Journal of Molecular Evolution
https://doi.org/10.1007/s00239-018-9869-1

ORIGINAL ARTICLE

Circular Permutation Obscures Universality of a Ribosomal Protein

Nicholas A. Kovacs[1] · Petar I. Penev[2] · Amitej Venapally[1] · Anton S. Petrov[1] · Loren Dean Williams[1,2]

Received: 23 August 2018 / Accepted: 28 September 2018

Abstract

Functions, origins, and evolution of the translation system are best understood in the context of unambiguous and phylogenetically based taxonomy and nomenclature. Here, we map ribosomal proteins onto the tree of life and provide a nomenclature for ribosomal proteins that is consistent with phylogenetic relationships. We have increased the accuracy of homology relationships among ribosomal proteins, providing a more informative picture of their lineages. We demonstrate that bL33 (bacteria) and eL42 (archaea/eukarya) are homologs with common ancestry and acute similarities in sequence and structure. Their similarities were previously obscured by circular permutation. The most likely mechanism of permutation between bL33 and eL42 is duplication followed by fusion and deletion of both the first and last β-hairpins. bL33 and eL42 are composed of zinc ribbon protein folds, one of the most common zinc finger fold-groups of, and most frequently observed in translation-related domains. Bacterial-specific ribosomal protein bL33 and archaeal/eukaryotic-specific ribosomal protein eL42 are now both assigned the name of uL33, indicating a universal ribosomal protein. We provide a phylogenetic naming scheme for all ribosomal proteins that is based on phylogenetic relationships to be used as a tool for studying the systemics, evolution, and origins of the ribosome.

Keywords Protein evolution · Zinc ribbon · Zinc finger · Ribosome · Tree of life · Translation

Abbreviations

DCC	Decoding center
PTC	Peptidyl transferase center
SSU	Small subunit
LSU	Large subunit
rRNA	Ribosomal RNA
rProtein	Ribosomal protein
PASE	Pairing adjusted sequence entropy
E. coli	*Escherichia coli*
EsCo	*Escherichia coli*
T. Thermophilus	*Thermus thermophilus*
ThTh	*Thermus thermophilus*
H. marismortui	*Haloarcula marismortui*
HaMa	*Haloarcula marismortui*
P. furiosus	*Pyrococcus furiosus*
PyFu	*Pyrococcus furiosus*
S. cerevisiae	*Saccharomyces cerevisiae*
SaCe	*Saccharomyces cerevisiae*
T. thermophila	*Tetrahymena thermophila*
TeTh	*Tetrahymena thermophila*
D. melanogaster	*Drosophila melanogaster*
DrMe	*Drosophila melanogaster*
H. sapiens	*Homo sapiens*
HoSa	*Homo sapiens*
P. falciparum	*Plasmodium falciparum*
PlFa	*Plasmodium falciparum*
T. brucei	*Trypanosoma brucei*
TrBr	*Trypanosoma brucei*

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00239-018-9869-1) contains supplementary material, which is available to authorized users.

✉ Anton S. Petrov
anton.petrov@biology.gatech.edu

✉ Loren Dean Williams
loren.williams@chemistry.gatech.edu

Nicholas A. Kovacs
Nicholas.kovacs@gatech.edu

Petar I. Penev
ppenev@gatech.edu

Amitej Venapally
v.amitej@gatech.edu

[1] School of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta, GA 30332-0400, USA

[2] School of Biological Sciences, Georgia Institute of Technology, Atlanta, GA 30332-0400, USA

Published online: 10 October 2018

Introduction

Ribosomes are ribonucleoprotein particles (Wimberly et al. 2000; Ben-Shem et al. 2010; Anger et al. 2013; Hashem et al. 2013; Amunts et al. 2014) that bring mRNAs, tRNAs, amino acids, initiation factors, and elongation factors together in the synthesis of coded proteins. The functional centers of the ribosome, called the Decoding Center (DCC) and the Peptidyl Transferase Center (PTC), are composed exclusively of ribosomal RNA (rRNA) (Noller et al. 1992; Ban et al. 2000; Selmer et al. 2006). The DCC is located in the small subunit (SSU) and the PTC is located in the large subunit (LSU). Ribosomal proteins (rProteins) facilitate the folding of rRNA and enable catalysis (Noller et al. 1992).

Structures and sequences of ribosomes from across the extant tree of life (Harris et al. 2003; Koonin 2003; Charlebois and Doolittle 2004) provide information on phylogeny (Hug et al. 2016) and on how rRNA and rProtein conformations, interactions, and functions originated and evolved over time (Lupas and Alva 2017; Kovacs et al. 2017). Information is provided by variation of sequence, conformation and molecular interactions in ribosomes from different species, and variation within a given ribosome. Here, we incorporate structure-based approaches to discover new phylogenetic relationships among rProteins. Historically, these relationships are determined by sequence comparisons. However, additional information is afforded by the exploding number of ribosomal structures. Advantages of combined approaches are seen in the orthogonal nature of structural and sequence information and the greater conservation of structure over sequence (Illergard et al. 2009). Structure can provide accurate information on deep ancestry. For example, the three-dimensional structure of rProtein uL14 (Davies et al. 1996) indicates an absence of homology to uL30, in contrast to interpretations of sequence alignments (Davies et al. 1994).

We demonstrate here that rProteins bL33 (found in bacteria) and eL42 (found in archaea and eukarya, also known as L44e or L36A), are homologs with acute similarities in sequence and structure. The sequence similarities and homology of these two rProteins were obscured by an unrecognized circular permutation that alters the connectivity between secondary structural elements, and changes the ordering of blocks of amino acid residues. Circular permutation has been documented previously in carbohydrate-binding proteins (Cunningham et al. 1979), saposins (Ponting and Russell 1995), and zinc ribbons, which are common in rProteins (Krishna et al. 2003). Circular permutations have not been observed previously in rProteins to our knowledge.

rProteins bL33 and eL42 are zinc ribbons (Fig. 1), one of the most common of the eight fold-groups of zinc fingers. Zinc ribbons are composed of a scaffold of two amphipathic β-hairpins linked by a zinc ion. A hydrophobic core is stabilized indirectly by the zinc ion, which is chelated by 4 amino acid sidechains, usually cysteines or histidines on the zinc-knuckles found on each of the loops of the β-hairpins (Krishna et al. 2003; D'Abrosca et al. 2016). The zinc-knuckles are the most conserved portions of these proteins, with the remainder of the protein exhibiting sequence variability (Krishna et al. 2003).

Zinc ribbons are frequently observed in translation-related proteins (Krishna et al. 2003). Here, zinc ribbon proteins bL33 and eL42 [the names used here are from the recent scheme by Ban et al. (2014), unless otherwise specified] are recognized as homologs with common ancestry. This increase in accuracy of phylogenetic relationships of rProteins helps provide a more informative picture of how rProteins evolved and are partitioned among the three domains of life. Here we map rProteins onto the tree of life and provide a nomenclature scheme for rProteins that is consistent with phylogenetic relationships. We provide a translation table for conversion of various non-evolutionary naming schemes to the phylogenetic rProtein naming scheme (See Supplementary Table S1, Additional File 1).

Results

Location and Fold of rProteins bL33 and eL42

rProtein bL33 in the bacterial ribosome and eL42 in the archaeal and eukaryotic ribosomes occupy the same position in the ribosome and assume the same global fold (Fig. 1). The similarities are apparent upon global superimpositions of three-dimensional structures of LSUs of species from the three primary branches of the tree of life. rProteins bL33 and eL42 are located on the surface of the LSU between the central protuberance and the L1 stalk.

The superimposition used here that establishes the common locations and folds of bL33 and eL42 use rRNA only and is not based on the atomic coordinates of rProteins. The superimposition uses 277 nucleotides that form and surround the PTC. This rRNA-based superimposition reveals similarities in positions of universal rProteins and the majority of LSU rRNA, as shown previously (Petrov et al. 2014, 2015; Kovacs et al. 2017). Positions of backbone atoms of universal rRNA nucleotides and universal rProtein amino acids deviate by less than an angstrom using this method. Many divalent cations also superimpose well.

Here we use a dataset of LSU structures that spans the tree of life, including two bacterial, two archaeal, and six eukaryotic ribosomes. Superimpositions of this structural dataset reveal that many polypeptide backbone atoms of rProteins bL33 and eL42 occupy near identical positions

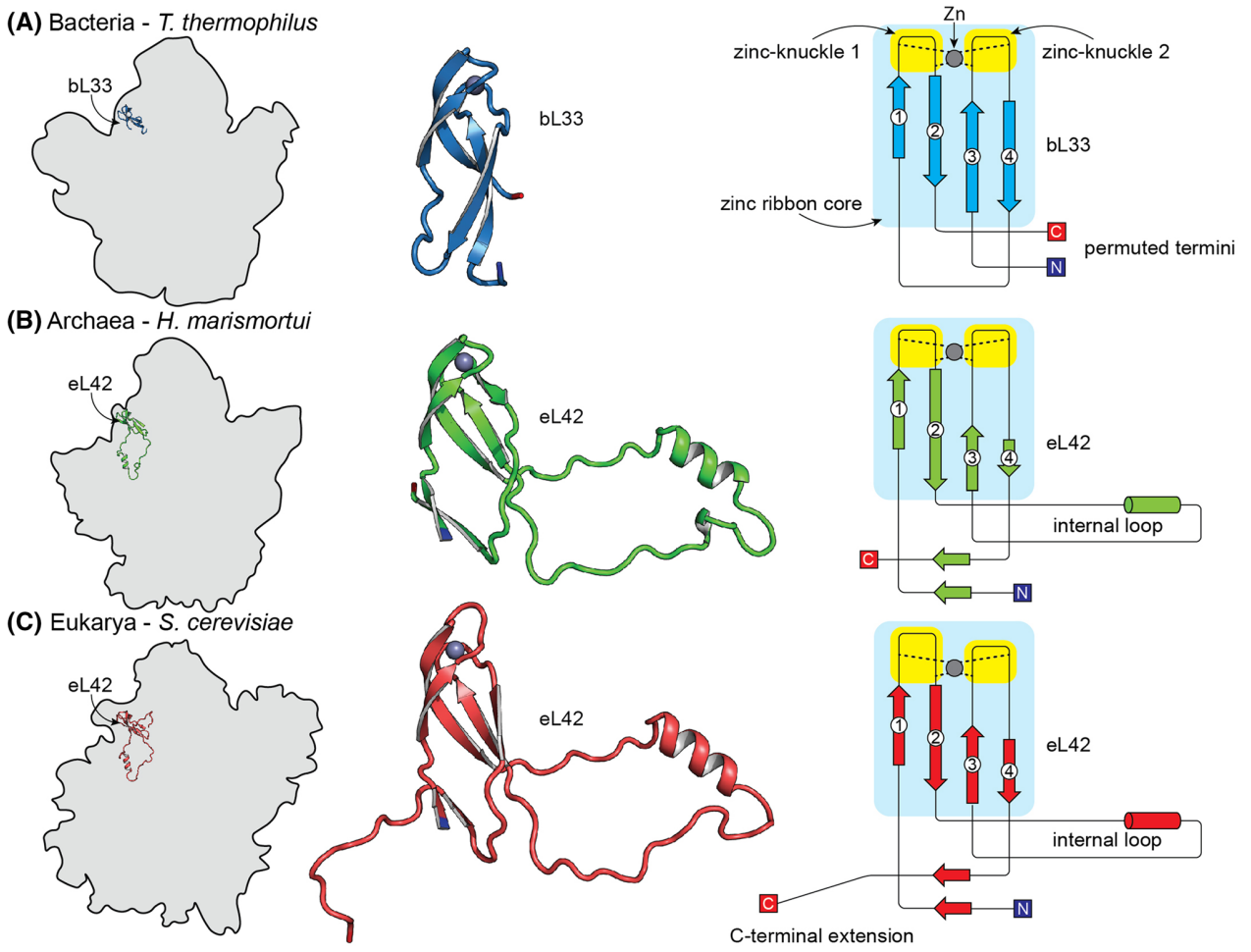

Fig. 1 Structures of rProteins bL33 and eL42 from each of the three domains of life reveal homology and evidence of common ancestry. **a** bL33 from the bacterium *T. thermophilus*. **b** eL42 from the archaeon *H. marismortui*, and **c** eL42 from the eukaryote *S. cerevisiae*. In the left panel, LSU rRNA is depicted by a contour line and bL33 and eL42 are drawn in 3-dimensional cartoon representation. In the middle panel, bL33 and eL42 are in cartoon representation. In the right panel are 2-dimensional protein structure cartoons of bL33 and eL42 with β-strands numbered using eL42 topology. In the middle and right panels, zinc atoms are indicated by gray spheres. Three-dimensional images are made from a common frame of reference provided by superimposition of rRNA

relative to the rRNA. The LSU structures in this dataset are bacterial [*Escherichia coli* (PDB ID 4V9D) (Dunkle et al. 2011) and *Thermus thermophilus* (thermophilic, PDB ID 1VY4) (Polikanov et al. 2014)]; archaeal [*Haloarcula marismortui* (halophilic, PDB ID 4V9F) (Gabdulkhakov et al. 2013) and *Pyrococcus furiosus* (thermophilic, anaerobic, PDB ID 4V6U) (Armache et al. 2013)]; and eukaryotic [unicellular fungi *Saccharomyces cerevisiae* (baker's yeast, PDB ID 4V88) (Ben-Shem et al. 2011); protozoa *Tetrahymena thermophila* (non-parasitic, PDB ID 4V8P) (Klinge et al. 2011); animals *Drosophila melanogaster* (fruit fly, PDB ID 4V6W) (Anger et al. 2013) and *Homo sapiens* (human, PDB ID 4UG0) (Khatter et al. 2015); and protozoan human pathogens *Plasmodium falciparum* (malaria, PDB ID 3JBN) (Sun et al. 2015) and *Trypanosoma brucei* (sleeping sickness, PDB ID 4V8M) (Hashem et al. 2013)].

bL33/eL42 Local Superimposition

Local superimpositions, using atoms within the rProteins only, support a model in which rProteins bL33 and eL42 share a common core of polypeptide (Fig. 1). This common core consists of the subset of amino acids whose backbone atoms are conserved in location relative to rRNA or relative to other rProtein backbone atoms (Table 1 and See Supplementary Table S2, Additional File 1). The common core defined by the superimpositions reveals that rProteins bL33 and eL42 are very similar to each other

Table 1 Defining the common core of bL33 and eL42

Species	Domain of life	Total length (aa)	Common core amino acids
E. coli	Bacteria	52	3–52
T. thermophilus	Bacteria	54	3–52
H. marismortui	Archaea	92	1–25, 63–87
P. furiosus	Archaea	94	1–25, 63–83, 85–88
T. thermophila	Eukarya	104	2–26, 64–88
S. cerevisiae	Eukarya	106	2–13, 16–28, 66–90
D. melanogaster	Eukarya	104	2–26, 64–88
H. sapiens	Eukarya	106	2–26, 64–74, 77–90
P. falciparum	Eukarya	96	2–13, 15–27, 65–89
T. brucei	Eukarya	106	2–13, 16–28, 66–90

As inferred by superimpositions and multiple sequence alignments

but are distinguished by circular permutation (Fig. 1). The common cores have different topologies; the strand termini do not superimpose.

The bL33/eL42 common core consists of 50 amino acids (Table 1) including a zinc ribbon (Krishna et al. 2003) containing two zinc-knuckles, formed by the loops of two β-hairpins (Fig. 1). The bL33/eL42 common core gives pairwise RMSDs of backbone atoms ranging from 0.50 to 3.33 Å in our structural dataset (See Supplementary Table S3, Additional File 1). The bL33/eL42 common core omits amino acids that are not structurally conserved in all 10 structures of the dataset (Table 1). Omitted peptide regions include (i) an internal loop of eL42 that is partially non-canonical and partially α-helical, and (ii) a structurally non-canonical C-terminal extension of eL42 present in eukaryotes only, and (iii) variable insertions of length from one to two amino acids in many bL33 and eL42 rProteins (Table 1).

The degree of similarity in the structures of the bL33 and eL42 common cores is consistent with the canonical tree of life, in which eukarya is more closely related to archaea than to bacteria. The lowest pairwise RMSDs, indicating the smallest differences in common core structure, are seen when comparing rProteins within a domain. The lowest pairwise RMSDs between the domains of life are between eukaryotic and archaeal eL42 rProteins, which range from 1.10 to 1.83 Å. The greatest pairwise RMSDs are between the bL33 common cores of bacteria and the eL42 common cores of archaea and eukarya, which range between 2.77 and 3.33 Å (See Supplementary Table S3, Additional File 1). Our results are in agreement with Grishin's structurally inferred evolutionary relationships of protein folds within the ECOD database (Cheng et al. 2014). In that work, the folds of bL33 and eL42 are closely related.

bL33/eL42 Sequence Alignment and Circular Permutation

A multiple sequence alignment (Figs. 2, 4) extracted from the superimposition supports the circular permutation of bL33 to give eL42 (and vice versa). The β-strands of eL42 are in topological order, N-terminus, then β-strand 1, β-strand 2, β-strand 3, β-strand 4, then C-terminus. By contrast, the termini of bL33 are located between β-strands 2 and 3. Homology is evident in the sequence alignment of bL33 and eL42 only after correction for circular permutation (Figs. 2, 4). We denote the sequence of bL33 that has been corrected for circular permutation as $bL33^{CP}$. eL42 aligns with $bL33^{CP}$ but not with bL33. The site of permutation in bL33 to form $bL33^{CP}$ is between amino acids 28 and 29 in *E. coli* and between amino acids 29 and 30 of *T. thermophilus*. The site of permutation is within the loop that links the two β-hairpins (Figs. 1, 2). An alternative permutation, of eL42 instead of bL33, is consistent with the data.

The zinc-binding sub-sites of $bL33^{CP}$ and eL42 superimpose in structure and align in the sequence. Zinc atoms are coordinated by four cysteines in the $bL33^{CP}$/eL42. Zinc is absent from some of the lower-resolution archaeal and eukaryotic LSU structures possibly due to errors in the structural models. The four conserved cysteines are present in all members of our structural dataset except for *E. coli*. Cadmium ions were contained in the crystallization solution and have replaced the zinc ions in the structure of *H. marismortui*.

Structure-Enhanced Multiple Sequence Alignments

In a multiple sequence alignment (MSA) of $bL33^{CP}$ and eL42, the zinc-knuckles and rRNA-binding sites appear universally conserved—across the tree of life. An alignment of $bL33^{cp}$ /eL42 sequences from 149 species that sparsely and efficiently sample the tree of life [the Sparse and Efficient Representation of Extant Biology (SEREB) database] (Bernier et al. 2018) was assembled using PROMALS3D (Pei et al. 2008), which incorporates both sequence and structural information into multiple sequence alignments (See Supplementary Table S4, Additional File 1). The PROMALS3D alignments show that zinc-knuckles have consensus sequences of -CTEC- in zinc-knuckle 1, and -CPXC- in zinc-knuckle 2. The zinc-knuckles of the bacterial proteins only align to the archaeal and eukaryotic proteins once the circular permutation is incorporated (See Supplementary Figures S4 and S5, Additional File 1). Sequences of bacterial bL33 and archaeal eL42 were obtained from Yutin et al., and from Uniprot for eukaryotic eL42 (Yutin et al. 2012; The UniProt Consortium 2017).

In several bacterial species, N-terminal extensions of bL33 are observed. After correction for the circular

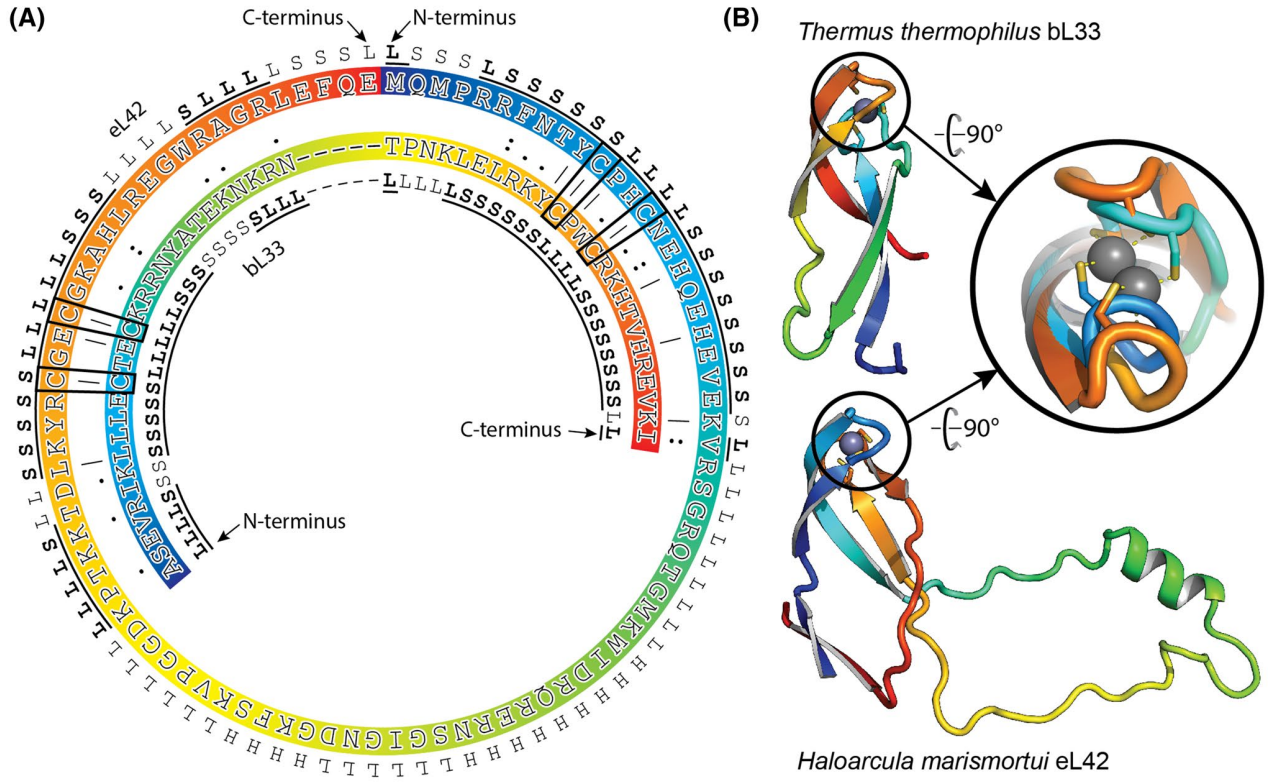

Fig. 2 Circular permutation between bacterial-specific rProtein bL33 and archaeal- and eukaryotic-specific rProtein eL42. A rainbow color gradient is applied to both the sequence and 3-dimensional structures of bL33 in *T. thermophilus* and eL42 of *H. marismortui* in **a** and **b** to illustrate the circular permutation of these proteins. **a** Going from the center to the periphery of the circle: secondary structure of bL33 (*L*-loop, *H*-Helix, *S*-Sheet), sequence of bL33, MUSCLE sequence conservation, sequence of eL42, secondary structure of eL42. Common secondary structural assignments are in bold and underlined. Conserved zinc-coordinating cysteine residues are highlighted in black boxes. **b** 3-dimensional structures of bL33 (top) and eL42 (bottom), with an inlet showing the zinc-binding site

permutation (i.e., in bL33^{cp}) these extensions align with parts of the internal loop of eL42. These bacterial species include *Deinococcus radiodurans, Propionibacterium acnes, Dehalococcoides ethenogenes* and *Symbiobacterium thermophilum* (See Supplementary Figure S1, Additional File 2). Although several structures of *D. radiodurans* ribosomes are contained within the PDB, only one of them (PDB ID 5DM6) resolves a portion of the N-terminal extension. The observed segment of the extension roughly follows the path of the internal loop of eL42.

Protein–Protein and Protein–tRNA Interaction Substitutions in bL33/eL42

The ribosomes of bacteria, archaea, and eukarya each contain rProteins that interact with the E-site tRNA. These rProteins, bL28 in bacteria and eL42 in archaea and eukarya, are not homologs and their E-site tRNA interfaces are not conserved. The common core of bL33/eL42 is located on the surface of the ribosome as noted above. In archaea and eukarya, the E-site tRNA interacts with the internal loop of eL42 (Fox 2010). The internal loop of eL42 (Fig. 3) forms a ring with an interior cavity. The two cytosines of the CCA-tail of the E-site tRNA interact with the interior surface of the cavity, whereas the 3′ adenine of the CCA-tail intercalates between two nucleobases of Helix 88 in LSU rRNA and base pairs with an unpaired nucleotide on the opposite strand of the rRNA helix (See Supplementary Figure S6, Additional File 1). The sequence of the internal loop of eL42 is highly conserved within eukaryotes but not within archaea (See Supplementary Figures S2 and S3, Additional File 1). In bacteria, the nearby bacterial-specific rProtein bL28 contains a β-hairpin extension that occupies the same region as the eL42 insertion, interacting with the bacterial E-site tRNA via amino acid residues located on the end of the β-hairpin. Although the specific E-site tRNA interactions in bacteria and archaea/eukarya are different (Fig. 3 and See Supplementary Figure S6, Additional File 1) and bL28 and eL42 do not share any sequence or structural homology; the internal loop of eL42 and the β-hairpin extension of bL28 appear to fulfill similar functions.

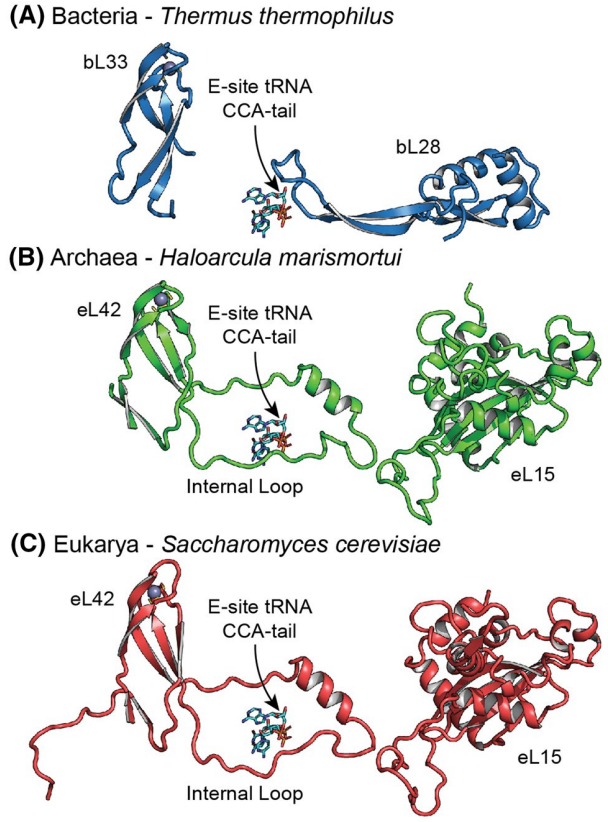

Fig. 3 Elaborations to the bL33/eL42 common core in archaea and eukarya account for differences in protein–protein interactions and protein–tRNA interactions. **a** rProteins bL33 and bL28 of *T. thermophilus*. **b** rProteins eL42 and eL15 of *H. marismortui*. **c** rProteins eL42 and eL15 of *S. cerevisiae*. The CCA-tail of E-site tRNA is colored cyan and is from the globally superimposed *T. thermophilus* ribosome

A segment of the eL42 internal loop composed of an α-helix and its adjacent amino acid residues is buried within a cleft between the central protuberance and the L1 stalk of the LSU 23S. This segment is in contact with rProtein eL15 (Fig. 3). The eL15–eL42 interface consists of two amino acid residues of eL42 in non-canonical conformation that are adjacent to the α-helix (Figs. 3, 4). The eL15–eL42 interface in eL42 is not well conserved in archaea but is well conserved in eukarya (See Supplementary Figures S2 and S3, Additional File 1).

In eukarya, the C-terminal extension of eL42 extends to the tip of the central protuberance and forms a eukaryotic-specific interaction with rProtein uL5 (See Supplementary Figure S7, Additional File 1). No amino acids belonging to any rProtein in the archaeal and bacterial structures analyzed are located in the same region as the C-terminal extension of eukaryotic eL42; this interaction is entirely absent in archaea and bacteria.

rRNA Interactions of bL33/eL42

rRNA–rProtein interactions are consistent with homology of bL33^{CP} and eL42. rRNA sequence and interactions are conserved in the first β-hairpin of the bL33^{CP}/eL42 zinc ribbon (Table 2). Amino acid residues 6–11 form β-strand 1, residues 11–14 form zinc-knuckle 1 (numbering and secondary structure assignments are from *H. marismortui*, and Cys11 is in β-strand 1 and zinc-knuckle 1). β-strand 1 forms hydrogen bonds with rRNA helices 31, 83, and 86. Zinc-knuckle 1 forms hydrogen bonds with rRNA helices 86 and 88 (See Supplementary Figure S8, Additional File 1). Lysine/arginine 8 and tyrosine/phenylalanine 10 are conserved and each forms hydrogen bonds with rRNA backbone atoms. Lysine/arginine is a consensus at position 15, interacting with either a base or the backbone of rRNA (Fig. 4; Table 2).

Amino acids 15–24 form β-stand 2 of bL33^{CP}/eL42. Histidine 17, which is conserved in all ten species of our structural dataset and is among the most highly conserved residues, forms conserved hydrogen bonds with rRNA backbone atoms. The other amino acids in β-strand 2 interact with rRNA helices 31 and 86 (See Supplementary Figure S8, Additional File 1). β-strand 1, zinc-knuckle 1, and β-strand

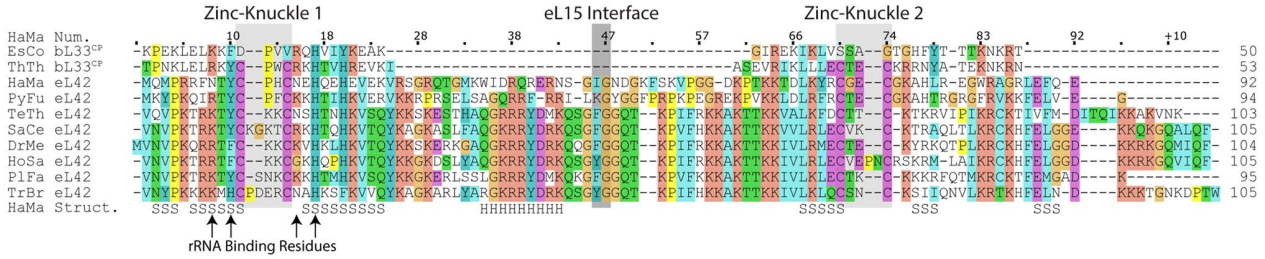

Fig. 4 Sequence alignment of ten organisms extracted from the superimpositions of bL33 (circularly permutated) and eL42. Zinc-knuckles are highlighted in light gray, while the interface with eL15 is highlighted in dark gray. EsCo = *E. coli*, ThTh = *T. thermophilus*, HaMa = *H. marismortui*, PyFu = *P. furiosus*, SaCe = *S. cerevisiae*, TeTh = *T. thermophila*, DrMe = *D. melanogaster*, HoSa = *H. sapiens*, PlFa = *P. falciparum*, TrBr = *T. brucei*. Amino acid residue numbering, and helix (H) and sheet (S) secondary structure annotations are from *H. marismortui*

Table 2 Conserved hydrogen bonds of bL33/eL42 (i.e., uL33) with rRNA

Amino acid residue number[a]	Organism	Amino acid number	Amino acid atom	RNA atom	RNA number (helix)	Distance
8	*T. thermophilus*	Arg37	NE	OP1	U2344 (H83)	2.9
			NH2	O3′	G2372 (H87)	3.6
			NH2	O2′	G2372 (H87)	3.9
	H. marismortui	Asn8	ND2	OP1	U2378 (H83)	3.5
	S. cerevisiae	Lys9	NZ	OP1	U2713 (H83)	4.2[b]
10	*T. thermophilus*	Tyr39	OH	O2′	G2370 (H86)	3.2
	H. marismortui	Tyr10	OH	O2′	G2407 (H86)	3.4
			OH	N4	C735 (H31)	3.9
	S. cerevisiae	–	–	–	–	–
15	*T. thermophilus*	Arg44	NH2	N9	A643 (H31)	3.4
			NH1	N7	A643 (H31)	3.6
	H. marismortui	–	–	–	–	–
	S. cerevisiae	Arg18	NH2	OP1	U759 (H31)	3.4
			NH1	OP1	C758 (H31)	3.7
17	*T. thermophilus*	His46	ND1	O2′	G2371 (H86)	2.7
			NE2	OP1	G2345 (H86)	3.4
	H. marismortui	His17	ND1	O2′	A2408 (H86)	2.8
			NE2	OP1	G2379 (H86)	2.9
	S. cerevisiae	His20	ND1	O2′	C2741 (H86)	2.5
			NE2	OP1	G2714 (H86)	2.8

[a]Nucleotide numbering from *H. marismortui*

[b]This hydrogen bond angle is less than 90° between heteroatoms

2 contact rRNA helices 83 and 86 of low pairing adjusted sequence entropy (PASE), and rRNA helices 31 and 88 of high PASE (See Supplementary Figures S8 and S9, Additional File 1) (Bernier et al. 2018).

β-Strand 3 of the common core of bL33^{CP}/eL42, containing amino acid residues 67–71, forms hydrogen bonds with β-strands 2 and 4, and makes minimal, non-specific, non-conserved interactions with Helix 88 of the rRNA. Amino acid residues 71–74 form zinc-knuckle 2 and are solvent exposed (Cys71 is in β-strand 3 and zinc-knuckle 2). The fourth β-strand containing amino acid residues 75–78 has the lowest sequence conservation of the bL33^{CP}/eL42 common core and makes non-specific interactions with rRNA Helix 88 which has a high PASE (See Supplementary Figure S9, Additional File 1) (Bernier et al. 2018).

In summary, the β-hairpin consisting of β-strand 1, zinc-knuckle 1, and β-strand 2 forms conserved hydrogen bonds with helices rRNA 31, 83, 86, and 88. rRNA Helices 83 and 86 have low PASE. The β-hairpin consisting of β-strand 3, zinc-knuckle 2, and β-strand 4 forms non-specific hydrogen bonds with Helix 88 which has high PASE. In bacteria, rRNA Helices 31, 82, 83, 86, and 88 are within 4 Å of bL33. In archaea, rRNA Helices 11, 13, 21, 31, 68, 74, 75, 82, 83, 86, 87, and 88 are within 4 Å of eL42. Eukaryotic eL42 contacts the same helices as archaeal eL42 despite having the eukaryotic-specific C-terminal extension (See Supplementary Figure S8, Additional File 1).

Discussion

The Zinc in Zinc Ribbons is Vestigial in the Bacterial Ribosome

Zinc ribbons are contained in many rProteins (See Supplementary Table S5, Additional File 1). Zinc binding to these proteins may be vestigial; amino acid substitutions of the coordinating amino acids that abrogate zinc binding do not alter global protein structure or ribosomal function (Dresios et al. 2005). Many bacteria have multiple zinc ribbon rProtein paralogs: some bind zinc and some do not. Phylogenetic analysis suggests that the zinc-binding rProteins are ancestors of non-zinc-binding paralogs (Makarova et al. 2001). rProtein paralogs arise from horizontal gene transfer and differential gene loss, a surprising result considering the importance and conservation of rProteins.

Our multiple sequence alignments (See Supplementary Figures S1-S3, Additional File 1) support the conclusion of Koonin that zinc-binding motifs are conserved in archaeal and eukaryotic rProteins but not in bacterial rProteins

(Makarova et al. 2001). Our analysis here suggests a structural basis for the differences. The archaeal/eukaryotic eL42 rProtein contains destabilizing insertions that in the absence of zinc would appear to abrogate folding. The importance of zinc to stability of zinc ribbons is highlighted by thermophilic bacterial and archaeal organisms which are enriched in zinc-binding proteins, suggesting that zinc is used to stabilize rProteins at high temperatures (McCall et al. 2000).

Circular Permutation

bL33 and eL42 are closely related by sequence and structure, after correcting for the circular permutation, and are clearly homologs. This relationship warrants a change to a common name of uL33. Bacterial uL33 and archaeal/eukaryotic uL33 are composed of the zinc ribbon protein fold and are related by circular permutation. The homology of these proteins is clear in structure-assisted approaches and is supported by the method of TopMatch (Sippl and Wiederstein 2012). Three mechanistic models have been proposed to achieve circularly permuted proteins (Bliven and Prlić 2012).

1. *Duplication/deletion mechanism* Adjacent duplication of genes is followed by partial deletion of motifs or domains on each of the termini. For example, ABC is duplicated to ABCABC. Then A|BCA|BC is truncated at "|" to form BCA. Evidence for this mechanism is the intermediates ABCA or BCABC.
2. *Independent fusion/fission mechanism* In fusion, two unrelated genes coding for multi-domain proteins combine in alternate combinations to yield proteins composed of the same constituents but in different orders. For example, AB and CD fuse to form ABCD and CDAB. In fission, a gene undergoes fission and reassembles in a different order. ABC splits into AB and C, then reassembles as CBA. Evidence for the fusion or fission mechanisms is N- or C-termini from one permutation being found within the sequence/structure of the other.
3. *Cut and paste mechanisms* can be either of the other methods but by means of a restriction-modification system present on DNA that is susceptible to restriction, such as a plasmid.

Duplication/deletion is the most likely mechanism of permutation here, between bacterial uL33 and archaeal/eukaryotic uL33. The wide distribution of zinc ribbon rProteins and the variety of their functions suggests their duplication is facile. A duplication of a zinc ribbon gene followed by deletion of the first β-hairpin of the first gene and the last hairpin of the second gene would give the observed permutation. The fusion/fission model is applicable only for multi-domain proteins with multiple-independent folding units (Weiner et al. 2006). The units of permutation of uL33 are simple β-hairpins, which are not independent folds.

It is likely that ancestral uL33, of LUCA, lacked the internal loop. The bacterial N-terminal extension of uL33 observed in some bacteria superimposes and aligns well with the internal loop in the archaeal and eukaryotic uL33 (See Supplementary Figure S5, Additional File 6). The N-terminal extension in uL33 is absent in most bacteria, whereas all uL33 proteins analyzed from archaea and eukarya contain this insertion. This pattern suggests bacterial uL33 proteins that contain the N-terminal extension are intermediates between the simplest bacterial uL33s and more complex uL33s of archaea and eukarya and that the internal loop was an elaboration on the zinc ribbon core that occurred after LUCA.

rProtein Distribution and Nomenclature

Much of our work on the ribosome is based ultimately on pioneering efforts of Ada Yonath, who demonstrated the possibility that ribosomal particles from a variety of species can be characterized in three dimensions (Wittmann et al. 1982; Shevack et al. 1985; Yonath 2002). The ribosome is the most ubiquitous molecular machine in the biological universe (Harris et al. 2003; Koonin 2003; Charlebois and Doolittle 2004; Ortiz et al. 2006; Scott et al. 2010). The translation system is a biological nexus, dominating the interactome in centrality, size, and complexity (Butland et al. 2005). The functions, origins, and evolution of the translation system frame some of the most vexing and engaging questions in biological science. These questions are best addressed in the context of unambiguous and phylogenetically based macromolecular taxonomy and nomenclature.

Traditional naming schemes for translational components are (i) inconsistent across organisms, (ii) based on physical properties such as gel mobility or rates of sedimentation (Stöffler and Wittmann 1971), and (iii) inconsistent with the ancestral relationships, especially of archaea to eukarya. Unfortunately, changes in nomenclature of the translation system, like evolutionary changes in structure and function of the translation system itself, are resisted by strong forces. For example, renaming of ribosomal components would require extensive and expensive changes of long-established databases (Ban et al. 2014). None-the-less, Ban and collaborators did successfully negotiate a significantly improved naming scheme for ribosomal proteins, which is consistent across species (Ban et al. 2014).

An optimal system of nomenclature is based on phylogeny, which provides a robust and logical framework consistent with biological principles (Chothia and Lesk 1986). Here, we provide a phylogenetic naming scheme for rProteins that is based on phylogenetic relationships. This rProtein nomenclature is intended as a tool for studying the

systemics, evolution, and origins of the ribosome rather than as an unrealistic proposal for global adoption by the ribosome community. In the supplementary materials we provide a translation table that relates various rProtein naming schemes (See Supplementary Table S1, Additional File 1).

The basic starting point for our naming scheme was to cease referring to archaeal rProteins as eukaryotic rProteins. Archaea is ancestral to eukarya; rProteins common to both are of archaeal lineage.

For our nomenclature here, rProteins fall into 4 classes and are identified by prefixes.

u; universal, found in all cytosolic ribosomes,

b; bacterial, absent from archaea and eukarya,

a; archaeal, common to archaea and eukarya but absent from bacteria,

e; eukaryotic, absent from archaea and bacteria.

These 4 prefixes are then followed by a letter identifying the ribosomal subunit.

L; large subunit,

S; small subunit.

A Venn diagram of rProteins (Fig. 5) mapped onto the tree of life illustrates their distribution among the three domains (Lecompte et al. 2002; Ban et al. 2014). Using our phylogenetic rProtein naming scheme, nineteen rProteins in the LSU and fifteen in the SSU are shown as universal. Fourteen LSU rProteins and seven SSU rProteins are specific to bacteria. Twenty-two LSU rProteins and twelve SSU rProteins are specific to archaea and eukarya. All archaeal rProteins are found in eukarya, which also contains six additional LSU and six additional SSU rProteins not found in other domains. No rProteins in either ribosomal subunits are shared between bacteria and eukarya except for those that are also found in archaea.

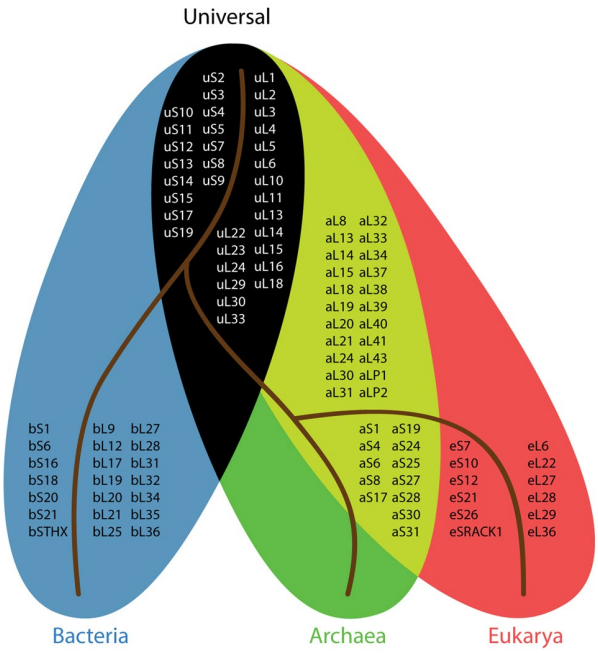

Fig. 5 Venn diagram of rProteins in the three domains of life. This rProtein nomenclature is consistent with the canonical tree of life. uL33 is incorporated as a universal rProtein. bL33 is removed from bacteria and aL42 (eL42 as defined in Ban et al. 2014) is removed from archaea/eukarya. THX, P1, P2, and RACK1 have been renamed to bSTHX, aLP1, aLP2, and eSRACK1, in order to reflect their ribosomal subunit and domain of life associations

The observed partitioning of rProteins, in which no archaeal rProteins are absent from eukarya, and none are shared between bacteria and eukarya unless they are universal, provides strong evidence for the Woese and Fox tree of life (Woese and Fox 1977). The incorporation of three-dimensional structural information into sequence-based phylogeny increases the accuracy and accessible range of phylogenetic relationships, and in our view can resolve many issues, including those related to the basic correctness of the Woese and Fox model.

We have assigned the name uL33 to proteins known previously bL33 and eL42, consistent with our observation that it is conserved in all three domains of life. A limited number of rProteins fit imperfectly into our classification system because they are absent from select species. For example, some bacteria lack genes encoding uL30, and aL13 is absent from some archaea (Lecompte et al. 2002). To the best of our knowledge, every species contains uL33. We hope that our naming scheme and discovery of homology and universality of uL33 (bL33 and eL42) will facilitate the accurate assignment of universal ribosome characteristics, and will be updated as new relationships between rProteins become apparent.

Conclusions

Bacterial rProtein bL33 and archaeal/eukaryotic rProtein eL42 are both composed of a zinc ribbon protein fold and are related to each other by circular permutation. *Duplication/deletion* appears to be the most likely mechanism of permutation between bL33 and eL42. This homology relationship warrants a change to a common name of uL33, consistent with the observation that this rProtein is conserved in all three domains of life. We provide a naming scheme for rProteins that is based on phylogenetic relationships. This scheme is intended for use as a tool for studying systemics, evolution, and origins of the translation system. Our naming scheme is similar to the previous nomenclature by Ban et al. (2014) with the exception that it does not refer to archaeal rProteins as eukaryotic rProteins, and every rProtein is assigned a letter corresponding to its ribosomal subunit.

Methods

Global Superimposition of LSU

A global superimposition of the LSU ribosome was performed using the CEAlign function of PyMOL. The superimposition uses 277 nucleotide residues that form and surround the PTC, no amino acid residues are considered in this method. Pro-origami (Stivala et al. 2011) was then used to make the 2-dimensional protein structure cartoons, and DSSP (Touw et al. 2015) was used to determine the secondary structure.

Defining the uL33 Common Core

We defined the common core of uL33 by first locating the most commonly conserved regions of zinc ribbons, the two zinc-knuckles. We labeled the zinc-knuckle of uL33 which interacts with rRNA helices 86 and 88 as "zinc-knuckle 1" and the zinc-knuckle that is solvent exposed as "zinc-knuckle 2." In archaeal and eukaryotic uL33, zinc-knuckle 1 is closest of the two zinc-knuckles to the N-terminus. In bacterial uL33, zinc-knuckle 2 is closest of the two zinc-knuckles to the N-terminus. 1 or 2 amino acid residues were removed from in between the cysteines of the zinc-knuckles in some eukaryotes so that all uL33 core structures would have 2 zinc-knuckle motifs of –CXXC–. In our dataset, the structure of uL33 in *E. coli* does not have a zinc-binding domain; however, visual inspection of the globally superimposed structures indicates that Asp39 and Val42 are equivalent to the cysteines of zinc-knuckle 1, while Ser12 and Gly15 are equivalent to the cysteines of zinc-knuckle 2.

We chose the common core of uL33 to be 50 amino acid residues long because uL33 of *E. coli* is the smallest of the 10 analyzed structures at only 50 resolved amino acid residues long. Amino acid residues were omitted from the structures so that all common core structures contained 50 amino acid residues which corresponded in position closest to those of *T. thermophilus*. We chose to circularly permutate uL33 of *E. coli* between amino acid residues 28 and 29, and *T. thermophilus* between amino acid residues 29 and 30, so that zinc-knuckles 1 and 2 aligned in all 10 structures.

Superimposition of the uL33 Common Core

PyMOL was used to calculate the RMSD between all residues of the ten globally superimposed uL33 common core structures, followed by aligning the structures irrespective of their amino acid sequences, and then recalculating the RMSD. A sequence alignment was then extracted from this superimposition.

Structurally Based Multiple Sequence Alignments

Structurally based multiple sequence alignments were performed on a species list previously applied in our lab which sparsely samples the tree of life (See Supplementary Table S4, Additional File 1) (Bernier et al. 2018). The list was expanded to 149 species to better represent all newly discovered organisms. The rProtein sequences for archaea and bacteria were downloaded from (Yutin et al. 2012), additionally any missing species' sequences together with the 30 eukaryotic proteins, were downloaded from Uniprot (2017). Separate alignments for bacteria, archaea, and eukarya were executed using PROMALS3D (Pei et al. 2008) with structural options provided by Dali (Holm and Sander 1996). Circular permutation was simulated on the aligned bacterial sequences corresponding to the circular permutation sites identified; residues after the circular permutation site were moved to the N-terminus of the corresponding sequence. By concatenating the sequences, two new alignment sets of 149 species each were created—one containing the original bacterial sequences and the other containing the permuted ones (See Supplementary Figures S4 and S5, Additional File 1).

rRNA Secondary Structures and PASE

rRNA secondary structures were made using RiboVision 2, apollo.chemistry.gatech.edu/RiboVision2/ (See Supplementary Figures S8 and S9, Additional File 1). PASE data were obtained from Bernier et al. and mapped onto rRNA secondary structures (See Supplementary Figure S9, Additional File 1) (Bernier et al. 2018).

Acknowledgements The authors would like to thank Dr. Hyman Hartman and Dr. Nenad Ban for discussions. The authors declare that they have no competing interests. This work was supported by National Aeronautics and Space Administration (Grant Numbers NNX16AJ28G and NNX16AJ29G) and NSF Grant 1713995.

Author Contributions NAK, PIP, ASP, and LDW conceived the study. NAK, PIP, and AV collected and analyzed the data. NAK and PIP generated all the figures and tables. NAK, PIP, ASP, and LDW wrote the manuscript.

References

Amunts A, Brown A, Bai XC, Llacer JL, Hussain T, Emsley P, Long F, Murshudov G, Scheres SH, Ramakrishnan V (2014) Structure of the yeast mitochondrial large ribosomal subunit. Science 343:1485–1489

Anger AM, Armache JP, Berninghausen O, Habeck M, Subklewe M, Wilson DN, Beckmann R (2013) Structures of the human and drosophila 80S ribosome. Nature 497:80–85

Armache J-P, Anger AM, Márquez V, Franckenberg S, Fröhlich T, Villa E, Berninghausen O, Thomm M, Arnold GJ, Beckmann R, Wilson DN (2013) Promiscuous behaviour of archaeal ribosomal proteins: implications for eukaryotic ribosome evolution. Nucleic Acids Res 41:1284–1293

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

Ban N, Beckmann R, Cate JH, Dinman JD, Dragon F, Ellis SR, Lafontaine DL, Lindahl L, Liljas A, Lipton JM, McAlear MA, Moore PB, Noller HF, Ortega J, Panse VG, Ramakrishnan V, Spahn CM, Steitz TA, Tchorzewski M, Tollervey D, Warren AJ, Williamson JR, Wilson D, Yonath A, Yusupov M (2014) A new system for naming ribosomal proteins. Curr Opin Struct Biol 24:165–169

Ben-Shem A, Jenner L, Yusupova G, Yusupov M (2010) Crystal structure of the eukaryotic ribosome. Science 330:1203–1209

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

Bernier CR, Petrov AS, Kovacs NA, Penev PI, Williams LD (2018) Translation: the universal structural core of life. Mol Biol Evol 34:2065–2076

Bliven S, Prlić A (2012) Circular permutation in proteins. PLoS Comput Biol 8:e1002445

Butland G, Peregrin-Alvarez JM, Li J, Yang W, Yang X, Canadien V, Starostine A, Richards D, Beattie B, Krogan N, Davey M, Parkinson J, Greenblatt J, Emili A (2005) Interaction network containing conserved and essential protein complexes in *Escherichia coli*. Nature 433:531–537

Charlebois RL, Doolittle WF (2004) Computing prokaryotic gene ubiquity: rescuing the core from extinction. Genome Res 14:2469–2477

Cheng H, Schaeffer RD, Liao Y, Kinch LN, Pei J, Shi S, Kim B-H, Grishin NV (2014) Ecod: an evolutionary classification of protein domains. PLoS Comput Biol 10:e1003926

Chothia C, Lesk AM (1986) The relation between the divergence of sequence and structure in proteins. EMBO J 5:823–826

Cunningham BA, Hemperly JJ, Hopp TP, Edelman GM (1979) Favin versus concanavalin a: Circularly permuted amino acid sequences. Proc Natl Acad Sci USA 76:3218–3222

D'Abrosca G, Russo L, Palmieri M, Baglivo I, Netti F, de Paola I, Zaccaro L, Farina B, Iacovino R, Pedone PV, Isernia C, Fattorusso R, Malgieri G (2016) The (unusual) aspartic acid in the metal coordination sphere of the prokaryotic zinc finger domain. J Inorg Biochem 161:91–98

Davies C, Gerchman SE, Kycia JH, McGee K, Ramakrishnan V, White SW (1994) Crystallization and preliminary X-ray diffraction studies of bacterial ribosomal protein l14. Acta Crystallographica Section D 50:790–792

Davies C, White SW, Ramakrishnan V (1996) The crystal structure of ribosomal protein l14 reveals an important organizational component of the translational apparatus. Structure 4:55–66

Dresios J, Chan Y-L, Wool IG (2005) Ribosomal zinc finger proteins: the structure and the function of yeast yl37a. In: Iuchi S, Kuldell N (eds) Zinc finger proteins: from atomic contact to cellular function. Springer, Boston, pp 91–98

Dunkle JA, Wang LY, Feldman MB, Pulk A, Chen VB, Kapral GJ, Noeske J, Richardson JS, Blanchard SC, Cate JHD (2011) Structures of the bacterial ribosome in classical and hybrid states of tRNA binding. Science 332:981–984

Fox GE (2010) Origin and evolution of the ribosome. Cold Spring Harb Perspect Biol 2:a003483

Gabdulkhakov A, Nikonov S, Garber M (2013) Revisiting the haloarcula marismortui 50S ribosomal subunit model. Acta Crystallogr Sect D 69:997–1004

Harris JK, Kelley ST, Spiegelman GB, Pace NR (2003) The genetic core of the universal ancestor. Genome Res 13:407–412

Hashem Y, des Georges A, Fu J, Buss SN, Jossinet F, Jobe A, Zhang Q, Liao HY, Grassucci RA, Bajaj C, Westhof E, Madison-Antenucci S, Frank J (2013) High-resolution cryo-electron microscopy structure of the *Trypanosoma brucei* ribosome. Nature 494:385–389

Holm L, Sander C (1996) Mapping the protein universe. Science 273:595–602

Hug LA, Baker BJ, Anantharaman K, Brown CT, Probst AJ, Castelle CJ, Butterfield CN, Hernsdorf AW, Amano Y, Ise K, Suzuki Y, Dudek N, Relman DA, Finstad KM, Amundson R, Thomas BC, Banfield JF (2016) A new view of the tree of life. Nat Microbiol 1:16048

Illergard K, Ardell DH, Elofsson A (2009) Structure is three to ten times more conserved than sequence—a study of structural response in protein cores. Proteins 77:499–508

Khatter H, Myasnikov AG, Natchiar SK, Klaholz BP (2015) Structure of the human 80S ribosome. Nature 520(7549):640–645

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

Koonin EV (2003) Comparative genomics, minimal gene-sets and the last universal common ancestor. Nat Rev Microbiol 1:127–136

Kovacs NA, Petrov AS, Lanier KA, Williams LD (2017) Frozen in time: the history of proteins. Mol Biol Evol 34:1252–1260

Krishna SS, Majumdar I, Grishin NV (2003) Survey and summary: structural classification of zinc fingers. Nucleic Acids Res 31:532–550

Lecompte O, Ripp R, Thierry JC, Moras D, Poch O (2002) Comparative analysis of ribosomal proteins in complete genomes: an example of reductive evolution at the domain scale. Nucleic Acids Res 30:5382–5390

Lupas AN, Alva V (2017) Ribosomal proteins as documents of the transition from unstructured (poly) peptides to folded proteins. J Struct Biol 198:74–81

Makarova KS, Ponomarev VA, Koonin EV (2001) Two c or not two c: recurrent disruption of zn-ribbons, gene duplication, lineage-specific gene loss, and horizontal gene transfer in evolution of bacterial ribosomal proteins. Genome Biol 2:RESEARCH 0033

McCall KA, Huang C-C, Fierke CA (2000) Function and mechanism of zinc metalloenzymes. J Nutr 130:1437S–1446S

Noller HF, Hoffarth V, Zimniak L (1992) Unusual resistance of peptidyl transferase to protein extraction procedures. Science 256:1416–1419

Ortiz JO, Förster F, Kürner J, Linaroudis AA, Baumeister W (2006) Mapping 70S ribosomes in intact cells by cryoelectron tomography and pattern recognition. J Struct Biol 156:334–341

Pei J, Kim B-H, Grishin NV (2008) Promals3d: a tool for multiple protein sequence and structure alignments. Nucleic Acids Res 36:2295–2300

Petrov AS, Bernier CR, Hsiao C, Norris AM, Kovacs NA, Waterbury CC, Stepanov VG, Harvey SC, Fox GE, Wartell RM, Hud NV, Williams LD (2014) Evolution of the ribosome at atomic resolution. Proc Natl Acad Sci USA 111:10251–10256

Petrov AS, Gulen B, Norris AM, Kovacs NA, Bernier CR, Lanier KA, Fox GE, Harvey SC, Wartell RM, Hud NV, Williams LD (2015) History of the ribosome and the origin of translation. Proc Natl Acad Sci USA 112:15396–15401

Polikanov YS, Steitz TA, Innis CA (2014) A proton wire to couple aminoacyl-tRNA accommodation and peptide-bond formation on the ribosome. Nat Struct Mol Biol 21:787–793

Ponting CP, Russell RB (1995) Swaposins: circular permutations within genes encoding saposin homologues. Trends Biochemical Sci 20:179–180

Scott M, Gunderson CW, Mateescu EM, Zhang Z, Hwa T (2010) Interdependence of cell growth and gene expression: origins and consequences. Science 330:1099–1102

Selmer M, Dunham CM, Murphy FV, Weixlbaumer A, Petry S, Kelley AC, Weir JR, Ramakrishnan V (2006) Structure of the 70S ribosome complexed with mRNA and tRNA. Science 313:1935–1942

Shevack A, Gewitz HS, Hennemann B, Yonath A, Wittmann HG (1985) Characterization and crystallization of ribosomal particles from halobacterium-marismortui. FEBS Lett 184:68–71

Sippl MJ, Wiederstein M (2012) Detection of spatial correlations in protein structures and molecular complexes. Structure 20:718–728

Stivala A, Wybrow M, Wirth A, Whisstock JC, Stuckey PJ (2011) Automatic generation of protein structure cartoons with pro-origami. Bioinformatics 27:3315–3316

Stöffler G, Wittmann H (1971) Ribosomal proteins. Xxv. Immunological studies on *Escherichia coli* ribosomal proteins. J Mol Biol 62:407–409

Sun M, Li W, Blomqvist K, Das S, Hashem Y, Dvorin JD, Frank J (2015) Dynamical features of the Plasmodium falciparum ribosome during translation. Nucleic Acids Res 43:10515–10524

The UniProt Consortium (2017) Uniprot: the universal protein knowledgebase. Nucleic Acids Res 45:D158–D169

Touw WG, Baakman C, Black J, te Beek TAH, Krieger E, Joosten RP, Vriend G (2015) A series of PDB-related databanks for everyday needs. Nucleic Acids Res 43:D364–D368

Weiner III J, Beaussart F, Bornberg-Bauer E (2006) Domain deletions and substitutions in the modular protein evolution. FEBS J 273:2037–2047

Wimberly BT, Brodersen DE, Clemons WM Jr, Morgan-Warren RJ, Carter AP, Vonrhein C, Hartsch T, Ramakrishnan V (2000) Structure of the 30 s ribosomal subunit. Nature 407:327–339

Wittmann HG, Mussig J, Piefke J, Gewitz HS, Rheinberger HJ, Yonath A (1982) Crystallization of *Escherichia coli* ribosomes. FEBS Lett 146:217–220

Woese CR, Fox GE (1977) Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proc Natl Acad Sci USA 74:5088–5090

Yonath A (2002) The search and its outcome: high-resolution structures of ribosomal particles from mesophilic, thermophilic, and halophilic bacteria at various functional states. Annu Rev Biophys Biomol Struct 31:257–273

Yutin N, Puigbò P, Koonin EV, Wolf YI (2012) Phylogenomics of prokaryotic ribosomal proteins. PLoS ONE 7:e36972