Archaeal Guide RNAs Function in rRNA Modification in the Eukaryotic Nucleus

Wayne A. Speckmann,1 Zhu-Hong Li,1 Todd M. Lowe,2 Sean R. Eddy,3 Rebecca M. Terns,1,4 and Michael P. Terns1,4
1Department of Biochemistry and Molecular Biology
University of Georgia
Athens, Georgia 30602
2Department of Genetics
Stanford University School of Medicine
Stanford, California 94305
3Howard Hughes Medical Institute
Department of Genetics
Washington University School of Medicine
St. Louis, Missouri 63110
4Correspondence: mterns@bmb.uga.edu [M.P.T.], rterns@bmb.uga.edu [R.M.T.]

Summary

In eukaryotes, many Box C/D small nucleolar RNAs base pair with ribosomal RNA through short complementary guide sequences, thereby marking up to 100 individual nucleotides of ribosomal RNA for 2'-O-methylation [1–3]. Function of the eukaryotic Box C/D RNAs depends upon interaction with at least six known proteins [4–10]. Box C/D RNAs are not known to exist in Bacteria but were recently identified in Archaea by biochemical analysis and computational genomic screens [11, 12] and have likely evolved independently in Archaea and Eukarya for more than 2000 million years [13–15]. We have microinjected Box C/D RNAs from Pyrococcus furiosus, a hyperthermophilic archaean, into the nuclei of oocytes from the aquatic frog Xenopus laevis. Our results show that Box C/D RNAs derived from this prokaryote are retained in the nucleus, localize to nucleoli, and interact with the X. laevis Box C/D RNA binding proteins fibrillarin, Nop56, and Nop58. Furthermore, we have demonstrated the ability of archaeal Box C/D RNAs to direct site-specific 2'-O-methylation of ribosomal RNA. Our studies have revealed the remarkable ability of archaeal Box C/D RNAs to assemble into functional RNA-protein complexes in the eukaryotic nucleus.

Results and Discussion

The numerous Box C/D RNAs that guide modification of rRNA in eukaryotic cells are characterized by two very short conserved sequence elements, Box C (UGAUGA) and Box D (CUGA), found near the 5' and 3' termini of the RNAs, respectively. Two additional elements, Box C' (UGAUGA) and Box D' (CUGA), are also present within Box C/D guide RNAs but are usually less highly conserved [16]. Box C/D RNAs base pair with rRNA via complementary guide sequences (10 to 21 nucleotides in length) found immediately upstream of Box D (or Box D'). Interaction of a Box C/D RNA with rRNA results in 2'-O-methylation of the rRNA at the base-paired position five nucleotides upstream of Box D (or D') [2, 17]. The eukaryotic Box C/D RNAs function as ribonucleoprotein (RNP) complexes. At least four proteins are core components of the RNPs: fibrillarin (the likely methyltransferase [18, 19]), Nop56, Nop58, and p15.5 kDa [4–6, 8, 9]. Two additional proteins, p50 and p55, also interact with the RNAs and are required for the stability and/or function of the RNAs [7, 10]. Box C and Box D are required for interaction of the RNAs with proteins and for the biogenesis, transport, and function of the eukaryotic Box C/D RNAs [1, 17, 20, 21].

The discovery of homologous RNAs containing Box C and Box D elements in at least six archaeal species [11, 12] suggested the remarkable possibility that the essential elements of the Box C/D RNAs had not diverged in Archaea and Eukarya over a vast period of independent evolution. However, the minimal features required for assembly of a functional Box C/D RNP in eukaryotes have not yet been defined. The archaeal Box C/D RNAs do not possess significant sequence homology with eukaryotic Box C/D RNAs outside of the Box C, C', D, and D' sequence elements [11, 12]. Moreover, the identified archaeal Box C/D RNAs are more compact than eukaryotic Box C/D RNAs (e.g., on average 56 nucleotides in P. furiosus versus 104 in Saccharomyces cerevisiae) [11, 22].

Analysis of the sequences of archaeal homologs of the eukaryotic Box C/D RNA-associated proteins suggests that the archaeal Box C/D RNAs interact with a simpler set of proteins. For example, archaeal fibrillarin homologs lack the glycine arginine-rich (GAR) domain that is characteristic of fibrillarin in eukaryotes [18, 23, 24]. Moreover, five of the eukaryotic proteins, Nop56, Nop58, p50, p55, and p15.5 kDa, appear to be products of gene duplications that have occurred since the divergence of Archaea and Eukarya [7, 20, 23]. For example, Nop56, Nop58, and the spliceosomal protein Prp31 are related eukaryotic proteins all represented by a single archaeal homolog [23]. Thus, the six eukaryotic Box C/D RNA-associated proteins are represented by just four homologous proteins in archaea [20]. All six of the eukaryotic Box C/D RNA-associated proteins are essential [6, 10, 25, 26], indicating that the duplicated genes (paralogs) have evolved to encode proteins with distinct and critical functions in eukaryotes. We were interested to know whether the eukaryotic proteins would recognize archaeal Box C/D RNAs.

In this work, we have tested the consequences of the evolutionary distance and extant differences between the Box C/D RNAs and associated proteins in Archaea and Eukarya. We cloned the coding sequence of four (of 52) Box C/D RNAs predicted to exist in P. furiosus: sR2, sR3, sR29, and sR42 [11] and equivalent to sR10, 25, 35, and 30, respectively, in [12]. We also confirmed by RT-PCR analysis that the computationally predicted P. furiosus sR2, sR3, sR29, and sR42 are expressed RNAs (data not shown). In eukaryotes, Box C/D RNAs are actively retained in the nucleus and function within
Figure 1. Archaeal Box C/D RNAs Are Stable and Retained in the Nucleus of Xenopus laevis Oocytes

In vitro transcribed, 32P-labeled *P. furiosus* sR29 (A) and sR2 (B) and coinjected *X. laevis* U3 Box C/D snoRNA and U1 sm- snRNA present in the nucleus (N) and cytoplasm (C) 1, 4, 8, and 24 hr after microinjection into *X. laevis* oocyte nuclei. Antisense (as) sR29 (A) and sR2 (B) RNA panels are from separate experiments. Marker lanes (M) show RNAs prior to microinjection.

the nucleolus [21, 27]—structures that are not known to exist in archaea. To test the fate of archaeal Box C/D RNAs in a eukaryotic cell, radiolabeled RNAs were produced by in vitro transcription and microinjected into nuclei of *Xenopus laevis* oocytes. The nuclei were dissected from the oocytes at various times after microinjection, and the nucleocytoplasmic distribution of the RNAs was determined (Figure 1). Remarkably, the four archaeal RNAs were stable and retained in the nucleus for up to 24 hr after microinjection (Figure 1, sR3 and sR42 not shown). In contrast, antisense transcripts of the archaeal RNAs were exported to the cytoplasm similar to spliceosomal U1Sm-snRNA (Figure 1). To determine whether the *P. furiosus* Box C/D RNAs localized to nucleoli, we analyzed nuclear spreads prepared 1 hr after microinjection of fluorescently labeled RNAs. The archaeal RNAs were targeted specifically to the central, dense fibrillar region of nucleoli in a pattern indistinguishable from eukaryotic snoRNAs (Figure 2, sR3 and sR42 not shown, and [28]). Thus, the archaeal RNAs were retained in the nucleus and targeted to nucleoli in *X. laevis* oocytes.

In eukaryotes, Box C/D RNAs interact with fibrillarin, Nop56, and Nop58 [4, 8, 9]. Fibrillarin shares structural features with methyltransferases and is thought to catalyze the 2’-O-methylation of rRNA, guided by the individual associated snoRNAs [5, 19]. Nop56 and Nop58 are both required for rRNA modification, but the precise roles of the proteins are not known [8, 9, 26]. We tested whether the archaeal RNAs interacted with eukaryotic Nop56, Nop58, and fibrillarin by immunoprecipitation of the endogenous *X. laevis* proteins 24 hr after microinjection of archaeal and control RNAs. Coinjected eukaryotic Box C/D RNAs (U8 and/or U14) but not spliceosomal snRNAs (U1 and U6) are associated with each of the proteins (Figure 3). Interestingly, U3 RNA, a Box C/D snoRNA involved in rRNA processing rather than modification, was associated with fibrillarin and Nop56 but not extensively with Nop58 (Figure 3). All four of the *P. furiosus* Box C/D RNAs were coimmunoprecipitated with endogenous *X. laevis* fibrillarin and both Nop56 and Nop58 (Figure 3, sR3 and sR42 not shown). Our results demonstrate that the endogenous eukaryotic proteins, including both the Nop56 and Nop58 paralogs, interact with the archaeal Box C/D RNAs.

Evidence suggests that the Box C/D RNAs also direct 2’-O-methylation of rRNA in archaea. Each archaeal RNA is predicted to guide modification of 1 or 2 rRNA sites that can be deduced from the internal complementary sequences using the “Box D + 5 nucleotide” rule derived from studies in eukaryotes [2, 3, 11, 12]. The total number of modifications, as well as the positions of some specific sites that have been mapped in archaeal species, are consistent with predictions based on the identified archaeal Box C/D RNAs [11, 12]. In addition,
Figure 3. Archaeal Box C/D RNAs Interact with Eukaryotic Proteins (Fibrillarin, Nop56, and Nop58)
P. furiosus sR29 and sR2 were co-injected with Xenopus laevis Box C/D RNAs U3, U8, and/or U14 and with spliceosomal snRNAs U1 and U6. The RNAs present in the nuclear extract prior to immunoprecipitation (20% of total, T), immunoprecipitated (P) with antibodies against fibrillarin (A), Nop56 (B), or Nop58 (C) and remaining in the supernatant (20% of supernatant, S) are shown.

The guide sequences upstream of Box D in sR29 and sR2 were replaced with sequences complementary to Xenopus laevis 28S rRNA to generate sR29ng3021 (A) and sR2ng3018 (B), respectively. The predicted sites of Xenopus rRNA modification (paired with the base five nucleotides upstream of Box D) are indicated (G3021 and A3018, circled). Reverse transcriptase primer extension pause assays [22, 35] performed on nuclear RNA from oocytes following injection of sR29ng3021, sR2ng3018, or no RNA (uninjected) are shown. Arrowheads indicate dNTP concentration-dependent pauses observed (typically one nucleotide of the modified base [35]) specifically in the presence of the archaeal RNAs. An endogenous rRNA modification (asterisk) is detected in the absence and presence of the archaeal RNAs. Adjacent lanes show reference sequence ladders produced with the primer utilized in the pause assays.

Figure 4. Archaeal Box C/D RNAs Guide 2'-O-Ribose Methylation of Xenopus laevis rRNA
The guide sequences upstream of Box D in sR29 and sR2 were replaced with sequences complementary to X. laevis 28S rRNA to generate sR29ng3021 (A) and sR2ng3018 (B), respectively. The predicted sites of Xenopus rRNA modification (paired with the base five nucleotides upstream of Box D) are indicated (G3021 and A3018, circled). Reverse transcriptase primer extension pause assays [22, 35] performed on nuclear RNA from oocytes following injection of sR29ng3021, sR2ng3018, or no RNA (uninjected) are shown. Arrowheads indicate dNTP concentration-dependent pauses observed (typically one nucleotide of the modified base [35]) specifically in the presence of the archaeal RNAs. An endogenous rRNA modification (asterisk) is detected in the absence and presence of the archaeal RNAs. Adjacent lanes show reference sequence ladders produced with the primer utilized in the pause assays.

The results described here are a unique demonstration of the ability of a class of RNAs to be recognized, be transported, and function in an organism from another domain of life. The accurate nuclear retention and sub-nuclear localization of RNAs from a prokaryote in a eukaryotic cell suggest that these processes are mediated by elements that existed in common ancestral RNAs prior to the divergence of Archaea and Eukarya and development of the eukaryotic nucleus. Our findings demonstrate that the RNA-protein interactions essential for the transport and function of Box C/D RNAs have not changed significantly since the divergence of Arch-
We propose that the remarkable recognition of archaean RNAs by eukaryotic proteins reflects a limitation on divergence imposed by the requirement for coevolution of the Box C/D proteins with each of multiple individual RNAs.

**Experimental Procedures**

**Synthesis of RNAs**

Sequences of putative Box C/D RNAs were obtained from a computational search of the *P. furiosus* genome [11] and are available at http://rna.wustl.edu/snoRNAdb/. Primers complementary to sR2, sR3, sR29 and sR42 were used for RT-PCR from purified *P. furiosus* total RNA (kindly provided by Gerti Schut, Dr. Amy Grunden, and Dr. Michael Adams, University of Georgia). Primers were designed to include sequences complementary to terminal stem structures [11] and exogenous T7 promoter sequences for in vitro transcription (underlined): sR2-5' oligo, TAATACGACTCACTATAGGGGATGATGAAATAG was supported by grants from the National Institutes of Health and the National Science Foundation to M.P.T. and R.M.T.

**Acknowledgments**

We are grateful to Michael Adams, Amy Grunden, and Gerti Schut (University of Georgia) for discussions, guidance, and *P. furiosus* RNA. We thank Michael Adams, John Avise, Claiborne V.C. Glover III, Richard Meagher (University of Georgia), Sean Carroll, James Dahlberg, Elisebeth Lund (University of Wisconsin), and John Logsdon (Emory University) for critical reading of this manuscript. This work was supported by grants from the National Institutes of Health and the National Science Foundation to M.P.T. and R.M.T.

Received: November 12, 2001
Accepted: November 23, 2001
Published: February 5, 2002

**References**


