Opinion TRENDS in Biochemical Sciences Vol.29 No.7 July 2004

Evolutionary origins of the RNA-guided nucleotide-modification complexes: from the primitive translation apparatus?

Elizabeth Tran[1], James Brown[2] and E. Stuart Maxwell[1]

[1]Department of Molecular and Structural Biochemistry, North Carolina State University, Raleigh, NC 27695, USA
[2]Department of Microbiology, North Carolina State University, Raleigh, NC 27695, USA

Eukarya and Archaea possess scores of RNA-guided nucleotide-modification complexes that target specific ribonucleotides for 2′-*O*-methylation or pseudouridylation. Recent characterization of these RNA-modification machines has yielded striking results with implications for their evolutionary origins: the two main classes of nucleotide-modification complex in Archaea share a common ribonucleoprotein (RNP) core element that has evolved from a progenitor RNP. The fact that this common RNP element is also found in ribosomes suggests that the origin of the progenitor RNP lies in the primitive translation apparatus. Thus, the *trans*-acting, RNA-guided nucleotide-modification complexes of the modern RNP world seem to have evolved from *cis*-acting RNA or RNP elements contained in the primitive translation apparatus during the transition from the ancient RNA world to the modern RNP world.

Although the first small nucleolar RNA (snoRNA) was reported >35 years ago [1,2], it has been only in the past decade that the diversity of this small RNA population and its importance in ribosome biogenesis have been fully appreciated (reviewed in Refs. [3–6]). The eukaryotic nucleolus possesses numerous snoRNAs, and predictions of their full complement number several hundred. The snoRNAs use complementary sequences to base-pair with ribosomal RNA (rRNA), and the site of snoRNA function is determined by the formation of a snoRNA–rRNA duplex. A select few snoRNAs facilitate the folding of precursor rRNA (pre-rRNA), whereas several others have been shown to have essential roles in the endonucleolytic cleavage of pre-rRNA. The primary role of most snoRNAs, however, is to guide the multitude of rRNA nucleotide modifications that occur during rRNA maturation and assembly of the ribosome subunits.

From sequence analysis, investigators have established two principal families of snoRNAs on the basis of conserved nucleotide 'boxes' (Box 1). The box C/D snoRNAs guide the 2′-*O*-methylation of targeted nucleotides, whereas the box H/ACA snoRNAs guide the conversion of targeted uridine nucleotides to pseudouridine. With the number of 2′-*O*-methylated and pseudouridylated nucleotides in mature vertebrate rRNAs each numbering about 100, it is easy to see how the snoRNAs have become one of the largest classes of small stable RNA in the eukaryotic cell.

Nucleotide-modification guide RNAs are not limited to Eukarya

Archaea also use 'snoRNA-like' RNAs (termed sRNAs) to guide 2′-*O*-methylation and pseudouridylation of targeted rRNA nucleotides [5,7,8]. The presence of box C/D and H/ACA RNAs in Archaea, a second domain of life that is distinct from Eukarya, strongly suggests that the use of guide RNAs for nucleotide modification is an evolutionarily ancient mechanism.

Although modified rRNA nucleotides in Archaea are typically fewer in number and differ in position from those in Eukarya, the common structural elements of the guide RNAs are well conserved between the two domains (Box 1). Archaeal and eukaryotic box C/D RNAs possess terminal box C/D core and internal C′/D′ motifs, both of which can guide 2′-*O*-methylation by using adjoining rRNA-complementary sequences [7,9,10]. The archaeal and eukaryotic box H/ACA RNAs possess pseudouridylation pockets, where an unpaired uridine in the H/ACA–rRNA duplex is converted to pseudouridine [11–13].

Guide RNAs function as RNP complexes

Eukaryotic snoRNAs and archaeal sRNAs function as ribonucleoprotein (RNP) particles, in which the enzymatic activities for nucleotide modification are located in the protein components of the RNP complex (Box 1). The eukaryotic box C/D snoRNP includes four core proteins: 15.5kD (also known as Snu13p), Nop56p, Nop58p and the methylase Fibrillarin [14,15]. Recent *in vitro* binding studies, as well as *in vivo* crosslinking analyses, have indicated that there is an 'asymmetric' distribution of core proteins between the terminal box C/D core and internal C′/D′ RNA motifs [16,17]. The proteins 15.5kD and Nop58p are bound to the box C/D core motif, whereas Nop56p binds the box C′/D′ motif. Only the methylase Fibrillarin

Corresponding author: E. Stuart Maxwell (stu_maxwell@ncsu.edu).
Available online 5 June 2004

 0968-0004/$ - see front matter doi:10.1016/j.tibs.2004.05.001

Box 1. Eukaryotic and archaeal RNA-guided nucleotide-modification complexes

RNA-guided nucleotide-modification complexes comprise small RNAs bound to core proteins to form ribonucleoprotein (RNP) complexes. The box C/D RNAs direct methylation at the 2′-OH position on the sugar of the targeted nucleotide, whereas the box H/ACA RNAs convert a targeted uridine to pseudouridine. The specificity for the nucleotide-modification reaction lies in the RNA component of these RNP enzymes, where complementary sequences in the guide RNA base-pair with target RNAs to determine the nucleotide for modification. The nucleotide-modification reactions are carried out by the associated core proteins, which possess the methylase or pseudouridine synthase activities, bound to the respective guide RNAs.

The box C/D RNAs follow the 'N + 5 rule' for methylation, where the fifth nucleotide of the target RNA upstream of box D and within the duplex of guide RNA and target RNA is methylated. Many box C/D RNAs contain two RNA motifs that guide 2′-*O*-methylation. Nucleotide boxes C and D, located near the 5′ and 3′ termini of the RNA, form the terminal box C/D core motif, whereas similar sequences located internally form the C′/D′ motif. Despite the sequence similarity of the C/D and C′/D′ boxes, it is not yet clear whether the folded structure of each RNA motif is identical. Analysis of core protein binding to the box C/D core and C′/D′ motifs has indicated symmetric versus asymmetric distributions of proteins on the archaeal and eukaryotic complexes, respectively (Figure I).

The box H/ACA RNAs use discontinuous guide sequences to hydrogen bond with the target RNA. The unpaired uridine in the duplex of guide RNA and target RNA is converted to pseudouridine. Eukaryotic box H/ACA RNAs characteristically show a bipartite RNA composed of two stem-loop structures, each possessing a bulged loop that contains the guide sequence and forms the pseudouridylation pocket. Analysis of archaeal box H/ACA RNAs has so far shown either one or three stem-loop structures with the signature H box sometimes absent (Figure I).

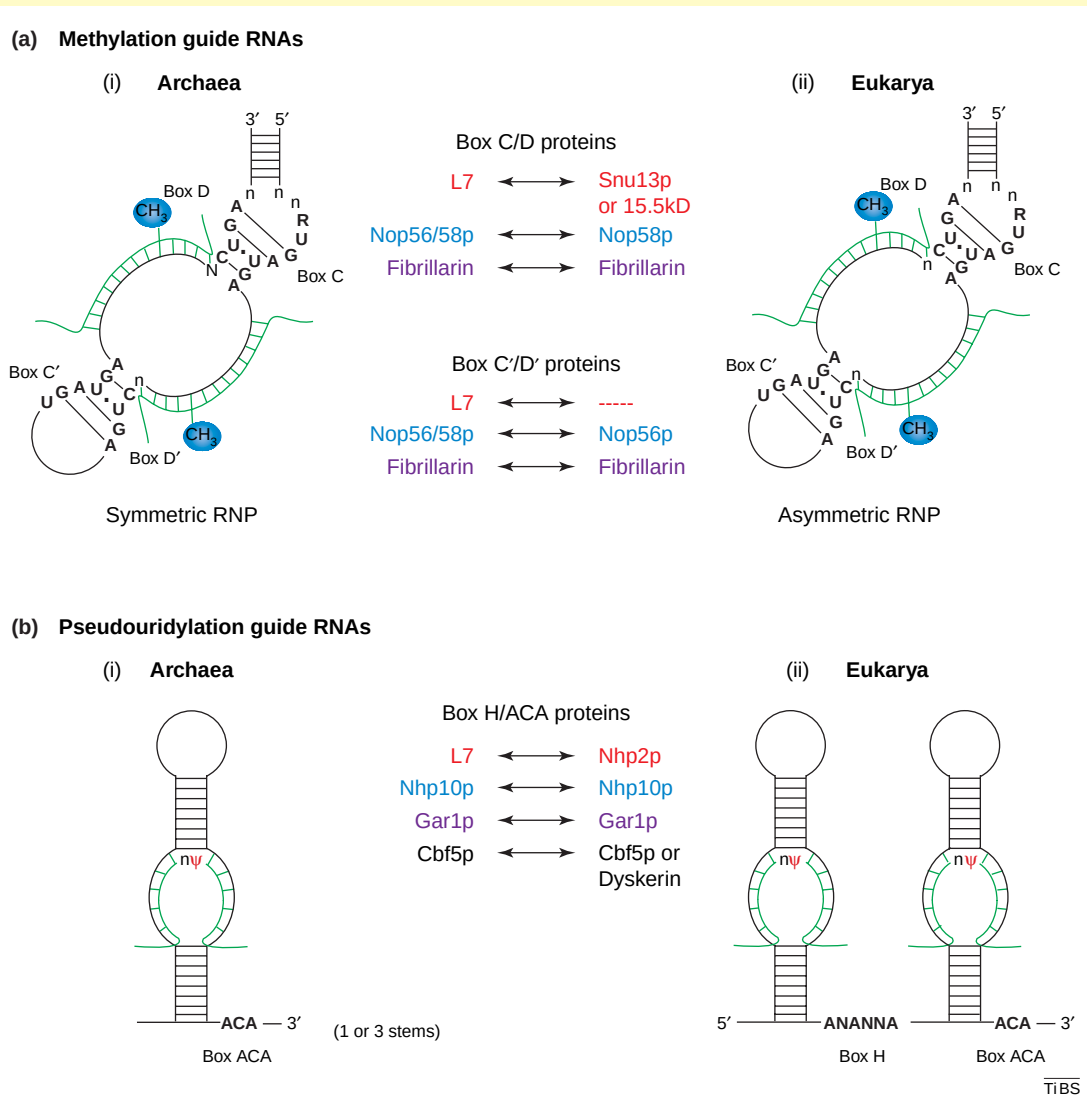

Figure I. Folded RNA structures and core proteins of the box C/D and H/ACA RNPs. **(a)** Box C/D methylation guide RNAs and core proteins. The folded secondary structures of archaeal (i) and eukaryotic (ii) box C/D RNAs are shown. Terminal boxes C and D and internal C′ and D′ sequences are expanded. Target RNAs (green) are base-paired to the box C/D guide sequences located upstream of boxes D and D′. The target nucleotide in the target RNA is methylated (blue). Core proteins associated with the terminal box C/D core motif and internal C′/D′ motif are listed, with homologous archaeal and eukaryotic proteins indicated in matching colors. **(b)** Box H/ACA pseudouridylation guide RNAs and core proteins. The folded secondary structures of archaeal (i) and eukaryotic (ii) box H/ACA RNAs are shown. Box H, located in the hinge region, and the 3′ terminal ACA triplet sequences are expanded. Target RNAs (green) are base-paired in a discontinuous RNA–RNA duplex to the box H/ACA guide sequences. The unpaired nucleotide (n, any nucleotide) and the modified pseudouridine nucleotide (red) of the target RNA are indicated. Core proteins are listed, with homologous archaeal and eukaryotic proteins indicated in matching colors.

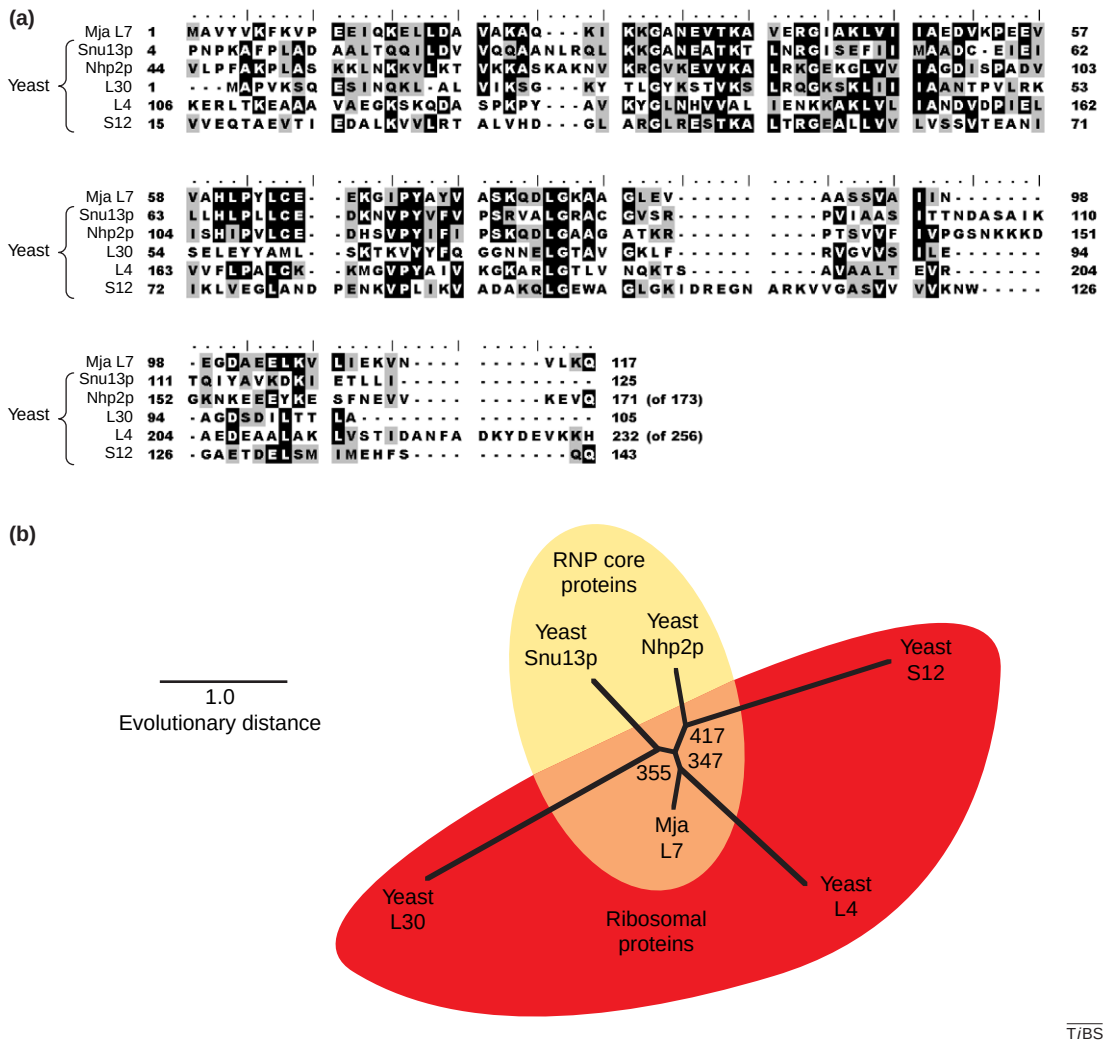

Figure 1. The L7 protein family. **(a)** Alignment of L7 protein family members. The amino acid sequence of archaeal *Methanococcus jannaschii* ribosomal protein L7 (B64450) is compared with the homologous yeast small nucleolar RNP (snoRNP) proteins Snu13p (NP 01088) and Nhp2p (NP 010073) and the yeast ribosomal proteins L30 (NP 011485), L4 (NP 011830) and S12 (NP 015014). GenBank accession numbers are given in parentheses. The L7 protein homologs from a single eukaryotic organism (*Saccharomyces cerevisiae)* are compared to emphasize sequence similarities. Identical amino acids are shown on a black background, similar amino acids are shaded in gray. This alignment shows the central regions of these proteins (indicated by amino acid numbering), which share high sequence similarity. **(b)** An evolutionary tree of L7 protein family members. In this tree, the snoRNP proteins are enclosed in yellow and ribosomal proteins are enclosed in red. Ribosomal protein L7, which is common to both groups, is enclosed in orange. The scale bar indicates an evolutionary distance of 1.0.

associates with both RNA motifs. Analysis of the archaeal box C/D sRNP complex has revealed a simpler protein composition and a 'symmetric' core protein distribution. Three core proteins that are highly similar to the eukaryotic core proteins – ribosomal protein L7 (the archaeal homolog of eukaryotic 15.5kD), a single Nop56/58p homolog and Fibrillarin – bind both the box C/D core and the internal C′/D′ motifs [18–20].

Eukaryotic box H/ACA snoRNPs also contain four core proteins: Nhp2p, Gar1p, Nop10p and the pseudouridyl synthase Cbf5p (also known as dyskerin) [21,22]. At present, little is known about the structure of this RNP complex; the binding sites of the core proteins on the conserved RNA sequences and structures that are characteristic of this snoRNA family have not been defined. Genome analysis has identified candidate archaeal homologs of Cbf5p, Gar1p and Nop10p [21,23]. Sequence analysis has also identified high similarity between eukaryotic Nhp2p and archaeal ribosomal protein L7 (Figure 1). Most recently, it has been shown that L7 is indeed the functional archaeal homolog of the eukaryotic H/ACA core protein Nhp2p [13]. Both of these guide RNP core proteins belong to a larger family of L7 homologous proteins that include several ribosomal proteins (Figure 1).

The 'kink-turn' of box C/D RNAs is an evolutionarily conserved motif

Investigators have recently begun to dissect the structure of the guide RNAs and bound core proteins, with most progress coming from analysis of the box C/D RNAs and RNPs. Initial insight into box C/D RNP structure came indirectly from structural analysis of the 15.5kD core protein bound to the U4 small nuclear RNA (snRNA) [24]. The 15.5kD protein binds both box C/D snoRNAs and spliceosomal U4 snRNA, and thus has a dual function in eukaryotes.

The crystal structure of the 15.5kD–U4 complex revealed a RNA motif consisting of two stems flanking

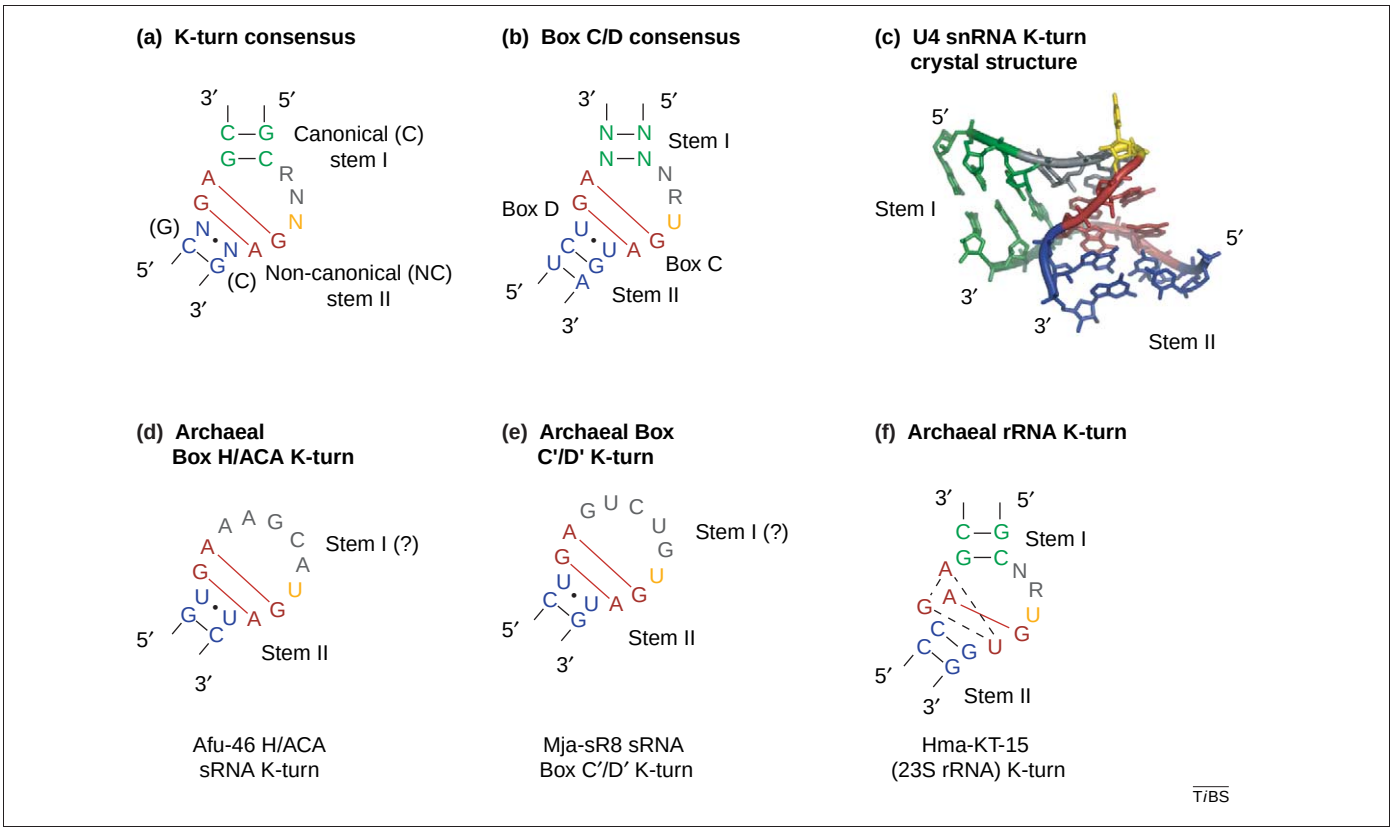

Figure 2. The kink- or K-turn: an evolutionarily conserved RNA motif. **(a,b)** Consensus secondary structure of the K-turn [25] and consensus box C/D terminal core motif folded as a K-turn [15], respectively. This motif consists of stem I or the canonical stem (green) and stem II or the non-canonical stem (blue) flanking the asymmetric (5 + 2) bulge region. Hydrogen bonding between the tandem-sheared G·A base pairs (red) of the asymmetric bulge are characteristic of K-turns and essential for RNA folding. Stems I and II, the tandem-sheared G·A base pairs, and the protruding nucleotide (yellow) characteristic of the asymmetric bulge are shown in the same colors in all panels. **(c)** Three-dimensional structure of the splicing U4 small nuclear RNA (snRNA) K-turn determined from the crystal structure of the U4 snRNA–15.5kD RNP complex [24]. **(d–f)** Three addition K-turns from archaeal box H/ACA sRNAs [13], archaeal box C′/D′ RNA motifs [20], and the L7-binding K-turn of archaeal 23S rRNA [25], respectively. Note the variation in these K-turn motifs and the use of a base triple (broken lines) in the KT-15 turn of 23S rRNA. Each of these three K-turns (d–f) bind archaeal ribosomal protein L7.

an asymmetric bulge region (2 + 5 nucleotides; Figure 2). Crucial for folding of this RNA structure are two tandem, sheared G·A base pairs that form hydrogen bonds across the asymmetric loop. These non-Watson–Crick pairs are stacked on the adjacent helix, and the resulting structure shows a sharp bend or kink in the RNA that has been designated the 'kink-' or 'K-turn' [25] (Figure 2). With knowledge of the 15.5kD–U4 snRNP structure, it was reasoned that the 15.5kD core protein interacts with the terminal box C/D of snoRNAs in an analogous manner [15]. Analysis of core protein binding to both eukaryotic and archaeal box C/D RNAs has confirmed the importance of the K-turn motif for the assembly of box C/D RNPs [26,27].

K-turns are widespread in nature and found in all three domains of life. In an elegant structural analysis of the archaeal *Haloarcula marismortui* ribosome, the K-turn was first defined by six examples found in the 50S subunit [25]. These K-turns are structural variants of the consensus structure (Figure 2), but they all depend on tandem, sheared G·A pairs for RNA folding. As a result of these variations, the K-turns of the archaeal 50S subunit interact with nine different ribosomal proteins, one of them being the sRNP core protein L7.

In eukaryotes, K-turn motifs have been found not only in the box C/D snoRNAs, U4 snRNA and rRNA, but also in the precursor messenger RNA (pre-mRNA) encoding ribosomal protein L30 [25,28]. In bacteria, K-turn motifs have been observed or predicted in 23S rRNA and in the leader sequences of mRNAs regulated by T-box and S-box antitermination mechanisms [29].

Archaeal box C/D and H/ACA sRNAs share a common RNP core

Although dissection of the box H/ACA RNP structure has lagged behind that of the box C/D complex, light has been recently shed on the organization of this RNA–protein complex, with rather striking implications for the evolution of nucleotide-modification complexes. The first archaeal box H/ACA sRNAs have been identified, and ribosomal protein L7 has been shown to be a core protein of this sRNA family [12,13]. The finding of a common core protein shared by both the box C/D and H/ACA RNPs supports their evolutionary origins in a common ancestral complex or progenitor RNP, as previously suggested [11].

The demonstration that L7 is a core protein of both archaeal sRNP families raised issues about the structure of the L7-binding site on the box H/ACA sRNAs. Logic dictated that L7 should also recognize a K-turn in the box H/ACA sRNAs, and mapping of the L7-binding site in several archaeal box H/ACA sRNAs has indeed now shown

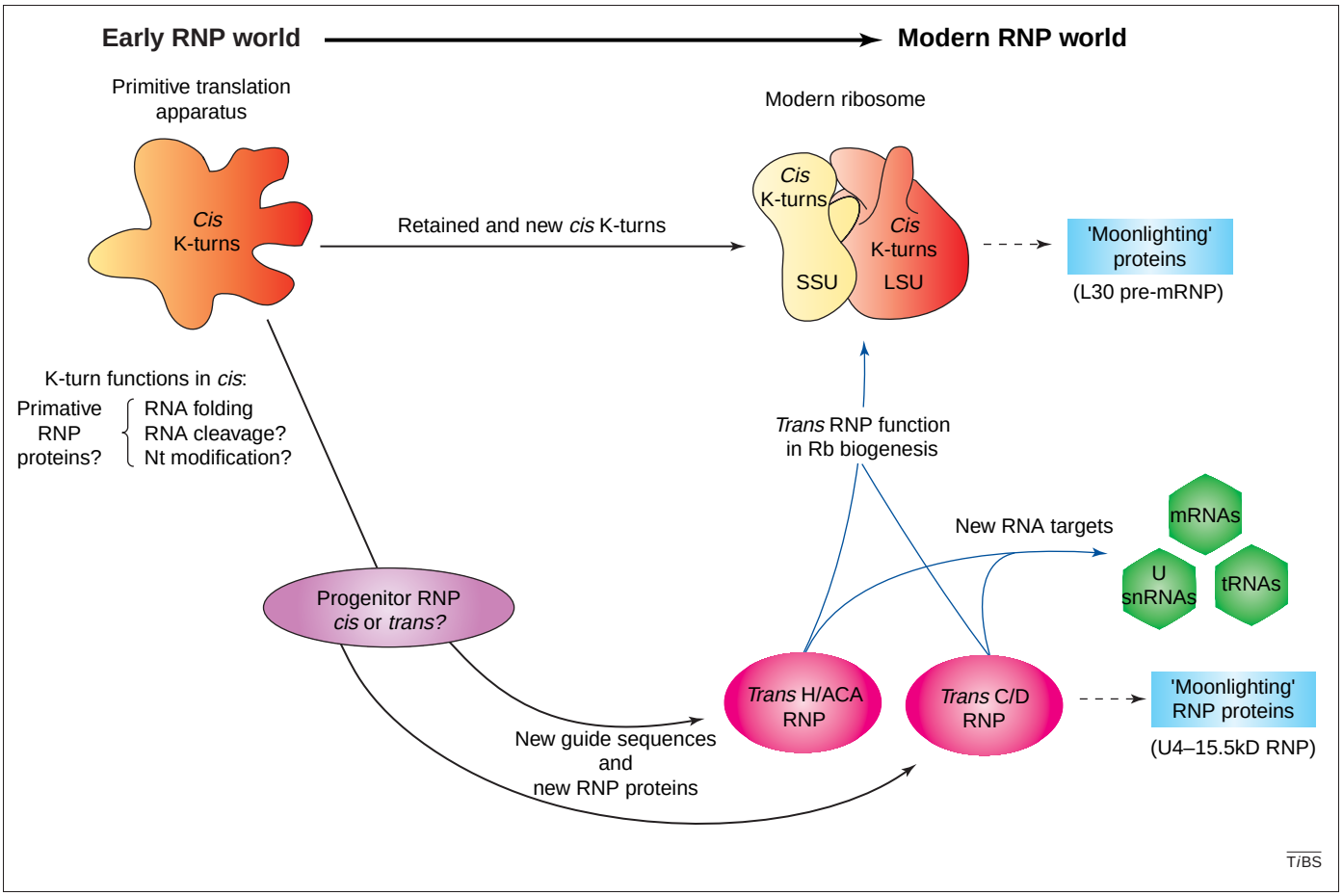

Figure 3. Evolution of RNA-guided nucleotide-modification complexes. Schematically shown is the proposed evolutionary pathway of the RNA-guided nucleotide-modification ribonucleoprotein (RNP) complexes, with origins in the primitive translation apparatus of the early RNP world. Bold, black arrows show evolution of the *cis*-positioned K-turns in a primitive translation apparatus proceeding to independent, *trans*-acting box C/D and H/ACA RNP complexes. This evolution is shown progressing through a common progenitor RNP that could have been positioned in *cis* within the primitive translation apparatus or could have already evolved as an independent, *trans*-acting complex. Some *cis*-positioned K-turns in the primitive RNP translation apparatus have been retained in present-day ribosomes of the modern RNP world. Blue arrows show functional activities of the *trans*-acting box C/D and H/ACA RNP on both modern-day ribosomes and other RNA targets such as tRNA, mRNA and the splicing small nuclear RNAs (snRNAs). Broken arrows indicate K-turn-binding proteins of the ribosome and the nucleotide-modification RNPs acting as 'moonlighting' proteins, in other words, binding the K-turn motifs of other RNA molecules and thereby serving alternative functions. Abbreviations: Nt, nucleotide; Rb, ribosome; SSU, small subunit; LSU, large subunit.

that L7 binds a K-turn [13]. Several of the archaeal box H/ACA K-turns are similar in folded structure to the internal C′/D′ RNA motif of the archaeal box C/D sRNAs (Figure 2).

Identification of a K-turn in archaeal box H/ACA sRNAs implies that eukaryotic Nhp2p should also bind a K-turn in the eukaryotic box H/ACA snoRNAs. Although a K-turn motif has not been identified in these snoRNAs, it might be premature to rule out this possibility because the K-turn motif shows great variation in folded structure. For example, in the archaeal 50S ribosome subunit, one variant uses a base triple for RNA folding, whereas another requires three separate RNA strands [25]. Definitive resolution of this issue will probably require biophysical analysis of the folded box H/ACA snoRNAs, both unbound and bound to core proteins.

A common ancestral RNP complex derived from the primitive translation apparatus?

The fact that L7 has a dual function in Archaea as both a core protein of the sRNP and a component of the ribosome large subunit has led several research groups to suggest that the common ancestral RNP of RNA-guided nucleotide-modification complexes had its evolutionary origins in the primitive translation apparatus (Figure 3) [3,13,27]. In this evolutionary pathway, one would predict that *cis* K-turns in the primitive translation apparatus first functioned as RNA folding elements. Recent analysis has shown that bound protein greatly stabilizes this RNA motif, suggesting that the participation of proteins and formation of a RNP would have occurred early in evolution [30]. These RNA elements or RNPs might have had additional roles at this early stage.

At some point, a RNA or RNP complex would have evolved that ultimately served as the progenitor complex of the box C/D and H/ACA RNPs. It is reasonable to suggest that this progenitor complex was a RNP, because the box C/D and H/ACA RNPs now share identical (archaeal L7) and highly homologous (eukaryotic 15.5kD and Nhp2p) core proteins (Figure 1). This progenitor RNP might have first functioned as a *cis* element in the primitive translation apparatus and evolved within this machinery into distinct box C/D and H/ACA RNPs.

Alternatively, the progenitor RNP could have first become an independent *trans* complex before evolving into the two structurally and functionally distinct families of guide RNPs.

…this fundamental RNA motif was probably present in the translation apparatus of organisms before the last common ancestor.

The apparent absence of RNA-guided nucleotide-modification complexes in Bacteria indicates that the progenitor RNP arose in the archaeal and eukaryotic branch after divergence from Bacteria. The finding of K-turns in bacterial RNAs including rRNA, however, now makes it plausible to suggest that this fundamental RNA motif was probably present in the translation apparatus of organisms before the last common ancestor [31]. The use of specific protein enzymes for nucleotide modification in Bacteria has resulted in a limited number of modified rRNA nucleotides [32]. By contrast, archaeal and eukaryotic rRNA possess greater numbers of 2′-*O*-methylated sugars and pseudouridine nucleotides.

It has been suggested that the far more numerous box C/D and H/ACA RNPs of Archaea and Eukarya are likely to have arisen through variation of the guide sequence [32]. Those mutant box C/D and H/ACA RNAs that proved beneficial to ribosome synthesis and/or function would have been retained by the organism. Such a model could explain the lack of conservation of modified nucleotides shared between Archaea and Eukarya and even among organisms in the same domain of life. The observation that thermophilic archaeal organisms seem to have higher numbers of modified nucleotides in response to their need to stabilize rRNA in extreme environments is consistent with this model [8].

Although the box C/D and H/ACA RNPs share common or homologous core proteins, each complex has developed a distinctive protein composition that is unique to each RNP. The C/D- and H/ACA-specific RNP proteins are highly similar between Archaea and Eukarya, indicating that they were present in each complex before the divergence of Archaea and Eukarya. As previously suggested, some of the C/D- and H/ACA-specific core proteins are likely to have evolved from ancestral coding sequences before the last common ancestor and so are now found in Bacteria, Archaea and Eukarya [32]. For example, the pseudouridine synthase Cbf5p of the box H/ACA RNPs is related to the bacterial transfer RNA (tRNA) pseudouridine synthase TruB. Gene duplication in eukaryotes explains two modification enzymes: one for tRNA modification (Pus4p) and the other for rRNA modification (Cbf5p). One would anticipate that enzymatically competent RNPs were established before variation in their respective guide sequences produced the large array of modification complexes now observed.

Continuing evolution of RNA-guided nucleotide-modification complexes

In the past several years, it has become clear that RNA-guided RNP complexes have continued to evolve in both structure and function. The organization of eukaryotic box C/D snoRNPs is more complex than that of archaeal box C/D sRNPs. Although the core proteins of the archaeal sRNP and eukaryotic snoRNP are homologous, the number of core proteins (three versus four) and their distribution on the C/D core and C′/D′ motifs (symmetric versus asymmetric) are so far characteristic of sRNPs and snoRNPs, respectively [16,17,20]. The differing number of core proteins and asymmetric structure of the eukaryotic snoRNP is the result of gene duplication (archaeal Nop56/58p to eukaryotic Nop56p and Nop58p), specialization of the individual core proteins, and divergence of the symmetrical halves of the ancestral RNP. Evolution of RNP structure is also seen in the additional, species-specific proteins of the U3 and U8 snoRNPs in eukaryotes [33,34].

Less is known about evolution of box H/ACA RNP structure. The number of H/ACA core proteins for the eukaryotic snoRNP and archaeal sRNP is the same (four), but the RNAs themselves show apparently distinctive folded structures (bipartite versus single-hairpin structures). Notably, the single RNA hairpin H/ACA snoRNAs of trypanosomes are strikingly similar to the archaeal box H/ACA sRNAs [35]. It thus seems that the bipartite H/ACA snoRNA structure was established after the divergence of this group of eukaryotes. It will be interesting to see how the box H/ACA RNP organization of trypanosomes and Archaea compares with those of plants, animals and fungi.

Eukaryotic snoRNP evolution is also seen in additional RNP functions in ribosome biogenesis. Several snoRNA species are essential for pre-rRNA processing and ribosome biogenesis [36–38]. Others have apparent ‘chaperone’ activity in which intermolecular base-pairing between snoRNA and rRNA has been suggested to promote specific rRNA folding or to prevent the formation of non-productive secondary or tertiary structures [36,39]. It seems likely that similar or alternative functions will be also guided by archaeal sRNAs.

SnoRNA and sRNAs guide modifications of target molecules other than rRNA. These RNA targets include archaeal tRNAs, the eukaryotic splicing U snRNAs, and apparently eukaryotic mRNAs [10,12,40–43]. The snoRNAs that modify splicing snRNAs have been shown to possess both box C/D and H/ACA motifs in a single snoRNA species, thereby guiding both 2′-*O*-methylation and pseudouridylation [41]. Signal sequences localize these species to Cajal bodies, where they carry out their modification reactions, and they have been therefore termed ‘small Cajal body RNAs’ or scaRNAs [44]. Even more striking and thought provoking are the snoRNA species that are expressed in a tissue-specific manner in the brain [42]. One such species shows complementarity to serotonin receptor mRNA and has been suggested to regulate mRNA maturation.

Finally, evolution of guide RNP function is not limited to the guide RNPs themselves, because some core proteins have taken on new roles as ‘moonlighting’ proteins [45]. For example, the 15.5kD core protein of the eukaryotic box C/D RNP functions as a structural protein of the U4

splicing RNP [15]. It is not clear whether the origin of the U4 K-turn is the same as those of the ribosome and the snoRNPs. Notably, the 61K protein of the U4 snRNP shares significant sequence similarity with the box C/D core proteins Nop56p and Nop58p [46]. Another example of a moonlighting protein is the yeast ribosomal protein L30 that binds to K-turns both in the 25S rRNA and its own pre-mRNA, thereby regulating pre-mRNA splicing to the mature mRNA [47].

Concluding remarks

The proposal that *cis*-acting elements important in RNA-modification reactions have evolved to become *trans*-acting RNPs while still maintaining their original functions is not unprecedented. A similar evolutionary scheme has been proposed for the splicing RNP complexes, in which present-day splicing snRNPs are derived from the folded RNA elements of self-splicing group II introns [48]. It has been also suggested that the mechanism of *trans*-splicing might be an intermediate step in, or an offshoot of, the evolution of *cis*- to *trans*-acting RNPs. A potentially analogous evolutionary intermediate has been found in Archaea, where *cis*-positioned box C/D elements in the pre-tRNA^{Trp} intron guide the 2′-*O*-methylation of two nucleotides in the tRNA molecule [40].

Particularly intriguing is the diversity of gene organization for the coding sequences of these *trans*-acting small RNAs. Some snoRNAs are transcribed directly from monocistronic or polycistronic genes; however, most snoRNA genes, particularly those in metazoan organisms, are encoded in the introns of protein-coding genes, and these intronic snoRNAs are processed from their pre-mRNA introns [3,5]. The recent identification of a putative box H/ACA pseudogene possessing a poly(A) tail suggests that intronic snoRNAs might be a consequence of retroposon events [49]. What insertion mechanism has resulted in the intronic snoRNAs becoming predominantly positioned in pre-mRNA introns of proteins that function in the ribosome or protein biosynthetic pathway? Does this biased positioning have a role in coordinating mRNA synthesis with ribosome biogenesis? What is the relationship between archaeal rRNA synthesis and intron splicing [50]?

Clearly, many additional questions concerning the evolution of the RNA-guided nucleotide-modification complexes are unanswered. What is the total complement of guide RNAs in a given organism? Why do the numbers of guide RNAs vary widely between Archaea and Eukarya, among different organisms within a domain, and among the different tissues of a given organism? Are these varying population sizes indicative of additional target RNAs and/or alternative functions? What roles do guide RNAs have in regulating gene expression? What other moonlighting roles have the RNP core proteins assumed? Unexpected answers are sure to arise from future dissection of the structure and function of these RNA-guided nucleotide-modification machines.

Acknowledgements

We thank M. J. Fournier and R. Hartmann for their helpful suggestions in the critical reading of this manuscript. This work was supported by NSF Grant MCB-0215545 to ESM.

References

1 Nakamura, T. *et al.* (1968) Studies on nucleolar 4 to 6S ribonucleic acid of Novikoff hepatoma cells. *J. Biol. Chem.* 243, 1368–1375
2 Weinberg, R. and Penman, S. (1968) Small molecular weight monodisperse nuclear RNA. *J. Mol. Biol.* 38, 289–304
3 Kiss, T. (2001) Small nucleolar RNA-guided post-transcriptional modification of cellular RNAs. *EMBO J.* 20, 3617–3622
4 Terns, M. and Terns, R. (2002) Small nucleolar RNAs: versatile *trans*-acting molecules of ancient evolutionary origin. *Gene Expr.* 10, 17–39
5 Bachellerie, J.P. *et al.* (2002) The expanding snoRNA world. *Biochimie* 84, 775–790
6 Fatica, A. and Tollervey, D. (2003) Insights into the structure and function of a guide RNP. *Nat. Struct. Biol.* 10, 237–239
7 Omer, A. *et al.* (2000) Homologs of small nucleolar RNAs in Archaea. *Science* 288, 517–522
8 Dennis, P. *et al.* (2001) A guided tour: small RNA function in Archaea. *Mol. Microbiol.* 40, 509–519
9 Kiss-Laszlo, Z. *et al.* (1998) Sequence and structural elements of methylation guide snoRNAs essential for site-specific ribose methylation of pre-rRNA. *EMBO J.* 17, 797–807
10 Gaspin, C. *et al.* (2000) Archaeal homologs of eukaryotic methylation guide small nucleolar RNAs: lessons from the *Pyrococcus* genomes. *J. Mol. Biol.* 297, 895–906
11 Ganot, P. *et al.* (1997) The family of box H/ACA small nucleolar RNAs is defined by an evolutionarily conserved secondary structure and ubiquitous sequence elements essential for RNA accumulation. *Genes Dev.* 11, 941–956
12 Tang, T. *et al.* (2002) Identification of 86 candidates for small non-messenger RNAs from the archaeon *Archaeoglobus fuldidus*. *Proc. Natl. Acad. Sci. U. S. A.* 99, 7536–7541
13 Roshdestvensky, T. *et al.* (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 Res.* 31, 869–877
14 Lafontaine, D. and Tollervey, D. (2000) Synthesis and assembly of the box C + D small nucleolar RNPs. *Mol. Cell. Biol.* 20, 2650–2659
15 Watkins, N. *et al.* (2000) A common core RNP structure shared between the small nucleolar box C/D RNPs and the spliceosomal U4 snRNP. *Cell* 103, 457–466
16 Cahill, N. *et al.* (2002) Site-specific cross-linking analyses reveal an asymmetric distribution for a box C/D snoRNP. *EMBO J.* 21, 3816–3828
17 Szewczak, L.W. *et al.* (2002) Exclusive interaction of the 15.5kD protein with the terminal box C/D motif of a methylation guide snoRNP. *Chem. Biol.* 9, 1095–1107
18 Omer, A. *et al.* (2002) *In vitro* reconstitution and activity of a C/D box methylation guide ribonucleoprotein complex. *Proc. Natl. Acad. Sci. U. S. A.* 99, 5289–5294
19 Aittaleb, M. *et al.* (2003) Structure and function of archael box C/D sRNP core proteins. *Nat. Struct. Biol.* 10, 256–263
20 Tran, E. *et al.* (2003) Efficient RNA 2′-*O*-methylation requires juxtaposed and symmetrically assembled archaeal box C/D and C′/D′ RNPs. *EMBO J.* 22, 3930–3940
21 Watkins, N. *et al.* (1998) Cbf5p, a potential pseudouridine synthase, and Nhp2p, a putative RNA-binding protein, are present together with Gar1p in all box H/ACA-motif snoRNPs and constitute a common bipartite structure. *RNA* 4, 1549–1568
22 Henras, A. *et al.* (1998) Nhp2p and Nop10p are essential for the function of H/ACA snoRNPs. *EMBO J.* 17, 7078–7090
23 Watanabe, Y. and Gray, M. (2000) Evolutionary appearance of genes encoding proteins associated with box H/ACA snoRNAs: Cbf5p in *Euglena gracilis*, an early diverging eukaryote, and candidate Gar1p and Nop10p homologs in archaebacteria. *Nucleic Acids Res.* 28, 2342–2352
24 Vidovic, I. *et al.* (2000) Crystal structure of the spliceosomal 15.5kD protein bound to a U4 snRNA fragment. *Mol. Cell* 6, 1331–1342
25 Klein, D. *et al.* (2001) The kink-turn: a new RNA secondary structure motif. *EMBO J.* 20, 4214–4221
26 Watkins, N. *et al.* (2002) Conserved stem II of the box C/D motif is essential for nucleolar localisation 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
27 Kuhn, J. *et al.* (2002) Archaeal ribosomal protein L7 is a functional

homolog of the eukaryotic 15.5kD/Snu13p snoRNP core protein. *Nucleic Acids Res.* 30, 931–941
28 Mao, H. *et al.* (1999) A novel loop–loop recognition motif in the yeast ribosomal protein L30 autoregulatory RNA complex. *Nat. Struct. Biol.* 6, 1139–1147
29 Winkler, W. *et al.* (2001) The GA motif: an RNA element common to bacterial antitermination systems, rRNA, and eukaryotic RNAs. *RNA* 7, 1165–1172
30 Goody, T. *et al.* (2004) The kink-turn motif in RNA is dimorphic, and metal ion dependent. *RNA* 10, 254–264
31 Penny, D. and Poole, A. (1999) The nature of the last universal common ancestor. *Curr. Opin. Genet. Dev.* 9, 672–677
32 Lafontaine, D. and Tollervey, D. (1998) Birth of the snoRNPs: the evolution of the modification-guide snoRNAs. *Trends Biochem. Sci.* 23, 383–388
33 Dragon, F. *et al.* (2002) A large nucleolar U3 ribonucleoprotein required for 18S ribosomal RNA biogenesis. *Nature* 417, 967–970
34 Tomasevic, N. and Peculis, B. (1999) Identification of a U8 snoRNA-specific binding protein. *J. Biol. Chem.* 274, 35914–35920
35 Liang, X. *et al.* (2001) Identification of the first Trypanosome H/ACA RNA that guides pseudouridine formation on rRNA. *J. Biol. Chem.* 276, 40313–40318
36 Peculis, B. (1997) The sequence of the 5′ end of the U8 small nucleolar RNA is critical for 5.8S and 28S rRNA maturation. *Mol. Cell. Biol.* 17, 3702–3713
37 Tycowski, K. *et al.* (1994) Requirement for intron-encoded U22 small nucleolar RNA for 18S ribosomal RNA maturation. *Science* 266, 1558–1561
38 Liang, W. and Fournier, M.J. (1995) U14 base-pairs with 18S rRNA: a novel snoRNA interaction required for rRNA processing. *Genes Dev.* 9, 2433–2443
39 Borovjagin, A. and Gerbi, S. (2001) *Xenopus* U3 snoRNA GAC-box A′ and box A sequences play distinct functional roles in rRNA processing. *Mol. Cell. Biol.* 21, 6210–6220
40 d'Orval, B.C. *et al.* (2001) Box C/D RNA guides for the ribose methylation of archaeal tRNAs. The $tRNA^{Trp}$ intron guides the formation of two ribose-methylated nucleosides in the mature tRNATrp. *Nucleic Acids Res.* 29, 4518–4529
41 Darzacq, X. *et al.* (2002) Cajal body-specific small nuclear RNAs: a novel class of 2′-*O*-methylation and pseudouridylation guide RNAs. *EMBO J.* 21, 2746–2756
42 Cavaille, J. *et al.* (2000) Identification of brain-specific and imprinted small nucleolar RNA genes exhibiting an unusual genomic organization. *Proc. Natl. Acad. Sci. U. S. A.* 97, 14311–14316
43 Liang, X. *et al.* (2002) The spliced leader-associated RNA is a trypanosome-specific sn(o)RNA that has the potential to guide pseudouridine formation on the SL RNA. *RNA* 8, 237–246
44 Darzacq, R. *et al.* (2003) A common sequence motif determines the Cajal body-specific localization of box H/ACA scaRNAs. *EMBO J.* 22, 4283–4293
45 Jeffery, C. (2003) Moonlighting proteins: old proteins learning new tricks. *Trends Genet.* 19, 415–417
46 Makarova, O. *et al.* (2002) Protein 61K, encoded by a gene (PRPF31) linked to autosomal dominant retinitis pimentosa, is required for U4/U6–U5 tri-snRNP formation and pre-mRNA splicing. *EMBO J.* 21, 1148–1157
47 Vilardell, J. and Warner, J. (1997) Ribosomal protein L32 of *Saccharomyces cerevisiae* influences the splicing of its own transcript and the processing of rRNA. *Mol. Cell. Biol.* 17, 1959–1965
48 Sharp, P. (1991) Perspective: five easy pieces. *Science* 254, 663
49 Vitali, P. *et al.* (2003) Identification of 13 novel human modification guide RNAs. *Nucleic Acids Res.* 31, 6543–6551
50 Tang, T. *et al.* (2002) RNomics in Archaea reveals a further link between splicing of archaeal introns and rRNA processing. *Nucleic Acids Res.* 30, 921–930