

## Peeling the Onion: Ribosomes Are Ancient Molecular Fossils

Chiaolong Hsiao, Srividya Mohan, Benson K. Kalahar, and Loren Dean Williams

School of Chemistry and Biochemistry, Parker H. Petit Institute of Bioengineering and Bioscience, Georgia Institute of Technology

We describe a method to establish chronologies of ancient ribosomal evolution. The method uses structure-based and sequence-based comparison of the large subunits (LSUs) of *Haloarcula marismortui* and *Thermus thermophilus*. These are the highest resolution ribosome structures available and represent disparate regions of the evolutionary tree. We have sectioned the superimposed LSUs into concentric shells, like an onion, using the site of peptidyl transfer as the origin (the PT-origin). This spherical approximation combined with a shell-by-shell comparison captures significant information along the evolutionary time line revealing, for example, that sequence and conformational similarity of the 23S rRNAs are greatest near the PT-origin and diverge smoothly with distance from it. The results suggest that the conformation and interactions of both RNA and protein can be described as changing, in an observable manner, over evolutionary time. The tendency of macromolecules to assume regular secondary structural elements such as A-form helices with Watson–Crick base pairs (RNA) and  $\alpha$ -helices and  $\beta$ -sheets (protein) is low at early time points but increases as time progresses. The conformations of ribosomal protein components near the PT-origin suggest that they may be molecular fossils of the peptide ancestors of ribosomal proteins. Their abbreviated length may have proscribed formation of secondary structure, which is indeed nearly absent from the region of the LSU nearest the PT-origin. Formation and evolution of the early PT center may have involved  $Mg^{2+}$ -mediated assembly of at least partially single-stranded RNA oligomers or polymers. As one moves from center to periphery, proteins appear to replace magnesium ions. The LSU is known to have undergone large-scale conformation changes upon assembly. The *T. thermophilus* LSU analyzed here is part of a fully assembled ribosome, whereas the *H. marismortui* LSU analyzed here is dissociated from other ribosomal components. Large-scale conformational differences in the 23S rRNAs are evident from superimposition and prevent structural alignment of some portions of the rRNAs, including the L1 stalk.

### Introduction

The ribosome, which synthesizes protein in all living systems, is one of life's most ancient molecular machines. The ribosome is our most direct macromolecular connection to the distant evolutionary past and to early life. "Translation is not just another molecular structure to be solved. It represents, it is, the evolutionary transition from some kind of nucleic acid-based world to the protein-based world of modern cells" (Woese 2001). It is believed that the ribosome in its present form was well established before the last universal common ancestor of life (LUCA), that is, beyond the root of the phylogenetic tree (Fox and Ashinikumar 2004). Much of the diversity of conformation and sequence between bacterial and archaeal ribosomes is believed to predate the LUCA. The LUCA represents a primitive cellular population with a diverse gene pool in spite of low barriers to lateral gene transfer (Woese 2000; Baymann et al. 2003).

Understanding of the ribosome has been dramatically advanced by the recent determination of high-resolution, three-dimensional structures from disparate regions of the evolutionary tree. The current structural database contains experimentally determined structures from six distinct ribosomes: *Thermus thermophilus*, X-ray, 2.8 Å (Selmer et al. 2006); *Haloarcula marismortui*, X-ray, 2.4 Å, large subunit (LSU) only (Ban et al. 2000); *Escherichia coli*, X-ray, 3.2 Å (Berk et al. 2006); *Deinococcus radiodurans*, X-ray, 3.1 Å, LSU only (Harms et al. 2001); *Saccharomyces cerevisiae*, cryo-EM, 11.7 Å (Spahn et al. 2004); and bovine mitochondrion, cryo-EM, 13.5 Å (Sharma et al. 2003). The ancient ori-

gins of the ribosome, combined with the greater conservation of three-dimensional structure than sequence over evolutionary time (Heinz et al. 1994; Rost 1999), suggest that structures of ribosomes might allow detection and inference of deep and distant evolutionary events.

Comparison of linear rRNA sequences is a well-established method for determination of phylogenetic relationships. Woese and Fox (1977) used sequence data in their discovery of Archaea, the third kingdom of life (Magrum et al. 1978; Woese et al. 1978). Their results produced the phylogenetic tree that includes prokaryotes, protozoa, fungi, plants, and animals (Woese 1987). Over 10,000 16S and 16S-like rRNA and over 1,000 23S and 23S-like rRNA genes have been sequenced (Cannone et al. 2002).

Here we describe comparative structural methods to develop and test models of ancient ribosomal evolution. The results provide guideposts along the evolutionary time line and suggest the possibility of novel evolutionary clocks. Using three-dimensional structures, we have aligned the 23S rRNA<sup>HM</sup> (23S rRNA of *H. marismortui*) and 23S rRNA<sup>TT</sup> (23S rRNA of *T. thermophilus*) and have performed accurate and objective local and global superimpositions of the two LSUs. These LSUs are the highest resolution structures available. We have sectioned the superimposed LSUs into concentric shells, like an onion, using the site of peptidyl transfer as the origin (fig. 1). We approximate ribosomal evolution by accretion of spherical layers. The approximation is shown here to capture significant information along the evolutionary time line revealing, for example, that sequence and conformational similarity of these 23S rRNAs are greatest near the PT-origin and diverge smoothly with distance from it (i.e., with increasing spherical shell radius). The results are consistent with previous proposals (Fox and Ashinikumar 2004) that the LSU is oldest in evolution near the PT-origin and younger near the surface.

Key words: RNA evolution, ribosome, structural alignment, tetraloop.

E-mail: loren.williams@chemistry.gatech.edu.

*Mol. Biol. Evol.* 26(11):2415–2425. 2009

doi:10.1093/molbev/msp163

Advance Access publication July 23, 2009

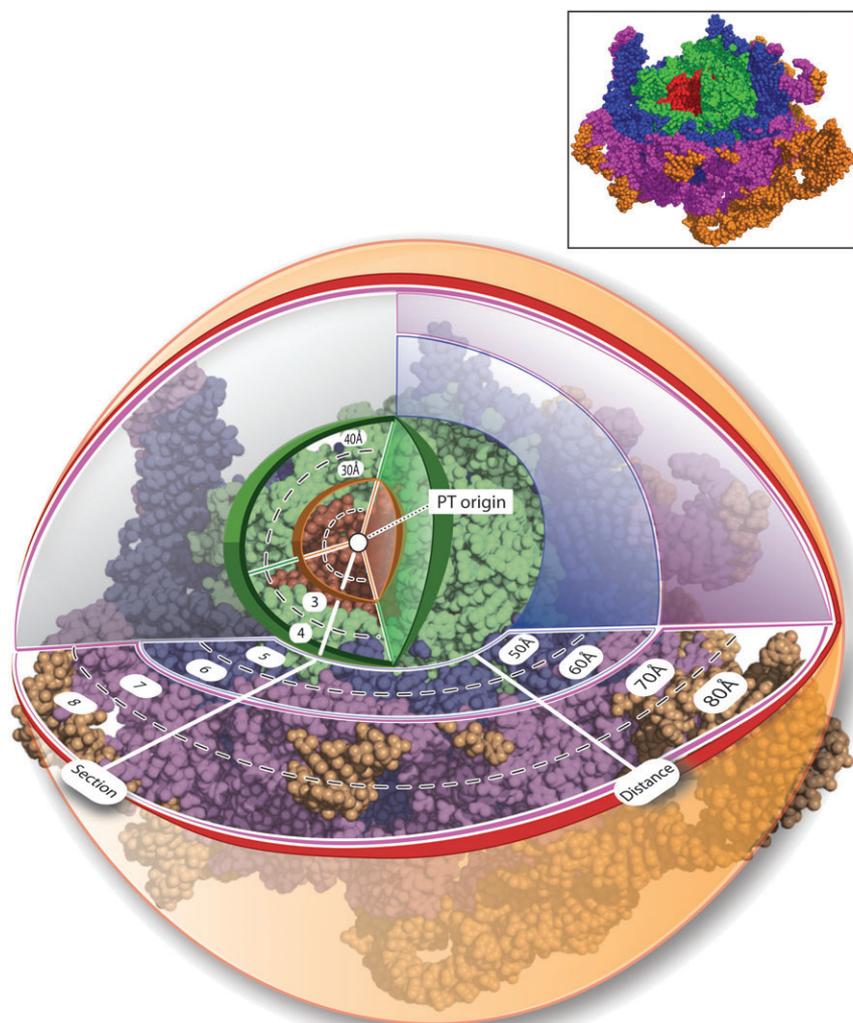


FIG. 1.—Peeling the ribosomal onion. The *Haloarcula marismortui* (and *Thermus thermophilus*, not shown) LSUs have been sectioned into concentric shells, with the origin at the site of peptidyl transfer (the PT-origin). The radius of any shell is 10 Å greater than the radius of the preceding shell. 23S rRNA<sup>HM</sup> is red in shells 1 and 2, green in shells 3 and 4, blue in shells 5 and 6, and purple in shells 7 and 8. Additional shells are shown in orange. Atoms are represented in spacefill. Ribosomal proteins, ions, and water molecules are omitted for clarity.

The “ribosome as onion” model helps explain ancient evolution and function from chemical and biophysical principles. Characteristics, such as 1) rRNA conformation, 2) rRNA base pairing interactions, 3) rRNA interactions with Mg<sup>2+</sup> ions, and 4) ribosomal protein conformation and interactions, vary with distance from the PT-origin. The results suggest that the conformation, environment, and interactions of both RNA and protein can be described as changing in an observable manner over evolutionary time. This information appears to have broad implications for the RNA World and origin of life models. The spherical analysis here is an approximation whose success should not be taken to indicate that the LSU literally evolved by the accretion of spherical layers.

## Materials and Methods

### PBR Space Analysis

Tetraloops in three-dimensional structures are detected here using a multiscaled pattern recognition approach de-

scribed previously (Hsiao et al. 2006). In this method, atomic positions are transformed into PBR space (P indicates phosphate, B indicates base, and R indicates ribose) where the resolution and complexity are attenuated in comparison to conventional all-atom representation. This change of scale reveals tetraloops adorned by four RNA deviations of local structure (DevLS), which are insertions, deletions, 3-2 switches, and strand clips. Visual inspection confirms that the PBR analysis identifies and classifies all tetraloops in the 3D structures of 23S rRNA<sup>HM</sup> and 23S rRNA<sup>TT</sup>.

### Structural Alignment

Structural alignment (SA) is iterated and optimized within each segment of the two large rRNAs (23S rRNA<sup>HM</sup> and 23S rRNA<sup>TT</sup>), followed by global rigid body superimposition of the entire 23S rRNAs and the entire LSUs. The process follows.

- 1) Identify all tetraloops in the 3D structures of 23S rRNA<sup>HM</sup> and 23S rRNA<sup>TT</sup>.

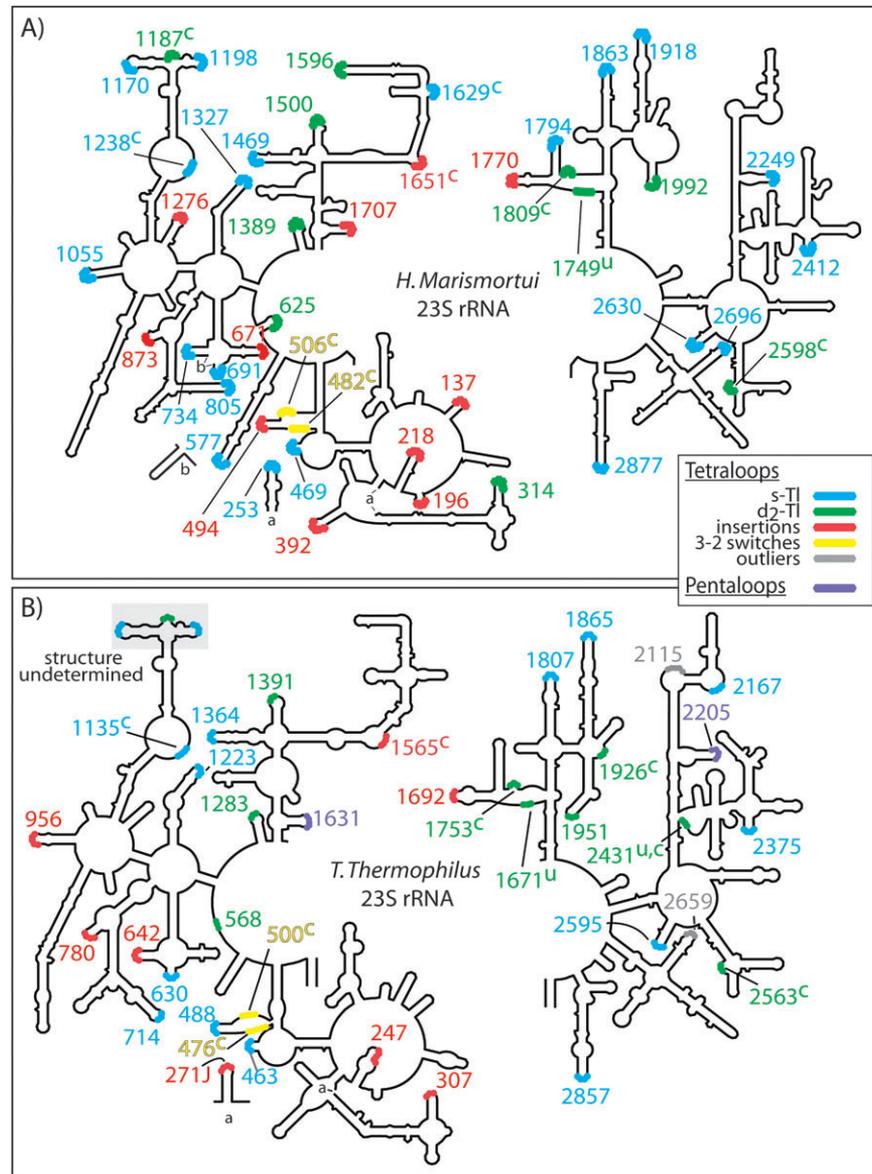


FIG. 2.—Secondary structures of (A) 23S rRNA<sup>HM</sup> and (B) 23S rRNA<sup>TT</sup>. Tetraloops are shaded and are colored by type, with cyan indicating standard tetraloop, green indicating deleted tetraloop, red indicating inserted tetraloop, yellow indicating 3-2 switch tetraloop, and the letter c indicating strand clipped tetraloop (as defined in reference Hsiao et al. 2006). Tetraloops are numbered by their first residue. In general, there is a close correspondence between the tetraloops of the two LSUs. However, there are exceptions. Inserted tetraloop 1707 in 23S rRNA<sup>HM</sup> corresponds to pentalooop 1631 in 23S rRNA<sup>TT</sup>. Standard tetraloop 2249 in 23S rRNA<sup>HM</sup> corresponds to pentalooop 2205 in 23S rRNA<sup>TT</sup>. Tetraloops 1651 of 23S rRNA<sup>HM</sup> and 1655 of 23S rRNA<sup>TT</sup> represent a newly discovered tetraloop subfamily that is elaborated with an insertion and a strand clip.

- 2) Mark the tetraloops (“anchors”) on the 2D representations of 23S rRNA<sup>HM</sup> and 23S rRNA<sup>TT</sup> (fig. 2) and confirm the correspondence of anchors between the two structures.
- 3) Define RNA segments, which terminate at anchors, and the correspondence (i.e., 1D pairing) of segments between the two rRNAs (fig. 3). The paired segments are homologous regions of RNA from 25 to 189 residues in length.
- 4) For each pair of RNA segments, determine the best SA and the locations of insertions and deletions using the following heuristical method.
  - a) The paired tetraloop anchors are corrected for subfamily differences (DevLS). Insertions within tetraloops are deleted, etc.
  - b) Paired segments of the same length are directly superimposed and visually inspected.
  - c) Heuristic fit: For paired segments that differ in length by one residue, each residue is systematically omitted from the longer segment, with a fit performed after each omission. The best fit determines the position of the insertion in the segment of greater length.
  - d) Enumerated heuristic fit: For paired segments that differ in length by two or three, residues are

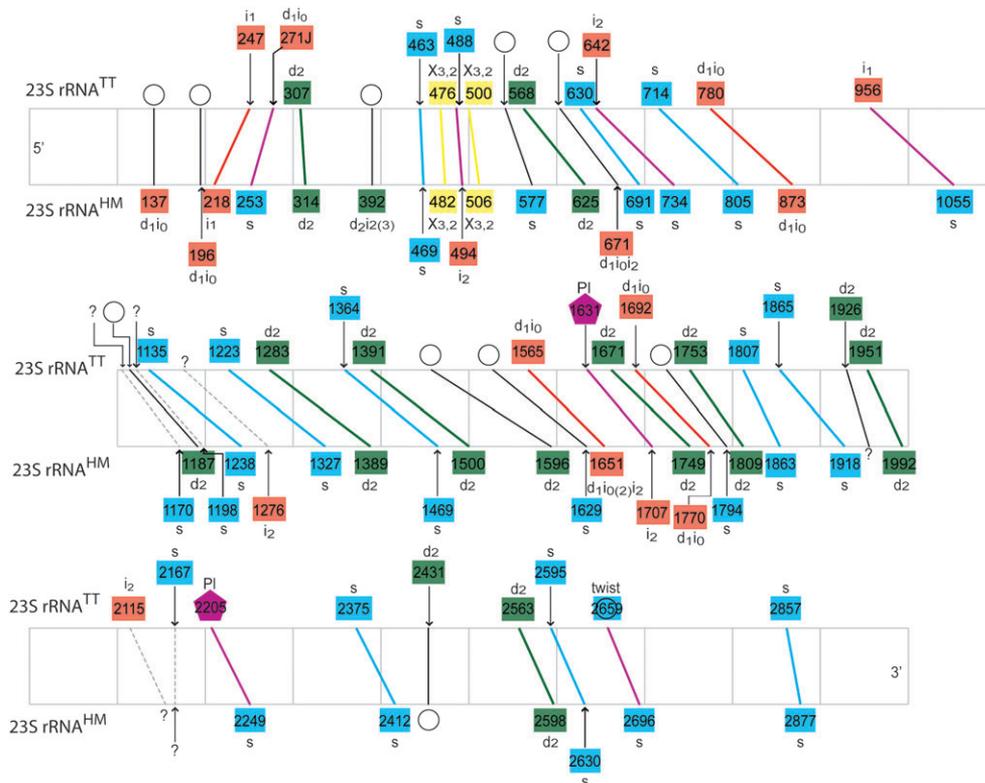


FIG. 3.—Mapping tetraloops between 23S rRNAs of bacteria and archaea. Tetraloops are proportionally spaced by number of rRNA nucleotides along the 23S rRNA<sup>TT</sup> (top) and 23S rRNA<sup>HM</sup> (bottom). Tetraloops are indicated by boxes and are numbered and colored as in figure 2. Pentaloops (PL) are indicated by pentagons. The correspondence of tetraloops/pentaloops between 23S rRNA<sup>HM</sup> and 23S rRNA<sup>TT</sup> is indicated by colored lines. The line color is the same as the tetraloop color where the line links tetraloops of the same type. Purple lines link different types of tetraloops or pentaloops. Dotted lines link to a region in one of the models that is missing that fraction of the rRNA, where the undetermined loop type is indicated by a question mark. The open circles indicate non-tetraloops.

enumerated from the longer segment, with a fit performed after each omission. For example, if the length difference is two, every pair of residues is stepwise omitted, in all combinations. The best fit determines the positions of the insertions. The workable limit for the RNA segment length difference is three residues.

- 5) Use the aligned pairs (APs) of rRNA for an initial rigid body superimposition of the LSUs. This initial superimposition employed 1,216 residues and 14,589 RNA backbone atoms (supplementary table 1S, Supplementary Material online).
- 6) Maximize the fraction of RNA used in the superimposition. A visual inspection of the two superimposed 23S rRNAs reveals that portions of nonaligned RNA are sufficiently similar that they can be included in the next iteration of the superimposition. “Secondary anchors” are placed, which define the limits of these regions of the rRNA. Secondary anchors are nontetraloop RNA anchors at sites where reasonable superimposition terminates, as determined by visual inspection. These secondary anchors are used to increase the amount of RNA used in the fit.
- 7) Return to step (v), iterate the rigid body superimposition including the additional RNA from step (vi).
- 8) Use all aligned rRNAs to perform a rigid body superimposition of the complete *H. marismortui* and

*T. thermophilus* LSUs, including nonaligned rRNA, proteins, ions, and solvent.

#### Creating the Onion

The site of peptidyl transfer, as determined by Steitz and coworkers (Ban et al. 2000), was set to be the origin (i.e., the PT-origin). RNA nucleotides were partitioned into concentric shells of 10 Å in width, centered on the PT-origin. Working out from the center, an rRNA residue is designated as a shell member if it contains one or more atoms within the boundaries of the shell. A single atom in a smaller radius shell allocates an entire residue to that shell (residues are not fragmented). Each residue is a member of one shell only, with priority to the innermost shell. The shell surfaces are not smooth. The number of 23S rRNA<sup>HM</sup> residues within each shell is shell 0–10 Å, 14 residues; 10–20 Å, 76 residues; 20–30 Å, 161 residues; 30–40 Å, 269 residues; 40–50 Å, 346 residues; 50–60 Å, 449 residues; 60–70 Å, 459 residues; 70–80 Å, 450 residues; 80–90 Å, 310 residues; 90–100 Å, 167 residues; 100–110 Å, 39 residues; and 110–120 Å, 5 residues. The number of *H. marismortui* amino acid residues within each shell is shell 0–10 Å, 0 residues; 10–20 Å, 2 residues; 20–30 Å, 45 residues; 30–40 Å, 144 residues; 40–50 Å, 259 residues; 50–60 Å, 376 residues; 60–70 Å, 593 residues; 70–80 Å, 792 residues; 80–90 Å, 703 residues; 90–100 Å, 441 residues; and 100–110 Å, 249 residues.

## Results

Tetraloops are employed here as *in silico* anchors, at which large RNAs are split into tractable segments. Tetraloops are terminal loops with characteristic four-residue sequences first observed in phylogenetic comparisons of RNAs (Woese et al. 1983, 1990; Tuerk et al. 1988). Tetraloops were seen to connect two antiparallel chains of double-helical RNA, and so cap A-form stems (Moore 1999), although “unhinged” tetraloops that do not cap helices have more recently been observed (Hsiao et al. 2006). Isolated stem/tetraloops 1) show well-defined structure and exceptional thermodynamic stabilities (Tuerk et al. 1988; Cheong et al. 1990; Varani et al. 1991; Antao and Tinoco 1992), 2) are thought to initiate folding of complex RNA molecules (Tuerk et al. 1988), 3) stabilize helical stems (Tuerk et al. 1988; Selinger et al. 1993), and 4) provide recognition elements for tertiary interactions and protein binding (Michel and Westhof 1990; Puglisi et al. 1992; Jaeger et al. 1994; Cate et al. 1996).

SA has been used here to determine how sequence and conformation within the LSUs of *H. marismortui* and *T. thermophilus* vary with location in three-dimensional space. SA as described here is a generally applicable process for objectively and accurately aligning and superimposing homologous RNAs and RNA–protein assemblies based on their three-dimensional structures. Here 73% of 23S rRNA<sup>HM</sup> and 23S rRNA<sup>TT</sup> (2,129 RNA residues, 25,545 backbone atoms) were successfully aligned. After global rigid body superimposition of the aligned RNA backbones, the root mean square deviation (RMSD) of atomic positions is 1.2 Å, a remarkably close fit considering the large number of atoms used in the fit, the vast evolutionary distance between *H. marismortui* and *T. thermophilus*, and differences in the states of assembly of the two ribosomes.

### Tetraloop Mapping

After identifying tetraloops (Hsiao et al. 2006) in 23S rRNA<sup>HM</sup> and 23S rRNA<sup>TT</sup> and annotating them on the 2D maps, we establish their correspondence in the two structures (fig. 2). The correspondence is close but not absolute. Tetraloop 218<sup>HM</sup> (fig. 2A) corresponds to tetraloop 247<sup>TT</sup> (fig. 2B), tetraloop 253<sup>HM</sup> corresponds to tetraloop 271<sup>TT</sup>, etc. Tetraloops 137<sup>HM</sup> and 196<sup>HM</sup> are absent from 23S rRNA<sup>TT</sup>. Tetraloop 1707<sup>HM</sup> corresponds to pentaloop 1631A<sup>TT</sup>, whereas tetraloop 2249<sup>HM</sup> corresponds to pentaloop 2205<sup>TT</sup>. A total of 47% of tetraloops are conserved in both position and type. A total of 63% of tetraloops are conserved in position but may vary in type (standard tetraloop vs. deleted tetraloop). These fractions are underestimates because certain regions of each LSU are undetermined or may contain errors.

### Alignment by Structure

Homologous segments of 23S rRNA<sup>HM</sup> and 23S rRNA<sup>TT</sup> were paired. A map of tetraloop correspondence in an one-dimensional representation (fig. 3) indicates how tetraloops serve as anchors to “divide” the 23S rRNAs

into segments of manageable length, the head, and tail of which are capped by tetraloops. Each RNA segment pair (one segment from 23S rRNA<sup>HM</sup> and one from 23S rRNA<sup>TT</sup>) was aligned via a heuristic determination of locations of inserted and deleted residues. The maximum length difference for alignment of a segment pair is three residues. Greater differences in length consume computational resources during the heuristic determination of insertions and deletions beyond that available.

### Superimposition

In an initial “conquer” process, 16 segments were aligned (supplementary table 1S, Supplementary Material online) and employed to calculate the initial fit. Visual inspection of the nonaligned RNA segments (that were omitted from the global fit) was performed to determine which might be alignable based on the initial superimposition. Secondary anchors were placed, defining the boundaries of these regions of RNA, where reasonable superimposition terminates (supplementary fig. 1S, Supplementary Material online). These secondary anchors (not tetraloops) increase the amount of aligned RNA. Using this additional RNA, the rigid body superimposition was thus iterated to achieve the final global superimposition (supplementary fig. 2S, Supplementary Material online).

### Local versus Global Superimposition

The alignment and superimposition process described here gives both local (segment level) and global (all aligned rRNAs) superimpositions. After local superimposition, deviations indicate differences in local conformation. After global superimposition, deviations include contributions from larger scale differences, such as net movement of segments. Small local and small global deviations of APs indicate that the local conformation and global position are conserved. In contrast, a small local and larger global deviations of APs indicate that the local conformation is conserved but that the position of the segment is different in the two ribosomes.

#### *AP<sub>14</sub>: Conserved Conformation and Position*

The longest aligned pair (AP<sub>14</sub>) is 189 residues in length in 23S rRNA<sup>HM</sup> and 191 residues in length in 23S rRNA<sup>TT</sup>. In 23S rRNA<sup>HM</sup>, this fragment starts at residue (G)2412 and ends at residue (A)2600. The heuristic fit indicates that residues (U)2431 and (A)2432 are insertions of 23S rRNA<sup>TT</sup> (or are deletions of 23S rRNA<sup>HM</sup>). These two residues were excluded from the alignment. The local superimposition of AP<sub>14</sub> gives an RMSD of backbone atomic positions of 1.23 Å (fig. 4), indicating that the local backbone conformation of this segment is highly conserved between *H. marismortui* and *T. thermophilus*. The global superimposition gives an RMSD of backbone atomic positions of 1.28 Å. The similarity of the local and global RMSDs indicates that the position of this segment is unchanged in 23S rRNA<sup>HM</sup> and 23S rRNA<sup>TT</sup>.

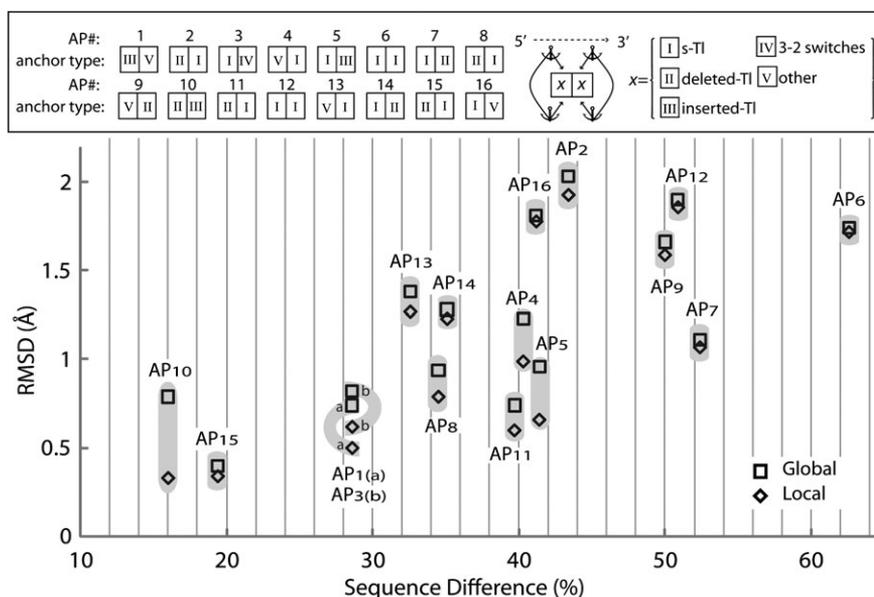


FIG. 4.—Sequence and Structure. The relationship between sequence divergence and conformational divergence for the segments of 23S rRNA<sup>HM</sup> and 23S rRNA<sup>TT</sup> aligned in this work. Shown is the relationship of sequence difference and RMSD of backbone atomic positions for 16 aligned pairs of rRNA. These rRNA segments are superimposed locally (diamonds) and globally (squares). In the boxed legend (top), the Roman Numerals in the rectangles indicate the types of tetraloops at the termini of the RNA segments.

#### AP<sub>10</sub>: Conserved Conformation with Differing Position

AP<sub>10</sub>, the shortest pair of segments aligned, is 25 residues in length in both rRNAs. In 23S rRNA<sup>HM</sup>, AP<sub>10</sub> starts at residue (U)1749 and ends at residue (G)1773. The heuristic fit confirms the absence of insertions and deletions. The local superimposition gives an RMSD of backbone atomic positions of 0.33 Å (fig. 4) indicating highly conserved conformation of these two segments. The global superimposition is 0.79 Å. The 0.46 Å difference between the local and global AP superimpositions indicates that the position of the segment is different in 23S rRNA<sup>HM</sup> and 23S rRNA<sup>TT</sup>. The difference may originate during folding or during ribosomal assembly. The location of this segment within Domain IV, at the LSU/SSU interface, with direct interactions with the 16S rRNA, suggests that the observed shift in position might occur during assembly (the *H. marismortui* and *T. thermophilus* LSUs are in different states of assembly). The *T. thermophilus* LSU (PDB entry: 2J00, 2J01) but not the *H. marismortui* LSU (PDB entry: 1JJ2) is part of a fully assembled ribosome.

#### Peeling the Onion

With the site of peptidyl transfer as the PT-origin, we have sectioned the superimposed *H. marismortui* and *T. thermophilus* LSUs into a series of concentric shells, each with thicknesses of 10 Å (fig. 1). The core region, the first shell, is a sphere of 10 Å radius. The second shell has an inner radius of 10 Å and an outer radius of 20 Å. This sectioning of the LSUs allows one to analyze how important characteristics of rRNA and other ribosomal components vary with distance from the PT-origin (i.e., with shell number). In principle, one can study sequence conservation (*H. marismortui* vs. *T. thermophilus*),

RNA conformational conservation, interactions with ions and water, RNA conformation, RNA modification, protein content and conformation, RNA–protein interactions, etc.

The extent of rRNA sequence conservation between 23S rRNA<sup>HM</sup> and 23S rRNA<sup>TT</sup> is high (>90%) within 10 Å of the peptidyl transferase center (PTC) (i.e., within the core region, fig. 5A). The extent of conservation falls to around 75% in the second shell, with an inner radius of 10 Å and an outer radius of 20 Å, then to less than 70% in the second shell. Moving outward from the PT-origin, sequence conservation continues to fall until around the fifth shell (which has an inner radius of 40 Å and an outer radius of 50 Å). From the fifth shell outward, the sequence conservation appears to plateau at around 60%.

Similarly, the conformations of 23S rRNA<sup>HM</sup> and 23S rRNA<sup>TT</sup>, as indicated by RMSD of atomic positions of the superimposed backbones, are very similar within the core region and first several shells (fig. 5B). The RMSDs of atomic positions are 0.7 (core), 0.5 (shell 2), 0.6 (shell 3), and 0.6 (shell 4). In the core region, the conformations of the two rRNAs differ more than predicted by the other shells. This elevated difference in the core regions appears to be caused by differences in the state of assembly of the two LSUs (Selmer et al. 2006). From shell 4 outward, the deviations of backbone atoms rise monotonically until the outer region of the LSU (shell 9). The deviations in the outer regions are underestimated in this analysis because significant portions of the outer shells are too divergent to superimpose.

The preferred state of the rRNA changes with distance from the PT-origin. Here results are given for the *H. marismortui* LSU although results are similar for the *T. thermophilus* LSU (data not shown). The propensity of rRNA to form base pairs increases with distance from

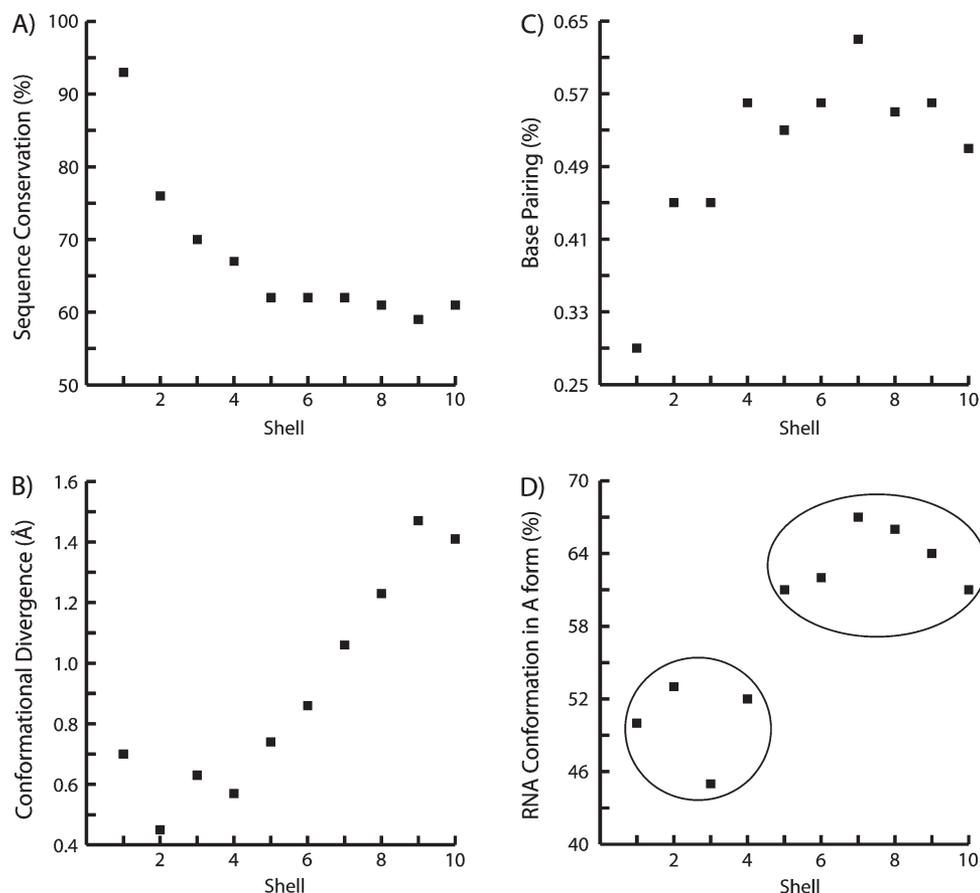


FIG. 5.—LSU as onion: analysis of 23S rRNA sequence, conformation, and base pairing. (A) The sequences of 23S rRNA<sup>HM</sup> and 23S rRNA<sup>TT</sup> diverge with distance from the PT-origin. In the core region around the PT-origin, the sequence is over 90% conserved. Beyond the fourth shell, the sequence conservation plateaus at around 60%. (B) The conformations of 23S rRNA<sup>HM</sup> and 23S rRNA<sup>TT</sup> highly conserved in the inner shells and increasingly diverge with distance from the PT-origin. Conformational divergence is indicated by the RMSD of the aligned backbone atoms after global superimposition. (C) The fraction of the rRNA involved in Watson–Crick base pairing is low near the PT-origin and increases with distance from the PT-origin. (D) The fraction of 23S rRNA<sup>HM</sup> in the A-conformation, defined by torsion angles, increases with distance from the PT-origin. In the first four shells, around 50% of rRNA is in the A-conformation. In the outer shells, around 60% of rRNA is in the A-conformation.

the origin (fig. 5C). Less than 30% of the bases in the core region of the *H. marismortui* LSU are engaged in Watson–Crick base pairs. The propensity of rRNA to form Watson–Crick base pairs increases with shell number until shells 4 and 5, where it plateaus at slightly less than 60% of bases paired. Base pairing is determined by Leontis et al. (2002).

RNA backbone conformational preferences are different in shells 1–4 from in shells 5 and greater (fig. 5D). RNA conformation is characterized here by torsion angles (see Hershkovitz et al. 2003, 2006; Richardson et al. 2008). In shells 1–4, the RNA is predominantly in conformations other than that characteristic of A-form helices. In shells 5 and greater, the RNA is predominantly (~60%) in A-form conformation.

The interactions of rRNA with Mg<sup>2+</sup> vary with distance from the PT-origin. “Mg<sup>2+</sup> density” is defined here as number of Mg<sup>2+</sup> ions with direct phosphate interactions per RNA residue. Mg<sup>2+</sup> density is greatest in the core region and falls off with increasing distance from the origin (fig. 6A). In the core region, there are around 0.21 Mg<sup>2+</sup> ions with direct phosphate interactions per rRNA nucleo-

tide. The ratio falls to nearly zero in the outer regions of the LSU.

The extent of interaction of ribosomal proteins with rRNA varies with distance from the PT-origin. “Ribosomal protein density” is defined as the number of ribosomal protein amino acids per rRNA nucleotide within a given shell. Protein density is at a minimum in the inner regions of the LSU and increases with increasing distance from the PT-origin (fig. 6B). The *H. marismortui* LSU, at least in the current model, lacks protein altogether in the core region (Ban et al. 2000; Klein et al. 2004).

We have found it useful to define an Mg<sup>2+</sup> dilution parameter, which is simply the reciprocal of the Mg<sup>2+</sup> density (fig. 6C). The core region of the LSU is characterized by the greatest Mg<sup>2+</sup> density (the least Mg<sup>2+</sup> dilution) and the lowest ribosomal protein density. A near linear relationship between Mg<sup>2+</sup> dilution and ribosomal protein density is observed. Moving out from the PT-origin, as Mg<sup>2+</sup> is diluted, ribosomal protein density increases.

Ribosomal protein conformation varies with distance from the PT-origin (fig. 6D). The protein observed in

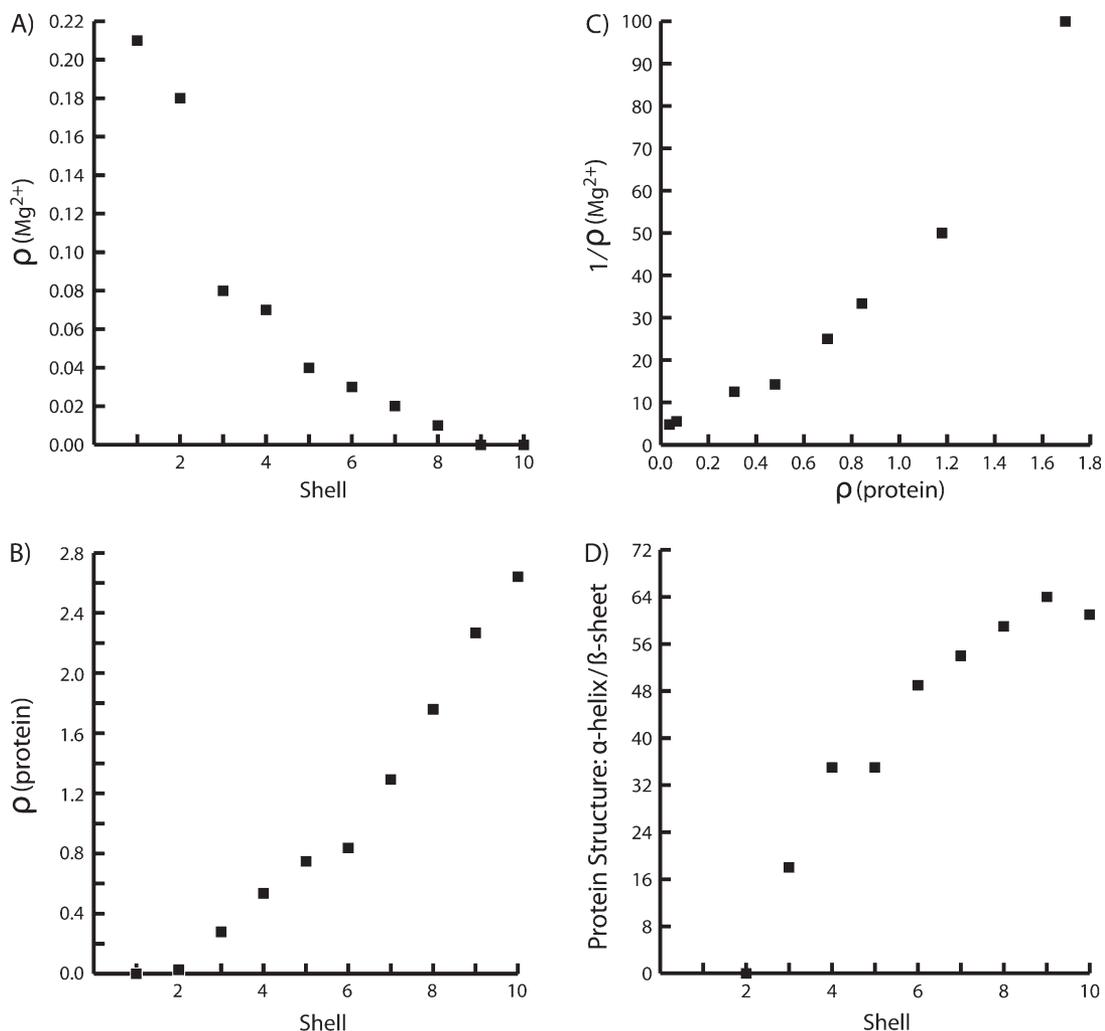


FIG. 6.—LSU as onion: analysis of rRNA interactions with proteins and ions within the LSU of *Haloarcula marismortui*. (A) Mg<sup>2+</sup> density decreases with distance from the PT-origin. Mg<sup>2+</sup> density is given by the number of Mg<sup>2+</sup> ions with inner shell phosphate ligands normalized to the number of nucleotides. (B) Ribosomal protein density increases with distance from the PT-origin. Ribosomal protein density is given by the number of amino acids normalized to the number of nucleotides. (C) Ribosomal protein density within the LSU increases with Mg<sup>2+</sup> dilution. Mg<sup>2+</sup> dilution is given by the reciprocal of the Mg<sup>2+</sup> density. (D) Ribosomal proteins near the PT-origin are unlikely to form secondary structure ( $\alpha$ -helices and  $\beta$ -sheets), as defined by torsion angles phi and psi. The fraction of protein involved in secondary structure increases with distance from the PT-origin.

the second shell lacks secondary structure, although there is very little protein there (two amino acids in the *H. marismortui* LSU). In the third shell, the torsion angles of only 20% of amino acids are consistent with  $\alpha$ -helix or  $\beta$ -sheet. The fraction of protein in  $\alpha$ -helix or  $\beta$ -sheet increases smoothly with increasing distance from the PT-origin to around 60% in the outer regions of the LSU.

#### Structure-Based and Sequence-Based Alignments

In many respects, structural (here) and sequence-based (Cannone et al. 2002) alignment methods give similar results. For example, both the structure-based and sequence-based translation tables show residue (C)2701<sup>TT</sup> to be an insertion (in comparison to 23S rRNA<sup>HM</sup>). However, in some locations, there are clear differences between the structure-based and sequence-based alignment tables. As illustrated in supplementary figure 3S (Supplementary

Material online), (U)2701<sup>TT</sup> and (U)2702<sup>TT</sup> appear to be insertions in three dimensions, where (C)2737<sup>HM</sup> aligns best with (C)2700<sup>TT</sup>, whereas (G)2738<sup>HM</sup> aligns best with (C)2703<sup>TT</sup>. By contrast, in the sequence-based alignment (Cannone et al. 2002), (G)2738<sup>HM</sup> aligns best with (U)2702<sup>TT</sup>.

The results of SA should allow one to apply additional constraints to increase the accuracy and extent of sequence alignment. SA can provide information that is absent or ambiguous in the sequence alignment. Facile alignment by structure of some segments is achieved even when there is no obvious alignment of sequence. For example, residues 236–242 of 23S rRNA<sup>HM</sup> are aligned with residues 265–271 of 23S rRNA<sup>HM</sup> by both the structural and the sequence-based methods. The sequence-based alignment terminates at residues A242<sup>HM</sup>/A271<sup>TT</sup>, whereas the structure-based alignment continues for six residues beyond residues A242<sup>HM</sup> and A271<sup>TT</sup>. Thus, the alignment table obtained from structure can be used to extend the alignment table obtained from linear sequence.

## Discussion

The comparative approach allows one to reconstruct history by degree of similarity between homologous genes, proteins, or RNA (Lio and Goldman 1998; Templeton 2001). Such comparisons are most commonly made on the basis of sequence. The rRNA sequence, because of conservation over long evolutionary times, allows inference of ancient events. Structures of ribosomes (i.e., three-dimensional X-ray structures) can provide information on the earliest events because structure changes more slowly than sequence over evolutionary time (Heinz et al. 1994; Rost 1999). Here relationships between sequence, conformation, and molecular interactions of two LSUs are determined using SA and superimposition, combined with a spherical approximation that allows comparison on the basis of internal location.

The LSUs of *H. marismortui* and *T. thermophilus* were accurately and objectively superimposed using over 70% of 23S rRNA backbone atoms. The site of peptidyl transferase was defined as the origin (i.e., the PT-origin). The superimposed LSUs were converted to “onions” by sectioning into 10 Å shells of 0, 10, 20, 30 Å... radius, with each shell 10 Å thick, and centered on the PT-origin (fig. 1). RNA sequence, conformation, and ion binding were characterized within each shell, as was ribosomal protein conformation and abundance. The ancestral peptidyl transferase activity is thus modeled as a sphere, which increased in size by accretion of shells during evolution. Although in reality neither the LSU nor its early progenitors are spherical, the approximation is shown here to be close enough to reality to expose clear and interpretable variation between shells.

### rRNA Sequence, rRNA Conformation, Ions, Proteins, and Time

Shell-dependent patterns of 23S rRNA sequence, conformation, and interactions suggest, as anticipated, that rRNA is evolutionarily oldest on average near the PT-origin and decreases in age with distance from the PT-origin (i.e., with shell radius). Relative evolutionary age is indicated by comparison of sequences of 23S rRNA<sup>HM</sup> and 23S rRNA<sup>TT</sup>, which are most conserved near the PT-origin, and increasingly diverge with distance from the PT-origin (fig. 5A). Likewise, the conformations of 23S rRNA<sup>HM</sup> and 23S rRNA<sup>TT</sup> are most similar near the PT-origin and increasingly diverge with distance from the PT-origin (fig. 5B). The extreme conservation of sequence and conformation near the PT-origin is consistent with rigorous requirements for function. Base Pairing: The propensity to form base pairs in 23S rRNA is different near the PT-origin than in more remote regions of the LSU. The frequency of base pairing (per nucleotide) is low near the PT-origin and increases until the fourth shell, after which it plateaus at around 60%. RNA Conformation: The conformational preference of 23S rRNA is different near the PT-origin (in the first four shells) than in more remote regions of the LSU (fig. 5D). A substantial proportion of the rRNA near the PT-origin is found to be in diverse and unusual conformations, not in A-form conformation. In contrast,

the more remote shells are predominantly (around 60%) in A-conformation. RNA-Mg<sup>2+</sup> Interactions: The interactions of rRNA with Mg<sup>2+</sup> near the PT-origin differ from those in the more remote regions. Near the PT-origin, phosphate oxygens more frequently act as inner sphere Mg<sup>2+</sup> ligands (fig. 6A). As distance from the PT-origin increases, the frequency of direct Mg<sup>2+</sup>-phosphate interactions decreases. Mg<sup>2+</sup> ions that interact directly with phosphate oxygens are particularly important in RNA structure and assembly (Hsiao et al. 2008; Hsiao and Williams 2009). Ribosomal Proteins: The density of ribosomal proteins (density = amino acid/nucleotide) varies with distance from the PT-origin. Ribosomal proteins are observed with the greatest density in the remote regions of the LSU but are absent near the PT-origin (fig. 6B). The *H. marismortui* LSU lacks protein in the core region. Only one amino acid (Met-1 of ribosomal protein L27) is observed in the core region of the *T. thermophilus* LSU. Mg<sup>2+</sup> dilution: As distance from the PT-origin increases, Mg<sup>2+</sup> density decreases, whereas ribosomal protein density increases (fig. 6C). The near linear relationship between Mg<sup>2+</sup> dilution and ribosomal protein density appears to illuminate fundamental relationships within the LSU (fig. 6C). It appears that interactions of rRNA with Mg<sup>2+</sup> ions are effectively replaced by those with ribosomal proteins, with increasing distance from the origin. Protein Conformation: Ribosomal protein conformation within the LSU is different near the PT-origin than in more remote regions of the LSU (fig. 6D). Ribosomal proteins in the inner regions of the LSU do not form  $\alpha$ -helices or  $\beta$ -sheets. The fraction of protein residues found within  $\alpha$ -helices and  $\beta$ -sheets increases with distance from the PT-origin.

### Models of Ribosomal Evolution

The results here broadly support the model of ribosomal evolution proposed by Fox and Ashinikumar (2004). In Fox's model, small RNAs with RNA aminoacylation activity evolved into portions of the aboriginal PT center (23S Domain V). In this model of ribosomal evolution, the initial PT center catalyzed nonspecific (nontemplated) synthesis of short peptides.

The conformations of ribosomal protein components near the PT-origin suggest that they are molecular fossils of peptide ancestors whose short length proscribed secondary structure, which is indeed absent from the region of the LSU nearest the PT-origin. Formation and early evolution of the PTC appears to have involved Mg<sup>2+</sup>-mediated assembly of single-stranded RNA oligomers or polymers. We observe a low frequency of base pairing near the PT-origin, along with a high frequency of inner sphere phosphate-Mg<sup>2+</sup> interactions and a preference against A-form conformation. It is known that Mg<sup>2+</sup> ions bind preferentially to single-stranded RNA over double-stranded RNA (Kankia 2003) and associate preferentially with non-A-form RNA conformations (Klein et al. 2004; Hsiao et al. 2008). In sum, results here are consistent with observations of Steitz and coworkers (Klein et al. 2004), who noted that Mg<sup>2+</sup> ions in the LSU are most abundant in the region surrounding the peptidyl transferase center,

and suggested that unusual RNA conformations, stabilized by  $Mg^{2+}$ , are molecular fossils.

A time line of ancestral RNA addition to the LSU proposed by Bokov and Steinberg (2009) using analysis of A-minor interactions corresponds closely with more course-grained models we inferred (Hsiao et al. 2008; Hsiao and Williams 2009) from  $Mg^{2+}$  interactions and which Gutell and Harvey deduced via phylogeny (Mears et al. 2002). This correspondence of results from truly orthogonal methods supports the validity of the consensus result.

#### Alignment by Structure/Alignment by Sequence

The SA allows superimposition of 73% of rRNA 23S backbone atoms of 23S rRNA<sup>HM</sup> and 23S rRNA<sup>TT</sup> to give an overall RMSD of 1.2 Å. Sequence and backbone conformation are experimentally orthogonal, in that they are determined independently of each other and contain non-overlapping information. In general, insertions and deletions identified by sequence alignment correspond to insertions and deletions observed in three dimensions, where insertions generally cause local perturbations in conformation that do not propagate over greater distances. Combining sequence with structural information appears to increase the alignable fraction of the rRNA and the accuracy of the alignment.

#### Relationships between Sequence and Conformation

The LSU is known to undergo large-scale conformation changes upon assembly and tRNA binding (Korostelev and Noller 2007). The *T. thermophilus* LSU is part of a fully assembled ribosome, whereas the *H. marismortui* LSU is dissociated from other ribosomal components. LSU rearrangements include movement of the L1 stalk upon interaction with the E-tRNA. Those large-scale conformational differences in 23S rRNA<sup>HM</sup> and 23S rRNA<sup>TT</sup> prevent SA of some portions of the rRNAs, including the L1 stalk.

The SA highlights more subtle conformational differences between 23S rRNA<sup>HM</sup> and 23S rRNA<sup>TT</sup>, which appear to be related primarily to differences in sequence. Local structural divergence between 23S rRNA<sup>HM</sup> and 23S rRNA<sup>TT</sup> generally increases with the extent of local divergence of 23S within the sequence database (supplementary fig. 4S, Supplementary Material online). Here we are using sequence divergence as determined by Gutell and coworkers (Cannone et al. 2002) who performed covariation analysis of over 500 23S ribosomal sequences, from the three phylogenetic kingdoms, along with mitochondria and chloroplasts. The relationship of local sequence divergence to structural divergence (between 23S rRNA<sup>HM</sup> and 23S rRNA<sup>TT</sup>) is summarized in supplementary figure 4S (Supplementary Material online), where the 2D map of 23S rRNA<sup>TT</sup> is annotated to indicate the degree of sequence conservation (from Gutell) and is also colored by local RMSD of atomic positions of 23S rRNA<sup>TT</sup> versus 23S rRNA<sup>HM</sup>. Regions of rRNA with the lowest sequence divergence generally show the smallest local structural divergence. The region with the greatest sequence divergence has the largest structural divergence; however, the relationship is complex and the signal is noisy. Some RNA regions

with relatively low sequence conservation (40%) show highly conserved structure (RMSD of atomic positions of backbone atoms of 0.6 Å). For example in Domain IV, helices 64, 65, and 66 (AP<sub>11</sub>, supplementary table 1S and fig. 4S, Supplementary Material online), the structure is more conserved than predicted by the sequence. In domain II, helix 27 (AP<sub>4</sub>), the sequences are highly divergent, whereas the structures are conserved (global RMSD: 0.61 Å). By contrast in Domain I, helix 13 (AP<sub>1</sub>), the sequences are conserved, whereas the structures are divergent (local RMSD: 10.6 Å).

#### Summary

We describe a method to establish chronologies of ancient ribosomal evolution. The method uses structure-based and sequence-based comparisons of the LSUs of *H. marismortui* and *T. thermophilus*. The results suggest that the conformation and interactions of both RNA and protein change, in an observable manner, over evolutionary time.

#### Supplementary Material

Supplementary figures 1S–4S and table 1S are available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>).

#### Acknowledgments

The authors thank Jessica Bowman and Drs Steve Harvey, Nicholas Hud, Roger Wartell, and Andrew Huang for helpful discussions. This work was supported by NASA Astrobiology Institute.

#### Literature Cited

- Antao VP, Tinoco I Jr. 1992. Thermodynamic parameters for loop formation in RNA and DNA hairpin tetraloops. *Nucleic Acids Res.* 20:819–824.
- 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.
- Baymann F, Lebrun E, Brugna M, Schoepp-Cothenet B, Giudici-Orticoni MT, Nitschke W. 2003. The redox protein construction kit: pre-last universal common ancestor evolution of energy-conserving enzymes. *Philos Trans R Soc Lond B Biol Sci.* 358:267–274.
- Berk V, Zhang W, Pai RD, Cate JH. 2006. Structural basis for mRNA and tRNA positioning on the ribosome. *Proc Natl Acad Sci USA.* 103:15830–15834.
- Bokov K, Steinberg SV. 2009. A hierarchical model for evolution of 23S ribosomal RNA. *Nature.* 457:977–980.
- Cannone JJ, Subramanian S, Schnare MN, et al. (14 co-authors). 2002. The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinformatics.* 3:2.
- Cate JH, Gooding AR, Podell E, Zhou K, Golden BL, Kundrot CE, Cech TR, Doudna JA. 1996. Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science.* 273:1678–1685.

- Cheong C, Varani G, Tinoco I Jr. 1990. Solution structure of an unusually stable RNA hairpin, 5' GGAC(UUCG)GUCC. *Nature*. 346:680–682.
- Fox GE, Ashinikumar KN. 2004. The evolutionary history of the translation machinery. In: de Pouplana LR, editor. *The genetic code and the origin of life*. New York: Kluwer Academic / Plenum. p. 92–105.
- Harms J, Schluenzen F, Zarivach R, Bashan A, Gat S, Agmon I, Bartels H, Franceschi F, Yonath A. 2001. High resolution structure of the large ribosomal subunit from a mesophilic eubacterium. *Cell*. 107:679–688.
- Heinz DW, Baase WA, Zhang XJ, Blaber M, Dahlquist FW, Matthews BW. 1994. Accommodation of amino acid insertions in an alpha-helix of T4 lysozyme. Structural and thermodynamic analysis. *J Mol Biol*. 236:869–886.
- Hershkovitz E, Sapiro G, Tannenbaum A, Williams LD. 2006. Statistical analysis of the RNA backbone. *IEEE/ACM Trans Comput Biol Bioinform*. 3:33–46.
- Hershkovitz E, Tannenbaum E, Howerton SB, Sheth A, Tannenbaum A, Williams LD. 2003. Automated identification of RNA conformational motifs: theory and application to the HM LSU 23S rRNA. *Nucleic Acids Res*. 31:6249–6257.
- Hsiao C, Mohan S, Hershkovitz E, Tannenbaum A, Williams LD. 2006. Single nucleotide RNA choreography. *Nucleic Acids Res*. 34:1481–1491.
- Hsiao C, Tannenbaum M, VanDeusen H, Hershkovitz E, Perng G, Tannenbaum A, Williams LD. 2008. Complexes of nucleic acids with group I and II cations. In: Hud N, editor. *Nucleic acid metal ion interactions*. London: The Royal Society of Chemistry. p. 1–35.
- Hsiao C, Williams LD. 2009. A recurrent magnesium-binding motif provides a framework for the ribosomal peptidyl transferase center. *Nucleic Acids Res*. 37:3134–3142.
- Jaeger L, Michel F, Westhof E. 1994. Involvement of a GNRA tetraloop in long-range tertiary interactions. *J Mol Biol*. 236:1271–1276.
- Kankia BI. 2003. Binding of Mg<sup>2+</sup> to single-stranded polynucleotides: hydration and optical studies. *Biophys Chem*. 104:643–654.
- Klein DJ, Moore PB, Steitz TA. 2004. The contribution of metal ions to the structural stability of the large ribosomal subunit. *RNA*. 10:1366–1379.
- Korostelev A, Noller HF. 2007. The ribosome in focus: new structures bring new insights. *Trends Biochem Sci*. 32:434–441.
- Leontis NB, Stombaugh J, Westhof E. 2002. The non-Watson-Crick base pairs and their associated isostericity matrices. *Nucleic Acids Res*. 30:3497–3531.
- Lio P, Goldman N. 1998. Models of molecular evolution and phylogeny. *Genome Res*. 8:1233–1244.
- Magrum LJ, Luehrsen KR, Woese CR. 1978. Are extreme halophiles actually “bacteria”? *J Mol Evol*. 11:1–8.
- 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.
- Michel F, Westhof E. 1990. Modeling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. *J Mol Biol*. 216:585–610.
- Moore PB. 1999. Structural motifs in RNA. *Annu Rev Biochem*. 68:287–300.
- Puglisi JD, Tan R, Calnan BJ, Frankel AD, Williamson JR. 1992. Conformation of the TAR RNA-arginine complex by NMR spectroscopy. *Science*. 257:76–80.
- Richardson JS, Schneider B, Murray LW, et al. (15 co-authors). 2008. RNA backbone: consensus all-angle conformers and modular string nomenclature (an RNA Ontology Consortium contribution). *RNA*. 14:1–17.
- Rost B. 1999. Twilight zone of protein sequence alignments. *Protein Eng*. 12:85–94.
- Selinger D, Liao X, Wise JA. 1993. Functional interchangeability of the structurally similar tetranucleotide loops GAAA and UUCG in fission yeast signal recognition particle RNA. *Proc Natl Acad Sci USA*. 90:5409–5413.
- Selmer M, Dunham CM, Murphy FV, Weixlbaumer A 4th, Petry S, Kelley AC, Weir JR, Ramakrishnan V. 2006. Structure of the 70S ribosome complexed with mRNA and tRNA. *Science*. 313:1935–1942.
- Sharma MR, Koc EC, Datta PP, Booth TM, Spremulli LL, Agrawal RK. 2003. Structure of the mammalian mitochondrial ribosome reveals an expanded functional role for its component proteins. *Cell*. 115:97–108.
- Spahn CM, Gomez-Lorenzo MG, Grassucci RA, Jorgensen R, Andersen GR, Beckmann R, Penczek PA, Ballesta JP, Frank J. 2004. Domain movements of elongation factor eEF2 and the eukaryotic 80S ribosome facilitate tRNA translocation. *EMBO J*. 23:1008–1019.
- Templeton AR. 2001. Using phylogeographic analyses of gene trees to test species status and processes. *Mol Ecol*. 10:779–791.
- Tuerk C, Gauss P, Thermes C, et al. (11 co-authors). 1988. CUUCGG hairpins: extraordinarily stable RNA secondary structures associated with various biochemical processes. *Proc Natl Acad Sci USA*. 85:1364–1368.
- Varani G, Cheong C, Tinoco I Jr, Wimberly B. 1991. Structure of an unusually stable RNA hairpin—conformation and dynamics of an RNA internal loop. *Biochemistry*. 30:3280–3289.
- Woese CR. 1987. Bacterial evolution. *Microbiol Rev*. 51:221–271.
- Woese CR. 2000. Interpreting the universal phylogenetic tree. *Proc Natl Acad Sci USA*. 97:8392–8396.
- Woese CR. 2001. Translation: in retrospect and prospect. *RNA*. 7:1055–1067.
- Woese CR, Fox GE. 1977. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci USA*. 74:5088–5090.
- Woese CR, Gutell R, Gupta R, Noller HF. 1983. Detailed analysis of the higher-order structure of 16S-like ribosomal ribonucleic acids. *Microbiol Rev*. 47:621–669.
- Woese CR, Magrum LJ, Fox GE. 1978. Archaeobacteria. *J Mol Evol*. 11:245–251.
- Woese CR, Winker S, Gutell RR. 1990. Architecture of ribosomal-RNA—constraints on the sequence of tetra-loops. *Proc Natl Acad Sci USA*. 87:8467–8471.

Michele Vendruscolo, Associate Editor

Accepted June 23, 2009