Supporting Information for

Evolution of the Ribosome at Atomic Resolution

Anton S. Petrov[a,b,1], Chad R. Bernier[a,b], Chiaolong Hsiao[a,c], Ashlyn M. Norris[a,b], Nicholas A. Kovacs[a,b], Chris C. Waterbury[a,b], Victor G. Stepanov[a,d], Stephen C. Harvey[a,e], George E. Fox[a,d], Roger M. Wartell[a,e], Nicholas V. Hud[a,b] and Loren Dean Williams[a,b,1]

[a]*Center for Ribosomal Origins and Evolution, Georgia Institute of Technology, Atlanta, GA, 30332, USA.*

[b]*School of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta, GA, 30332, USA.*

[c]*Institute of Biochemical Sciences, National Taiwan University, Taipei, 10617, Taiwan.*

[d]*Department of Biology and Biochemistry, University of Houston, Houston, TX, 77204, USA.*

[e]*School of Biology, Georgia Institute of Technology, Atlanta, GA, 30332, USA.*

Corresponding authors:

Name: Anton S. Petrov, Loren Dean Williams

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

Phone: 404-894-9752

Fax: 404-894-7452

Emails: anton.petrov@biology.gatech.edu, loren.williams@chemistry.gatech.edu

Key words: RNA evolution, RNA structure, translation, origin of life, phylogeny

Contents

Materials and Methods

Alignments and phylogenetic trees. We aligned LSU rRNA sequences from 135 organisms intended to represent, as far as available data allow, the broadest sparse sampling of the phylogenetic tree of life, including all three domains of life. Aligned sequences are derived from thirty-two eukaryotic species, sixty-seven bacterial species, and thirty-six archaeal species. The aligned sequences are mostly complete and with low sequence error frequencies, as indicated by cross-validation. Sequences were compiled, aligned, and validated in an extensive iterative process. Initial automated alignments generated by SILVA-ARB (1) were inspected and manually adjusted. Incomplete sequences were manually patched with missing fragments located with BLASTN (2). The alignment is provided in FASTA format (DataSet S2). The phylogenetic tree was generated from sTOL (3) and graphed using the ITOL (4) web server. The sizes of incomplete sequences were estimated with high confidence using the alignment by Mallot (5).

Secondary structures. The historical *E. coli* secondary structures of LSU and SSU rRNAs were downloaded from http://rna.ucsc.edu/rnacenter/ribosome_images.html, revised and extended with the program XRNA (rna.ucsc.edu/rnacenter/xrna/xrna.html), and finalized with Adobe Illustrator as described (6, 7). Secondary structures of all other species presented here were built from the *E. coli* template. Secondary structures were derived by applying geometric and molecular interaction criteria derived from the corresponding three-dimensional structures of ribosomes. We use historical representations as far as possible except where conflicts arise with correct helical assignments. Mapping of information on to secondary structures was performed with the webapp RiboVision (apollo.chemistry.gatech.edu/RiboVision) (8).

Three-dimensional structures. Three-dimensional structures of ribosomal particles were obtained from the PDB database. The x-ray structure of Steitz (9) was used for *H. marismortui* (PDB entry 1JJ2, resolution 2.4 Å). The x-ray structure of Cate (10) was used for *E. coli* (PDB entries 3R8S, 4GD1, resolution 3.0 Å), and the x-ray structure of Yusupov (11) was used for *S. cerevisiae* (PDB entries 3U5B, 3U5C, 3U5D, 3U5E, resolution 3 Å). Ribosomes of *D. melanogaster* and *H. sapiens* are the cryo-EM structures of Beckmann (12) (PDB entries (3J38, 3J3C, 3J39, 3J3E for *D. melanogaster*, resolution 6 Å; PDB entries 3J3A, 3J3B, 3J3D, 3J3F, resolution 5 Å for *H. sapiens*). Initial visualization and inspection of the three-dimensional structures, and final visualization and image preparation were performed with PyMOL (13).

Analysis of the three-dimensional representations of the expansion segment structures and their superposition for different species were also performed using PyMOL.

Computational ligation of AESs in three-dimensions. Ancestral fragments AES 1-5, were excised by extracting corresponding nucleotides (as defined in Table 1 of the manuscript) from the crystal structure of the *E. coli* LSU (PDB entry 3R8S) (10). Branching segments as described in Figure 5 were excised, and the corresponding trunk segments were sealed at the positions of excised branches by adding the 03'-P bond. To seal the nick, four nucleotides (two from each side of the new bond) were subjected to a partial minimization, while the remaining nucleotides remained intact.

Energy Minimization. Partial minimization of the re-ligated rRNAs was performed with Sybyl-X 1.2 software (Tripos International, St. Louis, MO, USA) with the AMBER FF99 force field (14) using an implicit solvent model with the distance dependent dielectric function D(r) = 20r. The non-bonded cut-off distance was set to 12 Å. Each system was minimized by 1000 steps of steepest decent followed by 5000 steps of conjugate gradient minimization.

A model of LSU rRNA evolution by the accretion of Ancestral Expansion Segments

Criteria for Identifying Eukaryotic Insertion Fingerprints. Eukaryotic expansion segments, shown in Table S2, were identified by comparison of common core rRNA to eukaryotic rRNA. Insertion fingerprints are evident upon inspection of the structures in three-dimensions. We observe that rRNA helices of trunk segments show minimal distortion at branch sites. Bases are paired and stacked in a trunk helix on either side of the branch point, and the trunk helical axis is linear. The helical axis of the branch segment deviates acutely from that of the trunk helix. On either side of the branch point, the sugar and phosphate moieties of the trunk helix are in close proximity to each other. Branch segments can be excised and the trunk segments can be computationally sealed with a minimal perturbation of the nucleotides near the insertion site. Some eukaryotic expansion segments were augmented by extension of an existing helix, rather than by insertion of a branch segment. Elongations were identified by superposition of the corresponding bacterial and eukaryotic helical elements. One can infer that the ordering of expansions is from inside the LSU to outside, with new segments added to surfaces rather than to interior spaces.

Criteria for Identifying Ancestral Expansion Segments. Insertion fingerprints, with structural characteristics similar to those described above for eukaryotic expansions, are observed throughout the ribosomal common core. These ancestral insertion fingerprints allow us to infer ancestral expansion segments (AESs, Table S3), which appear to build up the common core. Continuity of the sugar-phosphate backbone, stacking interactions, and helical axis was used as additional criteria to define AESs. Our model probably identifies an under-sampling of the actual ancestral expansion events. In our model, the majority of the AESs were added instantaneously in their present size without elongation. In reality, many AESs probably budded and increased in size over time. However, helical elongation does not leave fingerprints and so is not included in the model. In a few instances, we could tentatively identify helix elongations either by irregularities in a helical structure or by sharp turns of helical axes. In these cases, the elongated portion of an AES was assigned the same number as their founding AES with addition of the suffix ”a”.

Temporal Ordering of AES Addition. We use insertion fingerprints, structural dependencies, and functional data to suggest the portion of the LSU rRNA that forms the ultimate primordial ancestor of all other LSU rRNA. Identification of this element allows us to perform a recapitulation process for growth of the entire LSU rRNA, from small to the common core.

The chronology of the addition of the AESs to the LSU rRNA is based on:

a) insertion fingerprints- trunks are older than branches (a branch is not added until after its trunk has formed);

b) inside to outside growth, with new segments added to surfaces rather than to interior spaces,

c) structural dependencies - the directionality of A-minor interactions is dependent to independent, young to old [see Steinberg (15)],

d) on functionality of the expansion segments, with the machinery required for translocation such as EF-G, and EF-Tu binding sites, L7/L12 stalk, and the central protuberance appearing at some stages of ribosomal development as suggested by the minimal model (16). The segments are numbered in the order of their appearance in Table S3. The numbers and the exact temporal order of peripheral AESs are indeterminate, as they evolved independently one another. Their

ordering cannot be inferred from the rules listed above, and numbers of these AESs are somewhat arbitrary. This ambiguity does not affect conclusions of the study.

Evolutionary Model of the LSU rRNA.

Using a proposed temporal ordering of AESs we partitioned the evolution of LSU into phases. In this model, phases can be associated with the emergence of various functions. The detailed boundaries between phases are not rigorously defined and are not overly significant.

- *Phase 1*) Rudimentary Binding and Catalysis. AES 2 is inserted into AES 1. These two segments form the P-region. The combined AES 1-2 fragment may have catalytic activity (17) and/or an ability to bind to ions and small molecules (Figure S11A).
- *Phase 2*) Mature PTC and Exit Pore. Insertion of AES 3-5 adds the A-region to the P-region, in concert with formation of an exit pore (18). The exit pore will ultimately evolve into the exit tunnel. AES 4 consists of two portions: AES 4 (Helices 73 and 93) and AES 4a (Helix 1). AES 4a extends AES 4 (Figure S11B).
- *Phase 3*) Early Tunnel Extension. Inclusion of AES 6-10 transforms the exit pore into a short tunnel. AES6 (Helix 72 and Helix 2) as well as AES 7 (Helices 94 and 95) are inserted at the junction of AES 4 and AES 4a and give rise to Domains I and VI. AES 8 (Helix 25a) is inserted into AES 6. Helices 26, 26a, 32, 33, and 61 comprise AES 9, which is inserted into AES 8. AES 9 form a circular loop, which binds to the PTC around the exit pore and extends it to a short exit tunnel. A portion of Helix 34 and Helix 35 form AES 10, which locks the circular AES 9 in place (Figure S11C).
- Phase 4) Acquisition of the SSU Interface with Tunnel Extension. During this phase, AES 11-28 are acquired. Most of Domain IV (AES 11, 12, 15), responsible for association with the SSU, is added. Portions of Domains II and V (AESs 13, 14, 16, 17, 18, 20, 22, 23, 25, and 27) stabilize the PTC. Domains I (AESs 6a, 19, 21, 24, 26) and Domain III (AESs 28) stabilize the PTC and extend the exit tunnel (Figure S11D).
- Phase 5) Acquisition of Translocation Function with Tunnel Extension. Inclusion of AES 29-39 adds essential components of the modern energy-driven translational machinery (the

L7/L12 stalk, AESs 32, 32a, 38), including the central protuberance (AES 34, 36) (19, 20). AES39 is added to form a portion of the L1 stalk. The binding site for Elongation Factor G and Elongations Factor TU are acquired (sarcin-ricin loop, AES 30) (19, 20), and the SSU interface is increased by adding AES 29, 12a, 15a, 37 to Domain IV, and AES 10a to Domain II. The exit tunnel is further elongated (AES 31, 33, 35) (Figure S11E).

- Phase 6) Tunnel Extension. Accretion of AES 40-59 results in the maturation of the common core of the LSU. AESs 43 and 44 (together with AES 39) form the L1 stalk, responsible for translocation of tRNA to the E site, and the central protuberance (AESs 40, 46). AESs 41, 42, 48, 52, 53, 55 and 59 complete Domain I, and AESs 47, 49, 51, 56 complete Domain III and finalize the exit tunnel. Domains II and VI are completed upon addition of AESs 10a, 45, 50, 54, and 58 (Figure S11F).

Among the last ancestral segments to appear are AESs 41, 47, 49, 50, 52, 53, 54, 57, and 58. They are all located at the LSU surface and later evolve into eukaryotic Expansion Segments, making a smooth transition from pre to post LUCA evolution.

The expansion of the rRNA is also accompanied by the congruent development of the ribosomal proteins, which are bound to the rRNA and grow inside out. The evolution of rProteins is beyond the scope of this work, and therefore, not discussed here (Figure S11A).

Comparison with Previous Models. The proposed model is consistent with, and extends, a previous model by Fox (21). On a finer scale, some AESs are similar to previous elements of rRNA growth as proposed by Steinberg (15). Specifically, AESs 7, 14, 15, 16, 17, 19, 21, 22, 26, 32, 34, 37, 40, 49, 50, 51, 53, 54, and 57 are essentially identical to Steinberg elements. However, our model differs from Steinberg's in several ways. Our model i) addresses the origin of the PTC; ii) is constructive (it builds up the rRNA), as opposed to Steinberg's model, which peels off segments located on the rRNA surface regardless of their function; iii) treats portions of Domains II, III, IV, and V on a fine-grained level, revealing additional growth elements, iv) suggests that Helices 1, 73, and 93 formed a single AES 4 (with AES 4a) that emerged from the PTC at an early stage of evolution, and from which Domains I and VI evolved by insertions; v) suggests that Helices 2, 3, 4, and 72 evolved as a single expanding AES 6 (with AES 6a) inserted

into AES 4, vi) identifies Helices 26, 26a, 32, 33, and 61 as a single evolving segment AES 9 that binds to the PTC around the exit pore forming the ancestral exit tunnel, and from which Domains II, III, and VI evolved. The model describes the evolution of the LSU as a process of rRNA expansion acquisition of new rRNA segments that in isolation form thermodynamically stable elements. We also note that some expansion segments at the ribosomal surface may get shorter or disappear as some (families of) species evolve. rRNA buried under the surface remains the same in all species across the tree of life.

Special comments on some distorted AESs. Most AESs were defined by the insertion fingerprints as described above. This method allows easy identification of insertions but not elongations. The model treats the addition of each AES as an instantaneous event; even if it contains multiple helices. For example, AES 9 is formed by Helices 26, 26a, 32, 33, and 61. However, in a few cases, if a helix bends sharply, it has been divided into two segments: the main segment an appendage, denoted by suffix "a" (e.g. AES 12, and AES 12a). The appendage is assigned the same number because the criterion of the helix continuity between the main segment and appendage is met. The appendages may appear after the founding AES, if this is supported by structural constraints such as A-minor interactions.

AES 4 (Helices 93 and 73) appears to be extended by AES 4a (Helix 1). The elongation of AES 4 by AES 4a is considered a separate event from the insertion of AES 4 into AES 3. AES 4 and AES 4a represent a continuous helix, which is disrupted and slightly twisted by insertion of AESs 6 and 7. The insertions of AESs 6 and 7 are proposed to be followed by elongation of AES 4 with AES 4a, explaining the direct proximity of 3' and 5' ends of the 23S rRNA.

We infer that AES 6, which is inserted between AES 4 and 4a in Phase 3, bends and is extended by addition of AES 6a (Portion of Helix 2, Helices 3 and 4). In our model, AES 6a is added later in Phase 4, indicated by A-minor interactions with AES 9 and AES 14.

Helices 34 and 35 were partitioned into AES 10 and AES 10a based on structural and functional arguments. While AES 10 strongly interacts with AES 9 stabilizing the exit tunnel during Phase 3, AES 10a is involved in interactions with the SSU, and expected to appear during the late stages of LSU evolution (phase 5). A bulge in the middle Helix 34 provides a tentative boundary between these elements.

Helices 65 and 66 are partitioned into segments AES 12 and 12a because AES 12a sharply bends away from AES 12, and interacts with portions of Domain III (AES 35) that were added during later stages of evolution.

Helices 68 and 71 form a continuous helix that is divided here into AES 15 and AES 15a. Dependency requirements suggest that AES 15a follows AES 27, in Phase 5, while AES 15 precedes AES 27. AES 15 forms part of the interfacial region between two subunits, which is established at the beginning of Phase 4.

Helices 42 and 44 are partitioned into AES 32 and 32a because they are separated by a sharp helical turn of nearly 90°.

References.

1. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO (2013) The Silva Ribosomal RNA Gene Database Project: Improved Data Processing and Web-Based Tools. *Nucleic Acids Res* 41:D590-D596.
2. Madden TL, Tatusov RL, Zhang J (1996) Applications of Network Blast Server. *Methods Enzymol* 266:131-141.
3. Gough J, Karplus K, Hughey R, Chothia C (2001) Assignment of Homology to Genome Sequences Using a Library of Hidden Markov Models That Represent All Proteins of Known Structure. *J Mol Biol* 313:903-919.
4. Letunic I, Bork P (2011) Interactive Tree of Life V2: Online Annotation and Display of Phylogenetic Trees Made Easy. *Nucleic Acids Res* 39:W475-478.
5. Mallatt J, Chittenden KD (2014) The Gc Content of Lsu rRNA Evolves across Topological and Functional Regions of the Ribosome in All Three Domains of Life. *Mol Phylogen Evol* 72:17-30.
6. Petrov AS, Bernier CR, Hershkovitz E, Xue Y, Waterbury CC, Grover MA, C. HS, Hud NV, Wartell RM, Williams LD (2013) Secondary Structure and Domain Architecture of the 23S rRNA. *Nucleic Acids Res* 41:7522-7535.
7. Petrov AS, Bernier CR, Gulen B, Waterbury CC, Hershkovitz E, Hsiao C, Harvey SC, Hud NV, Fox GE, Wartell RM, Williams LD (2014) Secondary Structures of rRNAs from All Three Domains of Life. *PLoS One* 9:e88222.
8. 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 Discussions. (in press). DOI: 10.1039/C3FD00126A
9. Ban N, Nissen P, Hansen J, Moore PB, Steitz TA (2000) The Complete Atomic Structure of the Large Ribosomal Subunit at 2.4 Å Resolution. *Science* 289:905-920.
10. Dunkle JA, Wang LY, Feldman MB, Pulk A, Chen VB, Kapral GJ, Noeske J, Richardson JS, Blanchard SC, Cate JHD (2011) Structures of the Bacterial Ribosome in Classical and Hybrid States of tRNA Binding. *Science* 332:981-984.

11. Ben-Shem A, de Loubresse NG, Melnikov S, Jenner L, Yusupova G, Yusupov M (2011) The Structure of the Eukaryotic Ribosome at 3.0 Å Resolution. *Science* 334:1524-1529.
12. Anger AM, Armache JP, Berninghausen O, Habeck M, Subklewe M, Wilson DN, Beckmann R (2013) Structures of the Human and Drosophila 80S Ribosome. *Nature* 497:80-85.
13. Schrodinger L (2010) The Pymol Molecular Graphics System, Version 1.3r1.
14. Wang JM, Cieplak P, Kollman PA (2000) How Well Does a Restrained Electrostatic Potential (RESP) Model Perform in Calculating Conformational Energies of Organic and Biological Molecules? *J Comput Chem* 21:1049-1074.
15. Bokov K, Steinberg SV (2009) A Hierarchical Model for Evolution of 23S Ribosomal RNA. *Nature* 457:977-980.
16. Mears JA, Cannone JJ, Stagg SM, Gutell RR, Agrawal RK, Harvey SC (2002) Modeling a Minimal Ribosome Based on Comparative Sequence Analysis. *J Mol Biol* 321:215-234.
17. Wolf YI, Koonin EV (2007) On the Origin of the Translation System and the Genetic Code in the RNA World by Means of Natural Selection, Exaptation, and Subfunctionalization. *Biol Direct* 2:14.
18. Fox GE, Tran Q, Yonath A (2012) An Exit Cavity Was Crucial to the Polymerase Activity of the Early Ribosome. *Astrobiology* 12:57-60.
19. Valle M, Zavialov A, Sengupta J, Rawat U, Ehrenberg M, Frank J (2003) Locking and Unlocking of Ribosomal Motions. *Cell* 114:123-134.
20. Lancaster L, Lambert NJ, Maklan EJ, Horan LH, Noller HF (2008) The Sarcin-Ricin Loop of 23S rRNA Is Essential for Assembly of the Functional Core of the 50S Ribosomal Subunit. *RNA* 14:1999-2012.
21. Fox GE (2010) Origin and Evolution of the Ribosome. *Cold Spring Harb Perspect Biol* 2:a003483.

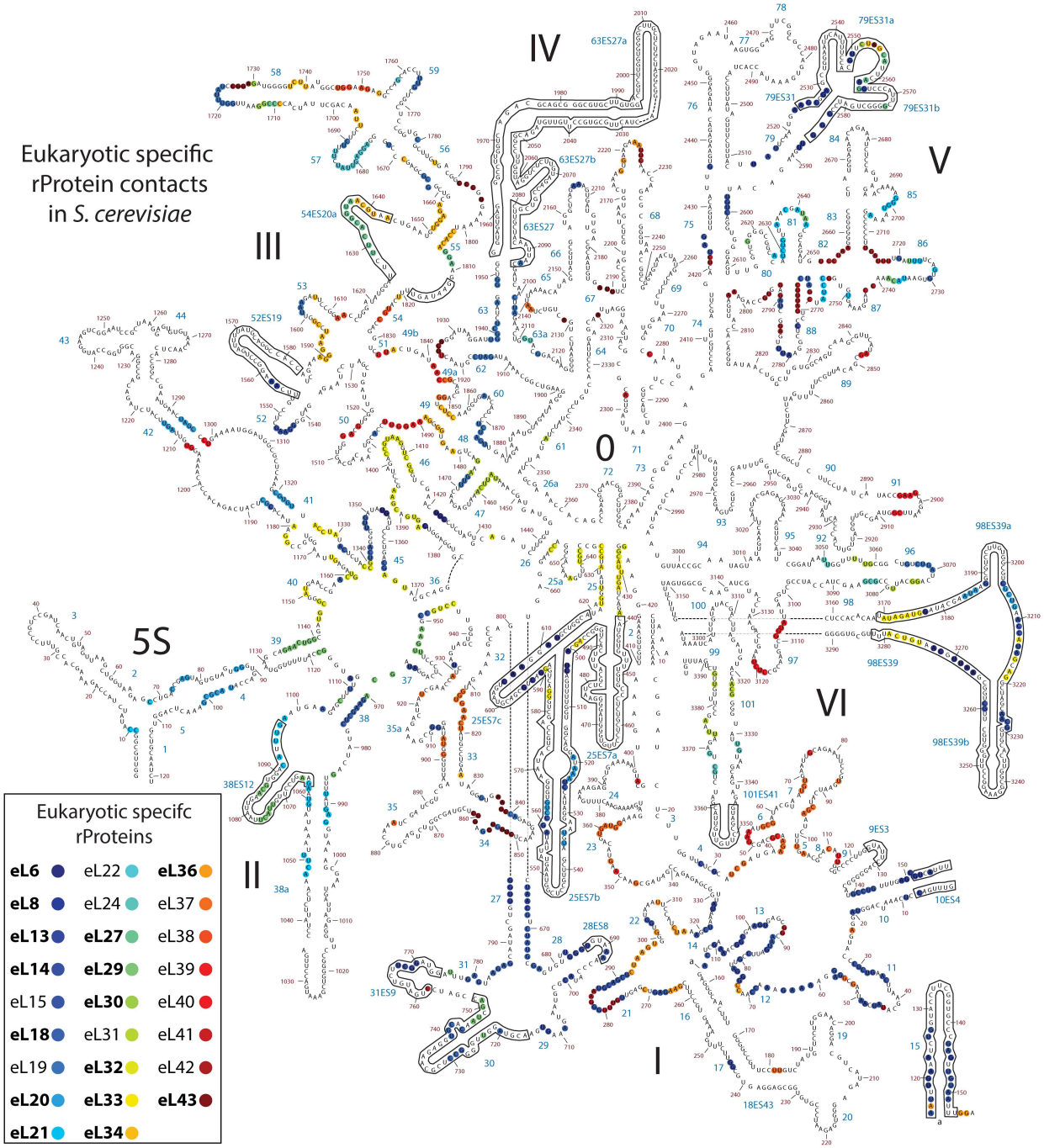


Figure S1. Eukaryotic-specific ribosomal protein contacts mapped onto the secondary structure of *S. cerevisiae* LSU rRNA. Ribosomal protein contacts within a cutoff distance of 3.4 Å are indicated. This figure was prepared using the webapp RiboVision (8) and elaborated in Adobe Illustrator. Eukaryotic expansion segments are embodied by black outlines. Ribosomal proteins that contact eukaryotic expansion segments are bold in the legend. A list of rRNA nucleotides, their contacts with eukaryotic-specific ribosomal proteins, and their presence in expansion segments are provided in DataSet S1.

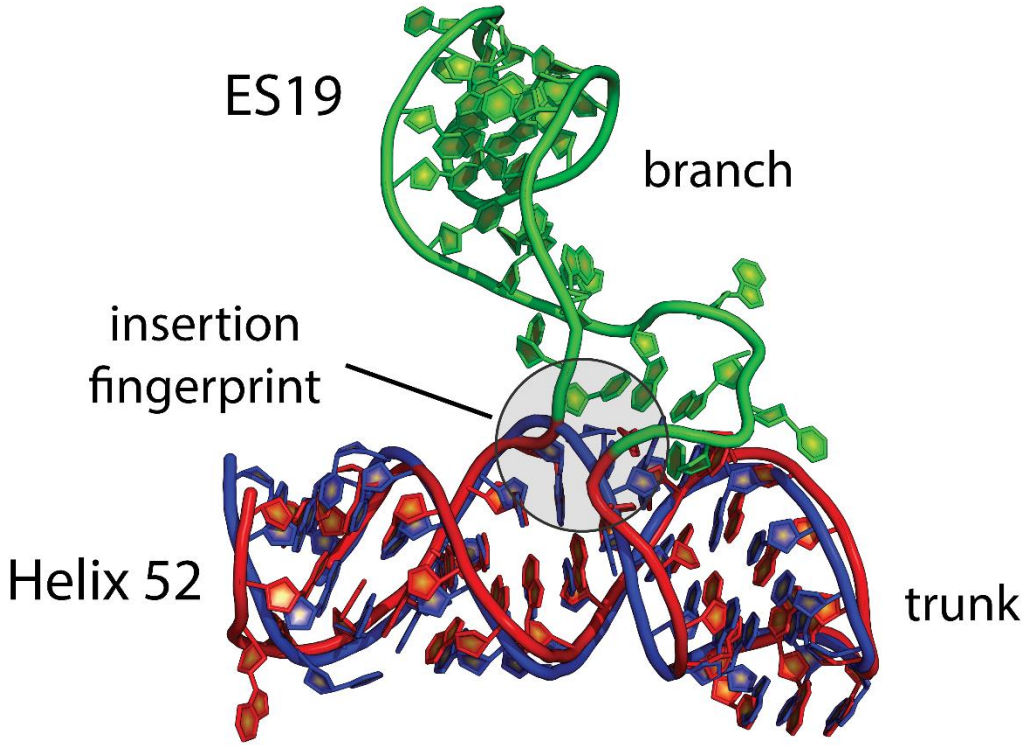


Figure S2. Helix 52 is expanded by insertion. The common core rRNA is blue for *E. coli* and red for *S. cerevisiae*. The expansion segment of *S. cerevisiae* rRNA is green. The insertion fingerprint is highlighted by a gray circle.

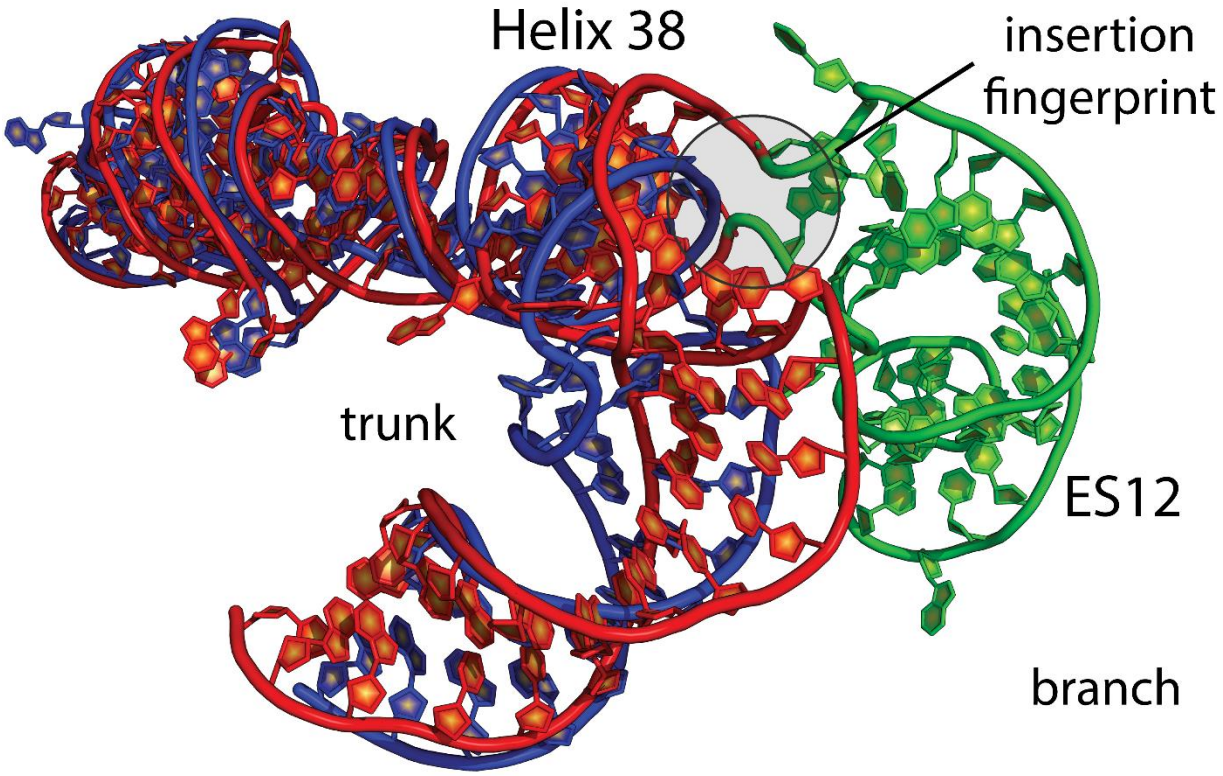


Figure S3. Helix 38 is expanded by insertion. The common core rRNA is blue for *E. coli* and red for *S. cerevisiae*. The expansion segment of *S. cerevisiae* rRNA is green. The insertion fingerprint is highlighted by a gray circle.

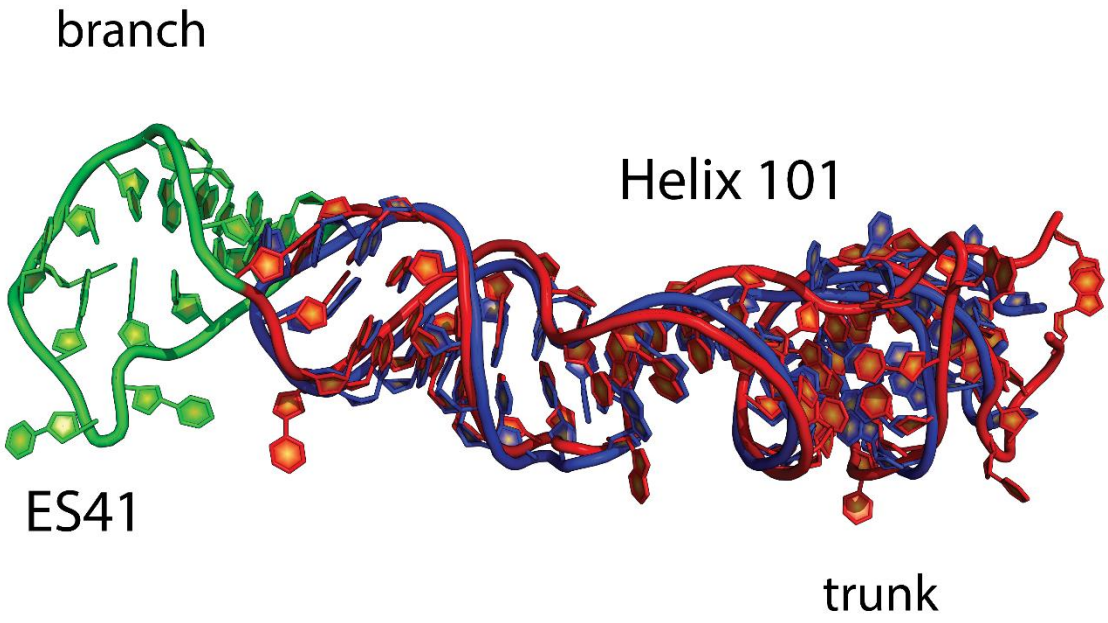


Figure S4. Helix 101 is expanded by extension. The common core rRNA is blue for *E. coli* and red for *S. cerevisiae*. The expansion segment of *S. cerevisiae* rRNA is green.

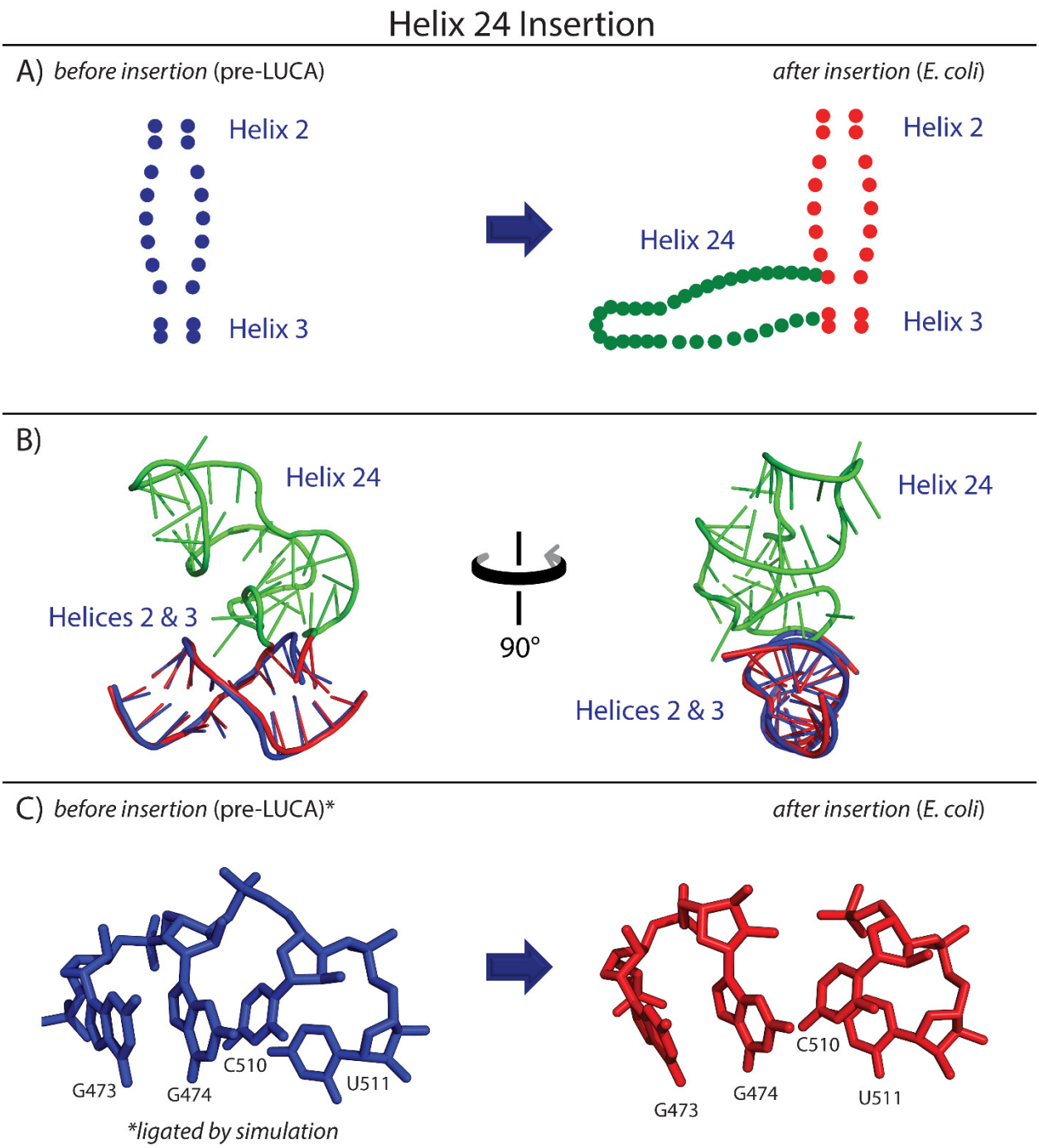


Figure S5. Expansion by helix insertion in the common rRNA core. Helices 2-3 (trunk) are expanded by insertion of Helix 24 (branch). A) Secondary structures of the trunk and branch fragments. B) 3D structures of the trunk and branch fragments. C) Atomic resolution representation of the insertion site. The pre-insertion state (blue) was modeled by computational ligation. Inserted branch is green and post-inserted trunk is red. The insertion process, moving forwards in time, is symbolized by blue arrows.

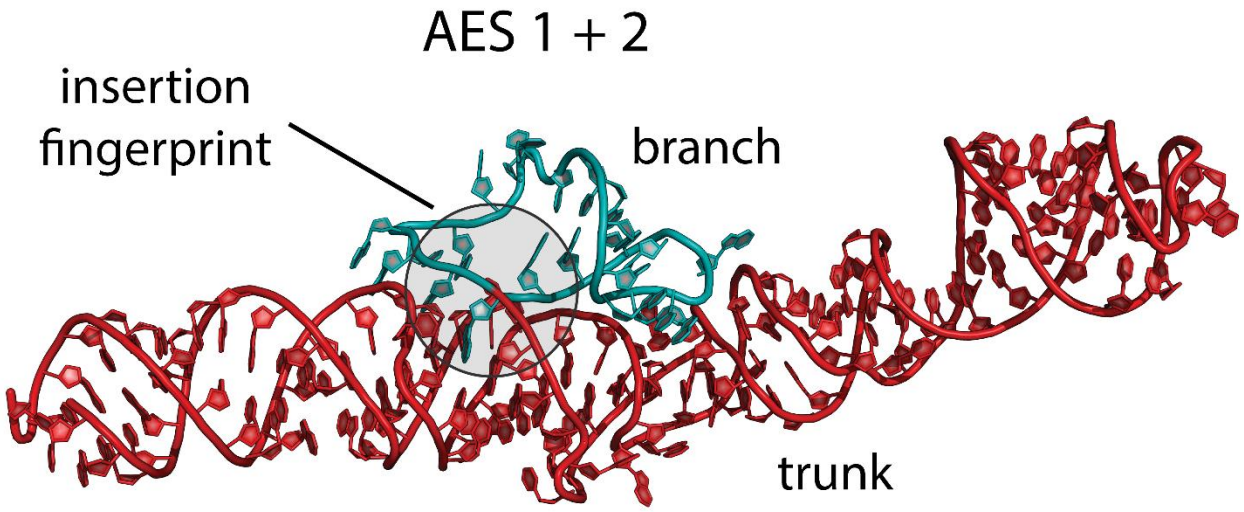


Figure S6. AES2 is inserted into AES1. AES1 is red and AES 2 is teal. The insertion fingerprint is highlighted by a gray circle.

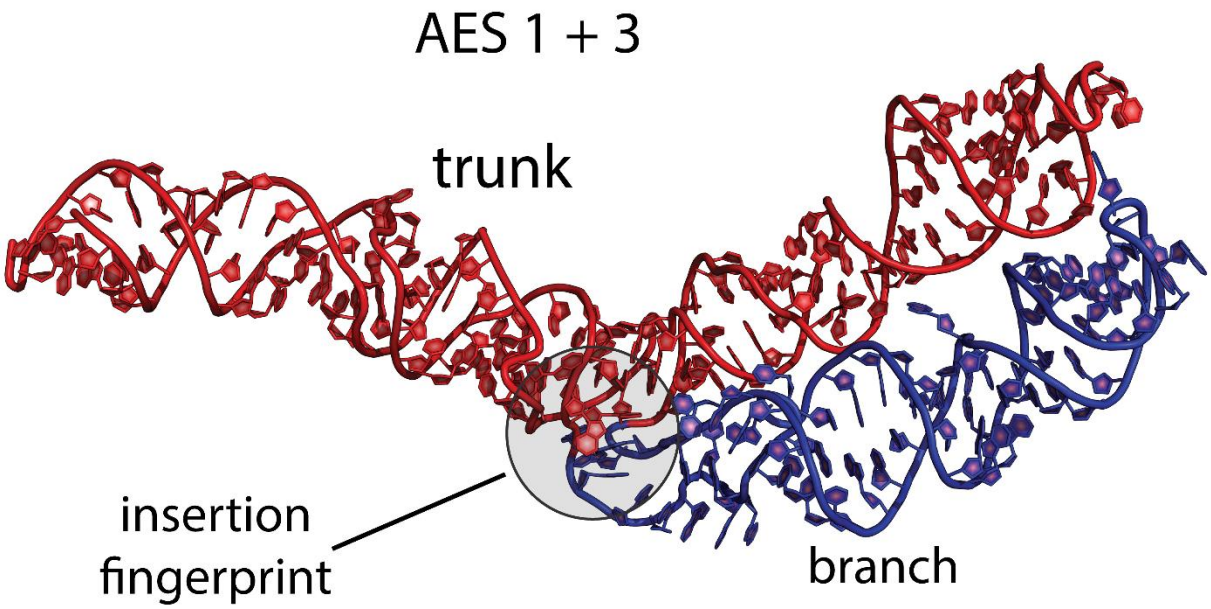


Figure S7. AES3 is inserted into AES1. AES1 is red and AES3 is blue. The insertion fingerprint is highlighted by a gray circle.

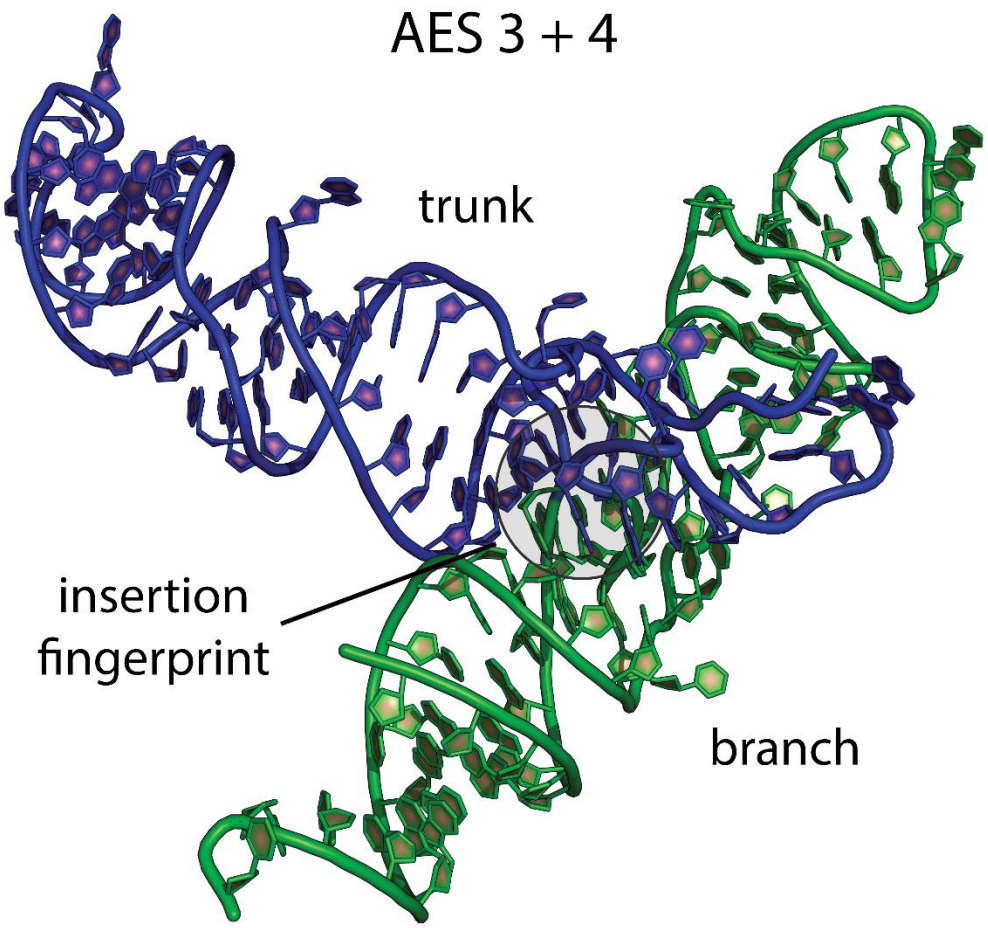


Figure S8. AES4 is inserted into AES3. AES3 is blue and AES4 is green. The insertion fingerprint is highlighted by a gray circle.

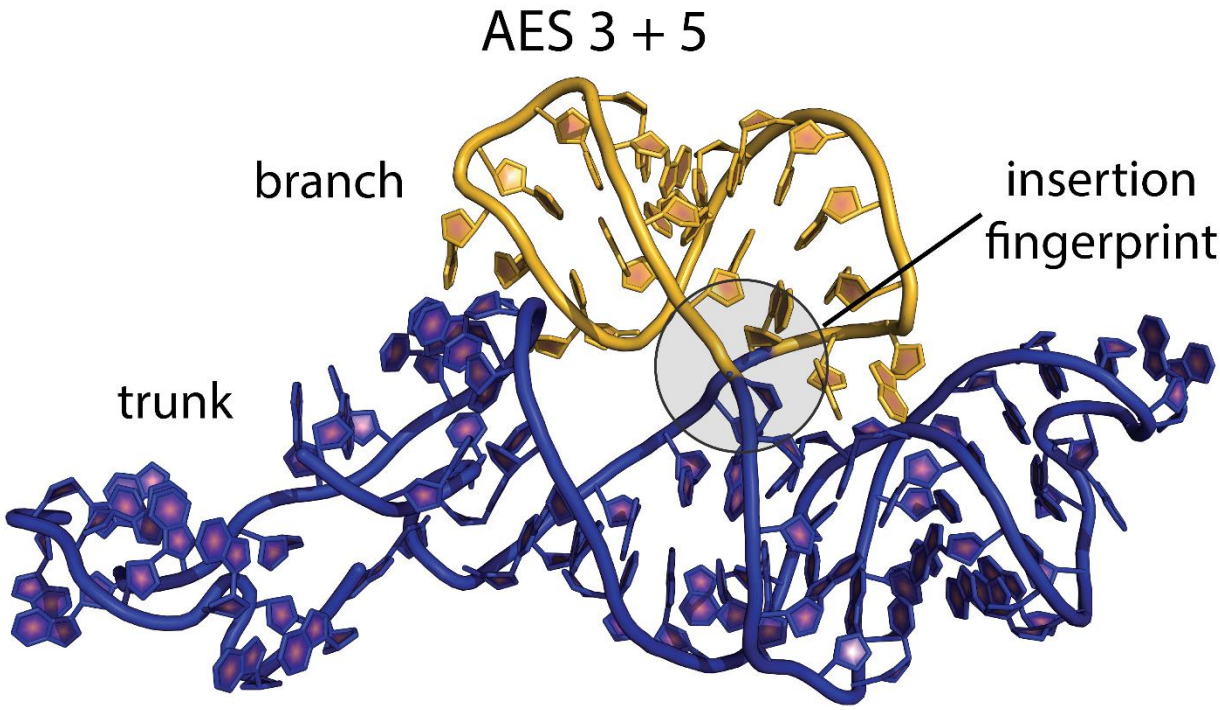


Figure S9. AES5 is inserted into AES3. AES3 is blue and AES5 is yellow. The insertion fingerprint is highlighted by a gray circle.

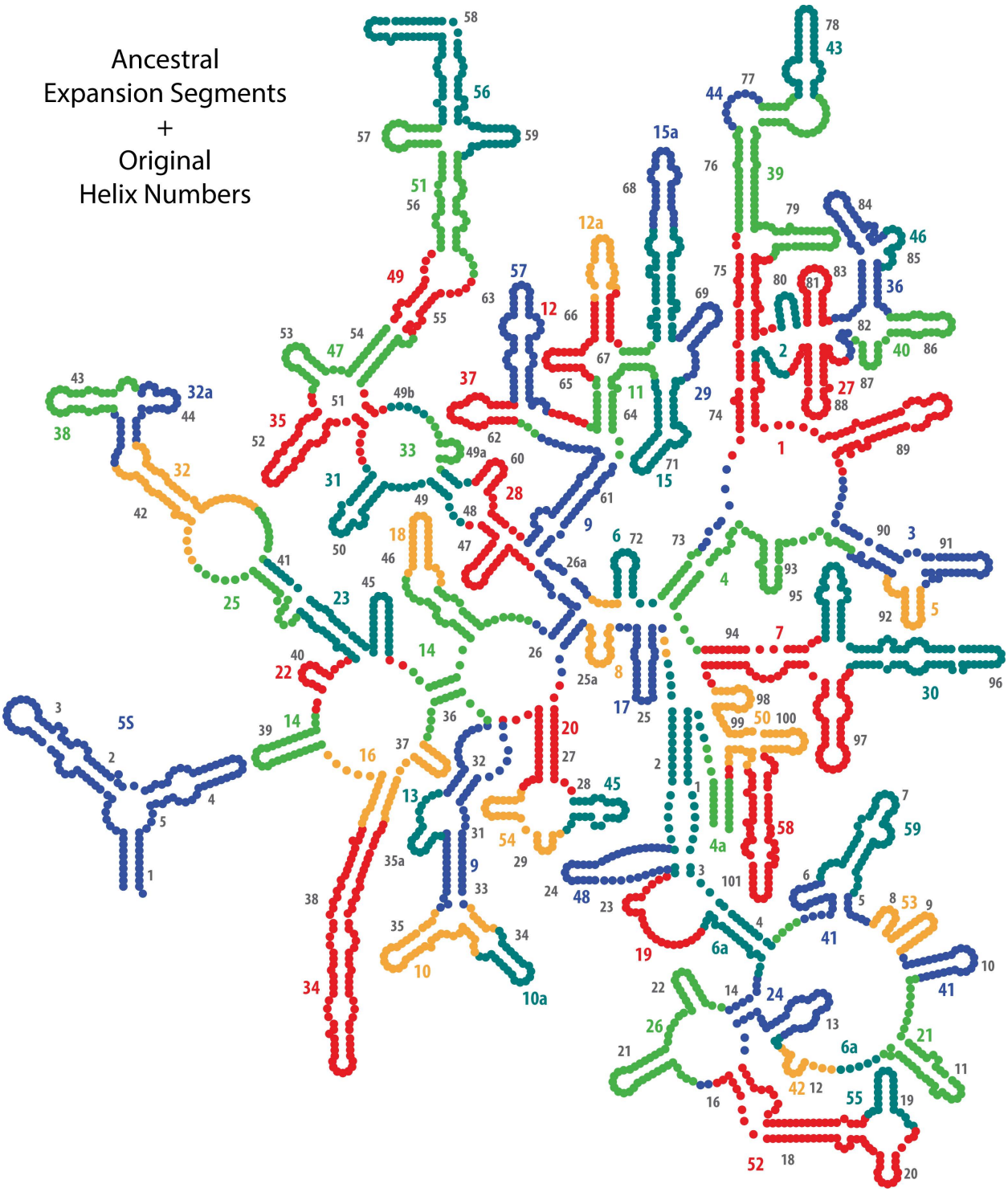


Figure S10. rRNA evolution mapped onto the LSU rRNA secondary structure of *E. coli*. The common core is built up in phases, by stepwise addition of ancestral expansion segments (AESs) at sites marked by insertion fingerprints. Each AES is individually colored and labeled by temporal number. AES colors are arbitrary, chosen to distinguish the expansions, such that no AES is the color of its neighbor. Original helix numbers are provided in gray for reference.

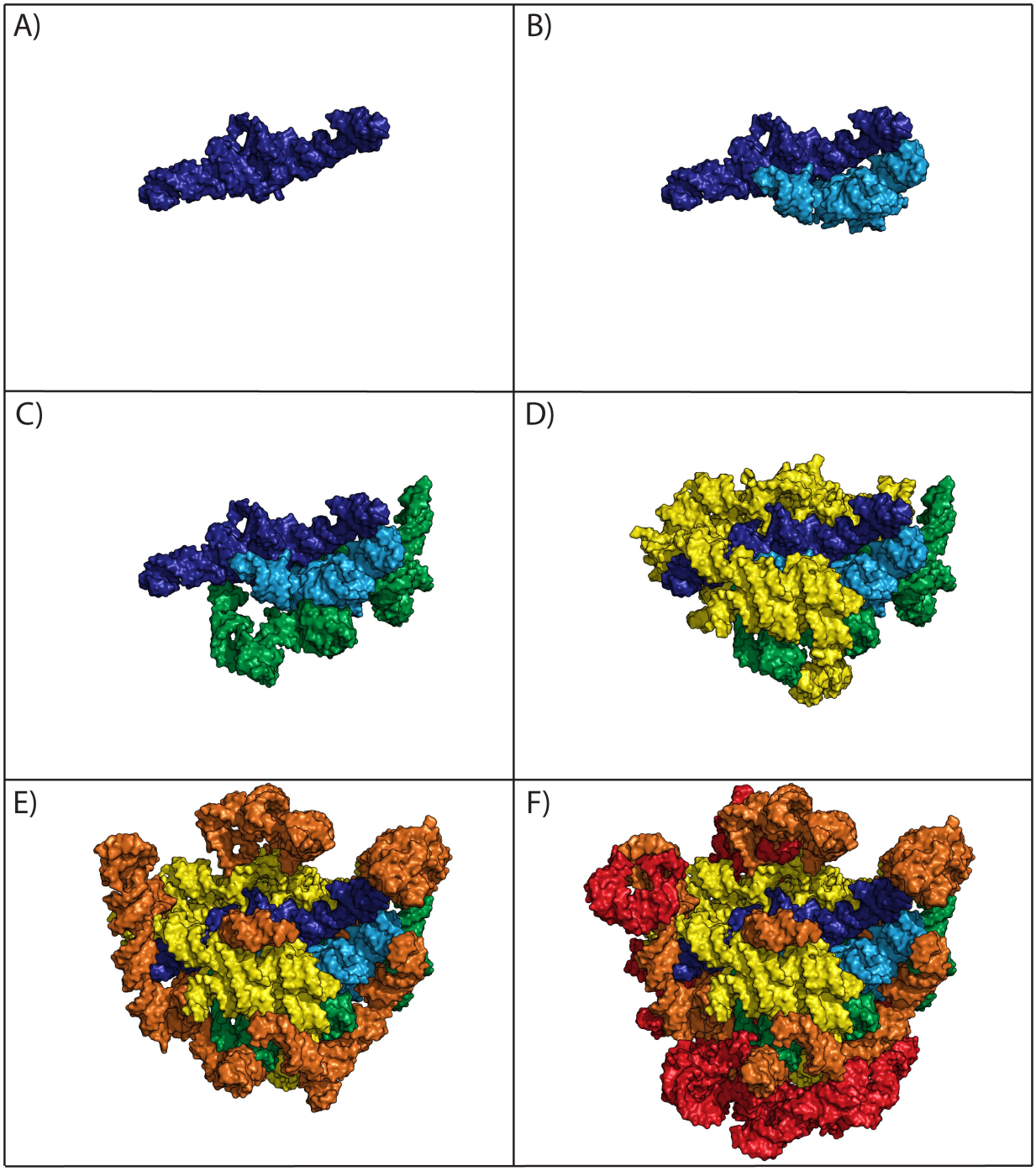


Figure S11. rRNA evolution mapped onto the 23S rRNA 3D structure. The common core is built up in phases, by stepwise addition of ancestral expansion segments (AESs) at sites marked by insertion fingerprints. Accretion of ancestral and eukaryotic expansion segments is distributed into eight phases, associated with ribosomal functions: Phase 1) Rudimentary Binding and Catalysis, dark blue; Phase 2) Maturation of the PTC and Exit Pore, light blue; Phase 3) Early Tunnel Extension, green; Phase 4) Acquisition of the SSU Interface, yellow; Phase 5) Acquisition of Translocation Function, orange; Phase 6) Late Tunnel Extension, red. The correspondence between AES and Phase is given in Table S3.

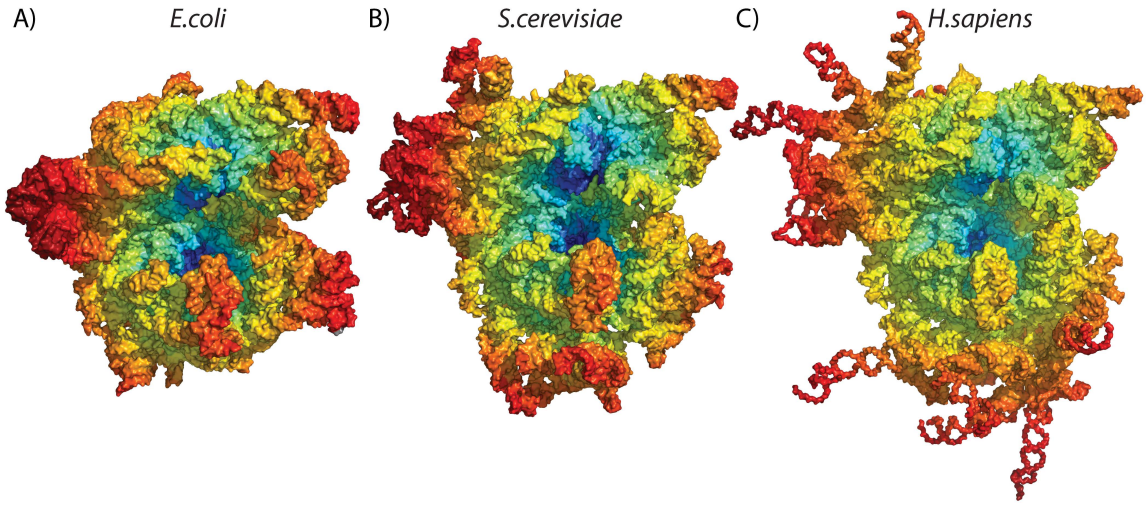
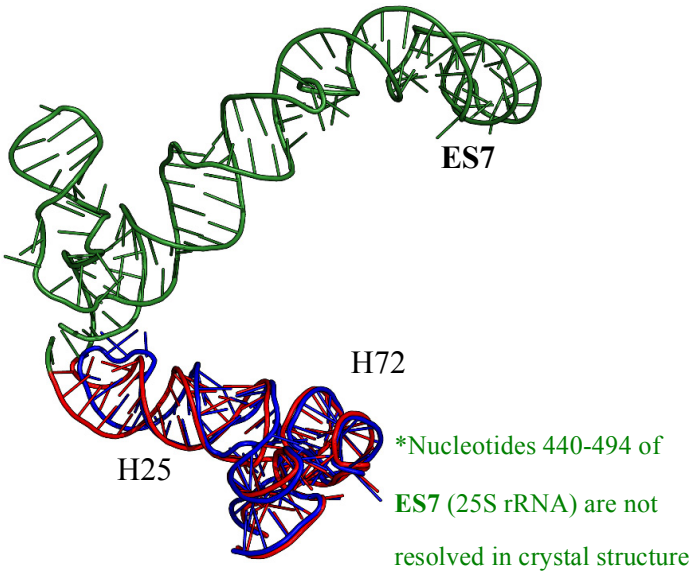
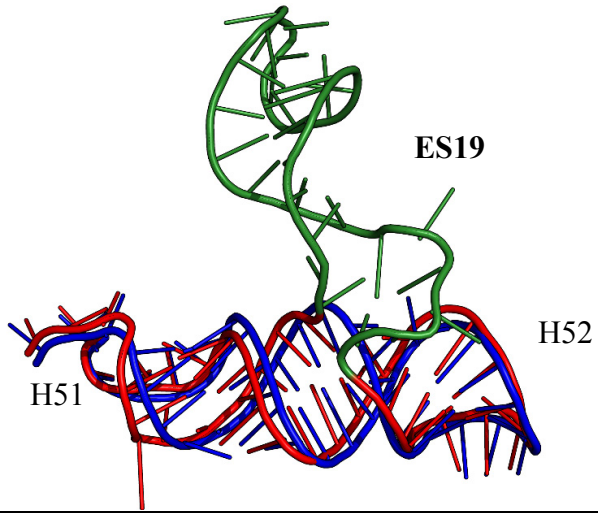


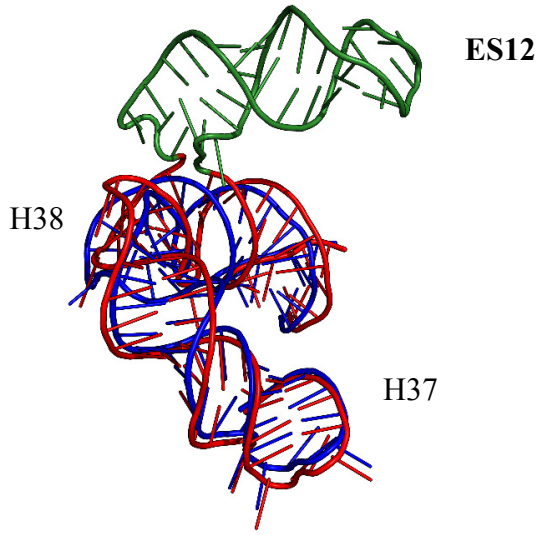
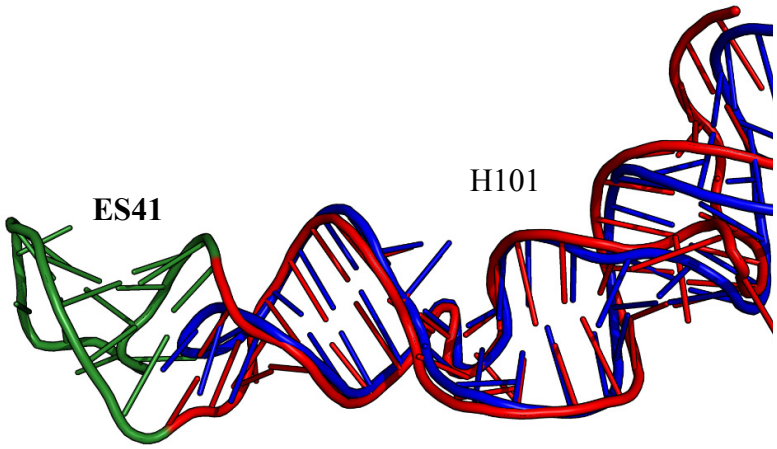
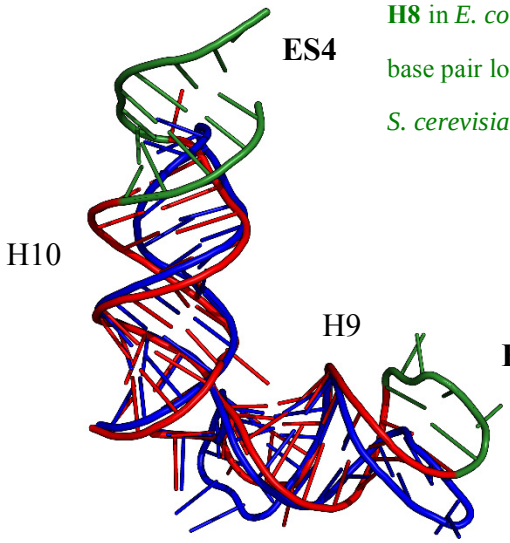
Figure S12. LSU and SSU rRNA structures in three-dimensions. A) *E. coli* represening the common core (phases 1-6). B) *S. cerevisiae* showing encasement of the common core in simple eukaryotes (phases 1-7). C) *H. sapiens* demonstrating surface elaboration in complex eukaryotes (phases 1-8). The color indicates the proximity in three-dimensions to central site of function for each subunit: the site of peptidyl transfer for the LSU (bottom half), and the site of the decoding center (tRNA recognition) for the SSU (top half). Dark blue is closest to these sites and dark red is furthest. This figure corresponds with Figure 2 in the main paper.

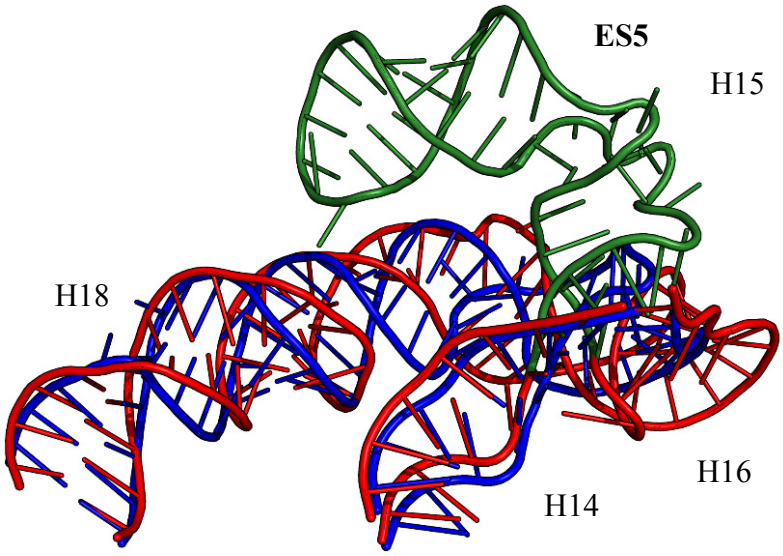
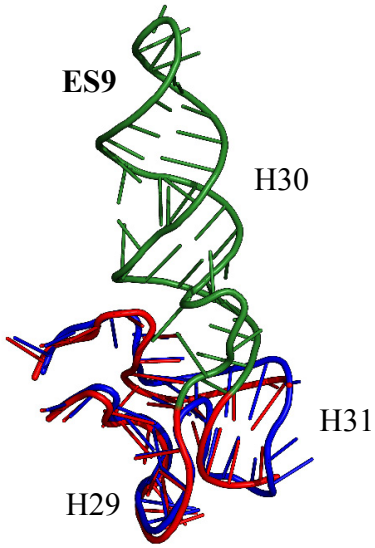
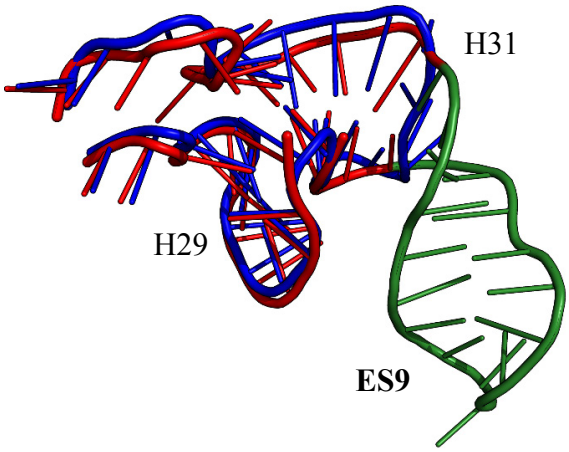
Table S1. LSU, SSU, and genome length values for species used in this study. Archaea and Bacteria are reported as averages (standard deviation). See the alignment file (DataSet S2) for the complete list of the bacterial and archaeal species.

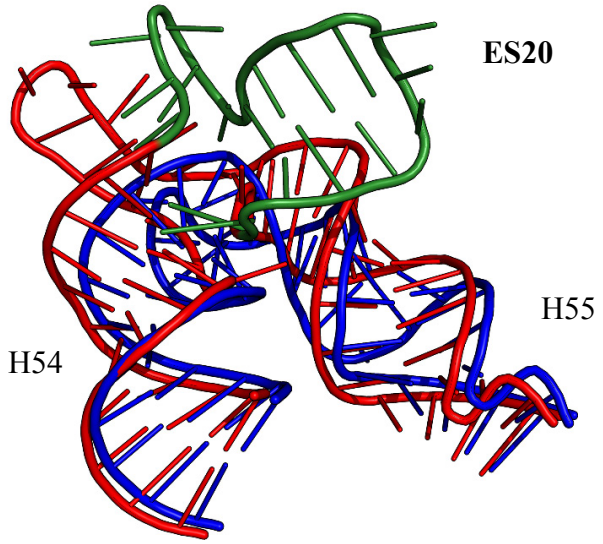
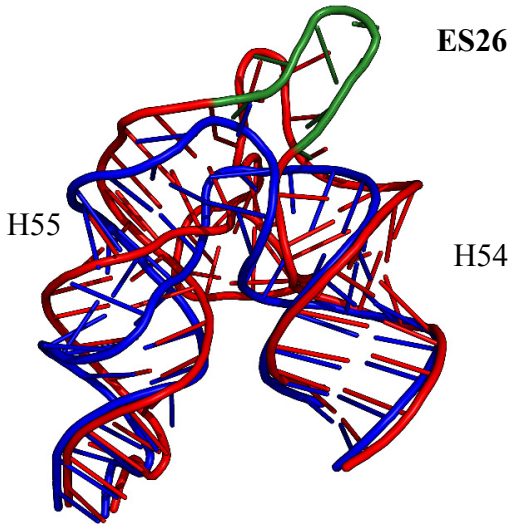
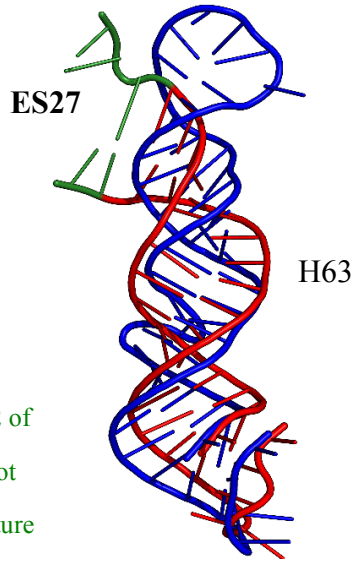
Species	LSU	SSU	C-value
H. sapiens	5347	1869	3.5
P. troglodytes	5125	1862	3.6
R. norvegicus	5063	1874	3.4
M. musculus	5021	1871	3.3
M. domestica	5236	1911	3.7
G. gallus	4845	1823	1.3
A. carolinensis	4200	1810	1.8
X. laevis	4361	1826	3.1
P. aethiopicus	3908	n.a.	133
L. chalumnae	4024	1808	2.9
D. rerio	4418	1887	1.8
A. albopictus	4331	1960	1.1
D. melanogaster	4198	1995	0.2
C. elegans	3782	1758	0.1
C. briggsae	3774	1754	0.1
A. vaga	3971	1812	0.2
S. cerevisiae	3674	1800	0.01
E. gossypii	3667	1795	0.01
Y. lipolytica	3459	1677	0.02
S. pombe	3762	1842	0.01
D. discoideum	4023	1875	0.03
T. thermophila	3628	1753	0.2
C. hominis	3647	1689	0.01
P. falciparum	4066	2092	0.03
T. brucei	4213	2251	0.03
L. major	4207	2204	0.03
T. pseudonana	3610	1805	0.03
G. theta	3992	1775	0.09
O. sativa	3700	1811	0.5
A. thaliana	3659	1808	0.4
P. glauca	3695	n.a.	24
C. merolae	3765	1809	0.02
Archaea	3019 (70)	1484 (15)	0.01 (0)
Bacteria	3087 (84)	1522 (28)	0.01 (0)

Table S2. Eukaryotic expansion segments accreted onto the common core. *E. coli* (common core) trunk helices are in blue. *S. cerevisiae* (eukaryote) trunk helices are in red and the expansion segments (branch helices) are in green. Definitions of trunk nucleotides of *E. coli* are shown in regular font; those of *S. cerevisiae* are in italics; the expansion segments of *S. cerevisiae* are in bold.

ES	Helix	E. coli (trunk)// *S. cerevisiae (trunk)*/ **Expansion Segment**	Three Dimensional View
7	25	23S:(528-578); 23S:(2018-2042)// *25S:(421-438); 25S:(620-652); 25S:(2361-2384)*/ **25S:(439-619)**	
19	52	23S:(1344-1385)// *25S:(1526-1553); 25S:(1583-1595)*/ **25S:(1554-1582)**	

12	38	23S:(822-862); 23S:(915-945)// *25S:(954-997);* *25S:(1052-1062);* *25S:(1098-1116)/* **25S:(1063-1097)**	
41	101	23S:(2835-2882)// *25S:(3320-3344);* *25S:(3361-3386)/* ***25S:(3345-3360)***	
3, 4	9,10	23S:(119-175)// *5.8S:(105-123);* *5.8S:(130-151);* *25S:(8-19)/* **5.8S:(124-129)** **5.8S:(152-158)** **25S:(1-7)**	 **H8** in *E. coli* is only one base pair longer than that in *S. cerevisiae*

5	15	23S:(263-297); 23S:(341-370); 23S:(424-430)// *25S:(107-111);* *25S:(155-186);* *25S:(230-267);* *25S:(319-325)/* **25S:(112-154)**	
9	30	23S:(625-656)// *25S:(703-715);* *25S:(753-757);* *25S:(775-787)/* **25S:(716-752)**	
9	31	23S:(625-656)// *25S:(703-715);* *25S:(753-757);* *25S:(775-787)/* **25S:(758-774)**	

20	54	23S:(1405-1428); 23S:(1569-1597)// 25S:(1615-1625); *25S:(1644-1657);* *25S:(1797-1829)/* **25S:(1626-1643)**	
26	54	23S:(1405-1428); 23S:(1569-1597)// *25S:(1615-1657);* *25S:(1797-1811);* *25S:(1819-1829)/* **25S:(1812-1818)**	
27	63	23S:(1706-1757)// *25S:(1938-1952);* *25S:(2095-2115)/* **25S:(1953-2094)**	 *Nucleotides 1956-2092 of **ES27** (25S rRNA) are not resolved in crystal structure

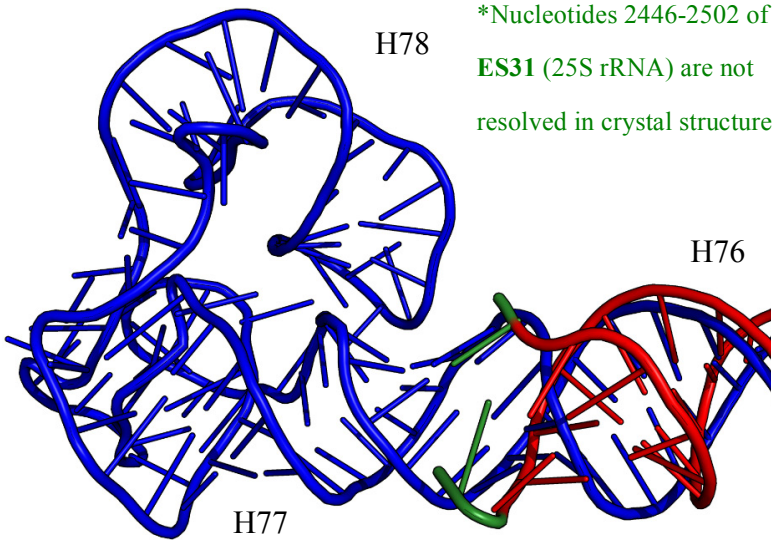
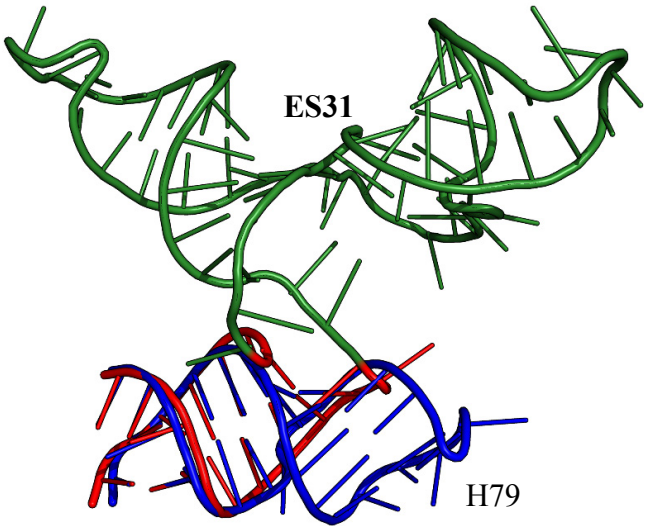
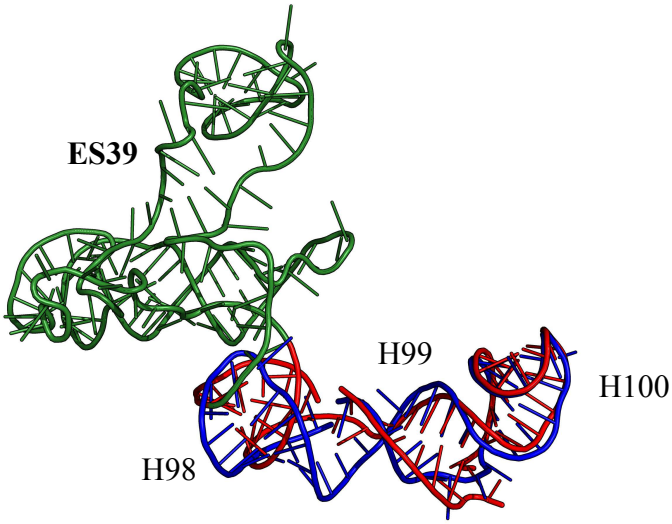
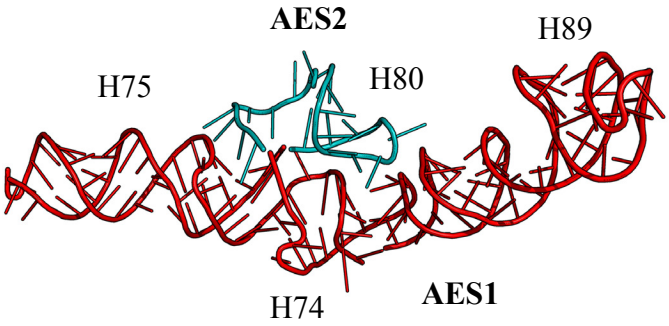
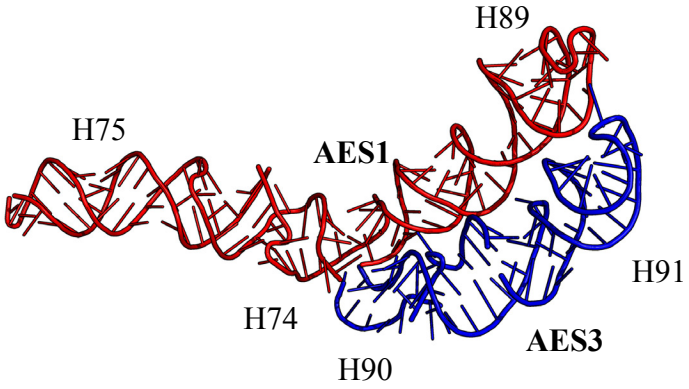
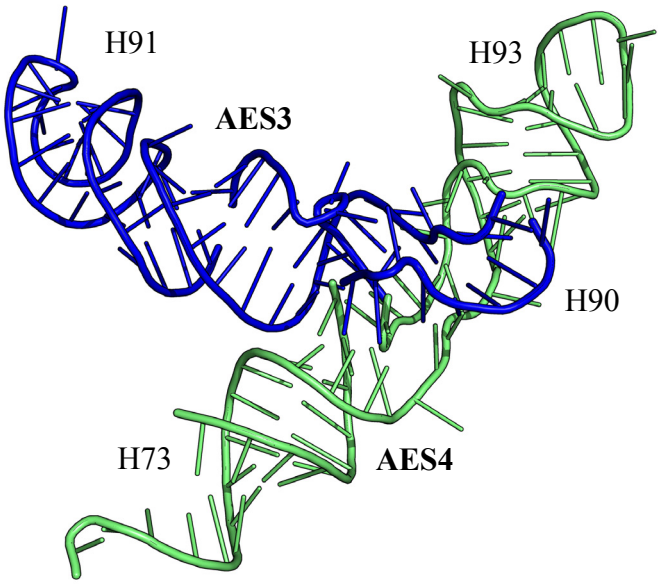
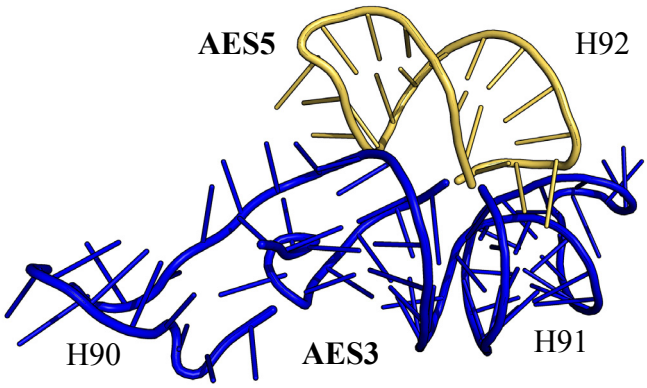
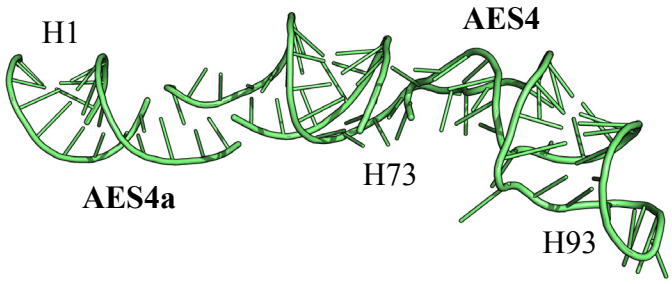
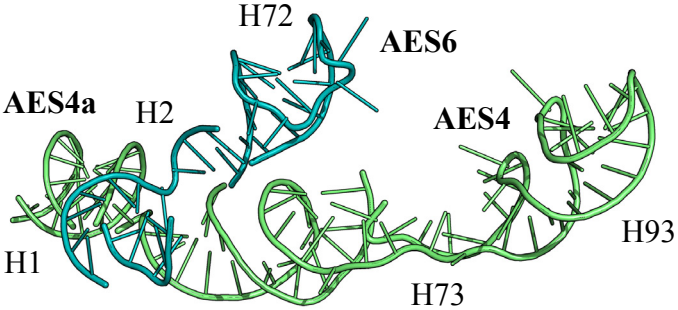
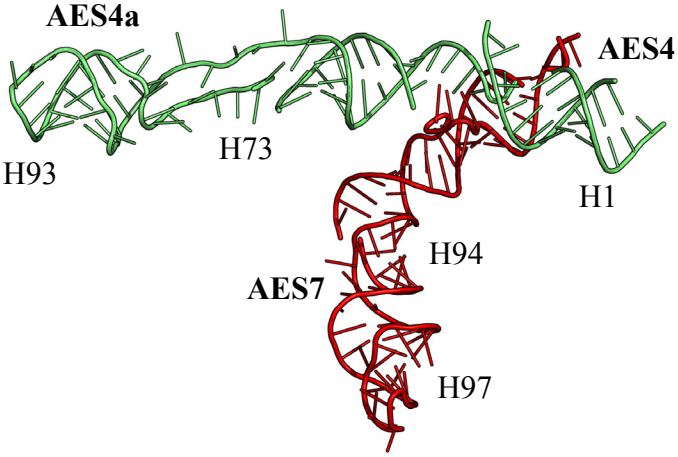
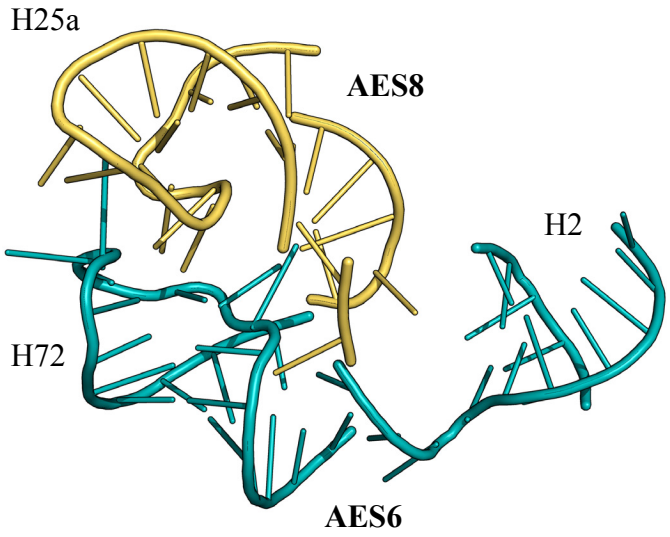
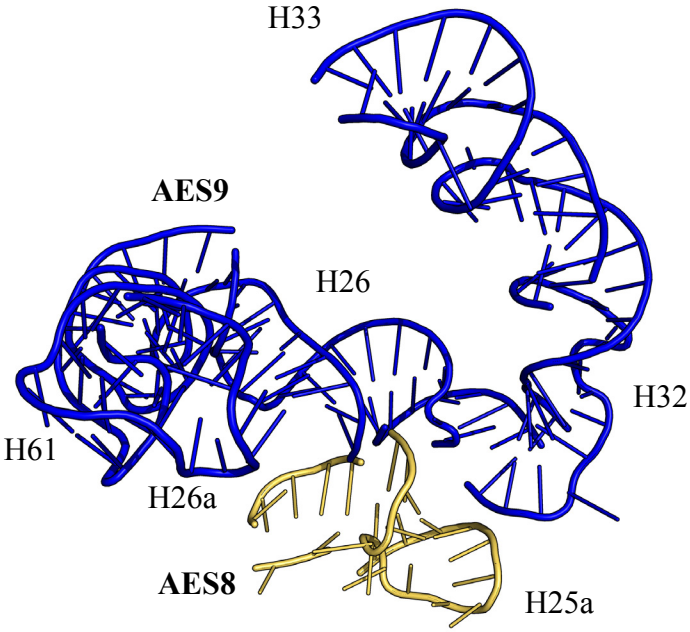
NA	78	23S:(2093-2195)// *25S:(2435-2444); 25S:(2502-2511)/* **25S:(2445-2501)**	
31	79	23S:(2196-2225)// *25S:(2512-2523); 25S:(2587-2593)/* **25S:(2524-2586)**	
39	98	23S:(2790-2832); 23S:(2884-2891)// *25S:(3159-3167); 25S:(3283-3317); 25S:(3388-3395)/* **25S:(3168-3282)**	

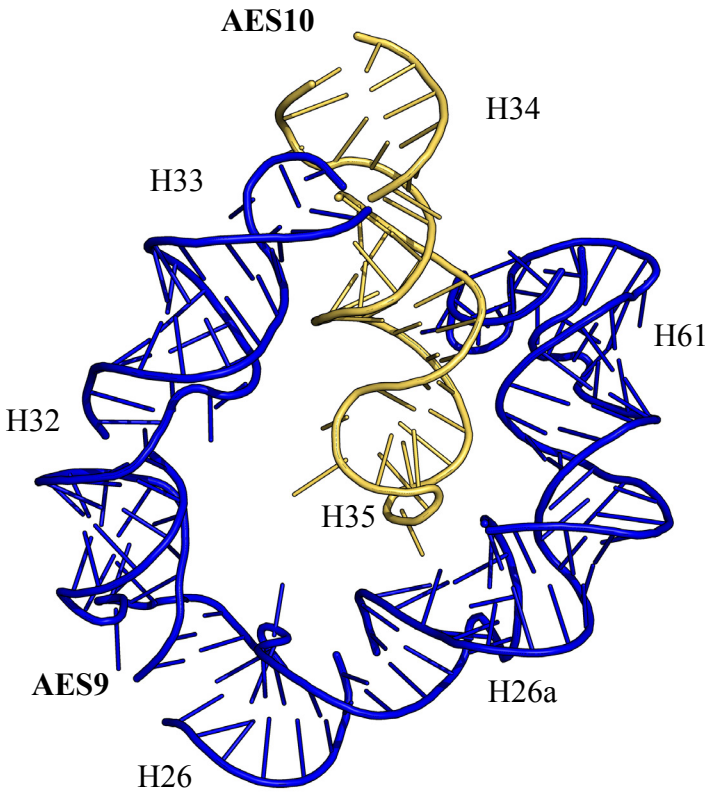
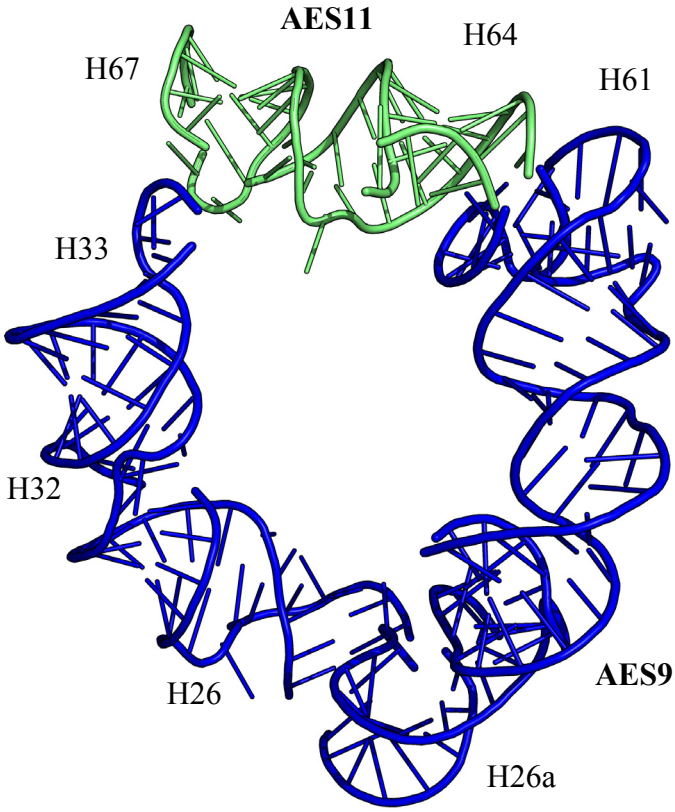
Table S3. Ancestral expansion segments (AESs) of a bacterial (*E. coli*) LSU rRNA. AESs are colored according to the scheme established in Figure 6a. Original helixes numbers and residue numbers are also provided.

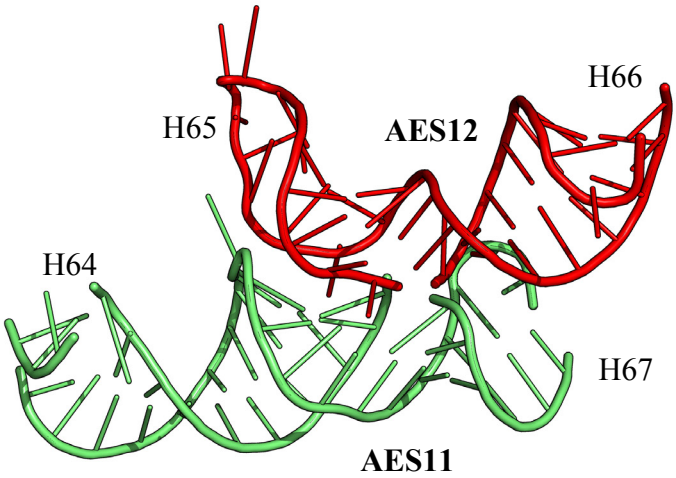
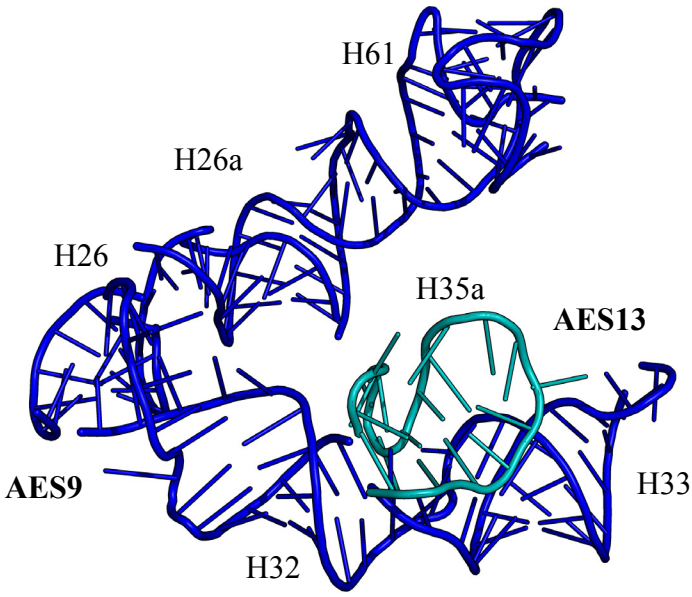
AES trunk/ *AES branch*	**Helices trunk/** *Helices branch*	**Nucleotides trunk/** *Nucleotides branch*	Three Dimensional View
Phase 1			
1 / *2*	**74, 75, 89** / *80*	**23S:(2061-2092); 23S:(2226-2245); 23S:(2436-2501)** / *23S:(2427-2435); 23S:(2246-2258)*	
1 / *3*	**74, 75, 89** / *90, 91*	**23S:(2061-2092); 23S:(2226-2245); 23S:(2436-2501)** / *23S:(2502-2546); 23S:(2053-2060); 23S:(2567-2576)*	

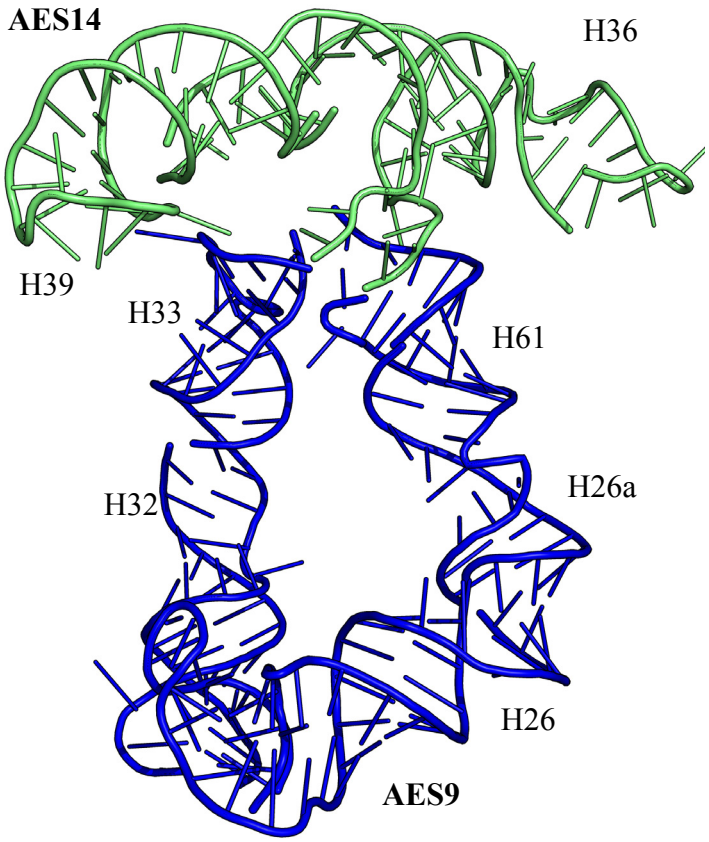
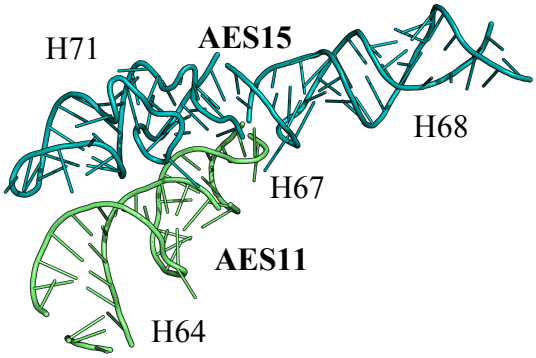
Phase 2			
3 / *4*	**90, 91** / *73, 93*	**23S:(2502-2546); 23S:(2053-2060); 23S:(2567-2576)** / *23S:(2043-2052); 23S:(2626-2629); 23S:(2577-2625)*	
3 / *5*	**90, 91** / *92*	**23S:(2502-2546); 23S:(2053-2060); 23S:(2567-2576)** / *23S:(2547-2566)*	

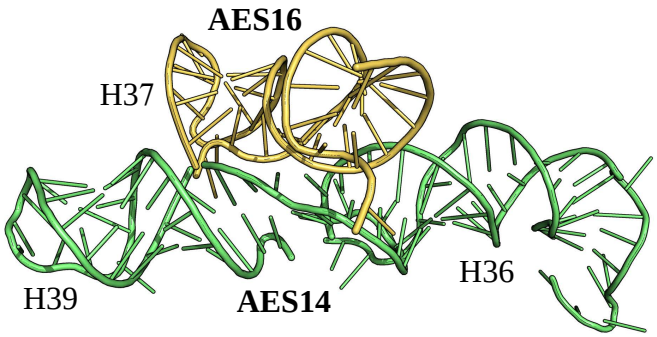
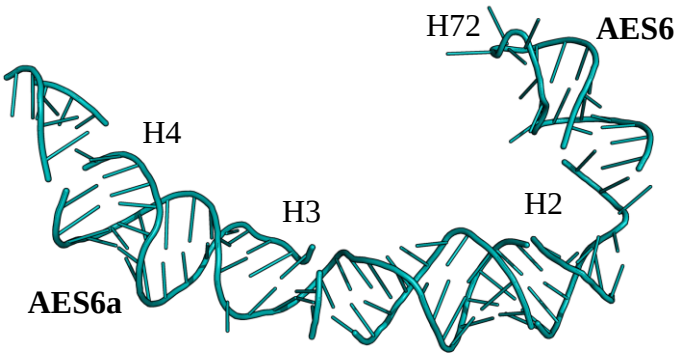
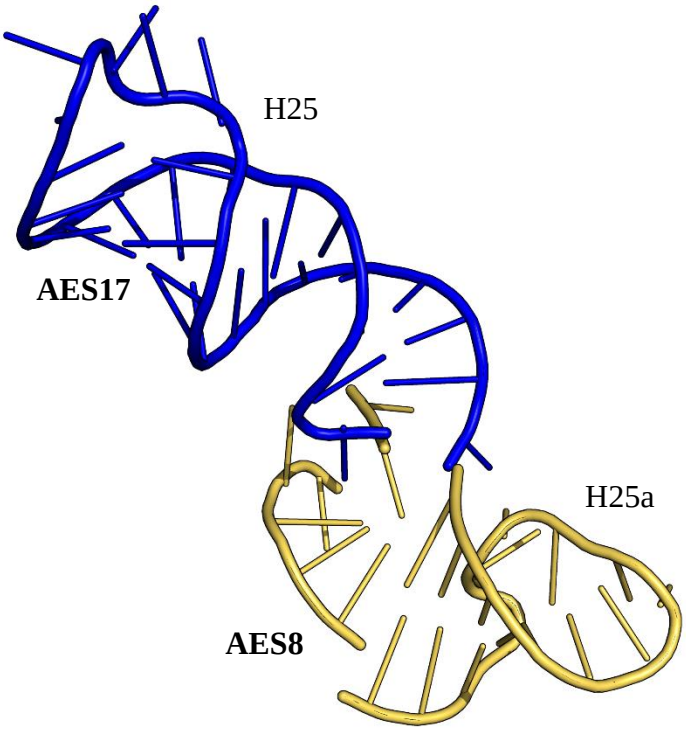
Phase 3			
4 / *4a*	**73, 93** / *1*	**23S:(2043-2052); 23S:(2626-2629); 23S:(2577-2625)** / *23S:(1-11); 23S:(2895-2903)*	
4, 4a / *6*	**73, 93, 1** / *72, 2*	**23S:(2043-2052); 23S:(2626-2629); 23S:(2577-2625) 23S:(1-11); 23S:(2895-2903)** / *23S:(527-529); 23S:(2023-2042); 23S:(12-16); 23S: (520-526)*	
4, 4a / *7*	**73, 93, 1** / *94, 97*	**23S:(2043-2052); 23S:(2626-2629); 23S:(2577-2625); 23S:(1-11); 23S:(2895-2903)** / *23S:(2630-2645); 23S:(2771-2790); 23S:(2893-2894); 23S:(2733-2770)*	

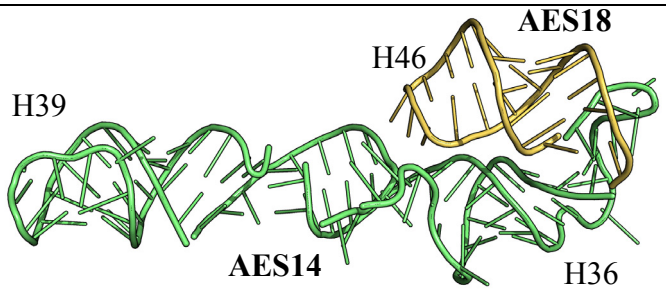
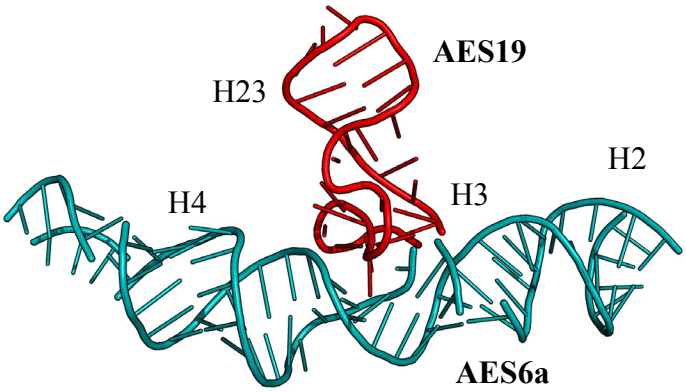
6 / *8*	**72, 2** / *25a*	**23S:(527-529); 23S:(2023-2042); 23S:(12-16); 23S:(520-526)** / *23S:(530-531); 23S:(563-578); 23S:(2018-2022)*	
8 / *9*	**25a** / *33, 32, 26, 26a, 61*	**23S:(530-531); 23S:(563-578); 23S:(2018-2022)** / *23S:(579-586); 23S:(1254-1261); 23S:(1262-1270); 23S:(2010-2017); 23S:(794-808); 23S:(672-683); 23S:(684-698); 23S:(762-775); 23S:(1648-1678); 23S:(1991-2009)*	

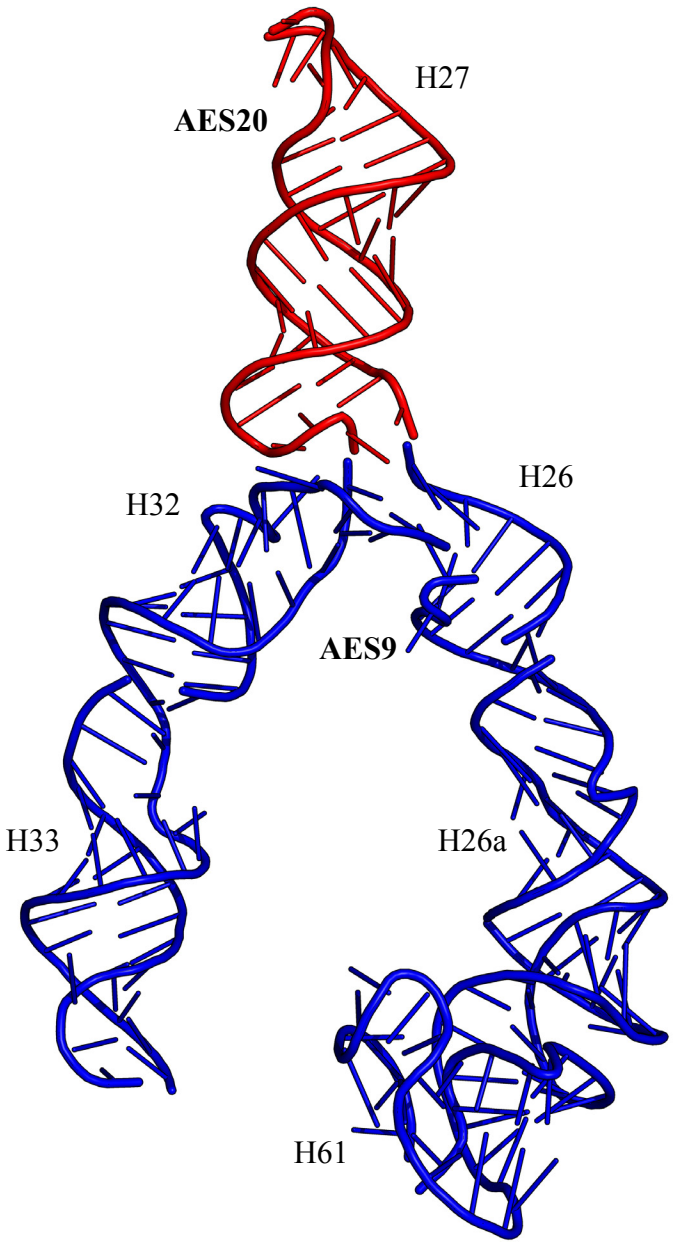
9 / *10*	**33, 32, 26, 26a, 61** / *34, 35*	**23S:(579-586); 23S:(1254-1261); 23S:(1262-1270); 23S:(2010-2017); 23S:(794-808); 23S:(672-683); 23S:(684-698); 23S:(762-775); 23S:(1648-1678); 23S:(1991-2009)** / *23S:(699-704); 23S:(727-761)*	
Phase 4			
9 / *11*	**33, 32, 26, 26a, 61** / *64, 67*	**23S:(579-586); 23S:(1254-1261); 23S:(1262-1270); 23S:(2010-2017); 23S:(794-808); 23S:(672-683); 23S:(684-698); 23S:(762-775); 23S:(1648-1678); 23S:(1991-2009)** / *23S:(1679-1681); 23S:(1763-1772); 23S:(1829-1834); 23S:(1970-1990)*	

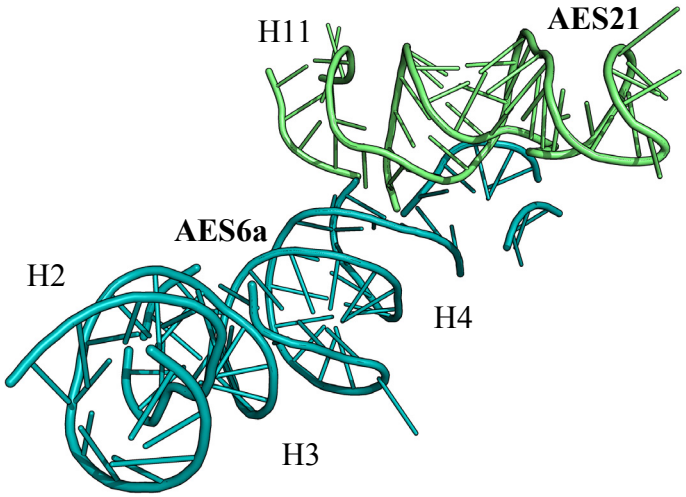
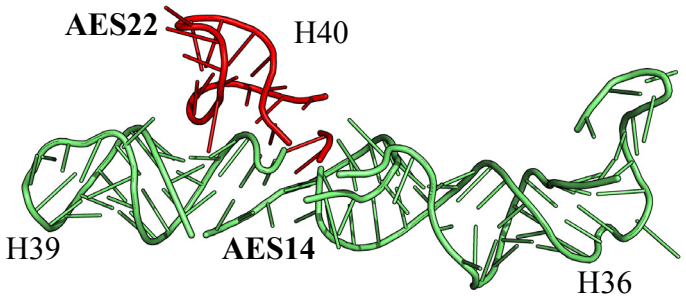
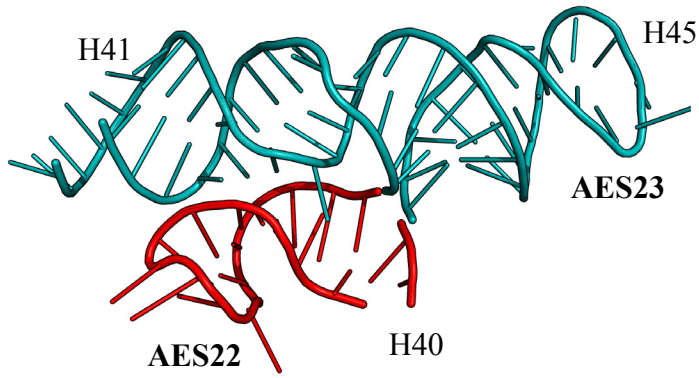
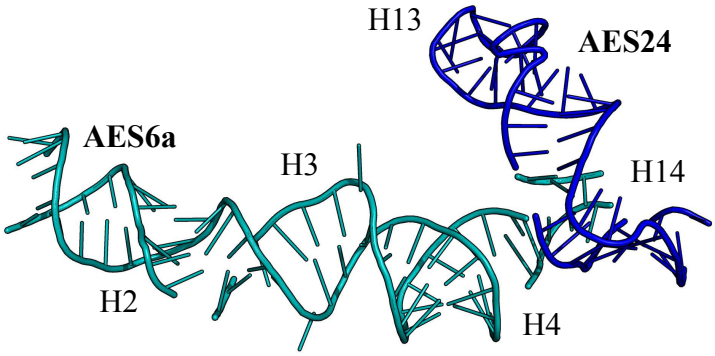
11 / *12*	**64, 67** / *65, 66*	**23S:(1679-1681); 23S:(1763-1772); 23S:(1829-1834); 23S:(1970-1990)** / *23S:(1773-1798); 23S:(1819-1828)*	
9 / *13*	**33, 32, 26, 26a, 61** / *35a*	**23S:(579-586); 23S:(1254-1261); 23S:(1262-1270); 23S:(2010-2017); 23S:(794-808); 23S:(672-683); 23S:(684-698); 23S:(762-775); 23S:(1648-1678); 23S:(1991-2009)** / *23S:(776-793)*	

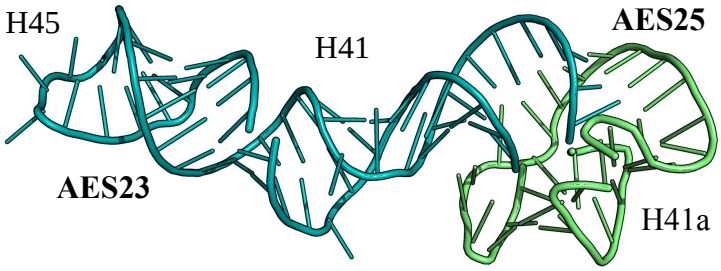
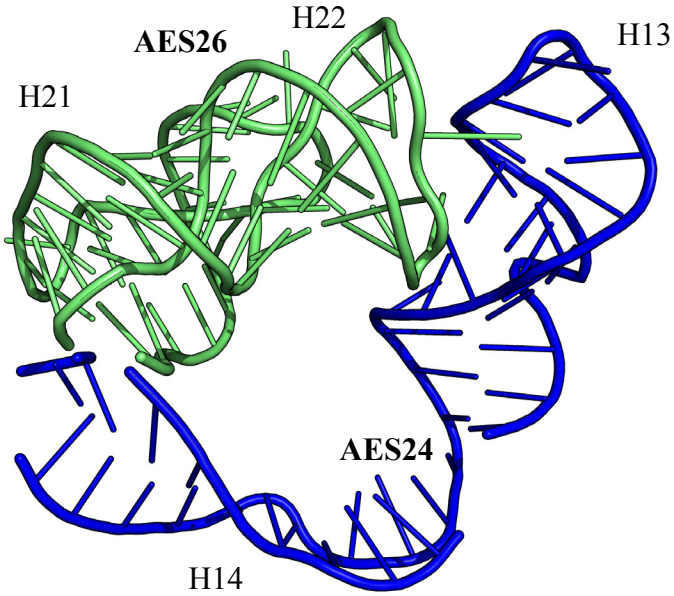
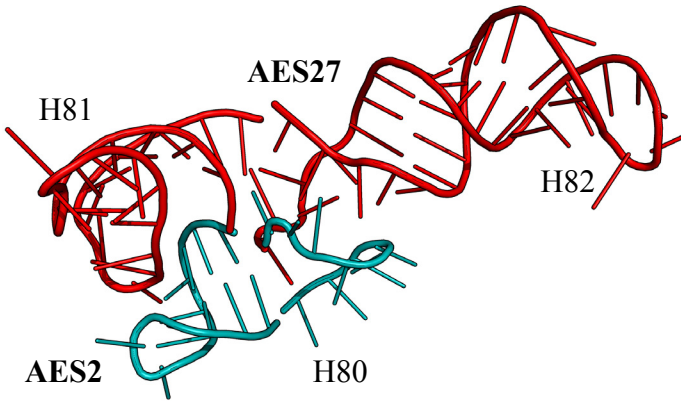
9 / *14*	**33, 32, 26, 26a, 61** / *36, 39*	**23S:(579-586); 23S:(1254-1261); 23S:(1262-1270); 23S:(2010-2017); 23S:(794-808); 23S:(672-683); 23S:(684-698); 23S:(762-775); 23S:(1648-1678); 23S:(1991-2009)** / *23S:(809-821); 23S:(946-973); 23S:(1188-1213); 23S:(1238-1253)*	
11 / *15*	**64, 67** / *68, 71*	**23S:(1679-1681); 23S:(1763-1772); 23S:(1829-1834); 23S:(1970-1990)** / *23S:(1835-1855); 23S:(1887-1905); 23S:(1930-1969)*	

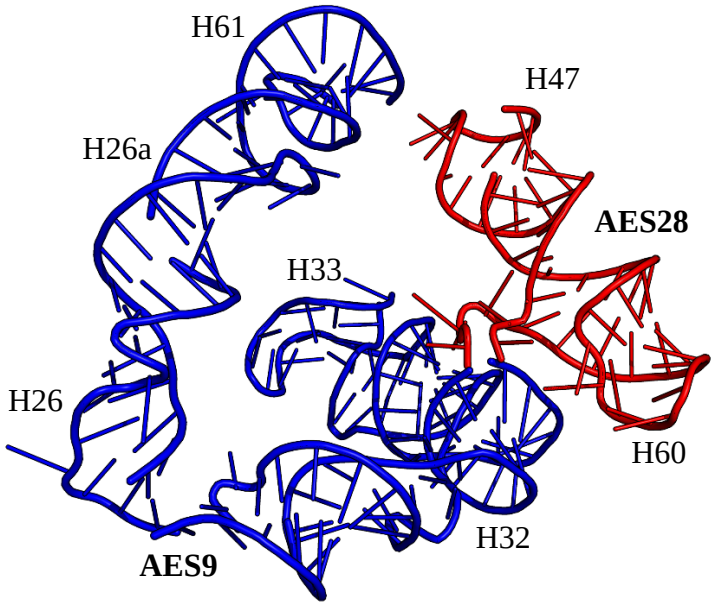
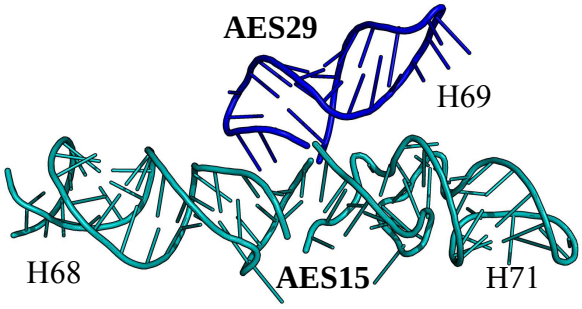
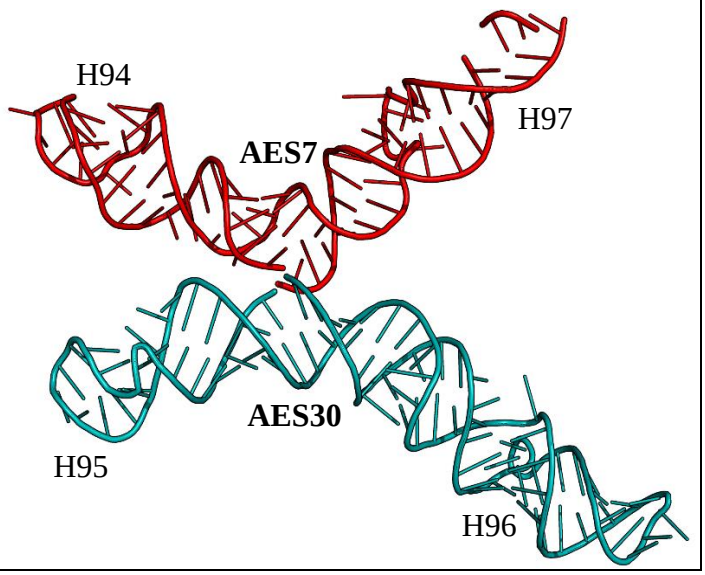
14 / *16*	**36, 39** / *37*	**23S:(809-821); 23S:(946-973); 23S:(1188-1213); 23S:(1238-1253)** / *23S:(822-845); 23S:(931-945)*	
6 / *6a*	**72, 2** / *2, 3, 4*	**23S:(527-529); 23S:(2023-2042); 23S:(12-16); 23S(520-526)** / *23S:(17-45); 23S:(216-220); 23S:(234-236); 23S:(510-519); 23S:(431-447); 23S:(473-474)*	
8 / *17*	**25a** / *25*	**23S:(530-531); 23S:(563-578); 23S:(2018); 23S:(2022)** / *23S:(532-562)*	

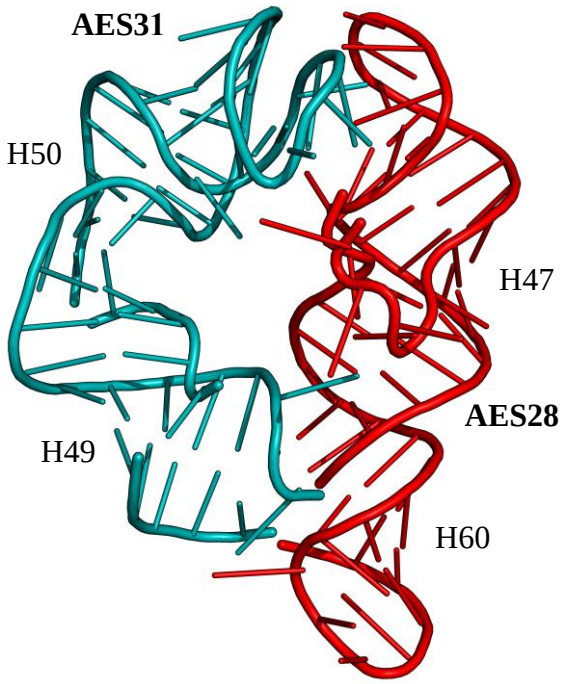
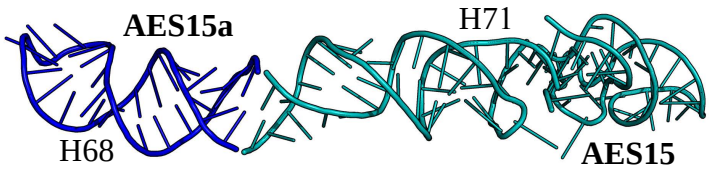
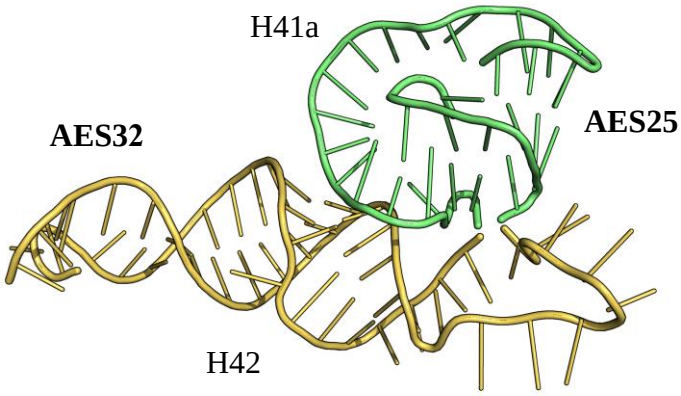
14 / *18*	**36, 39** / *46*	**23S:(809-821); 23S:(946-973); 23S:(1188-1213); 23S:(1238-1253)** / *23S:(1214-1237)*	
6a / *19*	**2, 3, 4** / *23*	**23S:(17-45); 23S:(216-220); 23S:(234-236); 23S:(510-519); 23S:(431-447); 23S:(473-474)** / *23S:(448-472)*	

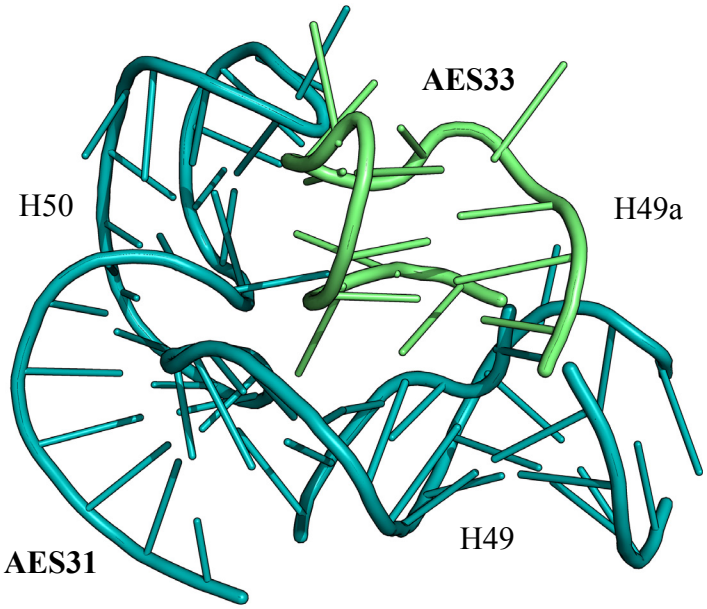
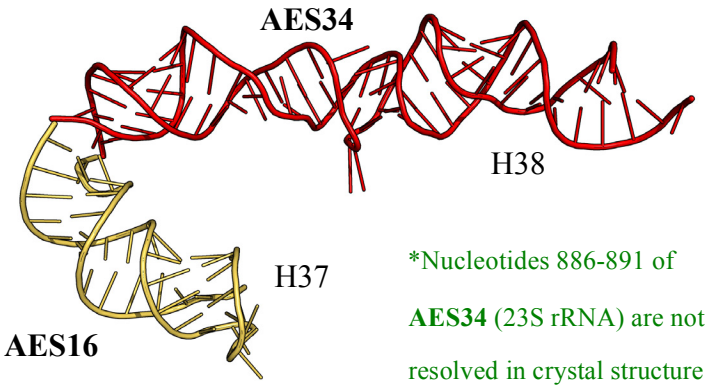
9 / *20*	**33, 32, 26, 26a, 61** / *27*	**23S:(579-586); 23S:(1254-1261); 23S:(1262-1270); 23S:(2010-2017); 23S:(794-808); 23S:(672-683); 23S:(684-698); 23S:(762-775); 23S:(1648-1678); 23S:(1991-2009)** / *23S:(587-603); 23S:(653-671)*	

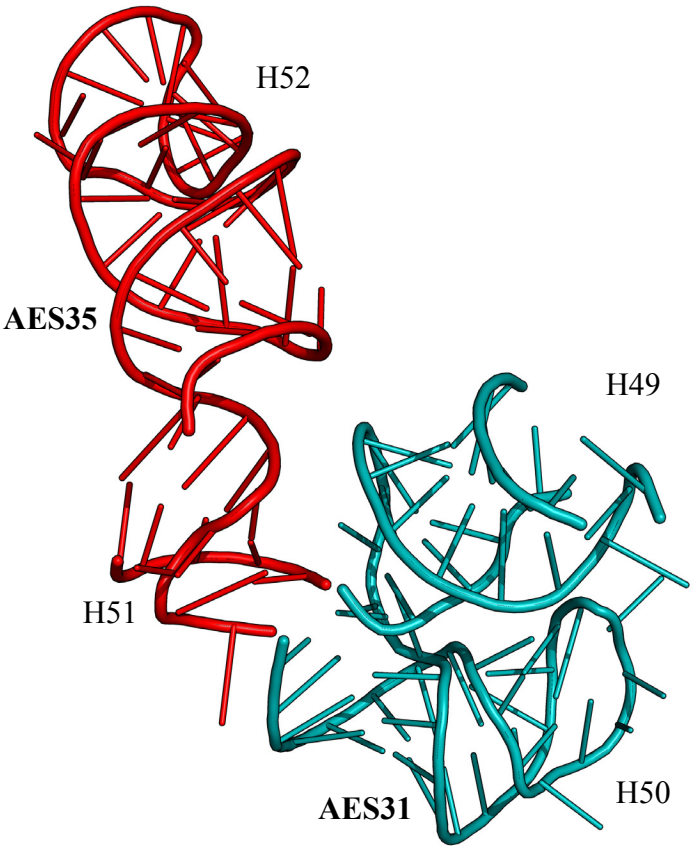
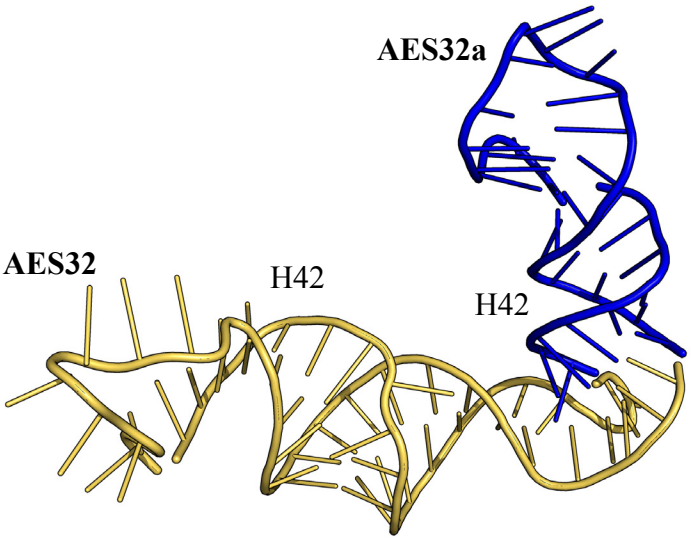
6a / *21*	**2, 3, 4** / *11*	**23S:(17-45); 23S:(216-220); 23S:(234-236); 23S:(510-519); 23S:(431-447); 23S:(473-474)** / *23S:(46-49); 23S:(175-215)*	
14 / *22*	**36, 39** / *40*	**23S:(809-821); 23S:(946-973); 23S:(1188-1213); 23S:(1238-1253)** / *23S:(974-990); 23S:(1186-1187)*	
22 / *23*	**40** / *41, 45*	**23S:(974-990); 23S:(1186-1187)** / *23S:(991-1004); 23S:(1144-1185)*	
6a / *24*	**2, 3, 4** / *13, 14*	**23S:(17-45); 23S:(216-220); 23S:(234-236); 23S:(510-519); 23S:(431-447); 23S:(473-474)** / *23S:(237-270); 23S:(369-370); 23S:(424-430)*	

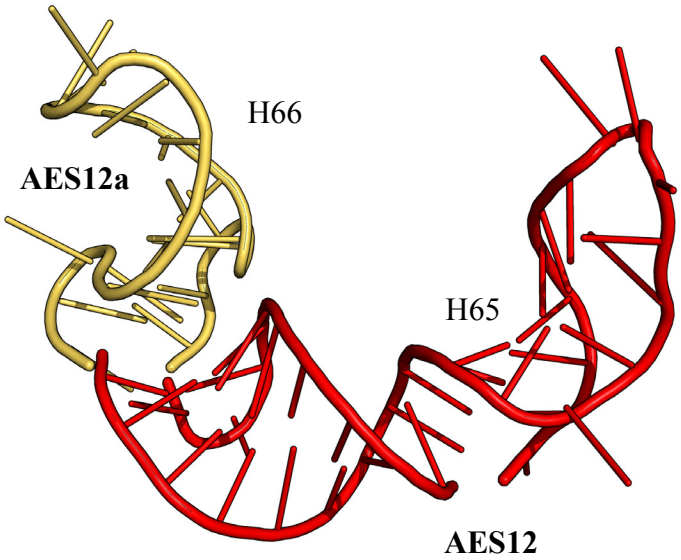
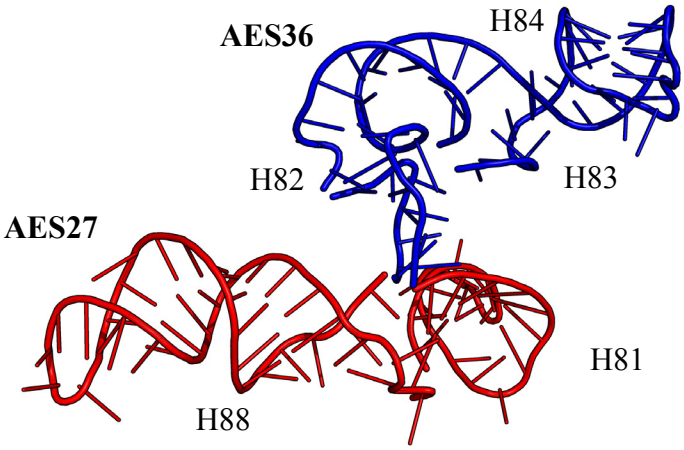
23 / *25*	**41, 45** / *41a*	**23S:(991-1004); 23S:(1144-1185)** / *23S:(1005-1025); 23S:(1135-1143)*	
24 / *26*	**13, 14** / *21, 22*	**23S:(237-270); 23S:(369-370); 23S:(424-430)** / *23S:(371-423)*	
2/ *27*	**80** / *81, 88*	**23S:(2427-2435); 23S:(2246-2258)** / *23S:(2259-2282); 23S:(2390-2426)*	

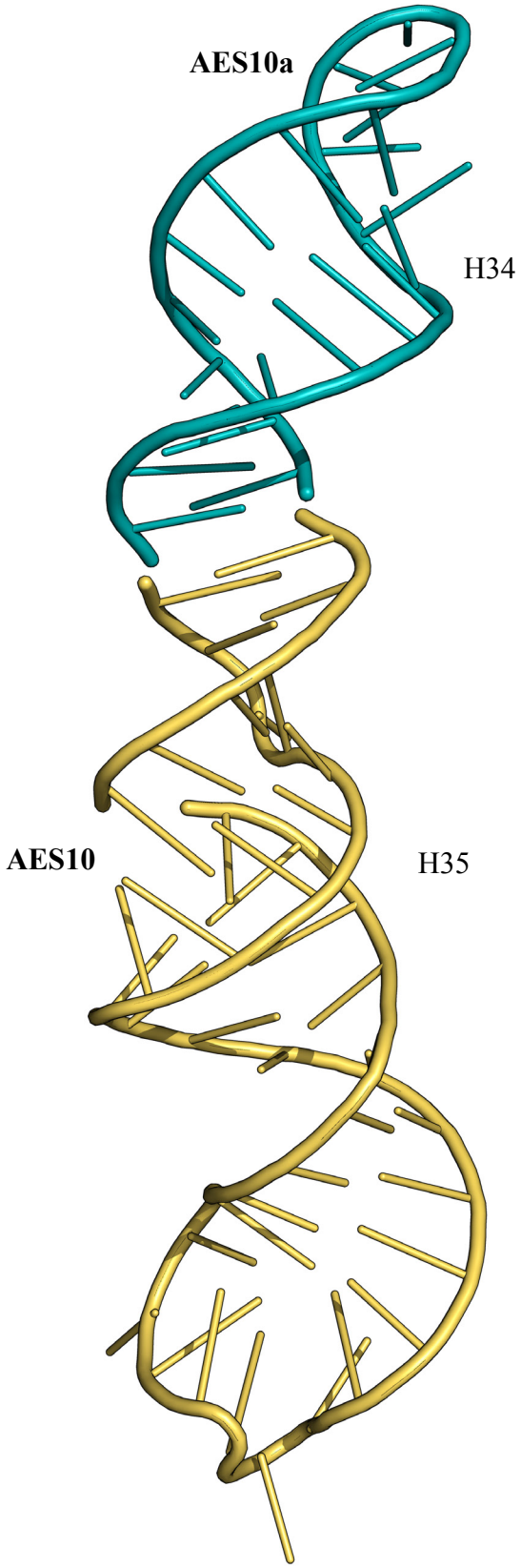
9 / *28*	**33, 32, 26, 26a, 61** / *47, 60*	**23S:(579-586); 23S:(1254-1261); 23S:(1262-1270); 23S:(2010-2017); 23S:(794-808); 23S:(672-683); 23S:(684-698); 23S:(762-775); 23S:(1648-1678); 23S:(1991-2009)** / *23S:(1271-1299); 23S:(1627-1647)*	
Phase 5			
15 / *29*	**68, 71** / *69*	**23S:(1835-1855); 23S:(1887-1905); 23S:(1930-1969)** / *23S:(1906-1929)*	
7 / *30*	**94, 97** / *95, 96*	**23S:(2630-2645); 23S:(2771-2790); 23S:(2893-2894); 23S:(2733-2770)** / *23S:(2646-2732)*	

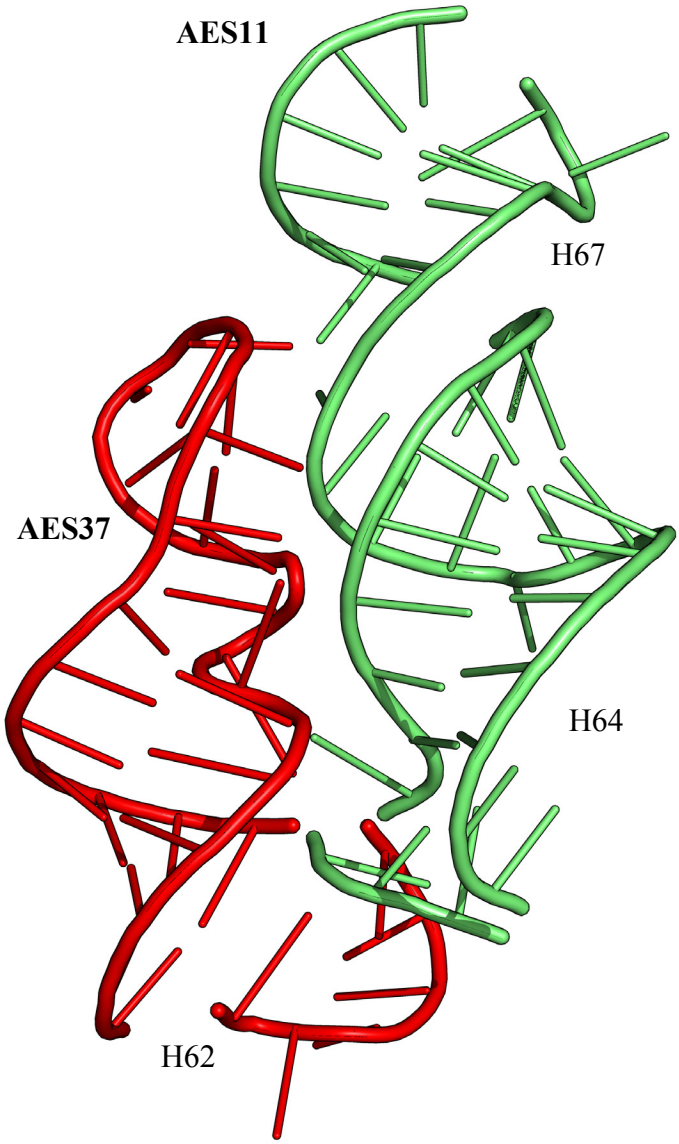
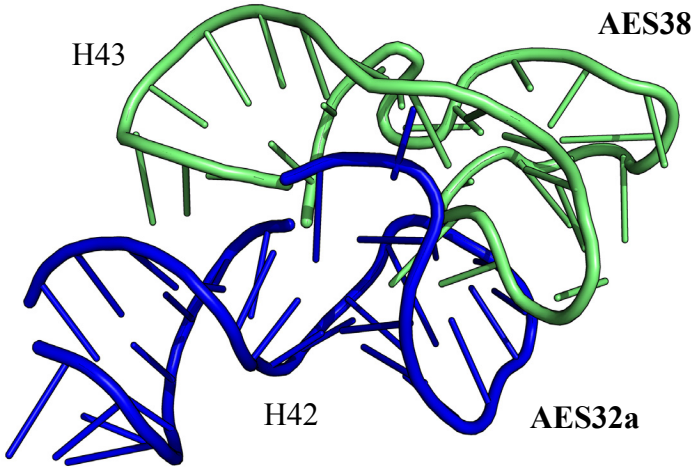
28 / *31*	**47, 60** / *49, 50*	**23S:(1271-1299); 23S:(1627-1647)** / *23S:(1300-1340); 23S:(1603-1607); 23S:(1622-1626)*	
15 / *15a*	**68, 71** / *68*	**23S:(1835-1855); 23S:(1887-1905); 23S:(1930-1969)** / *23S:(1856-1886)*	
25 / *32*	**41a** / *42*	**23S:(1005-1025); 23S:(1135-1143)** / *23S:(1026-1047); 23S:(1108-1134)*	

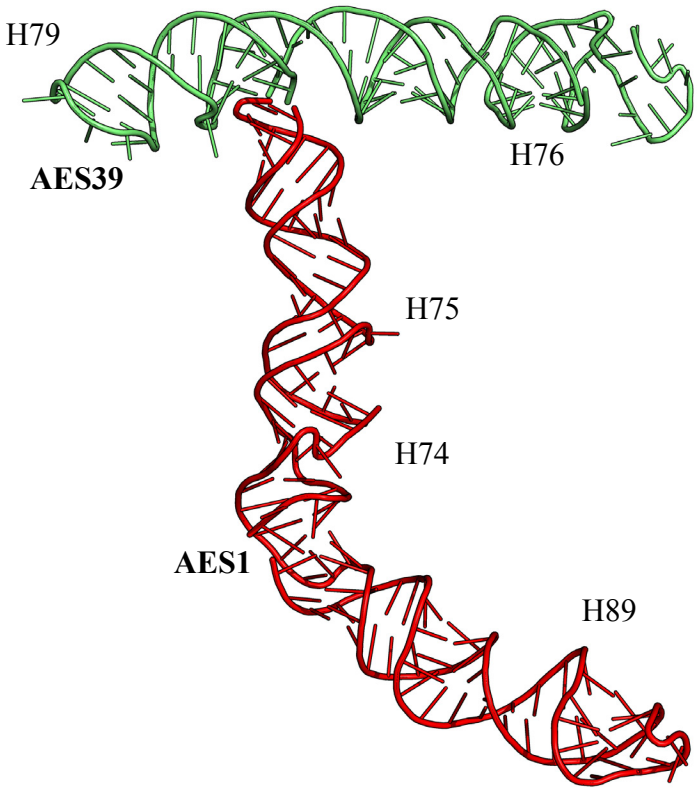
31 / *33*	**49, 50** / *49a*	**23S:(1300-1340); 23S:(1603-1607)** / *23S:(1622-1626); 23S:(1608-1621)*	
16 / *34*	**37** / *38*	**23S:(822-845); 23S:(931-945)** / *23S:(846-930)*	

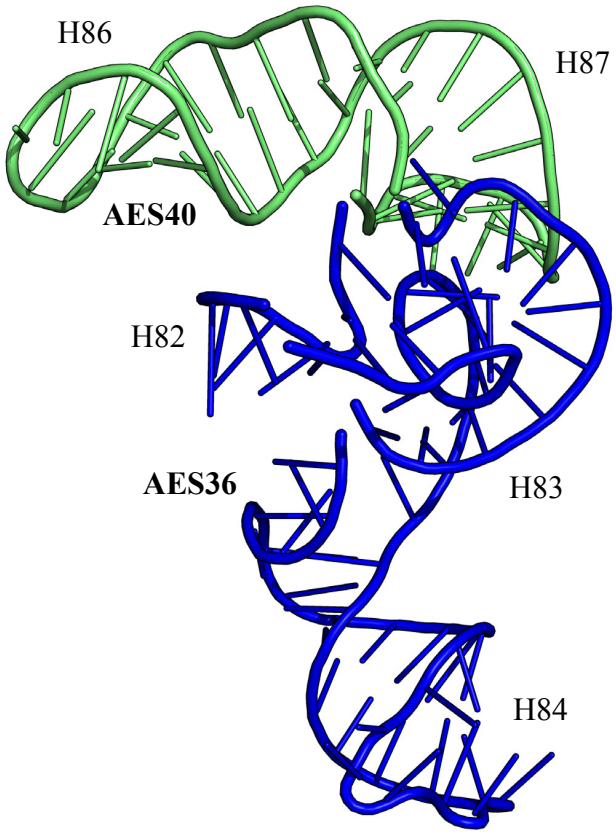
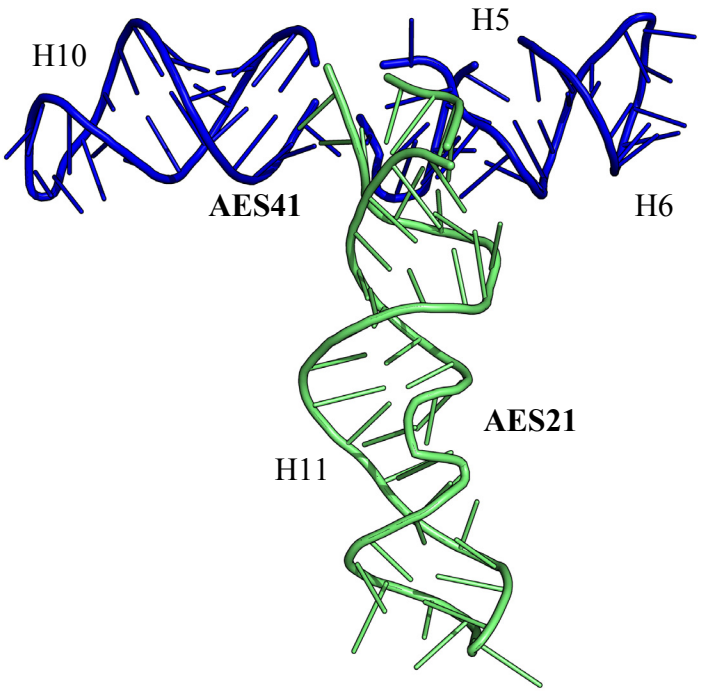
31 / *35*	**49, 50** / *51, 52*	**23S:(1300-1340); 23S:(1603-1607); 23S:(1622-1626)** / *23S:(1598-1602); 23S:(1341-1383)*	
32 / 32a	**42** / *42*	**23S:(1026-1047); 23S:(1108-1134)** / *23S:(1048-1056); 23S:(1087-1107)*	

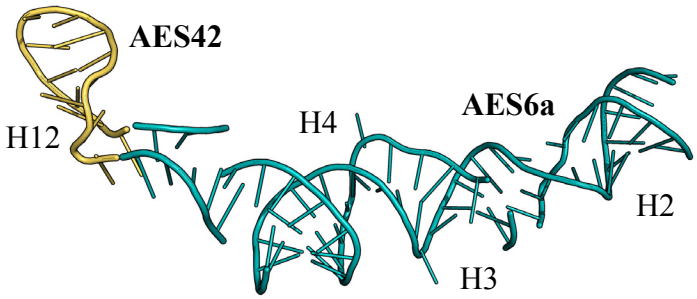
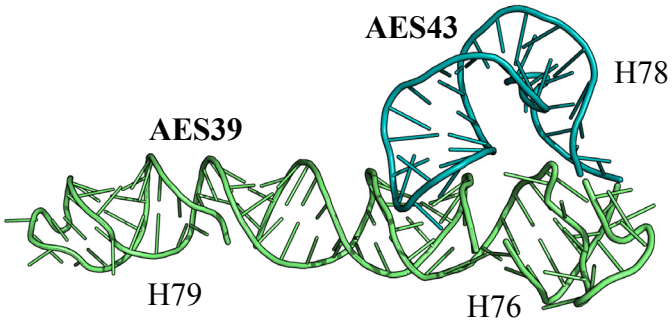
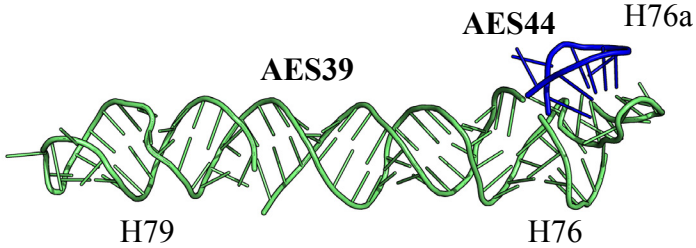
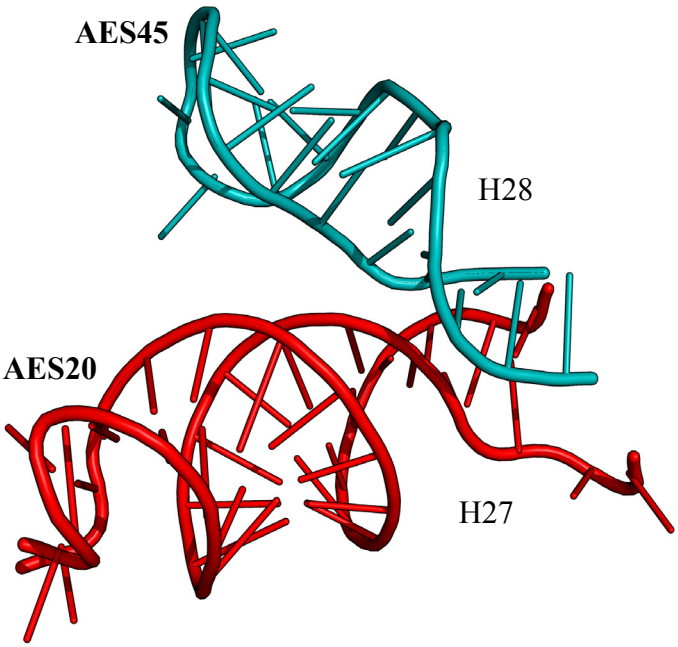
12 / 12a	**65, 66** / *65, 66*	**23S:(1773-1798); 23S:(1819-1828)** / *23S:(1799-1818)*	
27 / *36*	**81, 88** / *82, 83, 84*	**23S:(2259-2282); 23S:(2390-2426)** / *23S:(2283-2324); 23S:(2337-2346); 23S:(2383-2389)*	

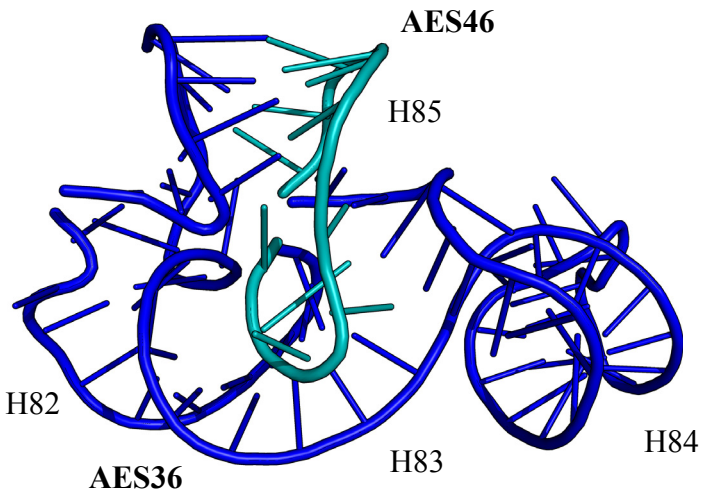
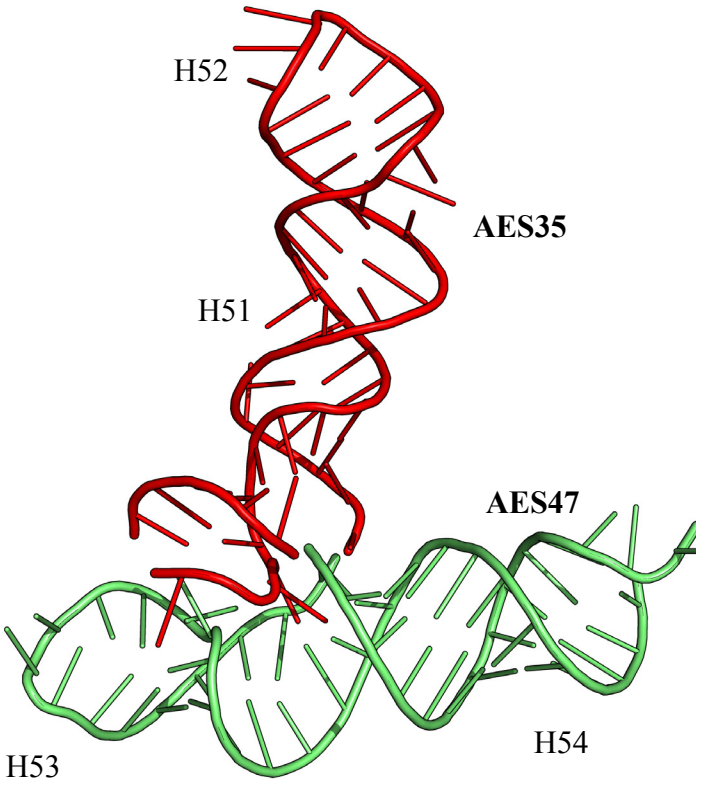
10 / *10a*	**34, 35** / *34, 35*	**23S:(699-704); 23S:(727-761)** / *23S:(705-726)*	

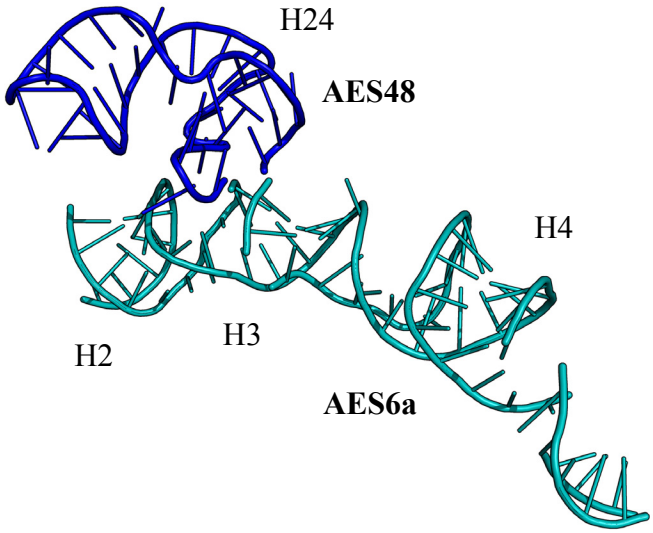
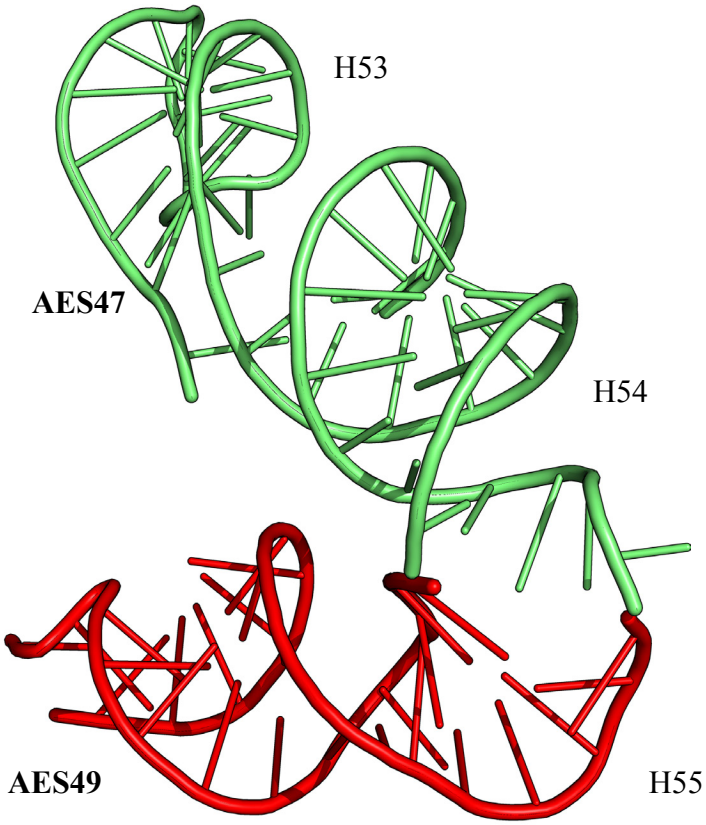
11 / *37*	**64, 67** / *62*	**23S:(1679-1681); 23S:(1763-1772); 23S:(1829-1834); 23S:(1970-1990)** / *23S:(1682-1706); 23S:(1757-1762)*	
32a / *38*	**42** / *43*	**23S:(1048-1056); 23S:(1087-1107)** / *23S:(1057-1086)*	

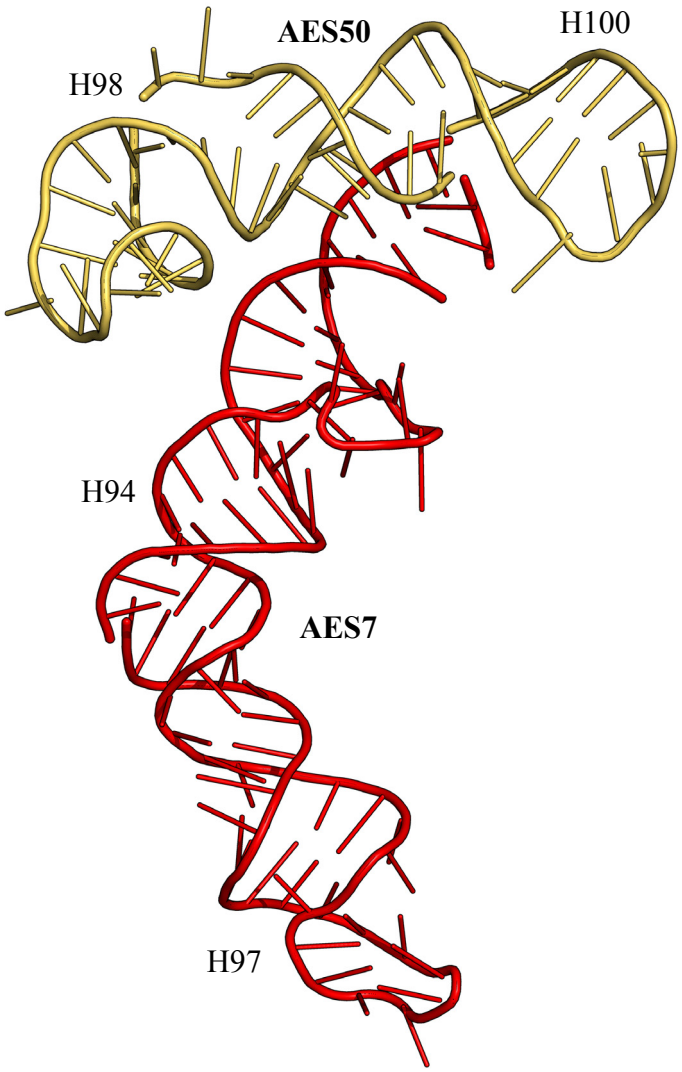
1 / *39*	**74, 75, 89** / *76, 79*	**23S:(2061-2092); 23S:(2226-2245); 23S:(2436-2501)** / *23S:(2093-2110); 23S:(2120-2126); 23S:(2173-2225); 23S:(2162-2172)*	

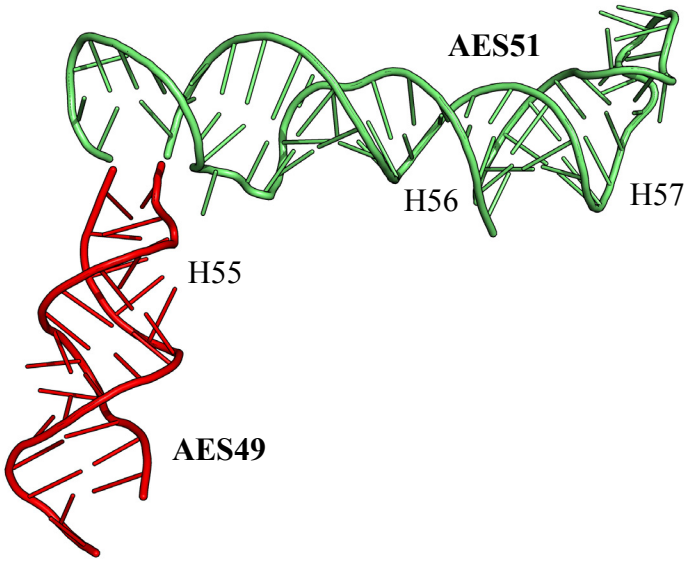
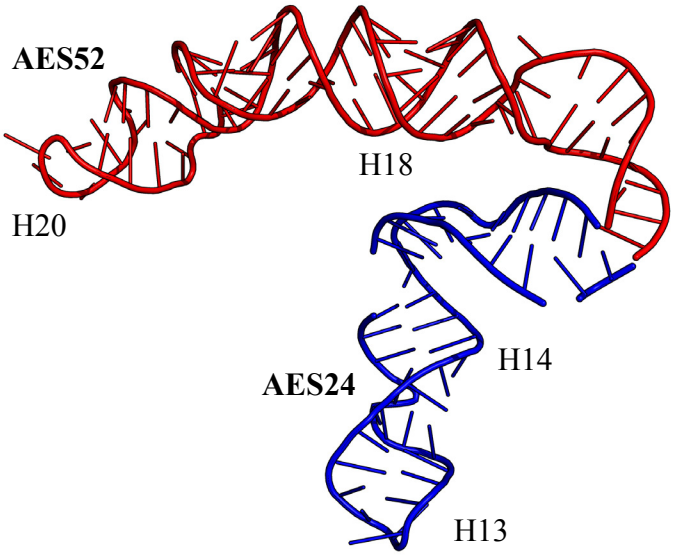
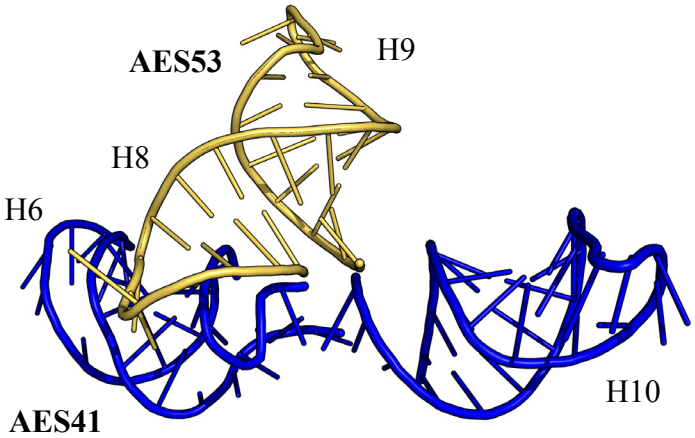
Phase 6			
36 / *40*	**82, 83, 84** / *86, 87*	**23S:(2283-2324); 23S:(2337-2346); 23S:(2383-2389)** / *23S:(2347-2382)*	
21 / *41*	**11** / *5, 6, 10*	**23S:(46-49); 23S:(175-215)** / *23S:(50-70); 23S:(114-120); 23S:(149-174)*	

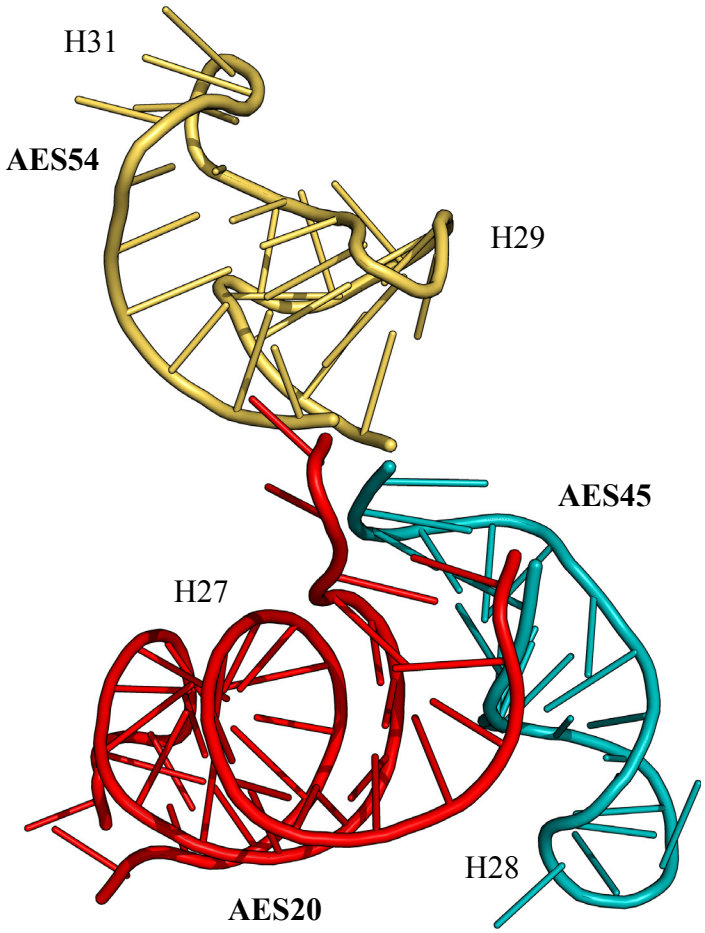
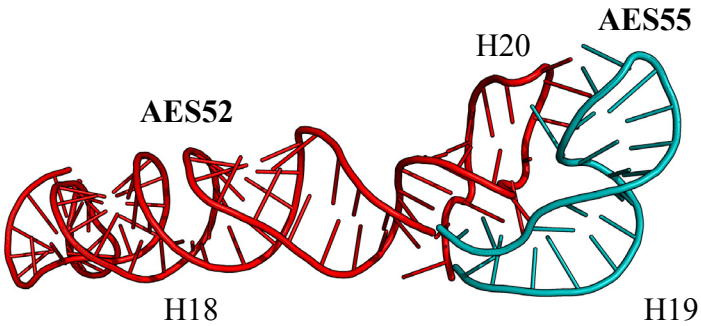
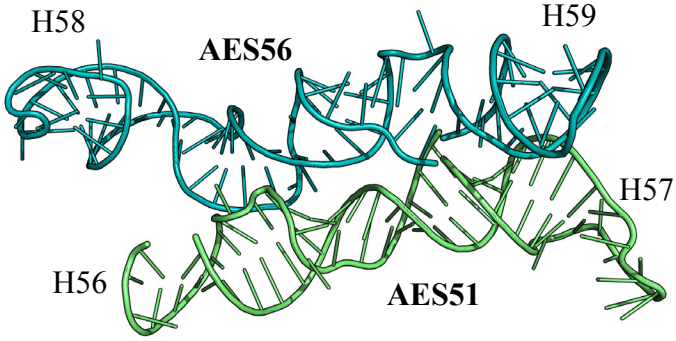
6a / *42*	**2, 3, 4** / *12*	**23S:(17-45); 23S:(216-220); 23S:(234-236); 23S:(510-519); 23S:(431-447); 23S:(473-474)** / *23S:(221-233)*	
39 / *43*	**76, 79** / *78*	**23S:(2093-2110); 23S:(2120-2126); 23S:(2173-2225); 23S:(2162-2172)** / *23S:(2127-2161)*	
39 / *44*	**76, 79** / *76a*	**23S:(2093-2110); 23S:(2120-2126); 23S:(2173-2225); 23S:(2162-2172)** / *23S:(2111-2119)*	
20 / *45*	**27** / *28*	**23S:(587-603); 23S:(653-671)** / *23S:(604-626)*	

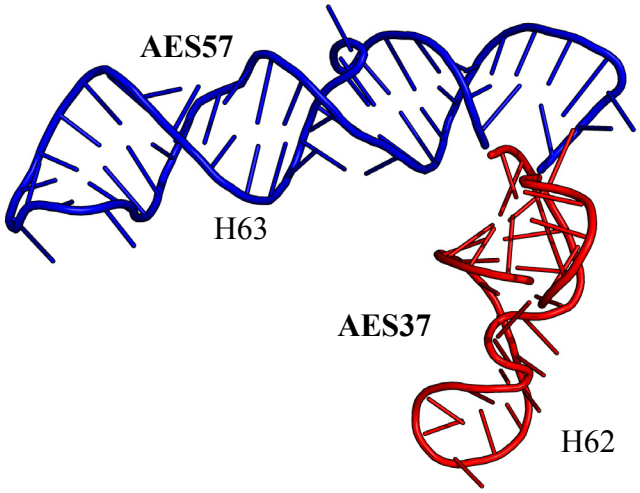
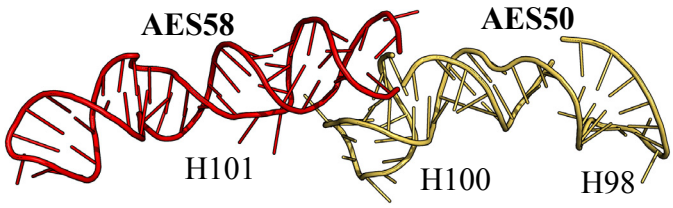
36 / *46*	**82, 83, 84** / *85*	**23S:(2283-2324); 23S:(2337-2346); 23S:(2383-2389)** / *23S:(2325-2336)*	
35 / *47*	**51, 52** / *53, 54*	**23S:(1598-1602); 23S:(1341-1383)** / *23S:(1384-1415); 23S:(1584-1597)*	

6a / *48*	**2, 3, 4** / *24*	**23S:(17-45); 23S:(216-220); 23S:(234-236); 23S:(510-519); 23S:(431-447); 23S:(473-474)** / *23S:(475-509)*	
47 / *49*	**53, 54** / *55*	**23S:(1384-1415); 23S:(1584-1597)** / *23S:(1416-1428); 23S:(1569-1583)*	

7 / *50*	**94, 97** / *98, 100*	**23S:(2630-2645); 23S:(2771-2790); 23S:(2893-2894); 23S:(2733-2770)** / *23S:(2791-2831); 23S:(2884-2892)*	

49 / *51*	**55** / *56, 57*	**23S:(1416-1428); 23S:(1569-1583)** / *23S:(1429-1466); 23S:(1547-1568)*	
24 / *52*	**13, 14** / *18, 20*	**23S:(237-270); 23S:(369-370); 23S:(424-430)** / *23S:(271-298); 23S:(321-368)*	
41 / *53*	**6, 10** / *8, 9*	**23S:(50-70); 23S:(114-120); 23S:(149-174)** / *23S:(121-148)*	

20, 45 / *54*	**27, 28** / *29, 31*	**23S:(587-603); 23S:(653-671); 23S:(604-626)** / *23S:(627-652)*	
52 / *55*	**18, 20** / *19*	**23S:(271-298); 23S:(321-368)** / *23S:(299-320)*	
51 / *56*	**56, 57** / *58, 59*	**23S:(1429-1466); 23S:(1547-1568)** / *23S:(1467-1546)*	

37 / *57*	**62** / *63*	**23S:(1682-1706); 23S:(1757-1762)** / *23S:(1707-1756)*	
50 / *58*	**98, 100** / *101*	**23S:(2791-2831); 23S:(2884-2892)** / *23S:(2832-2883)*	
41 / *59*	**5, 6, 10** / *7*	**23S:(50-70); 23S:(114-120); 23S: 149-174)** / *23S:(71-113)*