Homoiterons and expansion in ribosomal RNAs

Michael S. Parker a, Floyd R. Sallee b, Edwards A. Park c, Steven L. Parker a,⁎

a Department of Pharmacology, University of Tennessee Health Sciences Center, Memphis, TN 38163, USA
b Department of Psychiatry, University of Cincinnati School of Medicine, Cincinnati, OH 45226, USA
c Department of Microbiology and Molecular Cell Sciences, University of Memphis, Memphis, TN 38152, USA

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

The sequences of DNAs and most RNAs in all types of cells, plasmids and viruses contain multiple repeats of the same nucleotide. In bacterial plasmids such repeats are also present within the repeating mixed-nucleotide motifs labeled as iterons [1,2]. As will be shown in this work, numerous same-nucleotide repeats with three or more units are both differentially and non-randomly represented in ribosomal RNAs, and especially abundant in the expansion segments of the vertebrate LSU rRNAs. To avoid conflict with the bacterial plasmid usage, the identical repeats in polyribonucleotides will be termed homoiterons in this work. Homoi-terons with three or more units can stably pair with the canonical antisense counterparts [3,4], and therefore could be involved in significant structuring interactions, and would also be important in association with other RNAs, and with proteins. Multiple homoiterons are parts of DNA initiation sites, promoters and telomeres [5–7]. Homoi-terons in RNAs are mainly studied as poly(A) stretches involved in mRNA regulation and disposal (see Refs. [8,9]). However, the biological significance of RNA homoi-terons extends much further than polyadenylation-linked processing of mRNAs. Homoi-terons are among the codons for Lys, Pro, Gly and Phe, and the corresponding antisense homoi-terons are found in the anticodon loops in subspecies of the corresponding tRNAs. The “slippery” UUUUUUA motif in many RNA viruses induces an obligatory frameshift in viral protein synthesis [10,11]. The abundant 5’UTR and 3’UTR homoiterons should participate in the respective SSU RNAs. This is especially pronounced in the tetrapod vertebrates and seems to culminate in the hominid mammals. The stability of secondary structure in polyribonucleotides would significantly connect to GC content, and should also relate to G and C homoiteron content. RNA modeling points to considerable presence of homoiteron-rich double-stranded segments especially in vertebrate LSU RNAs, and homoi-terons with four or more nucleotides in the vertebrate and angiosperm LSU RNAs are largely confined to the expansion segments. These features could mainly relate to protein export function and attachment of LSU to endoplasmic reticulum and other subcellular networks.

Abbreviations: aa, amino acid residues; ES, an expansion segment; LSU, large cytoplasmic ribosome subunit (30S in prokaryotes and archaea, 60S in eukaryotes); mRNP, messenger ribonucleoprotein; ncRNA, non-coding RNA; nt, nucleotides; PCN, homionic motifs with >3% and >50% ionic residues, found especially in Polynucleotide-binding proteins, Carrier proteins and Nuclear localization signals; SSU, small cytoplasmic ribosome subunit (30S in prokaryotes and archaea, 40S in eukaryotes); u, nucleotide unit; XN or NX, [X = a number] a nucleotide unit with same nucleobases (homoiterons). In prokaryotes these repeats are much more frequent in thermophile compared to mesophile or psychrophile species, and have similar frequency in both large RNAs. These features point to use of prokaryotic homoiterons in stabilization of both ribosomal subunits. The two large RNAs of eukaryotic cytoplasmic ribosomes have expanded to a different degree across the evolutionary ladder. The big RNA of the larger subunit (60S LSU) evolved expansion segments of up to 2400 nucleotides, and the smaller subunit (40S SSU) RNA acquired expansion segments of not more than 700 nucleotides. In the examined eukaryotes abundance of rRNA homoiterons generally follows size and nucleotide bias of the expansion segments, and increases with GC content and especially with phylogenetic rank. Both the nucleotide bias and frequency of homoiterons are much larger in metazoan and angiosperm LSU compared to the respective SSU RNAs. This is especially pronounced in the tetrapod vertebrates and seems to culminate in the hominid mammals. The stability of secondary structure in polyribonucleotides would significantly connect to GC content, and should also relate to G and C homoiteron content. RNA modeling points to considerable presence of homoiteron-rich double-stranded segments especially in vertebrate LSU RNAs, and homoi-terons with four or more nucleotides in the vertebrate and angiosperm LSU RNAs are largely confined to the expansion segments. These features could mainly relate to protein export function and attachment of LSU to endoplasmic reticulum and other subcellular networks.

© 2015 The Authors. Published by Elsevier B.V. on behalf of the Federation of European Biochemical Societies. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
association of mRNAs with ribosomes [12,13], RNA-processing enzymes [14] and microRNAs [15]. The G homoiterons should help association with ionic amino acid sidechains [16–18]. The C repeats can accommodate ionic and hydrophobic amino acid clusters, and this should be even more pronounced with U repeats [19,20].

Large RNAs of both cytoplasmic ribosomal subunits of the contemporary prokaryotes and eukaryotes have extensive similarities in the conserved cores [21–23], indicating common ancestry from similar precursors. An internal expansion is evident in both large rRNAs of the eukarya, with some 15 common large expansion segments (ES) in LSU RNAs, and 11 in SSU RNAs. As will be shown in this work, both RNAs have expanded to a similar degree in most of the lower eukarya, plants and invertebrate metazoans. The vertebrate 28S RNAs however developed much larger and more GC-biased expansion segments (as partly documented in Ref. [24]).

We became interested in homoiterons of ribosomal RNAs through examination of the expansion segments of mammalian 28S rRNAs, (which were defined to a consensus in several modeling studies [23,25,26], and contain 45% homoiterons). As will be shown in this survey, internal homoiterons are highly represented in rRNAs. This subject is related to a number of studies of the expansion segments (ES) of eukaryotic rRNAs, but the ES homoiterons thus far were not directly studied. Localization and abundance of these repeats in rRNAs have, as will be shown, a considerable phylogenetic relatedness, and in vertebrate rRNAs parallel the highly regimented clustering of basic sidechains in ribosomal proteins [24].

2. Methods

2.1. The RNA sequences examined

Ribosomal RNA sequences were retrieved from Entrez nucleotide database (http://www.ncbi.nlm.nih.gov/nuccore), with the aid of access codes from Comparative RNA Web site (CRW: http://www.rna.icmb.utexas.edu). The list of examined rRNAs is presented as the first section of the Supplementary Data. The sequences include both free-living species and the intracellular parasites such as euglenozoan trypanosomes. Sequences of rRNAs of diplomonads (e.g. Giardia species) appear as not finalized, and were not included. To avoid mismatching, in all evaluations the insect sequences were sub-grouped based on the GC bias of the large LSU RNA. Human long nuclear RNAs were taken from IncRNAdb (http://incrnadb.com; see [70]). The human microRNAs were taken from mirBase (http://mirbase.org). The ribosomal protein mRNA sequences were taken from RPF (Ribosomal Protein Gene) database, http://ribosome.med.miyazaki-u.ac.jp. Other mRNA sequences were retrieved from the Ensembl database (http://www.ensembl.org).

2.2. Boundaries of the expansion segments

The segment boundaries of human 28S rRNA were searched for in clustalW alignments of other ribosomal RNAs to score the matching starting and ending nucleotides. The matches located in alignment interrupts were shifted to the nearest preceding nucleotide. This approach defined segments that for rRNAs in Table S2 correspond well with published values from modeling of RNA structure. Moreover, the size and nucleotide composition of the ES of RNAs not previously modeled (such as rat 28S rRNA) are predicted as very similar to those published for other species in the same taxonomic class (thus, mouse or human for the rat).

We find that boundaries derived by matching with human 28S rRNA correspond well with those defined by modeling in Mus musculus, Xenopus laevis and, over an immense phylogenetic distance, in Saccharomyces cerevisiae (Table S3). However, the ES boundaries for SSU rRNAs were thus far not defined to a consensus pattern between any eukaryotic phyla. We used the preliminary bounds for human 18S rRNA (Fig. S7 in Ref. [26]; also listed in our Table S1). The partial preliminary bounds suggested for yeast 18S rRNA (Table S3 in Ref. [35]) feature discontinuous sub-segments for the large ES6 and could not be employed for the alignment-based predictions.

2.3. RNA structure programs

Predictions of polyribonucleotide structures were obtained with RNAstructure program [3] and with RNAfold program [4]. These programs were also used for modeling the free energies of secondary structure formation/disbanding for homoiterons.

2.4. RNA–protein binding

The RNAbindR program [71] was used to evaluate RNA-binding potential of ribosomal proteins.

2.5. Statistical comparisons

The Schéfè t test following significant analysis of variance was used for evaluations within LSU and SSU groups. This is justified in terms of normalcy of the distribution of most intra-group parameters. The Wilcoxon rank sum test was used for evaluation of differences between the paired LSU and SSU values. The two-tailed p < 0.05 differences were considered as significant.

3. Results

3.1. Homoiteron type, size and frequency in ribosomal RNAs

Homoiterons in RNAs can be broadly divided by length and location in two types. Type I homoiterons are found inside sequences of all RNA classes, feature any of the four main nucleotides, and rarely exceed 10 nt in length. Type II homoiterons (absent from ribosomal RNAs) are mainly found at the 3′ termini of mRNAs, frequently have more than 10 units and in most cases are constituted of A or U nucleotides. In long sequences, type I G homoiterons largely model to stems, and the A and U units of this type usually locate to loops or to stem/loop borders. The type II homoiterons (present at n > 10 at the 3′ end of less than 1% of human chromosomally defined mRNA sequences, but elaborated post-transcriptionally in most mRNAs) are essentially devoid of secondary structure, and could serve as parts of docking platforms for extraneous partners, including poly(A)-binding proteins and translation factors [27] and RNases [28].

The size of homoiterons in ribosomal RNAs is quite limited. In 84 eukaryotic cytoplasmic rRNAs examined in this work there are only fifteen 8-unit (u) and five 9–12u repeats. No homoiterons above 7u are found in 96 examined prokaryotic rRNAs. (Fifty-five human long non-coding RNAs (IncRNAs), selected as similar in size to rRNAs (1000–5000 nt), have 19 >10u A and U, and no >10u G and C homoiterons (Table S4). Both large ribosomal RNAs have large numbers of 3u and 4u homoiterons (Fig. 1), presenting a substantial capacity for interaction with either RNA or protein sequences. In view of their frequency in both LSU and SSU RNAs, long homoiterons could support RNA secondary structure and stability of association with proteins in both ribosomal subunits of the thermophilic prokaryotes.

Frequency of eukaryotic LSU G and C homoiterons is distinctly higher in tetrapod vertebrates compared to other groups (Fig. 1A and C, the left column of test significances). This is not found for SSU G repeats (Fig. 1B) and 3u or 4u C repeats
Fig. 1. Frequency of G, C, A and U homoiterons in ribosomal RNAs as number of 3-unit, 4-unit and >4-unit repeats per 100 sequence nucleotides. Graphs A, C, E, G: 23–28S RNAs of the larger ribosomal subunit. Graphs B, D, F, H: 16–18S RNAs of the smaller subunit. Inscriptions in the left column show Scheffé t test significance vs. mammalian group for the eukaryotic, and vs. archaeal or bacterial thermophile group for the prokaryotic groups as comparators; n (or empty position) indicates p > 0.05 (or non-applicable, in the case of the comparator groups) and L stands for p < 0.05 lower values for 3, 4 and >4-unit homoiterons, respectively. Inscriptions in the right column of graphs A, C, E and G show Wilcoxon rank sum test significance for LSU compared to SSU RNA homoiterons of the corresponding group (H, higher; L, lower; n or blank, p > 0.05). The number of species examined is indicated in brackets following the group labels.

(Fig. 1D), but >4C SSU homoiterons are more frequent in tetrapods (Fig. 1D). The LSU G and C homoiterons have larger frequency than their SSU counterparts in tetrapod vertebrates, but generally not in other eukaryote groups (Fig. 1A–D). Among prokaryotic RNAs, thermophile LSU and SSU RNAs have more G (Fig. 1A and B) and C (Fig. 1C and D) homoiterons than the respective mesophile or psychrophile groups, and RNAs of both subunits have a considerable excess of G over C repeats in any size range. This imbalance is also present, but less pronounced, in eukaryote LSU RNAs.

In both LSU and SSU RNAs frequencies of A repeats (Fig. 1E and F) and U repeats (Fig. 1G and H) in the invertebrate metazoa are above those in the vertebrates, some at high levels of significance. This is however not observed in lower eukarya.

Comparisons of homoiteron frequency between LSU and SSU sequences were done in Wilcoxon rank sum tests on the corresponding RNA groups. Differences in the frequency comparisons (Fig. 1) are quite similar to those in the homoiteron sequence fraction comparisons (Fig. 3). The G and C homoiterons in vertebrate LSU RNAs are generally more frequent than in the SSU sequences (Fig. 1, the right columns of test significances). This is in most cases not observed with A homoiterons (graphs E, F of Fig. 1). There are several large differences with U homoiterons, especially the >4-units (graphs G, H of Fig. 1). The *X. laevis* frequencies have high LSU/SSU G and C repeat ratios, similar to mammalian and fish groups. The angiosperm LSU G, C and even U repeat frequencies are higher in LSU than in SSU RNAs.

Frequencies of the archaeal LSU and SSU G and C repeats do not differ for the two groups of thermophiles. Frequencies of A repeats across prokaryotes show some inconsistent differences. The U homoiterons are too few for meaningful comparisons, and anyway test as different in only a third of the cases.

The ratios of sums of the numbers of G and C to those of A and U homoiterons in the examined groups (Fig. 2) show a striking difference between thermophilic and non-thermophilic species in both prokaryotic large RNAs (Fig. 2, graphs A and B), and especially in the >4u repeats. These ratios reflect the overall GC bias of homoiterons. As seen in Fig. 2, the vertebrate and angiosperm ratios are much above any other eukaryotic LSU values (and especially in the >4u homoiterons), and also much above the vertebrate SSU values (Fig. 1D), which also applies to other eukaryotic LSU compared to the respective SSU RNAs. For eukaryotic LSU RNAs, the ratios are in all groups much lower than those for the corresponding LSU RNAs, and also do not increase as dramatically in vertebrate RNAs (Fig. 2B).

In view of prior findings about contiguous purine and pyrimidine nucleotide tracts in some viral DNAs and RNAs [29,30], the flanking triplets about rRNA homoiterons were examined for purine nt content at A and G, and for pyrimidine nt content at C and U repeats. Across the major RNA groups, both triplet flanks average less than 50% purine nt content with A and G, and less than 44% pyrimidine nt content with C and U homoiterons, indicating a low incidence of the respective contiguous tracts (Supplementary Table S6). Differences between and within the prokaryotic and eukaryotic groups are small.

No repeats above 12 units are found in any of the examined RNAs. The 7u repeats are found more than once per sequence only in vertebrate LSU RNAs. The >4u A and U repeats are generally low in both LSU and SSU of the prokaryotic rRNAs examined (Fig. 1). In eukaryotic rRNAs the A and U >4u repeats are present at least than three units per sequence, excepting U in LSU and A in SSU of lower eukarya.

The large imbalance of >4G and >4C repeats in prokaryotic RNAs suggests an open type of structure for the long G homoiterons. This imbalance is small in eukaryotic >4u repeats, and especially in ES tracts (see the section on ES domains and Fig. 4). As apparent from the published rRNA models (see supplements in Refs. [26,31,32]), long homoiterons frequently have incomplete in situ canonical base pairing (with frequent thermodynamically weak G:U pairs).
and some are partly assigned to loops. This could be of importance for interactivity of the long homoiterons that are accumulated in the expansion segments of vertebrate LSU rRNAs.

3.2. Homoiteron content of ribosomal RNAs

Abundance of two-unit repeats within the compared eukaryotic, archaeal and bacterial rRNA groups is quite uniform, with ranges of 27–30.8% of all nucleotides (nt) for LSU, and 29.7–31.9% nt for SSU RNAs (Table S7), and with few significant differences (mostly involving the UU units of the alveolate and insect-2 RNAs). The prokaryote 2u ranges are 26.7–32% for LSU, and 27.1–32% for SSU RNAs (see Table S7). No significant difference in 2-unit repeats was found for any eukaryotic LSU–SSU pairs. Also, no consistent differences in 2u repeats were found within LSU and SSU groups. Repeats with three or more units (which further on will be referred to as homoiterons proper) however show significant differences among many groups of eukaryotic as well as prokaryotic RNAs. This was statistically tested for the RNA sequence fractions occupied by homoiterons (Fig. 3).

Abundance of two-unit repeats within the compared eukaryotic, archaeal and bacterial rRNA groups is quite uniform, with ranges of 27–30.8% of all nucleotides (nt) for LSU, and 29.7–31.9% nt for SSU RNAs (Table S7), and with few significant differences (mostly involving the UU units of the alveolate and insect-2 RNAs). The prokaryote 2u ranges are 26.7–32% for LSU, and 27.1–32% for SSU RNAs (see Table S7). No significant difference in 2-unit repeats was found for any eukaryotic LSU–SSU pairs. Also, no consistent differences in 2u repeats were found within LSU and SSU groups. Repeats with three or more units (which further on will be referred to as homoiterons proper) however show significant differences among many groups of eukaryotic as well as prokaryotic RNAs. This was statistically tested for the RNA sequence fractions occupied by homoiterons (Fig. 3).

Comparison of the abundance of nucleotides in homoiterons of the examined sequences is shown in Fig. 3. The prokaryotic rRNAs have quite similar fractions of LSU and SSU sequences in these repeats, with the LSU/SSU ratio range of 0.95–1.19 across the groups (and also quite similar GC contents, with LSU/SSU ratios of % GC ranging from 0.92 to 0.99; see Table 2). However, rRNAs
of the thermophilic species of archaea and bacteria examined have many more G and C repeats than the respective mesophilic groups (including halophilic archaea and psychrophilic bacteria). In eukarya, cytoplasmic 60S subunit RNAs of mammalian, amphibian and angiosperm groups have >30% larger homoioteron sequence fraction than the corresponding cytoplasmic 40S subunit RNAs, while no other LSU group exceeds the matching SSU by more than 20% in that fraction. As will be shown later, this excess of homoioteron content is mainly related to the LSU ES tracts.

The devolved RNAs of ribosomes of mammalian mitochondria have very few long G and C homoiterons. However, both 16S (the cytoplasmic ribosome 28S RNA-correspondent) and 12S (the cytoplasmic 18S RNA-correspondent) metazoan mitochondrial rRNAs have a high AU, and especially A, content (35.1% and 32.9% A in human 16S and 12S, respectively), and feature multiple internal A homoiterons (>10% of both 16S and 12S human sequences).

Homoiterons in 5.8S, 2S, 5S, 4S and other small ribosomal RNAs could not be compared adequately across phylogenetic ranks, due to the lack of sequence data for many species. However, it is of interest that a helix in all available vertebrate 5.8S RNAs (in four mammalian, one amphibian and three fish species) has matched double repeats CCCCGGG and GGGGCCC, which confer a high stability, about 1.2 kcal/nt pair. Preliminary modeling indicates that double homoiterons in stems are frequent in many vertebrate, but in few invertebrate large LSU rRNAs.

Any evaluations relating to evolution of eukarya should avoid considering rRNAs of the intracellular parasite euglenozoans as products of typical lower eukarya genes. These rRNAs are composed of multiple pieces, some of which could be related to host genomes [33,34].

The LSU/SSU size ratio ranges from 1.88 to 2.08 in prokaryotes, quite similar to 1.80–2.14 in the examined non-vertebrate eukarya (Table 1). This ratio however rises from 2.08 in fish to 2.59 in mammalian group, which could be a tetrapode-specific LSU enlargement. This is obviously linked to the sequence fraction of the LSU expansion segments, which is 24–31% in non-vertebrate eukaryotes, but 30–45% in the vertebrate, with a sharp increase from fish to mammal. No comparable increase is seen for the SSU ES fraction, which ranges from 25% to 28% (32% in euglenozoans).

Prokaryotic LSU rRNAs average 59–71% of length of the mammalian group (Table 1). The non-vertebrate eukaryotic LSU RNA range of 69–81% (86% in euglenozoans) is smaller than the mammalian. The SSU rRNAs of prokaryotes range tightly between 78% and 82% of mammalian SSU size, and eukaryotic SSU rRNAs show 90–100% of that size (excepting the euglenozoan group at 120%). In terms of size evolution, these data indicate an essentially monophasic increase for eukaryotic SSU RNAs, but a biphasic change for vertebrate and insect LSU RNAs.

It is of interest that the homoiteron content of the coding sections of human mRNAs averages 17.8% of sequence nt (Table S4), which is far below 28.5% in human 28S rRNA (Fig. 3 and Table S4) or other tetrapod vertebrate 28S rRNAs (see the Supplementary List of rRNAs, and also Fig. 3), while human mRNA 5’UTR and 3’UTR sections average 21.4% and 25.1% homoiteron nucleotides, respectively (Table S4).

3.3. The GC content and nucleotide bias of ribosomal RNAs

The thermophilic prokaryotes have in both large rRNAs higher GC content than the mesophiles (Table 2). The GC content of the large LSU RNAs is similar to the respective SSU RNAs in prokaryotes and lower eukaryotes, but much higher in most metazoa and in angiosperm plants (Table 2). The additional GC content is mainly related to GC enrichment of the expansion segments, which is low in the lower eukarya, larger in metazoa, and high in angiosperms and especially in tetrapod vertebrates (Table 2). Here it should be stressed that the ES boundaries for the tetrapod vertebrate and yeast RNAs as defined by alignment to human or yeast LSU RNA closely correspond to values from the published models (see the Supplementary Table S3). No similar enrichment is present.
in ES of SSU RNAs. However, the GC content of SSU RNAs increases about 10% in vertebrates compared to the invertebrate metazoans.

The binding of ribosomal proteins to RNA is mainly dependent on backbone phosphates, and may not critically depend on RNA base bias. The very large difference in GC content for ES of Drosophila melanogaster and Anopheles gambiae LSU rRNAs (31% vs. 60%) contrasted with a high similarity of LSU ribosomal proteins of these organisms suggests that the type of nucleotide bias is not critical in terms of the ES interaction with these proteins. Similar would apply to the respective rRNA cores (with 41% and 54% GC). However, as indicated by examination in the RNAbindR program, the overall strength of RNA–protein association could be less in Drosophila ribosomes. Also, there could be important differences in association with intracellular membrane proteins. Sequences of 37 LSU protein pairs show 84% identity between D. melanogaster and A. gambiae, but the basic PCN clusters match only 60% (data not shown). This might reflect different affinity of the two protein sets for RNA structures, as well as affinity changes adapting to structural differences. Experiments with RNA aptamers and with isolated ribosomal proteins should provide more clues about these subjects.

The ES of SSU rRNAs defined from alignment with human 18S rRNA increase by not more than 25% from fungi to primates, and there is no clear difference with cores in GC content or in the infrequent large G and C repeats (Table 4).

### 3.4. A comparison of the expansion segments in eukaryotic LSU and SSU rRNAs

Four of the five largest expansion segments of LSU RNAs, ES7, 15, 27 and 39, show a phylogenetically linked increase in size from lower eukarya to mammals (Table 3). However, the predicted ES9 is similar in size across metazoan RNAs, and shorter in lower eukarya and angiosperms. The predicted ES15 is very short in angiosperms, and ES39 much larger in euglenozoans and tetrapods compared to other eukaryotes. The GC content of most LSU expansion segments rises significantly between lower eukarya and invertebrate metazoans, and is above 70% in vertebrates and

| Table 1 | Size of rRNAs and of the predicted expansion segments in eukaryote rRNAs. |
|---------|----------------------------------------------------------------------|
| **Group** | **Total LSU nt** | **Predicted ES LSU nt** | **Total SSU nt** | **Predicted ES SSU nt** | **ES as % SSU total** | **ES as % LSU total** | **LSU nt as % mammalian** | **SSU nt as % mammalian** |
| Archaea | | | | | | | | |
| Halophile | 2911 | 1470 | 527 | 1.98 | 60.02 | 78.57 |
| Mesophile | 2961 | 1462 | 507 | 2.03 | 61.05 | 78.14 |
| Thermophile | 3117 | 1498 | 508 | 2.08 | 64.27 | 80.06 |
| Acidophile | 3037 | 1491 | 504 | 2.04 | 62.62 | 79.69 |
| Bacteria | | | | | | | | |
| Psychrophile | 2864 | 1524 | 512 | 1.88 | 59.05 | 81.45 |
| Mesophile | 2965 | 1522 | 510 | 1.95 | 61.13 | 81.35 |
| Thermophile | 3456 | 1540 | 516 | 2.24 | 71.26 | 82.31 |

| Vertebrates | | | | | | | | |
| Mammalian | 4850 | 2196 | 1871 | 527 | 2.59 | 45.28 | 28.17 | 100 | 100 |
| Amphibian | 4082 | 1431 | 1825 | 498 | 2.24 | 35.06 | 27.29 | 84.16 | 97.54 |
| Fish | 3711 | 1107 | 1779 | 507 | 2.08 | 29.83 | 28.5 | 76.52 | 95.08 |

| Invertebrates | | | | | | | | |
| Chordate | 3562 | 1012 | 1780 | 488 | 2.00 | 28.41 | 27.42 | 73.44 | 95.14 |
| Mollusk | 3657 | 947 | 1787 | 483 | 1.80 | 28.13 | 27.03 | 66.33 | 95.51 |

| Lower eukarya | | | | | | | | |
| Fungal | 3374 | 836 | 1759 | 471 | 1.92 | 24.78 | 26.78 | 69.57 | 94.01 |
| Euglenozoan | 4193 | 1670 | 2237 | 708 | 1.87 | 39.83 | 31.65 | 81.98 | 106.73 |
| Alveolate | 3348 | 789 | 1997 | 517 | 1.98 | 23.57 | 25.16 | 69.03 | 90.49 |

| Plant | | | | | | | | |
| Angiosperm | 3382 | 807 | 1792 | 487 | 1.89 | 23.86 | 27.18 | 69.73 | 95.78 |

Data are means for the groups examined in Fig. 1. All sequences are at least 95% complete according to CRW RNA database. ES size predictions were done as described in Section 2.

### Table 2 | The percent GC content of LSU and SSU RNAs. |
|---------|-------------------------------------------------|
| **Group** | **Total LSU ES** | **Core LSU ES** | **Total SSU ES** | **Core SSU ES** | **LSU/SSU ES GC%** |
| Archaea | | | | | | |
| Halophile | 56.3 | 57.9 | 0.97 |
| Mesophile | 51.5 | 55.7 | 0.93 |
| Thermophile | 66.0 | 65.6 | 1.01 |
| Acidophile | 62.2 | 63.3 | 0.98 |
| Bacteria | | | | | | |
| Psychrophile | 51.7 | 52.7 | 0.98 |
| Mesophile | 52.5 | 54.2 | 0.97 |
| Thermophile | 60.0 | 63.0 | 0.95 |

| Vertebrates | | | | | | | | |
| Mammalian full | 67.7 | 81.8 | 56.0 | 56.0 | 55.8 | 55.8 | 1.45 |
| Big ape 5’ half | 70.1 | 81.2 | 55.4 | 56.0 | 55.8 | 55.8 | 1.45 |
| Amphibian | 65.4 | 83.4 | 55.8 | 53.8 | 53.8 | 53.8 | 1.55 |
| Fish | 59.8 | 70.9 | 54.9 | 54.5 | 53.2 | 55.0 | 1.33 |

| Invertebrates | | | | | | | | |
| Chordate | 57.2 | 68.0 | 52.8 | 50.9 | 49.6 | 51.3 | 1.37 |
| Mollusk | 53.7 | 61.0 | 50.8 | 49.3 | 45.8 | 50.6 | 1.33 |
| Insect-1 | 54.3 | 59.7 | 51.8 | 49.1 | 45.9 | 50.2 | 1.3 |
| Insect-2 | 40.5 | 32.3 | 44.2 | 42.5 | 36.3 | 44.9 | 0.89 |
| Nematode | 49.0 | 52.5 | 47.7 | 47.0 | 40.8 | 49.1 | 1.29 |
| Sponge | 52.7 | 58.7 | 50.1 | 45.9 | 41.1 | 47.6 | 1.43 |

| Lower eukarya | | | | | | | | |
| Fungal | 49.7 | 54.1 | 48.3 | 46.8 | 41.0 | 48.9 | 1.32 |
| Euglenozoan | 49.7 | 50.1 | 49.5 | 49.9 | 49.3 | 50.2 | 1.02 |
| Alveolate | 44.7 | 43.6 | 45.1 | 42.9 | 36.2 | 45.2 | 1.20 |

| Plants | | | | | | | | |
| Angiosperm | 58.0 | 71.0 | 53.9 | 50.6 | 47.9 | 51.5 | 1.48 |
angiosperms, with a striking rise above 80% in tetrapod vertebrates. On the other hand, both prokaryotic RNAs have a quite random distribution of >3-unit homoi-terons within sequence (data not shown).

The ES of 18S rRNAs do not differ appreciably in size across the eukaryotic phyla, and show not more than 20% increase in GC content in vertebrates over other metazoans, and show not more than 20% increase in GC content across eukaryotes (Table 1) would favor the speculation about an expansion that was largely accomplished at an early evolutionary stage. The vertebrate and angiosperm LSU RNAs compared to other eukaryotes (Table 1) would favor the speculation about an expansion that was largely accomplished at an early evolutionary stage. The vertebrate and angiosperm LSU RNAs compared to other eukaryotic groups show a clear accumulation of G and C homoiterons in the expansion segments, and especially of >4u homoiterons (Fig. 4A and E; also see Table S5). The respective LSU cores however show significant differences only in 3-unit G and C homoiterons (Fig. 4B and F). The SSU RNAs have fewer long G or C homoiterons, and especially in the expansion segments (Fig. 4C and G). The core 4u and >4u SSU G and especially C homoiterons are however generally more represented in vertebrate RNAs (see graphs 4D and 4H, respectively).

The A and U homoiterons in both LSU and SSU RNAs are much better represented in cores than in ES (Fig. 5). The vertebrate LSU groups show very low densities of ES homoiterons of either A (Fig. 5A) or U nt (Fig. 5E). The SSU A and U repeats are also more abundant in cores (Fig. 5D and H).

The expansion segments of 18S rRNAs have not been defined to a consensus, as can be seen by comparing the modeled human and yeats 18S rRNA ES boundaries (Refs. [26,35]; see also Table S1). However, the size similarity of 18S rRNAs across eukaryotes (Table 1) would favor the speculation about an expansion that was largely accomplished at an early evolutionary stage. The expansion segments of eukaryotic 18S RNAs seem to have developed to a similar size in unicellular heterotrophs, mammals and angiosperm plants (Table 4). Excepting the intracellular parasite euglenozoans, the predicted sequence expansion compared to bacterial or archaean 16S rRNAs does not exceed 25% (Table 1). The 18S rRNA expansion segments also generally show no clear difference with the core sections in nucleotide composition, as is summarized in Table 4 (see also the models of human [26] and yeast [35] 18S rRNAs).

Much of the ES sequences in SSU rRNAs could be involved in handling eukaryotic-only ribosomal proteins, and there also is critical association with initiation/elongation factors [36,37] and mRNAs [38], but there is no evidence that these SSU sectors importantly affect ribosome association with membrane systems or translocons.

### 3.5. GC content above sequence co-localizes with ES in tetrapod LSU but not SSU RNAs

An assay estimating GC content of sequence “windows” of e.g. 1/100 sequence length with or without subtraction of the GC content baselines could be useful in distinguishing GC distribution among sequences, and assessing association of window segments

### Table 3

| Group            | #nt   | Core GC% | Core % total |
|------------------|-------|----------|--------------|
| Mammalian [6]    | 744 ± 29.3 | 105 ± 0.43 | 56.1 ± 0.1   |
| Amphibian [1]    | 442   | 124      | 34.6         |
| Fish [6]         | 404 ± 19.6 | 98.4 ± 1.08 | 30.7 ± 0.86  |
| Chordate [2]     | 325 ± 9  | 96 ± 2   | 28.5 ± 0.55  |
| Mollusk [1]      | 310   | 95       | 28.6         |
| Insect-1 [5]     | 404 ± 19.6 | 98.4 ± 1.08 | 31.6 ± 0.69  |
| Insect-2 [3]     | 300 ± 14.2 | 109 ± 5.81 | 30.2 ± 1.5   |
| Nematode [1]     | 213   | 101      | 27.1         |
| Sponge [3]       | 361 ± 3.7 | 96.3 ± 4.67 | 30.9 ± 0.95  |
| Fungal [7]       | 194 ± 4  | 67.1 ± 1.86 | 25.0 ± 0.35  |
| Euglenozoan [3]  | 350 ± 44.7 | 82.7 ± 1.2 | 25.0 ± 0.35  |
| Alveolate [2]    | 204 ± 1  | 69.5 ± 0.5 | 39.8 ± 0.17  |
| Angiosperm [4]   | 179 ± 3.5 | 65.3 ± 1.18 | 24.4 ± 0.39  |

* Only three sequences (human, mouse and rat) are available for ES27 and ES39 in this group.

** For pools of nucleotides from all 15 ES tracts. The group members are shown in the list at the beginning of Supplementary Data. Data for numbers of nucleotides (#nt) and GC% are means with standard errors. The number of sequences analyzed is shown in brackets after the group labels. The segment boundaries were defined from alignment to those of human 28S rRNA (which were adopted from Ref. [26]).
with features such as the expansion segments. In confirmation of findings with >3u homoiterons, we find that the human LSU GC content above average sequence % GC taken as baseline co-localizes quite precisely with the expansion segments (Fig. 6A). Similar distribution profiles are observed with other tetrapod vertebrate LSU RNAs, and to a lesser degree also with fish

### Table 4

Size and GC content in four large expansion segments of eukaryotic SSU 17–20S cytoplasmic-ribosome RNAs.

| Group       | ES3 #nt | ES4 GC% | ES6 Core GC% | ES12 ES % total |
|-------------|---------|---------|-------------|-----------------|
| Mammal [3]  | 83.7 ± 0.33 | 72      | 175         | 64              |
| Amphibian [1]| 61      | 72      | 173         | 63              |
| Fish [5]     | 63.6 ± 3.6 | 72 ± 0.2| 177 ± 2.2   | 63.2 ± 0.2      |
| Chordate [2] | 58      | 70.5 ± 1.5 | 171 ± 1.5  | 62.5 ± 0.5      |
| Mollusk [1]  | 55      | 72      | 173         | 63              |
| Insect-1 [5] | 59.8 ± 3.7 | 72 ± 8.8| 203 ± 9.5   | 65 ± 2.4        |
| Insect-2 [3] | 55 ± 1.53 | 74      | 208 ± 8.7   | 48 ± 15         |
| Nematode [1] | 51      | 71      | 169 ± 1     | 58.5 ± 0.5      |
| Sponge [2]   | 55 ± 1  | 71      | 169 ± 1     | 58.5 ± 0.5      |
| Fungal [6]   | 51.45   | 70.8 ± 3.2 | 166 ± 3    | 59.2 ± 0.31     |
| Euglenozoon [3] | 84 ± 2.29 | 74.3 ± 1.20 | 357 ± 25.1 | 38.0 ± 6.0      |
| Alveolate [2] | 52      | 71.5 ± 0.5 | 136 ± 3    | 41              |
| Angiosperm [4]| 55.7 ± 0.25 | 71.5 ± 0.5 | 173 ± 1    | 61.5 ± 0.5      |

**Fig. 5.** The >2-unit homoiterons of adenosine and uridine nucleotides in expansion segments (ES) and cores of eukaryotic ribosomal RNAs as percent of sequence nucleotides. Upper row: Adenosine nucleotides. (A) ES, (B) cores of 25–28S RNAs of the 60S subunit; (C) ES, (D) cores of 17–20S RNAs of the 40S subunit. Lower row: Uridine nucleotides. (E) ES, (F) cores of 25–28S RNAs of the 60S subunit; (G) ES, (H) cores of 17–20S RNAs of the 40S subunit. For assignation of significance in the Schefé t tests see the caption of Fig. 1.
LSU RNAs. This is however largely not found for the corresponding human SSU GC content (Fig. 6B), or that of other eukaryotic SSU RNAs. The archaeal and bacterial LSU and SSU rRNAs have quite random profiles of the excess % GC distribution, with multiple evenly spread peaks (data not shown).

3.6. Free energy estimates suggest large differences in stability of the secondary structure between vertebrate and other eukaryote LSU RNAs

Stability of the secondary structure in polyribonucleotides would significantly connect to GC content, and could also relate to G and C homoiteron content. In this connection, the predicted free energies of folding/unfolding expressed per 100 nt for LSU RNAs are indeed significantly higher in vertebrates compared to other eukarya (Fig. 7A). The non-vertebrate LSU RNAs show a moderate trend of increase from alveolate to chordate. The prokaryote thermophile LSU RNAs have clearly larger relative trend of increase from alveolate to chordate. The prokaryote LSU RNAs also have very few homoiterons (Fig. 1–3) and in eukaryote SSU RNAs (Fig. 7B), with the LSU/SSU ratios significantly above the mesophile, but not significantly, and no differences are found among the bacterial groups (Fig. 7C).

3.7. RNA modeling points to considerable presence of long homoiteron-rich double-stranded segments especially in vertebrate LSU RNAs

The RNA-binding proteins that prefer double-stranded RNA segments (stems), in particular the stau proteins, are significantly involved in transport, activity and regulation of mRNAs [39,40], copiously localize to subcellular networks including the ER [41], and are found in granules containing mRNAs and ribosomes [40,42]. Association with RNAs could prefer stems with about 12 nucleotide pairs and containing multiple homoiterons [43], although shorter stems can be used as well [44]. Such segments could be frequent in homoiteron-rich rRNAs.

RNA modeling defines up to 12% rRNA sequence as stems with 10 or more nucleotide pairs (see Fig. 8A and B for LSU and SSU RNAs, respectively). Models of the mammalian LSU RNAs show significantly more such segments than models of any other eukaryote rRNAs (Fig. 8A and B). Eukaryote SSU RNAs however in most groups have less than 10% sequence in predicted long stems (Fig. 8B). The predicted long stems are also frequent in prokaryote models, and especially in mesophile species.

Abundance of homoiterons in >9-pair stems of the models in LSU RNAs is again the largest in mammalian LSU RNAs (close to 8% of the entire sequence; Fig. 8C), and tetrapod vertebrate LSU RNAs clearly out-run all other eukaryote LSU RNAs in this content. The archaeal mesophile and halophile LSU RNAs have more homoiterons in the predicted long stems than the thermophiles (Fig. 8C). Eukaryote SSU RNA models however show quite uniformly only 3–4% sequence in >9-stem homoiterons (Fig. 8D). Both LSU and bacterial SSU RNAs have relatively low homoiteron content in long stems. However, both archaeal halophile RNAs have above 6% sequence nucleotides in these homoiterons (Fig. 8C and D).

The long stems of mammalian and halophile LSU RNAs have more than 60% nucleotides in form of homoiterons (Fig. 8E). This is also predicted for halophile SSU RNAs (Fig. 8F). Other eukaryote LSU RNAs show below 50% of homoiteron content in the predicted stems, and most eukaryote SSU RNAs below 40% (Fig. 8F). The G and C homoiterons are much more frequent than A and U homoiterons in the predicted stems of most groups (Fig. 8C and D).

4. Discussion

Same-nucleotide repeats currently seem not to be perceived as a distinct category of sequence motifs. This is appropriate in the case of nucleotide doublets, which are present in similar large proportions across at least the ribosomal RNAs (as shown in this work). Also, these doublets per se cannot form stable secondary structure. Three-unit G and C and all >3u homoiterons are independently able to form stable stems.

The evolution of eukaryotic rRNAs proceeded via insertions in ancestral prokaryotic sequences [21–23,25,26,35,45]. As suggested by expansion profiles in lower eukaryotes, these additions may have started with little or no nucleotide bias, possibly accommodating changes in ribosomal proteins mainly via increases in length of the inserts. Biased GC or AU expansion obviously developed in both plants and metazoa, with GC enrichment apparently being preferred. This strikingly contrasts the GC loss accompanying size reduction of mitochoneral rRNAs in metazoa [46]. A high cytoplasmic LSU GC enrichment evolved in vertebrates, and especially in tetrapods. The parallel enlargement of basic clusters in eukaryotic ribosomal proteins may have reached saturation already in the invertebrate metazoa, preceding the large GC bias of LSU rRNA expansion segments, and the massive enlargement of these segments in the vertebrate [24]. However, both the largest basic
PCN clustering in LSU proteins and the largest accumulation of G and C homoiterons in LSU 28S rRNAs are found in the tetrapod vertebrates, and especially in mammals [24].

The low iteron content in human mRNAs (Table S4) should connect to degeneracy of the genetic code, as well as to infrequent occurrence of long amino acid homoiterons in proteins. The
considerably higher homoiteron content of human 28S rRNA and 3’UTR sections of mRNAs compared to mRNA coding sections, and even to 18S rRNA (Figs. 1–3 and Table S4), might also point to interactive importance of homoiterons in the former two.

The highly conserved intra-sequence type 1 homoiterons may serve both stabilizing and shaping/chaperoning or guiding functions in association of rRNAs with proteins and other RNAs. The stabilizing function appears self-evident in thermostophilic prokaryote rRNAs, where it obviously is not specific for either subunit. Currently there is no experimental evidence for stabilizing, chaperoning or guiding roles of homoiterons in vertebrate rRNAs. However, the very large difference between 28S and 18S rRNAs in content and localization of >3-unit homoiterons and the virtual confinement of big homoiterons in tetrapod 28S RNAs to the expansion segments indicate 60S subunit-specific functions. The two largest ESs of 28S rRNAs are importantly involved in formation of the ribosomal export tunnel, and in association with initiation factors. Both export and translocation of newly synthesized proteins could use chaperoning help by the repeats, linked to dynamic low-affinity associations. Association with the ER (which is definitely much stronger and more complex in vertebrates relative to lower eukarya and bacteria; compare Refs. [47–49]) could use the repeats for chaperoning as well as for specific binding.

The ER-attached 60S subunits constitute more than a half of all LSU in the rat (and likely also in other mammals). The “free” polyribosomes (polysomes) probably also are at least loosely anchored to intracellular networks (including the ER and cytoskeleton [50,51]) and components of LSU RNAs would interact with anchorages. The sizable parts of the large ES that show no association with proteins in free ribosomes examined by freeze electron microscopy [26,45] could conceivably be involved in these interactions. This also seems to be supported by patching of RNase-treated ER-bound ribosomes [52]. The need to use both puromycin and high salt for detachment of ribosomes from ER (Ref. [47] and our observations) is also compatible with involvement of ES homoiterons in the anchoring. The anchoring could be helped by GC bias of the ES homoiterons. The G and C units may focus respectively on ionic and hydrophobic clusters in the anchorages, and U units (enriched in AU-biased insect-2 and euglenozoan LSU RNAs) could prefer hydrophobic targets, possibly including proline-rich proteins [19].

The minimalized RNAs of ribosomes of mammalian mitochondria (which produce only non-exported inner mitochondrial membrane proteins [53]) have very few long G and C homoiterons, and this indirectly supports use of GC-rich homoiterons on cytoplasmic ribosomes for association with matrices. The multiple A homoiterons in LSU and SSU RNAs of human mitochondria (constituting about 10% of both the 16S and the 12S sequence) obviously are not involved in protein export and should serve mainly in organization of ribosomal proteins.

The increase in size of expansion segments together with segregation of >4u homoiterons into these segments in vertebrate cytoplasmic 28S RNAs could be linked to fashioning of the export channel/tunnel as well as to interactions with ER proteins. Also, new ribosome receptors might have evolved between fish and mammal. The export tunnel-associated (and thus also ER-proximal) ES27 [26] is more than doubled in size in the mammal compared to the fish (Table 3), and this could relate to interaction with novel partners. The ES7 is enlarged in the mammal by more than 65% relative to any other eukaryotic group (Table 3), and this should relate to interactions in involving both mRNAs and initiation proteins, as well as other thus far unidentified major partners. The initiation complex (or rather the initiation particle) in the mammal has a number of protein components (see e.g. Ref. [54]), some of which could interact with the large ES of the LSU 28S RNA.

Interaction with RNA-binding protein sequences could be an important role of RNA homoiterons, and especially of long G and C homoiterons in ES segments. The particulate translation initiation factors [55], SRPs [56], p180 [48] and staufen proteins [40,57] could be among the partners. Homoiterons in the long stems (Fig. 8) might be significantly used in these interactions.

A note of caution about secondary structure predictions of rRNAs, and in particular those for the expansion segments, would be that many proposed features have low stability at 37° [58,59], and many short homoiterons that are predicted as stem parts could have considerable physiological single-stranded reactivity (as also is suggested by thermodynamic estimates in RNAstructure and RNAfold programs).

In modeling studies with free ribosome crystals, the large ES7 and ES27 tracts of LSU rRNAs are invariably found to significantly lack stable association with ribosomal proteins [26,35,45,60]. These findings point to physiological availability for interaction with external partners. At least in the mammalian 28S rRNAs, these tracts ought to include parts of the large ES, and should in situ engage both ribosomal and external partners.

Differential expansion of vertebrate and angiosperm LSU and SSU RNAs could largely reflect a functional adaptation of LSU to lower mobility and preferential association with subcellular matrices. This obviously evolved along with phylogenetic complexity, being much less prominent in the non-vertebrate chordates, and most expressed in primates (Table 3) along with a quadrupling in size of the expansion complement (Figs. 1 and 3 and Table 3). This jump has to reflect an increased utility of such a transformation. The benefit could be linked mainly to chaperoning. The chaperoning could include especially non-ribosomal RNAs [61], possibly in mRNPs associating with the ER. Polyguanylate is a good competitor of mRNA complexes with proteins [62] and may participate in activation of mRNP assemblies at the ER [63,64]. Long G homoiterons especially in ES7 may conceivably also be used for such activation, in conjunction with activity of the initiation/elongation factors. The mRNPs containing staufen and other RNA receptors (e.g. RRPBP1/p180 (Q9P299, Q28298)) are mainly located in cytoplasmic membrane elements [65,66], and could interact with ES7 and ES27.

ES27 typically shows low definition in crystallization units [26,31,32,45], which indicates that much of the segment is not stably associated with ribosomal proteins, or with other parts of ribosomal RNAs. ES27 is known to be indispensable in the mammal [67], and ES7 (V3) in yeast [68]. ES27 could be the principal agent in ribosome communication with the translocation machinery [69]. In the mammalian ribosome, ES27 is largely disordered [26], and in fungi also significantly dynamic (but four times smaller, Table 3). ES27 interacts with ES4 and with ES5 and ES6 in the free mammalian ribosome [26]. This however may differ in the 60S subunit associated with the ER membrane.

In eukaryotes the abundance of RNA homoiterons follows increase in size and nucleotide bias of the expansion segments, and increases with phylogenetic rank. Both the nucleotide bias and the frequency of homoiterons are much larger in metazoan and plant LSU rRNAs compared to the respective SSU rRNAs. This is pronounced in tetrapod vertebrates and appears to culminate in hominid mammals. The massive change in homoiteron content and nucleotide bias encountered across the eukaryote evolutionary ladder would argue for an as yet uncharacterized differentiation of protein production and export.

Conflict of interest

The authors declare no conflict of interest related to this study.
Acknowledgments

This work was partly supported by the U.S. National Institutes of Health Grants HD-20074 and HD-13703. All authors performed literature searches and data derivation and calculation, and participated in interpretations and discussion of the results. S.L.P wrote the final manuscript. S.L.P. also gratefully acknowledges a Grant-in-aid from Mr. William E. Gentry of Collierville, Tennessee, USA.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.fob.2015.10.005.

References

[1] Zahn, K. and Blattner, F.R. (1985) Binding and bending of the lambda.
[2] Konieczny, I., Bury, K., Wawrzycka, A. and Wegrzyn, K. (2014) Iteron plasmids.
[3] Konieczny, I., Bury, K., Wawrzycka, A. and Wegrzyn, K. (2014) Iteron plasmids.
[4] Frey, S., Pool, M. and Seedorf, M. (2001) Scp160p, an RNA-binding, polysome-
[5] Frey, S., Pool, M. and Seedorf, M. (2001) Scp160p, an RNA-binding, polysome-
[6] Konieczny, I., Bury, K., Wawrzycka, A. and Wegrzyn, K. (2014) Iteron plasmids.
[7] Chappell, S.A., Edelman, G.M. and Mauro, V.P. (2004) Biochemical and
[8] Zambetti, G., Wilming, L., Fey, E.G., Penman, S., Stein, J. and Stein, G. (1990)
[9] Marion, R.M., Fortes, P., Beloso, A., Dotti, C. and Ortin, J. (1999) A human
[10] Morrow, M.W. and Brodsky, J.L. (2001) Yeast ribosomes bind to highly purified
[11] Mathews, D.H., Turner, D.H. and Zuker, M. (2007) RNA secondary structure
[12] Konieczny, I., Bury, K., Wawrzycka, A. and Wegrzyn, K. (2014) Iteron plasmids.
[13] Michot, B., Hassouna, N. and Bachereller, J.P. (1984) Secondary structure of mouse
[14] Parker, M.S., Sah, R., Balasubramaniam, A., Park, E.A., Salle, F.R. and Parker, S. L. (2014) On the expansion of ribosomal proteins and RNAs in eukaryotes.
[15] Amino Acids 46, 1589–1604.
Ojakian, G.K., Kreibich, G. and Sabatini, D.D. (1977) Mobility of ribosomes bound to microsomal membranes. A freeze-etch and thin-section electron microscope study of the structure and fluidity of the rough endoplasmic reticulum. J. Cell Biol. 72, 530–551.

von Heijne, G. (1986) Why mitochondria need a genome. FEBS Lett. 198, 1–4.

ElAntak, L., Tzakos, A.G., Locker, N. and Lukavsky, P.J. (2007) Structure of eIF3b RNA recognition motif and its interaction with eIF3j: structural insights into the recruitment of eIF3b to the 40 S ribosomal subunit. J. Biol. Chem. 282, 8165–8174.

Hiraihi, H., Shin, B.S., Udagawa, T., Nemoto, N., Chowdhury, W., Graham, J., Cox, C., Reid, M., Brown, S.J. and Asano, K. (2013) Interaction between 25S rRNA A loop and eukaryotic translation initiation factor 5B promotes subunit joining and ensures stringent AUG selection. Mol. Cell. Biol. 33, 3540–3548.

Miller, J.D., Wilhelm, H., Gierasch, L., Gilmore, R. and Walter, P. (1993) GTP binding and hydrolysis by the signal recognition particle during initiation of protein translocation. Nature 366, 351–354.

Duchaine, T.F., Hemraj, I., Furic, L., Deitinghoff, A., Kiebler, M.A. and DesGroseillers, L. (2002) Staufen2 isoforms localize to the somatodendritic domain of neurons and interact with different organelles. J. Cell Sci. 115, 3285–3295.

Pattanayak, V., Gifford, L.K., Lu, P. and Gewirtz, A.M. (2008) Observed versus predicted structure of fluorescent self-quenching reporter molecules (SQRM): caveats with respect to the use of “stem-loop” oligonucleotides as probes for mRNA folding. RNA 14, 657–665.

Fabian, M.R. and White, K.A. (2008) Solution structure probing of RNA structures. Methods Mol. Biol. 451, 243–250.

Armache, J.P., Jarasch, A., Anger, A.M., Villa, E., Becker, T., Bhushan, S., Jossinet, F., Habeck, M., Dindar, G., Frankenberg, S., Marquez, V., Mielke, T., Thomm, M., Berninghausen, O., Beutler, B., Soding, J., Westhof, E., Wilson, D.N. and Beckmann, R. (2010) Localization of eukaryote-specific ribosomal proteins in a 5.5-A cryo-EM map of the 80S eukaryotic ribosome. Proc. Natl. Acad. Sci. U.S.A. 107, 19754–19759.

Sweeney, R., Fan, Q. and Yao, M.C. (1994) Anti-ribosome RNAs: a vehicle for antisense RNAs. Proc. Natl. Acad. Sci. U.S.A. 93, 8518–8523.

Barrioux, A. and Rosenfeld, M.G. (1977) Comparison of mRNA binding by Met-tRNAf binding protein and mRNA-associated proteins. J. Biol. Chem. 252, 392–398.

Wilhelm, J.E., Buszczak, M. and Sayles, S. (2005) Efficient protein trafficking requires trailer hitch, a component of a ribonucleoprotein complex localized to the ER in Drosophila. Dev. Cell 9, 675–685.

Beckmann, C., Hilliker, A., Cziko, A.M., Noursey, A., Ramaswami, M. and Parker, R. (2008) The DEAD-box RNA helicase Ded1p affects and accumulates in Saccharomyces cerevisiae P-bodies. Mol. Biol. Cell 19, 984–993.

Kiebler, M.A., Hemraj, I., Verkade, P., Kohrmann, M., Fortes, P., Marion, R.M., Ortin, J. and Dotti, C.G. (1999) The mammalian staufen protein localizes to the somatodendritic domain of cultured hippocampal neurons: implications for its involvement in mRNA transport. J. Neurosci. 19, 288–297.

Ohashi, S., Koike, K., Omori, A., Ichinose, S., Ohara, S., Kobayashi, S., Sato, T.A. and Anzai, K. (2002) Identification of mRNA/protein (mRNP) complexes containing Puralpha, mStaufen, fragile X protein, and myosin Va and their association with rough endoplasmic reticulum equipped with a kinesin motor. J. Biol. Chem. 277, 37804–37810.

Sweeney, R., Chen, L. and Yao, M.C. (1994) An rRNA variable region has an evolutionarily conserved essential role despite sequence divergence. Mol. Cell. Biol. 14, 4203–4215.

Jeunega, R.E., Van Deflt, Y., de Graaff-Vincent, M., Dirks-Mulder, A., Venema, J. and Raue, H.A. (1997) Variable regions V13 and V3 of Saccharomyces cerevisiae contain structural features essential for normal biogenesis and stability of 5.8S and 25S rRNA. RNA 3, 476–488.

Beckmann, R., Spahn, C.M., Esvar, N., Helmers, J., Penczek, P.A., Sali, A., Frank, J. and Blobel, G. (2001) Architecture of the protein-conducting channel associated with the translating 80S ribosome. Cell 107, 361–372.

Amaral, P.P., Clark, M.B., Gascoigne, D.K., Dinger, M.E. and Mattick, J.S. (2011) LnRnAdb: a reference database for long noncoding RNAs. Nucleic Acids Res. 39, D146–D151.

Terrorihini, M., Sander, J.D., Lee, J.H., Zaback, P., Jernigan, R.L., Honavar, V. and Dobb, D. (2007) RNABindR: a server for analyzing and predicting RNA-binding sites in proteins. Nucleic Acids Res. 35, W578–W584.