Supplementary Materials

Frozen in Time: The History of Proteins

Nicholas A. Kovacs*, Anton S. Petrov*, Kathryn A. Lanier and Loren Dean Williams†

School of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta, GA 30332-0400, USA.

* These authors contributed equally to this work

† To whom correspondence may be addressed. Email: loren.williams@chemistry.gatech.edu
Fig. S1. Contacts between universal rProteins and LSU rRNA mapped onto the secondary structure of *T. thermophilus* LSU rRNA. Nucleotides within 4 Å of any atom of a universal rProtein are indicated by black circles with white centers. rRNA is colored by phase. Phase 1, the most ancient phase is dark blue, Phase 2 is light blue, Phase 3 is green, Phase 4 is yellow, Phase 5 is orange, and Phase 6, the most recent prokaryotic phase, is red. Each phase contains rRNA that is associated by time of acquisition (Petrov, et al. 2014; Petrov, et al. 2015). This figure was generated with the program RiboVision (Bernier, et al. 2014).
Table S1. Segmentation of universal rProteins in *T. thermophilus*, *E. coli*, and *P. furiosus*. Each rProtein segment is colored by phase following the coloring the scheme in figure 1B. Empty cells indicate that no protein segment belongs to that evolutionary phase for that protein. The numbering of the amino acids is given for each segment. rProtein images are from the *T. thermophilus* crystal structure when present in the structure. Otherwise, the images are from the *E. coli* or *P. furiosus* structures. TT = *T. thermophilus*, EC = *E. coli*, PF = *P. furiosus* (*T. thermophilus*, PDB entry 1VY4; *E. coli*, PDB entry 4V9D; *P. furiosus*, PDB entry 4V6U).

| Phase 3 | Phase 4 | Phase 5 | Phase 6 | Phase 3 | Complete |
|---------|---------|---------|---------|---------|----------|
| uL1 (PF) |         |         |         |         |          |
| TT residues |         |         |         |         |          |
| EC residues |         |         |         |         |          |
| PF residues |         |         |         |         |          |
| uL2 |         |         |         |         |          |
| TT residues | 226-242 | 217-225, 243-260 | 120-216, 261-276 | 2-119 | 2-276 |
| EC residues | 224-239 | 215-223, 240-257 | 118-214, 258-271 | 1-117 | 1-271 |
| PF residues | 197-211 | 187-196, 212-239 | 86-186 | 1-86 | 1-239 |
| uL3 |         |         |         |         |          |
| TT residues | 123-155 | 113-122, 156-159 | 1-30, 92-112, 160-204 | 31-91 | 1-204 |
| EC residues | 128-160 | 18-127, 161-164 | 1-31, 97-117, 165-209 | 32-96 | 1-209 |
| PF residues | 240-269 | 229-239, 270-274 | 1-40, 41-77, 208-228, 275-365 | 78-207 | 1-365 |
| uL4 |         |         |         |         |          |
| TT residues | 54-87 | 44-53, 88-98 |         | 6-43, 99-207 | 6-207 |
| EC residues | 49-82 | 39-48, 83-93 |         | 1-38, 94-201 | 1-201 |
| PF residues | 50-90 | 42-49, 91-102 |         | 1-41, 103-255 | 1-255 |
|      | TT residues | EC residues | PF residues |
|------|-------------|-------------|-------------|
| **uL5** |             |             |             |
| TT residues |             |             |             |
| EC residues |             |             |             |
| PF residues |             |             |             |
| **uL6** |             |             |             |
| TT residues |             |             |             |
| EC residues |             |             |             |
| PF residues |             |             |             |
| **uL10** (PF) | | | |
| TT residues |             |             |             |
| EC residues |             |             |             |
| PF residues |             |             |             |
| **uL11** (EC) | | | |
| TT residues |             |             |             |
| EC residues |             |             |             |
| PF residues |             |             |             |
| **uL13** |             |             |             |
| TT residues |             |             |             |
| EC residues |             |             |             |
| PF residues |             |             |             |
|        | TT residues | EC residues | PF residues |
|--------|-------------|-------------|-------------|
| uL14   | 1-6         | 7-87        | 88-122      |
|        | 1-6         | 7-87        | 88-122      |
|        | 20-23       | 10-19, 26-106| 107-141     |
|        |             |             | 10-141      |
| uL15   | 20-40       | 1-19, 41-62 | 64-78       |
|        | 20-40       | 1-19, 41-62 | 63-72       |
|        | 9-28        | 1-8, 29-59  | 60-72       |
|        |             |             | 73-147      |
| uL16   | 77-88       | 1-76, 89-141| 1-141       |
|        | 76-87       | 1-75, 88-136| 1-136       |
|        | 97-173      | 5-96, 118-170| 3-170      |
| uL18   |             |             |             |
|        |             |             |             |
| uL22   | 81-99       | 1-80, 100-112| 1-112      |
|        | 81-99       | 1-80, 100-110| 81-110     |
|        | 125-141     | 6-124, 142-153| 6-153     |
|        | TT residues | EC residues | PF residues |
|--------|-------------|-------------|-------------|
| uL23   |             |             |             |
|        | 1-95        | 1-95        |             |
|        | 3-86        | 3-86        |             |
| uL24   |             |             |             |
|        | 1-107       | 1-107       |             |
|        | 1-102       | 1-102       |             |
|        | 1-121       | 1-121       |             |
| uL29   |             |             |             |
|        | 1-70        | 1-70        |             |
|        | 1-63        | 1-63        |             |
|        | 1-72        | 1-72        |             |
| uL30   |             |             |             |
|        | 2-60        | 2-60        |             |
|        | 1-58        | 1-58        |             |
|        | 1-58        | 1-58        |             |
Fig. S2. A) A fragment of rRNA from Phase 1 (dark blue), Phase 2 (light blue) and Phase 3 (green) shown in surface representation. Phase 3 rProtein (gray) that is in proximity to this rRNA is also shown in surface representation. B) Same as Panel A, but rotated by 180°. C) The Phase 3 rProtein segment of rProtein uL4 associates with Phase 3 rRNA. The zoomed view shows the rProtein segment in stick representation. (from *T. thermophilus*, PDB entry 1VY4).
Fig. S3. Properties of segments of universal rProteins in Phases 3-6. The number of IMHB$^{BA}$ per amino acid is plotted in red and the SSA per amino acid is plotted in blue. IMHB$^{BA}$ increases from Phase 3 through 6, indicating increase in secondary structure. SSA, calculated when rRNA is computationally omitted, decreases from Phase 3 through 6, indicating collapse to globular domains. There are no segments in Phase 1 and 2. rProteins are from the *E. coli* and *P. furiosus* ribosomes (*E. coli*, PDB entry 4V9D; *P. furiosus*, PDB entry 4V6U).
Fig. S4. Frequency of secondary structural elements in each Phase. Universal rProtein segments transition from random coil to isolated secondary elements to globular domains as the phase increases from 3 to 5. The area of each pie chart is proportional to the number of amino acids within that Phase. Domains convert from predominantly $\beta$-domains in Phase 5 to mixed $\beta$-strand and $\alpha$-helix domains in Phase 6. Isolated helices in Phase 4 are composed of $\alpha$-helices and $3_{10}$-helices of three to seven residues in length. (E. coli, PDB entry 4V9D; P. furiosus, PDB entry 4V6U).
Table S2. Segmentation of bacterial rProteins in *T. thermophilus* and *E. coli*. Each rProtein segment is colored by evolutionary phase by the scheme in figures 1B. Empty cells indicate no protein segment belongs to that evolutionary phase. The numbering of the amino acids is given for each segment. rProtein images are from the *T. thermophilus* crystal structure when present in the structure. Otherwise, the images are from the *E. coli* structure. TT = *T. thermophilus*, EC = *E. coli* (*T. thermophilus*, PDB entry 1VY4; *E. coli*, PDB entry 4V9D).

| Protein | Phase 3 | Phase 4 | Phase 5 | Phase 6 | Complete |
|---------|---------|---------|---------|---------|----------|
| bL9     |         |         |         |         |          |
| TT residues |       |         |         | 1-146   | 1-146    |
| EC residues |     |         |         | 1-149   | 1-149    |
| bL17    |         |         |         |         |          |
| TT residues |       |         |         | 1-118   | 1-118    |
| EC residues |     |         |         | 1-120   | 1-120    |
| bL19    |         |         |         |         |          |
| TT residues |       |         |         | 1-131   | 1-131    |
| EC residues |     |         |         | 1-114   | 1-114    |
| bL21    |         |         |         |         |          |
| TT residues |       | 77-80   | 68-76, 81-88 | 1-67, 89-101 | 1-101 |
| EC residues |     | 79-82   | 69-78, 83-92 | 1-68, 93-103 | 1-103 |
| bL25    |         | 1-91    | 92-171  | 1-171   |          |
| TT residues |       |         |         |         |          |
| EC residues |     | 1-94    | Domain non-existant | 1-94 |
|     | TT residues | EC residues |     |     |     |     |
|-----|-------------|-------------|-----|-----|-----|-----|
| bL27 | 2-9         | 10-18       | 19-84 | 2-84 |     |     |
|     |             | 6-14        | 15-81 |     | 6-81 |     |
| bL28 | 10-45       | 2-9, 46-98  |     | 2-98 |     |     |
|     |             | 9-31        | 1-8, 32-77 |     | 1-77 |     |
| bL31 | 2-12        | 13-24       | 25-60 | 2-60 |     |     |
|     |             | 1-8         | 9-20  | 21-56 | 1-56 |     |
| bL32 | 1-8         |             | 2-54  | 3-52 | 2-54 | 3-52 |
| bL33 |             |             | 2-54  | 3-52 | 2-54 | 3-52 |
| bL34 | 1-7         | 8-48        | 1-48  |     |     |     |
|     | 1-7         | 8-46        | 8-46  |     |     |     |
Table S3. Segmentation of eukaryotic rProteins in *P. furiosus*. Each rProtein segment is colored by evolutionary phase by the scheme in figures 1B. Empty cells indicate no protein segment belongs to that evolutionary phase. The numbering of the amino acids is given for each segment. (*P. furiosus*, PDB entry 4V6U).

| Protein | Phase 3 | Phase 4 | Phase 5 | Phase 6 | Complete |
|---------|---------|---------|---------|---------|----------|
| eL8     |         |         |         | ![Image](4V6U.png) | 1-123    |
| Residues|         |         |         |          |          |
| eL14    | ![Image](4V6U.png) |         | ![Image](4V6U.png) | 1-123    |
| Residues| ![Image](4V6U.png) |         | ![Image](4V6U.png) |          |
| eL15    | ![Image](4V6U.png) | ![Image](4V6U.png) | ![Image](4V6U.png) | 1-194    |
| Residues| 77-85   | 67-76, 86-95 | 59-66, 96-133 |          |
| eL18    | ![Image](4V6U.png) | ![Image](4V6U.png) | ![Image](4V6U.png) | 1-120    |
| Residues| ![Image](4V6U.png) | ![Image](4V6U.png) | ![Image](4V6U.png) |          |
| eL19    | ![Image](4V6U.png) | ![Image](4V6U.png) | ![Image](4V6U.png) | 1-120    |
| Residues| ![Image](4V6U.png) | ![Image](4V6U.png) | ![Image](4V6U.png) |          |
| eL20    | ![Image](4V6U.png) | ![Image](4V6U.png) | ![Image](4V6U.png) | 1-77     |
| Residues| ![Image](4V6U.png) | ![Image](4V6U.png) | ![Image](4V6U.png) |          |
| eL21 | Residues | 3-13 | 14-97 | 3-97 |
|------|----------|------|-------|------|
| eL24 | Residues | 1-66 | 1-66  |      |
| eL30 | Residues | 1-99 | 1-99  |      |
| eL31 | Residues | 4-93 | 4-93  |      |
| eL32 | Residues | 25-47| 20-24, 48-70 | 3-19, 71-129 | 3-129 |
| eL33 | Residues | 11-21| 1-10, 22-87 | 1-87 |
| eL34 | Residues | 1-29 | 30-89 | 1-89 |
| eL37 |  |  |  |
|------|---|---|---|
| Residues | 1-13 | 14-62 | 1-62 |
| eL39 |  |  |  |
| Residues | 1-51 | 1-51 |  |
| eL40 |  |  |  |
| Residues | 3-47 | 3-47 |  |
| eL42 |  |  |  |
| Residues | 45-55 | 34-44, 56-62 | 1-33, 63-94 | 1-94 |
| eL43 |  |  |  |
| Residues | 6-18 | 19-35 | 36-83 | 6-83 |
Fig. S5. Properties of the full set of rProtein segments in Phases 3-6 of ribosomal evolution. The number of IMHB\textsuperscript{BA} per amino acid is plotted in red and the SSA per amino acid is plotted in blue. IMHB\textsuperscript{BA} increases from Phase 3 through 6, indicating increase in secondary structure. SSA, calculated when rRNA is computationally omitted, decreases from Phase 3 through 6, indicating collapse to globular domains. There are no segments in Phase 1 and 2. rProtein bL20 was excluded from the analysis of \textit{T. thermophilus} and \textit{E. coli}, and eL41 was excluded from the analysis of \textit{P. furiosus}. (\textit{T. thermophilus}, PDB entry 1VY4; \textit{E. coli}, PDB entry 4V9D; \textit{P. furiosus}, PDB entry 4V6U).
Fig. S6. Frequency of secondary structural elements in each Phase. The full set of rProtein segments transition from random coil to isolated secondary elements to globular domains as the phase increases from 3 to 5. Domains convert from predominantly $\beta$-domains in Phase 5 to mixed $\beta$-strand and $\alpha$-helix domains in Phase 6. Isolated helices in Phase 4 are composed of $\alpha$-helices and $3_{10}$-helices of three to seven residues in length. rProtein bL20 was excluded from the analysis of *T. thermophilus* and *E. coli*, and eL41 was excluded from the analysis of *P. furiosus*. (*T. thermophilus*, PDB entry 1VY4; *E. coli*, PDB entry 4V9D; *P. furiosus*, PDB entry 4V6U).
Each ribosome contains both domain-specific and universal proteins (Ban, et al. 2014). Universal rProtein segments are characterized here in figures S3 and S4. The full set of rProtein segments in a given ribosome are characterized in figures S5 and S6. rProtein bL20 is anomalous by several criteria and was omitted from the analysis of *E. coli* and *T. thermophilus*. (i) bL20 is an extended α-helical structure that cannot be segmented without breaking secondary structural elements. (ii) bL20 is the only rProtein that would require cleavage within a secondary structural element to achieve rProtein segments consistent with rRNA Phases. (iii) The differences between the bacterial and archaean/eukaryotic proteins at this location (bL20 versus eL33) are so profound as to suggest relatively recent remodeling of the bacterial protein. Therefore, the accretion model may not apply to bL20.

The eukaryotic rProtein eL41 was excluded from the analysis of *P. furiosus*. (i) This 24-residue long protein is found embedded deeply into the Archean and Eukaryotic Decoding Center of the SSU and makes minimal interactions with the subunit interface of the LSU-SSU. (ii) There is no analogous bacterial protein that occupies the same location in the bacterial ribosome. (iii) Sequence conservation among *P. furiosus, S. cerevisiae, D. melanogaster, and H. sapiens* is very high. (iv) 13 of the 24 residues in *P. furiosus* are either arginine or lysine, and 17 of the 25 residues in *S. cerevisiae, D. melanogaster, and H. sapiens* are arginine or lysine. For these reasons, the accretion model may also not apply to this protein.

eL40 in Archaea and bL36 in Bacteria were assigned to Phase 5. Both of these proteins are found in equivalent positions and interact with loops that terminate helices H89, H91, H97 and H42. These four helices are assigned to Phases 1-5 (Fig. S1) in the rRNA accretion model (Petrov, et al. 2015). H42 is the most recent, belonging to Phase 5, while H89, H91, H97 belong to earlier phases. eL40/bL36 link all these helices and appears dependent on the presence of all four. It is likely that H89, H91 and H97 were linearly extended via evolutionary processes that did not leave fingerprints (Petrov, et al. 2015, SI). The possibility of lengthening is supported by the presence of pivots within H89 at nucleotides U2460-G2490 of 23S *E. coli* (Paci and Fox 2016), as the dynamic elements tend to emerge during late phases of LSU evolution.
References

Ban N, Beckmann R, Cate JH, Dinman JD, Dragon F, Ellis SR, Lafontaine DL, Lindahl L, Liljas A, Lipton JM, Mcalear MA, Moore PB, Noller HF, Ortega J, Panse VG, Ramakrishnan V, Spahn CM, Steitz TA, Tchorzewski M, Tollervey D, Warren AJ, Williamson JR, Wilson D, Yonath A, Yusupov M. 2014. A New System for Naming Ribosomal Proteins. Curr. Opin. Struct. Biol. 24:165-169.

Bernier C, Petrov AS, Waterbury C, Jett J, Li F, Freil LE, Xiong B, Wang L, Le A, Milhouse BL, Hershkovitz E, Grover M, Xue Y, Hsiao C, Bowman JC, Harvey SC, Wartel JZ, Williams LD. 2014. Ribovision: Visualization and Analysis of Ribosomes. Faraday Discuss. 169:195-207.

Paci M, Fox GE. 2016. Centers of motion associated with EF-Tu binding to the ribosome. RNA Biol. 13: 524–530.

Petrov AS, Bernier CR, Hsiao C, Norris AM, Kovacs NA, Waterbury CC, Stepanov VG, Harvey SC, Fox GE, Wartell RM, Hud NV, Williams LD. 2014. Evolution of the Ribosome at Atomic Resolution. Proc. Natl. Acad. Sci. U.S.A. 111:10251-10256.

Petrov AS, Gulen B, Norris AM, Kovacs NA, Bernier CR, Lanier KA, Fox GE, Harvey SC, Wartell RM, Hud NV, Williams LD. 2015. History of the Ribosome and the Origin of Translation. Proc. Natl. Acad. Sci. U.S.A. 112:15396–15401.