Published online 6 October 2022

Nucleic Acids Research, 2022, Vol. 50, No. 18 ***10717–10732***
https://doi.org/10.1093/nar/gkac844

rRNA expansion segment 7 in eukaryotes: from Signature Fold to tentacles

Marcin Biesiada[1], Michael Y. Hu[2,3], Loren Dean Williams [2,3,*], Katarzyna J. Purzycka[1,*] and Anton S. Petrov [2,3,*]

[1]Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan 61-704, Poland, [2]Center for the Origins of Life, Georgia Institute of Technology, Atlanta, GA 30332, USA and [3]School of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta, GA 30332, USA

Received December 27, 2021; Revised September 13, 2022; Editorial Decision September 16, 2022; Accepted September 22, 2022

ABSTRACT

The ribosomal core is universally conserved across the tree of life. However, eukaryotic ribosomes contain diverse rRNA expansion segments (ESs) on their surfaces. Sites of ES insertions are predicted from sites of insertion of micro-ESs in archaea. Expansion segment 7 (ES7) is one of the most diverse regions of the ribosome, emanating from a short stem loop and ranging to over 750 nucleotides in mammals. We present secondary and full-atom 3D structures of ES7 from species spanning eukaryotic diversity. Our results are based on experimental 3D structures, the accretion model of ribosomal evolution, phylogenetic relationships, multiple sequence alignments, RNA folding algorithms and 3D modeling by RNAComposer. ES7 contains a distinct motif, the 'ES7 Signature Fold', which is generally invariant in 2D topology and 3D structure in all eukaryotic ribosomes. We establish a model in which ES7 developed over evolution through a series of elementary and recursive growth events. The data are sufficient to support an atomic-level accretion path for rRNA growth. The non-monophyletic distribution of some ES7 features across the phylogeny suggests acquisition via convergent processes. And finally, illustrating the power of our approach, we constructed the 2D and 3D structure of the entire LSU rRNA of *Mus musculus*.

INTRODUCTION

The ribosome, a massive assembly of rRNAs and ribosomal proteins (rProteins), lies at the heart of translation, synthesizing all coded protein. The cytosolic ribosome contains a 'common core' with universal structure in all organisms (1–3). Common core rRNA is approximated by bacterial rRNA (4); around 90% of *Escherichia coli* rRNA is universal in secondary (2D) and three-dimensional (3D) structure in cytosolic ribosomes of extant organisms (2).

The ribosomal common core grew by accretion during six phases of evolution, culminating with the prokaryotic ribosome (5). The expansion of ribosomes then paused for several billion years, between the last universal common ancestor and the rise of eukarya.

Eukaryotic rRNAs are larger than prokaryotic rRNAs, and contain a diverse array of rRNA expansion segments [ESs, (6–9)] (Figure 1, http://apollo.chemistry.gatech.edu/RibosomeGallery/index.html) that emerge from a small number of sites on the surface of the common core, and are remote from functional centers, as shown by chemical probing and confirmed by structure determination (10–14). The acquisition of the eukaryotic shell, on the surface of the common core, is a seventh phase of ribosomal evolution (2,4). Protist and fungal ESs combine with eukaryotic-specific rProteins to form a secondary shell around the common core (2). The growth of 'rRNA tentacles', which are long double-helical elements observed in endothermic eukaryotes, marks an eighth phase of ribosomal evolution. In birds and mammals, tentacles extend for hundreds of Ångstroms from the ribosomal surface (2,4) (Figure 1B). Variation of ES size and structure over evolution is evident by comparison of ribosomes of eukarya (15–18) and prokaryotes (19–22). An accretion mechanism of ESs growth over evolution was initially proposed by Michot (23).

Expansion segment 7 (ES7) is a nexus of ribosomal growth and diversity (1,4,6–9,24–26); it arose in the seventh phase of evolution and continued growth in the eighth phase. ES7 is the largest eukaryotic ES, with the most complex structure (Figure 2). ES7 tends to increase in size in concert with species complexity, and with overall size of the large ribosomal subunit (LSU), accounting for ∼20% of LSU rRNA in mammals (2,5). Obligate parasites show

*To whom correspondence should be addressed. Tel: +1 404 894 8338; Email: anton.petrov@biology.gatech.edu
Correspondence may also be addressed to Loren Dean Williams. Tel: +1 404 385 6258; Email: loren.williams@chemistry.gatech.edu
Correspondence may also be addressed to Katarzyna J. Purzycka. Tel: +48 61 8 528503; Email: purzycka@ibch.poznan.pl

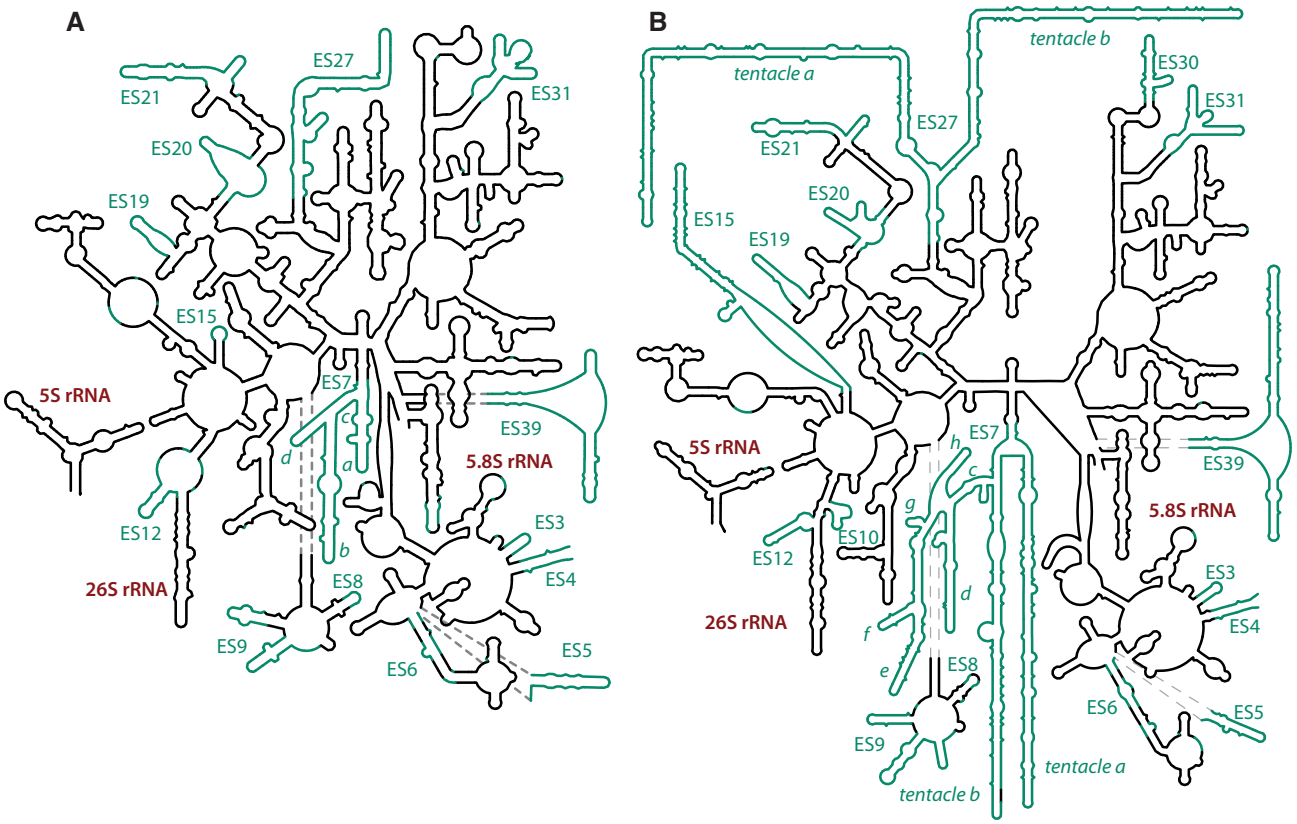

Figure 1. Secondary structures of two eukaryotic LSU rRNAs. (**A**) *S. cerevisiae,* and (**B**) *H. sapiens*. rRNA of ribosomal common core as defined in (2) is black. ESs are highlighted in teal and labeled according to Gerbi (8).

reduced rRNA expansion segments and genomes (27), and are not discussed in detail here.

In this report, we determine 2D and 3D structures of ES7 across phylogeny using a hybrid method that combines evolutionary mechanisms with RNA folding methods (Figure 3). We introduce the ES7 Signature Fold, which is conserved in 3D structure in all eukaryotes and is the foundation for broad diversity over phylogeny. We also improve previous ES7 structural and phylogenetic predictions (4,23,28–32). The structures here have been integrated into R2DT webserver (33) and can serve as templates for rRNA secondary structure for a broad variety of species.

MATERIALS AND METHODS

Modeling ES7: the approach

We have established a hybrid method to construct 2D and 3D models of ES7s of species broadly distributed across eukarya. The method introduces a novel combination of conventional 2D and 3D RNA folding algorithms with the accretion model of rRNA growth over evolution.

The accretion model stipulates that rRNA accumulates new elements onto pre-existing structure in the absence of remodeling (5), allowing identification of ancestral (structurally invariant) elements of ESs across all species.

We filtered the evolutionary information from a multiple sequence alignment (MSA, Supplementary Data Set 1) through folding algorithms (35), and predicted an initial set of secondary structures (Figure 3). Addition of structural constraints from the accretion model substantially reduced the number of plausible structures. We used the resulting secondary structures as input for 3D structure prediction by RNAComposer (36) and applied the 3D models to discriminate between several secondary structural using the plausibility of the 3D models and lack of the steric clashes as selection criteria. The results allow us to describe ES7 in 2D and 3D for a variety of eukaryotic species (Supplementary Figure S1). The resulting secondary and 3D structures of ES7 are given in Supplementary Data Sets 2 and 3.

ES7 Signature Fold

We have identified the ES7 Signature Fold (Figure 4), a structurally conserved portion of ES7 seen in essentially all eukaryotic ribosomes. The ES7 Signature Fold contains H25 in addition to the archaeal expansion (ES7b, numbered here and after according to *Homo sapiens* schema, Figure 2, inset), plus two additional helices (ES7a and ES7c), and junctions that link them. The ES7 Signature Fold is conserved in 2D and 3D structure. Exceptions are ES7 from *Yarrowia lipolytica*, which lacks ES7c and ES7 of kinetoplastic parasites, which lack ES7a (Figure 2).

ES7 in protists generally consists of the ES7 Signature Fold and little else. Variability of ES7 across phylogeny is focused on animals, and arises from elongation of helices of the Signature Fold (ES7a–ES7c) and insertion of new helices (ES7d–ES7h, inset of Figure 2). The largest ES7s, in

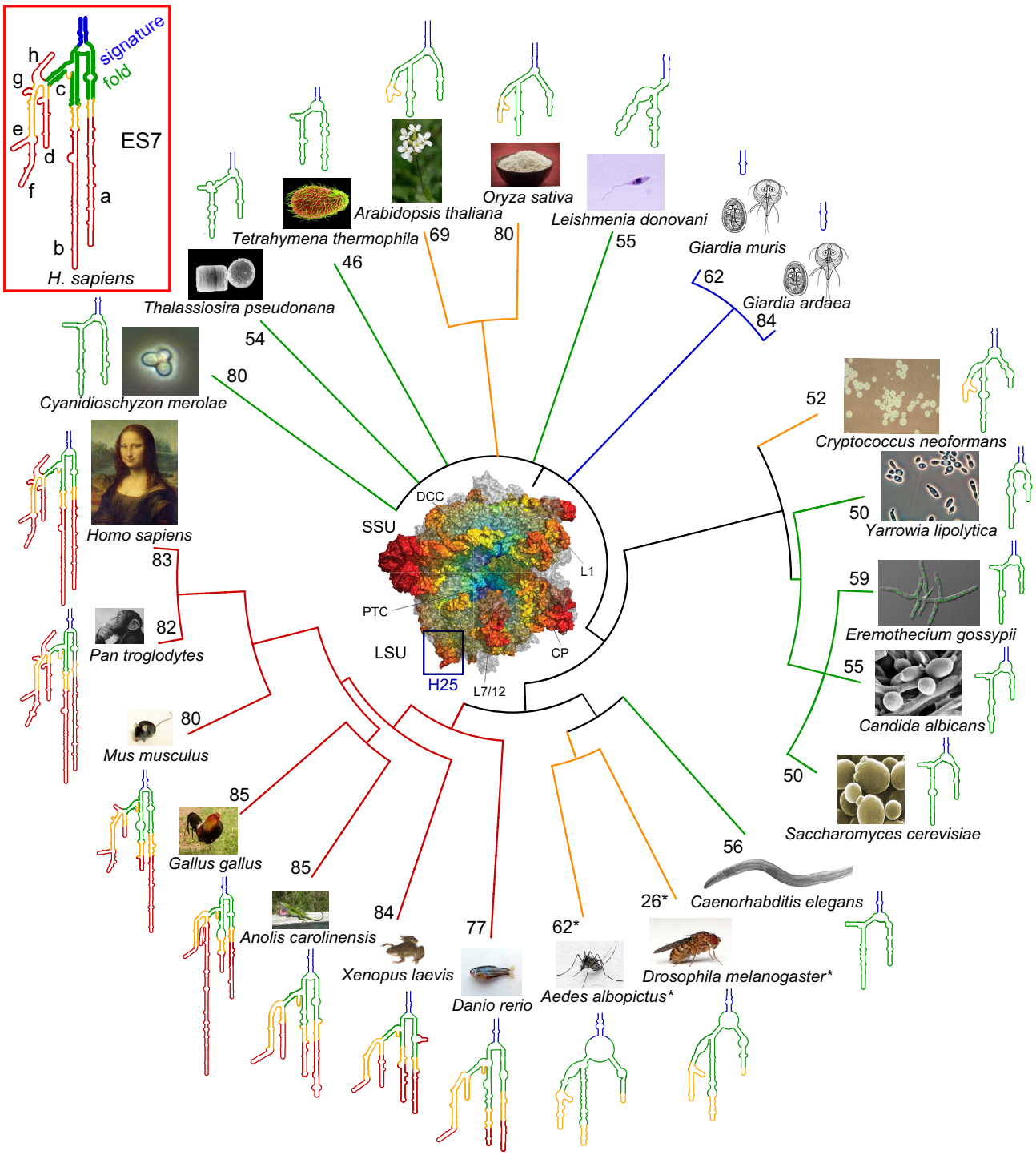

Figure 2. Secondary structures of ES7 mapped onto the canonical eukaryotic tree of life. Basal Helix 25 (boxed in the center) is part of the universal core; it is shown in blue in secondary structures as well as in the inset of *H. sapiens* ES7. Green indicates the ES7 Signature Fold, which is universal to rRNA of all eukaryotes. Yellow is universal to metazoans. Red is metazoan specific rRNA extensions, some of which reach extreme lengths in birds and mammals (tentacles). Colors of branches and internal nodes represent allocation of species (leaves) into Groups 0–3. The GC content (shown as percentage) for each ES7 is displayed next to each leaf of the tree. The asterisks indicate related species (*D. melanogaster* and *A. albopictus*) with a highly divergent GC content. Inset: the structure of ES7 of *H. sapiens* with the Signature Fold in bold green and blue. Helices labeled according to *H. sapiens* schema (4). All secondary structures in this document were visualized with the web portal RiboVision (34). The 3D structure of a prokaryotic ribosome with H25 highlighted in the blue box. In this image, colored by distance from the functional centers (peptidyl transferase center in the LSU and decoding center in the SSU) the ribosomal proteins are in gray.

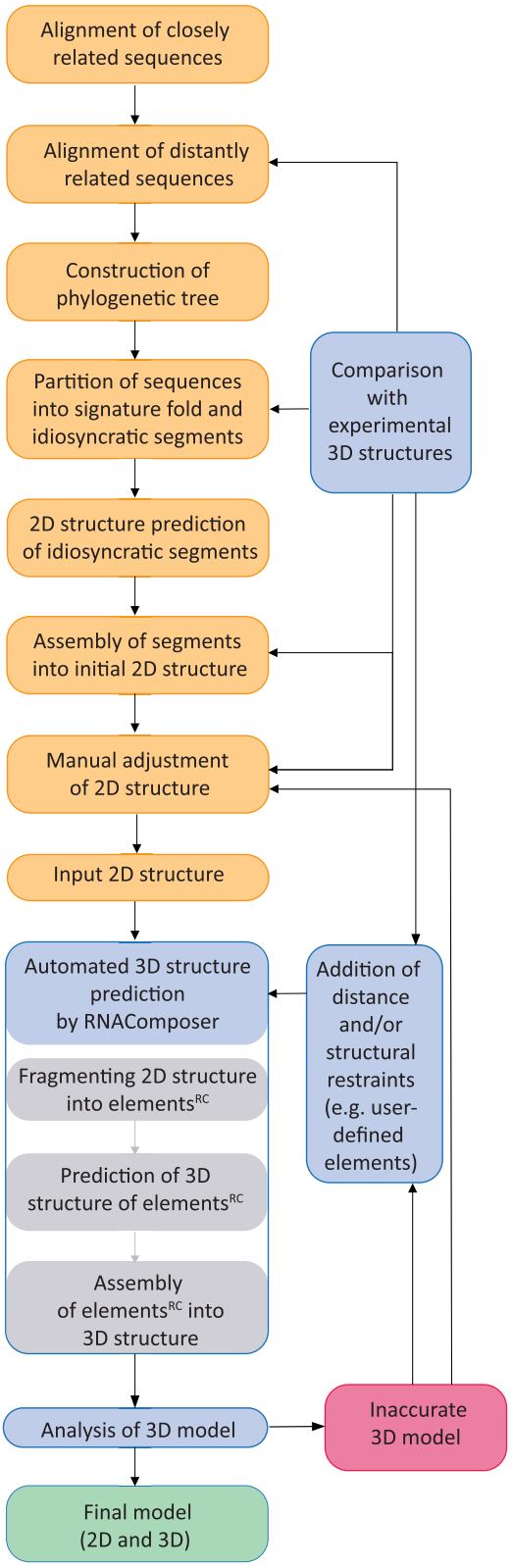

Figure 3. A schematic of the work-flow used to model 2D and 3D structures of eukaryotic rRNAs. RC indicates automated steps performed by RNAComposer.

birds and mammals, contain the ES7 Signature Fold elaborated by linked arrays of junctions, helices, bulges and loops (Figure 2). The ES7 Signature Fold is absent only from a few obligate pathogenic eukaryotes (Figure 2: Giardia) that are also characterized by reduced genomes and ribosomes (37,38).

Modeling ES7: methodology

The ES7 Signature Fold is maintained over evolution as helices accrete and elongate. This critical finding enabled us to partition sections of ES7s of various species into sub-fragments and to independently predict sub-fragment secondary structures. We have integrated evolutionary and structural constraints into a workflow of RNA 2D and 3D structure prediction, surmounting limitations of the size, complexity and heterogeneity of ES7s. Our modeling is enabled in part by a) the process of accretion, as opposed to remodeling, as ES7 increased in size over evolution, and b) structural conservation of the ES7 Signature Fold. Specific modeling information is provided in the Supplementary Text.

The MSA was used to identify the Signature Fold in each ES7. The starting seed MSA was derived from the structural superimposition and contained the regions within the Signature Fold and enriched by additional sequences using mafft (39). The Signature Fold-annotated MSA (Supplementary DataSet S4) allowed us to pinpoint variable helical stem loops. By combining the information from the MSA with known 2D structures, 2D ES7 models of closely related species were predicted. Finally, secondary structures of ES7s of more distantly related species were predicted by bootstrapping the available information from the MSA. Predicted 2D and 3D structures were constrained by the conserved Signature Fold and variable helical stems. Secondary structures of individual helical elements were predicted by mfold (35) and merged with the ES7 Signature Fold to form the complete ES7 secondary structure. Variable stem loop regions of the MSA were manually adjusted when necessary, based on the 2D predictions.

Three-dimensional models were predicted by inputting sequences and secondary structures into RNAComposer (36) (Figure 3). Experimental 3D structures of ES7 are known for several eukaryotic ribosomes (4,16,17,40). We leveraged these experimental ribosomal structures and mined a library of RNA structural elements (36,41,42). In instances, when RNAComposer's dictionary lacked required elements (within the Signature Fold or helical stems, Supplementary Figure S2), the missing segments were modeled using standard protocols of comparative RNA 3D structure prediction based on 3D RNA templates (43,44), subsequently introduced into the prediction protocol as user-defined elements (45), and further refined by RNAComposer.

RNAComposer uses secondary structural topology over the sequence homology to assemble small 3D elements into a 3D model (see Supplementary Text and Supplementary Figure S2). Application of the hybrid methodology for the prediction of 2D structures limited the number of plausible topologies and eliminated many 3D structures of

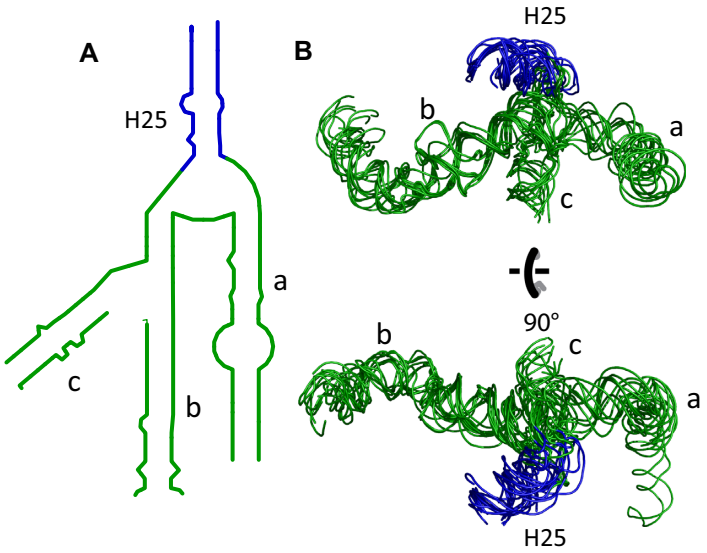

Figure 4. The ES7 Signature Fold. (**A**) Secondary structure. (**B**) Three-dimensional structure. The coloring scheme is the same as in Figure 2. In panel (B), all experimental and modeled structures from Table 1 are superimposed.

unfavorable steric interactions between the structural domains. Multiple models were predicted for each structure studied.

Final models were selected based on (i) lowest energy, (ii) preservation of relative orientations of helices at junctions (46) and (iii) consistency with the evolutionary model (the underlying core tends to be preserved in all ES7 structures). The use of RNAComposer for 3D structure prediction allowed to overcome the problem of computational cost for larger RNA molecules, demanding addition of atomic details to coarse-grain models or prediction dependent on structural templates. Furthermore, the application of the hybrid approach eliminates the problem of homologs for RNAs that differ vastly in sequence or contain unique expansions which are not found within the closest homologs. Therefore, our approach allows us to overcome challenges to structure prediction seen in other methods (43,44,47–52).

Validation of the approach

To validate the hybrid approach, we performed two tests. First, using ES7 of *Saccharomyces cerevisiae* as a template, we predicted the structure of ES7 of *Candida albicans* and experimentally validated it by a chemical probing using Selective 2'-Hydroxyl Acylation analyzed by Primer Extension (SHAPE) (53). The structure predicted by the hybrid approach is in good agreement with the experimental SHAPE data (54) and significantly improved the model of the *C. albicans* ES7 structure compared to a previous effort (55) (Supplementary Figure S3). Second, we performed a reciprocal prediction, by modeling the known structure of ES7 of *S. cerevisiae* based on ES7 of another fungus (*Eremothecium gossypii*) and comparing the reverse prediction with experimentally determined structures (PDB entry 4V6I, 5JUO, 4U4R) (15,56,57) (Supplementary Figure S4). The accuracy of the *S. cerevisiae* ES7 model was high as determined by comparison with ES7 of the experimentally determined structure (PDB entry 4V6I), resulting in Matthews correlation coefficient = 83.8%, sensitivity = 98.5%, specificity = 90.3% (58). Additional details of this validation of the hybrid approach are reported in the Supplementary Text.

Allocation of ES7s to Groups based on topology and length of variable expansions

To systematize ES7 modeling, ES7s from 23 species were allocated into groups based on size and complexity (Figure 2, Table 1). Group 0 represents ES7s in species with H25 alone. Groups 1–3 (Figure 2) represent ES7s with various extents of recursive accretion onto the ES7 Signature Fold, with Group 1 ES7 < Group 2 ES7 < Group 3 ES7. Representatives from each group are compiled in Table 1. Auxiliary MSAs for ES7s within each group are compiled in Supplementary Data Set 4.

Group 0 ES7s are simple stem loops, seen in obligate pathogens with reduced genomes and ribosomes similar to those of bacteria.

Group 1 ES7s are up to 250 nts in length and are composed essentially of the Signature Fold. *Y. lipolytica* is an exception that lacks ES7c (Figure 2). Parasitic kinetoplasts (*Leishmania donovani*, *Tripanosoma brucei* and *Tripanosoma cruzi*) lack ES7a and contain elongated ES7c (18,59).

Group 2 ES7s, seen in invertebrate metazoans, plants and some fungi, are larger and more complex than Group 1 ES7s (Figure 2). ES7s of Group 2 contain the ES7 Signature Fold, with extensions of helices ES7a–c and the addition of ES7d–e, which branch off of ES7c.

Group 3 ES7s are found in chordates (fish, amphibians, reptiles, mammals). Each Group 3 ES7 contains a subset of five core helices (ES7a–e, Figure 2, inset). Additional growth (ES7a–b; d–e) and further branching (ES7f–h) of ES7 is observed in endothermic vertebrates (birds and mammals), which contain GC rich tentacles. In these species, helices of ES7 break through the rRNA-protein

Table 1. ES7 groups

Group 0	nts	GC%	Group 1	nts	GC%	Group 2	nts	CG%	Group 3	nts	GC%
Giardia ardeae	*31*	*82*	*Caenorhabditis elegans*	*232*	*56*	***Aedes albopictus***[a]	***352***	***62***	*Anolis carolinensis*	*521*	*85*
Giardia intestinalis	*8*	*100*	***Candida albicans***[a]	***209***	***55***	*Arabidopsis thaliana*	*224*	*69*	*Danio rerio*	*473*	*77*
Giardia muris[a]	***29***	***62***	*Cyanidioschyzon merolae*	*255*	*80*	*Cryptococcus neoformans*	*235*	*52*	***Gallus gallus***[a]	***799***	***85***
			Eremothecium gossypii	*212*	*59*	*Drosophila melanogaster*[b]	*340*	*26*	*Homo sapiens*[b]	*876*	*83*
			Saccharomyces cerevisiae[b]	*210*	*50*	*Oryza sativa*	*214*	*80*	***Musmusculus***[a]	***690***	***80***
			Tetrahymena thermophilia[b]	*220*	*46*				*Pan troglodytes*	*866*	*82*
			Thalassiosira pseudonana	*212*	*54*				*Xenopus laevis*	*510*	*84*
			Yarrowia lipolytica	*161*	*50*						
			Leishmania donovani	*200*	*55*						

[a]Species whose ES7 3D structures were predicted here by RNAComposer (bold).
[b]Species whose ES7 3D structures were determined experimentally by others (4,16,17).

core and extend out from the characteristic protist shell of eukaryotic ribosomes (1,4).

RESULTS

ESs are highly diverse across phylogeny and can be dynamic and disordered *in vivo*. Modeling of ESs presents challenges at the levels of both 2D (30) and 3D structure (4). Significant portions of ESs are omitted from most experimental structures. We have predicted 2D structures of ES7s of 23 species in Groups 0–3 from across the phylogenetic tree, and modelled 3D structures for representative species from each group (Movie 1, Supplementary Table S1, Supplementary Data Set 3). Group 0 contains ES7s of obligate parasite species with reduced rRNAs. Mapping of ES7 structures onto the eukaryotic phylogenetic tree (Figure 2) reveals non-monophyletic distribution for Groups 1–2, while ES7 of Group 3 represents a monophyletic group of vertebrate species.

ES7 predictions

ES7 structures modeled in the current study differ in size, branching topologies and sequence compositions. These differences posed a number of modeling challenges that have been independently addressed in the process of predicting the ES7 structures of species from each group.

Sequences from Group 1 were very similar to each other in length and in 2D and 3D structure. ES7 in Group 1 is essentially composed of the Signature Fold. Due to the similarity of size and secondary structure, identification of the individual helical regions within the MSA and overall 2D and 3D modelling was straightforward.

The ES7s of Group 2 have accreted beyond the Signature Fold. Most of the sequences from Group 2 have high GC content, ranging from 60% to 80% (Figure 2) and are represented within multiple phylogenetic groups, revealing a substantial sequence variability. *Drosophila melanogaster*, the conventional biological model whose 3D structure has been determined (4), is anomalous in ES7 GC composition, and is the only metazoan we have discovered thus far with low GC content of ES7 (26%, Supplementary Figure S5). In the current study, we used the Signature Fold of *D. melanogaster* as a guide to model the 2D structure of ES7 for another insect, *A. albopictus*, whose ES7 contains a representative GC content for Group 2 (Supplementary Figure S6, Supplementary Data Set 5). This modeled ES7 structure was further used as a template to model 2D ES7s of other species from Group 2 that are similar in sequence and its GC content (Supplementary Data Sets 2 and 3).

Group 3 contains ES7s of endothermic vertebrates, which are characterized by large size and high complexity, with further diversity in helical branching. The tentacle-like regions are hypervariable in length and branching topology (Figure 2; compare *Xenopus laevis*, *Gallus gallus*, *Mus musculus* and *Homo sapiens*) among the members of the group, representing the main modeling challenge. The conservation of the ES7 Signature Fold in the 3D structure of the *H. sapiens* ribosome and the high sequence similarity among Chordates in this core region (Supplementary Figure S7) allowed us to model these ES7s using our hybrid approach in spite of the complexity. Here, we illustrate the general features of the modeling approach by presenting the specific method used to determine the structure ES7 of *G. gallus* (Supplementary Text). This ES7 contains a unique extension segment ES7d1 located in between ES7c and ES7d. We further describe the 2D and 3D models of the entire LSU rRNA for *M. musculus* and compare it with the previous predictions in the context of the experimental structures available for *H. sapiens*.

Comparative analysis of ES7s from Groups 1–3

Validation of the 2D models of ES7 (Figure 2) was accomplished by computing sequence conservation and covariation as well as by quantitative comparison of predicted and experimental 3D structures. Conservation and co-variation analyses were performed independently for ES7s of each Group 1–3 from MSAs in Supplementary Data Set 4. Structural comparisons were performed for species whose ES7 structures were modeled here or were previously determined, as specified in Table 1.

Conservation and co-variation scores. To estimate the accuracy of 2D structures we computed conservation and co-variation scores from the ES7 MSA (see Supplementary Materials and Methods). The conservation score used here is a gap adjusted Shannon entropy (GASE), which is a measure of conservation of nucleotide identity (2). The co-variation score PASE (pair adjusted Shannon entropy) characterizes paired nucleotides by their adherence to co-variation. Thus, a nucleotide position is considered conserved if it is always canonically paired, even if its nucleotide identity is not conserved. The computed scores (ranging from 0 to 2, where 0 represents universally conserved) were mapped onto ES7 representatives for Group 1 (*S. cerevisiae*), Group 2 (*D. melanogaster*) and Group 3 (*H. sapiens*), (Figure 5; Supplementary Figures S8–S10).

The structure of the Signature Fold is highly conserved. Base pairing within the Signature Fold is reasonably predicted in all groups despite variability of sequence. Conservation/co-variation metrics of ES7s are highest within the Signature Fold, and the covariation score is substantially higher than the conservation score (Figure 5). The base pair Shannon Entropy (BPSE, Figures S8B, S9B and S10B) and combined PASE score (Figures 5B, D and F) reveal a moderate to high degree of the base pairing conservation (regardless of the specific identity of the base pairs) within the Signature Fold, thus providing the relationship between the structural data and the sequence information within the MSAs for this nearly universal region of the ES7 Helical junctions in Groups 1 and 2 tend to be less conserved than the base paired regions. ES7s of group 1 exhibit moderate conservation and covariation despite their non-monophyletic origins.

Helical stems (ES7a–d) from Group 2, outside the Signature Fold, exhibit little conservation or co-variation, reflective of the polyphyletic nature of this group (Figure 5C and D). An abundance of irregularities of helical stems by bulges or internal loops pose challenges for predicting these elements by a conventional co-variational approach for remotely related species, as the irregularities vary between the species and complementary strands cannot be easily identified from the MSAs. We note that the co-variational method has been successfully applied to a subset of closely related species from Group 2 (e.g. within Hymenoptera) (60) due to high structural similarity of the extended regions.

Conservation/co-variation metrics for the respective helical stems of Group 3 (ES7a–d, Figure 5E and F) are higher for than those for Group 2 The Group 3 specific stems (ES7e–h) also reveal moderate to high conservation/covariation (if present). Overall, spatial alignment of the 3D elements of Group 3 is outstanding (Supplementary Data Set 3), especially in a subgroup of species that represents endothermal vertebrates. A majority of tentacles exhibit substantial sequence conservation (especially in mammalian species). Tentacle ES7b is highly variable is structure and ES7d is the least conserved in sequence. Due to high sequence conservation, co-variational analysis has limited power for species in Group 3, as co-variation signals cannot be detected.

Structural analysis by RMSD metrics and long-range interactions. Comparison of the root-mean square deviations

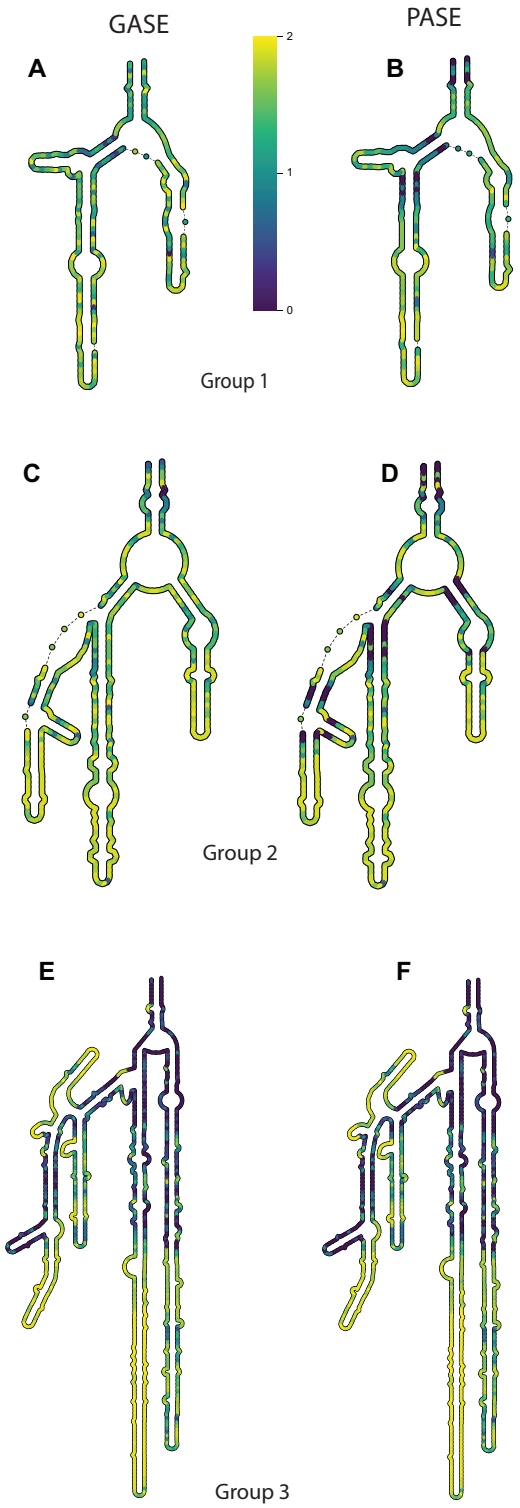

Figure 5. Conservation (GASE) and co-variation (PASE) scores mapped onto the secondary structures of ES7 from Groups 1–3. (**A**) GASE and (**B**) PASE computed from ES7 MSA of Group 1 and mapped onto ES7 of *S. cerevisiae*; (**C**) GASE and (**D**) PASE computed from ES7 MSA of Group 2 and mapped onto ES7 of *D. melanogaster*; (**E**) GASE and (**F**) PASE computed from ES7 MSA of Group 3 and mapped onto ES7 of *H. sapiens*. Both scores range from 0 (dark blue, absolute conservation or co-variation) to 2 (yellow, random signal, no conservation or covariation). Intermediate values are indicated by the color bar. The ES7 MSAs of Groups 1–3 are given in Supplementary Data Set 4.

Table 2. Structural analysis of experimental and modeled structures by RMSD metrics

Modeled/ experimental structures	RMSD, Å	Experimental/ experimental structures	RMSD, Å
C. albicans (1)[a]/ *T.thermophilia* (1)	6.713	*T. thermophilia* (1)/ *D.melanogaster* (2)	5.794
A. albopictus (2)/ *D.melanogaster* (2)	4.934	*T. thermophilia* (1)/ *H.sapiens* (3)	5.700
M. musculus (3)/ *H.sapiens* (3)	4.224	*D. melanogaster* (2)/ *H.sapiens* (3)	6.831

[a]Group numbers as defined in Table 1 are shown in parentheses.

(RMSDs) of atomic positions confirms that the structure of the ES7 Signature Fold is conserved among all species; structural differences between predicted and experimental structures (RMSD = 4–7 Å) are comparable to differences between experimental structures (RMSD = 5–7 Å), Table 2, Supplementary Table 2 and Supplementary Figure S11.

The RMSDs between experimental and modeled structures within each group reveal important patterns:

(a) The Signature Fold is structurally conserved (Figure 4).
(b) Junctions within the Signature Fold for Groups 1 and 2 vary in sequence and reveal subtle variations in 2D structure; they may substantially differ within a group (e.g. *T. thermophilia* versus *C. albicans*, Group 1; *D. melanogaster* versus *A. albopictus*; Group 2); 3D junctions of species within Group 3 tend to be highly conserved (*H. sapiens* versus *M. musculus*).
(c) In Group 2, helical stems of ES7 that link to the Signature Fold vary substantially in their structures because of irregular local elements. Each RNA helix (7a, 7b or 7c) is characterized by an idiosyncratic set of irregularities, making them quite different from the helical elements within the common core, where local motifs within the helices are generally well-defined (within a given taxonomic domain).
(d) Structural deviation between helical stems of Group 2 is greater than between those of Group 3. Species allocated to Group 2 are not monophyletic, and different local patterns likely reflect convergent evolution. Conservation within the helical elements of the Signature Fold varies between the groups, with Group 3 being most conserved.

More similar structures tend to preserve long range interactions in 3D. Base-phosphate and base-sugar contacts have been extracted from the modeled and experimental 3D structures of ES7 for Groups 1–3 and mapped onto the 2D structure (Supplementary Figures S12 and S13. These long-range interactions tend to be somewhat similar for the structures that belong to the same group and vary between the groups. Thus, despite the overall high structural conservation of the ES7 Signature Fold, the specific features (base pairing pattern, junction architecture, specific long-range interactions) are not fully preserved.

Patterns and exceptions

We observed informative patterns within the ES7 Groups, which show group-specific personalities. Comparison of sequence and structure revealed that across Groups 1–3 (Figure 5, Table 2), the size and complexity of ES7 is interdependent with sequence conservation, co-variation, heterogeneity of the local motifs within helical elements, and structural variability of the junctions and extended regions.

Group 1 ES7s are moderately conserved in sequence and highly conserved in structure. Group 2 ES7s are diverse in sequence with modest diversity in structure. Group 3 ES7s are conserved in both sequence and structure, in common regions. Group 3 ES7s show extensive variation in morphology and length of tentacles, especially in mammals, suggesting that the rate of accretion is high relative to the rate of change of sequence. It was instructive to model ES7 of each group independently, accounting for their unique sets of features.

rRNA growth mechanisms

The acquisition of the eukaryotic shell, on the surface of the common core, is a seventh phase of ribosomal evolution (2,4). This phase of rRNA evolution produced ES7 Groups 1 and 2. The growth of tentacles, which are long double-helical protrusions seen in endothermic eukaryotes, marks an eighth phase of ribosomal evolution. This phase of rRNA evolution produced ES7 Group 3.

The accretion model is consistent with the following constraints on ES7: (i) the core of ES7, the Signature Fold, is conserved in 2D and 3D structure and is universal to eukaryotes, (ii) new rRNA is added onto pre-existing ES7 rRNA by accretion processes such that addition of new rRNA fragments does not remodel pre-existing ES7 rRNA, (iii) 2D and 3D structures of ES7 in common between daughter species are ancestral and (iv) 3D structure of ES7 is more conserved than sequence (5).

The diversity of ribosomal structures and their known phylogenetic relationships have allowed construction of ES7 accretion pathways. We assume that rRNA ancestral states can be reasonably modeled by elements that are shared among all daughter species. This assumption suggests that ES7s of simple extant species reasonably represent those of extinct ancestors. We have ranked extant ES7s by size and complexity.

The *H. sapiens* accretion pathway (*hs*-accretion pathway) of ES7 describes a process that initiates with H25 of the Last Universal Common Ancestor (LUCA) (61) and terminates with ES7 of extant *H. sapiens* (Figure 6). Our best estimate of ES7 at LUCA is H25 of bacterial ribosomes (*E. coli* is used here as a prototype). Helix 25 is universal to all cytosolic ribosomes. The first step from LUCA in the *hs*-accretion pathway is inferred from archaeal ribosomes where H25 contains a micro-ES (μ-ES). μ-ES's of 5–30 nucleotides are commonly observed in archaeal ribosomes at sites of much larger eukaryotic ESs (Figure 6). An insertion fingerprint in H25 of archaeal ribosomes at the base of the μ-ES marks the site of archaeal expansion of the H25 stem. Additional structures of ES7 over the eukaryotic tree of life support an *hs*-accretion pathway in which further

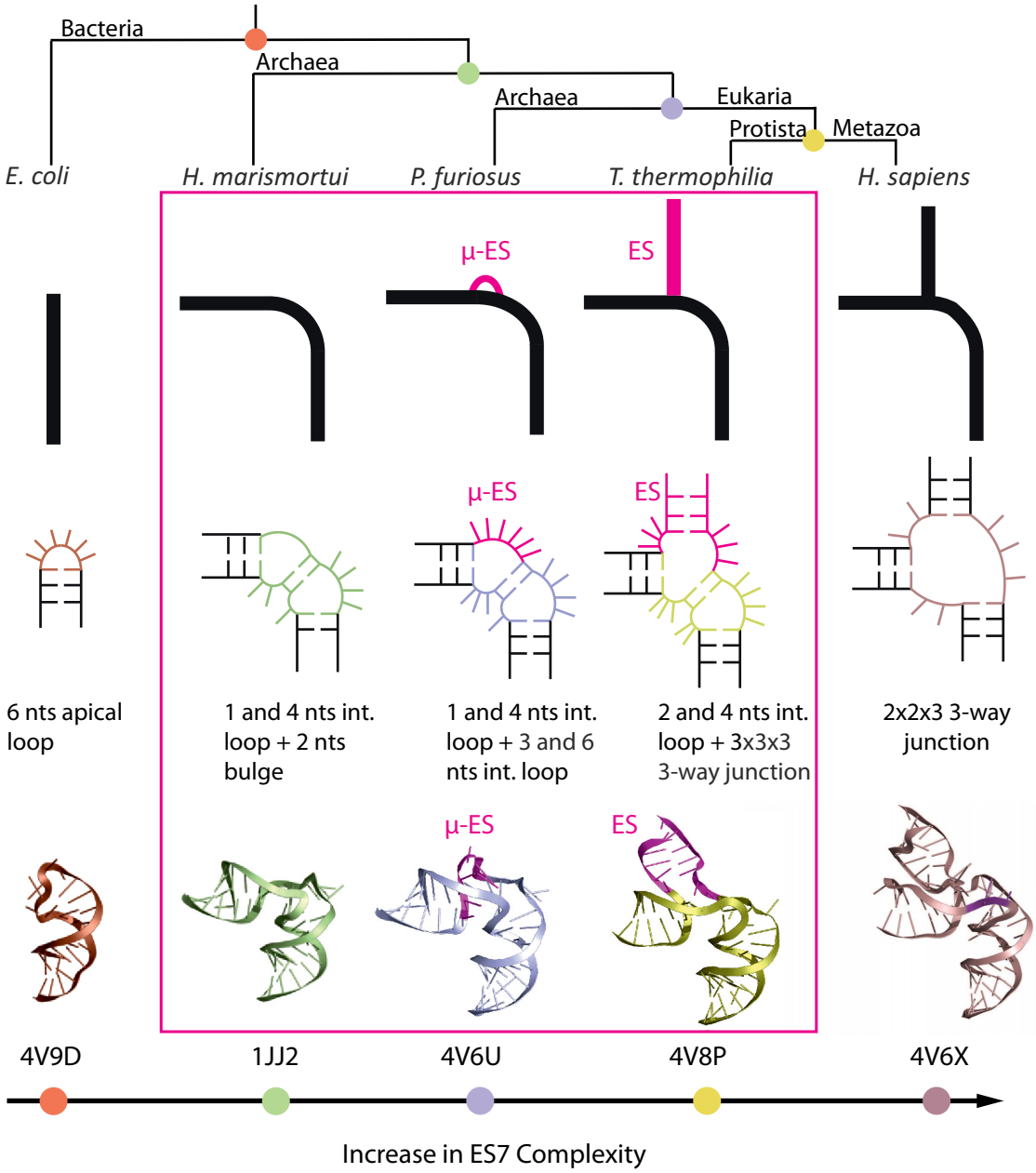

Figure 6. RNA accretion in ES7, based on comparison of experimental 3D structures and their evolutionary relationships. The lineage leading to *H. sapiens (Metazoan)* is highlighted. The rRNA at each ancestral node, highlighted by colored circles, is approximated by rRNA that is common to daughters (2). rRNA is depicted schematically, in secondary structures, and in three dimensional structures. A stem loop of H25 (*E. coli*) establishes the base of ES7. An expansion of the apical loop extends H25 in *H. marismortui;* a bulge expands into an internal loop in *P. furiosus*, which extends into a new expansion ES7a in *T. thermophilia*; this element undergoes further adjustments in *H. sapiens*. PDB IDs of the source structures are indicated below the 3D structures. Chain IDs and nucleotide numbers are given in Supplementary Table S3. The 3D structures are compiled in Movie 2.

extension of this internal loop leads to a three-way junction (ES7a/ES7b, *T. thermophilia*). This three-way junction is further elaborated by tertiary interactions between the branched stem and the underlying trunk (*H. sapiens*; Figure 6; Movie 2).

The *hs*-accretion pathway can be expanded to a fine-grained progression (Figure 7), of elementary steps, by interleaving structures of the phylogenetic accretion process with transitional structures extracted from the RNA structural database (41,42). Elementary steps of RNA growth include the accretion of one (or several) nucleotides, yielding a bulge. The bulge is expanded by insertion of additional nucleotides. Repetition of nucleotide accretion into the same bulge region results in more complex bulges (if insertion occurs in the same RNA strand) or internal loops (if the insertion appears in the opposing strand). Finally, accretion of a critical number of nucleotides (7–9) results in extrusion of a new helix and formation of a three-way junction. This process converts one RNA topology to another in a series of elementary steps that maintain the underlying struc-

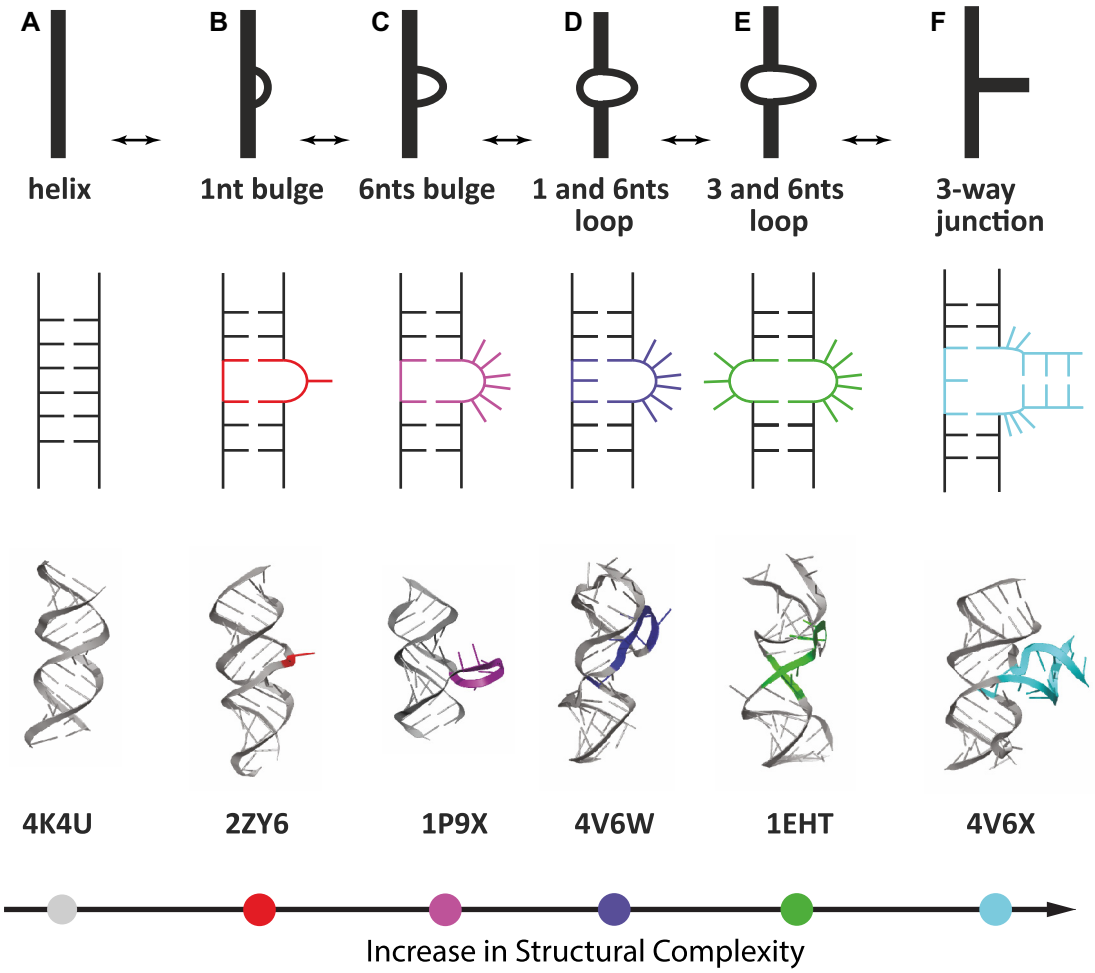

Figure 7. A three-way junction emerges from a helix through a series of elementary steps of rRNA accretion. Existing and emerging elements are depicted by their secondary and 3D structures and by schematic representations. (**A**) Double helical region; (**B**) 1 nt bulge; (**C**) 6 nt bulge; (**D**) 1 and 6 nt loop; (**E**) 3 and 6 nt loop; (**F**) three-way junction. This trajectory is based on experimental 3D structures that are ordered by size. The evolutionary relationship is inferred. PDB IDs of the source structures are indicated below the structures. Chain IDs and nucleotide numbers are given in Supplementary Table S4.

ture. The construction of this fine-grained molecular progression during rRNA evolution is analogous to structure-based descriptions of evolutionary processes (62,63) and macromolecular dynamics (64–66) that are inspired by the concept of reaction coordinates for chemical transformation (67).

LSU of *M. musculus*

To illustrate the utility and generality of our approach, we modeled the 2D and 3D structures of the LSU rRNA of *M. musculus*. This rRNA contains large eukaryotic ESs that contribute to a total LSU rRNA length of 4887 nts. 2D structures of all ESs were modeled by methods analogous to those described above for ES7, using the LSU rRNA of *H. sapiens* (68), which has been experimentally determined by Cryo-EM (4), as a guide.

The 2D model of *M. musculus* LSU rRNA presented here differs from the previous models (23,28).The differences are discussed in the SI text and Supplementary Figure S14. Comparison of the 2D structures of *M. musculus* LSU rRNA (Figure 8) and *H. sapiens* (Figure 1B) demonstrates their similarity. Yet, some helices of *M. musculus* rRNA are shorter than in *H. sapiens* (e.g. ES7a,b,d,h, ES15, ES27a, ES27b), while some are absent (ES7g, ES15a, ES30a). The result lends support to the accretion model and is consistent with a rapid growth rate of rRNA in mammalian lineages (2). We anticipate that the *M. musculus* LSU rRNA 2D and 3D structures will be useful for explaining and predicting a variety of data.

ES7 of M. musculus. *M. musculus* LSU rRNA contains one of the largest known ES7s (691 nts). 3D modeling produced a *M. musculus* ES7 structure nearly superimposable on that of *H. sapiens* (Movie 1, Supplementary Data Set 3). A comparison of the tentative 2D structure with the final 2D structure extracted from a 3D model reveals minor peculiar discrepancies discussed in the Supplementary Text and shown in Supplementary Figure S15. The tips of expansion segments ES7a,b in both *H. sapiens* and *M. musculus* are enriched in G and C and polarized in their distribution, so that one strand is G rich, while the other is C rich. Such polar-

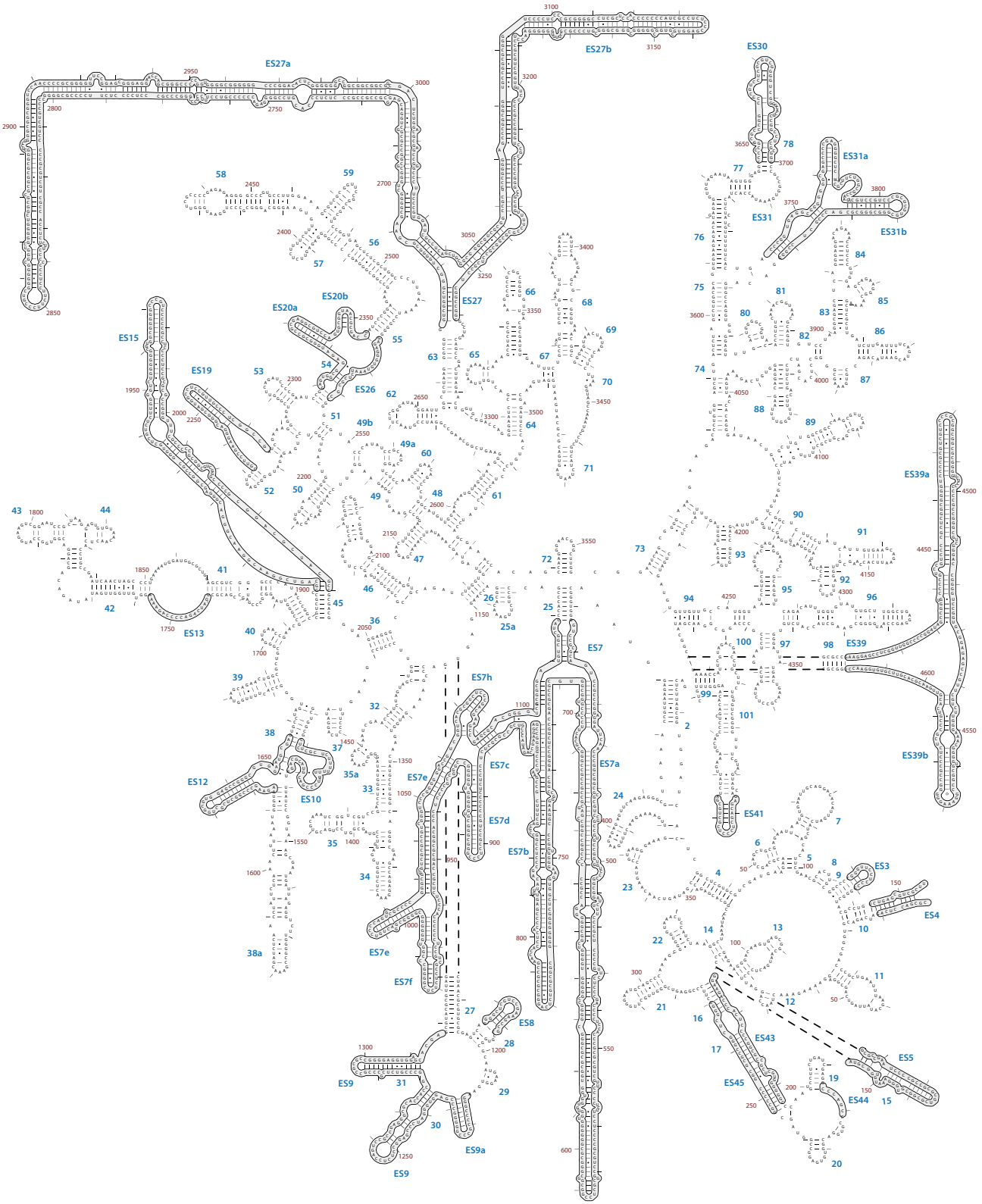

Figure 8. Secondary structure of the LSU rRNA of *M. musculus.* Expansion segments are outlined.

ization may promote the folding of the long expansions and may accelerate the formation of G- quadruplexes (25,26).

***Complete 3D model of the LSU rRNA of* M. musculus.** To further illustrate the utility and universality of our approach, we modeled the 3D structure of the *M. musculus* LSU rRNA, using the 2D structure as a basis. The resulting *M. musculus* rRNA model is nearly superimposable on that of *H. sapiens* (PDB ID 4V6X), revealing an overall good agreement (RMSD = 4.020 Å) of the common core regions and several tentacle-like expansion segments (ES7a, ES27a,b, ES39a). Yet, some of the tentacles were found to have different directionalities in the two structures. Examples of such tentacles include ES7b,e,f. These differences could arise from the lack of interactions with proteins, variability in the local base pairing, high mobilities and polymorphism of the tentacles themselves, or limitations of our modelling approach.

ES7 forms pseudoknots with ES15 and ES39. In rRNA from *H. sapiens* (PDB ID 4V6X), the first pseudoknot is formed by C1289 from ES7b and G4941 and C4942 from ES39 (Supplementary Figure S16). In the modeled structure of the LSU rRNA from *M. musculus*, the nucleotides equivalent to those in the ES7/ES39 pseudoknot are found to be in proximity (∼7 Å) (Supplementary Figure S16). The structural analysis of ribosomal complexes available in RiboVision reveals that the ES7/ES39 pseudoknot exists in eukaryotic species that represent major phylogenetic groups (mammalians, insects, fungi, plants, SAR).

The second pseudoknot is formed by base pairing between G963 from ES7 and C2250 from ES15. This pseudoknot has been observed in the ribosomes of mammalian species (4) and some kinetoplastids (18).

The corresponding nucleotide region in ES15 of *M. musculus* is significantly shorter than that of *H. sapiens* (Supplementary Figure S17). However, our prediction locates corresponding regions of ES15 and ES7 in proximity to each other (∼3 Å).

Both ES7/ES39 and ES7/ES15 pseudoknots are also stabilized by eukaryotic specific protein eL6 (Supplementary Figure S18), whose globular domain binds to ES39 and the intrinsically disordered tail wraps around ES7 and ES15 (in mammalian species) providing the tight structural integration of the eukaryotic specific motif ES7/ES15/ES39/eL6. Previous analyses have suggested that the components of this region of eukaryotic LSU have co-evolved (4,59). While formation of intact pseudoknots in the absence of rProteins is not expected in our models without introduction of specific constraints (69) (see Discussion below), the relative proximity of the corresponding fragments illustrates the power, universality, and robustness of the presented hybrid approach for ribosomal systems of any size.

DISCUSSION

The ES7 Signature Fold

ES7 is largest and the most complex ES and contains the most diverse rRNA over phylogeny. Yet, with the exception of ribosomes from some parasites, the ES7 Signature Fold (Figure 4) is highly conserved in 2D and 3D structures in all eukaryotes. The ES7 Signature Fold provides a foundation for rRNA elaboration and is a platform for helical elongation and branching. ES7 of species in Group 1 (Table 1) is essentially comprised of the Signature Fold.

Despite overall conservation, the ES7 Signature Fold varies over phylogeny by local differences in secondary structure (single nucleotide insertions, base pair disruptions) and long range interactions (base-sugar and base-phosphate). ES7 Signature Fold rRNA is engaged in the interactions with other components of the LSU, such as rRNA of ES39 and rProtein eL6. We used invariant Base Pair Shannon Entropy (2) to validate the ES7 Signature Fold, which was initially identified at the level of 3D structure. Our data confirms that covariation can predict 2D structures of some closely related species that contain divergent sequences. ES7 size correlates with size of other ESs, especially ES15, ES27 and ES39.

Beyond the ES7 Signature Fold

Elaboration of ES7 over phylogeny is achieved primarily by elongation of Signature Fold helices and insertion of stem loops into the elongated helices. Stem loops that insert outside the Signature Fold (as in Groups 2 and 3, Table 1) exhibit moderate conservation of position. These stem loops demonstrate substantial variation in length, and sequence composition over phylogeny, and are characterized by small insertions as well as variations in base pairing (even for closely related species such as *D. melanogaster* and *A. albopictus*).

2D models of ES7 were described previously by Gutell (30,70) for a small subset of phyla. These models are accurate for Hymenoptera (29,60), Acari (71) and some protists (72) even though sequence divergence, size and branching complexity presented challenges. However, structures of ES7s from many mammals and birds remained unresolved until now. We note that the 2D structure of ES7 of *M. musculus* presented here ($ES7^{Petrov}$) is substantially different from an early model by Bachellerie ($ES7^{Bachellerie}$) (9). $ES7^{Bachellerie}$ contains two four-way junctions, while $ES7^{Petrov}$ contains several three-way junctions (Supplementary Figure S14), and resembles a morphology of ES7 of *H. sapiens*.

Convergent evolution of ES7

Phylogenetic comparison of ES7s reveals evidence of convergence. Polyphyletic distribution of ES7 features from Groups 2 and 3 across phylogeny suggests that some insertions took place in parallel. Analogous elaborations, that project from the Signature Fold and result in extension of ES7abc and branching of ES7d, are broadly dispersed over phylogeny and are found in insects (*D. melanogaster* and *A. albopictus*), plants (*A. thaliana* and *O. sativa*) and fungi (*C. neoformans*). These elaborations are not observed in rRNAs of the deepest ancestral nodes. Indicators of convergent evolution are also seen for tentacles of Group 3.

Long tentacles are attached at different sites in rRNAs of different species. ES7 of *G. gallus* contains one long tentacle (ES7d1) attached at the position of a small stem-loop (that branches form ES7d) in mammalian rRNAs. ES7 of *H. sapiens* contains two long tentacles (ES7a and ES7b),

which attach at sites of shorter tentacles in the rRNA of *G. gallus* (Supplementary Figure S7). The comparison suggests similar function of highly elongated extensions that transcends specific sites of attachment.

Biological roles of ES7

Ribosomal expansion in eukaryotes might arise from conventional adaptive pressures, but might also be explained in part by Constructive Neutral Evolution (CNE) (73–75). CNE is a multi-step process in which neutral, non-adaptive change opens capacities for complementation and co-dependency. In a first CNE step, an intrinsic or environmental change would alter ribosomal structure, expanding the rRNA, with little or no gain in function. This rRNA expansion might not initially confer advantage. Subsequent steps would stumble upon functions for the expansion segment, ratcheting the complexity.

Experimental manipulation of rRNA of eukaryotic ribosomes *in vivo* is problematic, complicating the determination of functions (76). Nevertheless, expansion segment ES7 is associated with multiple functions, many of which have been recently highlighted for fungal and protist ribosomes (Group 1). Thus, removal of ES7 has been found to be lethal for fungi (77–79). ES7 interacts with ribosomal protein eL14 via ES7b (16). The tip of ES7b can be in proximity to a flexible L7/12-stalk and may transiently interact with ribosomal protein uL10 (80). It has been suggested that conformational changes of fungal ES7 (together with ES39) affect binding of the Signal Recognition Particle, facilitating a conformational switch of ES7/ES39 (81). ES7 has been found to interact with ribosomal assembly factors Nsa1p, Mak16p and Rpf1p in fungi, highlighting its functional importance in early ribosomal assembly (82–85). Recently, it has been shown that ES7 in fungi interacts with the assembly factor Noc2p (86) in functionally coordinated modules in the early 66S pre-ribosome, suggesting that ESs may have coevolved with ribosomal assembly factors to drive differentiation of the pathways of ribosomal biogenesis. ES7 is cleaved by Fe^{2+}, causing a rapid hydrolysis of ribosomal particles under oxidative stress (87). ES7a with ES27a help to localize N-terminal acetylases (88) and influence translation fidelity (78).

ES7 of higher eukaryotes gained additional functions. Ribosomes of Group 3 (Table 1), but not those of Groups 1 and 2 contain ES7e. This stem-loop plays a role in selenoprotein synthesis by interacting with SECIS-binding protein SBP2 via a K-rich motif found to be highly conserved in vertebrates (89). Sequences of the K-rich motif of SBP2 vary between Group 2 and Group 3 (90). Based on these observations we suggest a co-evolution of ES7e and SBP2 (90).

A recent bioinformatics study suggests that the substantial growth of ESs (including ES7) in mammals may reflect a gain of function related to interactions with non-ribosomal RNAs (91). ES7 contains a multitude of antisense matches (longer than 10nts) to many 5'-UTRs of human mRNAs as well as G-rich tracts which can facilitate triple helix RNA formation.

The structural models of ES7 provided here could be useful for designing, testing, confirming, or rejecting biological or evolutionary hypotheses as well as for modeling the interactions between the expansion segments and other constituents of the translational machinery at the atomic level of details (*e.g.* rationalizing the data from the pull down experiments).

Evaluation of the hybrid modeling approach

Our hybrid approach predicts 2D and 3D structure despite impediments by dynamics and conformational heterogeneity, large size, sequence hypervariability, high GC content, and absence of covariation signal, and thus is a powerful pipeline for modeling both 2D and 3D structures of rRNAs. The ES7 models presented here appear to correctly capture the vast majority of 2D and 3D elements of ES7, and are useful in interpreting and predicting a variety of data. The hybrid method can be applied to other eukaryotic rRNA expansions as exemplified by modeling of the rRNA LSU from *M. musculus*. Inaccuracies are most pronounced in tertiary interactions between helical motifs, helical and unpaired motifs, or non-paired motifs (such as A-platforms, ribose zippers or pseudoknots). If such interactions are experimentally detected, they could be constrained by incorporation of additional data into RNAComposer modeling (69). The methodology proposed in the current work cannot predict G-quadruplexes. Our 3D modelling may be biased towards high levels of topological similarity between related species (*H. sapiens vs S. cerevisiae or H. sapiens vs C.* albicans*)*; stem c (*D. melanogaster* and *A. albopictus*). 2D models are based on conservation of junctions, and RNAComposer predictions rank topology over sequence. Models are predicted in isolation from proteins. The absence of proteins can result in errors when compared with experimental structures, since proteins can stabilize RNA structure or alternatively prohibit formation of canonical base pairs.

CONCLUSIONS

The ES7 is the most diverse expansion segment within eukaryotic ribosomes. It is the product of intense restraints and convergent evolution, and is stringently maintained as helices accrete and elongate. ES7 structures have been produced here by efforts that integrated techniques and constrains from 1D, 2D, and 3D-based algorithms and datasets. The data support an evolutionary model in which ES7 of metazoan lineages emerged through a series of elementary and recursive growth events. As a general pattern, a given stem-loop, once introduced into an ancestral rRNA, appears to freeze, and remain conserved in rRNAs of daughter species. Similar 2D topologies and 3D morphologies can be achieved by a variety of sequences. Common ES7 morphology is observed among species with significant differences in sequence.

Expansions that perturb underlying ribosomal structure or functional regions of the ribosome are not observed. The number of distinct topological and morphological solutions exploited by nature appears to be limited. Comparison of ES7 structures across phylogeny reveals evidence of convergent evolution. The existence of a broad range intermediate structures in non-related RNAs suggest that nucleotide-level growth of the rRNA expansion segments was an incre-

mental and reversible process in which many closely related states were sampled.

Our work is inspired by pioneering efforts of Gutell (29,30,72,92). The identification of the ES7 Signature Fold of eukaryotes and the inference of rRNA accretion during evolution (5,93) enabled our prediction of 2D and 3D structures for diverse eukaryotes. Our hybrid method is facilitated by the availability of 3D structures of ES7 for diverse eukaryotic species and the ability of RNAComposer to predict RNA 3D structures when structural templates are missing. Our results provide a framework for further development of high-throughput predictions using statistical approaches (Infernal, R-scape) (94–96).

The structures modelled here reveal principles of rRNA accretion observed in the experimentally derived structures, and support the Accretion Model of ribosomal evolution (5,93,97). A recent experimental Cryo-EM structure of a plant Solanum lycopersicum (98) confirmed our predictions for ES7 of Group 3. Since the hybrid approach proposed here appears to successfully predict the largest eukaryotic expansion segments, it should be capable of resolving structures of ES's of essentially any rRNA, as demonstrated by the model of the LSU rRNA from *M. musculus*. The models predicted in the current study can now be used as templates in R2DT webserver (33) to derive accurate secondary structures of the other species.

DATA AVAILABILITY

The 2D and 3D models produced in the current study are available as Supplementary Datasets.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We would like to acknowledge Dr Elena Rivas for helpful discussions.

FUNDING

National Aeronautics and Space Administration [80NSSC18K1139 to L.D.W., A.S.P.]; National Science Center Poland [2012/06/A/ST6/00384]; Polish Ministry of Science and Higher Education, under the KNOW programme; Foundation for Polish Science [HOMING PLUS/2012–6/12 to K.J.P.]. Funding for open access charge: NASA [80NSSC18K1139].
Conflict of interest statement. None declared.

REFERENCES

1. Melnikov,S., Ben-Shem,A., Garreau de Loubresse,N., Jenner,L., Yusupova,G. and Yusupov,M. (2012) One core, two shells: bacterial and eukaryotic ribosomes. *Nat. Struct. Mol. Biol.*, **19**, 560–567.
2. Bernier,C.R., Petrov,A.S., Kovacs,N.A., Penev,P.I. and Williams,L.D. (2018) Translation: the universal structural core of life. *Mol. Biol. Evol.*, **35**, 2065–2076.
3. Bowman,J.C., Petrov,A.S., Frenkel-Pinter,M., Penev,P.I. and Williams,L.D. (2020) Root of the tree: the significance, evolution, and origins of the ribosome. *Chem. Rev.*, **120**, 4848–4878.
4. Anger,A.M., Armache,J.P., Berninghausen,O., Habeck,M., Subklewe,M., Wilson,D.N. and Beckmann,R. (2013) Structures of the human and drosophila 80S ribosome. *Nature*, **497**, 80–85.
5. Petrov,A.S., Bernier,C.R., Hsiao,C., Norris,A.M., Kovacs,N.A., Waterbury,C.C., Stepanov,V.G., Harvey,S.C., Fox,G.E., Wartell,R.M. *et al.* (2014) Evolution of the ribosome at atomic resolution. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, 10251–10256.
6. Ware,V.C., Tague,B.W., Clark,C.G., Gourse,R.L., Brand,R.C. and Gerbi,S.A. (1983) Sequence analysis of 28S ribosomal DNA from the amphibian *xenopuslaevis*. *Nucleic Acids Res.*, **11**, 7795–7817.
7. Clark,C.G., Tague,B.W., Ware,V.C. and Gerbi,S.A. (1984) *Xenopuslaevis* 28S ribosomal RNA: a secondary structure model and its evolutionary and functional implications. *Nucleic Acids Res.*, **12**, 6197–6220.
8. Gerbi,S.A. (1996) In: Zimmermann,R.A. and Dahlberg,A.E. (eds). *Ribosomal RNA—Structure, Evolution, Processing, and Function in Protein Synthesis*. CRC Press, Boca Raton, FL, pp. 71–87.
9. Hassouna,N., Michot,B. and Bachellerie,J.P. (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.
10. Han,H., Schepartz,A., Pellegrini,M. and Dervan,P.B. (1994) Mapping RNA regions in eukaryotic ribosomes that are accessible to methidiumpropyl-edta-fe(ii) and EDTA-Fe(II). *Biochemistry*, **33**, 9831–9844.
11. Larsson,S.L. and Nygård,O. (2001) Proposed secondary structure of eukaryote specific expansion segment 15 in 28S rRNA from mice, rats, and rabbits. *Biochemistry*, **40**, 3222–3231.
12. Alkemar,G. and Nygård,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 (New York, NY)*, **10**, 403–411.
13. Alkemar,G. and Nygård,O. (2006) Probing the secondary structure of expansion segment ES6 in 18S ribosomal RNA. *Biochemistry*, **45**, 8067–8078.
14. Spahn,C.M., Beckmann,R., Eswar,N., Penczek,P.A., Sali,A., Blobel,G. and Frank,J. (2001) Structure of the 80S ribosome from *saccharomycesCerevisiae*:trna-Ribosome and subunit-subunit interactions. *Cell*, **107**, 373–386.
15. Armache,J.P., Jarasch,A., Anger,A.M., Villa,E., Becker,T., Bhushan,S., Jossinet,F., Habeck,M., Dindar,G., Franckenberg,S. *et al.* (2010) Cryo-EM structure and rRNA model of a translating eukaryotic 80S ribosome at 5.5-Angstrom resolution. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 19748–19753.
16. Ben-Shem,A., Garreau de Loubresse,N., Melnikov,S., Jenner,L., Yusupova,G. and Yusupov,M. (2011) The structure of the eukaryotic ribosome at 3.0 Å resolution. *Science*, **334**, 1524–1529.
17. Klinge,S., Voigts-Hoffmann,F., Leibundgut,M., Arpagaus,S. and Ban,N. (2011) Crystal structure of the eukaryotic 60S ribosomal subunit in complex with initiation factor 6. *Science*, **334**, 941–948.
18. Vicens,Q., Bochler,A., Jobe,A., Frank,J. and Hashem,Y. (2021) In: Harris,J.R. and Marles-Wright,J. (eds). *Macromolecular Protein Complexes III: Structure and Function*. Springer International Publishing, Cham, pp. 433–450.
19. Ban,N., Nissen,P., Hansen,J., Moore,P.B. and Steitz,T.A. (2000) The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science*, **289**, 905–920.
20. Watson,Z.L., Ward,F.R., Méheust,R., Ad,O., Schepartz,A., Banfield,J.F. and Cate,J.H. (2020) Structure of the bacterial ribosome at 2 Å resolution. *Elife*, **9**, e60482.
21. Harms,J., Schluenzen,F., Zarivach,R., Bashan,A., Gat,S., Agmon,I., Bartels,H., Franceschi,F. and Yonath,A. (2001) High resolution structure of the large ribosomal subunit from a mesophilic eubacterium. *Cell*, **107**, 679–688.
22. Selmer,M., Dunham,C.M., Murphy,F.V., Weixlbaumer,A., Petry,S., Kelley,A.C., Weir,J.R. and Ramakrishnan,V. (2006) Structure of the 70S ribosome complexed with mRNA and tRNA. *Science*, **313**, 1935–1942.
23. Michot,B., Hassouna,N. and Bachellerie,J.P. (1984) Secondary structure of mouse 28S rRNA and general model for the folding of the large rRNA in eukaryotes. *Nucleic Acids Res.*, **12**, 4259–4279.
24. Parker,M.S., Sallee,F.R., Park,E.A. and Parker,S.L. (2015) Homoiterons and expansion in ribosomal RNAs. *FEBS Open Biol*, **5**, 864–876.

25. Mestre-Fos,S., Ito,C., Moore,C.M., Reddi,A.R. and Williams,L.D. (2020) Human ribosomal G-quadruplexes regulate heme bioavailability. *J. Biol. Chem.*, **295**, 14855–14865.
26. Mestre-Fos,S., Penev,P.I., Suttapitugsakul,S., Hu,M., Ito,C., Petrov,A.S., Wartell,R.M., Wu,R. and Williams,L.D. (2019) G-quadruplexes in human ribosomal RNA. *J. Mol. Biol.*, **431**, 1940–1955.
27. Melnikov,S.V., Manakongtreecheep,K., Rivera,K.D., Makarenko,A., Pappin,D.J. and Söll,D. (2018) Muller's ratchet and ribosome degeneration in the obligate intracellular parasites microsporidia. *Int. J. Mol. Sci.*, **19**, 4125.
28. Holmberg,L., Melander,Y. and Nygård,O. (1994) Probing the structure of mouse ehrlich ascites cell 5.8S, 18S and 28S ribosomal RNA in situ. *Nucleic Acids Res.*, **22**, 1374–1382.
29. Gillespie,J.J., Johnston,J.S., Cannone,J.J. and Gutell,R.R. (2006) Characteristics of the nuclear (18S, 5.8S, 28S and 5S) and mitochondrial (12S and 16S) rRNA genes of *apismellifera* (Insecta: hymenoptera): structure, organization, and retrotransposable elements. *Insect Mol. Biol.*, **15**, 657–686.
30. Cannone,J.J., Subramanian,S., Schnare,M.N., Collett,J.R., D'Souza,L.M., Du,Y., Feng,B., Lin,N., Madabusi,L.V., Muller,K.M. *et al.* (2002) The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinf.*, **3**, 2.
31. Hashem,Y., des Georges,A., Fu,J., Buss,S.N., Jossinet,F., Jobe,A., Zhang,Q., Liao,H.Y., Grassucci,R.A., Bajaj,C. *et al.* (2013) High-Resolution cryo-electron microscopy structure of the *trypanosomabrucei* ribosome. *Nature*, **494**, 385–389.
32. Will,S., Joshi,T., Hofacker,I.L., Stadler,P.F. and Backofen,R. (2012) LocARNA-P: accurate boundary prediction and improved detection of structural RNAs. *RNA (New York, NY)*, **18**, 900–914.
33. Sweeney,B.A., Hoksza,D., Nawrocki,E.P., Ribas,C.E., Madeira,F., Cannone,J.J., Gutell,R., Maddala,A., Meade,C.D., Williams,L.D. *et al.* (2021) R2DT is a framework for predicting and visualising RNA secondary structure using templates. *Nat. Commun.*, **12**, 3494.
34. Bernier,C.R., Petrov,A.S., Waterbury,C.C., Jett,J., Li,F., Freil,L.E., Xiong,X., Wang,L., Migliozzi,B.L., Hershkovits,E. *et al.* (2014) Ribovision suite for visualization and analysis of ribosomes. *Farad. Discuss.*, **169**, 195–207.
35. Zuker,M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, **31**, 3406–3415.
36. Popenda,M., Szachniuk,M., Antczak,M., Purzycka,K.J., Lukasiak,P., Bartol,N., Blazewicz,J. and Adamiak,R.W. (2012) Automated 3D structure composition for large RNAs. *Nucleic Acids Res.*, **40**, e112.
37. Moran,N.A. (2002) Microbial minimalism: genome reduction in bacterial pathogens. *Cell*, **108**, 583–586.
38. McCutcheon,J.P. and Moran,N.A. (2012) Extreme genome reduction in symbiotic bacteria. *Nat. Rev. Microbiol.*, **10**, 13–26.
39. Katoh,K. and Standley,D.M. (2016) A simple method to control over-alignment in the MAFFT multiple sequence alignment program. *Bioinformatics*, **32**, 1933–1942.
40. Jenner,L., Melnikov,S., Garreau de Loubresse,N., Ben-Shem,A., Iskakova,M., Urzhumtsev,A., Meskauskas,A., Dinman,J., Yusupova,G. and Yusupov,M. (2012) Crystal structure of the 80S yeast ribosome. *Curr. Opin. Struct. Biol.*, **22**, 759–767.
41. Popenda,M., Błażewicz,M., Szachniuk,M. and Adamiak,R.W. (2008) RNA FRABASE version 1.0: an engine with a database to search for the three-dimensional fragments within RNA structures. *Nucleic Acids Res.*, **36**, D386–D391.
42. Popenda,M., Szachniuk,M., Blazewicz,M., Wasik,S., Burke,E.K., Blazewicz,J. and Adamiak,R.W. (2010) RNA FRABASE 2.0: an advanced web-accessible database with the capacity to search the three-dimensional fragments within RNA structures. *BMC Bioinf.*, **11**, 231.
43. Flores,S.C., Wan,Y.Q., Russell,R. and Altman,R.B. (2010) Predicting RNA structure by multiple template homology modeling. *Pac. Symp. Biocomput.*, **2010**, 216–227.
44. Rother,M., Rother,K., Puton,T. and Bujnicki,J.M. (2011) ModeRNA: a tool for comparative modeling of RNA 3D structure. *Nucleic Acids Res.*, **39**, 4007–4022.
45. Antczak,M., Popenda,M., Zok,T., Sarzynska,J., Ratajczak,T., Tomczyk,K., Adamiak,R.W. and Szachniuk,M. (2016) New functionality of RNAComposer: an application to shape the axis of mir160 precursor structure. *Acta Biochim. Pol.*, **63**, 737–744.
46. Laing,C. and Schlick,T. (2009) Analysis of four-way junctions in RNA structures. *J. Mol. Biol.*, **390**, 547–559.
47. Touzet,H. and Perriquet,O. (2004) CARNAC: folding families of related RNAs. *Nucleic Acids Res.*, **32**, W142–W145.
48. Yao,Z.Z., Weinberg,Z. and Ruzzo,W.L. (2006) CMfinder - a covariance model based RNA motif finding algorithm. *Bioinformatics*, **22**, 445–452.
49. Lindgreen,S., Gardner,P.P. and Krogh,A. (2007) MASTR: multiple alignment and structure prediction of non-coding RNAs using simulated annealing. *Bioinformatics*, **23**, 3304–3311.
50. Kiryu,H., Tabei,Y., Kin,T. and Asai,K. (2007) Murlet: a practical multiple alignment tool for structural RNA sequences. *Bioinformatics*, **23**, 1588–1598.
51. Xu,Z.J. and Mathews,D.H. (2011) Multilign: an algorithm to predict secondary structures conserved in multiple RNA sequences. *Bioinformatics*, **27**, 626–632.
52. Wu,Y., Shi,B.B., Ding,X.Q., Liu,T., Hu,X.H., Yip,K.Y., Yang,Z.R., Mathews,D.H. and Lu,Z.J. (2015) Improved prediction of RNA secondary structure by integrating the free energy model with restraints derived from experimental probing data. *Nucleic Acids Res.*, **43**, 7247–7259.
53. Low,J.T. and Weeks,K.M. (2010) Shape-Directed RNA secondary structure prediction. *Methods*, **52**, 150–158.
54. Ramos,L.M.G., Degtyareva,N.N., Kovacs,N.A., Holguin,S.Y., Jiang,L.W., Petrov,A.S., Biesiada,M., Hu,M.Y., Purzycka,K.J., Arya,D.P. *et al.* (2017) Eukaryotic ribosomal expansion segments as antimicrobial targets. *Biochemistry*, **56**, 5288–5299.
55. Srikantha,T., Gutell,R.R., Morrow,B. and Soll,D.R. (1994) Partial nucleotide-sequence of a single Ribosomal-RNA coding region and secondary structure of the large subunit 25S-Ribosomal RNA of *candida-albicans*. *Curr. Genet.*, **26**, 321–328.
56. Abeyrathne,P.D., Koh,C.S., Grant,T., Grigorieff,N. and Korostelev,A.A. (2016) Ensemble cryo-em uncovers inchworm-like translocation of a viral ires through the ribosome. *Elife*, **5**, e14874.
57. Garreau de Loubresse,N., Prokhorova,I., Holtkamp,W., Rodnina,M.V., Yusupova,G. and Yusupov,M. (2014) Structural basis for the inhibition of the eukaryotic ribosome. *Nature*, **513**, 517–522.
58. Matthews,B.W. (1975) Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochim. Biophys. Acta*, **405**, 442–451.
59. Li,Z., Guo,Q., Zheng,L., Ji,Y., Xie,Y.-T., Lai,D.-H., Lun,Z.-R., Suo,X. and Gao,N. (2017) Cryo-EM structures of the 80S ribosomes from human parasites trichomonas vaginalis and toxoplasma gondii. *Cell Res.*, **27**, 1275–1288.
60. Gillespie,J.J., Munro,J.B., Heraty,J.M., Yoder,M.J., Owen,A.K. and Carmichael,A.E. (2005) A secondary structural model of the 28S rRNA expansion segments D2 and D3 for chalcidoid wasps (Hymenoptera: chalcidoidea). *Mol. Biol. Evol.*, **22**, 1593–1608.
61. Kyrpides,N., Overbeek,R. and Ouzounis,C. (1999) Universal protein families and the functional content of the last universal common ancestor. *J. Mol. Evol.*, **49**, 413–423.
62. Grishin,N.V. (2001) Fold change in evolution of protein structures. *J. Struct. Biol.*, **134**, 167–185.
63. Hsiao,C., Mohan,S., Hershkovitz,E., Tannenbaum,A. and Williams,L.D. (2006) Single nucleotide RNA choreography. *Nucleic Acids Res.*, **34**, 1481–1491.
64. Vargason,J.M., Henderson,K. and Ho,P.S. (2001) A crystallographic map of the transition from B-DNA to A-DNA. *Proc. Natl. Acad. Sci. U.S.A.*, **98**, 7265–7270.
65. Mohan,S., Hsiao,C., VanDeusen,H., Gallagher,R., Krohn,E., Kalahar,B., Wartell,R.M. and Williams,L.D. (2009) Mechanism of RNA double helix-propagation at atomic resolution. *J. Phys. Chem. B*, **113**, 2614–2623.
66. Sundaralingam,M. and Sekharudu,Y.C. (1989) Water-Inserted alpha-helical segments implicate reverse turns as folding intermediates. *Science*, **244**, 1333–1337.
67. Burgi,H.B. (1973) Chemical reaction coordinates from crystal-structure data 1. *Inorg. Chem.*, **12**, 2321–2325.
68. Petrov,A.S., Bernier,C.R., Gulen,B., Waterbury,C.C., Hershkovits,E., Hsiao,C., Harvey,S.C., Hud,N.V., Fox,G.E., Wartell,R.M. *et al.* (2014) Secondary structures of rRNAs from all three domains of life. *PLoS One*, **9**, e88222.

69. Purzycka,K.J., Popenda,M., Szachniuk,M., Antczak,M., Lukasiak,P., Blazewicz,J. and Adamiak,R.W. (2015) Automated 3D RNA structure prediction using the RNAComposer method for riboswitches. *Methods Enzymol.*, **553**, 3–34.
70. Gutell,R.R. and Fox,G.E. (1988) A compilation of large subunit RNA sequences presented in a structural format. *Nucleic Acids Res.*, **16**(Suppl), r175–r269.
71. Zhao,Y.-E., Wang,Z.-H., Xu,Y., Wu,L.-P. and Hu,L. (2013) Secondary structure prediction for complete rdna sequences (18S, 5.8S, and 28S rdna) of demodex folliculorum, and comparison of divergent domains structures across acari. *Exp. Parasitol.*, **135**, 370–381.
72. Schnare,M.N., Damberger,S.H., Gray,M.W. and Gutell,R.R. (1996) Comprehensive comparison of structural characteristics in eukaryotic cytoplasmic large subunit (23S-Like) ribosomal RNA. *J. Mol. Biol.*, **256**, 701–719.
73. Muñoz-Gómez,S.A., Bilolikar,G., Wideman,J.G. and Geiler-Samerotte,K. (2021) Constructive neutral evolution 20 years later. *J. Mol. Evol.*, **89**, 172–182.
74. Stoltzfus,A. (1999) On the possibility of constructive neutral evolution. *J. Mol. Evol.*, **49**, 169–181.
75. Lynch,M. and Walsh,B. (2007) In: *The Origins of Genome Architecture*. Sinauer Associates Sunderland, MA.
76. Hariharan,N., Ghosh,S. and Palakodeti,D. (2022) The story of rRNA expansion segments: finding functionality amidst diversity. *Wiley Interdiscip. Rev. RNA*, e1732.
77. Shankar,V., Rauscher,R., Reuther,J., Gharib,W.H., Koch,M. and Polacek,N. (2020) rRNA expansion segment 27Lb modulates the factor recruitment capacity of the yeast ribosome and shapes the proteome. *Nucleic Acids Res.*, **48**, 3244–3256.
78. Fujii,K., Susanto,T.T., Saurabh,S. and Barna,M. (2018) Decoding the function of expansion segments in ribosomes. *Mol. Cell*, **72**, 1013–1020.
79. Ramesh,M. and Woolford,J.L. Jr. (2016) Eukaryote-specific rRNA expansion segments function in ribosome biogenesis. *RNA*, **22**, 1153–1162.
80. Diaconu,M., Kothe,U., Schlünzen,F., Fischer,N., Harms,J.M., Tonevitsky,A.G., Stark,H., Rodnina,M.V. and Wahl,M.C. (2005) Structural basis for the function of the ribosomal L7/12 stalk in factor binding and GTPase activation. *Cell*, **121**, 991–1004.
81. Nilsson,J., Sengupta,J., Gursky,R., Nissen,P. and Frank,J. (2007) Comparison of fungal 80S ribosomes by cryo-em reveals diversity in structure and conformation of rRNA expansion segments. *J. Mol. Biol.*, **369**, 429–438.
82. Greber,B.J., Boehringer,D., Montellese,C. and Ban,N. (2012) Cryo-EM structures of arx1 and maturation factors rei1 and jjj1 bound to the 60S ribosomal subunit. *Nat. Struct. Mol. Biol.*, **19**, 1228–1233.
83. Wu,S., Tutuncuoglu,B., Yan,K., Brown,H., Zhang,Y., Tan,D., Gamalinda,M., Yuan,Y., Li,Z., Jakovljevic,J. *et al.* (2016) Diverse roles of assembly factors revealed by structures of late nuclear Pre-60S ribosomes. *Nature*, **534**, 133–137.
84. Kater,L., Thoms,M., Barrio-Garcia,C., Cheng,J., Ismail,S., Ahmed,Y.L., Bange,G., Kressler,D., Berninghausen,O., Sinning,I. *et al.* (2017) Visualizing the assembly pathway of nucleolar Pre-60S ribosomes. *Cell*, **171**, 1599–1610.
85. Bradatsch,B., Leidig,C., Granneman,S., Gnädig,M., Tollervey,D., Böttcher,B., Beckmann,R. and Hurt,E. (2012) Structure of the Pre-60S ribosomal subunit with nuclear export factor arx1 bound at the exit tunnel. *Nat. Struct. Mol. Biol.*, **19**, 1234–1241.
86. Wang,X., Yue,Z., Xu,F., Wang,S., Hu,X., Dai,J. and Zhao,G. (2021) Coevolution of ribosomal RNA expansion segment 7L and assembly factor Noc2p specializes the ribosome biogenesis pathway between *Saccharomyces cerevisiae* and *Candida albicans*. *Nucleic Acids Res.*, **49**, 4655–4667.
87. Smethurst,D.G.J., Kovalev,N., McKenzie,E.R., Pestov,D.G. and Shcherbik,N. (2020) Iron-Mediated degradation of ribosomes under oxidative stress is attenuated by manganese. *J. Biol. Chem.*, **295**, 17200–17214.
88. Knorr,A.G., Schmidt,C., Tesina,P., Berninghausen,O., Becker,T., Beatrix,B. and Beckmann,R. (2019) Ribosome–NatA architecture reveals that rRNA expansion segments coordinate N-terminal acetylation. *Nat. Struct. Mol. Biol.*, **26**, 35–39.
89. Kossinova,O., Malygin,A., Krol,A. and Karpova,G. (2014) The SBP2protein central to selenoprotein synthesis contacts the human ribosome at expansion segment 7L of the 28S rRNA. *RNA*, **20**, 1046–1056.
90. Takeuchi,A., Schmitt,D., Chapple,C., Babaylova,E., Karpova,G., Guigo,R., Krol,A. and Allmang,C. (2009) A short motif in drosophila secis binding protein 2 provides differential binding affinity to secis RNA hairpins. *Nucleic Acids Res.*, **37**, 2126–2141.
91. Parker,M.S., Balasubramaniam,A. and Parker,S.L. (2018) The expansion segments of human 28S rRNA match micrornas much above 18S rRNA or core segments. *MicroRNA*, **7**, 128–137.
92. Gutell,R.R. (1992) Evolutionary characteristics of 16S and 23S rRNA structures. In: Hartman,H. and Matsuno,K. (eds). *The Origin and Evolution of the Cell: Proceedings of the Conference on the Origin and Evolution of Prokaryotic and Eukaryotic Cells, Shimoda, Japan, 22-25 April 1992*. World Scientific, Singapore, pp. 243–309.
93. Petrov,A.S., Gulen,B., Norris,A.M., Kovacs,N.A., Bernier,C.R., Lanier,K.A., Fox,G.E., Harvey,S.C., Wartell,R.M., Hud,N.V. *et al.* (2015) History of the ribosome and the origin of translation. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, 15396–15401.
94. Nawrocki,E.P. and Eddy,S.R. (2013) Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics*, **29**, 2933–2935.
95. Rivas,E. (2020) RNA structure prediction using positive and negative evolutionary information. *PLoS Comp. Biol.*, **16**, e1008387.
96. Rivas,E., Clements,J. and Eddy,S.R. (2017) A statistical test for conserved RNA structure shows lack of evidence for structure in lncRNAs. *Nat. Methods*, **14**, 45–48.
97. Kovacs,N.A., Petrov,A.S., Lanier,K.A. and Williams,L.D. (2017) Frozen in time: the history of proteins. *Mol. Biol. Evol.*, **34**, 1252–1260.
98. Cottilli,P., Itoh,Y., Nobe,Y., Petrov,A.S., Lisón,P., Taoka,M. and Amunts,A. (2022) Cryo-EM Structure and rRNA Modification Sites of a Plant Ribosome. *Plant Communications*, **3**, 100342.