RESEARCH ARTICLES

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 $z$-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, and 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 Ashinikum 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 (Ber 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 origins 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, protzoa, 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$^\text{HM}$ (23S RNA of H. marismortui) and 23S rRNA$^\text{TT}$ (23S RNA 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 Ashinikum 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.

doi:10.1093/molbev/msp163
Advance Access publication July 23, 2009

© The Author 2009. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution. All rights reserved.
For permissions, please e-mail: journals.permissions@oxfordjournals.org
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\(^{2+}\) 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 described 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\(^{\text{HM}}\) and 23S rRNA\(^{\text{TT}}\).

**Structural Alignment**

Structural alignment (SA) is iterated and optimized within each segment of the two large rRNAs (23S rRNA\(^{\text{HM}}\) and 23S rRNA\(^{\text{TT}}\)), 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\(^{\text{HM}}\) and 23S rRNA\(^{\text{TT}}\).
2) Mark the tetraloops (“anchors”) on the 2D representations of 23S rRNAHM and 23S rRNATT (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
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 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. thermophillus 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 rRNAHM and 23S rRNATT (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. thermophillus, and differences in the states of assembly of the two ribosomes.

Tetraloop Mapping

After identifying tetraloops (Hsiao et al. 2006) in 23S rRNAHM and 23S rRNATT and annotating them on the 2D maps, we establish their correspondence in two structures (fig. 2). The correspondence is close but not absolute. Tetraloop 218HM (fig. 2A) corresponds to tetraloop 247TT (fig. 2B), tetraloop 253HM corresponds to tetraloop 271TT, etc. Tetraloops 137HM and 196HM are absent from 23S rRNAHM, and 196TT is absent from 23S rRNATT. Tetraloop 170TT corresponds to pentaloop 1631TT, whereas tetraloop 2249HM corresponds to pentaloop 2205TT. 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 rRNAHM and 23S rRNATT were paired. A map of tetraloop correspondence (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 rRNAHM and one from 23S rRNATT) 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.

AP14: Conserved Conformation and Position

The longest aligned pair (AP14) is 189 residues in length in 23S rRNAHM and 191 residues in length in 23S rRNATT. In 23S rRNAHM, 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 rRNATT (or are deletions of 23S rRNAHM). These two residues were excluded from the alignment. The local superimposition of AP14 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. thermophillus. 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 rRNAHM and 23S rRNATT.
AP10: Conserved Conformation with Differing Position

AP10, the shortest pair of segments aligned, is 25 residues in length in both rRNAs. In 23S rRNAHM, AP10 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 rRNAHM and 23S rRNA TT. 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 (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. 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 configuration, RNA modification, protein content and conformation, RNA–protein interactions, etc.

The extent of rRNA sequence conservation between 23S rRNAHM and 23S rRNA TT 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 rRNAHM and 23S rRNA TT, 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. 4.—Sequence and Structure. The relationship between sequence divergence and conformational divergence for the segments of 23s rRNAHM and 23S rRNA TT 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.
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-conformation.

The interactions of rRNA with Mg$^{2+}$ vary with distance from the PT-origin. “Mg$^{2+}$ density” is defined here as number of Mg$^{2+}$ ions with direct phosphate interactions per RNA residue. Mg$^{2+}$ 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$^{2+}$ ions with direct phosphate interactions per RNA nucleotide. 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$^{2+}$ dilution parameter, which is simply the reciprocal of the Mg$^{2+}$ density (fig. 6C). The core region of the LSU is characterized by the greatest Mg$^{2+}$ density (the least Mg$^{2+}$ dilution) and the lowest ribosomal protein density. A near linear relationship between Mg$^{2+}$ dilution and ribosomal protein density is observed. Moving out from the PT-origin, as Mg$^{2+}$ is diluted, ribosomal protein density increases.

Ribosomal protein conformation varies with distance from the PT-origin (fig. 6D). The protein observed in
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 an α-helix or β-sheet. The fraction of protein in α-helix or β-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)2701TT to be an insertion (in comparison to 23S rRNAHM). 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)2701TT and (U)2702TT appear to be insertions in three dimensions, where (C)2737HM aligns best with (C)2700TT, whereas (G)2738HM aligns best with (C)2703TT. By contrast, in the sequence-based alignment (Cannone et al. 2002), (G)2738HM aligns best with (U)2702TT.

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 rRNAHM are aligned with residues 265–271 of 23S rRNAHM by both the structural and the sequence-based methods. The sequence-based alignment terminates at residues A242HM/A271TT, whereas the structure-based alignment continues for six residues beyond residues A242HM and A271TT. 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 increases 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*TM* and 23S rRNA*TT*, 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*HM* and 23S rRNA*TT* 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*2+* Interactions: The interactions of tRNA with Mg*2+* 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*2+* ligands (fig. 6A). As distance from the PT-origin increases, the frequency of direct Mg*2+*-phosphate interactions decreases. Mg*2+* 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*2+* dilution: As distance from the PT-origin increases, Mg*2+* density decreases, whereas ribosomal protein density increases (fig. 6C). The near linear relationship between Mg*2+* dilution and ribosomal protein density appears to illuminate fundamental relationships within the LSU (fig. 6C). It appears that interactions of tRNA with Mg*2+* 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 α-helices or β-sheets. The fraction of protein residues found within α-helices and β-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*2+*-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*2+* interactions and a preference against A-form conformation. It is known that Mg*2+* 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*2+* ions in the LSU are most abundant in the region surrounding the peptidyl transferase center.
and suggested that unusual RNA conformations, stabilized by Mg\textsuperscript{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\textsuperscript{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\textsuperscript{HM} and 23S rRNA\textsuperscript{TT} 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\textsuperscript{HM} and 23S rRNA\textsuperscript{TT} prevent SA of some portions of the rRNAs, including the L1 stalk.

The SA highlights more subtle conformational differences between 23S rRNA\textsuperscript{HM} and 23S rRNA\textsuperscript{TT}, which appear to be related primarily to differences in sequence. Local structural divergence between 23S rRNA\textsuperscript{HM} and 23S rRNA\textsuperscript{TT} 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 co-variation 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\textsuperscript{HM} and 23S rRNA\textsuperscript{TT}) is summarized in supplementary figure 4S (Supplementary Material online), where the 2D map of 23S rRNA\textsuperscript{TT} is annotated to indicate the degree of sequence conservation (from Gutell) and is also colored by local RMSD of atomic positions of 23S rRNA\textsuperscript{TT} versus 23S rRNA\textsuperscript{HM}. Regions of tRNA 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\textsubscript{11}, 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\textsubscript{1}), the sequences are highly divergent, whereas the structures are conserved (global RMSD: 0.61 Å). By contrast in Domain I, helix 13 (AP\textsubscript{1}), 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


