed, as is the -1n sequence.

- **C** The conservation of -1n sequences (fraction of cytosine as percentile) for eukaryotic box C/D and C’/D’ sequences for which -1n = C in humans. Those that are methylated in humans are colored red while those that are not are colored blue.
the diffraction in most cases. Crystallographic statistics are presented in Materials and Methods.

The first species studied contained N6-methyladenine paired with uridine (PDB 5LR5). This sequence contains GGACU, which is methylated with the highest frequency in cellular RNA [9]. The RNA duplex adopts standard A-form helical geometry and the N6mA-U methylated with the highest frequency in cellular RNA [9]. The RNA sequence contains GGACU, which is 

Table 1. BoxC/D RNA A1n N6-methylation sites identified from RMBase. In human, 19 unique positions have been modified, comprising 14 unique boxC/D RNA (8 D box and 6 D' box). In mouse, one D' box was identified; importantly, this site is also modified in humans.

<table>
<thead>
<tr>
<th>Human</th>
<th>GeneName</th>
<th>modID</th>
<th>Chromosome</th>
<th>Position</th>
<th>Strand</th>
<th>SupportNum</th>
<th>Position in k-turn</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>SNORD13</td>
<td>m6A_site_124853</td>
<td>chr8</td>
<td>33571093</td>
<td>+</td>
<td>11</td>
<td>boxD 1n</td>
</tr>
<tr>
<td></td>
<td>SNORD13</td>
<td>m6A_site_136193</td>
<td>chrX</td>
<td>2352529</td>
<td>-</td>
<td>8</td>
<td>boxD 1n</td>
</tr>
<tr>
<td></td>
<td>SNORD13</td>
<td>m6A_site_24525</td>
<td>chr11</td>
<td>66988484</td>
<td>+</td>
<td>3</td>
<td>boxD 1n</td>
</tr>
<tr>
<td></td>
<td>SNORD13</td>
<td>m6A_site_130577</td>
<td>chr9</td>
<td>75142252</td>
<td>+</td>
<td>1</td>
<td>boxD 1n</td>
</tr>
<tr>
<td>2</td>
<td>SNORD46</td>
<td>m6A_site_4620</td>
<td>chr1</td>
<td>45242258</td>
<td>+</td>
<td>4</td>
<td>boxD 1n</td>
</tr>
<tr>
<td></td>
<td>SNORD46</td>
<td>m6A_site_122222</td>
<td>chr7</td>
<td>132437879</td>
<td>+</td>
<td>2</td>
<td>boxD 1n</td>
</tr>
<tr>
<td>3</td>
<td>SNORD48</td>
<td>m6A_site_111371</td>
<td>chr6</td>
<td>31803100</td>
<td>+</td>
<td>4</td>
<td>boxD 1n</td>
</tr>
<tr>
<td></td>
<td>SNORD101</td>
<td>m6A_site_115173</td>
<td>chr6</td>
<td>133136512</td>
<td>+</td>
<td>1</td>
<td>boxD 1n</td>
</tr>
<tr>
<td>5</td>
<td>SNORD5</td>
<td>m6A_site_25652</td>
<td>chr11</td>
<td>93466398</td>
<td>-</td>
<td>2</td>
<td>boxD 1n</td>
</tr>
<tr>
<td>6</td>
<td>SNORD67</td>
<td>m6A_site_22376</td>
<td>chr11</td>
<td>46783944</td>
<td>-</td>
<td>3</td>
<td>boxD 1n</td>
</tr>
<tr>
<td>7</td>
<td>SNORD71</td>
<td>m6A_site_50014</td>
<td>chr16</td>
<td>71792313</td>
<td>-</td>
<td>2</td>
<td>boxD 1n</td>
</tr>
<tr>
<td>8</td>
<td>SNORD89</td>
<td>m6A_site_74416</td>
<td>chr2</td>
<td>101889404</td>
<td>+</td>
<td>1</td>
<td>boxD 1n</td>
</tr>
<tr>
<td>9</td>
<td>SNORD2</td>
<td>m6A_site_95922</td>
<td>chr3</td>
<td>186502617</td>
<td>+</td>
<td>1</td>
<td>boxD' 1n</td>
</tr>
<tr>
<td>10</td>
<td>SNORD29</td>
<td>m6A_site_23425</td>
<td>chr11</td>
<td>62621408</td>
<td>+</td>
<td>4</td>
<td>boxD' 1n</td>
</tr>
<tr>
<td>11</td>
<td>SNORD41</td>
<td>m6A_site_64597</td>
<td>chr19</td>
<td>12817310</td>
<td>+</td>
<td>1</td>
<td>boxD' 1n</td>
</tr>
<tr>
<td>12</td>
<td>SNORD44</td>
<td>m6A_site_10668</td>
<td>chr1</td>
<td>173835140</td>
<td>-</td>
<td>9</td>
<td>boxD' 1n</td>
</tr>
<tr>
<td>13</td>
<td>SNORD57</td>
<td>m6A_site_79190</td>
<td>chr20</td>
<td>2637610</td>
<td>+</td>
<td>1</td>
<td>boxD' 1n</td>
</tr>
<tr>
<td>14</td>
<td>SNORD62A</td>
<td>m6A_site_233909</td>
<td>chr9</td>
<td>134361128</td>
<td>+</td>
<td>2</td>
<td>boxD' 1n</td>
</tr>
<tr>
<td></td>
<td>SNORD62B</td>
<td>m6A_site_133912</td>
<td>chr9</td>
<td>134365904</td>
<td>+</td>
<td>5</td>
<td>boxD' 1n</td>
</tr>
<tr>
<td>Mouse</td>
<td>SNORD2</td>
<td>m6A_site_29280</td>
<td>chr16</td>
<td>23108986</td>
<td>+</td>
<td>1</td>
<td>boxD' 1n</td>
</tr>
</tbody>
</table>

© 2017 The Authors

EMBO reports

Published online: June 16, 2017
Lin Huang et al N6-methyladenine blocks snoRNA assembly
N6-methyladenine is the most common covalent modification of mRNA in eukaryotes. In principle, there are two ways of expressing N6-methyladenine modification. One is recognition by a specific binding protein, frequently called a “reader” in the language of epigenetics. For example, He and colleagues have identified many Gm6AC sites in RNA that bind the YTHDF2 protein [9,11]. Alternatively, N6-methylation of adenine could alter the local conformation of RNA. We have shown that the Watson–Crick base pairs A-U and A-G can form normally despite methylation of adenine at N6, although the A-U base pair has been shown to be destabilized to some degree [43]. In marked contrast, we have found that the trans Hoogsteen-sugar A-G pair is completely disrupted by inclusion of N6-methyladenine. This base pair is frequently found in RNA structures that are not regular duplex, and form the core of the k-turn structure where the conserved adenine bases make key cross-strand hydrogen bonds. We have shown here that this provides the basis for a probable regulatory mechanism for box C/D snoRNP assembly. The 15.5k protein fails to bind to the k-turn when the A1n is methylated at the exocyclic N6. All the specific interactions between this class of protein and k-turns are made with the conserved guanine nucleotides at the 1b and 2n positions that lie in the major groove on
the outside of the kinked RNA, whereas the A1n is on the inside of the structure making no contact with the protein [35]. This class of protein therefore makes indirect recognition of the methylation, resulting from the overall RNA structure rather than specific contacts. Recently, it has been shown that RNA structure can be affected in vitro by N6-methylation of A or G [44], but we have not found evidence for such modification occurring naturally in box C/D snoRNA sequences.

The great majority of known k-turn-forming sequences have a C-G base pair in the -1b, -1n position, so creating a GAG sequence on the unbulged strand that is not a target for N6-methylation at A1n. This is true for almost all the known ribosomal and riboswitch

**Figure 4.** 15.5-kDa protein binding and induced folding of k-turn conformation is blocked by N6-methylation of adenine in box C/D and C’/D’ snoRNA.

A–D Two human snoRNA k-turns with -1n = C (thus creating a GAC methylation target on the non-bulged strand) have been chosen as examples. Human 15.5k protein binding was studied by gel electrophoretic retardation analysis (A and C) and induced folding analyzed by FRET between fluorescein and Cy3 terminally attached to a short duplex (B and D). The chosen snoRNA species were the SNORD13 (U13) box C/D (A, B) and SNORD62A (U62A) box C’/D’ (C, D) k-turns. Each species was studied with and without N6-methylation at the A1n position. 200 nM RNA was incubated with the indicated concentration of 15.5k, or A. fulgidus L7Ae proteins and applied to 10% polyacrylamide gels electrophoresed under non-denaturing conditions (A and C). Binding of either protein to the unmodified RNA (tracks 1 through 5) led to the formation of discrete retarded species. At higher concentrations of 15.5k, a continuous smear of complexes ran up the gel suggesting non-specific binding beyond stoichiometric conditions. By contrast, no specific RNA–protein complexes were observed when N6-methyladenine-containing RNA was used (tracks 6 through 10). FRET efficiency (EFRET) was measured as a function of 15.5k concentration for non-methylated RNA (closed circles) and N6-methyladenine-containing RNA (open circles). The data have been fitted to a simple binding model (line).
k-turn sequences. However, we have shown here that in a subset of human box C/D and C'/D' snoRNA the -1b/-1n base pair is reversed so placing a C at the -1n position, and thereby creating a GAC sequence that potentially targets the A1n for N6-methylation. Approximately half of these have been shown (generally in multiple experiments) to be methylated in vivo. Importantly, the -1n = C is strongly conserved for those that are methylated, whereas this is not true for those that are not methylated.

We have shown here that N6-methylation of A1n in human box C/D and C'/D' k-turns known to undergo methylation in vivo...
Figure 5. Crystal structures of duplex RNA containing base pairs involving adenine or N<sup>6</sup>-methyladenine. In each case, the relevant base pair is shown with its 2F<sub>c</sub>-F<sub>c</sub> map contoured at 1.2σ. The sequences of each self-complementary duplex are shown, with the relevant base pair highlighted in green, and N<sup>6</sup>-methyl modification in red. S-Bromocytosine nucleotides are shown blue; each structure was solved by SAD using the anomalous scatter from the two bromine atoms, except for 5LR3 that was determined by soaking with CuCl<sub>2</sub>. A A duplex containing a N<sup>6</sup>-methyladenine-uracil base pair (PDB 5LR5). Crystals of space group P6<sub>5</sub> were obtained that diffracted to 2.27 Å. A standard cis Watson–Crick base pair is formed in this duplex, despite the presence of the N<sup>6</sup>-methyl group. B, C Accommodation of N<sup>6</sup>-methyladenine in a cis Watson–Crick G/A base pair. Two duplex species were constructed containing adjacent G/A base pairs flanked by G-C base pairs, with (B, PDB 5LQO) and without (C, PDB 5LQT) N<sup>6</sup>-methyladenine at 1.87 and 1.50-Å resolution, respectively. In both structures, the G/A base pairs are cis Watson–Crick pairs connected by two hydrogen bonds. Thus, N<sup>6</sup>-methyladenine does not lead to disruption of these Watson–Crick G/A pairs. D, E Disruption of a trans sugar-Hoogsteen G/A base pair by N<sup>6</sup>-methyladenine. Crystal structures of RNA duplexes in which guanine opposes adenine or N<sup>6</sup>-methyladenine with flanking G/U base pairs. In the absence of the N<sup>6</sup>-methyl group (D, PDB 5LR3 at 1.65-Å resolution), a trans Hoogsteen-sugar G/A base pair is formed. In marked contrast, the N<sup>6</sup>-methyladenine does not form a base pair with the guanine (E, PDB 5LR4 at 1.72-Å resolution), and there is no hydrogen bonding between the two nucleobases. Thus, N<sup>6</sup>-methylation of adenine prevents the formation of the trans Hoogsteen-sugar G/A base pair.

Figure 6. Superposition of RNA helices with potential trans sugar-Hoogsteen G/A pairs (PDB 5LR3 and 5LR4), with and without N<sup>6</sup>-methylation of adenine. The unmethylated structure is shown (green) and the methylated structure (pink). Images are shown as parallel-eye stereoscopic pairs. A Stereo image of superimposed structures, showing the central G/A pairs and flanking G-U base pairs. B Superimposed view of one of the G/A pairs. Note that the G:6mA base pair is disrupted, such that the N<sup>6</sup>-methyladenine is significantly translated away from the opposing guanine base. The RMSD for the four G, A nucleotides is 0.962 Å.
prevent the binding of the 15.5k protein and the concomitant folding of the k-turn into the kinked conformation. This is the first stage of the assembly of the box C/D snoRNP, without which formation of an active methylation complex cannot proceed further [30]. Although assembly factors are also involved [32,45–47], the initial binding of 15.5k protein is fundamental, and blocking the required
RNA structure is likely to be fatal to the assembly regardless. Thus, the assembly of these box C/D snoRNPs could plausibly be regulated by N^6-methylation of the guide RNA species. It has been shown that binding 15.5k protein stabilizes box C/D snoRNA [48], and if complex formation fails to occur, the RNA is unstable to degradation [49]. In the absence of any of the core, snoRNP proteins box C/D snoRNA failed to accumulate in the nucleolar body of the nucleolus [50].

Is this restricted solely to box C/D snoRNA? 15.5k additionally binds the U4/U6.U5 tri-snRNP in spliceosome assembly [51], where it also interacts with a k-turn in the RNA [52] as the first stage of assembly [53]. Given the similarities to box C/D snoRNP assembly, could this be subject to the same kind of regulation? In fact the U4 k-turn sequence is standard with -1n = G, and this is strongly conserved from humans down to yeast. So this cannot be subject to regulation by methylation in the same way as the box C/D snoRNP species.

However, we have found data that are indicative of potential regulation by N^6-methylation at G-A base pairs in other systems. In fact one of the box C/D k-turns we identified as methylated (SNORD71, HBII 239; Fig 3) has been shown to be a precursor of the human microRNA mir768 [54] so connecting this with regulatory RNA species. Searching more widely we noted that the human signal recognition particle (SRP) RNA Alu domain contains a sequence that had been considered as a putative k-turn, with the normal consecutive G-A and A-G base pairs (Fig 7). Moreover, like box C/D snoRNP, the SRP is assembled in the nucleolus [55,56]. This has a HS₂HS₁ structure [57], but the known Haloarcula marismortui k-turn Kt-S8 has the same secondary structure and a closely similar sequence. The nucleotide immediately 3' of the GA sequence of SRP on the unbulged strand (C108 on the lower strand as drawn in Fig 7) is strongly conserved as cytosine, so creating a potential GAC methylation target. Examination of a recent crystal structure of SRP [58] (Appendix Fig S4B) shows that while this sequence did not adopt k-turn conformation, consecutive G₂U₂ sugar-Hoogsteen G-A and A-G base pairs were formed that directed the adenine nucleobases into the minor groove of the juxtaposed helix 1 to make an A-minor interaction to stabilize the tertiary structure. A107 N3 accepts an H-bond from the O2' of U24, and its O2' donates an H-bond to O2 of U24 (Fig 7B). This is closely analogous to the A-minor interactions found in the core of a standard k-turn structure, but in this case, they are making a longer-range tertiary contact. We examined the RMBase database [39] to see whether the central adenine (A107) is methylated in vivo. This revealed that this adenine has indeed been demonstrated to be N^6-methylated in 14 independent experiments (Appendix Fig S4C). A107 would correspond to A1n were this forming a k-turn and is the nucleotide that makes the key cross-helix H-bonds that fixes the tertiary structure (Fig 7B). From our crystallography, we know that N^6-methylation is incompatible with the required trans sugar-Hoogsteen G-A base pair and that the adenine must translocate substantially. It is therefore most likely that methylation would destabilize the folded structure of the SRP.

Generalizing a little further, the left-hand side of SRP (as depicted in Fig 7A) is closely related to the Alu retrotransposon, which is an extremely widespread human mobile genetic element. We examined one Alu element BC200 using RMBase and discovered that the equivalent adenine is indeed N^6-methylated in vivo (Appendix Fig S4). We further examined the first 30 human Alu elements (from a total of 1,862 entries) in the RMBase database and found that 11 were methylated at the corresponding adenine (i.e. the position corresponding to A1n in the k-turn) (Appendix Fig S4C). Evidently, this adenine is subject to N^6-methylation at a high frequency in Alu elements in human cells.

In summary, crystallographic investigation shows that the major structural effect of N^6-methylation of adenine is likely to be manifest through the trans sugar-Hoogsteen G-A sheared base pair rather than Watson–Crick base pairs. These are widely used in RNA structures, including rather prominently in k-turns. We find that a subset of box C/D snoRNAs are methylated at the 1n position and that this blocks both 15.5k binding and the formation of the kinked conformation in vivo. But these effects are not restricted to k-turns, and can apply to any structure that uses trans sugar-Hoogsteen G-A base pairs, and we have found equivalent N^6-methylation of adenine in human SRP and in related Alu elements. It is likely that control of RNA conformation by N^6-methylation of adenine in G-A base pairs is quite general.

Materials and Methods

Bioinformatic analysis

Human box C/D k-turn WebLogo

A full alignment of 262 of human box C/D sequences taken from snoRNABase [37] (https://www-snorna.biotoul.fr/) was made using Jalview (Jalview 2.9) [59]. The k-turn region was aligned manually based on the known pattern of conserved nucleotides. A WebLogo plot [60] showing the occurrence of box C/D sequences in human snoRNA was made using the website http://weblogo.threeplusone.com.

Identification of box C/D RNAs with a 2n = G, 1n = A, and -1n = C sequence

Human box C/D sequences extracted from snoRNABase [37] and snOPy [38] databases (http://snoopy.med.miyazaki-u.ac.jp) were checked with the Microsoft Word search function and visual alignment to identify box C/D RNAs with 2n = G, 1n = A, and -1n = C sequence.

-InC conservation analysis

Box C/D sequences extracted from snoRNABase [37] and snOPy [38] databases were further aligned and manually inspected by Genedoc [61] and Jalview [59]. Jalview was used to calculate the number and percentage of -1n = C. The quality of the alignment was also checked by comparison with a recent alignment of box C/D sequences deposited on Rfam [62,63].

Identification of sites of N^6-adenine methylation using RMBase

Candidate box C/D snoRNAs were used to search RMBase [39] http://mirlab.sysu.edu.cn/rmbase/index.php in order to discover their A1n methylation status.

RNA synthesis

RNA oligonucleotides were synthesized using t-BDMS phosphoramidite chemistry [64] as described in Wilson et al [65],
implemented on an Applied Biosystems 394DNA/RNA synthesizer. RNA was synthesized using ribonucleotide phosphoramidites with 2′-O-t-butyldimethyl-silyl (t-BDMS) protection [66,67] (Link Technologies). t-BDMS-protected N6-methyl-A-CE phosphoramidite was obtained from Glen Research. Fluorescein (Link Technologies) and Cy3 (GE Healthcare) were attached to the 5′ termini of the oligonucleotides as phosphoramidites in the final cycle of synthesis as required. Unmodified RNA was deprotected in a 25% ethanol/ammonia solution at room temperature for 3 h and evaporated to dryness. Oligonucleotides containing N6-methyladenine were further deprotected for 2 h at 65°C. Oligonucleotides containing 5-bromocytidine (ChemGenes) were deprotected for 36 h at 20°C. Cytidine (ChemGenes) were deprotected for 3 h and evaporated to dryness. Oligonucleotides containing N6-methyladenine were further deprotected for 2 h at 65°C. Oligonucleotides containing 5-bromocytidine (ChemGenes) were deprotected for 36 h at 20°C. All oligoribonucleotides were redissolved in 115 µl of anhydrous DMSO, 60 µl triethylamine (Aldrich), and 75 µl triethylamine trihydrofluoride (Aldrich) to remove t-BDMS groups, and agitated at 65°C in the dark for 2.5 h. After cooling on ice for 10 min, 250 µl RNA quenching buffer (Glen Research) was added to stop the reaction, and the oligonucleotides were desalted using NAP-10 columns (GE Healthcare).

RNA for crystallization was purified by gel electrophoresis in polyacrylamide under denaturing conditions in the presence of 7 M urea. The full-length RNA product was visualized by UV shadowing. The band was excised and electroeluted using an Elutrap Electroelution System (GE Healthcare) into 45 mM Tris-borate (pH 8.5), 5 mM EDTA buffer for 8 h at 200 V at 4°C. The RNA was precipitated with ethanol, washed once with 70% ethanol, and suspended in water.

Fluorophore-labeled oligoribonucleotides were purified by gel electrophoresis under denaturing conditions (as described below) and subjected to further purification by reversed-phase HPLC on a C18 column (ACE 10-300, Advanced Chromatography Technologies), using an acetonitrile gradient with an aqueous phase containing 0.1% triethylamine (Aldrich), and 75% (v/v) acetonitrile (Aldrich) to remove t-BDMS protection [66,67] (Link Technologies). RNA solutions were prepared by diluting concentrated stocks into the binding buffer containing 40 mM HEPES (pH 7.5), 150 mM NaCl, by slow cooling from 90°C to 4°C. They were purified by electrophoresis in 12% polyacrylamide under non-denaturing conditions and recovered by electroelution, followed by ethanol precipitation.

Preparation of proteins

L7Ae protein

Archeoglobus fulgidus L7Ae protein was expressed and purified as previously published [35].

15.5-kDa protein

Human 15.5k protein was expressed as an N-terminal GST fusion in Escherichia coli BL21 (DE3) pLysS induced with 0.2 mM isopropyl β-D-1-thiogalactopyranoside at 22°C. Harvested cells were resuspended in PBS buffer and lysed by sonication. Insoluble protein was removed by ultracentrifugation at 45,000 g for 60 min at 4°C. The solution was then applied to Glutathione Sepharose 4B resin (GE Healthcare) equilibrated in buffer PBS. The GST tag was cleaved by incubation with 40 U/mg thrombin in 20 mM Tris–HCl (pH 8.0), 50 mM NaCl for 8 h at 25°C. In order to remove contaminating RNA the protein was applied to a heparin column (GE Healthcare) in 20 mM Tris–HCl (pH 8.0) with a 50-2,000 mM NaCl gradient; 15.5k protein eluted at 500 mM NaCl. The purified protein was concentrated to 10 mg/ml.

FRET analysis of k-turn folding

Fluorescence resonance energy transfer efficiency was measured from duplex RNA terminally 5′-labeled with fluorescein and Cy3. RNA duplexes contained a central box C/D sequence or N6-A1n-substituted variant (from position -2 to 4). The sequences used were (written 5′–3′):

| 5′-TCCCUCAGCCCGUGGGAUCAAGG | 3′-CUCCUGAGAAGG | U13top: F –CCAGUCAGUGACGUGAUCAGGUAGG
| 5′-CUCCUGAGAAGG | 3′-CUCCUGAGAAGG | U62Atop: F –CCAGUCAGUGAGGAUAAGCAGUAGGC
| 5′-CUCCUGAGAAGG | 3′-CUCCUGAGAAGG | U48top: F –CCAGUCAGUGAGGAUAAGCAGUAGGC
lower strand: Cy3-3CCUGACAGUCUGACCCACGUAGUG

Note that the lower strand is the same in each case.

Absorption spectra were measured in 90 mM Tris-borate (pH 8.3) in a 5-mm path-length cuvette in 2 µl volumes using a NanoDrop 2000c spectrophotometer (Thermo Scientific). Spectra were deconvoluted using a corresponding RNA species labeled only with Cy3, and fluorophore absorption ratios calculated using a MATLAB program. Fluorescence spectra were recorded in 90 mM Tris-borate (pH 8.3) at 4°C using an SLM-Aminco 8100 fluorimeter. Spectra were corrected for lamp fluctuations and instrumental variations, and polarization artifacts were avoided by setting excitation and emission polarizers crossed at 54.7°. Values of FRET efficiency (EFRET) were measured using the acceptor normalization method [68] implemented in MATLAB.

EFRET as a function of 15.5k protein concentration was fitted to:

\[
EFRET = E_0 + \Delta EFRET \cdot \frac{(1 + K_A P_T + K_A R_T) - \sqrt{(1 + K_A P_T + K_A R_T)^2 - 4 K_A P_T^2}}{2 K_A P_T}
\]

where \(E_0\) is the initial FRET efficiency in the absence of added protein, \(\Delta EFRET\) is the full range of the change in FRET efficiency, \(K_A\) is the apparent association constant, and \(P_T\) and \(R_T\) are the total concentration of 15.5k and RNA, respectively.

Electrophoretic analysis of protein binding to snoRNA k-turns

Binding of recombinant 15.5k and Afl7Ae protein to the U13 and U62A k-turn motif RNA was analyzed by gel electrophoresis. 200 nM Cy3-labeled RNA (the same species as used in the FRET experiments, see above) was incubated in 45 mM Tris-borate (pH 8.3) with various concentrations of proteins (15.5k or Afl7Ae) in a final volume of 10 µl at 7°C for 60 min. An equal volume of loading buffer containing 10% glycerol and 45 mM Tris-borate (pH 8.3) was added to the reactions, which were then electrophoresed in 10% native polyacrylamide gels in 45 mM Tris-borate (pH 8.3) at 20 V/cm at 7°C for 1 h. Fluorescent RNA was visualized using a Typhoon 9500 fluorimeter (GE Healthcare).

Isothermal titration calorimetry

Titrations were performed at 298 K using an ITC-200 microcalorimeter (GE). RNA solutions were prepared by diluting concentrated stocks into the binding buffer containing 40 mM HEPES (pH 7.5),
100 mM KCl, 10 mM MgCl₂ to a final concentration of 20 μM. 15.5k was prepared in the same binding buffer with a concentration of 200 μM. Solutions were degassed for 2–5 min before loading. The sample cell was filled with 200 μl of RNA. 15.5k protein was injected into the RNA solution, using 0.4 μl for the first injection and 2 μl for the subsequent 19 injections using a computer-controlled 40-μl microsyringe with an injection interval of 120 s. Titration of protein into the binding buffer or titration of the binding buffer into the RNA solution produced negligible heat evolution. Integrated heat data were analyzed using a one-set-of-sites model in MicroCal Origin following the manufacturer’s instructions. The first data point was excluded in the analysis. The binding parameters ΔH (cal mol⁻¹), K (M⁻¹) and n (moles of 15.5k bound per RNA) were fitted. The binding free energy ΔG and reaction entropy ΔS were calculated using the relationships ΔG = −RTln K (R = 1.9872 cal mol⁻¹ K⁻¹, T = 298 K) and ΔG = ΔH − TΔS. The dissociation constant K_d was calculated as 1/K.

Crystallization, structure determination, and refinement

The sequences are shown in Fig 4. All RNA used in crystallization in this study was formed by hybridization of self-complementary strands. A solution of 1 mM RNA in 5 mM Tris–HCl (pH 8.0), 100 mM NaCl was heated to 95°C for 1 min. The solution was slowly cooled to 20°C, and MgCl₂ was added to a final concentration of 10 mM. The complete conditions used for crystallization of each species are given in Appendix Table S3.

All data were collected on beamline I02 of Diamond Light Source (Harwell, UK), and processed by XIA2 version 0.4.0.0 [69]. The coordinates of the crystal structures described in this work are deposited with the PDB with accession numbers 5LR5, 5LQO, 5LQT, 5LR3, and 5LR4.

Expanded View for this article is available online.

Data availability and accession numbers

Coordinates of the crystal structures described in this work are deposited with the PDB with accession numbers 5LR5, 5LQO, 5LQT, 5LR3, and 5LR4.

Acknowledgements

We thank our colleague Dr. Tim Wilson for discussion and Dr. Alessio Ciubili for use of the calorimeter. This study was supported by Cancer Research UK (CRUK) program grant A18604.

Author contributions

LH performed bioinformatic analysis and crystallography, SA carried out RNA synthesis, JW performed bioinformatic analysis, and LH and DMJL planned experiments, analyzed data and wrote the paper.

Conflict of interest

The authors declare that they have no conflict of interest.

References


License: This is an open access article under the terms of the Creative Commons Attribution 4.0 License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.