Differences in the path to exit the ribosome across the three domains of life

Khanh Dao Duc[1], Sanjit S. Batra[1], Nicholas Bhattacharya[2], Jamie H. D. Cate[3,4,5], and Yun S. Song[1,6,7,*]

[1] Computer Science Division, University of California, Berkeley, CA 94720
[2] Department of Mathematics, University of California, Berkeley, CA 94720
[3] Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720
[4] Department of Chemistry, University of California, Berkeley, CA 94720
[5] Molecular Biophysics and Integrated Bioimaging Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720
[6] Department of Statistics, University of California, Berkeley, CA 94720
[7] Chan Zuckerberg Biohub, San Francisco, CA 94158

* To whom correspondence should be addressed: yss@berkeley.edu

June 28, 2018

ABSTRACT

Recent advances in biological imaging have led to a surge of fine-resolution structures of the ribosome from diverse organisms. Comparing these structures, especially the exit tunnel, to characterize the key similarities and differences across species is essential for various important applications, such as designing antibiotic drugs and understanding the intricate details of translation dynamics. Here, we compile and compare 20 fine-resolution cryo-EM and X-ray crystallography structures of the ribosome recently obtained from all three domains of life (bacteria, archaea and eukarya). We first show that a hierarchical clustering of tunnel shapes closely reflects the species phylogeny. Then, by analyzing the ribosomal RNAs and proteins localized near the tunnel, we explain the observed geometric variations and show direct association between the conservations of the geometry, structure, and sequence. We find that the tunnel is more conserved in its upper part, from the polypeptide transferase center to the constriction site. In the lower part, tunnels are significantly narrower in eukaryotes than in bacteria, and we provide evidence for the existence of a second constriction site in eukaryotic tunnels. We also show that ribosomal RNA and protein sequences are more likely to be conserved closer to the tunnel, as is the presence of positively charged amino acids. Overall, our comparative analysis shows how the geometric and biophysical properties of the exit tunnel play an important role in ensuring proper transit of the nascent polypeptide chain, and may explain the differences observed in several co-translational processes across species.

INTRODUCTION

Ribosomes are the key actors of mRNA translation, a fundamental biological process underlying all forms of life. While decoding the mRNA nucleotides into their associated polypeptide sequence, ribosomes regulate the dynamics of translation and other central co-translational processes such as the translocation to cell membranes or protein folding [1, 2, 3]. These processes rely on the structural properties of the ribosome, through interactions with different elements such as binding factors, tRNAs, or the nascent polypeptide chain. For example, it has been shown that some specific sequence motifs associated with antibiotic treatment could stall the ribosome and subsequently arrest translation [4, 5, 6, 7]. This phenomenon is caused by interactions between the ribosome and the nascent polypeptide chain itself: prior to leaving the ribosome, nascent polypeptides first pass through a structure called the ribosome exit tunnel, spanning from the peptidyl-transferase center (PTC) to the surface of the ribosome. As the tunnel can accommodate about 40 amino acids or so [8], its geometry and biophysical properties potentially impact the translation dynamics [9].

Being a functionally important structure, the ribosome exit tunnel needs to be well conserved across species to ensure proper translation of mRNA sequences. On the other hand, the selectivity of arrest sequences to specific species [6, 8], or differences of translational and co-translational mechanisms between eukarya and bacteria [10, 11, 12, 13], for example, suggest that important variations of the exit tunnel structure exist. As such variations have potentially important consequences on the regulation of translation or antibiotic resistance, it is thus crucial to identify and catalog these differences, and more generally understand the evolution of the ribosome exit tunnel.

For the past several years, the ribosome exit tunnel has been the object of intense study to unravel characteristic features, such as its geometric or electrodynamic properties, solvent behavior, rigidity, and dynamic domains [14, 15, 16, 17, 18, 19], to name just a few. As an essential molecular machine present in all living systems, the ribosome has also been extensively used in the past to elucidate phylogenetic relationships via sequence analysis [20]. More recently, several studies have shed light on the relation between the evolution of the ribosome and its function [21, 22]. Specifically, by taking advantage of the sequence information and availability of 3D structures from X-ray crystallography and cryo-electron microscopy (cryo-EM), it has been shown how the evolution of ribosomal RNA (rRNA) has been locally constrained at the beginning of the tunnel around the PTC, where amino acids coded by the mRNA sequence get added to the polypeptide chain. Over the past few years an increasing number of new ribosome structures has been obtained at an atomic resolution of a few Angströms (see below). Hence, it is now possible to extend our understanding of the relation between the biophysical structure of the entire exit tunnel and its evolution across many different species, thereby unraveling local specificities of the tunnel function.

In the present work, we provide a quantitative analysis and comparison of the ribosome exit tunnel structure across a diverse set of species. We compile and compare 20 recently obtained fine-resolution ribosome structures, coming from all three domains of life (bacteria, archae, and eukarya). Upon extracting the coordinates of the tunnels from these structures, we investigate the relation between the geometry of the tunnel and the evolution of the ribosomal structure and its constituent sequences. To achieve this, we introduce and apply a suite of computational methods to study the geometric properties of the tunnel, the local structure of the ribosome near the tunnel, and the conservation of rRNA and ribosomal protein sequences. Our comparative analysis highlights the essential role of the geometric and biophysical properties of the exit tunnel in ensuring proper transit of the nascent polypeptide chain, while revealing potential specificities of the tunnel function across different species.

RESULTS

Extraction of the ribosome exit tunnel structure

We compiled and analyzed publicly available ribosome structures obtained from cryo-EM and X-ray crystallography (see Methods and Table 1). These recent structures (9 published in 2017) accounted for 20 different organisms, including 2 archaea, 2 organelles, 8 bacteria and 8 eukaryotes. Most of these structures (16 out of 20) came from cryo-EM maps, with an average resolution of 3.74 Å; the X-Ray structures had an average resolution of 2.92 Å. Replicates of *E. coli*, *T. thermophilus* and *H. sapiens* ribosome structures were also included in our analysis to examine the robustness of our results (see Supporting Information). For each structure, we applied a procedure described in Methods and illustrated in Figure 1, to extract the ribosome exit tunnel geometry, encoded as a set of coordinates describing the trajectory of the centerline and the tunnel radius at each point of the centerline. Upon obtaining these data, we checked the quality of each tunnel reconstruction by comparing the density-fitting score (see Methods) between residues located close to the tunnel and the rest of the structure. All the structures in our dataset showed an increase of the density-fitting score for tunnel residues, suggesting the good quality of maps in the tunnel region.

Analysis of global features indicates larger bacterial tunnels and more variation in the lower part

To compare the geometry of the ribosome exit tunnel across different species, we first examined global geometric features, such as the length, tunnel-wide average radius, and volume (see Supplementary Data for specific values). The range of values for the tunnel length (88.8 ± 6.0 Å) and that for average radius (5.4 ± 0.4 Å) were both consistent with previous observations [8, 23]. Upon ordering the species by their exit tunnel volume (see Figure 2), we found a perfect separation between bacteria and eukaryotes, with archaea in between. Specifically, bacterial tunnels (including the ones from organelles) are larger than eukaryotic ones, with mean $(3.85 \pm 0.37) \times 10^4$ Å^3 for bacteria, compared to $(2.78 \pm 0.13) \times 10^4$ Å^3 for eukaryotes. We similarly analyzed the length and average radius (Figure 2) to see if the volume variation could be mainly explained by one of these two variables. We obtained a less clear separation among the three domains of life, but still observed a similar trend for the length (91.6 ± 3.6 Å for bacteria and 83.3 ± 2.9 Å for eukarya) and average radius (5.7 ± 0.3 Å for bacteria and 5.1 ± 0.1 Å for eukarya), suggesting that both contribute to the observed difference in volume.

   To study how these geometric features vary along the tunnel, we carried out a more refined analysis by partitioning the tunnel into two sub-parts separated by the "constriction site" (see Figure 1), a conserved central region constricted by the uL4 and uL22 protein loops. Upon locating the constriction site (see Methods) for each tunnel, which were at position 34.1 ± 4.8 Å from the start of the tunnel (see Figure 2), we studied how the volumes in the upper (from start to constriction site) and lower (from constriction site to exit) parts respectively correlated with the volume of the entire tunnel. We found a strong correlation for the lower part (Pearson correlation $r^2 = 0.9649$, p-value $p < 10^{-4}$), while no significant correlation was observed for the upper part ($r^2 = 0.085$), suggesting that the upper part of the tunnel is geometrically quite conserved and most of the variation across species comes from the lower part. The total length also shows a better correlation with the length of the lower part of the tunnel ($r^2 = 0.72$, $p < 10^{-3}$, compared with $r^2 = 0.32$, $p < 10^{-3}$ for the upper part). While the upper part accounts for $38.4 \pm 5.0\%$ of the total tunnel length, it captures only $30.0 \pm 6.9\%$ of the total tunnel volume, suggesting an increase of the average radius in the lower part. Computing the average radius in the two regions indeed showed an increase from 4.7 ± 0.3 Å

in the upper part to 5.8 ± 0.6 Å in the lower part.

Hierarchical clustering of tunnel radius variation plots reflects the species phylogeny

As we found domain-specific geometric variations of the tunnel that are more amplified in the lower part of the tunnel, we aimed to quantify these variations more precisely. We first examined the 3D coordinates of the tunnel centerline and found that $96.7 \pm 1.0\%$ of their variations could be explained by fitting to a straight line (see Supplementary Data). We therefore simply parametrized the centerline of the tunnel by its arc length and studied the associated radii (see Methods), leading to a radius variation plot for each species (Figure 3A and Figure S1). To compare these plots, we introduced a distance function (see Methods) and evaluated it for each pair of species. The resulting pairwise distance matrix was then used to cluster the species, yielding the hierarchical tree shown in Figure 3B.

As in the aforementioned global feature analysis, we obtained a clear separation of domains, with bacteria clustering separately from archaea and eukaryotes. Among eukaryotes, *T. vaginalis* is rather special in that it clustered with archaea (*P. furiosus* and *H. marismortui*), which then grouped together with a cluster of trypanosomes (*L. donovani* and *T. cruzi*); separated from these, the remaining eukaryotes formed a larger cluster. This separation of *T. vaginalis* is supported by the properties of its rRNAs, which are comparable in size to prokaryotic counterparts, and with nearly all the eukaryote-specific rRNA expansion segments missing [24]. More generally, the separations of *T. vaginalis* and trypanosomes from the other eukaryotes is also consistent with the evolutionary relationships obtained from 16S-like rRNA sequences (from the ribosome small subunit) [24]. We confirmed this result by carrying out a phylogenetic analysis for the species in our dataset, based on their 16S/18S rRNA sequences (see Figure S2A). At a finer resolution, our hierarchical clustering from tunnel geometry comparison (see Figure S2B) starts to deviate from the phylogenetic tree from sequence analysis, suggesting that smaller differences of the tunnel geometry are more difficult to relate evolutionarily. By including ribosomes from mitochondria and chloroplast (Figure S2C), we also found that they clustered with prokaryotes, which is consistent with their bacterial origin [25]. To assess the robustness of these results, we used replicate structures from same species and studied the sensitivity of the clustering to the parameters of the geometric distance (see Supporting Information, Figure S3 and Table S1). Overall, this did not lead to significant changes, demonstrating the robustness of our results to replicate structures and the choice of the comparison metric.

The contribution of the different tunnel sub-regions to the intra- and inter- domain distance

To understand why we obtained a clear separation between the prokaryotic and eukaryotic tunnels, we more closely studied the variation of tunnel geometric distances across the two domains. Distinguishing intra- and inter-domain pairs (respectively accounting for pairs of species from the same domain, and pairs with one from each), we found that while intra-domain distances were on average similar for bacteria and eukaryotes (Figure S4), the inter-domain distances were noticeably larger (with an increase of 40% for the average). As the distance that we introduced integrates the geometric variation along the tunnel, we sought to examine which part of the tunnel contributes most significantly to the intra- and inter-domain distances. To this end, we divided up the tunnel for each pair of species into 4 quarters (the first and fourth respectively corresponding to the start and the exit parts), and applied the same metric as before to compute the geometric distance in

each of these subregions (Figure 3C). While the inter-domain distance was on average always larger than the intra-domain one in every subregion, we found that across these regions, the inter-domain distance was substantially larger for the last two quarters (each carrying on average 30% of the total variation) compared to the first two ($\sim 20\%$ each). A similar trend was observed among eukaryotes (with 33% and 26% of the total variation respectively carried by the third and last quarters). In stark contrast, in bacteria the third quarter was the one carrying the least variation (21%, compared to 24%, 26% and 29% for the first, second and fourth subregions). Overall, our local comparative analysis of the tunnel geometry confirmed that the tunnel is more conserved in the upper part of the tunnel.

Existence of a second constriction site in eukaryotes and the role of ribosomal protein uL4

To understand the local geometric variation of the exit tunnel observed across different species, we sought to determine how the ribosomal structure can explain the aforementioned clustering pattern of bacterial and eukaryotic tunnels. In the region associated with most inter-domain variation (see Figure 3C), we first looked at the constriction site, where proteins uL4 and uL22 meet. Interestingly, we found that the structure of uL4 at the tunnel generally differs in eukarya and bacteria, due to an extension of the uL4 loop in eukarya that yields a second constriction site. Such an extension is also present but less prominent in archaea (see Figure S5). We illustrate this structural difference in Figure 4A for *E. coli* and *H. sapiens*. The presence of a second constriction site in *H. sapiens* exit tunnel is clearly illustrated in Figure 4B, which shows that the second trough of the radius plot is even lower than the first trough for *H. sapiens*, while the opposite is true for *E. coli*.

To confirm that such a difference persists between other bacteria and eukarya, we analyzed in Figure 4C,D the relative positions and the tunnel radii associated with the first and second troughs of the radius plot for each species in our entire dataset. We found the distance between the troughs to be on average larger in bacteria ($23.6 \pm 2.7\,\text{Å}$) than in eukarya ($19.1 \pm 2.3\,\text{Å}$), suggesting structural changes in the constriction site region. While at the first trough (which corresponds to the universally shared constriction site) bacteria and eukarya have similar radii ($4.2 \pm 0.6\,\text{Å}$ for bacteria and $4.2 \pm 0.4\,\text{Å}$ for eukarya), the radius at the second trough is significantly larger in bacteria ($5.0 \pm 0.5\,\text{Å}$) than in eukarya ($3.9 \pm 0.6\,\text{Å}$). Furthermore, direct comparison between the first and second troughs for each species shows that the second trough radius is larger than the first one for all archaea and bacteria. In contrast, we observed the opposite in eukarya, except for *L. donovani* and *T. cruzi* (also explaining the clustering of these two with archaea in Figure 3B). Therefore, the second constriction site in eukarya is in general narrower than the first one.

To explain the discrepancy observed for *L. donovani* and *T. cruzi*, we looked at the ribosomal structure in the constriction site region. Aside from ribosomal proteins uL4 and uL22, the tunnel is surrounded there by ribosomal RNA (rRNA). In trypanosomes like *L. donovani* and *T. cruzi*, the large subunit (LSU) rRNA breaks from the standard 28S rRNA into six smaller chains [26, 27]. The two main chains precisely meet in the constriction site region (see Figure S6), suggesting that less constraint is locally applied to the ribosomal proteins and tunnel structure, potentially increasing the size of the second constriction site. To summarize, we concluded that an important part of the geometrical differences and clustering observed between the ribosome exit tunnels come from the structure at the constriction site. More precisely, there exists a second constriction site specific to eukarya, which narrows the tunnel after the first, universally shared constriction site.

Replacement of uL23 by eL39 ribosomal protein in eukarya affects the tunnel geometry

In addition to the structure of the constriction site region, we further examined the structure of the lower part of the tunnel, where we also detected some important geometric variations. In bacteria, the tunnel in this region is mainly surrounded by rRNA and the protein uL23. In eukarya and archaea, uL23 is also present, but the segment covering the tunnel region is replaced by the protein eL39 [19, 23, 28]. Upon close examination and comparison of these structures, we found that eL39 does not only cover the region originally occupied by uL23 in bacteria, but further extends to the tunnel exit, as illustrated in Figure 5A. More precisely (Figure 5B), the coverage distance of uL23 in bacterial ribosomes is 19.0 ± 2.8 Å, compared to 31.6 ± 2.3 Å for eL39 in eukarya. Such a difference affects the tunnel geometry (Figure 5B), as the tunnel radius in the corresponding regions is significantly larger for bacteria (6.0 ± 0.4 Å) than for eukarya (4.6 ± 0.4 Å). As a result, the exit region of the tunnel is wider in bacteria than in eukarya, with an average radius difference of approximately 1 Å in the last 30 Å of the tunnel (see Figure S7), hence contributing to the clustering of the radius plots that we previously obtained.

Association between rRNA sequence conservation and the exit tunnel

After finding evidence that the geometric variation of the tunnel across different species can be explained by rRNA and protein structural variations that have emerged through evolution, we sought to study how the tunnel and its geometry directly relate to the evolution of the ribosome at the sequence level. First, we investigated the conservation of rRNAs, focusing on the main chain (23S for bacteria, and 28S for archaea and eukarya) that constitutes the ribosome LSU. Comparative analysis of rRNA sequences has been extensively used in the past to elucidate phylogenetic relationships [29, 30], and has recently [22] led to the identification of stretches of evolutionarily conserved sequences, the so-called "conserved nucleotide elements" (CNE). In particular, it has been suggested that a large part of the CNEs may have a function in nascent polypeptide transit through the tunnel [22]. To verify this, we compared the presence of CNEs with their distance to the tunnel (see Figure 6A and Figure S8). For each species and for a given distance d, we computed the frequency of CNE among all rRNA nucleotides located within distance d from the tunnel. Plotting this frequency as a function of d for all the species, we observed in Figure 6B a global decrease in the frequency of CNEs as d increases, which means that nucleotides located farther from the tunnel are less conserved. For much divergent rRNAs such as the ones from organelles or *T. vaginalis* [24], the low number of CNEs yields no association with the distance to the tunnel. For bacteria, archaea, and eukaryotic trypanosomes, we found that the frequency of CNEs within 25 Å from the tunnel is $21 \pm 5\%$, which is more than the double the frequency of CNEs located between 25 and 50 Å ($10 \pm 2\%$) and also that located farther than 50 Å from the tunnel ($8 \pm 2\%$). In the remaining eukaryotes, we found a much larger frequency of CNEs within 25 Å from the tunnel ($45 \pm 12\%$), with a sharp decrease for the region between 25 and 50 Å ($18 \pm 5\%$) and the region beyond 50 Å ($10 \pm 3\%$). Distinguishing universal and domain-specific CNEs [22], we found a similar trend for both, with a larger contribution from domain-specific CNEs (Figure S9). Overall, these results confirm the association between the tunnel and the conservation of surrounding rRNA nucleotides.

To determine whether such conservation is homogeneous or specific to some local parts of the tunnel, we computed for each species the local frequency of CNEs and compared the level of conservation across the tunnel, as shown in Figure 6C and Figure S10. Overall, we found that the sequence conservation along the tunnel is heterogeneous and strongly enhanced towards the upper part of the tunnel, as more than 80% of the CNEs were associated with the first 40 Å of the tunnel

(while all the nucleotides positioned along the first 40 Å of the tunnel represent only 48% of all nucleotides). In this region, we also observed a stronger conservation closer to the PTC, as 43% of the CNEs were located less than 10 Å from the tunnel start position. These results are in agreement with the comparative geometric study of the tunnels, which also showed more conservation in the upper part region.

Conservation of ribosomal protein sequence and positive charges at the exit tunnel

Lastly, we studied the relation between the geometric and sequence conservations at the tunnel for ribosomal proteins. To do so, we aligned across species the main proteins located close to the tunnel, namely uL4, uL22, and uL23 for bacteria and eL39 for eukarya (see Methods, Figure 7A and Figure S11). For uL4, because of multiple insertions that prevent a good alignment of all sequences (like the one leading to the aforementioned second constriction site), we separately aligned uL4 for bacterial and eukaryotic sequences. Upon computing a conservation score [31] for each alignment, we found peaks of conservation in regions located in proximity with the ribosomal tunnel, suggesting that association between the sequence conservation and the exit tunnel also occurs for proteins (Figure 7B and Figure S12A). Upon closer examination of the consensus motif sequences at the tunnel, we found a large fraction ($\sim 30\%$) of positively charged amino acids (arginine R and lysine K). Interestingly, these positively charged amino acids sometimes co-occur at the same position of an alignment, e.g., position 159 of the uL22 alignment (with 8 arginines and 8 lysines, see Figure 7A), position 69 of bacterial uL4 alignment, and position 34 of eL39 alignment (Figure S11). This suggests that not only the sequence but also the charge properties are important to maintain the integrity of the tunnel.

Therefore, we introduced a measure of charge conservation (see Methods) that specifically accounts for the local presence of positively charged amino acids in an alignment. Recent studies have shown enrichment of positively charged amino acids in ribosomal proteins [32]. Upon computing the charge score for our sequence alignments, we indeed found some significant correlation between the sequence conservation and charge conservation scores for bacterial proteins with Pearson's correlation $r^2 = 0.70$, 0.51 and 0.47 (p-values $< 10^{-4}$) for bacterial alignments of uL4, uL23 and uL22, respectively (see Supplementary Data). We also found that overall, regions with the highest presence of positive charges are located in the tunnel region (see Figure 7B and Figure S12A). In uL22, the residues in direct contact with the tunnel (Figure 7B and C) are highly conserved, with the highest charge conservation score. Larger charge and sequence conservations were also found in uL4, with more conservation at the first constriction site compared to the second one in eukarya (Figure S12A and B). In the lower part of the tunnel, we detected a strong peak of both charge and sequence conservation in bacterial uL23. Interestingly, while we did not observe in eukarya a conservation signal as strong in the whole region of the tunnel covered by eL39, the region where eL39 overlaps with uL23 in bacteria also exhibits large charge and sequence conservation (Figure S12A and B), suggesting persistence of the local charge properties in this region although the ribosome structure differs between bacteria and eukarya. Overall, we concluded that ribosomal proteins are more locally conserved at the proximity of the tunnel, with a bias towards positively charged amino acids, suggesting that these amino acids and the electrostatic environment are important for the tunnel function (see Discussion).

DISCUSSION

In this study, we carried out a comparative analysis of the ribosome exit tunnel geometry that is most comprehensive to date. Recent advances in cryo-EM [33, 34] allowed us to observe the ribosome at a resolution fine enough to characterize the local structure of the exit tunnel, thereby making it feasible to detect significant differences across species. Despite the eukaryotic 80S ribosome being approximately 50% greater in mass and rRNA lengths compared to the prokaryotic 70S ribosome [21, 35], we found the bacterial exit tunnel to be larger, in both length and average radius (Figure 2). Such differences mostly come from the lower part of the tunnel.

We classified the tunnels using their radius variation plots and interestingly found that the resulting hierarchical clustering closely reflects the species phylogeny (Figure 3). We then found evidence of two notable evolutionary changes that affect the tunnel structure and explain our clustering result. The first is the insertion of an extended loop sequence in eukaryotic ribosomal protein uL4 (Figure 4), which leads to a second constriction site that is narrower than the first one (except in trypanosomes). The second is the replacement of the uL23 loop in the lower part of the tunnel in bacteria by the small protein eL39 in archaea and eukarya (Figure 5). While the existence of the first constriction site and the overlapping positions of eukaryotic eL39 and prokaryotic uL23 proteins have been previously well established [23, 28, 36], our comparative analysis shows that the above evolutionary changes have led to important differences among species that significantly affect the tunnel geometry and hence the path of the nascent chain through the tunnel.

Implications on the evolution of the ribosome

Since both ribosomal RNA and protein sequences located near the exit tunnel are more conserved (Figure 6, Figure 7, and Figure S12), there is a strong relationship between the evolution of ribosome components and the variation in the geometry of the exit tunnel [22]. The high geometric conservation of the upper part of the tunnel coincides with the high sequence conservation of the associated rRNAs and ribosomal proteins, suggesting that it is an important part of the ribosome core. The observed contrast in conservation between the upper and lower parts of the tunnel is in agreement with previously proposed models of the evolution of the rRNA structure [21], which distinguishes an early phase of creation of a short tunnel, subsequently completed by a phase of tunnel extension and expansion of the LSU. These two phases are preceded by the maturation of the PTC, commonly considered as the oldest ribozyme [21, 37, 38] and which also appeared in our analysis as the most universally conserved region [22, 39]. While there is a clear association between the geometry of the exit tunnel and the evolution of its components, some other biophysical properties of the tunnel may also be influencing the evolution of the ribosome. Upon closer examination of the conserved ribosomal protein sequences at the tunnel, we indeed found a significant contribution of positively charged amino acids (see Figure 7 and Figure S12), which are generally found to be highly enriched in ribosomal proteins [32, 40]. While it has been proposed that this general enrichment may be needed to create electrostatic interaction during ribosome assembly [40], our analysis suggests that the conservation of positively charged amino acids near the tunnel could be tied to maintaining an appropriate electrostatic environment inside the tunnel (see below).

Impact on the nascent polypeptide chain transit

The geometric and structural variation of the exit tunnel may have several functional and biophysical implications, as the nascent chain transit can be affected. On the one hand, a too narrow tunnel could obstruct the elongation of the nascent chain and generate tunnel-peptide interactions that

might alter the elongation rate during protein synthesis [4, 41, 42]. On the other hand, a too large tunnel radius could lead to an increase of conformational sampling [43, 44, 45] and hence misfolding events. Our results indicate that before reaching the constriction site in the upper tunnel, the nascent chain propagates in an environment that seems relatively well conserved, with an increase of the tunnel radius followed by a decrease near the constriction site. In the context of a diffusing particle inside the tunnel, we previously found [9] that this radial increase creates a strong entropic barrier, which can be compensated by the electrostatic potential if the particle is positively charged. More generally, it was also experimentally shown that electrostatics in the ribosomal tunnel modulate chain elongation rates [9, 17] and can even induce ribosome stalling [46]. At the proteome-wide level, averaging the charged amino acid frequency over all translated genes in *S. cerevisiae* revealed an elevated (respectively, decreased) amount of positively (respectively, negatively) charged amino acids in the first ~ 20 codons [9]. These patterns also hold in other species (Figure S13). Together with the conservation of ribosomal protein charge at the tunnel (which is also more significant at the first than at the second constriction site, see Figure 7 and Figure S12), they suggest co-evolution of the tunnel and the charge distributions of the proteome, to maintain a favorable electrostatic environment for the nascent polypeptide during its initial pass through the exit tunnel.

Once the N-terminus exits from the tunnel, the elongation of the nascent chain may be affected by additional factors, such as the co-translational processes that generate pulling forces on the nascent polypeptide chain [47]. For example, the translocon, involved in the translocation of membrane proteins, can relieve elongation arrest due to the SecM sequence, thereby enabling translation restart [48]. In eukaryotes, a similar mechanism has been suggested for the chaperone Hsp70 [49]. More generally, proteins that start to fold co-translationally while still in contact with the ribosome, exert some pulling force sufficient to weaken or even abolish stalling [50, 51]. As shown by theoretical studies of polymer dynamics in response to external pulling [52, 53], the efficiency of the pulling depends on how tension propagates from the ends into the bulk, and, in our case, how the free energy released by the folding reaction can get stored as an increased tension in the nascent chain [54]. This phenomenon is evidenced by the increase of the pulling force with the nascent chain hydrophobicity (found with the formation of α-helix structure and chain stiffness [47]) and the positive impact of the hydrophobicity on the elongation rate (obtained by analyzing ribosome profiling data [9]). As the confined geometry of the tunnel can actually play a role in stabilizing α-helices [55], we therefore hypothesize that the presence of a second narrower constriction site in eukaryotes can improve the tension propagation across the nascent chain, and thus facilitate the elongation process.

Functional implications of the tunnel variation

We detected more variation in the lower part of the tunnel, as structural modifications of the ribosomal proteins lead to a decrease of the eukaryotic tunnel size. Various experiments and simulations have shown that the tunnel can be large enough to accommodate a substantial degree of protein structure (e.g., α-helix, tertiary hairpins and even protein folding), especially in the so-called "folding vestibule" located near the exit port [36, 54, 56, 57, 58, 59]. As the shape of the exit tunnel in this region can affect co-translational protein folding [60], our results suggest that eukaryotic tunnels are less favorable for such folding to occur inside the ribosome. Such a difference could actually reflect the differences in complexity and division of labor between the eukaryotic and prokaryotic chaperone networks [13, 56]. In prokaryotes, both co-translational folding and denaturation of proteins during stress are ensured by an overlapping set of chaperones that primarily relies on the bacterial trigger factor [61]. In contrast, distinct specialized networks of downstream chaperones evolved in eukaryotes to separately carry out these processes [62]. In particular, the existence of a more efficient chaperone network devoted to folding may be the result of eukaryotes

9

having a higher proportion of larger multidomain proteins, as well as more complex protein folds [63]. While the prokaryotic tunnel can assist the chaperone network by pre-folding the nascent chain inside the tunnel [64], the Prefoldin family of proteins [65] plays such a role instead in eukarya and archaea, which could explain the observed reduced size of the tunnel for these domains.

Another example of the direct implication of the bacterial ribosome tunnel on co-translational processes is the translocation of membrane proteins, which is mediated in both eukarya and bacteria by the Signal Recognition Particle (SRP), a ribonucleoprotein that recognizes the signal peptide emerging from the ribosome [1, 8, 66, 67]. Such a process is constrained by the short time window to recognize and target the signal peptide [66, 68]. In bacteria, the SRP can be recruited prior to the emergence of the peptide from the tunnel [69], via uL23 that both recognizes the nascent sequence inside the tunnel and binds to the SRP at the outer surface of the ribosome. In eukaryotes, however, forward signaling cannot similarly occur, as the replacement of uL23 by eL39 disrupts the original coupling between the signal peptide inside the tunnel and the SRP recruitment [8]. To compensate for this lacking function, it has been shown that eukaryote-specific signaling mechanisms can promote the progressive binding of the SRP [68], while the usage of non-optimal codons, associated with translation slowdown, also increases the binding opportunity time window for the SRP [66].

Finally, while the reduction of the tunnel size in eukarya seems to reflect the externalization of major co-translational functions, the increase of tunnel confinement may also provide some other advantages. Besides the aforementioned facilitation of nascent chain elongation, the reduction of the tunnel radius at the exit port by eL39 and the addition of a second constriction site also contribute to restricting the access to the tunnel and the PTC from external threats. For example, as the insertion of amino acids in the loop of uL4 in *E. coli* confers resistance to larger-size macrolides [70, 71], it has been suggested that the narrower size of the constriction site in eukarya can block the access of these antibiotics to the targeted PTC; we actually found a second eukaryote-specific constriction site, located below the universal constriction site, to be responsible for this narrower access. Similarly, mutants lacking eL39 are more vulnerable against ribosome targeting antibiotics [72], in addition to an increased translation error rate and cold sensitive phenotype. Molecular dynamic simulations have predicted interaction between the 28S rRNA tetraloop and eL39, potentially leading to even more obstruction of the tunnel, suggesting that eL39 acts as an energy barrier to ensure protein quality control and protect the ribosome from deleterious external agents [19].

Future directions

Based on the present study and the aforementioned involvement of the tunnel structure in various co-translational processes, it would be interesting to thoroughly investigate the inter-domain differences in the chaperone machinery engaging the ribosome exit tunnel, especially near the exit port. For example, the class of chaperones involved in guiding *de novo* protein folding includes multiple domain-specific actors, such as the bacterial trigger factor, the archaeal and eukaryotic nascent polypeptide-associated complex and specialized eukaryotic heat shock proteins [61], the diversity of which may be the result of different strategies developed in conjunction with the tunnel structure. Elucidating the evolutionary basis of these processes might also require the development of more refined tools to extract and analyze the tunnel geometry and its properties. It is quite remarkable that only considering the radius variation and a simple metric allowed us to recapitulate the general phylogenetic relationship. Integrating more information from the geometry and structure of the tunnel should allow to study and interpret smaller variations of the tunnel, which should also become more reliable as the resolution of the structures improves. In particular, it would be interesting to build a stochastic model describing the evolution of the tunnel geometry, to decipher the larger variation detected in bacteria and to determine if it is mainly due to the difference in evolutionary

time scale or whether natural selection played a role.

As the tunnel wall responds to the nascent peptide contained within it [73], it would be important to analyze and compare the geometry of the tunnel in the presence of a nascent peptide, and specifically arrest sequences known to interact with the tunnel. While the responsiveness of the tunnel to the chain is in general likely to be at the level of small movements [74], we interestingly found that the tunnel structure of *E. coli* containing the ErmCL chain seemed to deviate more substantially from the original one in the region where the sequence spans (Figure S3). The methods presented here could be extended and applied to more structures in the future, and as a result provide useful insights into many of the essential biological processes in which the ribosome takes part.

METHODS

Ribosome structures

Cryo-EM reconstructions and X-ray crystallography structures of ribosomes were downloaded from the protein data bank https://www.rcsb.org/. References and details are given in Table 1. The fitting of the model's residues to the original map was evaluated for each structure using coot [75] density-fit-score function (2Fo-Fc maps were used for X-Ray data).

Extraction of the ribosome tunnel geometry

To extract the tunnel coordinates, we used a tunnel search algorithm developed by Sehnal *et al.* [76]. The tunnel search was initiated at the PTC. To locate the PTC, we aligned the sequences of 23S and 28S rRNAs and selected the nucleotide aligned with U4452 in human. The tunnel search algorithm was applied after editing the structure to contain atoms located less than 80 Å from the constriction site. For the human ribosome, the constriction site was obtained by computing the center of mass of amino acids G70, R71, in uL4 and H133 in uL22. The same procedure was applied for the other species after alignment of proteins uL4 and uL22. As the algorithm gives several possible tunnels, we manually picked the ribosome exit tunnel by looking at the shape, length and relative position to uL4 and uL22. Coordinates were extracted using Pymol and Python custom scripts. The origin of the tunnel was set at a distance of 5 Å from the nucleotide used to initiate the tunnel search algorithm, to remove sensitive regions generated by the tunnel search algorithm close to the initial point. Downstream analysis for computing global and local geometric features of the tunnel were done using Matlab and Python (More details in Supporting Information).

Distance metric for pairwise comparison and clustering of radius plots

For two tunnels T_1 and T_2 parametrized by $T_i = (\bar{S}^{(i)}, \bar{R}^{(i)})$, where $i = 1, 2$, $\bar{S}$ is an arc length parametrization of the tunnel centerline 3D coordinates and $\bar{R}$ is the associated radius (in ångström), we introduced the following distance metric $D(T_1, T_2)$ given by

$$D(T_1, T_2) = \min_{|\delta| \leq \ell} d_\delta(T_1, T_2), \tag{1}$$

where ℓ is the maximum shift length and

$$d_\delta(T_1, T_2) = \frac{\int_{\bar{S}_\delta^{(1)} \cap \bar{S}^{(2)}} \left[\bar{R}_\delta^{(1)}(s) - \bar{R}^{(2)}(s)\right]^p ds}{\int_{\bar{S}_\delta^{(1)} \cap \bar{S}^{(2)}} ds} + \varepsilon|\delta|, \tag{2}$$

where $\bar{S}_\delta^{(1)} = \bar{S}^{(1)} - \delta$, and $\forall s \in \bar{S}$, $\bar{R}_\delta(s - \delta) = \bar{R}(s - \delta)$. In other words, we looked for the best alignment of the two tunnels such that it minimizes the average L^p difference between the two aligned radius plots, with a negative penalty for the size of the alignment shift (which is also limited to be less than ℓ). We then take this measurement as the distance between the two tunnels. In practice, we took $\ell = 20$ Å, $\varepsilon = 0.01$ and $p = 2$ (for more details on how sensitive our results are with these parameters, see Supporting Information), and evaluated the integrals by computing their Riemann sum. After computing the pairwise distance matrix associated with our dataset, we constructed the associated phylogenetic tree using the Unweighted Pair Group Method Average (UPGMA) algorithm.

Sequence alignment and conservation of ribosomal proteins

Ribosomal proteins are named according to the system set by Ban *et al.* [77]. Sequences were aligned using MAFFT [78] with default parameters and visualized in Jalview [79]. Conservation scores were computed using Jensen-Shannon divergence and software provided by Capra and Singh [31], using the default parameters.

Conserved motifs in ribosomal RNA

For each of the three domains (bacteria, archaea and eukarya), conserved motif sequences were obtained from [22], with their positions along the rRNA sequence computed using exact string matching. For each nucleotide, the distance from the tunnel (or PTC) was computed using MATLAB.

Charge conservation score

To quantify charge conservation among the ribosomal proteins in our study, we introduce the following metric: For a given multiple sequence alignment (MSA) M containing N sequences, let M_C denote the Cth column of the MSA and $M_C^{(i)}$ denote the symbol in column C of sequence i, such that $M_C^i \in AA$, where AA is the set of 21 amino acids plus the gap symbol. For each element x of AA, we assign a charge $c(x) = +1$ for lysine and arginine, -1 for aspartic acid and glutamic acid and 0 for all other amino acids plus the gap symbol. For each column C of the MSA, we define the conservation charge score $S(C)$, as

$$S(C) = \sum_{n=C-w}^{C+w} \sum_{i=1}^{N} c(M_n^{(i)}),$$

where the window size w is set in our analysis to be 2. Analysis of conservation scores was done using Matlab custom scripts.

Visualization tools

Structures were visualized using Pymol. Maps of ribosomal RNA secondary structures, distance of the rRNA nucleotides to the tunnel (or PTC) and conserved motifs were visualized in RiboVision [80].

12

Acknowledgments

REFERENCES

1. Nyathi Y, Wilkinson BM, Pool MR. Co-translational targeting and translocation of proteins to the endoplasmic reticulum. Biochimica et Biophysica Acta (BBA)-Molecular Cell Research. 2013;1833(11):2392–2402.

2. Thommen M, Holtkamp W, Rodnina MV. Co-translational protein folding: progress and methods. Current opinion in structural biology. 2017;42:83–89.

3. Rodnina MV, Wintermeyer W. Protein elongation, co-translational folding and targeting. Journal of molecular biology. 2016;428(10):2165–2185.

4. Choi J, Grosely R, Prabhakar A, Lapointe C, Wang J, Puglisi J. How Messenger RNA and Nascent Chain Sequences Regulate Translation Elongation. Annual review of biochemistry. 2018;87:421–449.

5. Vazquez-Laslop N, Thum C, Mankin AS. Molecular mechanism of drug-dependent ribosome stalling. Molecular cell. 2008;30(2):190–202.

6. Davis AR, Gohara DW, Yap MNF. Sequence selectivity of macrolide-induced translational attenuation. Proceedings of the National Academy of Sciences. 2014;111(43):15379–15384.

7. Tenson T, Ehrenberg M. Regulatory nascent peptides in the ribosomal tunnel. Cell. 2002;108(5):591–594.

8. Ito K, editor. Regulatory Nascent Polypeptides. Springer; 2014.

9. Dao Duc K, Song YS. The impact of ribosomal interference, codon usage, and exit tunnel interactions on translation elongation rate variation. PLoS genetics. 2018;14(1):e1007166.

10. Kisselev L. Polypeptide release factors in prokaryotes and eukaryotes: same function, different structure. Structure. 2002;10(1):8–9.

11. Kozak M. Initiation of translation in prokaryotes and eukaryotes. Gene. 1999;234(2):187–208.

12. Ryazanov A, Ovchinnikov L, Spirin A. Development of structural organization of protein-synthesizing machinery from prokaryotes to eukaryotes. Biosystems. 1987;20(3):275–288.

13. Pechmann S, Willmund F, Frydman J. The ribosome as a hub for protein quality control. Molecular cell. 2013;49(3):411–421.

14. Voss N, Gerstein M, Steitz T, Moore P. The geometry of the ribosomal polypeptide exit tunnel. Journal of molecular biology. 2006;360(4):893–906.

15. Ishida H, Hayward S. Path of nascent polypeptide in exit tunnel revealed by molecular dynamics simulation of ribosome. Biophysical journal. 2008;95(12):5962–5973.

16. Steitz TA. A structural understanding of the dynamic ribosome machine. Nature Reviews Molecular Cell Biology. 2008;9(3):242–253.

17. Lu J, Deutsch C. Electrostatics in the ribosomal tunnel modulate chain elongation rates. Journal of Molecular Biology. 2008;384(1):73–86.

18. Lucent D, Snow CD, Aitken CE, Pande VS. Non-bulk-like solvent behavior in the ribosome exit tunnel. PLoS Comput Biol. 2010;6(10):e1000963.

19. Petrone PM, Snow CD, Lucent D, Pande VS. Side-chain recognition and gating in the ribosome exit tunnel. Proceedings of the National Academy of Sciences. 2008;105(43):16549–16554.

20. Olsen GJ, Woese CR. Ribosomal RNA: a key to phylogeny. The FASEB journal. 1993;7(1):113–123.

21. Petrov AS, Bernier CR, Hsiao C, Norris AM, Kovacs NA, Waterbury CC, et al. Evolution of the ribosome at atomic resolution. Proceedings of the National Academy of Sciences. 2014;111(28):10251–10256.

22. Doris SM, Smith DR, Beamesderfer JN, Raphael BJ, Nathanson JA, Gerbi SA. Universal and domain-specific sequences in 23S–28S ribosomal RNA identified by computational phylogenetics. RNA. 2015;21(10):1719–1730.

23. Wilson DN, Cate JHD. The structure and function of the eukaryotic ribosome. Cold Spring Harbor perspectives in biology. 2012;4(5):a011536.

24. Li Z, Guo Q, Zheng L, Ji Y, Xie YT, Lai DH, et al. Cryo-EM structures of the 80S ribosomes from human parasites Trichomonas vaginalis and Toxoplasma gondii. Cell Research. 2017;27(10):1275.

25. Gray MW, Burger G, Lang BF. The origin and early evolution of mitochondria. Genome biology. 2001;2(6):reviews1018–1.

26. White TC, Rudenko G, Borst P. Three small RNAs within the 10 kb trypanosome rRNA transcription unit are analogous to domain VII of other eukaryotic 28S rRNAs. Nucleic acids research. 1986;14(23):9471–9489.

27. Gao H, Ayub MJ, Levin MJ, Frank J. The structure of the 80S ribosome from Trypanosoma cruzi reveals unique rRNA components. Proceedings of the National Academy of Sciences of the United States of America. 2005;102(29):10206–10211.

28. Harms J, Schluenzen F, Zarivach R, Bashan A, Gat S, Agmon I, et al. High resolution structure of the large ribosomal subunit from a mesophilic eubacterium. Cell. 2001;107(5):679–688.

29. Yilmaz P, Parfrey LW, Yarza P, Gerken J, Pruesse E, Quast C, et al. The SILVA and "all-species living tree project (LTP)" taxonomic frameworks. Nucleic acids research. 2013;42(D1):D643–D648.

30. Yarza P, Ludwig W, Euzéby J, Amann R, Schleifer KH, Glöckner FO, et al. Update of the All-Species Living Tree Project based on 16S and 23S rRNA sequence analyses. Systematic and applied microbiology. 2010;33(6):291–299.

31. Capra JA, Singh M. Predicting functionally important residues from sequence conservation. Bioinformatics. 2007;23(15):1875–1882.

32. Wyant GA, Abu-Remaileh M, Frenkel EM, Laqtom NN, Dharamdasani V, Lewis CA, et al. NUFIP1 is a ribosome receptor for starvation-induced ribophagy. Science. 2018; p. eaar2663.

33. Frank J. Advances in the field of single-particle cryo-electron microscopy over the last decade. Nature protocols. 2017;12(2):209.

34. Razi A, Britton RA, Ortega J. The impact of recent improvements in cryo-electron microscopy technology on the understanding of bacterial ribosome assembly. Nucleic acids research. 2016;45(3):1027–1040.

35. Berg JM, Tymoczko JL, Stryer L, et al.. Biochemistry; 2002.

36. Nissen P, Hansen J, Ban N, Moore PB, Steitz TA. The structural basis of ribosome activity in peptide bond synthesis. Science. 2000;289(5481):920–930.

37. Fox GE. Origin and evolution of the ribosome. Cold Spring Harbor perspectives in biology. 2010;2(9):a003483.

38. Krupkin M, Matzov D, Tang H, Metz M, Kalaora R, Belousoff MJ, et al. A vestige of a prebiotic bonding machine is functioning within the contemporary ribosome. Philosophical Transactions of the Royal Society B: Biological Sciences. 2011;366(1580):2972–2978.

39. Agmon I, Bashan A, Yonath A. On ribosome conservation and evolution. Israel Journal of Ecology & Evolution. 2006;52(3-4):359–374.

40. Lott BB, Wang Y, Nakazato T. A comparative study of ribosomal proteins: linkage between amino acid distribution and ribosomal assembly. BMC biophysics. 2013;6(1):13.

41. Wilson DN, Beckmann R. The ribosomal tunnel as a functional environment for nascent polypeptide folding and translational stalling. Current opinion in structural biology. 2011;21(2):274–282.

42. Kannan K, Vázquez-Laslop N, Mankin AS. Selective protein synthesis by ribosomes with a drug-obstructed exit tunnel. Cell. 2012;151(3):508–520.

43. Cabrita LD, Dobson CM, Christodoulou J. Protein folding on the ribosome. Current opinion in structural biology. 2010;20(1):33–45.

44. Fedyukina DV, Cavagnero S. Protein folding at the exit tunnel. Annual review of biophysics. 2011;40:337–359.

45. Chen C, Wang E, Liu P, Xiao Y. Simulation study of the role of the ribosomal exit tunnel on protein folding. Physical Review E. 2013;87(2):022701.

46. Javed A, Christodoulou J, Cabrita LD, Orlova EV. The ribosome and its role in protein folding: looking through a magnifying glass. Acta Crystallographica Section D: Structural Biology. 2017;73(6):509–521.

47. Ismail N, Hedman R, Schiller N, Von Heijne G. A biphasic pulling force acts on transmembrane helices during translocon-mediated membrane integration. Nature Structural and Molecular Biology. 2012;19(10):1018.

48. Butkus ME, Prundeanu LB, Oliver DB. Translocon "pulling" of nascent SecM controls the duration of its translational pause and secretion-responsive secA regulation. Journal of bacteriology. 2003;185(22):6719–6722.

49. Liu B, Han Y, Qian SB. Cotranslational response to proteotoxic stress by elongation pausing of ribosomes. Molecular cell. 2013;49(3):453–463.

50. Goldman DH, Kaiser CM, Milin A, Righini M, Tinoco I, Bustamante C. Mechanical force releases nascent chain–mediated ribosome arrest in vitro and in vivo. Science. 2015;348(6233):457–460.

51. Fritch B, Kosolapov A, Hudson P, Nissley DA, Woodcock HL, Deutsch C, et al. Origins of the mechanochemical coupling of peptide bond formation to protein synthesis. Journal of the American Chemical Society. 2018;140(15):5077–5087.

52. Seifert U, Wintz W, Nelson P. Straightening of thermal fluctuations in semiflexible polymers by applied tension. Physical review letters. 1996;77(27):5389.

53. Hallatschek O, Frey E, Kroy K. Propagation and relaxation of tension in stiff polymers. Physical review letters. 2005;94(7):077804.

54. Nilsson OB, Hedman R, Marino J, Wickles S, Bischoff L, Johansson M, et al. Cotranslational protein folding inside the ribosome exit tunnel. Cell reports. 2015;12(10):1533–1540.

55. Ziv G, Haran G, Thirumalai D. Ribosome exit tunnel can entropically stabilize α-helices. Proceedings of the National Academy of Sciences of the United States of America. 2005;102(52):18956–18961.

56. Kramer G, Boehringer D, Ban N, Bukau B. The ribosome as a platform for co-translational processing, folding and targeting of newly synthesized proteins. Nature Structural and Molecular Biology. 2009;16(6):589.

57. Lu J, Deutsch C. Folding zones inside the ribosomal exit tunnel. Nature Structural and Molecular Biology. 2005;12(12):1123.

58. O'Brien EP, Christodoulou J, Vendruscolo M, Dobson CM. New scenarios of protein folding can occur on the ribosome. Journal of the American Chemical Society. 2011;133(3):513–526.

59. Marino J, Heijne G, Beckmann R. Small protein domains fold inside the ribosome exit tunnel. FEBS letters. 2016;590(5):655–660.

60. Kudva R, Pardo-Avila F, Sandhu H, Carroni M, Bernstein HD, Von Heijne G. The Shape of the Ribosome Exit Tunnel Affects Cotranslational Protein Folding. bioRxiv. 2018; p. 274191.

61. Preissler S, Deuerling E. Ribosome-associated chaperones as key players in proteostasis. Trends in biochemical sciences. 2012;37(7):274–283.

62. Albanese V, Yam AYW, Baughman J, Parnot C, Frydman J. Systems analyses reveal two chaperone networks with distinct functions in eukaryotic cells. Cell. 2006;124(1):75–88.

63. Agashe VR, Guha S, Chang HC, Genevaux P, Hayer-Hartl M, Stemp M, et al. Function of trigger factor and DnaK in multidomain protein folding: increase in yield at the expense of folding speed. Cell. 2004;117(2):199–209.

64. Kudlicki W, Coffman A, Kramer G, Hardesty B. Ribosomes and ribosomal RNA as chaperones for folding of proteins. Folding and Design. 1997;2(2):101–108.

65. Vainberg IE, Lewis SA, Rommelaere H, Ampe C, Vandekerckhove J, Klein HL, et al. Prefoldin, a chaperone that delivers unfolded proteins to cytosolic chaperonin. Cell. 1998;93(5):863–873.

66. Pechmann S, Chartron JW, Frydman J. Local slowdown of translation by nonoptimal codons promotes nascent-chain recognition by SRP in vivo. Nature Structural and Molecular Biology. 2014;21(12):1100.

67. Jan CH, Williams CC, Weissman JS. Principles of ER cotranslational translocation revealed by proximity-specific ribosome profiling. Science. 2014;346(6210):1257521.

68. Berndt U, Oellerer S, Zhang Y, Johnson AE, Rospert S. A signal-anchor sequence stimulates signal recognition particle binding to ribosomes from inside the exit tunnel. Proceedings of the National Academy of Sciences. 2009;106(5):1398–1403.

69. Bornemann T, Jöckel J, Rodnina MV, Wintermeyer W. Signal sequence–independent membrane targeting of ribosomes containing short nascent peptides within the exit tunnel. Nature Structural and Molecular Biology. 2008;15(5):494.

70. Zaman S, Fitzpatrick M, Lindahl L, Zengel J. Novel mutations in ribosomal proteins L4 and L22 that confer erythromycin resistance in Escherichia coli. Molecular microbiology. 2007;66(4):1039–1050.

71. Tu D, Blaha G, Moore PB, Steitz TA. Structures of MLSBK antibiotics bound to mutated large ribosomal subunits provide a structural explanation for resistance. Cell. 2005;121(2):257–270.

72. Dresios J, Derkatch IL, Liebman SW, Synetos D. Yeast ribosomal protein L24 affects the kinetics of protein synthesis and ribosomal protein L39 improves translational accuracy, while mutants lacking both remain viable. Biochemistry. 2000;39(24):7236–7244.

73. Lu J, Hua Z, Kobertz WR, Deutsch C. Nascent peptide side chains induce rearrangements in distinct locations of the ribosomal tunnel. Journal of molecular biology. 2011;411(2):499–510.

74. Seidelt B, Innis CA, Wilson DN, Gartmann M, Armache JP, Villa E, et al. Structural insight into nascent polypeptide chain–mediated translational stalling. Science. 2009;326(5958):1412–1415.

75. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and Development of Coot. Acta Crystallographica Section D - Biological Crystallography. 2010;66:486–501.

76. Sehnal D, Vařeková RS, Berka K, Pravda L, Navrátilová V, Banáš P, et al. MOLE 2.0: advanced approach for analysis of biomacromolecular channels. Journal of cheminformatics. 2013;5(1):1.

77. Ban N, Beckmann R, Cate JH, Dinman JD, Dragon F, Ellis SR, et al. A new system for naming ribosomal proteins. Current opinion in structural biology. 2014;24:165–169.

78. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecular biology and evolution. 2013;30(4):772–780.

79. Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ. Jalview Version 2 multiple sequence alignment editor and analysis workbench. Bioinformatics. 2009;25(9):1189–1191.

80. Bernier CR, Petrov AS, Waterbury CC, Jett J, Li F, Freil LE, et al. RiboVision suite for visualization and analysis of ribosomes. Faraday discussions. 2014;169:195–207.

81. Ahmed T, Shi J, Bhushan S. Unique localization of the plastid-specific ribosomal proteins in the chloroplast ribosome small subunit provides mechanistic insights into the chloroplastic translation. Nucleic acids research. 2017;45(14):8581–8595.

82. Amunts A, Brown A, Toots J, Scheres SH, Ramakrishnan V. The structure of the human mitochondrial ribosome. Science. 2015;348(6230):95–98.

83. Beckert B, Abdelshahid M, Schäfer H, Steinchen W, Arenz S, Berninghausen O, et al. Structure of the Bacillus subtilis hibernating 100S ribosome reveals the basis for 70S dimerization. The EMBO journal. 2017;36(14):2061–2072.

84. Fischer N, Neumann P, Konevega AL, Bock LV, Ficner R, Rodnina MV, et al. Structure of the E. coli ribosome–EF-Tu complex at< 3 Å resolution by C s-corrected cryo-EM. Nature. 2015;520(7548):567.

85. Krupkin M, Wekselman I, Matzov D, Eyal Z, Posner YD, Rozenberg H, et al. Avilamycin and evernimicin induce structural changes in rProteins uL16 and CTC that enhance the inhibition of A-site tRNA binding. Proceedings of the National Academy of Sciences. 2016;113(44):E6796–E6805.

86. Franken LE, Oostergetel GT, Pijning T, Puri P, Arkhipova V, Boekema EJ, et al. A general mechanism of ribosome dimerization revealed by single-particle cryo-electron microscopy. Nature communications. 2017;8(1):722.

87. Hentschel J, Burnside C, Mignot I, Leibundgut M, Boehringer D, Ban N. The complete structure of the mycobacterium smegmatis 70S ribosome. Cell reports. 2017;20(1):149–160.

88. Yang K, Chang JY, Cui Z, Li X, Meng R, Duan L, et al. Structural insights into species-specific features of the ribosome from the human pathogen Mycobacterium tuberculosis. Nucleic acids research. 2017;45(18):10884–10894.

89. Matzov D, Eyal Z, Benhamou RI, Shalev-Benami M, Halfon Y, Krupkin M, et al. Structural insights of lincosamides targeting the ribosome of Staphylococcus aureus. Nucleic acids research. 2017;45(17):10284–10292.

90. Polikanov YS, Melnikov SV, Söll D, Steitz TA. Structural insights into the role of rRNA modifications in protein synthesis and ribosome assembly. Nature Structural and Molecular Biology. 2015;22(4):342.

91. Gabdulkhakov A, Nikonov S, Garber M. Revisiting the Haloarcula marismortui 50S ribosomal subunit model. Acta Crystallographica Section D: Biological Crystallography. 2013;69(6):997–1004.

92. Armache JP, Anger AM, Márquez V, Franckenberg S, Fröhlich T, Villa E, et al. Promiscuous behaviour of archaeal ribosomal proteins: implications for eukaryotic ribosome evolution. Nucleic acids research. 2012;41(2):1284–1293.

93. Natchiar SK, Myasnikov AG, Kratzat H, Hazemann I, Klaholz BP. Visualization of chemical modifications in the human 80S ribosome structure. Nature. 2017;551(7681).

94. Zhang X, Lai M, Chang W, Yu I, Ding K, Mrazek J, et al. Structures and stabilization of kinetoplastid-specific split rRNAs revealed by comparing leishmanial and human ribosomes. Nature communications. 2016;7:13223.

95. Wong W, Bai Xc, Brown A, Fernandez IS, Hanssen E, Condron M, et al. Cryo-EM structure of the Plasmodium falciparum 80S ribosome bound to the anti-protozoan drug emetine. Elife. 2014;3:e03080.

96. Schmidt C, Becker T, Heuer A, Braunger K, Shanmuganathan V, Pech M, et al. Structure of the hypusinylated eukaryotic translation factor eIF-5A bound to the ribosome. Nucleic acids research. 2016;44(4):1944–1951.

97. Gogala M, Becker T, Beatrix B, Armache JP, Barrio-Garcia C, Berninghausen O, et al. Structures of the Sec61 complex engaged in nascent peptide translocation or membrane insertion. Nature. 2014;506(7486):107.

98. Liu Z, Gutierrez-Vargas C, Wei J, Grassucci RA, Ramesh M, Espina N, et al. Structure and assembly model for the Trypanosoma cruzi 60S ribosomal subunit. Proceedings of the National Academy of Sciences. 2016;113(43):12174–12179.

99. Arenz S, Meydan S, Starosta AL, Berninghausen O, Beckmann R, Vázquez-Laslop N, et al. Drug sensing by the ribosome induces translational arrest via active site perturbation. Molecular cell. 2014;56(3):446–452.

100. Osterman IA, Khabibullina NF, Komarova ES, Kasatsky P, Kartsev VG, Bogdanov AA, et al. Madumycin II inhibits peptide bond formation by forcing the peptidyl transferase center into an inactive state. Nucleic acids research. 2017;45(12):7507–7514.

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

TABLE

Table 1. Ribosomes structures used in our study. First column contains the species, with the domain they belong to (b: bacteria, a: archaea, e: eukarya). The upper part of the table contains the main structures of the study. Replicates for *H. sapiens*, *E. coli* and *T. thermophilus* are in the lower part.

| Species/Organelles	Resolution	Reference	PDB codename
Chloroplast (*Spinacia*)	3.8 Å (EM)	Ahmed *et al.* (2017) [81]	5X8T
Mitochondria (*H. Sapiens*)	3.1 Å (EM)	Amunts *et al.* (2015) [82]	3J9M
B. subtilis (b)	3.8 Å (EM)	Beckert *et al.* (2017) [83]	5NJT
E. coli (b)	2.9 Å (EM)	Fischer *et al.* (2015) [84]	5AFI
D. radiodurans (b)	3.4 Å (X-Ray)	Krupkin *et al.* (2016) [85]	5JVG
L. lactis (b)	5.6 Å (EM)	Franken *et al.* (2017) [86]	5MYJ
M. smegmatis (b)	3.2 Å (EM)	Hentschel *et al.* (2017) [87]	5O60
M. tuberculosis (b)	3.4 Å (EM)	Yang *et al.* (2017) [88]	5V7Q
S. aureus (b)	3.4 Å (X-Ray)	Matzov *et al.* (2017) [89]	5NRG
T. thermophilus (b)	2.5 Å (X-Ray)	Polinakov *et al.* (2015) [90]	4Y4P
H. marismortui (a)	2.4 Å (X-Ray)	Gabdulkhakov *et al.* (2013) [91]	4V9F
P. furiosus (a)	6.6 Å (EM)	Armache *et al.* (2012) [92]	4V6U
H. sapiens (e)	2.9 Å (EM)	Natchiar *et al.* (2017) [93]	6EK0
L. donovani (e)	2.9 Å (EM)	Zhang *et al.* (2016) [94]	5T2A
P. falciparum (e)	3.2 Å (EM)	Wong *et al.* (2014) [95]	3J79
S. cerevisiae (e)	3.9 Å (EM)	Schmidt *et al.* (2016) [96]	5GAK
T. aestivum (e)	5.5 Å (EM)	Gogala *et al.* (2014) [97]	4V7E
T. cruzi (e)	2.5 Å (EM)	Liu *et al.* (2016) [98]	5T5H
T. gondii (e)	3.2 Å (EM)	Li *et al.* (2017) [24]	5XXB
T. vaginalis (e)	3.4 Å (EM)	Li *et al.* (2017) [24]	5XY3
E. coli (b)	3.9 Å (EM)	Arenz *et al.* (2014) [99]	3J7Z
T. thermophilus (b)	2.8 Å (X-Ray)	Osterman *et al.* (2017) [100]	5VP2
H. sapiens (e)	3.6 Å (EM)	Khatter *et al.* (2015) [101]	4UG0

FIGURES
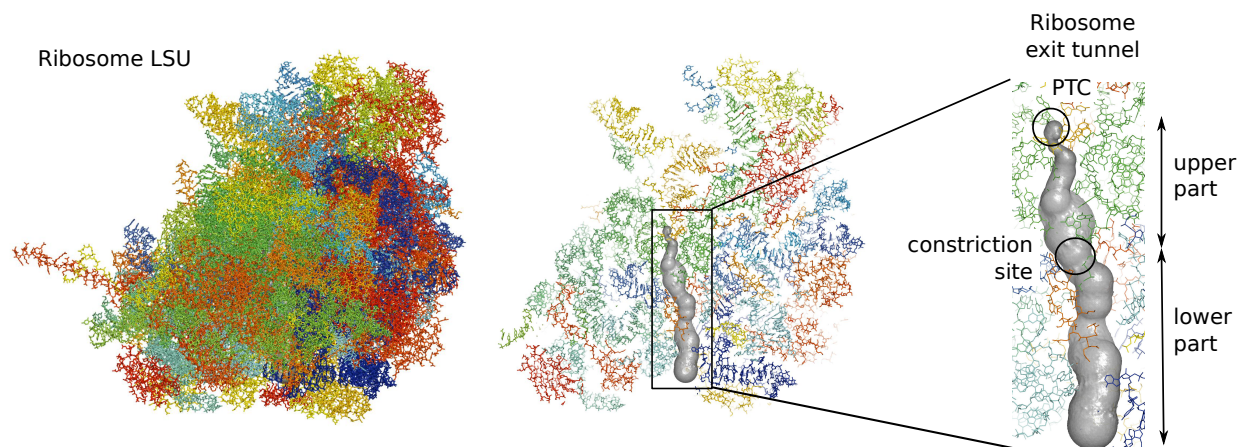
Figure 1. Extraction of the ribosome exit tunnel coordinates. For a given structure of the ribosome large subunit (LSU) represented in the left panel (from Schmidt *et al.* [96]), we first locate the polypeptide transferase center (PTC) and then apply a tunnel search algorithm [76] to reconstruct the geometry of the exit tunnel, illustrated on the right (more details in the Methods section). The inset in the right panel shows the exit tunnel with the PTC and the constriction site that separates the tunnel between its upper and lower parts.
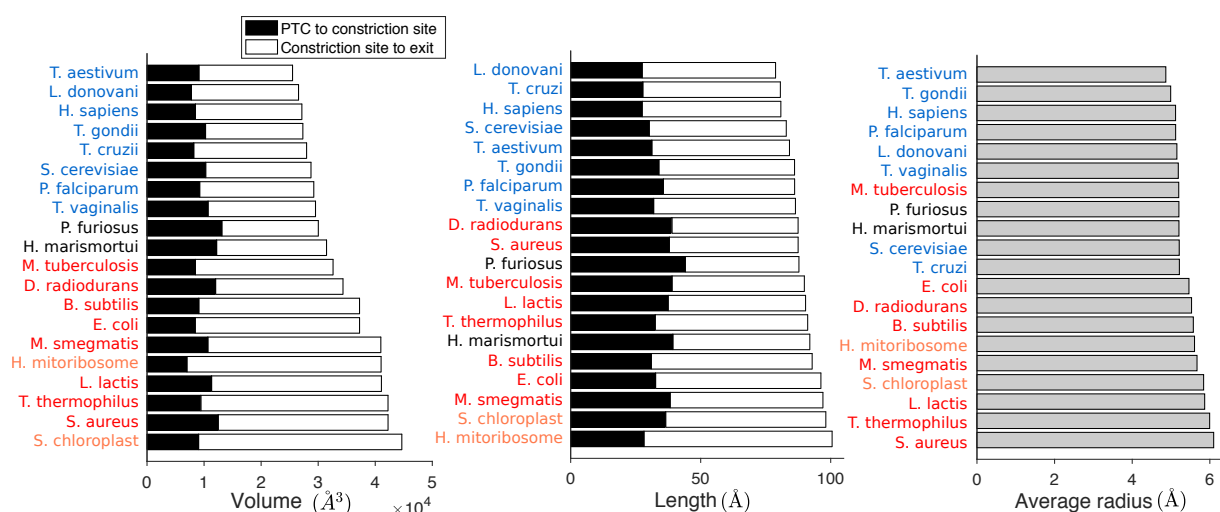
Figure 2. Volume, length and average radius of the ribosome exit tunnel across different species. Horizontal bar plots represent the ordered volume (left), length (middle) and average radius (right) of the tunnels from our dataset. The volume and length are decomposed into two subparts, separated by the constriction site (see Methods). Species are specified and colored by their respective domains (plus organelles, in orange): bacteria (red), archaea (black), eukarya (blue).
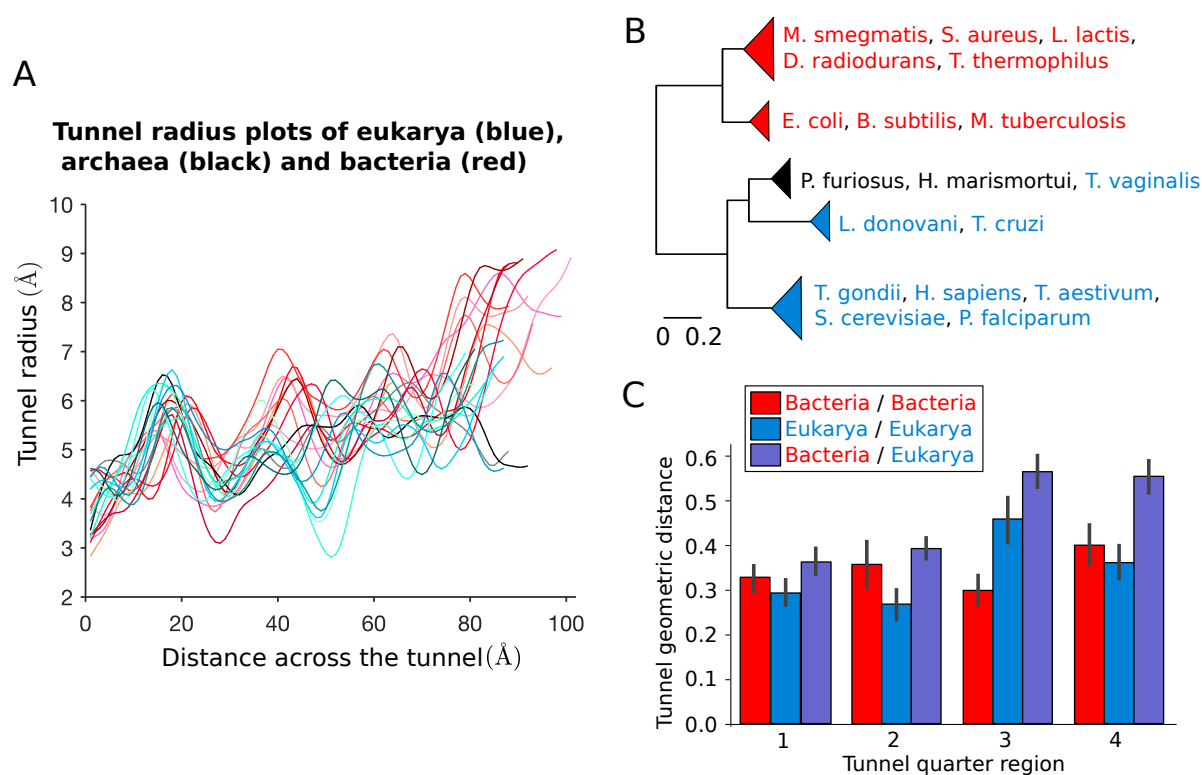
Figure 3. Clustering of species obtained from pairwise comparison of the tunnel geometry. **A:** For all our structures, we plot the tunnel radius as a function of the distance across the tunnel. These plots are used to compare the tunnel geometries (more details in Figure S1). **B:** Clustering obtained after applying our tunnel geometric distance metric to bacteria, eukarya and archaea of our dataset (see Methods). The first main branch encompasses the bacterial ribosomes, highlighted in red, while the second contains eukarya (blue) and archaea (black). For the full clustering and phylogenetic trees obtained from 16S/18S rRNA sequences, see Figure S2. **C:** We divide for each couple of species their common domain after alignment in 4 quarters (see Methods), and use the same metric to compute the distance in each of the subregions. The bar plots represent for each quarter the average and std of the geometric distance for subset of pairs made of 2 prokaryotes (red), 2 eukaryotes (blue), and 1 prokaryote and 1 eukaryote (violet).
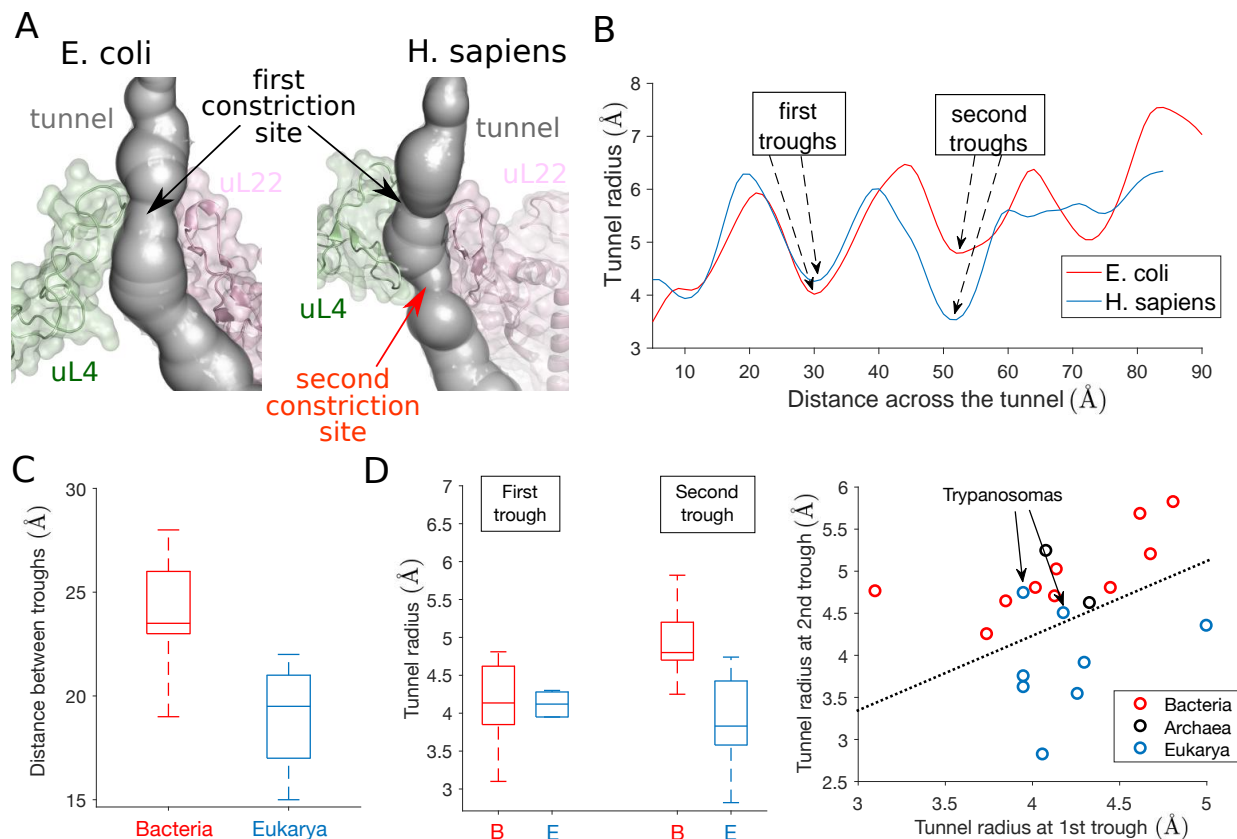
Figure 4. The presence of a second constriction site in eukarya explains the geometric difference observed between bacterial and eukaryotic tunnels. A: We show the constriction site region in *E. coli* (left), obtained from Fischer *et al.* [84] and *H. Sapiens* (right), obtained from Natchiar *et al.* [93]. The structure is surrounded by ribosomal proteins uL4 (green) and uL22 (pink). An extended arm in *H. sapiens* uL4 produces a second constriction site (see also Figure S5). **B:** The plots of the tunnel radius as a function of the tunnel distance shows a first trough associated with the constriction site (around position 30), common to *E. coli* and *H. sapiens*. A second trough appears around position 50, lower than the first trough in *H. sapiens*. **C:** We compare the distance between the troughs for bacteria (left box plot, in red) and eukarya (right box plot, in blue) in our dataset. The interquartile range is indicated by the box, the median by a line inside, and upper and lower adjacent values by whiskers. **D:** In left, we provide the same comparison as in **C** for the tunnel radii associated with the first and second troughs. In right, we compare the radius of the first and second troughs for each species of our dataset. The second trough radius is larger than the first one for all archaea (black dots) and bacteria (red dots). In contrast this is only the case in eukarya for trypanosome species *L. donovani* and *T. cruzi*.
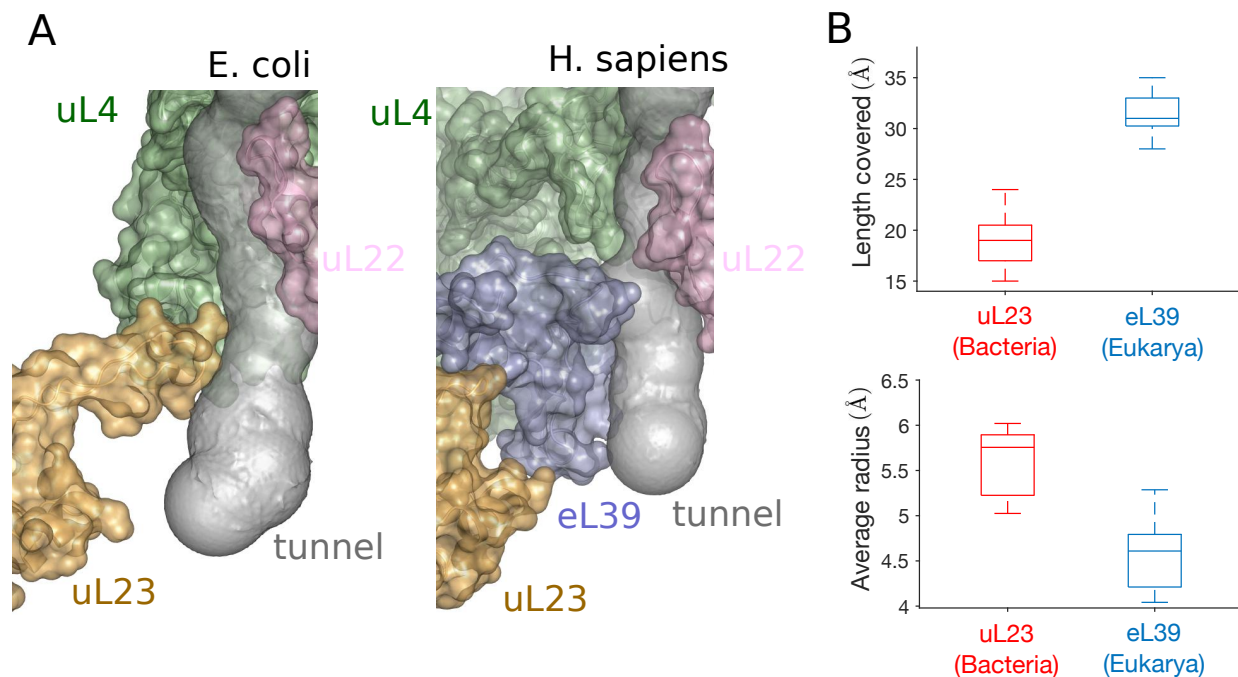
Figure 5. The replacement of uL23 by eL39 in eukarya impacts the tunnel geometry. **A:** The structures of the lower part of the tunnel in *E. coli* (left) and *H. sapiens* (right) show the replacement of ribosomal protein uL23 (in yellow) by el39 (in violet) in *H. sapiens*, which also covers a larger portion of the tunnel. **B:** Upper plot shows a comparison of the distance covered by uL23 and eL39 in bacteria (red box plot) and eukarya (left box plot). Lower plot shows the same comparison for the average radius.
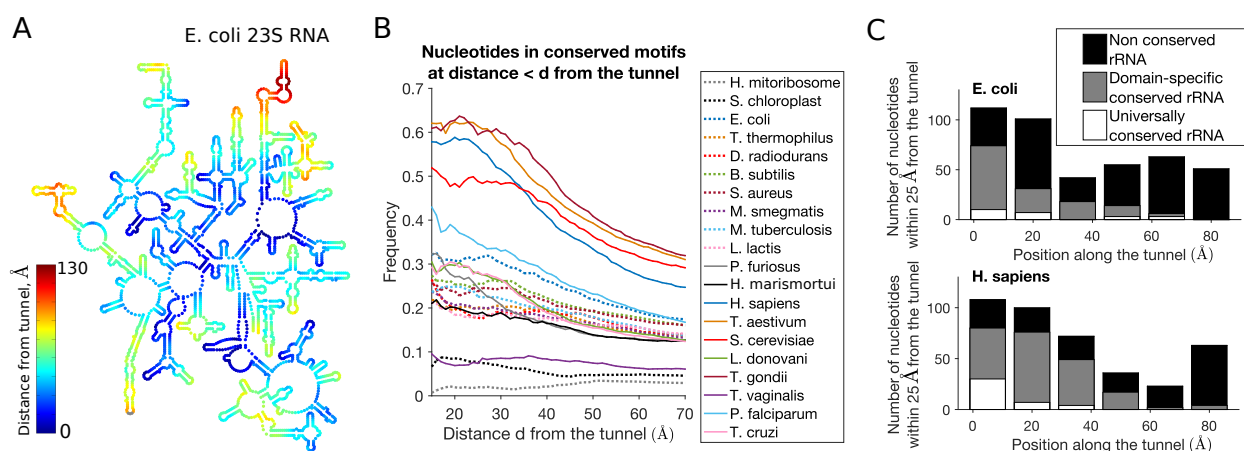
Figure 6. Association between geometric and sequence conservations of ribosomal rRNA. **A:** A map of the secondary structure of the 23S rRNA in *E. coli*, colored by the distance from tunnel (see also Figure S8). **B:** For a given species and distance d, we look at all the rRNA nucleotides located within distance d from the tunnel, and we compute the frequency of conserved elements [22]. We plot this frequency as a function of d for all the species of our dataset (see also Figure S9). **C:** We study the local conservation of rRNA nucleotides along the tunnel: Upon dividing the tunnel into regions of 15 Å along the centerline, we consider for each region all the rRNA nucleotides that are the closest and located within 25 Å, and we compute the associated number of conserved, domain-specific and universally conserved elements. We show here the resulting plots for *E. coli* (up) and *H. sapiens* (down) (for other species, see Figure S10).
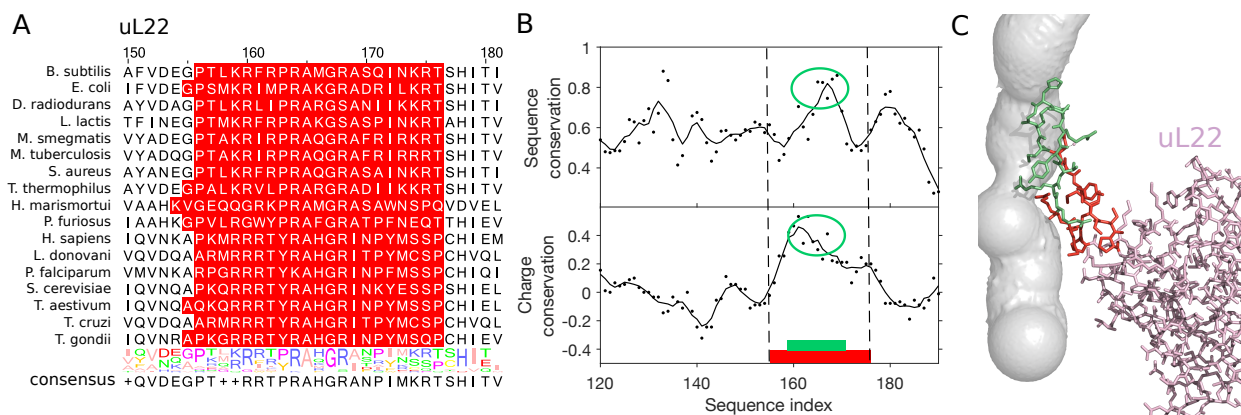
26

Figure 7. Conservation of sequence and positive charge of ribosomal protein uL22 at the tunnel. **A:** We show the multiple sequence alignment of ribosomal protein uL22 close to the tunnel. Residues highlighted in red are the ones located within 10 Å from the tunnel. **B:** We plot the sequence and charge conservation scores (see Methods) along the sequence alignment. Continuous lines represent the signal averaged over a window of 5 sites. Region in red is the same as in **A**. In green, we highlight a subregion of residues close to the tunnel, with a peak in charge or sequence conservation. **C:** The associated structure of uL22 in *H. sapiens*, where residues in green and red correspond to the ones highlighted in **B**. In particular, the green region of high charge and sequence conservation is also in direct contact with the constriction sites. For the other ribosomal proteins associated with the tunnel, see Figure S11 and Figure S12.