Orig life Evol Biosph (2006) 36:421–429
DOI 10.1007/s11084-006-9011-z

Ribosome origins: The relative age of 23S rRNA Domains

James Hury · Uma Nagaswamy · Maia Larios-Sanz · George E. Fox

Received: 17 October 2005 / Accepted in revised form: 10 January 2006

Abstract The modern ribosome and its component RNAs are quite large and it is likely that at an earlier time they were much smaller. Hence, not all regions of the modern ribosomal RNAs (rRNA) are likely to be equally old. In the work described here, it is hypothesized that the oldest regions of the RNAs will usually be highly integrated into the machinery. When this is the case, an examination of the interconnectivity between local RNA regions can provide insight to the relative age of the various regions. Herein, we describe an analysis of all known long-range RNA/RNA interactions within the 23S rRNA and between the 23S rRNA and the 16S rRNA in order to assess the interconnectivity between the usual Domains as defined by secondary structure. Domain V, which contains the peptidyl transferase center is centrally located, extensively connected, and therefore likely to be the oldest region. Domain IV and Domain II are extensively interconnected with both themselves and Domain V. A portion of Domain IV is also extensively connected with the 30S subunit and hence Domain IV may be older than Domain II. These results are consistent with other evidence relating to the relative age of RNA regions. Although the relative time of addition of the GTPase center can not be reliably deduced it is pointed out that the development of this may have dramatically affected the progenotes that preceded the last common ancestor.

Keywords Connectivity · Long-range contacts · Origins of translation · Progenote · Protein synthesis · Ribosomes

Introduction

Modern organisms use ribosomes to synthesize defined sequence peptides of a single chirality. Genomic comparisons have revealed that all three kingdoms of life share a significant portion of the ribosomal machinery (Olsen and Woese, 1997). Studies of structure and function have revealed that the ribosome is almost certainly an RNA machine (Nissen *et al*., 2000).

J. Hury · U. Nagaswamy · M. Larios-Sanz · G. E. Fox (✉)
Dept. Biology and Biochemistry, University of Houston, Houston, Tx. 77204-5001
e-mail: fox@uh.edu

Together, these results suggest that functional ribosomal machinery was already present in the last universal common ancestor of the three Domains of life as defined by sequence comparisons and therefore strongly support the notion that the ribosomal machinery had its origins in an RNA World. In fact, the emergence of the ribosomal machinery may have played a significant role in terminating the RNA World by making protein intensive living systems possible (Fox and Naik, 2004). The ribosomal machinery is far too complex to spring forth fully formed and unchanged and hence not all of its components can be of the same age. From the RNA perspective, it seems clear that the original ribosomal RNAs, (rRNAs), were much smaller and that the RNA component has grown more complex possibly as an amalgamation of originally independent functional fragments (Clark, 1987). Thus, some regions of the modern rRNAs are likely to be older than others. This is in fact clearly true for a number of hyper-variable regions of the RNA that differ in size and structure in the large subunit unit rRNAs from different Domains of life (Clark, 1987). The oldest parts of the rRNAs would be especially central to the early evolution of the translation machinery and therefore it is of interest to identify them. Herein we are attempting to identify important features that provide insight into the relative timeframe for additions of various ribosomal components. If many of these origins of important features can be identified, it will be possible to assemble a cogent hypothesis about ribosomal origins.

Methods

23S rRNA is generally considered to have six major domains that are defined by secondary structure (Cannone *et al.*, 2002). Each of these Domains was further sub classified into logical subdomains suggested by secondary structure (Figure 1). Long-range base-base interactions were catalogued by examination of the 23S rRNA tertiary structure as determined by X-ray crystallography (Ban *et al.*, 2000) using the Swiss-Pdb Viewer (Guex and Peitsch, 1997). An interaction was considered to be long-range if it connects two or more bases from different subdomains. Atomic resolution studies of the 70S ribosome as a whole are not yet at sufficient resolution to identify specific base-base contacts between 23S rRNA and 16S rRNA. Nevertheless, major sites of interactions between 23S and 16S have been localized (Yusupov *et al.*, 2001). For the purpose of the analysis presented here, each of these contacts was considered to be equivalent to one interaction. Certain regions of the 23S rRNA, e.g. *E. coli* positions 851–926 and 2095–2194, are not well resolved in the crystal structure. This is presumably because their correct conformation requires the presence of the 30S subunit or tRNAs. These regions were not included in the analysis. The number of bases contained in each subdomain, excluding unresolved regions, was used to normalize the interactions involving a particular domain or subdomain. Each subdomain's interactions were recorded and the eight subdomains with the most interactions were isolated with the Swiss-Pdb Viewer to determine the relative locations of the different subdomains.

Results

Figure 2 illustrates the long-range interactions emanating from Domain II. The numbers of long range interactions involving each of the major domains are summarized in Table I. Table II summarizes the results for the eight most extensively connected subdomains. Initial crystallographic studies of whole ribosomes (Yusupov *et al.*, 2001) suggest that the vast majority of the interactions between 16S and 23S rRNA involve Domain IV (particularly

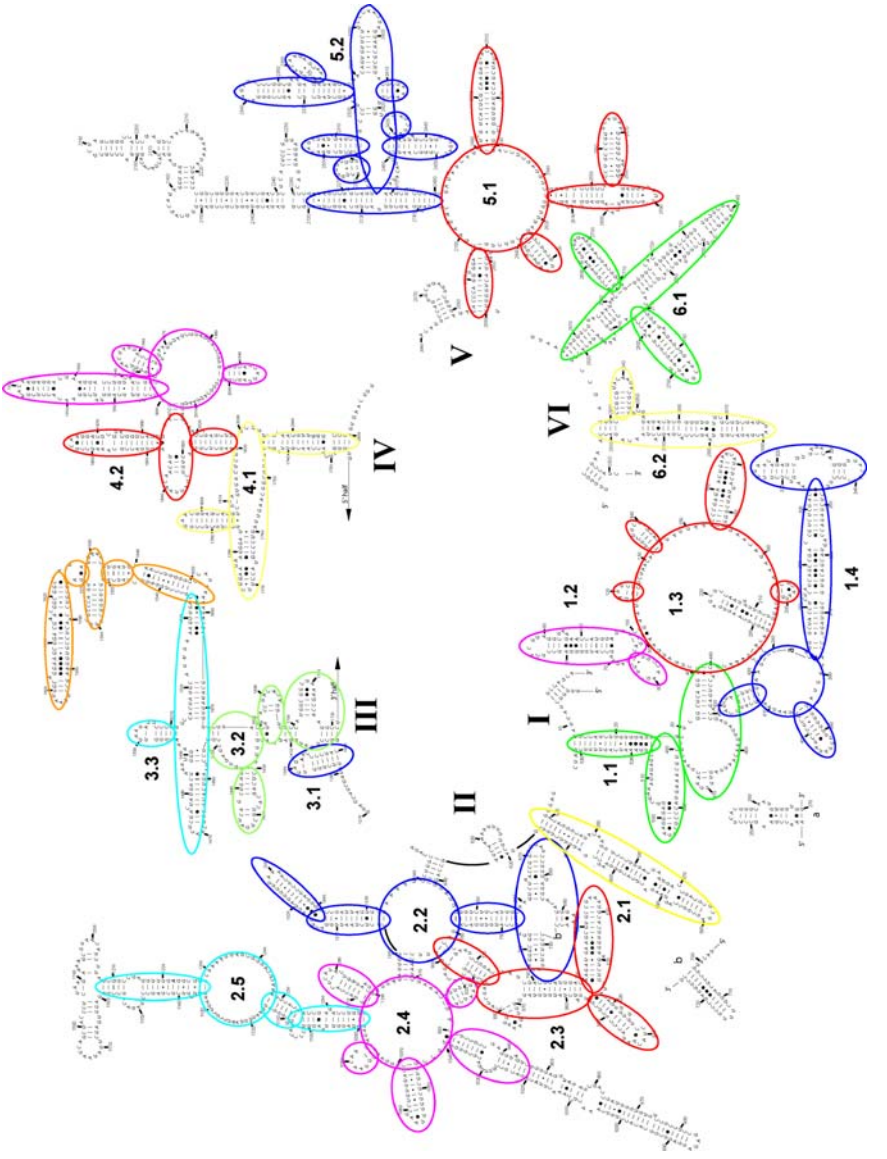

Fig. 1 *Haloarcula marismortui* 23S rRNA secondary structure. The six major domains and logical subdomains based on secondary structure are shown

Fig. 2 Tertiary interactions involving residues in Domain II of *Haloarcula marismortui* 23S rRNA. Each line connects a base in Domain II that forms a base-base interaction with a residue outside of Domain II

Table 1 Domain connectivity

Domain connectivity				
Domain	# of residues	# of Interactions	Ratio	# of tRNA/RNA Contacts
I	510	17	0.033	2
II	503	31	0.062	2
III	350	8	0.023	0
IV	361	32	0.089	6
V	477	36	0.075	12
VI	226	14	0.062	0

Table 2 Most connected subdomains

Subdomain	# of Residues	# of Interactions	Ratio
4.2	77	10	0.13
5.1	145	16	0.11
4.3	145	15	0.103
2.3	138	12	0.087
5.2	237	19	0.08
6.1	158	11	0.07
1.4	162	10	0.062
4.1	137	7	0.051

subdomain 4.3) in 23S rRNA. This is consistent with earlier studies (Frank and Agrawal, 2000; Gabashvili *et al*., 1999) that identified multiple RNA/RNA bridges involving nucleotides in subdomain 4.2 and 4.3. The crystal structure results also suggest some contacts occur between 23S rRNA Domain II and the 3′ mini domain of 16S rRNA. The most interconnected regions in 23S rRNA are Domains IV and V. Domains II and VI also have considerable connectivity with both of these being primarily connected to Domain IV and V. The least connected regions are clearly Domains III and much of Domain I. Figure 3 shows the location of the eight most connected subdomains in 23S rRNA. As summarized by Mears *et al*., 2002 the rRNAs also have conserved contacts with the tRNAs in the A, P and E-sites. Most of these involve subdomains tRNAs in terms of proximity to the tRNAs in the A, P and E site.

Discussion

The results show a large number of interconnections between selected regions of Domains II, Domain IV and Domain V. Domain V contains the RNA region responsible for the peptidyl transferase activity, which is at the heart of translation and is among the most integrated and therefore likely has existed since very early times. All of Domain IV and much of Domain II are also highly interconnected and are likely to have a longstanding history. Although, Domain II has essentially the same number of total connections as Domain IV, it is minimally associated with the 30S subunit and typically does not have direct proximity to the tRNAs and is therefore overall less connected than Domain IV. Domain III and many parts of Domain I appear to be minimally connected. The results suggest a large subunit RNA history from oldest to youngest: Subdomains 5.1 and 5.2, followed by essentially all of

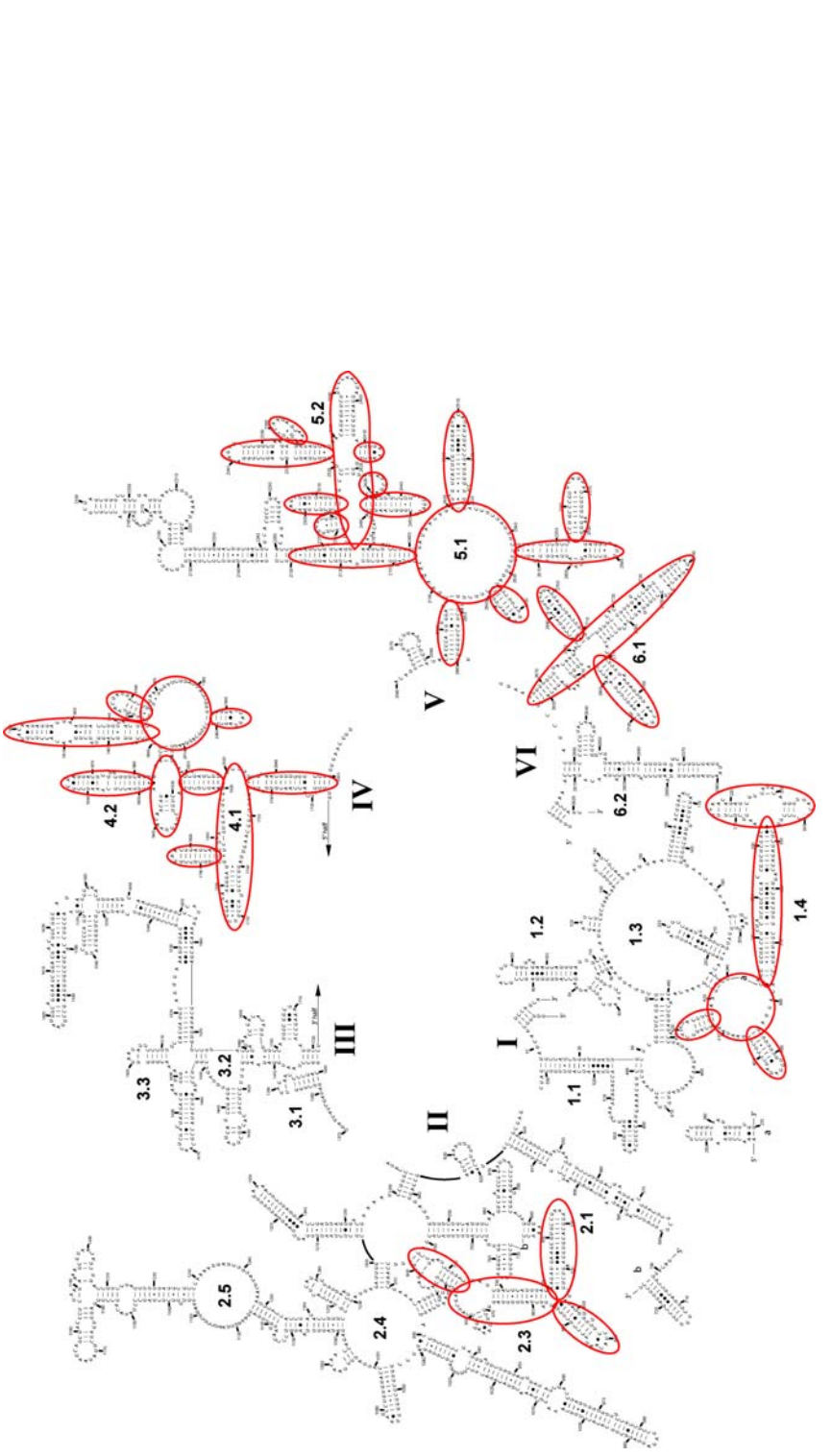

Fig. 3 Most strongly connected regions. The *Haloarcula marismortui* 23S rRNA secondary structure is highlighted to indicate the location of the most connected regions listed in Table II

Domain IV, followed in some order by subdomain 6.1, portions of Domain II, and/or Domain II and the 3′ mini domain of 16S rRNA. Portions of Domain I are also rather old whereas all of Domain III likely was added quite late.

The results presented here are consistent with earlier work in which the conservation at individual nucleotide positions were examined. In that study, it was concluded that many of the most conserved and hence likely oldest positions were located near the center of the molecule where they could serve as a foundation for the incorporation of new additions over time (Wuyts *et al*., 2001). The regions identified here as being the most interconnected contain many of the most conserved regions of the 23S rRNA and hence the considerable agreement between the results is not unexpected.

A second test of the reasonableness of the results represented in Figure 3 is provided by a favorable comparison of the results presented here with earlier efforts to define minimal rRNAs by comparative analysis (Mears *et al*., 2002). Such minimal rRNAs presumably contain all the regions which are functionally essential. Assuming the core ribosome functions have not changed over evolutionary time, one might then infer these regions might date to earliest times. RNAs very similar to the predicted minimal rRNAs (Mears *et al*., 2002) actually occur in the mitochondria of various organisms including nematodes. The highly minimalized large subunit rRNA of *Trypanosoma brucei* (Sloof *et al.*, 1985) encompasses essentially only the areas highlighted in Figure 3. The one exception is a section of the subdomain 2.5 which is present in both the minimal RNAs and all the mitochondrial rRNAs but not the current analysis.

This region of the RNA encompasses the GTPase center of the ribosome where the elongation factors EF-Tu and EF-G bind competitively (Spahn and Nierhaus, 1998; Willson and Noller, 1998). This region is also the target of thiazole antibiotics (Sopori and Lengyel, 1972). The GTPase center of the 23S rRNA is located at the top of Domain II overlaying the binding site of ribosomal protein L11. As indicated in the current analysis, the region folds independently without any tertiary interactions with other domains (Wimberly *et al*., 1999; Klein *et al*., 2004). Within the GTPase center, there are extensive tertiary interactions involving the major and minor grooves of the local RNA strands. After the peptidyl transferase center, this region of the 23S rRNA has the most obvious functional significance and it therefore is initially surprising that it appears to be a late addition to the ribosome. Since this region of the RNA is known to be subject to significant conformational changes during ribosome function (Gonzalo and Reboud, 2003), this allows two *ad hoc* explanations. First, it might be that in order to allow the required movements the GTPase center can not form stable long range interactions. Less likely, the interactions might only be seen when an alternative conformation not seen in the crystal structures is assumed. In either case, the results presented here may not be meaningful for subdomain 2.5. Finally, the possibility may exist that old contacts have been broken by shifts in three-dimensional location over evolutionary time. Such shifts have been suggested for the L11 binding domain and alpha sarcin loop by comparative studies (Mears *et al*., 2002).

Is it, nevertheless, possible that the GTPase center could be a relatively modern addition? Consistent with this, it has been shown that ribosomes can function at a greatly reduced rate in the absence of the GTPase center (Spirin, 2002; Gavrilova and Spirin, 1971; Gavrilova *et al.*, 1971). It also appears that the assembly of the large ribosomal subunit may partially recapitulate its history (Fox and Naik, 2004). From this second perspective, the ribosomal proteins associated with the GTPase center are incorporated relatively late in the assembly process (Nierhaus, 1991). Arguing for an early addition is the fact that r-proteins L7/L12 and L11 are both universal and is one of the few operons whose organization is the same in both Archaea and Bacteria. One might also argue this position from the fact the GTPase center

associated proteins EF-Tu and EF-G are likely to have evolved from a gene duplication event that occurred before the emergence of the last common ancestor. This is less convincing when one recalls that most of the rRNA structure is also shared by all lineages leading to the last common ancestor. Thus, although the analysis presented here really does not address it, we opine that the GTPase domain was likely added at a mid point in subunit evolution.

Of special significance is the likely effect that addition of the GTPase domain would have had on the entity (possibly a progenote-Woese and Fox, 1977) that preceded the last common ancestor defined by sequence comparisons. The addition of the GTPase center to ribosomes would have dramatically increased the rate of peptide synthesis in the progenotes possessing it and consequently the types and size of peptide products that could be made. This would be a major breakthrough on the way to an RNA/protein World.

The connectivity survey described here is just one of many approaches that might provide suggestive information about the relative age of various ribosomal components. It remains necessary to fully consider the consequences of this connectivity survey in the context of other information. For example, some ribosomal proteins are readily identified in organisms from all three Domains of life and hence likely to be older than proteins that are only uniquely found in one Domain of life (Mushegian, 2005). The newest proteins likely help fold the newer RNA regions in which case they would interact primarily with the less tightly integrated regions of the RNA. Alternatively, the newer proteins might also assist in incorporating the newer RNA sections into the structure, in which case they will be seen to bind with both newer and older regions of the RNA. The problem is further complicated by the fact that the portions of individual ribosomal proteins may also be of different ages than other portions. We are currently examining the protein/RNA contact information in detail to examine the extent to which apparently older proteins may or may not be interacting with the more strongly connected regions of the RNA. We are optimistic that it will ultimately be possible to infer much about the evolutionary history of the ribosomal machinery from detailed study of the structures and interactions of the various ribosomal components.

Acknowledgments

This work was supported in part by grants from the National Aeronautics and Space Administration Exobiology Program (grant NNG05GN75G and its predecessor NAG5-12366) and the University of Houston Institute of Space Systems Operations to G.E.F..

References

Ban N, Nissen P, Hansen J, Moore PB, Steitz TA (2000) The complete atomic structure of the large ribosomal subunit at 2.4 A resolution. Science 289:905–920.

Cannone JJ, Subramanian S, Schnare MN, Collett JR, D'Souza LM, Du Y, Feng B, Lin N, Madabusi LV, Muller KM, Pande N, Shang Z, Yu N, Gutell RR (2002) The Comparative RNA Web (CRW) Site: An online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. BMC Bioinformatics 3:2.

Clark CG (1987) On the evolution of ribosomal RNA. J Molec Evol 25:343–350.

Fox GE, Naik AK (2004) The Evolutionary History of the Ribosome. IN: Ribas de Poplana, L. (Ed.). The Genetic Code and the Origin of Life. Landes Bioscience, Georgetown and New York, pp. 92–105.

Gavrilova LP, Spirin AS (1971) Stimulation of "non-enzymic" translocation in ribosomes by *p*-chloromercuribenzoate. FEBS Lett 17:324–326.

Gavrilova LP, Kostiashkina OE, Koteliansky VE, Rutkevich NM, Spirin AS (1976) Factor-free ("Non-enzymic") and factor-dependent systems of translation of polyuridylic acid by *Escherichia coli* ribosomes. J Mol Biol 101:537–552.

Guex N, Peitsch MC (1997) SWISS_MODEL and the Swiss-Pdb Viewer: An environment for comparative protein modeling. Electrophoresis 18:2714–2723.
Gonzalo P, Reboud JP (2003) The puzzling lateral flexible stalk of the ribosome. Biol Cell 95:179–193.
Klein DJ, Moore PB, Steitz TA (2004) The roles of ribosomal proteins in the structure assembly, and evolution of the large ribosomal subunit. J Mol Biol 340:141–177.
Mears JA, Cannone JJ, Stagg SM, Gutell RR, Agrawal RK, Harvey SC (2002) Modeling a minimal ribosome based on comparative sequence analysis. J Mol Biol 321:215–234.
Mushegian A (2005) Protein content of minimal and ancestral ribosome. RNA 11:1400–1406.
Nierhaus KH (1991) The assembly of prokaryotic ribosomes. Biochimie 73:739–755.
Nissen P, Hansen J, Moore PB, Steitz TA (2000) The structural basis of ribosomal activity in peptide bond synthesis. Science 289:920–930.
Olsen GJ, Woese CR (1997) Archaeal genomics: An overview. Cell 89:991–994.
Sloof P, Van den Burg J, Voogd A, Benne R, Agostinelli M, Borst P, Gutell R, Noller H (1985) Further characterization of the extremely small mitochondrial ribosomal RNAs from trypanosomes: A detailed comparison of the 9S and 12S RNAs from *Crithidia fasciculata* and *Trypanosoma brucei* with rRNAs from other organisms. Nucleic Acids Res 13:4171–4190.
Sopori ML, Lengyel P (1972) Components of the 50S ribosomal subunit involved in GTP cleavage. Biochem Biophys Res Commun 46:238–244.
Spahn CM, Nierhaus KH (1998) Models of the elongation cycle: An evaluation. Biol Chem 379:753–772.
Spirin AS (2002) Ribosome as a molecular machine. FEBS Lett 514:2–10.
Woese CR, Fox GE (1977) Phylogenetic structure of the prokaryotic domain: The primary kingdoms. Proc Natl Acad Sci USA 74:5088–5090.
Wilson KS, Noller HF (1998) Mapping the position of translational elongation factor EF-G in the ribosome by directed hydroxyl radical probing. Cell 92:131–139.
Wimberly BT, Guymon R, McCutcheon JP, White SW, Ramakrishnan V (1999) A detailed view of a ribosomal active site: The structure of the L11-RNA complex. Cell 97:491–502.
Wuyts J, Van de Peer Y, De Watcher R (2001) Distribution of substitution rates and location of insertion sites in the tertiary structure of ribosomal RNA. Nucleic Acids Res 29:5017–5028.
Yusupov MM, Yusupova GZ, Baucom A, Leiberman K, Earnest TN, Cate JH, Noller HK (2001) Crystal structure of the ribsome at 5.5 A resolution. Science 292:883–896.