Amino Acids (2014) 46:1589–1604
DOI 10.1007/s00726-014-1704-4

REVIEW ARTICLE

On the expansion of ribosomal proteins and RNAs in eukaryotes

Michael S. Parker · Renu Sah · Ambikaipakan Balasubramaniam · Floyd R. Sallee · Edwards A. Park · Steven L. Parker

Received: 15 February 2014 / Accepted: 18 February 2014 / Published online: 15 March 2014

Abstract While the ribosome constitution is similar in all biota, there is a considerable increase in size of both ribosomal proteins (RPs) and RNAs in eukaryotes as compared to archaea and bacteria. This is pronounced in the large (60S) ribosomal subunit (LSU). In addition to enlargement (apparently maximized already in lower eukarya), the RP changes include increases in fraction, segregation and clustering of basic residues, and decrease in hydrophobicity. The acidic fraction is lower in eukaryote as compared to prokaryote RPs. In all eukaryote groups tested, the LSU RPs have significantly higher content of basic residues and homobasic segments than the SSU RPs. The vertebrate LSU RPs have much higher sequestration of basic residues than those of bacteria, archaea and even of the lower eukarya. The basic clusters are highly aligned in the vertebrate, but less in the lower eukarya, and only within families in archaea and bacteria. Increase in the basicity of RPs, besides helping transport to the nucleus, should promote stability of the assembled ribosome as well as the association with translocons and other intracellular matrix proteins. The size and GC nucleotide bias of the expansion segments of large LSU rRNAs also culminate in the vertebrate, and should support ribosome association with the endoplasmic reticulum and other intracellular networks. However, the expansion and nucleotide bias of eukaryote LSU rRNAs do not clearly correlate with changes in ionic parameters of LSU ribosomal proteins.

Keywords Basic cluster · Homoionic segregation · Ionic sequestration · Protein sequence alignment · RNA expansion segment · Nucleotide bias

Abbreviations

aa	Amino acid(s)
nt	Nucleotide(s)
EO	Ribosomal proteins found only in eukaryotes
EA	Ribosomal proteins aligning between eukaryotes and archaea
EAB	Ribosomal proteins aligning among eukaryotes, archaea and bacteria
ES	Expansion segment(s) of ribosomal RNAs
LSU	The large cytoplasmic ribosome subunit (50S in bacteria and archaea, 60S in eukaryotes)
SSU	The small cytoplasmic ribosome subunit (30S in bacteria and archaea, 40S in eukaryotes)
sec	A translocon component
srp	A signal recognition particle component
PCN	Homoionic cluster with >2 and ≥50 % ionic residues

Electronic supplementary material The online version of this article (doi:10.1007/s00726-014-1704-4) contains supplementary material, which is available to authorized users.

M. S. Parker
Department of Microbiology and Molecular Cell Sciences, University of Memphis, Memphis, TN 38152, USA

R. Sah · F. R. Sallee
Department of Psychiatry, University of Cincinnati College of Medicine, Cincinnati, OH 45263, USA

A. Balasubramaniam
Department of Surgery, University of Cincinnati College of Medicine, Cincinnati, OH 45263, USA

E. A. Park · S. L. Parker (✉)
Department of Pharmacology, University of Tennessee Health Sciences Center, Memphis, TN 38163, USA
e-mail: stevenleonardparker@msn.com

Introduction

Ribosomes are responsible for protein synthesis and have analogous structure and constitution in cytoplasm of all cell types (Alberts et al. 2002), as well as in organelles such as mitochondria (Christian and Spremulli 2012) and chloroplasts (Harris et al. 1994). The contemporary ribosomes represent the end product of a primordial development of protein synthesis apparatus, possibly predating, adapting to, and developing with the DNA-based genetic information system. The remarkable poly-enzymatic activity of ribosomes (Whitesides 2001) is based on a small number of specific non-coding RNA components (acting as protein organizers and supporters of energy generation), and on multiple protein species with energy-releasing, anchoring/structuring and dynamic/micromechanical functions.

Three cytoplasmic ribosomal RNA (rRNA) species are found in prokaryotes. These include the 16S rRNA [1,400–1,600 nucleotides (nt)] of the smaller ribosomal subunit, and the 5S (~120 nt) and 23S (2,900–3,100 nt) rRNAs of the larger subunit. In eukaryotes, three RNAs match the prokaryote rRNA complement [see (Clark et al. 1984; Prinz et al. 2000)]. These rRNAs, while retaining a core structure similar to prokaryote rRNAs (Cannone et al. 2002; Chandramouli et al. 2008) are larger than the respective non-eukaryote molecules. The increase consists of insertions within the core ancestral sequences [as exemplified by those of *Escherichia coli* rRNAs (Gutell et al. 1994)], or of archaea such as *Methanopyrus kandleri*). One should note that expansion segments (ES) are encountered in non-eukaryote rRNAs as well (Greber et al. 2012; Shasmal and Sengupta 2012). The ESs in most eukaryotes add about 400 nt to RNA of the smaller subunit (which is usually labeled as 18S rRNA), and about 40 nt to 5S rRNA to produce the 5.8S RNA of the large subunit. The 25–28S ribosomal RNAs, which assemble the larger ribosomal subunit, are expanded differentially across eukaryote kingdoms and phyla, by about 500 nucleotides in plants and fungi, and by as many as 2,400 in vertebrates.

Ribosomal proteins in eukarya differ considerably from archaeal or bacterial (e.g., (Klinge et al. 2011)]. The RPs of eukaryote groups also show some differences, which have not been studied in detail due to insufficient availability of verified sequences. This has, however, improved considerably over the last few years, permitting direct comparisons of the entire sets of RPs for many phylogenetically distant species. Of a particular interest are the extent of homoionic segregation and abundance of extended homoionic motifs (Parker et al. 2012). The present survey finds that the PCN (polynucleotide-binding, protein carrier-associating, nuclear localization-intensive) ionic clusters of ribosomal proteins are highly aligned in the entire vertebrate phylum, but less in other metazoan phyla, much less in lower eukarya, and only within families in archaea and bacteria. This has an evolutionary interest, and also medical aspects, e.g., related to the vertebrate tolerance of antibiotics that affect the bacterial ribosome.

An increase in basicity of eukaryote ribosomal proteins could be anticipated in relation to multiple factors. The ribosomal proteins in nucleated cells need to travel en masse to the nucleolar sites of ribosome assembly using counterionic carriers (Rout et al. 1997) and this depends selectively on the size and sequence of basic aa clusters (Plafker and Macara 2002; Tai et al. 2013). Increases in cell size and compartmentalization augment and extend contact with ribonucleases, enzymes that ubiquitously depend on basic clusters in attachment to, and even in action upon substrate RNAs [thus in sarcins (Plantinga et al. 2011) and exosomal nucleases (Liu et al. 2006)].

Ribosomes in all biota associate with intracellular membrane systems, especially at the stage of translocation of the newly synthesized protein [see (Herskovits and Bibi 2000; Kuhn et al. 2011) for eubacterial, (Sabatini et al. 1966; Kruppa and Sabatini 1977; Menetret et al. 2008) for mammalian, (Osborne et al. 2005) for pushing and ratcheting modes of ribosomal protein export, and Moore 2009 for a general review]. This is linked to both release of the newly synthesized, "nascent" protein from the ribosome's elongation/export tunnel (Nissen et al. 2000) and to a triage that differentiates the newly synthesized proteins into, roughly, ER membrane resident, ER lumen imported, and cytosol released. Connection to the membrane should also matter in the resetting (including dissociation) of the ribosome for a new cycle of protein synthesis, and in association with messenger ribonucleoproteins (mRNPs) and mRNAs (Lande et al. 1975; Adesnik and Maschio 1981; Ohashi et al. 2002; Mallardo et al. 2003; Stephens et al. 2005; Pyhtila et al. 2008; Cui et al. 2011). This connection might be enhanced by expansion of basic motifs, and especially in metazoa.

In agreement with the above predictions, this survey finds a large increase in cationic parameters of LSU ribosomal proteins in metazoa, with lesser such increases in plants and lower eukaryotes. These parameters do not show a clear relation to the expansion of ribosomal RNAs.

Materials

Ribosomal proteins and RNAs examined

The comparisons included non-acidic LSU and SSU RPs from ten metazoan species (five mammals, three non-mammalian vertebrates, and two insects), three angiosperm

plants, three lower eukaryotes, 12 species of bacteria and 12 of archaea (see the supplementary tables S1 (LSU) and S2 (SSU) for the listings of species, type labels and access codes). The choice of species was restricted by the need to have confirmed sequences of both LSU and SSU rRNAs and of as complete as possible sets of ribosomal proteins. Some RP orthologs are not detected in all species of a biotic domain, which could be due to deletion, or conversion to pseudogenes. For LSU RPs the included ortholog count per species is 31–39 in prokaryotes, and 41–49 in eukaryotes. For SSU RPs, the count per species is 20–21 in bacteria, 23–27 in archaea, and 26–33 in eukaryotes (with 19 in *Giardia lamblia*, which was not statistically evaluated). In view of gene deletion and inactivation it was felt that for proper global comparisons the use of RP sets lacking certain orthologs is more justified than limiting all sets to only the orthologs described as active in all species. The acidic ribosome-attaching proteins and the initiation/elongation factors across species vary considerably in the currently recognized types and could not be adequately compared.

In choosing the taxonomic species an effort was made to include different major taxonomic and other groups, as well as species with different rRNA size and composition. Thus, the archaeal species are from phyla of crenarchaeota and euriarchaeota and include the habitat groups such as halophiles, mesophiles, thermophiles and hyperthermophiles. The bacterial species are from phyla of proteobacteria, firmicutes, actinobacteria and spirochetes, and include non-pathogenic and pathogenic, Gram-positive and Gram-negative, enteric, lung- and epidermis-preferring species. The two insect species differ radically in GC content of the large LSU rRNAs and in the nucleotide bias (Klukas and Dawid 1976; Escobar et al. 2011) of the expansion segments. The lower eukaryotes include an alveolate ciliophore and a fungal ascomycete with similar and small rRNAs, and a euglenozoan with much larger rRNAs and expansion segments. A diplomonad with short ribosomal RNAs lacking ESs was also examined.

As seen in the supplementary tables S3A.1–7, accrual of the ionic parameters of LSU proteins generally indicates a good overall semblance of LSU RP sequences within the examined groups of species. The most variability is encountered in archaea (Table S3A.1). The two vertebrate groups [the three non-mammalian (Table S3A.5) and especially the five mammalian LSU RP sets (Table S3A.6)] show an excellent stability in accrual of the parameters, and similar seems to apply even to the taxonomically heterogeneous group of lower eukaryote heterotrophs (Table S3A.3). The within-group stability of the bulk acidic and basic residue content is high in all LSU groups. A high accrual stability is also found for the SSU groups (data not shown).

Methods

Sequence alignments

SSEARCH3 program (Pearson 2000) was used for sequence alignments. The alignments in ClustalW program (Thompson et al. 1994) corresponded well with SSEARCH3 values. However, the SSEARCH3 output features individual identity scores, enabling more detailed comparisons. Sets of the RP orthologs retrieved from Entrez or UniProtein databases were aligned with the corresponding human proteins, which in metazoa and plants have similar consensus type labels. Orthologs of the lower eukarya, in some cases not sharing the consensus labels with human proteins, were identified based on data from Armache et al. (2010a, b), Ben-Shem et al. (2011) and Klinge et al. (2011) after confirmation in SSEARCH3. The archaeal and bacterial quasi-orthologs of human proteins are the LSU or SSU RP sequences (from the respective 12 species in either group) that align best with the human orthologs. The access codes and type labels are listed in the supplementary tables S1, S2 and S4.

Homoionic motifs

Segments with two or more ionic sidechains of the same type of charge were scored as previously described (Parker et al. 2012). Briefly, these are contiguous sequence segments containing two or more ionizable residues of the same type of potential charge in the physiological [H^+] range (Asp and Glu for the homoacidic, His, Lys and Arg for the homobasic segments) and starting and ending by such residues. These segments provide a measure of the homoionic sequestration in protein sequences.

Based on similarity with nuclear localization signals and nuclear export motifs, we also tabulated homoionic segments with three or more, and at least 50 %, ionic sidechains. We abbreviate such segments as PCN, to indicate their ubiquitous presence in polynucleotide-binding proteins, protein carrier proteins (such as importins and exportins) and nuclear localization signal/import tracts. Most of ribosomal proteins feature abundant basic PCNs (bPCNs), and these segments are present in all mammalian LSU RPs. The above criterion in most cases also finds bPCNs in the mixed-ionic (bipartite) NLSs (thus, in all bipartite NLSs of Table S5). The bPCNs defined as above are found in 170 of 174 examined human/human viral nuclear localization motifs (Table S5), averaging 70.3 % of NLS length. Acidic PCNs (aPCNs) are quite scarce in a large majority of ribosomal proteins, but frequent in acidic ribosome-associating proteins, translocon proteins and the N-terminal exocellular segments of peptide receptors [see

(Parker et al. 2005, 2012)]. The scoring was performed by Microsoft Excel macros.

The identity between PCN segments across species was assessed in order of occurrence in the sequence, and then by positional identity within the individual segments. Identities of 100 and 50 % were assigned for the matching and similar residues, respectively (similarity being defined as in the SSEARCH3 and clustalW programs).

Boundaries of rRNA expansion segments

The boundaries are based on segments defined for human 28S LSU rRNA and 18S SSU rRNA by modeling of the secondary structure (Chandramouli et al. 2008). The segment boundaries of human 28S rRNA or 18S rRNA were applied to the SSEARCH3 or clustalW alignments of other ribosomal RNAs to locate the matching parts. As shown in Table S6.1, this approach suggests segment boundaries that for three vertebrate and a fungal large LSU rRNA correspond well with the bounds derived by modeling of secondary structures. Boundaries of LSU 25–28S rRNA ES segments defined by alignment with those of yeast 25S rRNA (Ben-Shem et al. 2011) also compare well with the modeling boundaries (Table S6.2), in accordance with an eukarya-wide insertion pattern (Michot et al. 1982; Ware et al. 1983). This will be considered in more detail in a separate communication.

Predicted association with RNAs

The predictions were made in RNABindR 2.0 (Terribilini et al. 2007) program, which is available on the Internet.

Hydrophobicity

Hydrophobicity was assessed using the coefficients of (Kyte and Doolittle 1982).

Statistical

The various parameters of protein and RNA sequences were compared in post hoc Bonferroni or Scheffé *t* tests, as indicated in the corresponding graphs and tables.

Surveys

Ribosomal proteins are generally larger in both subunits of eukaryote ribosomes as compared to non-eukarya

As seen in Fig. 1a, the average size of LSU RPs is significantly larger in all eukaryote groups as compared to bacteria and archaea (see the supplementary tables S1 (LSU) and S2 (SSU) for the lists of species and access codes). This is also found for 15 eukaryotic LSU proteins that show alignment (and may share ancestry) with archaeal and (to a quite limited extent) bacterial LSU RPs ["EAB" proteins (Armache et al. 2010b; Klinge et al. 2011); the supplementary Table S4.3], and for 19 additional eukaryotic LSU RPs that align with archaeal RPs ("EA" proteins; Table S4.2). A lesser size increase over prokaryotic RPs is found for SSU proteins of eukaryotic groups, and the vertebrate SSU RPs are not significantly larger than those of the surveyed archaea and bacteria (Fig. 1a, Table S3.1). The average size of LSU and SSU RPs in the same group is quite similar (Fig. 1a; Table S3.1). These findings corroborate and extend previous reports (Chandramouli et al. 2008; Armache et al. 2010b; Ben-Shem et al. 2011; Klinge et al. 2011). It should be emphasized that sizes of the confirmed sequences of several RP orthologs in multiple additional biotic species within the same clades closely resemble the sizes for these orthologs in the above groups, and similar is observed for the ionic aa content, segregation and clustering.

Hydrophobicity of LSU proteins in eukaryotes is less than in prokaryotes

A comparison of the Kyte–Doolittle scale hydropathies (Kyte and Doolittle 1982) somewhat surprisingly reveals a considerable drop in mean hydrophobicity for eukaryotic LSU RPs as compared to non-eukarya [-0.74 ± 0.04 (SD) vs. -0.54 ± 0.07] (Fig. 1b; see Table S3.2 for detail). The *G. lamblia* LSU hydropathy (for 37 available orthologs) is also lower than bacterial (but higher than in archaea), as shown in Table S3.2. No similar difference is found for the SSU parameters, which tightly average -0.56 ± 0.01 for the eukarya tested, very close to bacterial and archaeal SSU averages (Fig. 1b and Table S3.2). The mean hydrophobicity of eukaryotic SSU RPs is consistently higher than that of the corresponding LSU proteins.

The above findings correlate with the lower fraction in LSU proteins of eukaryotes, as compared to prokaryotes, of the large neutral (I, L, M, V) residues, (archaea 23.5 % and bacteria 23.6 %; eukarya from 20.9 % in insects to 21.6 % in mammals) and of the small neutral (A, G) residues (eukarya from 14 % in vertebrates to 14.7 % in plants; archaea, 15.8 %; bacteria 17.6 %). Clusters of three or more ILMV residues represent at most about 3 % LSU RP sequence (archaea 3 %, bacteria 2.8 %, eukarya from 1.9 % in insects to 2.5 % in non-mammalian vertebrates) (Fig. 1c; Table S3.3). In comparisons including all LSU proteins, differences in this parameter are not significant among eukaryote groups. The prokaryote LSU AGILMV >2 clusters are much more frequent, representing 12.1 % sequence in bacteria, 11 % in archaea, and significantly

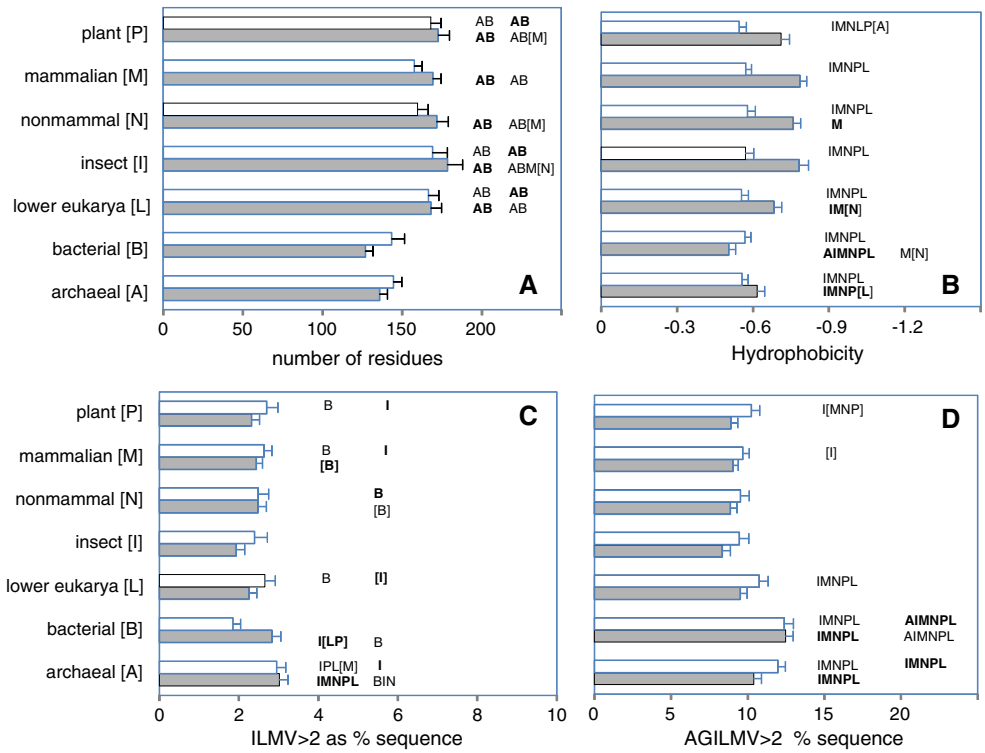

Fig. 1 A comparison of size, hydrophobicity and clustering of aliphatic neutral sidechains for LSU and SSU proteins of the groups surveyed. LSU data are shown by *shaded* and SSU by *blank bars*. The data are means with standard errors. Inscriptions within the graphs indicate means significantly lower than the specified ($p < 0.05$; $p < 0.1$ where bracketed) in post hoc Bonferroni t tests. The abbreviations are identified in *brackets* after group names at the *left* of *graph rows*. The *left* and *right columns* of the inscriptions identify, respectively, the LSU and SSU means that are significantly lower than the current group mean; the boldfaced tests refer to subunits of the same type. For other details see the "Methods" section and the text

less in all eukaryote groups, from 8.4 % in insects to 9.5 % in lower eukarya (Fig. 1d; Table S3.4).

The SSU proteins of eukarya and archaea show a slightly higher content of ILMV >2 clusters than bacteria, with no significant differences among eukaryote groups (Fig. 1c; Table S3.3). These clusters are quite scant in all SSU RPs, averaging only 1.9 % sequence in bacteria, and 2.4–2.9 % in other groups. The SSU AGILMV >2 clusters are much more frequent, respectively, constituting 12.1 % sequence in both archaea and bacteria, and 9.4–10.7 % in eukaryotic SSU RPs (Fig. 1d; Table S3.4). The prokaryotic parameters are significantly larger than those in eukaryotic groups (Fig. 1d).

The mean number of aa per sequence is very close across groups in ILMV >2 clusters (Table S3.5A), and in AGILMV >2 clusters (Table S3.5B) is slightly higher in bacteria as compared to other groups. Thus, in the eukarya studied there is a decrease in the frequency rather than the number of hydrophobic clusters. Irrespective of the clustering, the number of ILMV residues per the entire LSU RP sequence in bacteria and archaea is actually lower than the mammalian (79 and 90 % of the mammalian, respectively; Table S3.6), and the number of AG sidechains is close to the mammalian (92 % in archaeal, 93 % in bacterial LSU RPs; Table S3.6).

Both bulk basic and homobasic complement increase in eukaryotic ribosomal proteins, with a decrease in acidic character as compared to prokaryotes

The basic sidechain fraction is consistently higher in eukaryote LSU RPs compared to bacteria and archaea, and the metazoan groups also have this parameter significantly above lower eukarya and plant groups (Fig. 2a; Table S3.7). The LSU bulk HKR content of sequences is 21.7 % in archaea, 22.2 % in bacteria, and from 26 % in the lower eukaryote to 28.3 % in the mammals (Fig. 2a; Table S3.7). The mean LSU HKR aa numbers per sequence range from 25.7 in bacteria to 47.8 in insects, a remarkably high difference in cumulative basicity. The high LSU HKR content in metazoan groups is significantly above lower eukaryote (and especially *Giardia*; Table S3.7) and plant group values. LSU RPs of all higher eukarya have larger basic content than the corresponding SSU proteins (Table S3.7). The SSU RPs of archaea have significantly lower bulk HKR (19.6 %) than either bacteria (22.6 %) or eukaryotes (which range

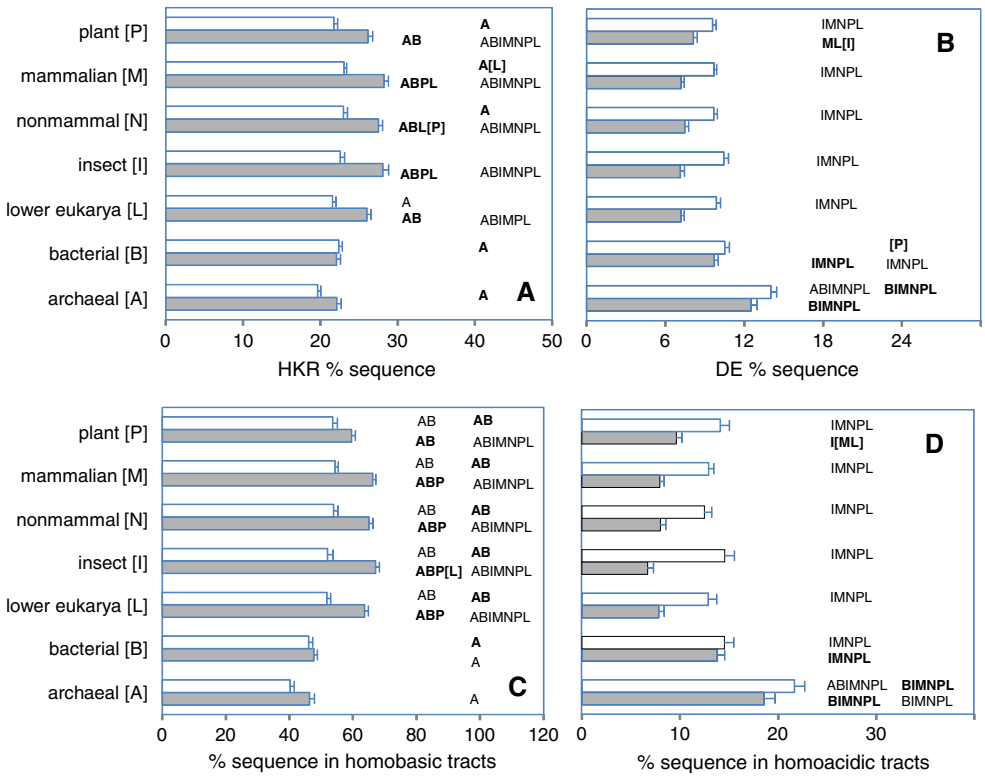

Fig. 2 Comparison of ionic sidechain content and sequence homoionicity in LSU and SSU proteins. The LSU data are in *shaded*, and the SSU in *blank bars*. Inscriptions in the *left* and *right columns*, respectively, identify LSU and SSU means lower than the current mean. For other details see the caption of Fig. 1, the "Methods" section, and the text

from 21.8 % in lower eukarya to 23 % in non-mammal vertebrate RPs (Fig. 2a; Table S3.7). This parameter of SSU RPs is quite close for all eukaryotic groups.

The acidic aa fraction of sequence is significantly lower in all eukaryotic LSU proteins compared to either bacteria or archaea (Fig. 2b; Table S3.8). The mean number of DE residues per LSU RP sequence is larger in archaea (17.5) than in any other group (ranging from 13 in bacterial to 15.1 in plant LSU RPs). The acidic fraction in SSU RP is consistently above that for LSU RPs, and significantly in most groups (Fig 2b; Table S3.8).

Homobasic segments with two or more HKR residues constitute less than 50 % of the LSU RP sequences in prokaryotic groups (46 % in archaeal, 48 % in bacterial), but from 57 % in Giardia to 66 % in mammalian RPs (Fig. 2c; Table S3.9). As will be shown in the next section, this considerable difference in sequestration mainly reflects an increased clustering of basic sidechains in eukaryotic LSU RPs. The heterotroph eukaryote groups appear to have a larger LSU homobasic sequestration than plant LSU RPs. The SSU proteins also show increased homobasic fractions in eukaryotes (from 52 % in lower eukarya to 54 % in mammalian; Fig. 2c; Table S3.9) compared to either archaea (41 %) or bacteria (48 %), but differences in sequestration are rather less than found for the LSU proteins.

Homoacidic segments with two or more DE residues occupy much smaller part of sequence in eukaryotic LSU RPs (the highest being 9.6 % in plants) than in archaea (18.7 %) or bacteria (13.9 %), and in general represent less than a fifth of the sequences (Fig. 2d; Table S3.10). The SSU proteins of archaea have 21 % sequence in homoacidic segments, quite above any other group (Fig. 2d; Table S3.10). In eukaryotic groups, the homoacidic SSU fraction ranges from 12.5 % in non-mammalian to 14.6 % in plant RPs, and is at least 35 % above the LSU fraction (Fig. 2d; Table S3.10).

The interdomain-aligning EAB LSU proteins show lower basicity and quite larger acidity and hydropathy than EO and EA LSU proteins, and also are significantly larger than these (Tables S4.1–3). As with the entire LSU RP sets, these differences from prokaryotes are most prominent in the mammal, and lowest for the lower eukaryote group. These findings will be considered in detail in a separate communication.

The basic clusters and PCN tracts are highly augmented in metazoa even relative to lower eukarya

The acidic PCN motifs with at least three DE sidechains and having at least 50 % DE content (aPCNs) are poorly

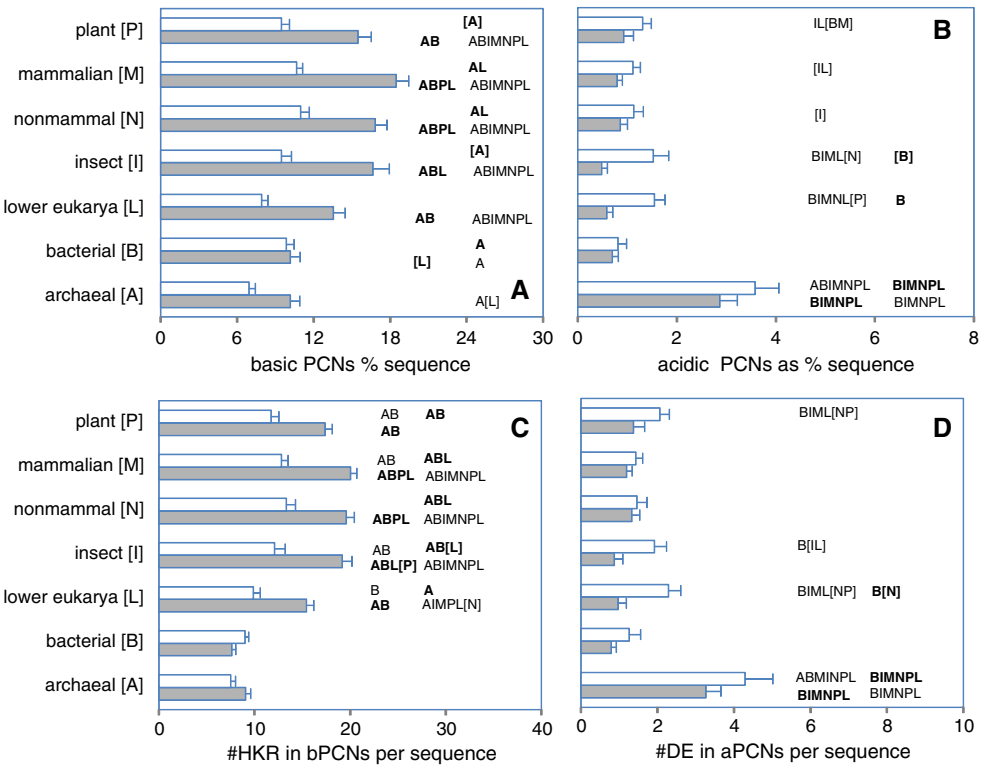

Fig. 3 Parameters of the PCN motifs in LSU and SSU proteins. The LSU data are in *shaded*, and the SSU in *blank bars*. Inscriptions in the *left* and *right columns*, respectively, identify the LSU and SSU means lower than the current mean. For other details see the caption of Fig. 1, the "Methods" section, and text

represented in the ribosomal proteins surveyed. (The acidic LSU ribosomal proteins (UniProtein designations P0–P7; not included in this survey) have numerous acidic clusters and PCN motifs, but are essentially devoid of basic clusters and PCN motifs, and the initiation/elongation factors (also not surveyed) frequently have numerous acidic as well as basic PCNs). Even the archaeal RPs show aPCN motifs in only 49 % LSU and 55 % SSU sequences, and eukaryotes have these in, at the most, 29 % LSU (in non-mammalian vertebrates) and 48 % SSU RPs (in plants). Both LSU and SSU RPs of the archaea have more acidic PCN motifs than any other group, but these motifs comprise less than 4 % sequence aa (Fig. 3a; Table S3.12). The lower eukarya and insect SSU show more acidic PCN motifs than other SSU groups (Fig. 3a; Table S3.12). The acidic PCNs are very low in both LSU and SSU RPs of bacteria (Fig. 3a).

The basic PCNs are found in all of the examined mammalian LSU RPs, in virtually all other eukaryotic LSU proteins and in 91–97 % RPs across eukaryotic SSU protein groups; the corresponding percentages in archaea are 86.4 for LSU and 84.3 for SSU, and in bacteria 83.3 for LSU and 98.1 for SSU RPs (Table S3.11). These motifs represent <12 % LSU or SSU sequence in archaea, and <10 % in bacteria, but 13.4 % in the lower eukaryotes (though only 10 % in the incomplete *Giardia* set) and 18.5 % in the mammalian LSU RPs (Fig. 3b; Table S3.11). In the LSU of eukaryotes, this parameter is significantly above non-eukarya in all groups, and in all metazoan LSU groups is significantly higher than in the lower eukaryote group. The SSU RPs in most groups have much lower basic PCNs than the LSU proteins (Fig. 3b; Table S3.11).

The numbers of ionic residues within sequence PCN motifs (graphs c and d in Fig. 3) could be useful as indicators of the overall RP interactive homoionic impact. The basic counts (Fig. 3c; Tables S3.13) reach 20 per sequence (or 11.8 % of all sequence aa) in the mammalian LSU and are significantly higher than the lower eukarya and the plant counts in all metazoan groups. The LSU counts are about 1.5 times larger than SSU counts in all eukaryote groups. All eukaryotic LSU counts significantly outnumber the archaeal and bacterial, and the lower eukaryote SSU values are also significantly above the archaeal (Fig. 3c). The acidic counts (Fig. 3d; Tables S3.14) average less than three residues per SSU sequence, and less than two per LSU sequence in all groups except archaeal (which has the respective averages of 4.3 and 3.3).

Probability of interaction with RNA as judged by RNABindR program (Terribilini et al. 2007) is similarly high for bPCN motifs in 15 "EAB" LSU RPs (Table 1; see Table S4.3 for the list of these proteins). Similar values were obtained in BindN program (Wang and Brown 2006). The estimated probabilities are similar and high for bPCNs of the three protein sets. The averaged probabilities for the sequence parts outside of bPCNs are barely significant.

Table 1 Compared probabilities of association with RNA for basic PCN motifs in 15 LSU EAB proteins

Species	bPCNs present in	bPCN count	% Sequence in bPCNs	*p* for bPCN	*p* outside
H sapiens	15	91	14.8 ± 1.4	0.821 ± 0.024	0.535 ± 0.015
M. jannaschii	13	49	9.21 ± 1.5	0.842 ± 0.038	0.545 ± 0.025
M. tuberculosis	14	37	7.71 ± 1.2	0.870 ± 0.035	0.528 ± 0.022

p is the average probability of association with RNA from RNAbind2 program

p outside is the average probability for sequence parts not in bPCN clusters

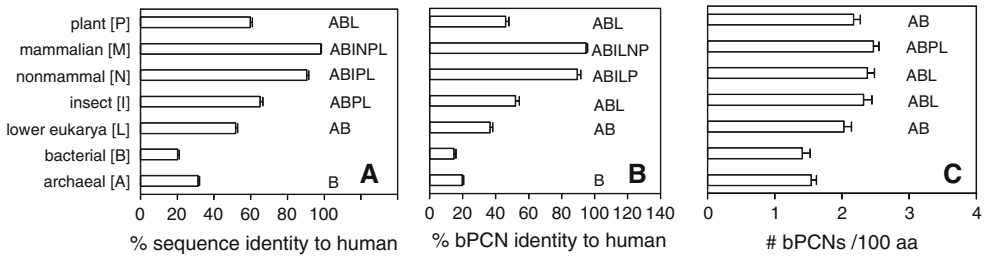

Fig. 4 Basic PCN motifs in LSU ribosomal proteins as compared with the corresponding human proteins across the groups examined. The eukaryote data are averages for 41 matching LSU RP orthologs, and the archaeal and bacterial data are for LSU RPs that align best with the human. **a** Percent identity of ribosomal proteins to matching human LSU RPs (estimates from SSEARCH3 program). **b** Percent identity of bPCN motifs to human bPCNs, assessed in order of occurrence in the sequence (see "Methods"). **c** The number of basic PCNs per 100 residues in sequence. See the caption of Fig. 1 about the assignation of significance

Patterns of LSU basic clusters are highly conserved in the vertebrate, uneven in lower eukarya, and quite random in prokaryotes

Sequences of the individual orthologs of vertebrate ribosomal proteins are known to be highly aligned across families and orders, and this of course is verified with 41 LSU orthologs in the seven groups of species examined in this survey (Tables S4.1–3). The whole-sequence identity with human orthologs exceeds 98 % for the five mammalian and is 91 % for the three non-mammalian species examined (Fig. 4a; Table S4.4.1). The two insect and three plant LSU ortholog sets have, respectively, 65 and 60 % sequence identity to the human set and this identity is 52 % in lower eukarya eukaryotes (Fig. 4a; Table S4.4.1). The sequence identity with human quasi-orthologs is above 30 % for the examined archaea, but only 20 % in bacteria (Fig. 4a; Table S4.4.1).

The mammalian bPCNs are highly conserved, with about 95 % identity to the human (Fig. 4b), which is close to that for the entire sequences (Fig. 4b). The bPCN identity with human RP orthologs decreases to 89 % in non-mammalian vertebrate RPs examined, and then to 52 % in insects, 46 % in plants and 37 % in lower eukarya (Fig. 4b; Table S4.4.2). The archaeal and bacterial bPCNs show, respectively, only 20 and 15 % identity to human bPCNs. The number of basic PCN clusters per 100 aa in LSU RP sequences is much higher in eukaryote as compared to non-eukaryote groups, and

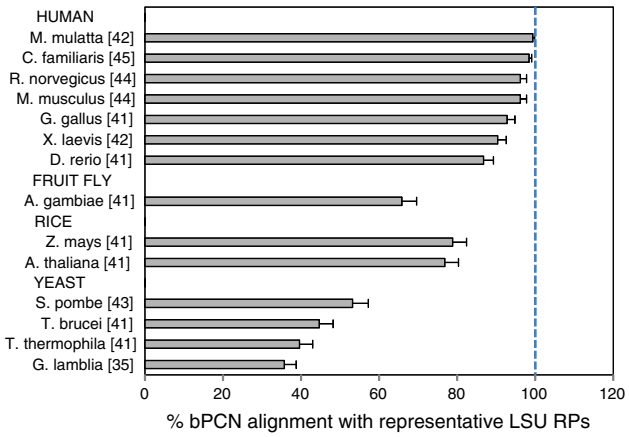

Fig. 5 Basic PCN clusters in eukaryotic LSU ribosomal proteins compared within groups of species with orthologs from a representative species. The eukaryote data are averages for the indicated numbers of matching LSU RP orthologs. The within-group comparator sequences were from species indicated by capital inscriptions above the species names in the group (HUMAN, *H. sapiens*; FRUIT FLY, *D. melanogaster*; RICE, *O. sativa*; YEAST, *S. cerevisiae*)

the metazoan RPs have significantly higher frequency of these clusters than the lower eukarya (Fig. 4c; Table S4.4.3). Due to the larger size of eukaryotic LSU RPs, difference in the number of bPCN clusters per sequence is much larger than the difference in cluster frequency (Table S4.4.3).

A detailed matching of basic PCN clusters in eukaryotic LSU orthologs with comparator sequences from the same

Table 2 25–28S LSU and 18–20S SSU rRNAs of species examined for relation of RP basicity and constitution of RNA expansion segments

Group	Version	Species	Total nt	Core nt	ES nt	Total GC%	Core GC%	ES GC%
25–29S LSU rRNAs								
Mammalian	U13369.1	*H. sapiens*	5 035	2,648	2,387	69.18	56.16	83.62
	NR_003279.1	*M. musculus*	4,730	2,657	2,073	66.62	56	80.22
	NR_046246.1	*R. norvegicus*	4,786	2,653	2,133	67.38	55.94	81.62
Amphibian	X02995.1	*X. laevis*	4,082	2,669	1,413	65.39	55.83	83.44
Insect	KC177663.1	*A. gambiae*[a]	3,440	2,291	1,149	54.51	53.34	56.83
	M21017.1	*D. melanogaster*	3,900	2,602	1,298	39.28	44.27	29.28
Lower eukarya	J01355.1	*S. cerevisiae*	3,396	2,538	858	47.94	47.12	50.35
	X14553.1	*T. brucei*	4,647	2,703	1,944	48.74	49.87	47.17
	JN547815.1	*T. thermophila*	3,354	2,593	761	44.75	45.05	43.76
Plant	X52320.1	*A. thaliana*	3,375	2,515	860	55.73	53.12	63.37
	M11585.1	*O. sativa*	3,377	2,577	800	59.43	54.52	75.25
	AJ309824.2	*Z. mays*	3,385	2,570	815	58.70	53.81	74.11
18–20S SSU rRNAs	K03432.1	*H. sapiens*	1,870	1,343	527	56.30	56.10	56.7
	NR_003278.3	*M. musculus*	1,870	1,344	526	56.04	55.81	56.65
	NR_046237.1	*R. norvegicus*	1,874	1,347	527	55.71	55.46	56.36
Amphibian	X02995.1	*X. laevis*	1,825	1,327	498	53.81	53.81	53.81
Insect	AM157179.1	*A. gambiae*	2,015	1,462	553	50.42	51.3	48.10
	M21017.1	*D. melanogaster*	1,995	1,482	513	42.51	43.79	38.79
Lower eukarya	Z75578.1	*S. cerevisiae*	1,800	1,328	472	44.83	47.13	38.35
	M12676.1	*T. brucei*	2,251	1,545	706	51.44	51.33	51.70
	2XZN_A	*T. thermophila*	1,751	1,319	432	42.83	44.96	36.34
Plant	GQ380689.1	*A. thaliana*	1,802	1,318	434	49.17	50.23	46.29
	AF069218.1	*O. sativa*	1,808	1,260	491	51.34	52.06	49.49
	K02202.1	*Z. mays*	1,809	1,324	485	51.02	51.66	49.28

The expansion segments (ES) were defined by alignment with human 28S and 18S rRNA boundaries modeled by Chandramouli et al. (structure 16:535–548)

[a] Partial sequence

group of species is presented in Fig. 5 (with data in Table S4.5). The large correspondence of human data with those from other vertebrates is seen to reflect the taxonomic position of the species, from >99 % in the monkey to 87 % in the fish. The plant bPCN alignments to rice do not exceed 79 %, and do not differ appreciably for monocotyledon *Z. mays* (in the same family with rice) and the dicotyledon *A. thaliana*. Basic clusters in the LSU RP sets of the two dipteran insects match only 66 % (Fig. 5; Table S4.5), which may or may not relate to the large difference in predicted GC content of the respective large LSU rRNA expansion segments (34 % in *D. melanogaster* and 57 % in *A. gambiae*, Table 2). The taxonomically highly diverse LSU RP sets of lower eukarya differ very much in bPCN clusters, with matchup to yeast of 53 % for *S. pombe* (from a different class of the same phylum), and of only 36 % for *G. lamblia*.

In the archaeal and bacterial species examined the LSU ribosomal proteins even at the level of taxonomic families differ extensively in overall sequence, and much more in basic cluster patterns. For six LSU tunnel proteins of 11 archaea (listed in Table S4.6) compared with the respective *M. jannaschii* orthologs, the family identities are in the range of 60–80 % (Fig. 6a; Table S.4.6.1) and the bPCN identities are as low as 10 % in Halobacteriaceae (Fig. 6a). As would be expected, the *M. jannaschii* identities are very low and uneven against bacterial LSU tunnel proteins (Fig. 6b; Table S.4.6.2). The identities of the bacterial tunnel proteins to the respective *M. tuberculosis* orthologs as expected are much higher, but again highly uneven, with the bPCN identities ranging from 22 % in the four Enterobacteriaceae to 88 % in two other Mycobacteriaceae tested (Fig. 6c; Table S.4.6.3).

The above findings could be consistent with a uniform increase in stability of association of LSU proteins with the large rRNA in multicellular eukaryotes as compared to lower eukaryotes and prokaryotes. Also, as seen in Figs. 4, 5 and Tables S.4.4–5, the sequence alignment and patterns of LSU basic clusters have an essentially phylum-wide identity in the vertebrate.

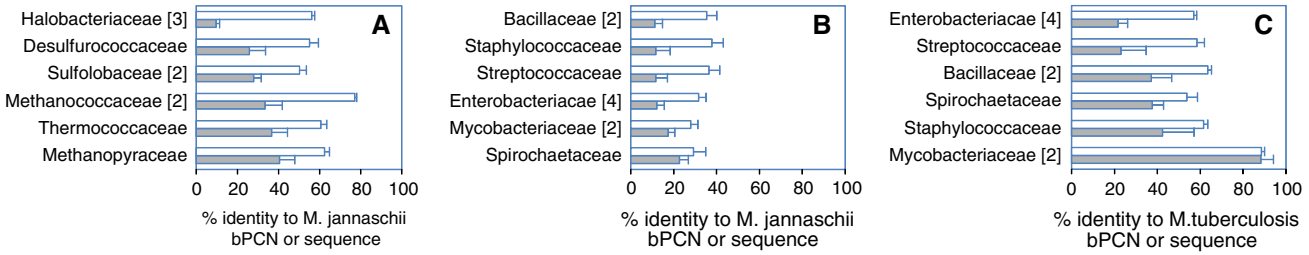

Fig. 6 Basic PCN motifs in prokaryote LSU export tunnel proteins differ considerably among the taxonomic families and in most cases have much lower identity to representative LSU orthologs than found for the entire sequences. The LSU tunnel proteins L4, L22, L23, L24, L29 and L31 from 12 species of archaea and bacteria were examined for bPCN and whole-sequence identity to the corresponding orthologs of *M. jannaschii* or *M. tuberculosis*. **a** Identities of archaeal LSU tunnel proteins to *M. jannaschii* orthologs. **b** Identities of bacterial LSU tunnel proteins to *M. jannaschii* orthologs. **c** Identities of bacterial LSU tunnel proteins to *M. tuberculosis* orthologs. The percent identities for bPCNs and the entire sequences are respectively shown in *shaded* and *blank bars*. The numbers of species per family where above 1 are shown in *brackets*. For other details see "Methods" section and text

It should be pointed out that aPCNs and bPCNs of *srp* and *sec* proteins, but especially acidic clusters, are much more abundant in vertebrates as compared to lower eukaryotes. Also, as with ribosomal proteins, the vertebrate *srp* and *sec* PCNs are highly aligned in orthologs from different classes, which again is not found for lower eukaryotes. This will be the subject of a separate communication.

Increase of basic parameters in eukaryote ribosomal proteins is similar in the N-terminal and the C-terminal halves of their molecules

Increases in basic residues and motifs of ribosomal proteins in eukaryotes (Figs. 2, 3) may have developed primarily by internal expansion of the prokaryotic progenitor sequences, or mainly via extension at the termini. Also, an uneven distribution could exist regardless of the ancestry considerations. An examination of this subject in 15 "EAB" LSU proteins shows that the average basic cluster position as percent of sequence length is 46.7 ± 3.1 (SE) in 91 human, 40.9 ± 4.2 in 49 archaeal, and 49.8 ± 4.5 in 37 bacterial motifs, which does not indicate important distribution differences among the groups, or preference for either the N-terminal or the C-terminal halves.

Examination of contingency of ionic clusters and homoionic segments in RPs is difficult to perform due to the low average length (120–180 aa, Table S3.1) and more than tenfold size range of the sequences. Also, the segments/clusters are frequently large and sometimes span most of the sequence. We, therefore, opted for examination of all ionic parameters in the N- and C-terminal halves of LSU RP molecules.

To get a more detailed insight, we examined the distribution of ionic elements in the N- and C-terminal halves of LSU RPs in the studied seven groups of species. No important differences were found in the halves for the bulk basic residue content, homobasic fraction and basic PCN motifs across the groups (Table 3), including a lack of difference in the halves of prokaryote RPs. The basic elements in eukaryote LSU RPs thus appear to be rather evenly distributed within sequences, generally favoring an expansive enlargement in eukaryote RPs that do align significantly with the prokaryote quasi-orthologs.

The much less represented acidic residues are, however, significantly more frequent in the C-terminal half of LSU RP sequences in all groups. This is also found for the acidic segregation. The acidic PCN motifs, especially in plant RPs, are quite depleted in N-terminal halves of eukaryotic LSU RPs as compared to prokaryotic. However, these motifs are found in not more than 21 % of the eukaryotic LSU RP sequences (and 19 % of plant sequences; Table 3).

The absence of important changes in abundance of the ionic parameters relative to prokaryotes between the halves of eukaryote RPs is also indirectly supported by the lack of difference in hydrophobicity for the N- and C-terminal halves across the groups (Table 3), again suggestive of a generally insertive enlargement of hydrophilic constituents.

Expansion of ribosomal RNAs and nucleotide bias do not seem to directly relate to increase in the basicity of ribosomal proteins

The vertebrate (and especially tetrapod) 28S rRNAs have massive GC-rich expansion segments (Clark et al. 1984; Hassouna et al. 1984; Wakeman and Maden 1989; Chandramouli et al. 2008) [however see (Nunn et al. 1996)], while the 25–26S rRNAs of lower eukaryotes typically have much smaller ES with lower GC content (Armache et al. 2010b; Ben-Shem et al. 2011; Klinge et al. 2011). Since a large fraction of ribosomal proteins is quite stably associated with rRNAs, and there is a considerable enrichment in multi-basic tracts in metazoa as compared to lower eukarya (Figs. 3, 4), the obvious disparities in LSU

Table 3 Sequence parameters of the LSU ribosomal protein N-terminal half as percent of values in the C-terminal half

Parameter	A	B	L	I	N	M	P
HKR % sequence	101.7	99.28	100.1	100.7	102.2	102.6	105.4
DE % sequence	87.42	82.51	71.79	72.3	67.46	72.53	60.2
>1-Homobasic segments	104.7	102.9	111.2	104.6	107.5	104.7	117.5
Found in % Nt	98.05	97.5	100	100	100	100	100
Found in % Ct	96.59	99.5	100	100	100	100	100
>1-Homoacidic segments	77.19	98.64	74.1	54.63	55.89	60.75	44.4
Found in % Nt	78.35	61.25	62.99	52.94	60.15	63.48	59.42
Found in % Ct	79.56	73	74.8	75.29	73.68	73.04	81.88
Basic PCN clusters	105.4	120.6	97.43	102.4	106.4	104.4	109.9
Found in % Nt	67.15	62.5	84.25	90.59	93.23	93.04	88.41
Found in % Ct	64.48	55.5	87.4	89.41	87.22	88.7	92.75
Acidic PCN clusters	76.73	58.13	34.94	30.59	23.9	33.22	19.02
Found in % Nt	31.87	8.75	7.09	5.88	9.02	9.13	9.42
Found in % Ct	34.55	13.75	16.54	14.12	21.05	20	16.67
Hydrophobicity	93.9	94.38	87.97	90.71	90.91	95.2	93.42

The one-letter column headings represent the seven groups of species examined, as identified in Figs. 1, 2, 3 and 4. ">1-homobasic segments" and ">1-homoacidic segments" are the percentages of sequence in the respective homoionic segments with two or more ionic sidechains. Basic residues are present in both halves of all sequences, and acidic in more than 89 %

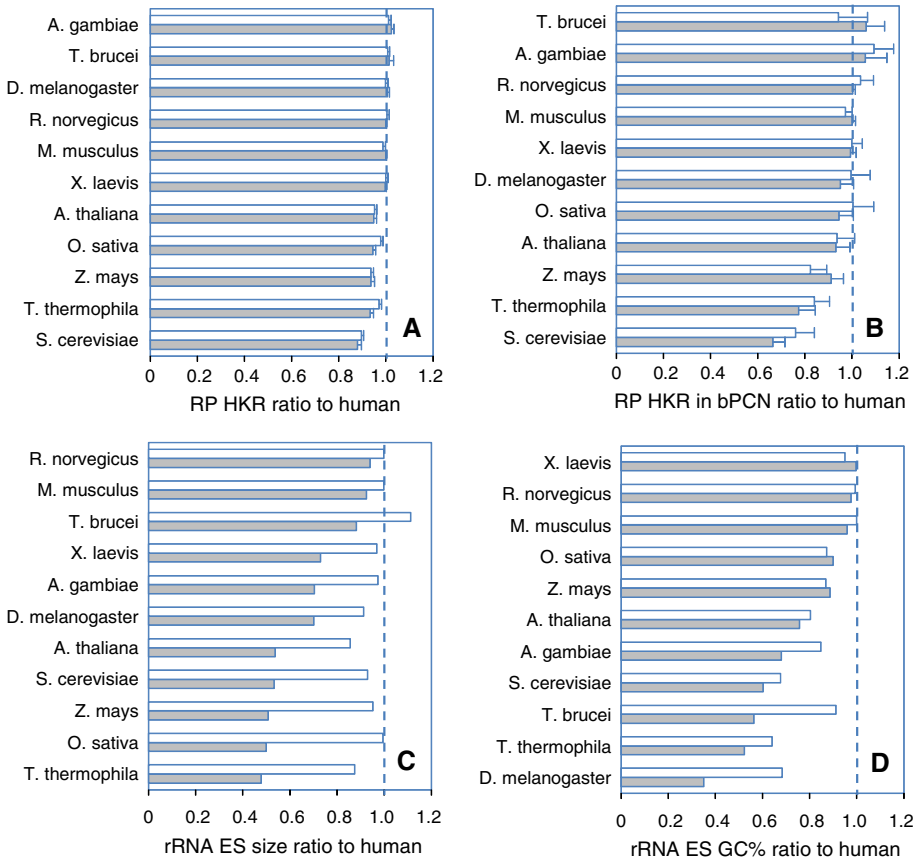

Fig. 7 Parameters of ribosomal protein basicity and ribosomal RNA expansion segments relative to the corresponding values in human RPs and RNAs. Data for LSU and SSU are shown in *shaded* and *blank bars*, respectively. The protein data are for ortholog sets of 41 LSU and 31 SSU ribosomal proteins of the species indicated. **a** Percent sequence in basic sidechains as ratio to the corresponding human RP value. **b** Percent sequence in basic sidechains within bPCNs as ratio to the corresponding human. **c** Percent rRNA sequence in expansion segments as ratio to the corresponding human rRNA value. **d** Percent ES GC content as ratio to the corresponding human ES value. The numerical data are in tables S4.6. For other details see "Methods" section and text

ES size and nucleotide bias would prompt checking the correlation of size and GC content with enlargement of the basic content.

Twelve eukaryote species examined in this survey have sequences elucidated for all rRNAs (see Table 2 for ES and core parameters), and for most of the consensus LSU and SSU ribosomal proteins. (Note that the following comparisons include a relativization to the corresponding human RPs, necessary to bring the data in register.) The sequences of matching orthologs of 41 LSU RPs are

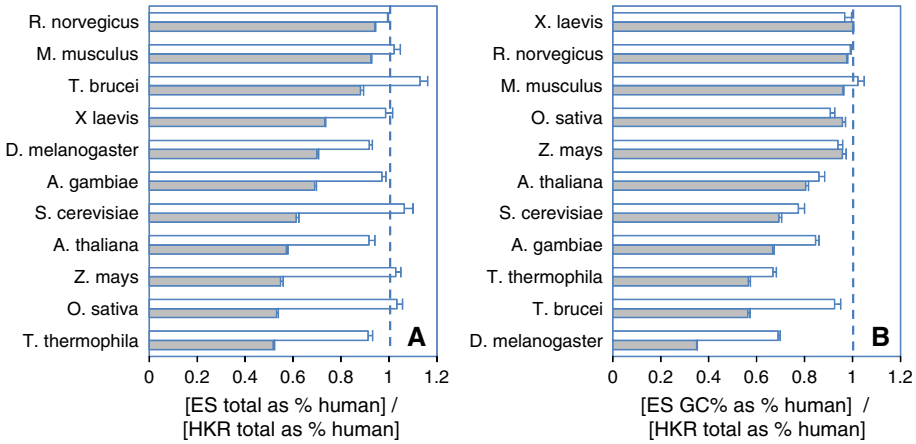

Fig. 8 A comparison of eukaryote rRNA expansion segments and basicity of ribosomal proteins. The data are ratios of the parameters relative to the corresponding human values for eukaryote LSU and SSU rRNA expansion segments (ES) and ribosomal proteins in 11 species from mammalian, non-mammalian vertebrate, insect, lower eukarya and angiosperm plant groups. 41 LSU RP and 31 SSU RP orthologs matching across the species were examined. LSU values are in *shaded*, SSU values in *blank bars*. **a** The ratio of percent sequence in ES segments expressed (as % human) to percent basic sidechains in sequence expressed as % human. **b** The ratio of percent GC in ES segments (as % human) to percent basic sidechains in sequence as % human. For other details see the "Methods" section and the text

confirmed for all, and 31 SSU RPs are confirmed for eleven of these species (27 in *T. brucei*). These proteins did not show important differences in the bulk HKR fraction for either subunit (Fig. 7a; Table S.4.7.1). The basic residue sequence fraction in bPCN motifs was significantly below human only for LSU RP of S. cerevisiae (Fig. 7b; Table S.4.7.2).

Expansion of rRNAs apparently proceeded in rather different ways for the LSU (25–28S) and the SSU (17–20S, mainly 18S) types. The expansion segments of LSU rRNAs show considerable increase in size from lower eukaryotes to metazoans, and especially in the mammal (Fig. 7c; Table S.4.7.3), leading to LSU rRNA and ES size much larger than is typical in the non-organismic eukaryotes. The ES size increment is generally tied to a biased nucleotide composition of ES, in most cases toward higher GC content (Fig. 7d; Table S.4.7.4). The SSU rRNAs, on the other hand, show little increase in size of the expansion segments from lower eukarya to the mammalian group (Fig. 7c). The variation in SSU GC content is less than 25 %, although there is a GC increment in the vertebrate species (Fig. 7d; Table S.4.7.4).

The above changes, however, do not correlate with the bulk HKR fraction or basic PCN content of ribosomal proteins (Fig. 8). This is especially clear in metazoan LSU comparisons. In the case of ratios of LSU ES size to bulk HKR content (Fig. 8a; Table S.4.8.1), the mammalian values are significantly higher (in post hoc Scheffé *t* tests) than any other, and any non-mammalian means are significantly larger than those positioned below in the graph. The SSU ratios do not show a consistent pattern of difference for either the ES size/HKR content (Fig. 8a) or the GC over HKR fraction ratio (Fig. 8b; Table S.4.8.2). Thus, the 26–28S rRNA expansion segments are enlarged in higher eukarya, and 18S ES are not, and the enlargement of LSU bulk basic fraction does not reflect ES composition. The type of nucleotide bias does not link to RP constitution, as is clear from the *D. melanogaster* parameters; this is also found in a number of other arthropods, and in vertebrate mitochondrial rRNAs (M.S. Parker and S.L Parker, in preparation). Increases in the tetrapod vertebrate ES may connect primarily to increased compartmentation of ribosomes and other components of the protein-synthesizing and export apparatus (see the "Discussion").

Discussion

Enlargement of the ribosomal subunits in eukaryotes involved addition of a number of proteins and increase in size of both the proteins and the rRNAs. The size difference with prokaryotes is more substantial in the large subunit. At the protein level, the size change appears to have been effected already at the level of the non-organismic eukarya. However, the metazoan LSU large rRNAs show a phylogeny-related elongation culminating in the primate and being mostly due to extension of the expansion segments. Expansion of the SSU rRNA appears to have saturated in lower eukaryotes. The discordant upsizing of euglenozoan rRNAs (Table 2) could perhaps link to interactions with host genomes [see (Busse and Preisfeld 2002; Alkemar and Nygard 2004)].

Increases in RP basicity in eukaryotes apparently developed in a biphasic manner. The first phase involved

mainly basic residue content and sequestration (Fig. 2a, c) and could relate significantly to the need for carrier-supported transport of RPs (Jakel and Gorlich 1998; Bange et al. 2013) to the nuclear sites of ribosome assembly. An expansion of both large rRNAs, typically by not more than 500 nucleotides, apparently accompanied this rise in RP basicity and sequestration. The second phase involved increased sequestration of basic residues for LSU RP of metazoa relative to prokaryotes, lower eukarya and plants (Fig. 2c) and also increases in basic clusters for both LSU and SSU RP compared with lower eukarya and prokaryotes (Fig. 3a, c). The very high alignment among vertebrate LSU PCNs may testify to a phylum-wide optimization of interactive tools that should affect association with both the carriers and the receptors, and from our preliminary examination could be reflected in a similar degree of alignment in *srp* proteins and translocons.

The coincident second phase expansion of LSU 26–28S rRNAs added another 500 nucleotides in insects, and up to 2,000 in mammals, while the SSU rRNA expansion largely remained at the level of lower eukarya and plants (Table S2-RNA). Expansion of the vertebrate large LSU RNA is obviously very much above that of SSU RNA and could be noticeable even in a contemporary timeframe [e.g., (Leffers and Andersen 1993)], what clearly is not found for the highly aligned vertebrate LSU RPs. Much higher expansion rates for LSU as compared to SSU rRNAs are also estimated in insects (Stage and Eickbush 2007), and even in plants (Kuzoff et al. 1998).

The increased sequestration and clustering of basic residues in metazoan LSU RPs and the large expansion of metazoan 26–28S rRNAs show no obvious correlation (Figs. 7, 8). However, both should improve contacts and enhance association with the internal as well as the external partners. The enrichment in translocon acidic clusters in the vertebrate as compared to fungi and prokaryotes could be envisaged as an evolutionary interactive match of the enlarged RP basicity.

The increment in basicity should affect ribosome organization and reactivity in a number of ways. Increase of homobasic clusters should competitively reduce sensitivity of rRNAs to typical RNases, which have abundant homobasic segments and clusters [see, e.g., Table 5 in (Parker et al. 2012)]. While the comparative studies are lacking, the reported rRNA sensitivity to pancreatic RNase is lower in rat liver than in E. coli ribosomes. This may also relate to the physiological stability of rRNAs, which in homeotherm vertebrate cells and tissues typically show half-lives in excess of 24 h (Bowman and Emerson 1977; Nwagwu and Nana 1980; Sameshima et al. 1981). The microbial rRNA turnover is difficult to measure, but could have half-periods in the range of hours (Hsu et al. 1994; Kalpaxis et al. 1998), what could partly relate to a lesser RNA protection due to the lower basicity of the bacterial RPs. The much larger comparative basicity of vertebrate RPs can also be protective in terms of access of large antibiotics (e.g., macrolides and aminonucleosides) to ribosome structures involved in the respective sensitivities of bacterial ribosomes (Chattopadhyay et al. 1999; Gabashvili et al. 2001; Davydova et al. 2002).

The large complement of basic clusters in eukaryote RPs should enhance protein binding to rRNA backbone phosphate (Ellis et al. 2007; Ciriello et al. 2010)to increase stability of the ribosome. The early-installed eukaryotic LSU proteins L4 (Kruger et al. 2007), L7 and L8 (Jakovljevic et al. 2012) have 10-residue bPCNs, and the strongly RNA-crosslinked L6 (Uchiumi et al. 1983) has the most dense basic clusters among the vertebrate RPs. Helix-destabilizing RP activity might be present in clusters that also contain proline (found in about 30 % of bPCNs in the vertebrate, but only 14 % in the lower eukarya). However, the examined RPs lack DEAD and the like helix-breaking motifs, and could rather enhance the RNA helix stability and decrease the constitutive activity (Qu et al. 2011), probably depending on the size of the clusters.

The boost in external partnerships of the 60S subunit is well supported in the mammal (Adelman et al. 1973; Kruppa and Sabatini 1977). This could also be linked to acquisition of membrane-ensconced ribosome receptors which may not be present in lower eukaryotes (Morrow and Brodsky 2001), and may also act as matrix organizers (Diefenbach et al. 2004; Ueno et al. 2012). Ribosome-binding protein p180 (RRBP1; Q28298), which appears to be restricted to mammals, enhances the ability of yeast microsomes to bind native (Morrow and Brodsky 2001) or dog (Wanker et al. 1995) ribosomes. This protein has >40 repeats containing NQGKKXE motif, a likely target of G or C repeats (Du et al. 2005), and also multiple acidic and basic clusters. This formidable polar/ionic complement could also significantly interact with translocons, providing support for the triage and translocation of the nascent polypeptides by the *srp*/*sec* system. Further research may reveal additional ribosome-associating proteins in the ER and other intracellular matrices. The significantly larger augmentation of LSU over the SSU RP basic parameters is unlikely to be caused primarily by differences in transport to the nucleus, or in RNA backbone phosphate access, but could relate significantly to association with membrane proteins, including those as yet unidentified.

The big ES7 and ES27 tracts of eukaryote 25–28S RNAs largely lack stable association with ribosomal proteins (Chandramouli et al. 2008; Armache et al. 2010a; Ben-Shem et al. 2011; Klinge et al. 2011). These segments could be localized mainly on ribosome surface (Larsson and Nygard 2001; Chandramouli et al. 2008), coordinating the access of external partners (Beckmann et al. 2001;

Armache et al. 2010a). The high GC bias of ES in vertebrate 28S rRNAs (Clark et al. 1984; Wakeman and Maden 1989; Chandramouli et al. 2008) affects up to a half of the molecule, and is significantly constituted of 4–11 nucleotide G and C repeats [e.g., (Leffers and Andersen 1993)]. These repeats can focus on a variety of targets. Guanine can bind Asp and Arg via both conventional and Hoogsteen pairing (Kondo and Westhof 2011), and also bind Trp (Ellis et al. 2007). Cytosine could contribute to phosphate binding via the Hoogsteen edge, and guanine via the Watson–Crick edge (Zirbel et al. 2009). Cytosine repeats may target lysine-rich motifs in proteins (Du et al. 2005). The targets could also include non-ribosomal RNAs (Sweeney et al. 1996), messenger ribonucleoproteins (Ohashi et al. 2002; Potter and Nicchitta 2002), and ribosome receptors with multiple ionic repeats and PCNs, such as p180 (Wanker et al. 1995; Ueno et al. 2012), analogous to their interactions with mRNAs (Cui et al. 2011). However, across the organismic eukaryotes surveyed, the nucleotide bias and size of rRNA expansion segments seem to increase consistently only in the vertebrate.

Acknowledgments This study was partly supported by the U.S. National Institutes of Health Grant HD13703.

Conflict of interest The authors declare no conflict of interest related to his study.

Note added in proof While this survey was considered for publication, Fedyukina, Jennaro and Cavagnero (J Biol Chem 2014, M113.507707) reported differences in hydropathy and basicity of ribosomal proteins between halophile and non-halophile prokaryotes, and also presented evidence for homoionic charge segregation in these proteins [in accord with our current and previous observations (Parker et al. 2012)].

References

Adelman MR, Sabatini DD, Blobel G (1973) Ribosome-membrane interaction. Nondestructive disassembly of rat liver rough microsomes into ribosomal and membranous components. J Cell Biol 56:206–229

Adesnik M, Maschio F (1981) Segregation of specific classes of messenger RNA into free and membrane-bound polysomes. Eur J Biochem 114:271–284

Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P (eds) (2002) Molecular biology of the cell. Garland, New York

Alkemar G, Nygard O (2004) Secondary structure of two regions in expansion segments ES3 and ES6 with the potential of forming a tertiary interaction in eukaryotic 40S ribosomal subunits. RNA 10:403–411

Armache JP, Jarasch A, Anger AM, Villa E, Becker T, Bhushan S, Jossinet F, Habeck M, Dindar G, Franckenberg S, Marquez V, Mielke T, Thomm M, Berninghausen O, Beatrix B, Soding J, Westhof E, Wilson DN, Beckmann R (2010a) Cryo-EM structure and rRNA model of a translating eukaryotic 80S ribosome at 5.5-A resolution. Proc Natl Acad Sci USA 107:19748–19753

Armache JP, Jarasch A, Anger AM, Villa E, Becker T, Bhushan S, Jossinet F, Habeck M, Dindar G, Franckenberg S, Marquez V, Mielke T, Thomm M, Berninghausen O, Beatrix B, Soding J, Westhof E, Wilson DN, Beckmann R (2010b) Localization of eukaryote-specific ribosomal proteins in a 5.5-A cryo-EM map of the 80S eukaryotic ribosome. Proc Natl Acad Sci USA 107:19754–19759

Bange G, Murat G, Sinning I, Hurt E, Kressler D (2013) New twist to nuclear import: when two travel together. Commun Integr Biol 6:e24792

Beckmann R, Spahn CM, Eswar N, Helmers J, Penczek PA, Sali A, Frank J, Blobel G (2001) Architecture of the protein-conducting channel associated with the translating 80S ribosome. Cell 107:361–372

Ben-Shem A, Garreau de Loubresse N, Melnikov S, Jenner L, Yusupova G, Yusupov M (2011) The structure of the eukaryotic ribosome at 3.0 A resolution. Science 334:1524–1529

Bowman LH, Emerson CP Jr (1977) Post-transcriptional regulation of ribosome accumulation during myoblast differentiation. Cell 10:587–596

Busse I, Preisfeld A (2002) Unusually expanded SSU ribosomal DNA of primary osmotrophic euglenids: molecular evolution and phylogenetic inference. J Mol Evol 55:757–767

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 Bioinform 3:2

Chandramouli P, Topf M, Menetret JF, Eswar N, Cannone JJ, Gutell RR, Sali A, Akey CW (2008) Structure of the mammalian 80S ribosome at 8.7 A resolution. Structure 16:535–548

Chattopadhyay S, Pal S, Pal D, Sarkar D, Chandra S, Das Gupta C (1999) Protein folding in *Escherichia coli*: role of 23S ribosomal RNA. Biochim Biophys Acta 1429:293–298

Christian BE, Spremulli LL (2012) Mechanism of protein biosynthesis in mammalian mitochondria. Biochim Biophys Acta 1819:1035–1054

Ciriello G, Gallina C, Guerra C (2010) Analysis of interactions between ribosomal proteins and RNA structural motifs. BMC Bioinform 11(Suppl 1):S41

Clark CG, Tague BW, Ware VC, Gerbi SA (1984) Xenopus laevis 28S ribosomal RNA: a secondary structure model and its evolutionary and functional implications. Nucleic Acids Res 12:6197–6220

Cui XA, Zhang H, Palazzo AF (2011) p180 promotes the ribosome-independent localization of a subset of mRNA to the endoplasmic reticulum. PLoS Biol 10:e1001336

Davydova N, Streltsov V, Wilce M, Liljas A, Garber M (2002) L22 ribosomal protein and effect of its mutation on ribosome resistance to erythromycin. J Mol Biol 322:635–644

Diefenbach RJ, Diefenbach E, Douglas MW, Cunningham AL (2004) The ribosome receptor, p180, interacts with kinesin heavy chain, KIF5B. Biochem Biophys Res Commun 319:987–992

Du Z, Lee JK, Tjhen R, Li S, Pan H, Stroud RM, James TL (2005) Crystal structure of the first KH domain of human poly(C)-binding protein-2 in complex with a C-rich strand of human telomeric DNA at 1.7 A. J Biol Chem 280:38823–38830

Ellis JJ, Broom M, Jones S (2007) Protein-RNA interactions: structural analysis and functional classes. Proteins 66:903–911

Escobar JS, Glemin S, Galtier N (2011) GC-biased gene conversion impacts ribosomal DNA evolution in vertebrates, angiosperms, and other eukaryotes. Mol Biol Evol 28:2561–2575

Gabashvili IS, Gregory ST, Valle M, Grassucci R, Worbs M, Wahl MC, Dahlberg AE, Frank J (2001) The polypeptide tunnel system in the ribosome and its gating in erythromycin resistance mutants of L4 and L22. Mol Cell 8:181–188

Greber BJ, Boehringer D, Godinic-Mikulcic V, Crnkovic A, Ibba M, Weygand-Durasevic I, Ban N (2012) Cryo-EM structure of the archaeal 50S ribosomal subunit in complex with initiation factor 6 and implications for ribosome evolution. J Mol Biol 418:145–160

Gutell RR, Larsen N, Woese CR (1994) Lessons from an evolving rRNA: 16S and 23S rRNA structures from a comparative perspective. Microbiol Rev 58:10–26

Harris EH, Boynton JE, Gillham NW (1994) Chloroplast ribosomes and protein synthesis. Microbiol Rev 58:700–754

Hassouna N, Michot B, Bachellerie JP (1984) The complete nucleotide sequence of mouse 28S rRNA gene. Implications for the process of size increase of the large subunit rRNA in higher eukaryotes. Nucleic Acids Res 12:3563–3583

Herskovits AA, Bibi E (2000) Association of Escherichia coli ribosomes with the inner membrane requires the signal recognition particle receptor but is independent of the signal recognition particle. Proc Natl Acad Sci USA 97:4621–4626

Hsu D, Shih LM, Zee YC (1994) Degradation of rRNA in Salmonella strains: a novel mechanism to regulate the concentrations of rRNA and ribosomes. J Bacteriol 176:4761–4765

Jakel S, Gorlich D (1998) Importin beta, transportin, RanBP5 and RanBP7 mediate nuclear import of ribosomal proteins in mammalian cells. EMBO J 17:4491–4502

Jakovljevic J, Ohmayer U, Gamalinda M, Talkish J, Alexander L, Linnemann J, Milkereit P, Woolford JL Jr (2012) Ribosomal proteins L7 and L8 function in concert with six A(3) assembly factors to propagate assembly of domains I and II of 25S rRNA in yeast 60S ribosomal subunits. RNA 18:1805–1822

Kalpaxis DL, Karahalios P, Papapetropoulou M (1998) Changes in ribosomal activity of *Escherichia coli* cells during prolonged culture in sea salts medium. J Bacteriol 180:3114–3119

Klinge S, Voigts-Hoffmann F, Leibundgut M, Arpagaus S, Ban N (2011) Crystal structure of the eukaryotic 60S ribosomal subunit in complex with initiation factor 6. Science 334:941–948

Klukas CK, Dawid IB (1976) Characterization and mapping of mitochondrial ribosomal RNA and mitochondrial DNA in *Drosophila melanogaster*. Cell 9:615–625

Kondo J, Westhof E (2011) Classification of pseudo pairs between nucleotide bases and amino acids by analysis of nucleotide-protein complexes. Nucleic Acids Res 39:8628–8637

Kruger T, Zentgraf H, Scheer U (2007) Intranucleolar sites of ribosome biogenesis defined by the localization of early binding ribosomal proteins. J Cell Biol 177:573–578

Kruppa J, Sabatini DD (1977) Release of poly A(+) messenger RNA from rat liver rough microsomes upon disassembly of bound polysomes. J Cell Biol 74:414–427

Kuhn P, Weiche B, Sturm L, Sommer E, Drepper F, Warscheid B, Sourjik V, Koch HG (2011) The bacterial SRP receptor, SecA and the ribosome use overlapping binding sites on the SecY translocon. Traffic 12:563–578

Kuzoff RK, Sweere JA, Soltis DE, Soltis PS, Zimmer EA (1998) The phylogenetic potential of entire 26S rDNA sequences in plants. Mol Biol Evol 15:251–263

Kyte J, Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. J Mol Biol 157:105–132

Lande MA, Adesnik M, Sumida M, Tashiro Y, Sabatini DD (1975) Direct association of messenger RNA with microsomal membranes in human diploid fibroblasts. J Cell Biol 65:513–528

Larsson SL, Nygard O (2001) Proposed secondary structure of eukaryote specific expansion segment 15 in 28S rRNA from mice, rats, and rabbits. Biochemistry 40:3222–3231

Leffers H, Andersen AH (1993) The sequence of 28S ribosomal RNA varies within and between human cell lines. Nucleic Acids Res 21:1449–1455

Liu Q, Greimann JC, Lima CD (2006) Reconstitution, activities, and structure of the eukaryotic RNA exosome. Cell 127:1223–1237

Mallardo M, Deitinghoff A, Muller J, Goetze B, Macchi P, Peters C, Kiebler MA (2003) Isolation and characterization of Staufen-containing ribonucleoprotein particles from rat brain. Proc Natl Acad Sci USA 100:2100–2105

Menetret JF, Hegde RS, Aguiar M, Gygi SP, Park E, Rapoport TA, Akey CW (2008) Single copies of Sec61 and TRAP associate with a nontranslating mammalian ribosome. Structure 16:1126–1137

Michot B, Bachellerie JP, Raynal F (1982) Sequence and secondary structure of mouse 28S rRNA 5′ terminal domain. Organisation of the 5.8S–28S rRNA complex. Nucleic Acids Res 10:5273–5283

Moore PB (2009) The ribosome returned. J Biol 8:8

Morrow MW, Brodsky JL (2001) Yeast ribosomes bind to highly purified reconstituted Sec61p complex and to mammalian p180. Traffic 2:705–716

Nissen P, Hansen J, Ban N, Moore PB, Steitz TA (2000) The structural basis of ribosome activity in peptide bond synthesis. Science 289:920–930

Nunn GB, Theisen BF, Christensen B, Arctander P (1996) Simplicity-correlated size growth of the nuclear 28S ribosomal RNA D3 expansion segment in the crustacean order Isopoda. J Mol Evol 42:211–223

Nwagwu M, Nana M (1980) Ribonucleic acid synthesis in embryonic chick muscle, rates of synthesis and half-lives of transfer and ribosomal RNA species. J Embryol Exp Morphol 56:253–267

Ohashi S, Koike K, Omori A, Ichinose S, Ohara S, Kobayashi S, Sato TA, 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

Osborne AR, Rapoport TA, van den Berg B (2005) Protein translocation by the Sec61/SecY channel. Annu Rev Cell Dev Biol 21:529–550

Parker SL, Parker MS, Sah R, Sallee F (2005) Angiogenesis and rhodopsin-like receptors: a role for N-terminal acidic residues? Biochem Biophys Res Commun 335:983–992

Parker MS, Balasubramaniam A, Parker SL (2012) On the segregation of protein ionic residues by charge type. Amino Acids 43:2231–2247

Pearson WR (2000) Flexible sequence similarity searching with the FASTA3 program package. Methods Mol Biol 132:185–219

Plafker SM, Macara IG (2002) Ribosomal protein L12 uses a distinct nuclear import pathway mediated by importin 11. Mol Cell Biol 22:1266–1275

Plantinga MJ, Korennykh AV, Piccirilli JA, Correll CC (2011) The ribotoxin restrictocin recognizes its RNA substrate by selective engagement of active site residues. Biochemistry 50:3004–3013

Potter MD, Nicchitta CV (2002) Endoplasmic reticulum-bound ribosomes reside in stable association with the translocon following termination of protein synthesis. J Biol Chem 277:23314–23320

Prinz A, Behrens C, Rapoport TA, Hartmann E, Kalies KU (2000) Evolutionarily conserved binding of ribosomes to the translocation channel via the large ribosomal RNA. EMBO J 19:1900–1906

Pyhtila B, Zheng T, Lager PJ, Keene JD, Reedy MC, Nicchitta CV (2008) Signal sequence- and translation-independent mRNA localization to the endoplasmic reticulum. RNA 14:445–453

Qu X, Wen JD, Lancaster L, Noller HF, Bustamante C, Tinoco I Jr (2011) The ribosome uses two active mechanisms to unwind messenger RNA during translation. Nature 475:118–121

Rout MP, Blobel G, Aitchison JD (1997) A distinct nuclear import pathway used by ribosomal proteins. Cell 89:715–725

Sabatini DD, Tashiro Y, Palade GE (1966) On the attachment of ribosomes to microsomal membranes. J Mol Biol 19:503–524

Sameshima M, Liebhaber SA, Schlessinger D (1981) Dual pathways for ribonucleic acid turnover in WI-38 but not in I-cell human diploid fibroblasts. Mol Cell Biol 1:75–81

Shasmal M, Sengupta J (2012) Structural diversity in bacterial ribosomes: mycobacterial 70S ribosome structure reveals novel features. PLoS One 7:e31742

Stage DE, Eickbush TH (2007) Sequence variation within the rRNA gene loci of 12 Drosophila species. Genome Res 17:1888–1897

Stephens SB, Dodd RD, Brewer JW, Lager PJ, Keene JD, Nicchitta CV (2005) Stable ribosome binding to the endoplasmic reticulum enables compartment-specific regulation of mRNA translation. Mol Biol Cell 16:5819–5831

Sweeney R, Fan Q, Yao MC (1996) Antisense ribosomes: rRNA as a vehicle for antisense RNAs. Proc Natl Acad Sci USA 93:8518–8523

Tai LR, Chou CW, Lee IF, Kirby R, Lin A (2013) The quantitative assessment of the role played by basic amino acid clusters in the nuclear uptake of human ribosomal protein L7. Exp Cell Res 319:367–375

Terribilini M, Sander JD, Lee JH, Zaback P, Jernigan RL, Honavar V, Dobbs D (2007) RNABindR: a server for analyzing and predicting RNA-binding sites in proteins. Nucleic Acids Res 35:W578–W584

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

Uchiumi T, Terao K, Ogata K (1983) Ribosomal proteins cross-linked to 28-S and 18-S rRNA separated by sedimentation after ultraviolet irradiation of rat-liver ribosomes. Eur J Biochem 132:495–499

Ueno T, Kaneko K, Sata T, Hattori S, Ogawa-Goto K (2012) Regulation of polysome assembly on the endoplasmic reticulum by a coiled-coil protein, p180. Nucleic Acids Res 40:3006–3017

Wakeman JA, Maden BE (1989) 28 S ribosomal RNA in vertebrates. Locations of large-scale features revealed by electron microscopy in relation to other features of the sequences. Biochem J 258:49–56

Wang L, Brown SJ (2006) BindN: a web-based tool for efficient prediction of DNA and RNA binding sites in amino acid sequences. Nucleic Acids Res 34:W243–W248

Wanker EE, Sun Y, Savitz AJ, Meyer DI (1995) Functional characterization of the 180-kD ribosome receptor in vivo. J Cell Biol 130:29–39

Ware VC, Tague BW, Clark CG, Gourse RL, Brand RC, Gerbi SA (1983) Sequence analysis of 28S ribosomal DNA from the amphibian Xenopus laevis. Nucleic Acids Res 11:7795–7817

Whitesides GM (2001) The once and future nanomachine. Sci Am 285:78–83

Zirbel CL, Sponer JE, Sponer J, Stombaugh J, Leontis NB (2009) Classification and energetics of the base-phosphate interactions in RNA. Nucleic Acids Res 37:4898–4918