Published online 25 January 2013 *Nucleic Acids Research, 2013, Vol. 41, No. 5* ***3373–3385***
doi:10.1093/nar/gkt023

Molecular paleontology: a biochemical model of the ancestral ribosome

Chiaolong Hsiao[1,2,3], Timothy K. Lenz[1,2,3], Jessica K. Peters[1,2,3], Po-Yu Fang[1,2,3], Dana M. Schneider[1,2,3], Eric J. Anderson[1,2,3], Thanawadee Preeprem[3,4], Jessica C. Bowman[1,2,3], Eric B. O'Neill[1,2,3], Lively Lie[2,3,4], Shreyas S. Athavale[2,3,4], J. Jared Gossett[2,3,4], Catherine Trippe[1,2,3], Jason Murray[1,2,3], Anton S. Petrov[1,2,3,4], Roger M. Wartell[2,3,4], Stephen C. Harvey[1,2,3,4], Nicholas V. Hud[1,2,3] and Loren Dean Williams[1,2,3,*]

[1]School of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta, GA 30332-0400, USA, [2]Parker H. Petit Institute for Bioengineering and Bioscience, Georgia Institute of Technology, Atlanta, GA 30332-0363, USA, [3]Center for Ribosomal Origins and Evolution, Georgia Institute of Technology, Atlanta, GA 30332-0400, USA and [4]School of Biology, Georgia Institute of Technology, Atlanta, GA 30332-0230, USA

Received September 5, 2012; Revised December 7, 2012; Accepted January 3, 2013

ABSTRACT

Ancient components of the ribosome, inferred from a consensus of previous work, were constructed *in silico*, *in vitro* and *in vivo*. The resulting model of the ancestral ribosome presented here incorporates ~20% of the extant 23S rRNA and fragments of five ribosomal proteins. We test hypotheses that ancestral rRNA can: (i) assume canonical 23S rRNA-like secondary structure, (ii) assume canonical tertiary structure and (iii) form native complexes with ribosomal protein fragments. Footprinting experiments support formation of predicted secondary and tertiary structure. Gel shift, spectroscopic and yeast three-hybrid assays show specific interactions between ancestral rRNA and ribosomal protein fragments, independent of other, more recent, components of the ribosome. This robustness suggests that the catalytic core of the ribosome is an ancient construct that has survived billions of years of evolution without major changes in structure. Collectively, the data here support a model in which ancestors of the large and small subunits originated and evolved independently of each other, with autonomous functionalities.

INTRODUCTION

The ribosome is a molecular machine that synthesizes all coded proteins. It is made of a small subunit (SSU) that decodes messenger RNA and a large subunit (LSU) that catalyzes formation of peptide bonds. Some components of the ribosome are highly conserved throughout extant life (1) and are considered to be among the oldest structures in biology (2–7). The peptidyl transferase center (PTC), for example, is thought to predate coded protein (8–12). If so, the PTC emerged from an ancient biological world, possibly an 'RNA World' (13–17), before life adopted all the processes of Crick's 'central dogma' (18). In this scenario, the PTC was an active participant in the origins of current biology and is one of our most direct biochemical links to the distant evolutionary past.

Several groups, including Fox (11), Noller (12), Steinberg (8), Williams (9) and Gutell and Harvey (1), have proposed molecular-level events in early ribosomal evolution or have determined universally conserved ribosomal components (Figure 1). These proposed evolutionary pathways can be used to predict specific sequences and structures of ancestral rRNAs and polypeptides. Here, we use a consensus of proposed evolutionary pathways and conserved ribosomal components (Figure 1) to design and construct a molecular-level model, *in silico*, *in vitro* and *in vivo*, of an ancestral PTC (a-PTC, Figure 2). This model is intended as a starting platform for an iterative hypothesis-testing approach for understanding the origins of translation. Our design process relies substantially on three-dimensional (3D) structures, which are more conserved than sequence over long evolutionary time frames (9,19,20). The a-PTC incorporates fragments of the 23S rRNA, fragments of ribosomal proteins and divalent cations.

The ancestral rRNA fragments inferred here from consensus (Figure 1) are joined together to form a single RNA

*To whom correspondence should be addressed. Tel: +1 404 894 9752; Fax: +1 404 894 7452; Email: loren.williams@chemistry.gatech.edu

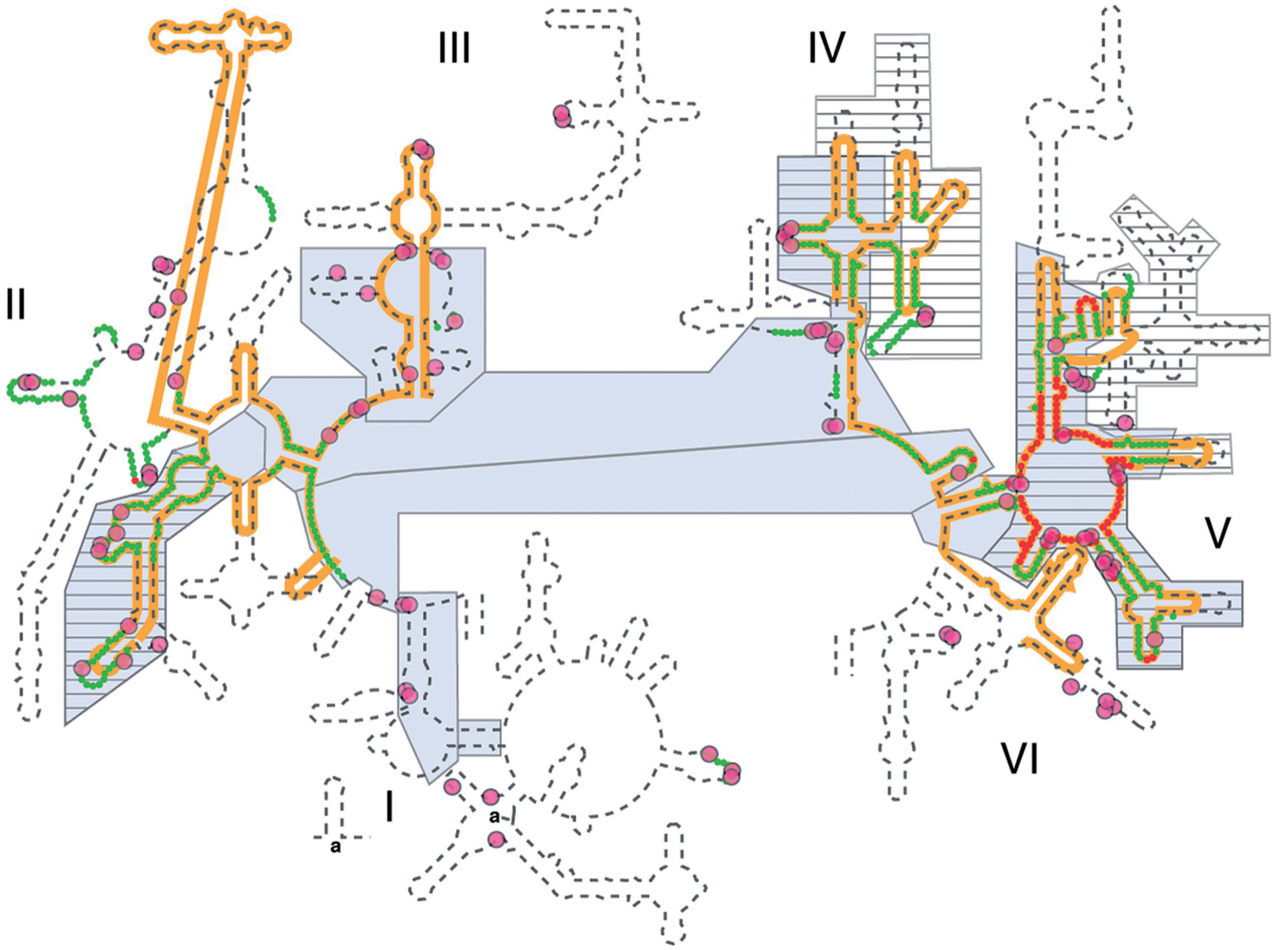

Figure 1. Various models of 23S rRNA evolution. The dashed line illustrates the canonical secondary structure of the *T. thermophilus* 23S rRNA. Secondary structural domains are indicated by roman numerals. The red and green lines show the two inner shells of the ribosomal onion of Hsiao and Williams, marking the rRNA that is in closest proximity, in three dimensions, to the site of peptidyl transfer. The gray boxes are ancient according to the 'A-minor' method of Steinberg. The hashed boxes (with black horizontal lines) are ancient according to the networking analysis of Fox. Multidentate Mg^{2+}-phosphate interactions, also proposed as an indicator of ancient rRNA, are indicated by magenta circles. The orange line shows the universally conserved portions of the 23S rRNA in bacteria, archaea, eukarya, and in mitochondria, as determined by Gutell and Harvey.

polymer (a-rRNA, Figure 2). The a-rRNA contains rRNA that forms and surrounds the PTC, which is composed of fragments from Domains II, IV and V. Using the 3D structure of the *Thermus thermophilus* LSU (21), the 23S rRNA was 'shaved' to a rough sphere of around 30 Å in radius, centered at the site of peptidyl transfer. This shaving process created 13 rRNA fragments. To facilitate reconnection of these fragments to form a polymer, termini were selected preferentially in A-form helical regions of the rRNA 3D structure and were capped with stem loops. The result is a single RNA polymer containing the most ancient 20% of the *T. thermophilus* 23S rRNA. The fragments are stitched together by stem loops in such a way that the 3D structure of the PTC will be maintained (Figure 2). The a-rRNA contains rRNA that is (i) universally conserved in secondary and 3D structure in extant organisms and organelles (1,9), (ii) tightly networked by molecular interactions (11), (iii) densely coordinated by Mg^{2+} ions (22) and (iv) found in the central shells of the ribosomal 'onion'. The ribosomal onion is a conceptual model in which the age of elements of the ribosome is considered to correlate inversely with radial distance from the site of peptidyl transfer (9).

Ribosomal proteins do not engage directly in catalytic processes in the ribosome (23,24). However, the non-globular 'tails' of some ribosomal proteins do penetrate the LSU and interact extensively with the rRNA that forms the catalytic PTC. The tails of ribosomal proteins L2, L3, L4, L15 and L22 penetrate the 30 Å sphere that defines the a-rRNA and are highly conserved in conformation throughout the three domains of life (25–27). We shaved those proteins at the 30 Å boundary used for the 23S rRNA, and included the resulting peptides in the a-PTC. The shaving process did not disrupt protein secondary structure; none of these protein tails are globular. The ancestral peptide components of the a-PTC are called a-rPeptide L2, a-rPeptide L3, a-rPeptide L4, a-rPeptide L15 and a-rPeptide L22 (see Table 1 for peptide sequences). It was previously proposed that these tails are more ancient than any protein with globular structure (9).

Using computation, *in vitro* and yeast three-hybrid experiments, we test the hypotheses that: (i) a-rRNA adopts LSU-like secondary structure, as shown in Figure 2A,

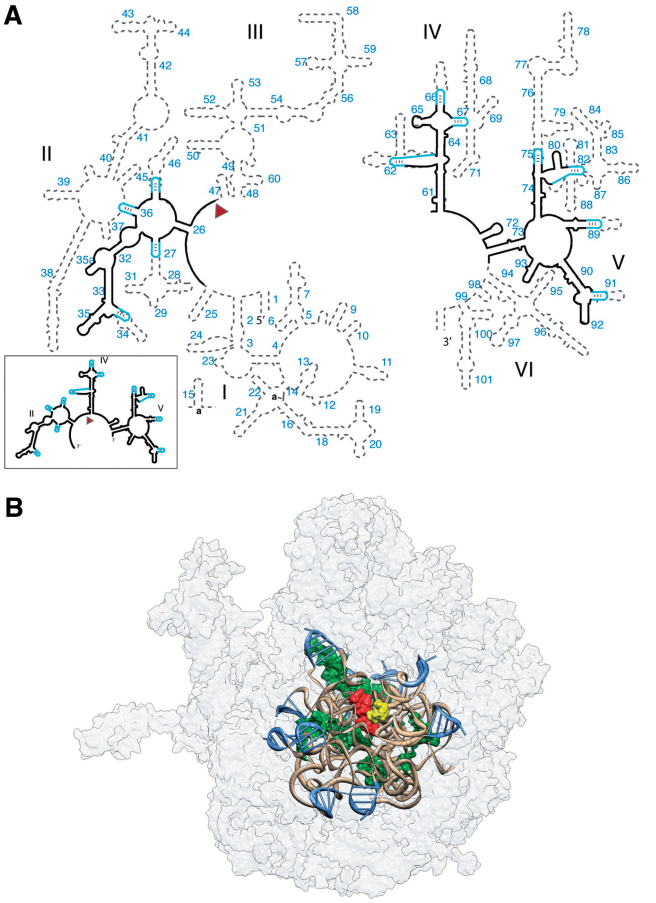

Figure 2. **(A)** Predicted secondary structure of the ancestral 23S rRNA. Ancestral fragments of rRNA, indicated by black lines in the secondary structure, are derived from a consensus of models of rRNA evolution. The ancestral rRNA elements are stitched together by stem loops (blue). The RNA sequences are from the *T. thermophilus* 23S rRNA. Helix numbers are indicated. The predicted secondary structure of the a-rRNA alone is highlighted in the outbox. **(B)** 3D model of the a-PTC. This 3D model contains the a-rRNA plus five a-rPeptides (ancestral fragments of ribosomal proteins L2, L3, L4, L15 and L22). a-rRNA is shown in ribbon (brown), the stem loops are blue and the peptides are in surface representation (green). For reference, A-site (yellow) and P-site (red) substrates are shown in the figure, but are not components of the a-PTC. The modern LSU surface is shown for comparison (light gray, transparent).

(ii) a-rRNA in association with Mg^{2+} adopts LSU-like tertiary interactions, as shown in Figures 2B and 3 and (iii) a-rRNA forms specific LSU-like complexes with a-rPeptides, as shown in Figure 2B. Components of the a-PTC were assembled *in silico*, *in vitro* and *in vivo* and these hypotheses tested by various modeling, footprinting and binding assays. A longer-range goal is determining if a-rRNA in association with a-rPeptides can form a functional a-PTC, as defined by catalytic activity.

MATERIALS AND METHODS

Computer modeling of a-PTC

Model building

The crystal structure of the ribosome from *T. thermophilus* (PDB 2J01) was used as the starting structure for building the *in silico* model of the a-PTC. This a-PTC model (Figure 2B) contains 615 nucleotides of RNA, 505 of which are from the *T. thermophilus* 23S rRNA and 110 of which are from 11 added stem loops of 10 nucleotides each. The a-PTC contains five peptides derived from ribosomal proteins (Table 1).

Table 1. a-rPeptides derived from *T. thermophilus* ribosomal proteins

Parent rProtein	rProtein residue range	Length	a-rPeptide sequence
L2	219–241	23	PHVRGAAMNPVDHPHGGGEGRAP (a-rPeptide L2)
L3	136–153	18	RHPGSIGNRKTPGRVYKG (a-rPeptide L3)
L4	52–92	41	KTRGEVAYSGRKIWPQKHTGRAR HGDIGAPIFVGGGVVFGP (a-rPeptide L4)
L15	19–63	45	VGRGPGSGHGKTATRGHKGQKSRS GGLKDPRRFEGGRSTTLMRLP (a-rPeptide L15)
L22	82–99	18	LKRVLPRARGRADIIKKR (a-rPeptide L22)

The *T. thermophilus* 23S rRNA was shaved to yield the rRNA fragments indicated by the black lines in Figure 2A. The shaving process left 13 rRNA fragments. Most of the fragment termini are within helical regions, allowing facile reconnection with stem loops to create a continuous RNA chain. For 11 of the connections, the stem loop is 5′-gccGUAAggc-3′, which is a GNRA tetraloop (capital letters) on a three base pair stem (lower case letters). The stem loops, indicated by blue lines in Figure 2A, were computationally docked onto the 3D a-rRNA model as described in Supplementary Figure S1. Two of the rRNA fragments were connected directly, without a stem loop.

Stem-loop geometry

We have previously described the consensus conformation of a GNRA tetraloop (28). Here, we searched the crystal structure of the *T. thermophilus* ribosome (PDB entries 2J00 and 2J01) for a 10-nucleotide GUAA tetraloop (5′-gccGUAAggc-3′) containing several additional base pairs on the stem. An rRNA sub-structure from the SSU was selected based on the highest sequence identity to our designated tetraloop sequence and nearly ideal GNRA tetraloop geometry as described (28). The tetraloop sequence from this substructure was mutated from GCAA to GUAA using SYBYL-X 1.2 (Tripos International, St. Louis, MO, USA). The VMD 1.8.7 superposition tool (29) was used to find the best docking position between each stem loop and the termini of the helices remaining after the shaving process (see Supplementary Materials for further information on the docking process).

Minimization

After addition of the stem loops to the a-rRNA *in silico* model, the structure was manually inspected for steric clashes and unfavorable backbone conformations. The solvent accessible surface areas of the stem loops were

calculated using a VMD tool to ensure that all stem loops were located on the surface of the a-PTC. After minor manual manipulation, the a-PTC model was energy minimized in the presence of 51 Mg^{2+} ions observed in the crystal structure to be coordinated by the a-rRNA. The solvated system was neutralized with Na^{+} ions. The nucleotides directly coordinating Mg^{2+} ions were constrained to fixed positions during minimization. The a-rPeptides were free from constraint during minimization. In total, 500 000 steps of conjugate gradient minimization were performed with NAMD (29) using the CHARMM22 force field (30). Structure validation of the energy minimized a-PTC model was performed using the ADIT Validation Server (http://deposit.rcsb.org/validate/), which returned no significant stereochemical issues. An analysis of the structure of the a-PTC 3D model can be found in the Supplementary Material.

Synthesis of a-rRNA

The DNA gene encoding a-rRNA was constructed by recursive PCR (31) (Supplementary Figure S2), and its sequence confirmed. The rRNA fragments were linked together at the DNA level with 11 stem loops. The RNA product of the *in vitro* transcription reaction runs as a tight band on a denaturing gel (Supplementary Figure S3). Gene synthesis and characterization, and RNA synthesis by *in vitro* transcription are described in detail in the Supplementary Materials.

RNase H cleavage reactions

Oligodeoxynucleotide probes

DNA probes for RNase H cleavage reactions were obtained from Integrated DNA Technologies, Inc. or Eurofins MWG Operon. The nomenclature for describing the DNA probes indicates if the RNA target is predicted to be double-stranded (D), single-stranded (S) or a stem-loop (T), and gives the helix number that either contains or succeeds the target site. The DNA probes are: 5′-TTTC GGGTCC-3′ (S32), ACCAGCTATC (S36), CATTCGCA CT (S26), CGTTACTCAT (S61), TGCAGAGTTC (S62), GTTCAATTTC (S72), GGTCTTTTCG (S74), CCTGTT ATCC (S89), ACATCGAGGT (S90), TCTGAACCCA (S93), GATAGAGACC (S73), GCCTTACGGC (T), G CCAGGGCTA (D33), CCCTCGCCGA (D61) and AG CTCCACGG (D74). Nucleotide numbers targeted by each probe are provided in Supplementary Table S1.

RNase H reaction

In vitro-transcribed a-rRNA was dephosphorylated with Antarctic Phosphatase (New England Biolabs). Ten picomoles of the a-rRNA was 5′-end labeled with 0.05 mCi of [γ-^{32}P]ATP using T4 polynucleotide kinase for 30 min at 37°C. The end-labeled RNA was separated from unreacted [γ-^{32}P]ATP with Ambion NucAway spin columns. Divalent cations were removed from the labeled a-rRNA by heating at 90°C for 2 min in the presence of Divalent Cation Chelating Resin (100–200 Mesh sodium form; Hampton Research). To anneal the a-rRNA, 17 pmol of unlabeled a-rRNA was mixed with trace amounts of end-labeled a-rRNA, dissolved in 24 µl of 50 mM Tris–HCl pH 8.0, 100 mM KCl, 33 µM $MgCl_2$, heated at 90°C for 30 s and cooled to room temperature. The a-rRNA solution was diluted to a total volume of 240 µl with water, 10× RNase H reaction buffer and 22 U of RNase H (New England Biolabs) to provide a master mix with 0.071 pmol/µl a-rRNA, 0.092 U/µl RNase H in 1.43× RNase H buffer [50 mM Tris–HCl (pH 8.3), 75 mM KCl, 3 mM $MgCl_2$ and 10 mM dithiothreitol]. This master mix was split into 7 µl aliquots, and 3 µl of a DNA probe solution was added to each tube to give 10 µl reaction volumes. Each reaction contained a final 1:30 RNA to DNA molar ratio. RNase H reactions were carried out at 37°C for 30 min and quenched by quick freezing. Reaction products were run on a 5% denaturing polyacrylamide gel (8 M urea), and imaged on a General Electric Typhoon Trio+ Imager. To calculate the extent of digestion in the presence of each probe, bands representing the digested and undigested a-rRNA were quantified (Supplementary Figure S4).

SHAPE reactions

Selective 2′-hydroxyl acylation analyzed by primer extension (SHAPE) methods were adapted from published protocols (32). *In vitro*-transcribed a-rRNA was prepared in TE buffer (10 mM Tris–HCl, 1 mM EDTA, pH 8.0) at 100 ng/µl a-rRNA. Thirty-two microliter aliquots of the RNA solution were added to 4 µl of 10× folding buffer (500 mM NaHEPES pH 8.0, 2 M NaOAc, varying $MgCl_2$) and incubated at 37°C for 20 min.

NMIA modification of a-rRNA

A 130-mM *N*-methylisatoic anhydride (NMIA; Tokyo Chemical Industry Co., Ltd.) solution in 2 µl dimethyl sulfoxide (DMSO) was added to solutions of 18 µl annealed a-rRNA. Control reactions contained DMSO only. Reactions were carried out for 1 h at 37°C. For NMIA modification under denaturing conditions, the reactions were run at 90°C for 4 min. NMIA-modified RNA was purified from reaction mixtures with the Zymo RNA Clean + Concentrator-25 kit (Zymo Research). a-rRNA was eluted with 25 µl TE buffer. Recovery after purification was >75%.

Reverse transcription of NMIA-modified a-rRNA

Four different 5′-[6-FAM]-labeled DNA oligonucleotides (Eurofins MWG Operon) were used as primers (each helix in parenthesis either contains or succeeds the primer binding site): 5′-TGCCCGTGGCGGATAGAGAC-3′ (helix 73), 5′-ACATCGAGGTGCCAAACCGCC-3′ (helix 89), 5′-GTTCAATTTCACCGGGTCCCTCG-3′ (helix 61) and 5′-CGTTACTCATGCCGGCATTCGC-3′ (helix 26). Modified RNA (20 µl) was added to 8 pmol of each primer in 10 µl of TE buffer. To anneal primers, samples were heated at 95°C for 1 min, held at 65°C for 3 min and then placed on ice. SuperScript III Reverse Transcriptase (Invitrogen) was used in reverse transcription (RT) reactions. RT buffer (19 µl) was added at 30°C to yield 50 mM Tris–HCl pH 8.3, 75 mM KCl, 3 mM $MgCl_2$, 2 mM DTT and 250 µM of each dNTP (final concentrations in 50 µl). RT mixtures were heated at 55°C for 1 min before addition of 1 µl Superscript III Reverse

Transcriptase enzyme mix (200 U). Reactions were incubated at 55°C for 2 h and terminated by heating at 70°C for 15 min. A sequencing control reaction of unmodified *in vitro*-transcribed RNA was dissolved in TE buffer at 31 ng/μl. Aliquots of RNA solution (20 μl) were annealed to the DNA primers as described earlier. RNA was sequenced by RT/chain termination using all four dideoxynucleotidetriphosphates (ddNTPs), at a ratio of 8:1 ddNTP to dNTP, and a control reaction without ddNTPs was also prepared.

Capillary electrophoresis of RT reaction products

One microliter of RT reaction mixture was mixed with 0.3 μl of ROX-labeled DNA sizing ladder (for alignment of disparate traces) and 8.7 μl Hi-Di Formamide (Applied Biosystems) in a 96-well plate. Plates were heated at 95°C for 5 min and the products were resolved by capillary electrophoresis using a 3130 Genetic Analyzer (Applied Biosystems) at 65°C with a custom fluorescence spectral calibration. The capillary array was loaded with Performance Optimized Polymer-4 (Applied Biosystems).

SHAPE data processing

SHAPE data were processed as described earlier (33). Several nucleotides were excluded from SHAPE analysis because the reverse transcriptase gave high background termination in the (−) NMIA reactions. These nucleotides were determined statistically [termination value ≥90% of the (+) NMIA value, where both values exhibit ≥0.25 normalized intensity].

Gel mobility of a-rRNA

The a-rRNA was first treated with divalent chelating beads (Divalent Cation Chelating Resin, 100–200 Mesh sodium form; Hampton Research) to remove Mg^{2+} ions. a-rRNA in 10 mM Tris-Cl, pH 8, was mixed with the chelating beads and heated at 90°C for 2 min. The mixture was then chilled on ice for 20 min. The a-rRNA was separated from the beads and buffer with a 0.2-μm centrifugal filter (Amicon). Aliquots of 180 ng a-rRNA were added to TE solutions of varying [Mg^{2+}], heated at 85°C for 30 s and allowed to cool at room temperature for 25 min before loading on a 5% native-PAGE gel.

a-rRNA/a-rPeptide L4 interaction by continuous variation analysis

a-rPeptide L4 (Table 1) was purchased from Peptides International (Louisville). In a series of solutions, the total concentration ([a-rRNA] + [a-rPeptide L4]) was held constant at 60 μM, while the mole fractions of the two components were varied from 0.0 to 1.0 (34,35). The mixtures were prepared in 10 mM Tris-Cl buffer, pH 8, heated at 85°C for 30 s and annealed by cooling to 30°C at a linear rate of 1.5°C/min. The annealed samples were transferred to a microplate. The fluorescence signal of the tryptophan contained in a-rPeptide L4 was measured at 334 nm at 25°C using a Biotek Synergy H4 Multi-Mode Plate Reader.

Gel shift assays with a-rRNA and a-rPeptides

Cloning of MBP-a-rPeptide fusion proteins

DNA genes encoding a-rPeptides L2, L3, L15 and L22 (Table 1) were synthesized by recursive PCR as described earlier (31). a-rPeptide genes were ligated into the pMAL-c5x vector (New England Biolabs) downstream of the *malE* gene, which encodes for maltose-binding protein (MBP). Each resultant MBP fusion protein bears an a-rPeptide (Table 1) on the C-terminus and is referred to here as an MBP-a-rPeptide. Information on gene sequences, cloning, expression and purification of MBP-a-rPeptide fusions is provided in the Supplementary Materials.

a-rRNA/MBP-a-rPeptide interactions

A solution of 1 μM a-rRNA was prepared in 20 mM Tris-Cl, pH 8. The a-rRNA solutions were heated at 85°C for 30 s and cooled linearly at a rate of 1.5°C/min to 30°C. At concentrations of 1–10 μM, each MBP-a-rPeptide fusion was incubated with 1 μM a-rRNA at 4°C for variable durations (several hours to overnight). Interactions between the MBP-a-rPeptides and a-rRNA were analyzed on a 5% native PAGE gel with 3% glycerol. Gels were visualized using a two-color fluorescence dye protocol (36).

Yeast three-hybrid assays

Hybrid cloning

The genes of *T. thermophilus* ribosomal proteins L2 (NCBI accession number NC_005835: c1265055-1264225), L3 (NC_005835:c1266643-1266023), L4 (U36480), L15 (NC_005835:c1258456-1258004) and L22 (NC_005835: c1263932-1263591) were amplified from genomic DNA and individually cloned into the pACTII vector (37). Amplification primer sequences are given in Supplementary Table S3. In the pACTII vector, ribosomal protein genes were fused to the C-terminal end of the GAL4 transcriptional activation domain (GAD). The a-rRNA gene was amplified from the transcription vector described in the Supplementary Materials and cloned into the T-cassette vector (38). The a-rRNA gene was ligated to the 5′-end of the gene encoding for MS2 RNA. Positive control RNA hybrid p50-MS2 cloned into the T-cassette vector and protein hybrid GAD-p53 cloned into pACTII (38,39) were a generous gift from Dr James Maher.

In vivo *binding assays*

Yeast three-hybrid assays were performed in the YBZ-1 yeast strain as described earlier (37,40,41). Double transformants with the protein and RNA-hybrid constructs were selected in medium lacking adenine and leucine (CM-AL). Interaction of RNA with protein results in activation of the GAL4 promoter and expression of *LacZ*. The strength of an RNA-protein interaction is indicated quantitatively by β-galactosidase (β-gal) activity (42). The positive control used hybrids p50-MS2 and GAD-p53, which are known to interact strongly in the yeast three-hybrid system (39). Negative controls include each protein hybrid with the MS2 RNA only (T-cassette vector without insert). Growth observed in negative

controls was considered to represent background signal. All experiments were performed in a minimum of six replicates.

RESULTS

a-PTC *in silico*

The a-PTC was modeled in three dimensions (Figure 2B). To form the *in silico* a-PTC model, the X-ray structure of the LSU of *T. thermophilus* (PDB 2J01) was shaved to a sphere of around 30 Å, and 12 of the 13 resulting rRNA fragments were stitched together with stem loops, each of sequence 5′-gccGUAAggc-3′. Each pair of adjacent rRNA fragments was connected with a stem loop as described in the Supplementary Materials. An exception is the rRNA fragments adjacent to helices 47 and 61, where the rRNA fragments were joined directly, without a stem loop (Figure 2A). The completed a-rRNA is composed of 505 nucleotides derived from the 23S rRNA plus 110 nucleotides from 11 stem loops. The a-PTC *in silico* contains a-rRNA (615 nucleotides), 51 Mg^{2+} ions and five a-rPeptides (Table 1).

In the 3D model, the components that are common between the a-PTC and the *T. thermophilus* LSU crystal structure have conserved conformation and interactions. The stem loops are constrained to canonical GNRA tetraloop conformation (28) and are located on the surface of the a-PTC. The stem loops do not engage in unfavorable steric contacts with other parts of the a-PTC. The model is stereochemically reasonable.

The predicted secondary structure of the a-PTC shown in Figure 2A is based on the canonical secondary structure of the 23S rRNA. There are some small differences between the canonical secondary structure and the actual secondary structure found in the crystal structure. These difference are responsible for some otherwise unexpected experimental observations, as discussed below.

RNase H characterization of a-rRNA secondary structure

RNase H and 15 different DNA oligonucleotide probes were employed to characterize the secondary structure of a-rRNA *in vitro* (Figure 3A). RNase H cleavage is broadly used to characterize RNA secondary structure, allowing discrimination between single-stranded and double-stranded RNA (43,44). RNase H recognizes and cleaves RNA/DNA heteroduplexes; double-stranded RNA inhibits cleavage by blocking heteroduplex formation. The RNase H cleavage pattern of the a-rRNA shows good overall agreement with predictions of the *in silico* model, suggesting formation of the secondary structure shown in Figures 2A and 3. Four DNA probes were designed to target regions of a-rRNA that are double-stranded in the predicted secondary structure, and so should not yield RNase H cleavage. Eleven DNA probes were designed to target regions of a-rRNA that are single-stranded in the predicted secondary structure, which should yield RNase H cleavage. Ten of 11 predicted single-stranded RNA target regions are cleaved partially or fully by RNase H. Three of the four predicted

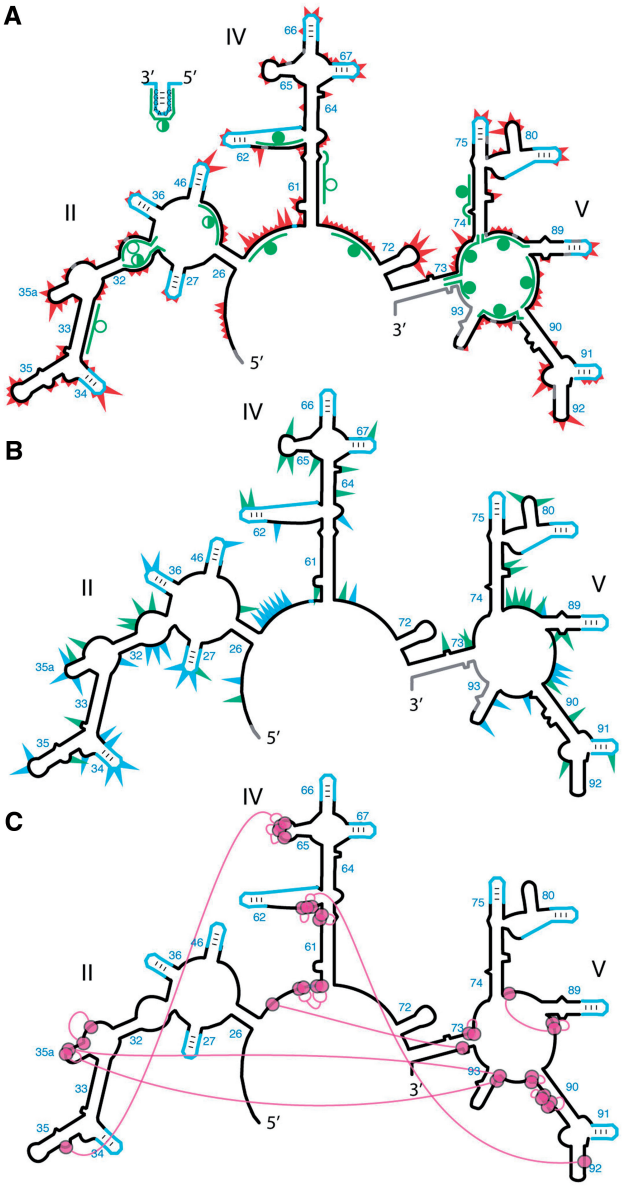

Figure 3. Probing the secondary and tertiary structure of a-rRNA. **(A)** SHAPE and RNase H mapping. Red triangles mark SHAPE reactivities in 250 mM Na^{+}, mapped onto the predicted secondary structure of a-rRNA. Larger triangles indicate greater SHAPE reactivity. RNase H DNA probes are indicated by green lines. Circles indicate extent of RNA digestion by RNase H: filled circles (more than 75%), half-filled circles (between 25 and 75%) and empty circles (<25%). **(B)** Effects of 10 mM Mg^{2+} on SHAPE reactivity suggest formation of tertiary structure. Green triangles show the greatest increases in SHAPE reactivity upon addition of Mg^{2+}. Blue triangles show the greatest decreases in reactivity. **(C)** Multidentate Mg^{2+}-phosphate interactions observed in the *T. thermophilus* LSU (PDB 2J01) are mapped onto the predicted secondary structure of a-rRNA. Magenta circles indicate first-shell Mg^{2+}-OP (magnesium-phosphate oxygen) interactions. Magenta lines indicate PO-Mg^{2+}-OP linkages. Gray shading in panels A and B indicates rRNA where SHAPE data were not accessible. SHAPE reactions were performed in 50 mM NaHEPES, pH 8.0, 200 mM NaOAc, 0 or 10 mM $MgCl_2$.

double-stranded target regions are protected from cleavage, or showed only partial cleavage.

RNase H cleavage in the presence of two of the DNA probes initially appeared to suggest inconsistencies

between the predicted and observed a-rRNA secondary structure, but further analysis indicated otherwise. Specifically, probe S36 targets RNA that is single-stranded in the predicted secondary structure, and therefore was anticipated to yield cleavage by RNase H. However, inspection of the *T. thermophilus* 50S crystal structure (21) reveals that the target region of probe S36 is engaged in secondary interactions not reflected in the canonical 23S secondary structure on which the predicted a-rRNA secondary structure is based. S36 does not yield cleavage of a-rRNA consistent with the secondary interactions observed in the LSU crystal structure.

Oligonucleotide D74 targets Helix 74, which is double-stranded in the predicted secondary structure, and thus D74 was not expected to yield cleavage. Contrary to this prediction, D74 yields significant cleavage (~88%) compared with the other double-stranded target regions. Comparing RNA/RNA versus DNA/RNA duplex free energies (45,46), we find that the stability of Helix 74 (−5.80 kcal/mol) is less than that of the DNA/RNA heteroduplex formed by D74 with the target RNA (−12.5 kcal/mol). This differential stability may allow the probe to invade the RNA/RNA duplex, accounting for the cleavage.

SHAPE characterization of a-rRNA secondary structure

SHAPE utilizes an electrophile, in this case NMIA, that reacts with 2′-hydroxyl groups of RNA (47). SHAPE reactivity is modulated by RNA flexibility. Base-paired nucleotides exhibit low reactivity as their flexibility is constrained. Single-stranded nucleotides are flexible and reactive to the electrophile. SHAPE has been shown to be accurate for mapping secondary structure of many RNAs, including *Escherichia coli* rRNA (48).

The SHAPE data show excellent correspondence with the predicted secondary structure of a-rRNA (Figure 3A). These data were obtained by probing the a-rRNA in the presence of monovalent cations (250 mM Na^+) and the absence of Mg^{2+}. These conditions are known to support the formation of RNA secondary structure, but not tertiary structure (49–51). SHAPE data were obtained for 459 of the 505 nucleotides (91%) of the a-rRNA derived from the 23S (stem-loop nucleotides are analyzed separately below). The values and statistics provided in the remainder of this section refer only to these nucleotides derived from the 23S rRNA. Data were not obtainable for RNA near the termini, or for a few nucleotides with high background (i.e. high reverse transcriptase termination in samples not treated with NMIA). These regions are shown in gray on the a-rRNA secondary structure in Figure 3A.

Nucleotides were binned into four groups according to their normalized SHAPE reactivity. U2574 (NCBI *E. coli* numbering) is the most reactive nucleotide, with an absolute reactivity of 3.24 (arbitrary units), and is considered 100% reactive. Bin assignments are indicated by the size of red triangles in Figure 3A. The largest triangles denote the 13 nucleotides with normalized SHAPE reactivities ranging from 38 to 100%. Triangles of intermediate size indicate the 24 nucleotides with reactivities from 25 to 37%. The smallest triangles indicate the 109 nucleotides with reactivities from 8 to 24%. Nucleotides with reactivities of $\leq$7% are considered to be unreactive. Full numerical SHAPE values are available in the Supplementary Material.

Of the 146 nucleotides that are reactive to the SHAPE reagent, 133 (91%) fall in regions expected to be single-stranded, in that they do not form base pairs in the predicted secondary structure. These nucleotides are in loops, bulges or other single-stranded regions. The resolution of SHAPE is sufficiently high to allow identification of small bulges such as those in helices 35, 61, 64, 73, 74 and 89 (Figure 3A). Thirteen reactive nucleotides are base paired in the predicted secondary structure. These nucleotides represent <6% of the 232 nucleotides that are expected to be double-stranded, the remaining 94% of which exhibit low SHAPE reactivity as predicted. The 133 nucleotides that SHAPE identifies as flexible compose 59% of the 227 single-stranded nucleotides in the predicted secondary structure.

gccGUAAggc stem-loops fold in all contexts

SHAPE data were obtained for 106 (96%) of the 110 nucleotides within the 11 stem loops. The stem regions are unreactive. Only 5 of the observable 62 nucleotides in the stem regions (8%) exhibit reactivity, which is low for each. The loop regions yield a consistent pattern of reactivity that is independent of the surrounding sequence. In 10 of the 11 stem loops, the UAA nucleotides are more reactive than the preceding G. This pattern of reactivity is consistent with known patterns of flexibility of GNRA tetraloops (28,52). Nucleotides NRA of a GNRA tetraloop are more polymorphic and flexible than the G. The N nucleotide is the most polymorphic of all.

The SHAPE results suggest that the stem-loop replacing helix 91 (Figure 3A) might not form the predicted secondary structure. The stem loop at this position exhibits a distinctive pattern of reactivity. Two nucleotides in the stem appear to be anomalously flexible. This SHAPE data will inform design changes in future iterations of a-rRNA.

SHAPE characterization of a-rRNA tertiary interactions

RNAs require Mg^{2+} ions to form compact structures (49–51). The core of the assembled LSU is particularly rich in Mg^{2+} ions (9,22,53). In our 3D model of the a-PTC, the conformation of the a-rRNA is stabilized in part by Mg^{2+} ions (Figures 2B and 3C) and also by tertiary interactions between RNA elements that are remote in the primary sequence. Here, we ask if a-rRNA *in vitro*, upon the addition of Mg^{2+}, forms a compact structure consistent with the *in silico* a-PTC model (Figure 2B).

A network of phosphate-Mg^{2+}-phosphate interactions is anticipated in the a-PTC, based on the network in the LSU crystal structure (PDB 2J01) (22). This phosphate-Mg^{2+}-phosphate network is indicated in Figure 3C, where magenta circles represent nucleotides with phosphate oxygens that make direct contacts with Mg^{2+} ions (<2.4 Å cutoff distance). Long-range phosphate-Mg^{2+}-

phosphate linkages are indicated by magenta lines. The proposed network includes all Mg^{2+} ions that interact with phosphates of two or more of the nucleotides. Formation of the phosphate-Mg^{2+}-phosphate network is coupled with formation of base-base tertiary interactions. Therefore, Mg^{2+} is expected to influence the rRNA flexibility and SHAPE reactivity of nucleotides that contact Mg^{2+} or are involved in long-range tertiary interactions. This pattern of Mg^{2+}-dependent SHAPE reactivity has previously been observed for tRNA, RNase P, the P4-P6 domain of the *Tetrahymena* Group I intron and Domain III of the 23S rRNA (33,47,54,55).

The influence of Mg^{2+} on a-rRNA SHAPE reactivity (Figure 3B) suggests that Mg^{2+} does indeed induce global folding. Some Mg^{2+}-induced changes in SHAPE reactivity of a-rRNA are clustered around hypothesized regions of direct RNA-Mg^{2+} contact similar to those in the folded LSU. Others are observed in regions of hypothesized long-range tertiary interactions. Absolute SHAPE data are graphed in the Supplementary Figure S5. The numerical data are contained in a Supplementary Excel file (Supplementary Dataset 1).

Changes in SHAPE reactivity upon addition of Mg^{2+} (10 mM) were determined. The greatest changes in SHAPE reactivity (>40% change) are mapped onto the predicted secondary structure of a-rRNA (increases: green; decreases: blue, Figure 3B). Nucleotides with large changes in reactivity are dispersed throughout the sequence (Supplementary Figure S5) and the secondary structure. Nucleotides for which the absolute SHAPE reactivity is small ($\leq$0.3) were omitted from the analysis to avoid attributing artificial significance to small changes in absolute value.

Of the 65 rRNA nucleotides (excluding stem-loops) with altered SHAPE reactivities upon addition of Mg^{2+} (Figure 3B), 25 (38%) are within three residues of a nucleotide that contacts a Mg^{2+} ion in the predicted interaction network in Figure 3C. Most of the other Mg^{2+} effects are at or near bulges and loops, assumed to be involved in tertiary interactions. The data are consistent with induction by Mg^{2+} of local and long distance interactions that alter nucleotide flexibility. The results suggest a transition of the a-rRNA from an extended secondary structure to a collapsed structure with tertiary interactions.

The SHAPE reactivities argue against significant alteration of the secondary structure upon addition of Mg^{2+}. Only a small number of helical nucleotides show Mg^{2+}-dependent changes. Of the 65 rRNA nucleotides that show significant Mg^{2+} effects, only a few (9 nucleotides) are found in regions that are paired in the predicted secondary structure, and several of these are adjacent to predicted bulges or other single-stranded regions. Changes in chemical reactivity upon folding in the presence of rProteins or a-rPeptides will be investigated in future work.

Magnesium alters the gel mobility of a-rRNA

The SHAPE data support the hypothesis that the conformation of the a-rRNA is altered by Mg^{2+}. This hypothesis is also tested by gel mobility. In the absence of Mg^{2+}, a-rRNA appears to have one predominant conformation and an ensemble of minor conformations (Figure 4), inferred from the smeared intensity between the lower band and fainter upper band on the native gel. The smear does not arise from heterogeneity in length, as a-rRNA runs as a tight single band on denaturing gels (Supplementary Figure S3). During the course of Mg^{2+} titration, the a-rRNA is seen in a native gel to transition to a single, slower-running state (Figure 4). This transition appears complete at 250 μM Mg^{2+}.

a-rRNA forms a specific 1:1 complex with a-rPeptide L4

The fluorescence of the tryptophan within a-rPeptide L4 (Table 1) was used to monitor the interaction of a-rRNA and a-rPeptide L4 in a Continuous Variation experiment (34,35). The two species were mixed in a series of solutions of varying mole fractions. The total concentration of the two species was held fixed throughout. Continuous variation of a-rRNA and a-rPeptide L4 shows a distinct inflection at 1:1 stoichiometry (Figure 5), where the molar

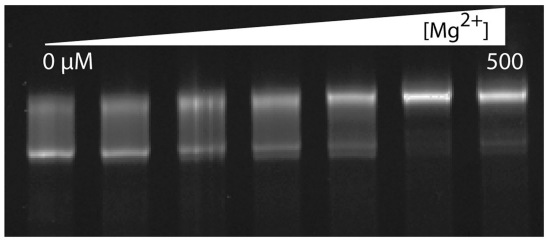

Figure 4. The effect of Mg^{2+} on gel mobility of the a-rRNA suggests Mg^{2+} induction of folding. Shown here is a-rRNA annealed in 10 mM Tris, pH 8.0, and varying [Mg^{2+}], resolved on a 5% native acrylamide gel. Lane 1, [Mg^{2+}] = 0 μM; Lane 2, 12.5; Lane 3, 25; Lane 4, 50; Lane 5, 100; Lane 6, 250; Lane 7, 500.

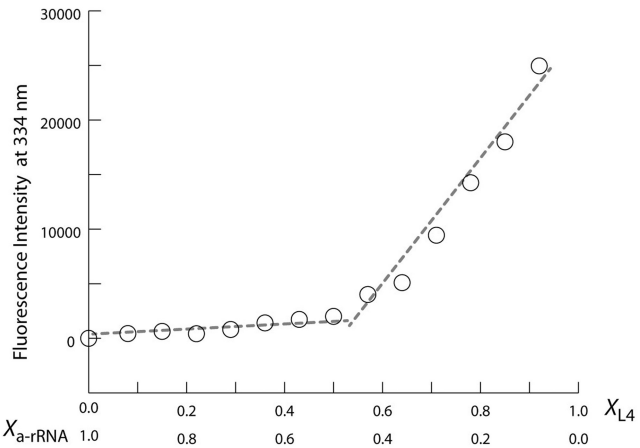

Figure 5. a-rRNA and a-rPeptide L4 are assembly competent and form a complex with 1:1 stoichiometry. Fluorescence signal at 334 nm is from the tryptophan residue of the a-rPeptide L4. In this plot of a continuous variation experiment of a-rPeptide L4 with a-rRNA, X_{L4} and $X_{a\text{-}rRNA}$ values on horizontal axis denote mole fraction of the peptide and RNA in each sample, where the total concentration ([a-rRNA] + [a-rPeptide L4]) was held constant at 60 μM. The discontinuity at equivalent mole fractions of peptide and RNA indicates a complex with 1:1 stoichiometry. These binding assays were performed in 10 mM Tris, pH 8.0, at 25°C.

fractions of a-rPeptide L4 and a-rRNA are both 0.5. Thus, continuous variation of a-rRNA and a-rPeptide L4 suggests a 1:1 stoichiometry of binding. In a control experiment, an inflection was not observed for the continuous variation of the P4-P6 domain of the *Tetrahymena* Group I intron with a-rPeptide L4 (Supplementary Figure S6).

a-rRNA forms complexes with MBP-a-rPeptides L3, L15 and L22

The interactions between a-rRNA and a-rPeptides L2, L3, L15 or L22 are not detectable by fluorescence due to the absence of fluorescent amino acids in these peptides. Instead, the interactions of a-rPeptides L2, L3, L15 or L22 with a-rRNA were evaluated by electrophoretic mobility shift assay (EMSA). The peptides alone are too small to cause a visible change in a-rRNA gel mobility. Therefore, the a-rPeptides were fused to the C-terminus of MBP using molecular cloning techniques. In a two-color EMSA (36), uncomplexed RNA is green, and free protein (MBP-a-rPeptide fusions) is red. If an RNA and a protein form a complex, they co-localize on the gel, producing a yellow band (Figure 6).

The results indicate that MBP fusions with a-rPeptides L3, L15 or L22 form complexes with a-rRNA in the range of 1–10 μM of the MBP-a-rPeptide fusion (Figure 6). Each of these three MBP-a-rPeptide fusions shifts the a-rRNA band on the gel. MBP-a-rPeptide L3 required overnight incubation to form a complex with a-rRNA, but MBP-a-rPeptides L15 and L22 required only a few hours incubation. MBP-a-rPeptide L2 did not form a complex with a-rRNA under any of the conditions tested, even at concentrations an order of magnitude greater than those sufficient for binding of MBP-a-rPeptide L3, L15 or L22. Here, we assayed a-rRNA interactions with a-rPeptides in the absence of Mg^{2+}. The a-rRNA exhibits two dominant conformations on EMSA gels (Figure 6) indicated by two bands. The binding of any a-rPeptide, like binding of Mg^{2+}, induces formation of a single state, indicated by a single band on the gel. A control experiment shows MBP alone does not bind to a-rRNA (Figure 6A). At high protein concentration (>50 μM), the

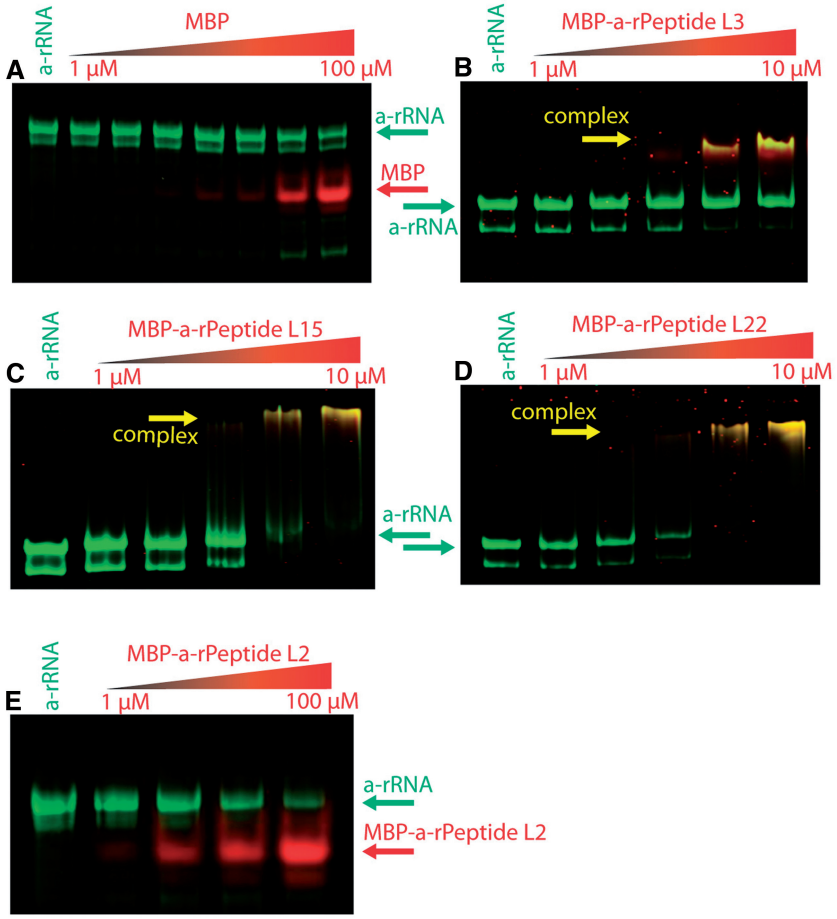

Figure 6. Gel shift analyses of interactions between a-rRNA and MBP-a-rPeptide fusions. RNA, protein and RNA–protein complexes were visualized on 5% native-PAGE gels by two-color EMSA. All binding reactions were performed with 1 μM a-rRNA, in 20 mM Tris buffer, pH 8.0. In each gel, the far left lanes contained a-rRNA only (no protein). **(A)** In a control assay, a-rRNA does not interact with MBP alone: a-rRNA incubated with MBP (left to right: 0, 1, 2, 4, 8, 10, 50 and 100 μM). **(B)** a-rRNA incubated with MBP-a-rPeptide L3 (left to right: 0, 1, 2, 4, 8 and 10 μM). **(C)** a-rRNA incubated with MBP-a-rPeptide L15 (left to right: 0, 1, 2, 4, 8 and 10 μM). **(D)** a-rRNA incubated with MBP-a-rPeptide L22 (left to right: 0, 1, 2, 4, 8 and 10 μM). **(E)** a-rRNA incubated with MBP-a-rPeptide L2 (left to right: 0, 1, 10, 50 and 100 μM).

a-rRNA is seen to degrade, as indicated by faint low-molecular weight RNA bands on the gels.

Using the method of Williamson (56), we have estimated the dissociation constants (K_d) of the complexes formed by MBP-a-rPeptides L3, L15 or L22 with a-rRNA. Data were obtained by integrating the band intensities from the EMSA gels (Figure 6). MBP-a-rPeptide L3 binds weakly, and the K_d is estimated to be >10 μM. The K_d is 4.6 μM for MBP-a-rPeptide L15 and 7.4 μM for MBP-a-rPeptide L22. Kimura previously measured a K_d of 1 μM for the interaction of intact ribosomal protein L2 with appropriate fragments of the 23S rRNA (57). Hachimori measured a K_d of 10 nM for the interaction of intact ribosomal protein L3 with the Sarcin/Ricin Domain of 23S rRNA (58). Nierhaus and others used filter binding and sedimentation to determine stoichiometries of binding of ribosomal proteins to rRNA (59). Ribosomal proteins L2, L3 and L4 are classified as normal binders to the 23S rRNA in the presence of 4 mM Mg^{2+}. L15 and L22 are classified as weak binders.

a-rRNA interacts with intact rProteins L3, L4, L15 and L22 in vivo

To determine if our *in vitro* results could be replicated *in vivo* we used the yeast three-hybrid system to investigate interactions of a-rRNA with a-rPeptides and intact rProteins. The interactions of full length rProteins L2, L3, L4, L15 or L22 with a-rRNA were characterized (schematic provided in Figure 7A). Each rProtein was fused to the C terminus of the GAD, which interacts with the *LacZ* reporter gene. These hybrids are referred to as GAD-L2, GAD-L3, GAD-L4, GAD-L15 and GAD-L22. The RNA–rProtein interactions were assayed by quantification of increased β-gal activity due to enhanced expression of the *LacZ* reporter gene (Figure 7B). By this assay, a-rRNA exhibited the strongest *in vivo* interaction with L4, with a β-gal average activity of 1150 Miller Units (MU). The second strongest *in vivo* interaction with a-rRNA was observed with L22 at 750 MU, and moderate *in vivo* interactions were observed with L15 (420 MU) and L3 (400 MU). Consistent with the *in vitro* assays, a-rRNA did not exhibit any measurable *in vivo* interaction with L2. With a similar experimental approach, interactions were not observed between the a-rPeptides and a-rRNA. Our current hypothesis is that the linker between the a-rPeptide and GAD may need to be lengthened to allow and observe interactions. Further work in this area is in progress.

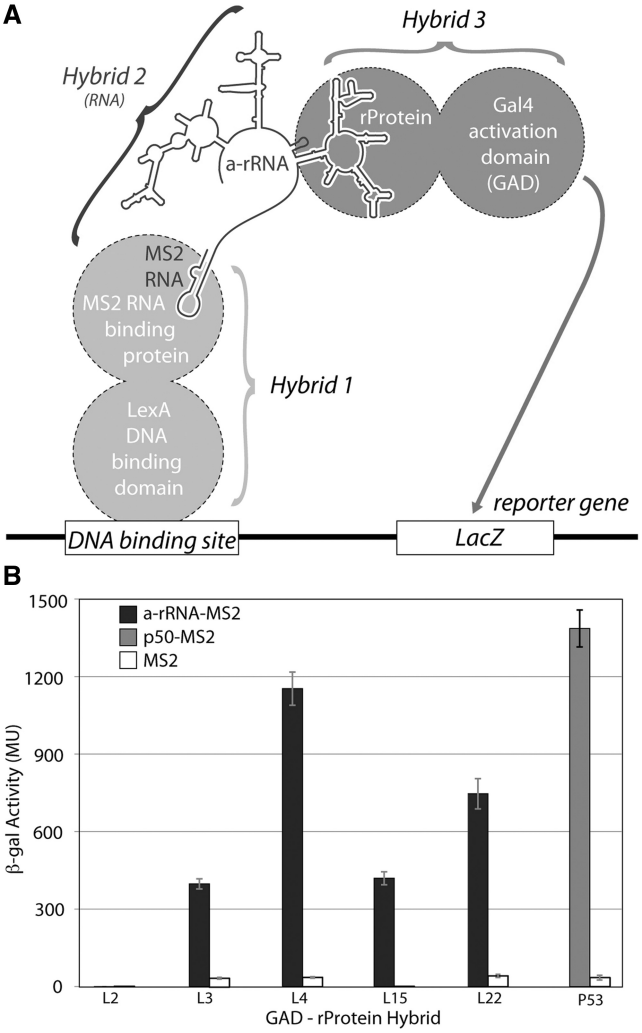

Figure 7. (A) Schematic of the yeast three-hybrid assay used to characterize interactions between a-rRNA and ribosomal proteins. In yeast strain YBZ-1, the *LacZ* reporter gene is controlled by the LexA operator. Hybrid 1, a LexA/MS2 binding protein fusion, binds to the DNA binding site. The MS2 coat protein domain binds tightly to the MS2 RNA, which is fused to the RNA sequence of interest (e.g. a-rRNA). The rProtein of interest is fused to the yeast GAL4 transcriptional activation domain (GAD). *In vivo* RNA-protein binding results in expression of the *LacZ* reporter gene, which is quantified by β-gal activity. **(B)** *In vivo* interactions between a-rRNA and individual rProteins. Interaction is quantified by β-gal activity, reported in MU. Interactions were assayed between a-rRNA-MS2 and GAD-L2, GAD-L3, GAD-L4, GAD-L15 and GAD-L22 (black bars). Positive control was RNA aptamer p50-MS2 and GAD-p53 (gray bar). Negative controls consisted of MS2 RNA and the indicated protein hybrid (white bars).

DISCUSSION

Translation and the ribosome are some of the most illuminating biochemical links between current and ancient biological systems (60). Translation provides a powerful experimental and theoretical system for exploring life's oldest biological processes and molecules (2–7). Some parts of the ribosome are thought to predate our current biology of DNA, RNA and coded protein. Following previous work (10,61–64), we have undertaken an effort to exploit the ribosome as a tool for learning about ancient biology. Ancient components of the LSU, as inferred from consensus among models of early ribosomal evolution, were constructed and characterized. The conclusion here is that the core of the LSU has retained an ancient ability to fold and assemble over vast evolutionary time frames of billions of years. Relatively small ancestral components of the LSU are folding competent and assembly competent.

The experimental system

Advances in technology allow us to address fundamental questions of paleobiochemistry, a discipline first conceived by Zuckerkandl and Pauling (65). Here, we have constructed a model, *in silico*, *in vitro* and *in vivo*, of an ancestral PTC. The a-PTC is designed from a consensus among previous proposals of ribosomal evolution by the laboratories of Fox (11), Steinberg (8), Noller (12), Williams (9) and others. The a-PTC is an experimental model of one of life's most ancient assemblies, composed of fragments of rRNA and protein from the oldest part of the extant ribosome (Figure 1). These rRNA fragments are joined together here by stem loops to form a single RNA polymer called the a-rRNA (Figure 2A). The protein fragments contained within the a-PTC (Table 1) are a-rPeptides (ancestral peptides), derived from portions of rProteins that penetrate deep into the ribosome.

Structure and assembly of the a-PTC was probed by a variety of computational, biophysical and *in vivo* methods. The results suggest that the a-PTC *in vitro* and *in vivo* folds and assembles in general accordance with the predictions of the *in silico* model. This conclusion is based on (i) RNase H cleavage and SHAPE reactivity, which support formation of the predicted secondary structure of a-rRNA, (ii) the Mg^{2+} dependence of the SHAPE profile and gel mobility, which indicate LSU-like Mg^{2+} interactions within a-rRNA, and Mg^{2+}-induced formation of tertiary interactions and (iii) specific binding of a-rPeptides L3, L4, L15 and L22 (*in vitro*) and rProteins L3, L4, L15 and L22 (*in vivo*) with a-rRNA. L2 appears to be an exception in that neither *in vitro* binding of a-rPeptide L2 nor *in vivo* binding of rProtein L2 to the a-rRNA is observed. Variants of a-rPeptide L2 will be constructed and assayed for binding in subsequent work.

Yonath and coworkers previously described a model for the proto-ribosome (10,63) that is significantly smaller than model our a-rRNA. In fact, Yonath's proto-ribosome is fully contained within our a-rRNA. In contrast to the a-PTC, there are no peptide components in the proto-ribosome. Yonath's proto-ribosome appears to represent a model of a smaller, even more primitive ancestor of the a-PTC.

The cooption model of early ribosomal evolution

The fundamental hypothesis that underlies our approach is called the '*cooption model*' (7,12). In the cooption model: (i) ancestors of the SSU and LSU originated and evolved independently of each other, with autonomous functionalities, (ii) an ancestor of the LSU, independent of the SSU, contained the PTC, which catalyzed production of short, non-coded peptide or combined peptide/ester (66) oligomers, (iii) some of the non-coded oligomer products of the LSU ancestor associated with the LSU, conferring advantage, (iv) an ancestor of the SSU had a function that is more tentative but as proposed by Noller may have involved RNA binding or polymerization, (v) ancestral LSUs and SSUs joined, in a cooption process, enabling coded synthesis and (vi) the peptide/ester oligomers associated with the ancestral LSU eventually became the tails of ribosomal proteins that penetrate deep within the extant LSU.

In the cooption model, and other models of ribosomal evolution, changes over evolution are restricted to those that maintain the structure and functionality of the PTC and the decoding center. The catalytic core of the LSU and the decoding center of the SSU are assumed to predate the cooperative relationship between the LSU and SSU. This cooption model predicts that an ancestor of the 23S rRNA, lacking more recent components required for association with the SSU, will retain folding, recognition and assembly capabilities. Here, we test that aspect of the cooption model by determining if the a-rRNA, a model of an ancestor of the 23S rRNA, can fold and assemble with proposed ancestral fragments of ribosomal proteins to form the a-PTC.

The results here support elements of the cooption model of ribosomal evolution. The proposed ancestral elements of the LSU are highly robust in folding and assembly. We have shaved around 2400 nucleotides from the 23S rRNA, and the vast majority of amino acids from the protein components, excising globular protein domains *in toto*. The remaining a-rRNA and a-rPeptides, except that derived from ribosomal protein L2, retain the ability to fold and specifically assemble. It is striking that the a-rPeptides, lacking α-helices, β-sheets and globular domains, retain the ability to interact specifically with the a-rRNA. This robustness is consistent with the premise that the ribosome is an ancient assembly that evolved in a primitive biological environment and has survived billions of years of evolution, growing in size and complexity, without major changes in core structure or function. Although the modern LSU has the appearance of a massive monolithic assembly (23), our results here, and elsewhere (33), indicate that it can be understood as much smaller elements of RNA and peptide that retain ancient abilities to fold and independently assemble.

Future work

The article described here is part of an iterative effort to recapitulate ancestral interactions, assemblies and catalytic activities of the PTC. This effort involves both systematic step-wise changes and the use of libraries for selection. Our long-term goal is to develop a biochemically functional ancestral ribosome.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Tables 1–3, Supplementary Figures 1–6, Supplementary Methods, Supplementary Dataset 1 and Supplementary References [67–70].

ACKNOWLEDGEMENTS

The authors thank Drs James Maher and George Fox for helpful discussions.

FUNDING

NASA Astrobiology Institute [NNA09DA78A]. Funding for open access charge: NASA and NAI.

Conflict of interest statement. None declared.

REFERENCES

1. Mears,J.A., Cannone,J.J., Stagg,S.M., Gutell,R.R., Agrawal,R.K. and Harvey,S.C. (2002) Modeling a minimal ribosome based on comparative sequence analysis. *J. Mol. Biol.*, **321**, 215–234.
2. Woese,C.R. (2000) Interpreting the universal phylogenetic tree. *Proc. Natl. Acad. Sci. USA*, **97**, 8392–8396.
3. Woese,C.R. (2001) Translation: in retrospect and prospect. *RNA*, **7**, 1055–1067.
4. Fournier,G.P., Neumann,J.E. and Gogarten,J.P. (2010) Inferring the ancient history of the translation machinery and genetic code via recapitulation of ribosomal subunit assembly orders. *PLoS One*, **5**, e9437.
5. Williams,D., Fournier,G.P., Lapierre,P., Swithers,K.S., Green,A.G., Andam,C.P. and Gogarten,J.P. (2011) A rooted net of life. *Biol. Direct*, **6**, 45.
6. Wolf,Y.I. and Koonin,E.V. (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.
7. Fox,G.E. (2010) Origin and evolution of the ribosome. *Cold Spring Harb. Perspect. Biol.*, **2**, a003483.
8. Bokov,K. and Steinberg,S.V. (2009) A hierarchical model for evolution of 23S ribosomal RNA. *Nature*, **457**, 977–980.
9. Hsiao,C., Mohan,S., Kalahar,B.K. and Williams,L.D. (2009) Peeling the onion: ribosomes are ancient molecular fossils. *Mol. Biol. Evol.*, **26**, 2415–2425.
10. Belousoff,M.J., Davidovich,C., Zimmerman,E., Caspi,Y., Wekselman,I., Rozenszajn,L., Shapira,T., Sade-Falk,O., Taha,L., Bashan,A. *et al.* (2010) Ancient machinery embedded in the contemporary ribosome. *Biochem. Soc. Trans.*, **38**, 422–427.
11. Hury,J., Nagaswamy,U., Larios-Sanz,M. and Fox,G.E. (2006) Ribosome origins: the relative age of 23S rRNA domains. *Orig. Life Evol. Biosph.*, **36**, 421–429.
12. Noller,H.F. (2010) Evolution of protein synthesis from an RNA world. *Cold Spring Harb. Perspect. Biol.*, **7**, 7.
13. Rich,A. (1962) On the problems of evolution and biochemical information transfer. In: Kasha,M. and Pullman,B. (eds), *Horizons in Biochemistry*. Academic, New York, pp. 103–126.
14. Woese,C.R. (1967) *The Genetic Code: The Molecular Basis for Genetic Expression*. Harper & Row, NY.
15. Gilbert,W. (1986) Origin of life: the RNA world. *Nature*, **319**, 618–618.
16. Crick,F.H. (1968) The origin of the genetic code. *J. Mol. Biol.*, **38**, 367–379.
17. Orgel,L.E. (1968) Evolution of the genetic apparatus. *J. Mol. Biol.*, **38**, 381–393.
18. Crick,F. (1970) Central dogma of molecular biology. *Nature*, **226**, 561–563.
19. Rost,B. (1997) Protein structures sustain evolutionary drift. *Fold Des.*, **2**, S19–S24.
20. Heinz,D.W., Baase,W.A., Zhang,X.J., Blaber,M., Dahlquist,F.W. and Matthews,B.W. (1994) Accommodation of amino acid insertions in an alpha-helix of T4 lysozyme. structural and thermodynamic analysis. *J. Mol. Biol.*, **236**, 869–886.
21. Selmer,M., Dunham,C.M., Murphy,F.V., Weixlbaumer,A., Petry,S., Kelley,A.C., Weir,J.R. and Ramakrishnan,V. (2006) Structure of the 70S ribosome complexed with mRNA and tRNA. *Science*, **313**, 1935–1942.
22. Hsiao,C. and Williams,L.D. (2009) A recurrent magnesium-binding motif provides a framework for the ribosomal peptidyl transferase center. *Nucleic Acids Res.*, **37**, 3134–3142.
23. Ban,N., Nissen,P., Hansen,J., Moore,P.B. and Steitz,T.A. (2000) The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science*, **289**, 905–920.
24. Noller,H.F., Hoffarth,V. and Zimniak,L. (1992) Unusual resistance of peptidyl transferase to protein extraction procedures. *Science*, **256**, 1416–1419.
25. Klein,D.J., Moore,P.B. and Steitz,T.A. (2004) The roles of ribosomal proteins in the structure assembly, and evolution of the large ribosomal subunit. *J. Mol. Biol.*, **340**, 141–177.
26. Ben-Shem,A., Jenner,L., Yusupova,G. and Yusupov,M. (2010) Crystal structure of the eukaryotic ribosome. *Science*, **330**, 1203–1209.
27. Petrov,A., Bernier,C., Hsiao,C., Okafor,C.D., Tannenbaum,E., Stern,J., Gaucher,E., Schneider,D.M., Harvey,S.C. and Williams,L.D. (2012) RNA-magnesium-protein interactions in large ribosomal subunit. *J. Phys. Chem. B*, **116**, 8113–8120.
28. Mohan,S., Hsiao,C., Bowman,J.C., Wartell,R. and Williams,L.D. (2010) RNA tetraloop folding reveals tension between backbone restraints and molecular interactions. *J. Am. Chem. Soc.*, **132**, 12679–12689.
29. Phillips,J.C., Braun,R., Wang,W., Gumbart,J., Tajkhorshid,E., Villa,E., Chipot,C., Skeel,R.D., Kale,L. and Schulten,K. (2005) Scalable molecular dynamics with NAMD. *J. Comput. Chem.*, **26**, 1781–1802.
30. Brooks,B.R., Brooks,C.L. 3rd, Mackerell,A.D. Jr, Nilsson,L., Petrella,R.J., Roux,B., Won,Y., Archontis,G., Bartels,C., Boresch,S. *et al.* (2009) CHARMM: the biomolecular simulation program. *J. Comput. Chem.*, **30**, 1545–1614.
31. Bowman,J.C., Azizi,B., Lenz,T.K., Roy,P. and Williams,L.D. (2012) Preparation of long templates for RNA in vitro transcription by recursive PCR. In: Conn,G.L. (ed.), *Recombinant and In Vitro RNA Synthesis: Methods and Protocols, Methods in Molecular Biology*, Vol. 941. Springer Science, LLC, pp. 19–41.
32. Wilkinson,K.A., Merino,E.J. and Weeks,K.M. (2006) Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. *Nat. Protoc.*, **1**, 1610–1616.
33. Athavale,S.S., Gossett,J.J., Hsiao,C., Bowman,J.C., O'Neill,E., Hershkovitz,E., Preeprem,T., Hud,N.V., Wartell,R.M., Harvey,S.C. *et al.* (2012) Domain III of the *T. thermophilus* 23S rRNA folds independently to a near-native state. *RNA*, **18**, 752–758.
34. Job,P. (1928) Studies on the formation of complex minerals in solution and on their stability. *Ann. Chimie France*, **9**, 113–203.
35. Cantor,C. and Schimmel,P. (1984) *Biophysical Chemistry (I-III)*. Academic Press, New York.
36. Shcherbakov,D. and Piendl,W. (2007) A novel view of gel-shifts: analysis of RNA-protein complexes using a two-color fluorescence dye procedure. *Electrophoresis*, **28**, 749–755.
37. Stumpf,C.R., Opperman,L. and Wickens,M. (2008) Analysis of RNA-protein interactions using a yeast three-hybrid system. *Methods Enzymol.*, **449**, 295–315.
38. Wurster,S.E., Bida,J.P., Her,Y.F. and Maher,L.J. 3rd. (2009) Characterization of anti-Nf-Kb RNA aptamer-binding specificity in vitro and in the yeast three-hybrid system. *Nucleic Acids Res.*, **37**, 6214–6224.
39. Cassiday,L.A. and Maher,L.J. 3rd. (2001) *In vivo* recognition of an RNA aptamer by its transcription factor target. *Biochemistry*, **40**, 2433–2438.
40. Hook,B., Bernstein,D., Zhang,B. and Wickens,M. (2005) RNA-protein interactions in the yeast three-hybrid system: affinity, sensitivity, and enhanced library screening. *RNA*, **11**, 227–233.
41. Bernstein,D.S., Buter,N., Stumpf,C. and Wickens,M. (2002) Analyzing mRNA-protein complexes using a yeast three-hybrid system. *Methods*, **26**, 123–141.
42. Nikolcheva,T. (1997) Association of a group 1 intron with its splice junction in 50S ribosomes: implications for intron toxicity. *RNA*, **3**, 1016–1027.
43. Kauffmann,A.D., Campagna,R.J., Bartels,C.B. and Childs-Disney,J.L. (2009) Improvement of RNA secondary structure prediction using Rnase H cleavage and randomized oligonucleotides. *Nucleic Acids Res.*, **37**, e121.
44. Donis-Keller,H. (1979) Site specific enzymatic cleavage of RNA. *Nucleic Acids Res.*, **7**, 179–192.
45. Hofacker,I.L., Fekete,M., Flamm,C., Huynen,M.A., Rauscher,S., Stolorz,P.E. and Stadler,P.F. (1998) Automatic detection of

conserved RNA structure elements in complete RNA virus genomes. *Nucleic Acids Res.*, **26**, 3825–3836.
46. Wu,P., Nakano,S. and Sugimoto,N. (2002) Temperature dependence of thermodynamic properties for DNA/DNA and RNA/DNA duplex formation. *Eur. J. Biochem.*, **269**, 2821–2830.
47. Merino,E.J., Wilkinson,K.A., Coughlan,J.L. and Weeks,K.M. (2005) RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE). *J. Am. Chem. Soc.*, **127**, 4223–4231.
48. Deigan,K.E., Li,T.W., Mathews,D.H. and Weeks,K.M. (2009) Accurate SHAPE-directed RNA structure determination. *Proc. Natl. Acad. Sci. USA*, **106**, 97–102.
49. Brion,P. and Westhof,E. (1997) Hierarchy and dynamics of RNA folding. *Annu. Rev. Biophys. Biomol. Struct.*, **26**, 113–137.
50. Tinoco,I. Jr and Bustamante,C. (1999) How RNA folds. *J. Mol. Biol.*, **293**, 271–281.
51. Bowman,J.C., Lenz,T.K., Hud,N.V. and Williams,L.D. (2012) Cations in charge: magnesium ions in RNA folding and catalysis. *Curr. Opin. Struct. Biol.*, **22**, 262–272.
52. Butcher,S.E., Dieckmann,T. and Feigon,J. (1997) Solution structure of a GAAA tetraloop receptor RNA. *EMBO J.*, **16**, 7490–7499.
53. Klein,D.J., Moore,P.B. and Steitz,T.A. (2004) The contribution of metal ions to the structural stability of the large ribosomal subunit. *RNA*, **10**, 1366–1379.
54. Mortimer,S.A. and Weeks,K.M. (2008) Time-resolved RNA SHAPE chemistry. *J. Am. Chem. Soc.*, **130**, 16178–16180.
55. Athavale,S.S., Petrov,A.S., Hsiao,C., Watkins,D., Prickett,C.D., Gossett,J.J., Lie,L., Bowman,J.C., O'Neill,E., Bernier,C.R. *et al.* (2012) RNA folding and catalysis mediated by iron (II). *PLoS ONE*, **7**, e38024.
56. Ryder,S.P., Recht,M.I. and Williamson,J.R. (2008) Quantitative analysis of protein-RNA interactions by gel mobility shift. *Methods Mol. Biol.*, **488**, 99–115.
57. Hayashi,T., Tahara,M., Iwasaki,K., Kouzuma,Y. and Kimura,M. (2002) Requirement for C-terminal extension to the RNA binding domain for efficient RNA binding by ribosomal protein L2. *Biosci. Biotechnol. Biochem.*, **66**, 682–684.
58. Uchiumi,T., Sato,N., Wada,A. and Hachimori,A. (1999) Interaction of the Sarcin/Ricin domain of 23 S ribosomal RNA with proteins L3 and L6. *J. Biol. Chem.*, **274**, 681–686.
59. Marquardt,O., Roth,H.E., Wystup,G. and Nierhaus,K.H. (1979) Binding of *Escherichia Coli* ribosomal proteins to 23S RNA under reconstitution conditions for the 50S subunit. *Nucleic Acids Res.*, **6**, 3641–3650.
60. Cech,T.R. (2009) Crawling out of the RNA World. *Cell*, **136**, 599–602.
61. Anderson,R.M., Kwon,M. and Strobel,S.A. (2007) Toward ribosomal RNA catalytic activity in the absence of protein. *J Mol Evol.*, **64**, 472–483.
62. Khaitovich,P., Mankin,A.S., Green,R., Lancaster,L. and Noller,H.F. (1999) Characterization of functionally active subribosomal particles from *Thermus aquaticus*. *Proc. Natl. Acad. Sci. USA*, **96**, 85–90.
63. Krupkin,M., Matzov,D., Tang,H., Metz,M., Kalaora,R., Belousoff,M.J., Zimmerman,E., Bashan,A. and Yonath,A. (2011) A vestige of a prebiotic bonding machine is functioning within the contemporary ribosome. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, **366**, 2972–2978.
64. Ostergaard,P., Phan,H., Johansen,L.B., Egebjerg,J., Ostergaard,L., Porse,B.T. and Garrett,R.A. (1998) Assembly of proteins and 5S rRNA to transcripts of the major structural domains of 23S rRNA. *J. Mol. Biol.*, **284**, 227–240.
65. Pauling,L. and Zuckerkandl,E. (1963) Chemical paleogenetics molecular restoration studies of extinct forms of life. *Acta Chem. Scand.*, **17**, 9–16.
66. Rich,A. (1971) The possible participation of esters as well as amides in prebiotic polymers. In: Buvet,R. and Ponnamperuma,C. (eds), *Chemical Evoltuion and the Origin of Life*. North-Holland Publishing Company, Amsterdam.
67. Yang,H., Jossinet,F., Leontis,N., Chen,L., Westbrook,J., Berman,H. and Westhof,E. (2003) Tools for the automatic identification and classification of RNA base pairs. *Nucleic Acids Res.*, **31**, 3450–3460.
68. Leontis,N.B. and Westhof,E. (2001) Geometric nomenclature and classification of RNA base pairs. *RNA*, **7**, 499–512.
69. Leontis,N.B., Stombaugh,J. and Westhof,E. (2002) The non-watson-crick base pairs and their associated isostericity matrices. *Nucleic Acids Res.*, **30**, 3497–3531.
70. Studier,F.W. (2005) Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.*, **41**, 207–234.