10946–10968 *Nucleic Acids Research, 2018, Vol. 46, No. 20* *Published online 12 September 2018*
doi: 10.1093/nar/gky762

How to fold and protect mitochondrial ribosomal RNA with fewer guanines

Maryam Hosseini[1], Poorna Roy[2], Marie Sissler[3], Craig L. Zirbel[4], Eric Westhof[3] and Neocles Leontis[1,*]

[1]Department of Chemistry, Bowling Green State University, Bowling Green, OH 43403, USA, [2]Center for RNA Biomedicine, Department of Chemistry, University of Michigan, Ann Arbor, MI 48109-1055, USA, [3]Architecture et Réactivité de l'ARN, Université de Strasbourg, Institut de Biologie Moléculaire et Cellulaire du CNRS, Strasbourg France and [4]Department of Mathematics and Statistics, Bowling Green State University, Bowling Green, OH 43403, USA

Received May 17, 2018; Revised August 06, 2018; Editorial Decision August 09, 2018; Accepted September 06, 2018

ABSTRACT

Mammalian mitochondrial ribosomes evolved from bacterial ribosomes by reduction of ribosomal RNAs, increase of ribosomal protein content, and loss of guanine nucleotides. Guanine is the base most sensitive to oxidative damage. By systematically comparing high-quality, small ribosomal subunit RNA sequence alignments and solved 3D ribosome structures from mammalian mitochondria and bacteria, we deduce rules for folding a complex RNA with the remaining guanines shielded from solvent. Almost all conserved guanines in both bacterial and mammalian mitochondrial ribosomal RNA form guanine-specific, local or long-range, RNA–RNA or RNA–protein interactions. Many solvent-exposed guanines conserved in bacteria are replaced in mammalian mitochondria by bases less sensitive to oxidation. New guanines, conserved only in the mitochondrial alignment, are strategically positioned at solvent inaccessible sites to stabilize the ribosomal RNA structure. New mitochondrial proteins substitute for truncated RNA helices, maintain mutual spatial orientations of helices, compensate for lost RNA–RNA interactions, reduce solvent accessibility of bases, and replace guanines conserved in bacteria by forming specific amino acid–RNA interactions.

INTRODUCTION

Mitochondria (mt) are the sites of cellular respiration, responsible for generating 90% of the ATP used by mammalian cells (1). This process generates hydroxyl radicals (OH•) and hydrogen peroxide (H_2O_2) as by-products, collectively known as reactive oxygen species (ROS), at the surface of the inner mitochondrial membrane, which is also the site of protein synthesis by mt-ribosomes (2,3). These ribosomes are distinct from those in the cytosol; they have evolved from ancestral bacterial ribosomes, thought to be most closely related to alpha-proteobacteria, by large-scale loss of peripheral RNA elements, while retaining the elements directly interacting with tRNA. A major peculiarity of the mammalian mitochondrial (mmt) translational apparatus is that all RNA components are encoded by the mitochondrial DNA (mtDNA), while all required proteins, including ribosomal proteins (rProteins) and translation factors, are encoded by the nuclear DNA, translated in the cytosol, and imported into mitochondria. It is also noteworthy that mmt genomes accumulate mutations at far greater rates than nuclear genomes (4). Significantly, the half-lives of mammalian mitochondrial (mmt) ribosomal RNA (rRNA) have been found to be considerably shorter than for cytoplasmic rRNA (5,6). The mmt-RNAs are significantly enriched in A nucleotides (nt), and to a lesser extent in U, at the expense of G (7), the nucleotide with the highest frequency in bacterial rRNA due to its versatility in forming diverse strong interactions. On the other hand, G is also the most easily oxidized, least chemically stable base (8,9). RNA oxidation damage can lead to strand breaks, loss of bases, and rapid loss of function (10). With these unusual characteristics, mmt ribosomes present a unique case study in molecular evolution of a truncated, G-poor RNA, selected to function in a highly oxidizing environment. The cryo-EM 3D structures obtained recently at near atomic-resolution for human and porcine mmt-ribosomes form the basis for comprehensive biochemical understanding of this evolution (11–13).

This study focuses on a central question linked to a second one. (i) How is it possible to reliably fold a large RNA into a complex 3D structure with fewer Gs together with a drastic reduction in rRNA? (ii) How does the ribosome maintain function in a highly oxidizing environment? We address these questions by coupling comparative study of

*To whom correspondence should be addressed. Tel: +1 419 372 2031; Fax: +1 419 372 9809; Email: leontis@bgsu.edu

RNA sequence alignments with 3D structural analysis to identify both the conserved and novel features of the mmt-ribosome. We confine our attention to the small subunit (SSU) rRNA, which mediates the crucial contacts between mRNA and tRNA that decode the mitochondrial mRNAs and ensure smooth translocation of mRNA after peptide bond formation.

We analyze how architectural RNA features are maintained in the mmt SSU 12S rRNA, despite the loss of several RNA parts and contacts, with the goal of delineating the limits in reduction of the mmt-RNAs and the mechanisms of potential compensation through increased protein content. This knowledge adds to our views on RNA structural modules and how they interact with other ribosomal components and substrates to maintain folding and stability. We compare the changes in base composition of the mmt-SSU, by reference to bacterial SSU from which they are derived, most notably, the massive overall decrease in Gs in mmt-SSU and the redistribution of some of the remaining Gs to new highly-conserved sites. We identify interactions and functional roles of conserved and altered bases. Finally, we show how individual amino acids (aa) stabilize RNA folding through specific interactions with the remaining, G-poor RNA elements.

MATERIALS AND METHODS

The strategy we applied contains the following steps: (i) Identify conserved, reduced and eliminated helical RNA elements by comparing 2D structures of mmt and bacterial small subunit (SSU) rRNA. (ii) Identify the functional network of the SSU by classifying the helical elements that interact directly ('primary') or indirectly ('secondary' or 'peripheral') with substrates (tRNA and mRNA). This involves two steps: (a) Identify structurally corresponding elements and nucleotides between the reference bacterial SSU 3D structure and the mmt-SSU 3D structure; and (b) identify mt-rProteins that substitute for reduced or lost structural elements to maintain the functional network in mmt-SSU. (iii) Identify the corresponding 3D motifs in the two structures that mediate RNA–RNA interactions in bacterial SSU and RNA–protein interactions in mmt-SSU. (iv) Compare the distributions of highly conserved nucleotides in bacterial and mmt-SSU. (v) Compare the Solvent Accessible Surface Area (SASA) of conserved nucleotides in bacterial and mmt-SSU and correlate with sequence conservation and functional roles. (vi) Identify and classify base-specific interactions in the 3D structure as G-specific, G-favorable or not requiring G.

Analysis of 3D structures

We analyzed high quality 3D structures of bacterial SSU rRNAs using PDB file 4YBB (2.1 Å (14)) for *Escherichia coli* and PDB file 4Y4P (2.5 Å (15)) for *Thermus thermophilus*. These structures were identified in the representative set of 3D structures (16,17). The *T. thermophilus* structure includes three tRNAs (one in each of the three standard functional sites). For analysis of mitochondrial structures we examined near atomic resolution cryo-EM structures with nominal resolution (3–4 Å) for porcine (*Sus scrofa*; *S. scrofa*); PDB entry: 5AJ3, 3.6 Å (13)) and human (*Homo sapiens*; *H. sapiens*); PDB entry: 3J9M, 3.5 Å (12)) mt-ribosomes, as no high resolution X-ray structures are available. Structures were visualized in SwissPDBViewer (18) and homologous proteins and RNA helices were colored identically in each structure to facilitate superposition and visualization of common elements.

3D structural alignment of bacterial and mitochondrial SSU rRNA structures

The 3D structures of the *E. coli* SSU rRNA (PDB entry: 4YBB) and mmt SSU rRNA (PDB entry: 5AJ3) were aligned manually using SwissPDBViewer. Corresponding (homologous) helices of both structures were colored in a consistent way, to facilitate comparison of superposed structures (19). Eight distinct colors suffice to color most RNA structures so that no two helices, in contact in 3D space or adjacent in the 2D diagram, share the same color. The same color scheme was used as previously presented (19). We call positions with equivalent nucleotides in the mmt and bacterial 3D structures 'Core positions' and those lacking correspondences in either of the structures, 'Peripheral positions' (shown in gray shading and black circles in Figure 1). There were 1533 and 961 positions in the reference bacterial and mmt 3D structures, respectively, of which we identified 803 as core positions. All nucleotides were manually categorized based on the type of their secondary structure element: "Helix" (Watson–Crick paired nucleotides within helices), "Flanking" (Watson–Crick paired nucleotides at the ends of helices), or "Loop" (nucleotides in hairpin, internal, and multi-helix junction loops, as well as single-stranded linkers that join RNA domains). The colored, superposed 3D structures are provided in the supplementary materials as a PDB file 'bact_mmt_SSU_rRNA_superposition.pdb' in SwissPDBViewer.

Analysis of sequence alignments

Bacterial and mitochondrial SSU rRNA alignments were downloaded as fasta files from the curated rRNA sequence alignments maintained by Robin Gutell's laboratory (http://www.rna.icmb.utexas.edu/DAT/3C/Alignment/). The bacterial SSU alignment consists of 1228 sequences. The mitochondrial alignments consist of 899 sequences of which 308 are mammalian mitochondria, comprising diverse and non-redundant representatives of mammalian evolution.

Eliminating redundancy of the alignments. Although these alignments were non-redundant in terms of sequence sampling, they were partially redundant in terms of organism sampling. A Python program was written to group the sequences by organism. The 1228 bacterial SSU rRNA sequences and 308 mmt SSU rRNA sequences grouped into 544 and 277 unique bacterial and mitochondrial organisms, respectively. When sequence microheterogeneity was present, more than one sequence was retained per organism. In these cases, the average counts of each base (G, A, C, U) were calculated for each position in the sequences of that specific organism, so that each organism was represented by

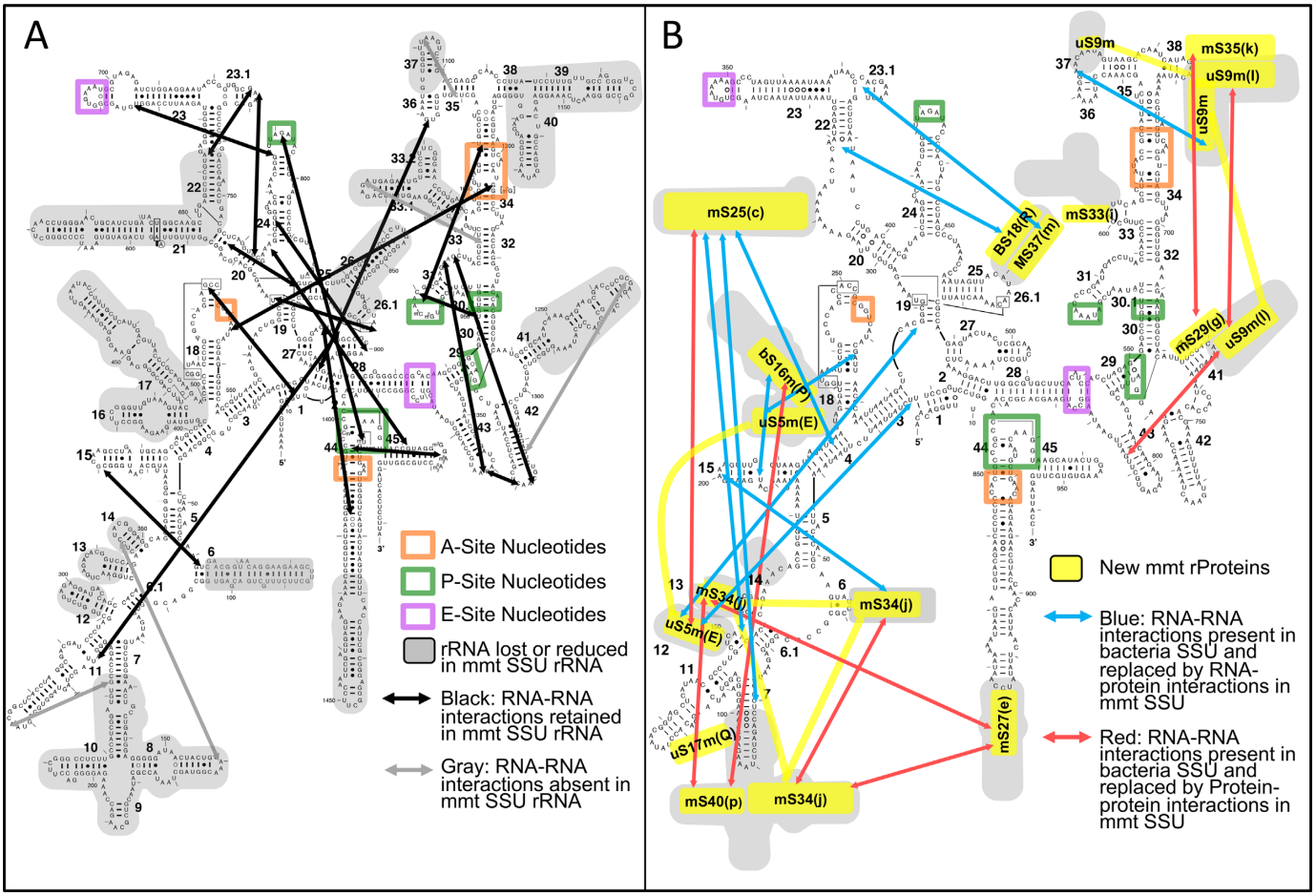

Figure 1. Long-range interactions mapped on the 2D structures of SSU rRNAs for (**A**) Bacterial SSU rRNA (*E. coli*) and (B) Mmt SSU rRNA (*S. scrofa*). Helical elements that are eliminated or drastically reduced in the mmt-SSU rRNA are shaded in grey. rProteins replacing rRNA helices are shown in yellow, with rProtein chain IDs indicated within parenthesis (PDB: 5AJ3). Long-range interactions are shown by colored arrows, with black arrows in panel (A) indicating RNA–RNA interactions found in both 3D structures. Blue and red arrows in panel (**B**) indicate RNA–RNA interactions found in the *E. coli* structure that are replaced in the *S. scrofa* structure by RNA–protein or protein-protein interactions, respectively. Nucleotides that form the tRNA A-site, P-site and E-site are indicated with orange, green and purple rectangles, respectively.

only one sequence. Thus, for example, a position can have 0.5 G and 0.5 A (see Excel file 'Data 1_ Nonredundant bacterial and mmt SSU rRNA Alignments' in supplementary materials).

Base compositions of the bacterial and mmt SSU rRNA alignments. To study the base composition of the positions in the aligned 3D structures of the bacterial and mmt SSU rRNA, all 1533 positions in the bacterial alignment that correspond to the *E. coli* SSU sequence (SeqID '>000736::*E. coli*.01 – *Escherichia coli*') and 961 positions in the mmt alignment that correspond to the *S. scrofa* sequence (SeqID '>00307::AF486866 – *Sus scrofa*') were selected. Only these columns of the alignments were used for analysis. To calculate the base composition at each position of each alignment, the total count of each type of base (including fractional counts from above) was divided by the total count of bases at that position. We refer to these compositions as %A, %C, %G, %U. The average base compositions for each alignment were calculated by averaging over all positions of the alignments. The standard deviations and ranges of base compositions were calculated from the percentage base compositions of each sequence in the alignments, see Table 1. The source data are provided in supplementary material as the Excel file 'Data 2 Bacterial and mmt SSU rRNA Nucleotide Alignment and Base composition—*E. Coli* (4YBB) and porcine mmt SSU (5AJ3).'

Base composition by structural context. To compare the base composition as a function of structural context in the bacterial and mmt SSU alignments, the 803 core positions in each alignment were grouped into three categories: helix interior, flanking, and loop positions, based on the *E. coli* (4YBB) and *S. scrofa* (5AJ3) 3D structures. The average of the base composition across the positions in each category was calculated (see Table 1).

Base composition by position. In order to compare the G and A composition by position in bacterial and mmt SSU rRNA alignments, plots were made to directly display the percentage G and A values across the 1533 and 961 columns of the two alignments. The values are shown in decreasing

Table 1. Nucleotide composition by structural context in the non-redundant bacterial and mmt SSU rRNA alignments

	Nucleotide composition in bacterial alignment (%)				
Structural Context	G	C	A	U	Total
Helix	11.4	8.8	5.6	7.4	33.2
Flanking	11.0	8.6	3.3	4.7	27.6
Loop	9.4	5.1	16.2	8.5	39.2
Total	31.7	22.6	25.2	20.6	100.0
Standard deviation of base composition across sequences in the alignment	1.54	1.42	1.55	1.40	
	Nucleotide composition in mmt alignment (%)				
Structural Context	G	C	A	U	Total
Helix	5.9	5.7	7.3	7.5	26.5
Flanking	8.2	8.3	6.7	6.1	29.3
Loop	4.3	8.7	21.5	9.6	44.2
Total	18.5	22.7	35.6	23.3	100.0
Standard deviation of base composition across sequences in the alignment	1.15	1.60	1.46	1.55	

The structural context is classified as Helix (for those nts within a helix), Flanking (for those nts at the ends of helices), and Loop (for internal or hairpin loops). The standard deviation is obtained, for example, by calculating the %G in each sequence of the alignment, and then calculating the standard deviation of those numbers.

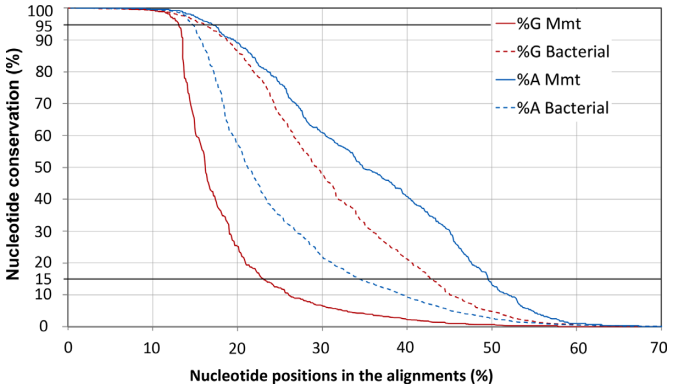

Figure 2. Nucleotide conservation (%) as a function of the fraction of nucleotide positions in the aligned rRNAs for bacterial (dotted lines) and mmt-SSUs (continuous lines). Along the X-axis is the number of positions in the sequence alignments normalized by the total number of positions in the reference rRNA, 1530 and 960 nts, respectively for *E. coli* and *S. scrofa*. Thus, the X-axis is demarcated by percentage of total positions rather than column count. The nucleotide positions are sorted separately for each type of nucleotide (G, in red, or A, in blue), from high to low fractional nucleotide composition. The Y-axis shows the conservation of A (in blue) or G (in red) in each column of the corresponding alignment. For each of the four curves, the columns are sorted from highest to lowest conservation.

order, and the x-axis is scaled according to percentage of columns rather than column count (see Figure 2).

Fraction of positions with given base composition range. For each base (A, C, G and U), the 1533 bacterial and 961 mmt nt positions (columns) in the alignments were each identified as falling into three base composition ranges: high (≥95%), medium (15–95%) and low (<15%). For each alignment, for each base, and for each composition range, the number of positions with the given composition range for the given base was divided by the number of positions in the alignment (1533 in bacteria and 961 in mmt). These numbers are reported as percentages in Supplemental Table S2A. For example, for the bacterial alignment, considering base G, we counted the number of the 1533 positions in the bacterial alignment in which the G composition was high (≥95%), divided this number by 1533, and reported the percentage in row 1 and column 1 of Supplementary Table S2A. Then, for each alignment, each base, and each composition range, we iterated over the columns which have the given base in the given composition range and added the counts of the given base in that column to a running total. At the end, we divided this running total by the number of times the base occurred in the whole alignment and reported these numbers as percentages in Supplemental Table S2B.

Counts of nucleotides by G composition range and structural context. The 803 core positions were grouped according to G composition range in each alignment: high-G (≥95%G), medium-G (15–95%G) or low-G (<15%G) and also structural context (helix, flanking and loop positions) based on *E. coli* (4YBB) and *S. scrofa* (5AJ3) structures. The number of positions that fit into each group were counted and reported in Table 2. Using the simple null model that columns of the alignment correspond randomly and uniformly to positions in the RNA structure, the expected number of columns/positions that would fall into each group was calculated using the following formula: expected number = (number of positions with the given G composition range) * (number of positions with the given structural context)/803. For example, the expected number of bacterial positions with ≥95%G and helix interior is (48+73+58)*(48+74+141)/803 = 58.6.

Correlation of G composition between bacterial and mmt SSU rRNA alignments by structural context in the core positions. The 803 core positions were divided into high, medium and low G composition ranges in bacterial and mmt alignments. There are nine possible combinations for the base composition ranges in the bacteria and mmt. Each group was assigned a color to facilitate visual display. These colors were used on the bacterial and mmt 2D structures and in 3D structures (see Figures 4 and 5). The total num-

Table 2. Number of positions with a given G composition depending on context in the core region (803 positions) of bacterial and mmt SSU rRNA alignments

		Bacterial alignment				mmt alignment			
G composition	Structural Context	Number of core positions	Expected value	Percent of core positions in group	Percent of total core positions	Number of core positions	Expected value	Percent of core positions in group	Percent of total core positions
ALL	Helix	263	-	32.8%	32.8%	238	-	29.6%	29.6%
	Flanking	230	-	28.6%	28.6%	258	-	32.1%	32.1%
	Loop	310	-	38.6%	38.6%	307	-	38.2%	38.2%
	Total	**803**	**-**	**100.0%**	**100.0%**	**803**	**-**	**100.0%**	**100.0%**
≥95%	Helix	48	58.6	26.8%	6.0%	37	34.4	31.9%	4.6%
	Flanking	73	51.3	40.8%	9.1%	56	37.3	48.3%	7.0%
	Loop	58	69.1	32.4%	7.2%	23	44.3	19.8%	2.9%
	Group total	**179**	**179.0**	**100.0%**	**22.3%**	**116**	**116.0**	**100.0%**	**14.4%**
≥15% <95%	Helix	74	42.3	57.4%	9.2%	33	27.3	35.9%	4.1%
	Flanking	32	36.9	24.8%	4.0%	36	29.6	39.1%	4.5%
	Loop	23	49.8	17.8%	2.9%	23	35.2	25.0%	2.9%
	Group total	**129**	**129.0**	**100.0%**	**16.1%**	**92**	**92.0**	**100.0%**	**11.5%**
<15%	Helix	141	162.1	28.5%	17.6%	168	176.4	28.2%	20.9%
	Flanking	125	141.8	25.3%	15.6%	166	191.2	27.9%	20.7%
	Loop	229	191.1	46.3%	28.5%	261	227.5	43.9%	32.5%
	Group total	**495**	**495.0**	**100.0%**	**61.6%**	**595**	**595.0**	**100.0%**	**74.1%**

The structural context is classified as Helix (for those nts within a helix), Flanking (for those nts at the ends of helices), and Loop (for internal or hairpin loops). Using the simple null model that columns of the alignment correspond randomly and uniformly to positions in the RNA structure, the expected number of positions that would fall into each group was calculated using the following formula: expected number = (number of positions with the given G composition range) * (number of positions with the given structural context)/803. For example, the expected number of bacterial positions with ≥95%G and helix interior is (48+73+58)*(48+74+141)/803 = 58.62.

ber of positions that fit into each group was counted. Then, each group was divided based on the mmt (5AJ3) structural context into helix interior, flanking, and loop positions and the numbers of positions were counted and all reported in Figure 3. Again, using the simple null model described above, expected counts were calculated this way:

Expected count = (number of positions in bacteria with the given G composition range) * (number of positions in mmt with the given G composition range)/803.

Analysis of solvent accessible surface areas (SASA)

Solvent accessible surface area (SASA) for the 12S and 16S rRNA from *E. coli* SSU (PDB entry: 4YBB, chain AA) and mmt-SSU (PDB entry: 5AJ3) were calculated for each atom in the presence and absence of the associated rProteins, using Gerstein's accessible surface algorithm (20) available at the High-Performance Computing server at NIH: https://hpcwebapps.cit.nih.gov/structbio/basic.html. Probe radius used for solvent accessibility calculations was 1.4 Å, typical for a water molecule.

Absolute SASA (A-SASA) of nucleobases. The 4YBB and 5AJ3 PDB files from *E. coli* (14) and *S. scrofa* (11), respectively, were saved once with and then without the rProteins. Each of these files were used as an input to the High-Performance Computing server at NIH to get the Absolute SASA (A-SASA) for each atom in the structures. A Python program was written to sum the A-SASA of the base atoms at each position and calculate the A-SASA of each base at each position in the 3D structure. The sum of the A-SASA for all nucleotides in mmt and bacterial RNA and also RNA with the protein was calculated (see Excel file 'Data 3_ Bacterial and mmt SSU rRNA Base composition and SASA-*E. Coli* (4YBB) and *S. scrofa* (5AJ3)' in supplementary materials).

Base-Weighted absolute SASA (BWA-SASA) of nucleobases in bacterial and mmt structures. For each position in the bacterial structure the absolute SASA of each base in the *E. coli* (4YBB) and *S. scrofa* (5AJ3) structures was multiplied by the percentage of the respective base (ACGU) composition of that position in the bacterial and mmt alignments to calculate the Base-Weighted Absolute SASA (BWA-SASA) of that position. The sum of the BWA-SASA for all and core positions of the bacterial and mmt alignments was calculated with and without ribosomal proteins and reported as Å^2 in Table 3A.

Estimated SASA of isolated nucleobases. In order to estimate the SASA of the base atoms of an isolated nt of each type (without any other nt or protein nearby), 30 nts of each type of nucleobase (A, U, G or C) were randomly selected and saved individually as PDB files using SwissPDBViewer, to serve as input for the SASA calculations as before. The average SASA of 30 selected isolated nts of each type of base was calculated. The numbers were G: 210.4, A:195.2, C:168.0, U:161.7 Å^2 (see Excel file 'Data 4_ SASA of thirty isolated nts in *E. coli* SSU rRNA (4YBB)' in supplementary materials).

Normalized SASA (N-SASA) of nucleobases in each position of bacterial and mmt structures. The absolute SASA for each nt in the reference bacterial and mmt 3D structures

Range of G composition in bacterial and mmt core positions (803 positions)		G Composition Color Code (Bact → mmt)	Structural Context in the mmt Structure	Observed Number of Aligned Core Positions			Expected Number of core Positions	Expected Number of Aligned Core Positions	
Bacteria	mmt			By mmt Context	Totals	% of Group		Expected Number by mmt Context	Expected Number by Bacterial Context
≥95%	≥95%	Blue (high→high)	Helix	15	**74**	41.3%	**25.9**	7.7	8.5
			Flanking	38				8.3	7.4
			Loop	21				9.9	10.0
	≥15% <95%	Cyan (high→med)	Helix	9	**40**	22.3%	**20.5**	6.1	6.7
			Flanking	16				6.6	5.9
			Loop	15				7.8	7.9
	<15%	Yellow (high→low)	Helix	17	*65*	36.3%	*132.6*	39.3	43.4
			Flanking	27				42.6	38.0
			Loop	21				50.7	51.2
≥15% <95%	≥95%	Red (medi→high)	Helix	11	20	15.5%	18.6	5.5	6.1
			Flanking	8				6.0	5.3
			Loop	1				7.1	7.2
	≥15% <95%	Pink (med→med)	Helix	13	28	21.7%	14.8	4.4	4.8
			Flanking	12				4.7	4.2
			Loop	3				5.7	5.7
	<15%	Green (med→low)	Helix	43	*81*	62.8%	*95.6*	28.3	31.3
			Flanking	20				30.7	27.4
			Loop	18				36.5	36.9
<15%	≥95%	Brown (low→high)	Helix	11	*22*	4.4%	*71.5*	21.2	23.4
			Flanking	10				23.0	20.5
			Loop	1				27.3	27.6
	≥15% <95%	Orange (low→med)	Helix	11	*24*	4.8%	*56.7*	16.8	18.6
			Flanking	8				18.2	16.2
			Loop	5				21.7	21.9
	<15%	White (low→low)	Helix	108	*449*	90.7%	*366.8*	108.7	120.1
			Flanking	119				117.8	105.1
			Loop	222				140.2	141.6
Totals			**Total**	**803**				**803.0**	**803.0**
			Helix	238				238.0	263.0
			Flanking	258				258.0	230.0
			Loop	307				307.0	310.0

Figure 3. Correlation between the G compositions of bacterial and mmt SSU rRNA alignments in the 803 core nt positions. The G composition was divided into high (>95% G), medium (15–95% G), and low (<15% G), giving rise to nine possible categories for a given core, aligned nt position. Each category is assigned a color code for use in 2D and 3D structure diagrams. For example, yellow indicates positions with high G in bacterial but low G in mmt rRNAs. The positions in each category are further separated based on the structural context (helix, flanking, and loop positions). Observed and expected numbers of positions that fit each category are reported. Bold font is used to indicate where the observed number of positions is much higher than the expected value, and italic font, that the observed number of positions is much lower than expected. The structural contexts are not exactly identical in the bacterial and mmt structures, because of truncation of some elements in the mmt rRNA, so we took the mmt structure as the reference. Using the simple null model that columns of the alignment correspond randomly and uniformly to positions in the RNA structure, the expected number of positions that would fall into each group was calculated using the following formula: expected number = (number of positions with the given G composition range) * (number of positions with the given structural context)/803.

was divided by the estimated SASA for an isolated base to determine a normalized SASA (N-SASA), which is more robust to the varying sizes of the bases.

Base-weighted normalized SASA (BWN-SASA) of nucleobases in bacterial and mmt structures. For each core position, for each 3D structure and the corresponding alignment, and for each base type (ACGU), the product of the base composition with N-SASA gives Base-Weighted Normalized SASA (BWN-SASA), reported as a percentage (Table 3B, Supplementary Tables S5A and S5B). Bases in decreasing order of GWN-SASA in bacterial and mmt structures were listed (Supplementary Tables S5A and S5B).

Table 3A. Base-Weighted absolute SASA (BWA-SASA) (Å^2) of all and core (803 positions) nts in the bacterial (*E. coli* PDB file 4YBB) and mmt (*S. scrofa* PDB file 5AJ3) SSU rRNA structures with and without proteins

Base-Weighted Absolute SASA (BWA-SASA) (Å^2) of all positions								
	RNA only				RNA + protein			
Structure	G	A	C	U	G	A	C	U
Bacteria (*E.c.* file 4YBB)	16971.8	12906.7	10999.8	12052.0	15840.5	11303.7	10376.4	10795.4
mmt (*S.sc.* file 5AJ3)	6099.0	15872.0	9267.5	9544.7	4876.2	11539.8	6960.5	7582.1
Base-Weighted Absolute SASA (BWA-SASA) (Å^2) of core positions (803 nucleotides)								
	RNA only				RNA + protein			
Structure	G	A	C	U	G	A	C	U
Bacteria (*E.c.* file 4YBB)	8617.2	6424.3	5729.4	4901.1	7784.0	5343.1	5219.0	4095.7
mmt (*S.sc.* file 5AJ3)	5435.6	10211.7	6369.3	6485.4	4387.1	7728.1	4889.7	5342.5

Table 3B. Base-weighted normalized SASA (BWN-SASA) (%^{2}) by structural context in the complete bacterial and mmt SSU rRNA alignments with and without protein. One-electron Redox Potential is reported in parenthesis for each type of base

Base-weighted normalized SASA (BWN-SASA) (%) in the complete bacterial SSU rRNA alignment								
	G (1.88)		A (2.04 V)		C (2.28 V)		U (2.48 V)	
Structural context	rRNA only	rRNA with protein	rRNA only	rRNA with protein	rRNA only	rRNA with protein	rRNA only	rRNA with protein
Helix	15.3 ± 5.7	14.4 ± 5.3	16.4 ± 6.2	15.1 ± 6.2	15.9 ± 6.3	15.1 ± 5.7	17.7 ± 5.4	16.7 ± 6.1
Flanking	14.1 ± 7.2	12.9 ± 7.3	13.0 ± 7.1	11.9 ± 7.1	14.2 ± 7.5	13.1 ± 7.4	18.6 ± 7.4	17.3 ± 13.6
Loop	22.8 ± 16.5	20.7 ± 15.5	18.8 ± 18.8	15.5 ± 15.7	29.7 ± 24.9	25.8 ± 23.9	29.0 ± 24.6	24.1 ± 22.8
All positions	17.1 ± 11.1	15.7 ± 10.6	17.5 ± 15.7	15.0 ± 13.2	18.4 ± 14.5	16.8 ± 13.7	22.6 ± 18.3	19.9 ± 16.9
Base-weighted normalized SASA (BWN-SASA)(%) in the complete mmt SSU rRNA alignment								
	G (1.88 V)		A (2.04 V)		C (2.28 V)		U (2.48 V)	
Structural Context	rRNA only	rRNA with protein	rRNA only	rRNA with protein	rRNA only	rRNA with protein	rRNA only	rRNA with protein
Helix	14.0 ± 5.7	12.4 ± 6.0	16.4 ± 4.9	14.9 ± 5.5	17.1 ± 5.5	15.8 ± 5.8	17.6 ± 5.1	16.4 ± 5.7
Flanking	13.6 ± 6.0	11.0 ± 6.4	18.9 ± 12.8	15.4 ± 11.3	14.9 ± 10.6	13.0 ± 10.4	17.8 ± 1.4	15.2 ± 11.8
Loop	29.7 ± 22.0	21.6 ± 20.3	28.8 ± 24.2	19.3 ± 17.5	39.4 ± 27.2	26.0 ± 23.0	35.9 ± 7.3	25.9 ± 20.9
All positions	17.5 ± 13.8	13.9 ± 12.7	24.4 ± 20.5	17.7 ± 14.8	24.9 ± 21.6	18.7 ± 16.9	25.2 ± 20.7	20.0 ± 15.8

Then, the average BWN-SASA of the positions in the interior helix, flanking, and loop positions were calculated. For this, the sum of the BWN-SASA of the positions in each structural context was calculated and then divided by the sum of the base composition of those positions.

$$\text{average BWN} - \text{SASA of nts in a structural context:} \frac{\Sigma\ \text{BWN–SASA of each position}}{\Sigma\ \text{Base composition of each position}}$$

The formula for the standard deviation (STDEV) calculation is shown below:

$$\sqrt{\frac{\Sigma\ (N - \text{SASA of each } nt - \text{average BW}N\text{\_SASA})^2 * \text{ Base composition of each position}}{\Sigma\ \text{Base composition of each position}}}$$

The calculated numbers were reported in Table 3B. The same method was used to calculate the GWN-SASA of the positions that belong to each of the nine colored groups (see Supplementary Table S6).

RNA–RNA and RNA-protein interactions

All RNA–RNA and RNA-protein interactions for aligned nts in the 8 colored groups in *E. coli* (PDB file 4YBB) and *S. scrofa* (5AJ3) were studied using SwissPDBviewer and compiled, together with G-specific thermodynamic stabilization data for each nucleotide position in Supplementary Table S4. The summary of these tables is reported as Figure 6. The numbers in Figure 6 are the fraction of the nts forming each type of interaction in each colored composition group. The average number of RNA–RNA interactions are reported as a percentage for non-Watson–Crick or tertiary Watson–Crick base-pair, base-phosphate interactions, junction and tertiary stacking, and perpendicular stacking interactions. Five colors are used to highlight the differences between the *E. coli* and *S. scrofa* structures. The numbers in the range of ±15% are considered as equal in *E. coli* and *S. scrofa* and are colored in blue. Differences are colored in: Dark yellow: two or more times higher number in *E. coli*; light yellow: somewhat (>15%) larger in *E. coli*, dark red: two or more times higher number in *S. scrofa*; light red: somewhat (>15%) larger in *S. scrofa.*

The interactions that were preserved in both structures and those that were not present in either the bacterial or the mmt structure were counted for the high G positions (blue, highly conserved in both bacteria and mmt, cyan, highly conserved in bacteria with variable range in mmt, and yel-

low, highly conserved in bacteria and very poorly in mmt, see Figure 3).

Geometric analysis of RNA–RNA and RNA–protein interactions. RNA–RNA interactions involving core nucleotides from bacterial and mmt alignments were identified from analysis using FR3D (21). Specific interactions between the nucleotide and amino acid (aa) residues were determined on the basis of the stereochemistry of the amino acid sidechain or peptide backbone and the part of the nucleotide participating in the interaction. A suite of Python programs were developed to detect, classify and annotate residue level RNA–protein interactions in ribosomes (22). Our programs analyze mmCIF files (PDB entries: 5AJ3 for mmt-SSU and 4YBB, chain AA for *E. coli* SSU) to detect and classify RNA–protein interactions as (i) RNA base-aa stacking interactions, (ii) complex 'bidentate' interactions where an amino acid residue bind to two stacked bases simultaneously, (iii) edge-to-edge interactions forming 'pseudo-pairs' by two H-bonds from the same part of the amino acid to the nucleotide. We further categorize these three types of interactions based on the interacting parts of the nucleotide (base or sugar-phosphate backbone) and of the amino acid (sidechain or peptide backbone).

Geometric analysis of protein-protein interactions. Protein-protein contacts were determined by careful inspection of the three-dimensional structure of the mmt SSU ribosome (PDB entry 5AJ3) and *E. coli* SSU ribosome (PDB entry 4YBB). Amino acid residues within 4 Å of any atom of each protein chain was defined as contacted.

RESULTS

Section One: Overview of the RNA–RNA and RNA–protein interaction networks

Compared to ∼1530 nts for the bacterial 16S rRNA, represented by *E. coli* in Figure 1A (14), the mmt-SSU 12S rRNA, represented in Figure 1B by the porcine cryo-EM structure (11), contains just ∼960 nts, a 37% reduction. The folding of the bacterial SSU rRNA body and head is stabilized by dense networks of tertiary RNA–RNA contacts. There are 56 RNA–RNA contacts in the body, 20 in the head, five between body and head, and three between the neck (h28) and the head or body (Figure 1A for bacteria). Loss or reduction of helical elements of 12S rRNA reduces the 84 all-RNA contacts in 16S rRNA from 56 to 25 in the mmt-SSU body, from 20 to 13 in the head and from 5 to 4 between body and head (black arrows in Figure 1A and Supplementary Table S1). All three RNA–RNA contacts of the neck are conserved in mmt. Remarkably, much of the contact map of bacterial SSU is maintained in mmt by replacing some of the 31 lost RNA-only interactions with new RNA–protein or protein-protein interactions (blue and red arrows in Figure 1B and Supplementary Table S1), for a total of 42 interactions of some kind in the mmt body and 16 in the head. A total of 18 RNA–RNA interactions, 14 in the body and 4 in the head, all of which involve peripheral elements, are absent in mmt (gray arrows in Figure 1A and Supplementary Table S1). Overall, 66 out of 84 long-range RNA–RNA contacts in bacterial SSU are either conserved in mmt SSU (45) or replaced by RNA–protein (13) or protein-protein interactions (8), thus largely maintaining the interaction networks in mmt-SSU. In summary, based on analysis of the published mmt SSU structures (11–13), we notice the following points. (i) Where two RNA elements that interact in bacteria are preserved in mmt SSU, the RNA–RNA interaction between them is also present in mmt, although it may be modified. (ii) Where one of the interacting elements is lost or reduced in mmt, the RNA–RNA interactions in bacteria are usually replaced by RNA–protein interactions in mmt. (iii) Many lost or reduced RNA elements in mmt SSU are replaced in part or in whole by new rProteins or rProtein extensions in mmt, some of which form protein-protein interactions that substitute for lost RNA interactions (see Figure 1 and Supplementary Table S1).

Section Two: Selective changes in nt composition of mmt-SSU rRNA

The mmt-SSU rRNAs have also suffered a large-scale, global loss of guanines. The average percentage of guanines in the mmt sequences is only 18.5 ± 1.15% of all nucleotides, compared to 31.7 ± 1.54% for bacteria (Table 1). The standard deviation (stdev) for mmt sequences (±1.15%), as well as the range (4.2%), are considerably smaller than for other mmt and for all bacterial nucleotides (stdev: 1.4–1.6%; range: 5.2–11.9%), suggesting that stronger constraints operate on the total number of Gs in mmt sequences. While G is the most abundant base in the bacterial SSU alignments, A is most common in the mmt alignments, increasing from 25% of nts in bacteria to 36% in mmt. Significantly more As and fewer Gs occur in mmt loops. The percentages of Us increase from 21% to 23%, but those for Cs remain at 23% in both alignments; Cs shift from helix to loop positions, with little change in flanking positions. In mmt, the number of WC paired Gs and Cs is almost identical, as there are far fewer GU pairs than in bacteria. Overall, for mmt sequences, Gs decrease in all contexts, but more so in loops and helix positions than at flanking positions.

Section Three: Concentration of Gs in selected positions of mmt SSU

Figure 2 shows A and G composition plotted against the percentage of aligned positions (x-axis) in the curated bacterial and mmt SSU rRNA alignments (23–25). These plots reveal the unusual distribution of G in the mmt SSU sequences: The curve for G in mmt drops sooner and more steeply than for any other nucleotide in either the mmt or bacterial alignments. Supplementary Table S2A quantifies the differences in the distribution of bases among sites with high- (≥95%), medium- (15–95%) and low- (<15%) G composition. While 76% of nt positions have low-G in mmt sequences (compared with 56% in bacteria), 65% of Gs in the 277 mmt sequences occur in vertically aligned positions with high-G composition (compared with 52% in the 577 bacterial sequences), but only 28% of Gs in mmt occur in positions with medium-G, compared to 45% in bacteria and 51% of As in mmt (Supplementary Tables S2A and B). In summary, Gs in the mmt alignment are concentrated in

fewer nt positions than are Gs in the bacterial alignment or any other base in either alignment. In subsequent sections, we attempt to account, not only for the reduction of the number of Gs, but also for their restriction to specific positions in the alignments and 3D structures.

Structural alignment of bacterial and mmt SSU rRNA. We superposed the *E. coli* (14) and *S. scrofa* (11) SSU rRNA 3D structures to identify all nt positions that are structurally equivalent at the nt level, and thus establish a correspondence between specific columns of the bacterial and mmt rRNA sequence alignments. We identified 803 out of 960 nt positions in the *S. scrofa* SSU rRNA and 1530 in the *E. coli* SSU rRNA, that are structurally equivalent. We refer to these as 'core aligned nucleotides' of the SSU rRNAs. The remaining positions, where bacterial RNA elements are lost or reduced or where new motifs have been inserted in mmt, lack corresponding aligned positions and are shaded gray in the 2D diagrams.

Overall, ∼30% of core nucleotides occupy 'Helix' and another 30% 'Flanking' positions in both mmt and bacteria (Table 2), while Loop positions make up almost 40% of core nucleotides. There are about 3% more flanking positions and 3% fewer helix positions in mmt, as result of the truncation of peripheral RNA elements (see Supplementary Table S3). Although fewer core positions in mmt have high G than in bacteria (116 versus 179 positions), a higher percentage of these occur in flanking base-pairs (48% mmt versus 41% bacteria) and far fewer belong to the loop category (20% mmt versus 32% bacteria). In absolute numbers, just 23 high-G positions remain in loops in mmt (Table 2).

To correlate the G composition between the bacterial and mmt alignments, we partitioned the core nucleotide positions into nine disjoint sets, according to the G composition of the corresponding columns: high-G (>95%), medium-G (15–95%) or low-G content (<15%). When we refer to one of these sets, we first give the range of G composition for bacteria and then for mmt, e.g. 'high→low' means high conservation of G in the bacterial and low conservation of G in mmt alignments. These sets are color-coded as shown in Figure 3 to facilitate display by nucleotide of the correlated G composition on 2D rRNA diagrams (see Figure 4 and Supplementary Figure S1). Positions for which the G composition is <15% in both alignments are left unmarked. A simple model for the high-, medium-, and low-G positions is that they are located randomly, uniformly, and independently in bacteria and mmt. From the total number of positions in each composition range in bacteria and mmt, we can calculate the expected number in each set based on this model, see Figure 3. This is a helpful reference point because the distribution over high-, medium-, and low-G composition is different in bacteria than in mmt.

We first consider the 179 core high-G positions in bacteria (colored blue, cyan, and yellow in Figure 3). If high-G core positions were randomly distributed in mmt, we would expect to find just 26 positions in the high→high set (colored in blue), but 74 are observed, significantly more than expected. Similarly, we expect just 20.5 positions in the high→medium set (cyan) but observe 40. However, fewer positions than expected are observed for the high→low set (yellow; 133 expected, 65 observed). In all, 41% of the high-G positions in bacteria are also high in mmt, 36% are low-, and just 22% are medium-G in mmt. Taken together, these data suggest that many highly conserved Gs in bacteria are essential to the function of rRNA, and therefore are also conserved in mmt. On the other hand, where Gs are not indispensable, they tend to be largely eliminated from mmt sequences. We consider reasons for these conservation patterns below.

Conversely, we consider the 116 core positions that have high-G in mmt (colored blue, red and brown): fully 64% (74/116) are also high-G in bacteria (blue), suggesting that at nearly $\frac{2}{3}$ of positions where Gs are indispensable in mmt they are also crucial in bacteria. Still, 17% of high-G positions in mmt are also medium in bacteria (red), and 19% are low in bacteria (brown). While medium→high (red) positions occur at the expected frequency, low→high (brown) occur less than expected (72 expected, 22 observed). New high-G positions (red or brown) occur almost exclusively at flanking or helix positions (Figure 3). Overall, most high-G positions in mmt coincide with those in bacteria, while new high G positions (red and brown) are relatively rare (42/803, 5.2% of core nucleotide positions, see Figure 3), but not insignificant.

Looking next at low-G positions in mmt (yellow, green and white), we find that just 18% of these (146) are medium to high-G in bacteria (yellow or green, see Figure 3). Below, we explore reasons for low-G in mmt where G appears to be indispensable or preferred in bacteria.

Interestingly, fully 74% of core positions that are medium G in mmt are either high-G (40 cyan/92 = 43.5%) or medium-G (28 pink/92 = 30.4%) in bacteria, considerably more than expected. This indicates that for significant numbers of core positions, where G is indispensable in bacteria, G is favored in mmt but not indispensable. We also explore reasons for this pattern of partial conservation below. By contrast, there are fewer than expected low→medium (orange; 24 versus 57), just as noted for low→high positions.

Focusing next on the 307 loop nucleotides in core positions of mmt (Figure 3), 91% of those with high-G in mmt (blue, red, or brown) also have high-G in bacteria (blue), and 78% of those with medium-G in mmt (cyan, pink, or orange) have medium- to high-G in bacteria (cyan or pink, see Figure 3). Nonetheless, almost half of core loop positions (49%) that are medium- to high-G in bacteria are low-G in mmt (green or yellow), pointing to the large-scale loss of Gs from loops in mmt. The key question that emerges from this analysis is to identify the factors that make certain Gs indispensable in mmt while fingering others for substitution by other bases.

In summary, we find that 64.9% of Gs in core positions of the mmt alignment occur at high-G positions, compared to just 28.2% at medium-G and 6.9% at low-G positions (Supplementary Table S2B). This is evidence that most Gs remaining in mmt sequences are highly constrained and concentrated at selected nucleotide positions with high-G composition in bacterial alignments. These findings suggest that

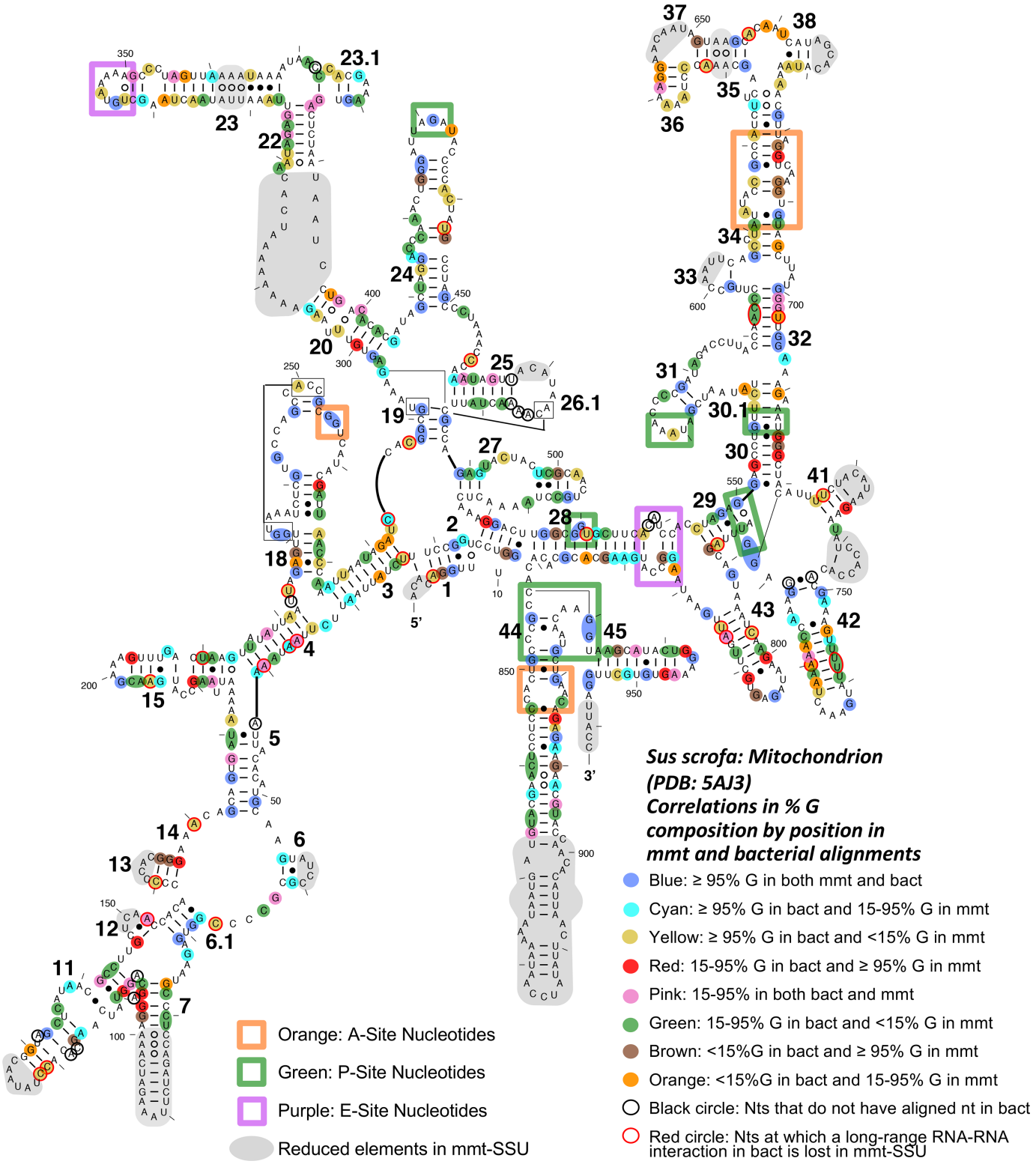

Figure 4. *S. scrofa* 12S rRNA 2D diagram showing the correlation between G conservation at aligned positions in the bacterial and mmt SSU rRNA alignments. Blue, Cyan, yellow, red, pink, green, brown and orange colors indicate for each aligned nt position the range of G composition in the bacterial and mmt alignments, using the categories defined in Figure 3. The tRNA A-, P- and E-sites are indicated by rectangles colored as in Figure 1. Black circles mark individual nucleotides that have been deleted (or inserted) in one of the SSU structures are indicated with black circles, while red circles indicate nt positions at which long-range interaction present in the bacterial 3D structure have been lost in the mmt SSU rRNA.

these positions are under strong positive selection because they require the presence of G either to stabilize the structure of the SSU rRNA or to enable its function. Conversely, positions having high-or medium-G in mmt but not in bacteria should be particularly informative with regard to the evolution of RNA structure and function. The structural and functional implications of these observations are explored in subsequent sections.

Section Four: Thermodynamic and structural stabilization of mmt RNA structure by conserved Gs

In most structured RNA molecules, G is the most common base. It can form up to seven hydrogen bonds, compared to just five for other bases. Through the two adjacent H-bond donors on its Watson–Crick edge (N1 and N2), G uniquely forms strong ionic interactions with phosphate groups (26). Finally, G has the largest dipole moment and largest planar surface area for base stacking interactions (27). Consequently, many base interactions are specific to G or more favorable than similar ones formed by other bases (see Supplementary Material Table S6 for descriptions of G-favored and G-specific interactions).

We hypothesize that at least some of the remaining Gs in mmt SSU rRNA are also selected by evolution to contribute to thermodynamic stability, to ensure correct folding and structural stability at ∼37°C in spite of far fewer Gs (28). Given that most Gs in mmt occur at high-G positions, and are present in almost all mmt sequences, we focus primarily on these locations, i.e. blue, red and brown positions in Figures 3 and 4.

A disproportionate number of high-G nucleotide sites in mmt form flanking G = C Watson–Crick pairs, thus stabilizing helices from fraying. The Turner thermodynamic parameters for nearest neighbor Watson–Crick pairs predict greater stabilization for Gs occupying the 5′-position of flanking pairs, regardless of which base the 5′-G pairs with (C, U, or A) or which base-pair follows in the helix (29). We find that one in five flanking positions in mmt are high-G, regardless of the conservation in bacterial alignments (i.e. blue, red, and brown, see Supplementary Table S3 and Figure 4). Of these, 5′-Gs occur twice as often as 3′-Gs in mmt, showing a clear preference for the more stable orientation of G in flanking base-pairs, see Supplementary Table S3. Moreover, we find that almost all flanking pairs in mmt with 3′-Gs form some kind of G-specific long-range RNA or protein interaction. Overall, more 5′-flanking nucleotide positions are high-G in both SSUs, while more 3′-flanking positions are high in pyrimidines.

In Figure 4, nt positions are colored using the G color scheme from Figure 3. (Positions with low percentage G in both bacteria and mmt are white and so uncolored.) We observe that most high-G positions forming Watson–Crick pairs in mmt (i.e. blue, red or brown flanking or helix positions, see Figure 4) occur adjacent to other high-G positions forming Watson–Crick pairs. Those that involve two blue (high→high) Gs are common to bacteria, but others are unique to mmt SSU. In fact, large fractions of red (medium→high) and brown (low→high) positions, show these associations, more than for cyan (high→medium) or yellow (high→low), which are high-G only in bacteria (see Thermodynamic enhancement column in Supplementary Table S4). We observe that they occur preferentially as the adjacent pairs 5′-GC/3′-CG (–3.42 kcal/mol) or 5′-GG/3′-CC (–3.26 kcal/mol), for which the experimentally determined, nearest neighbor free energy values are more negative than for the 5′-CG/3′-GC arrangement (–2.36 kcal/mol) (29). Thus, for blue positions, we observe thirteen 5′-GC/3′-CG and ten 5′-GG/3′-CC juxtapositions, compared to just eight of the less stable 5′-CG/3′-GC adjacent pairs (see Supplementary Table S3). Again, wherever the less stable arrangement occurs, long-range RNA or protein interactions are observed that require G at one of the positions (Supplementary Table S4). A number of new stabilizing interactions occur in mmt thanks to new high-G brown or red positions, which more than compensate for the loss of other stabilizing interactions present only in bacteria, due, for example, to adjacent yellow positions with low-G in mmt.

Taken together, we find that almost 75% (40 of 53) of helical blue (high→high) positions in mmt participate in one or more types of thermodynamic stabilization, including twenty six 5′-flanking Gs, thirteen 5′-GC/3′-CG and ten 5′-GG/3′-CC nearest neighbor interactions (Supplementary Table S3). Likewise, 19 of 21 brown (low→high) helical positions (90.5%) and 13 of 19 red (medium→high) helical positions (68.4%) exhibit one or more forms of thermodynamic stabilization. By contrast, for cyan (high→medium) and yellow (high→low) positions, the fraction of stabilizing arrangements is lower (56% for cyan positions and 54% for yellow positions, see Supplementary Table S3). In conclusion, thermodynamic stabilization plays an important role in the location of many high-G positions in both mmt and bacterial structures, but especially for mmt. Notably, there is a significant number of runs of three conserved paired Gs in mmt-SSU, something not observed to the same degree in bacterial SSU: In h1, G6, G7 and G18; in h7, G102, G103 and G105; in h13, G165-G167; in h19, G288, G289, G485, and G391; in h28, G516, G517, G 836 and G519; in h30, G553, G717, G716 and G715; in h34, G682, G683 and G618; and in h35 and h36 G642, G643 and G651.

Pink nucleotides (medium→medium) occur at flanking positions more than is expected statistically (12 versus 4.7), and eight of these nts (67%) occur at the 5′-ends of helices, contributing to the thermodynamic stability, (see Figure 3 and the column labeled ”Thermodynamic Enhancement in Mmt” in Supplementary Table S4). The pink positions located at the 5′-ends of helices are enriched in G in mmt alignments, with 45% average G composition compared to 35% for helix interior pink positions inside of helices. Pink nucleotides also exhibit the second highest tandem 5′-GG/3′-CC or 5′-GC/3′-CG of all the colored groups (68%), and are usually associated with other positions that have 15–95% G in mmt or bacteria, see Figure 6. While each of these stabilizing interactions is not present simultaneously in all mmt sequences, all mmt SSUs benefit from at least some of them.

A final observation is that mmt SSU rRNA sequences share a large reduction, compared to *E. coli* SSU, in the

number of GU pairs, retaining only those needed for crucial tertiary interactions (30–32). GU pairs are widely considered roughly comparable to AU pairs in thermodynamic stability, although the Turner nearest neighbor parameters suggest that, averaged over all nearest neighbor arrangements, AU pairs offer greater stabilization than GU in helices (33). This effect may be operating in both thermophilic and mmt rRNAs, leading to replacement of GU in mmt, except where needed to mediate tertiary interactions (31). In conclusion, it appears that a significant number of high-G positions in mmt provide thermodynamic stabilization to the 3D structure, in addition to other roles they may play, as discussed in subsequent sections.

Section Five: Localization of highly conserved Gs at functional sites

Next, we consider the hypothesis that guanines tend to be retained in mmt rRNA within or close to functional sites. The functional sites of the SSU rRNA are indicated approximately in the 2D diagrams by colored rectangles (see Figure 4 and Supplementary Figure S1). Remarkably, the helical elements of SSU rRNA that contain these sites are conserved intact in mmt SSU. These 'primary elements' include h18, h23, h24, h28, h29, h30, h31, h34, h44, and h45, and contain a significant fraction of the high-G positions in mmt (61/116 = 53%). Most of the bases that interact directly with the tRNA or mRNA substrates occupy loop positions. Tellingly, 10 of the 23 high-G loop positions found in the entire mmt rRNA occur at the functional sites. Of the 21 yellow (high→low) Gs in loops (see Figure 3), only 3 occur in functional sites and all of them are replaced by conserved As in mmt sequences (G693→A346 in h23, G966→A571 in h31, and G1053→A613 in h34). Two of these As interact directly with substrate RNAs (tRNA or mRNA) by base stacking, a less base-specific interaction that still requires purine. These As belong to the most exposed bases in the SSU (see below). The third A participates in a base triple in h34 that helps structure the A-site in the head (34,35). The GGC triple conserved in bacteria is replaced in mmt by the isosteric ACG triple that achieves the same structure with just a single G instead of two (also see section 7 below).

The primary elements containing substrate-binding sites interact extensively with 'supporting elements.' Listed with their primary elements, these include: h1 (h18), h2 (h28), h22 (h23), h20 (h24), h27 (h24, h44), h32 (h31, h34), h42 (h30, h31), h43 (h29, h31), and h45 (h24, h44). Many of the remaining high-G positions in mmt occur in supporting elements (32/55) for a total of 93/116 (80%) of high-G positions occurring either in primary or supporting elements. In conclusion, high-G sites, especially loop Gs, cluster near functional sites, either close in the 2D (primary elements) or in the 3D structure (supporting elements). In fact, 16 of 23 high-G loop positions occur in the primary or supporting elements. In Figure 4 and Supplementary Figure S1 the aligned nts in the *E. coli* and porcine 2D diagrams are colored according to G-composition, using the colors in Figure 3. In addition, Supplementary Figure S2 shows the 3D structure of the porcine SSU with nts colored by G composition (Panel A) or functional sites (Panel B).

Section Six: Evidence for selection pressure to reduce oxidative damage to Gs

In this section, we consider the hypothesis that guanines tend to be lost from sites of high solvent exposure, perhaps to minimize oxidative damage and reduce energetically costly ribosome turnover within mitochondria. We note that, among the four nucleobases, guanine is the most sensitive to oxidation (10). To test whether the loss of Gs in mmt is correlated with solvent exposure, we calculated the normalized solvent accessible surface area (N-SASA) for the RNA bases in the *E. coli* and *S. scrofa* 3D structures, with and without rProteins and substrates, weighting the accessibilities of aligned positions using the G-composition from the respective alignments (see Methods). We call this the G-weighted normalized SASA (GWN-SASA). GWN-SASA can be thought of as a susceptibility to G oxidation at each nucleotide position over all sequences in the alignment. The range of N-SASA values for Watson–Crick paired Gs in helices is 11–19% (the GWN-SASA values are all less than N-SASA values unless the percentage G is 100%). Thus, we use 20% GWN-SASA as a reference point to identify highly exposed nucleotides. The average GWN-SASA for the 30 most exposed positions in mmt SSU rRNA is just 26.5% (18.9% with protein), compared to 37.1% (33.6% with protein) in bacteria. Gs in mmt are therefore significantly less exposed, with a shielding effect by proteins greater for the most exposed positions of mmt compared to bacteria (see Supplementary Tables S5A and S5B).

Highly exposed positions. Relatively few positions in both alignments have GWN-SASA values ≥20%, but significantly more do so in bacteria (46 in bacteria versus 20 in mmt, see Supplementary Tables S5A and S5B). In mmt, all but two of the 30 most exposed G-positions are high-G (i.e. blue, red or brown) and seven out of nineteen with GWN-SASA ≥20%[2] occur within functional sites (Supplementary Table S5A). By contrast, of the 30 most exposed in bacteria, half are low-G in mmt (yellow or green) and seven are medium-G (cyan or pink), see Supplementary Table S5B. Most of the highly exposed Gs that remain in mmt are found at the tRNA or mRNA binding sites, including G256 in the A-site (*E. coli* equivalent G530), G530 (*E. coli* G926) and G782 (*E. coli* G1338) at the P-site, and G344 (*E. coli* G691) in the E-site (Supplementary Table S5A). Notably, some of the most highly exposed Gs in functional sites in bacteria have mutated to other bases in mmt, including *E. coli* G693 (*S. scrofa* A346) in the E-site, G966 (A571) in the P-site, and G1491 (C917) in the A-site. The exposure of G1491 depends on the conformation of the mobile bases, A1492 and A1493, which depends on A-site occupancy by tRNA and mRNA. These data clearly show a strong tendency among the most exposed G positions in bacteria to mutate to other bases in mmt or to recruit new protein protections, supporting the hypothesis that, at exposed positions, only the most functionally essential Gs are retained by mmt SSU.

Solvent exposure across all nucleotide positions. Next, we extended the analysis of solvent accessibility to all nucleotide positions to explore the validity of these conclusions for the structure as a whole, including the periph-

eral elements. Table 3A shows the Base-Weighted-Absolute-SASA (BWA-SASA) for all positions in the *E. coli* (14) and *S. scrofa* (11) structures with and without r-proteins. The GWA-SASA in presence of protein over all positions in mmt (4876 Å^2) is more than three times smaller than for Gs in bacteria (15840.5 Å^2) and significantly smaller than for the BWA-SASA of A, C or U in mmt, which ranges between 6960 and 11540Å^2. Remarkably, only 5 of the 60 most exposed nucleotide positions in the mmt structural alignment are Gs, compared to 15 out of 60 for bacteria, see Supplementary Table S5C.

Table 3B shows the base-weighted normalized SASA (BWN-SASA) values averaged over each structural context in the molecule (helix, flanking, loop), manually annotated for SSU 3D structures 4YBB (chain A, *E. coli*) and 5AJ3 (chain A, *S. scrofa*), and weighted according to the percentage G, A, C, or U in the corresponding columns of the alignments. The bases are listed in Table 3B in the order of decreasing reactivity to oxidation, G, A, C, U. Comparing values for all positions in the two alignments, we see that the BWN-SASA systematically increases (from ∼17% to ∼25%) as the ease of oxidation decreases: G has lowest values and is most easily oxidized, followed by A, then C, and finally U. This trend holds roughly for both the RNA alone, reflecting exposure during and shortly after transcription, and when protein is included. Protein binding reduces solvent exposure substantially for all bases, but especially for Gs and As in mmt. The distribution of BWN-SASA values by base type is much more uniform in bacteria than in mmt. Table 3B shows that Gs at flanking positions in mmt are less exposed than other bases, with or without protein (Cs and Us in loops are especially exposed, both in mmt and in bacteria). The Hoogsteen edge of G contains C8, the atom that is most sensitive to modification when G is oxidized and that is more exposed when the G is on the 5′-end of a helix than on the 3′-end because of the right-handedness of helices. The average N-SASA for high-G positions in mmt is 13.5% at 5′-ends of helices versus 11.1% at 3′-ends, a small but distinct difference (with protein this difference is largely eliminated (10.1% versus 9.4%)).

Supplementary Table S6 shows the average GWN-SASA for nucleotides belonging to the nine colored sets (i.e. the aligned core nucleotides of mmt and bacterial SSU rRNA). SASA values in parentheses are normalized but unweighted by G composition. The average GWN-SASA is lowest for blue (hight→high) positions (14.7% bacteria and 15.6% mmt), reflecting the large number of tertiary interactions they form that protect the bases from solvent (see below). In addition, many blue nts interact with rProteins, tRNA or mRNA, further decreasing their SASA to ∼12% average GWN-SASA, identical in the mmt and bacterial structures, consistent with the fact that essentially the same contacts are formed by blue bases in both structures. Low values of average GWN-SASA are also observed for red (medium→high) and brown (low→high) sites in both structures. Protein selectively reduces GWN-SASA for red and brown positions in mmt compared to bacteria, reflecting new protein protections at these sites. On average, the red nucleotides display the least solvent accessibility in mmt rRNA.

The highest N-SASA of all groups, ∼24% without protein, is seen for positions which are low-G in mmt (yellow, green and white) but which vary in bacteria from high- to medium- to low-G, respectively. While SASA is uniformly high in mmt for these positions, there is an increasing trend in SASA for bacteria, from yellow to green to white, consistent with the decrease in G-composition in bacterial alignments for these positions. Significant protection by protein is observed in mmt for this group of positions, especially for yellow positions, almost to the level seen in bacteria in the absence of protein (from 24% to 17.5% N-SASA), consistent with the substitution at many yellow mmt positions of protein-mediated interactions for lost tertiary RNA interactions (see below). The reductions in accessibility induced by protein at the yellow, green, and white positions are the largest of all groups in mmt, but the N-SASA values are still highest of all aligned positions. It is possible that the protein protections at these positions are a by-product of new protein-RNA interactions, which reinforce the RNA structure at locations where RNA–RNA interactions have been weakened by the removal of G.

In conclusion, these data show that as the percentage of G increases in the aligned positions, the solvent accessibility, measured by GWN-SASA, generally decreases for core positions in both alignments but most strongly for mmt. Nonetheless, two distinct trends may be operating with regard to solvent protection of RNA by protein:

1. Protein protection increases for mmt nucleotides at sites that are low-G in mmt but high-G in bacteria, perhaps to reinforce mmt structures that have lost stabilizing, G-specific interactions.
2. Protein protection also increases in mmt at sites with high-G where G is essential but highly exposed, and therefore must be protected from solvent to be retained.

Section Seven: Evolution of RNA modules to maintain 3D structure with fewer Gs

As documented above, mmt rRNAs tend to lose Gs from highly exposed nt positions, many of which occur in loops, and tend to substitute less easily oxidized bases for Gs at these positions. RNA 3D modules, recurrent structural features found in hairpin, internal and multi-helix junction loops, mediate most RNA tertiary interactions, protein binding, and functional interactions (36–39). In addition, 3D modules are responsible for architectural features, such as bends, kinks, and twists in the RNA helix and the specific helix stacking geometries of multi-helix junction loops (40,41). The extensive loss of Gs from positions in 3D modules that are high-G in bacteria is striking and raises several questions. Do the mmt 3D modules retain the same structures as seen in the corresponding bacterial motifs? If so, how do the substituting bases compensate for the loss of stabilizing interactions provided by Gs? In those cases where structural elements change significantly in mmt, are the long-range contacts retained? How do the altered modules mediate these interactions?

Comparisons between mmt and bacterial 3D structures are hampered by the lower resolution of the cryo-EM method, relative to the best available X-ray structures for

bacterial ribosomes (42). Thus, deviations from the bacterial structures, seen in a small number of 3D modules in the mmt structures, for example the loop E modules of h20 and h42, may be attributed to inconsistent modeling due to lack of adequate resolution. Setting such exceptions aside, we see remarkable conservation of the 3D structures of almost all the remaining hairpin, internal and multi-helix junction loops not affected by truncation of helical elements. The only significant exceptions are the internal loop in h28 that forms part of the E-site and the GNRA hairpin loops 27 and 36, which in most mmt sequences have lost the characteristic G that forms the *trans* Sugar/Hoogsteen (tSH) pair that structures canonical GNRA loops. In mmt hairpin loop 36, A637 replaces G and loops out to stack on U471 from the much reduced loop corresponding to h26.1 in bacteria, thereby helping to position the bases A473 and C474 that form the pseudoknot with h19. The interaction of h36 with the h2 pseudoknot is retained in a reduced form (one A-minor interaction instead of two). The conserved G of the h27 GNRA loop in bacteria is replaced by A or C in mmt (A502 in *S. scrofa*). It is also bulged out and interacts with h24 by ribose stacking and H-bonding to the phosphate backbone, instead of the A-minor interactions seen in bacteria. By contrast, the 3D structures of the GNRA loops of h15, h23.1 and h45 and their interactions are unchanged in mmt.

The internal loop in h28 has been described in previous work as 'C-loop-like' (38). Interestingly, in mmt, this loop undergoes sequence and structure changes that transform it into a canonical C-loop with both of the tertiary interactions characteristic of C-loops, i.e. the cis Watson–Crick/Sugar edge (cWS) (also present in bacteria) and the trans Watson–Crick/Hoogsteen (tWH) pairs (see Figure 5A). The sequence conservation supports the new tWH pair between C825 and A537 (100% CA) in mmt rRNA. C-loops are stable 3D modules that are unique in not requiring Gs (38).

Other structural changes involve loss or gain of bulged bases, indicated in the 2D structures by black circles (see Figure 4 and Supplementary Figure S1). Some of these changes are correlated. For example, both bulged G31 (*E. coli*) in h3 and C48 (*E. coli*), the nt in h5 with which G31 interacts in bacteria, are deleted in mmt. Others are not. For example, U49 (*E. coli*) is deleted in mmt, even though the base with which it interacts in bacteria, G362 in *E. coli*, is retained in mmt A181 (*S. scrofa*). G362 is 27% solvent exposed and conserved in bacteria where it forms a G-specific Watson–Crick edge base-phosphate interaction, in addition to the *cis* Sugar/Watson–Crick (cSW) pair with U49. These two tertiary interactions are replaced by the stable, A-specific A181–A184 cWW tertiary pair in mmt.

There are also new bulged bases inserted in mmt. For example, A105 in *S. scrofa* is inserted in h7 and provides a platform for a new RNA-protein interaction in mmt, stacking on A105 of a Trp sidechain.

Isosteric base-pair substitutions that conserve 3D structure. Mmt makes extensive use of isosteric base substitutions in non-Watson–Crick pairs of all types to eliminate exposed Gs, while maintaining the 3D structures of modules. We now provide two examples illustrating how internal loops responsible for architectural features retain their structures despite loss of exposed Gs. In the first and most common way, exemplified by the kink-turn of h11, this is achieved by base substitutions that form isosteric base-pairs. In the second example, the 90° bend in h24, base substitutions create an alternative H-bonding network to the one present in bacteria, with fewer Gs.

The *E. coli* and *S. scrofa* versions of the h11 kink-turn are compared in Figure 5B. The mmt kink-turn has three fewer Gs, yet the same Watson–Crick and non-Watson–Crick pairs form in mmt as in bacteria. G281 (mmt A139) is highly exposed in both structures (32–43%) and forms a *trans* Sugar/Hoogsteen (tSH) pair with A246 (*S. scrofa* A114). The G281/A246 (*E. coli*) and A139/A114 (*S. scrofa*) pairs are isosteric. Half the mmt sequences have G at position 139, while the remainder have A (42%), C (3%), or U (4%). A246 (*S. scrofa* A114) also forms a *trans* Sugar/Sugar (tSS) pair with G278 (*S. scrofa* C136), which is ∼44% exposed in both structures. Protein interactions reduce exposure of C136 to 20% in the *S. scrofa* structure. G278 is 75% G in bacteria but entirely substituted by C (82%), A (10%) or U (8%) in mmt. The AC tSS pair between A and C is isosteric to the pair between A and G but less stable due to loss of the strong GN2-AN3 H-bond. A new Watson–Crick C-edge pseudo-pair is observed between C136 and Arg 198, from the mmt-specific, C-terminal extension of rProtein S15, that stabilizes the A114/C136 interaction. Interestingly, G278 in *T.th.* is also protected by stacking of a Tyr sidechain (of S17), an interaction that is absent in *E. coli*. Finally, G247 in *E. coli* (*S. scrofa* A115 with 35% G and 64% A composition in mmt sequences) with 27% solvent exposure in bacteria forms a tSS base-pair with A282 (*S. scrofa* A140). In this way, mmt retains the kink-turn structure while reducing or eliminating three exposed Gs, at positions 139 (*E. coli* 281), 136 (*E. coli* 278), and 115 (*E. coli* 247).

The compound internal loop in h34 provides a second example. Within internal loop 34, which forms part of the A-site that resides in the Head Domain, G1053 in *E. coli*, is relatively exposed to solvent and forms a base triple with G1057 and C1203 (cWH/cWW). Both Gs are conserved in bacteria, but not in mmt-SSU where the equivalent bases, A613/C617/G688, form an isosteric base triple (cWH/cWW) that maintains the required 3D structure. Consequently, G688 is one of the new brown (low→high) conserved Gs in the mmt-SSU alignment. This isosteric change replaces two conserved Gs with one conserved G and places it in a more protected position in the 3D structure, see Figure 5C.

Use of alternative H-bonding networks. The first internal loop in h24 forms a 90° bend to align the hairpin loop 24 to interact with P-site tRNA. This internal loop also contains exposed Gs in bacteria, but achieves reduction of Gs and 3D structure conservation in mmt by a different mechanism. Green (medium→low) G776 (C416 in *S. scrofa*) is highly exposed in bacteria (36% N-SASA); its N2 forms a G-specific H-bond with O6 of yellow (high→low) G803 to stabilize the 90° bend in bacteria. The high SASA explains

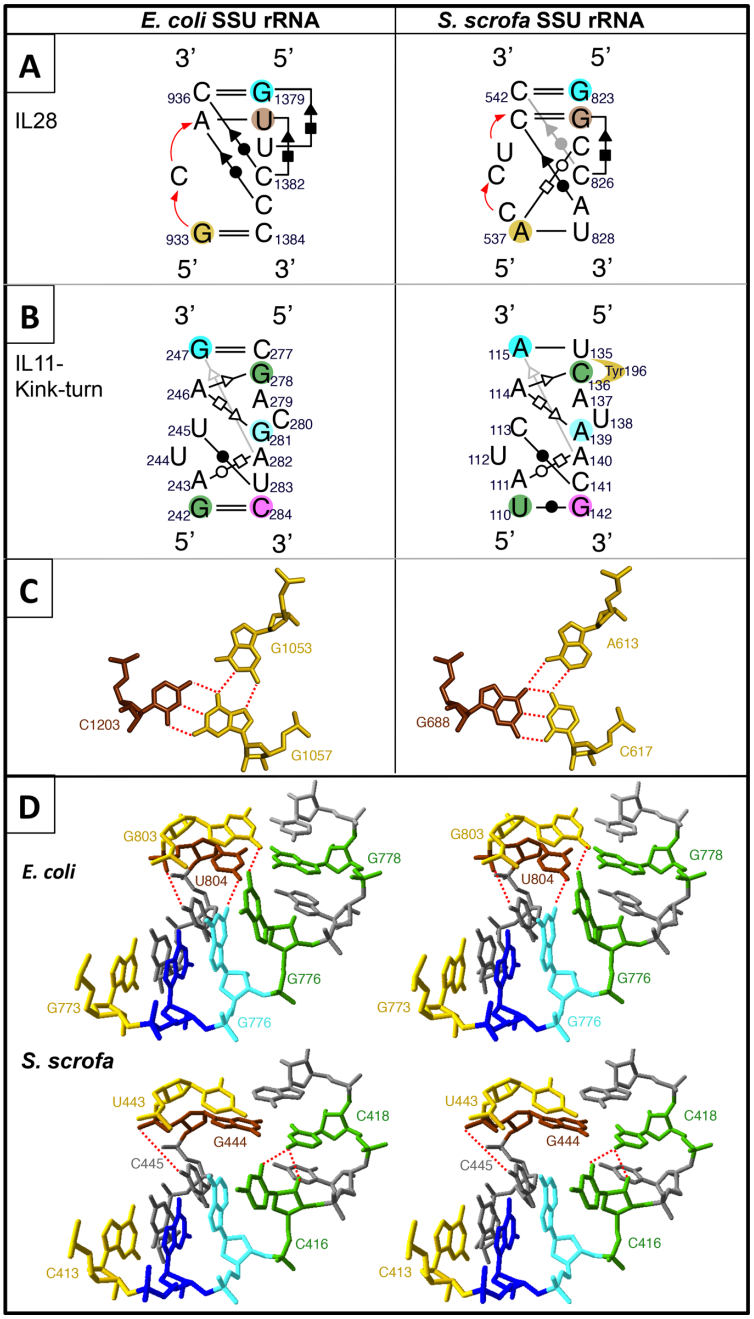

Figure 5. Examples of sequence, structure, and motif changes in *S. scrofa* to maintain 3D structure with fewer Gs. (**A**) An example of motif change in *E. coli* and *S. scrofa.* The 'C-loop-like' internal loop in h28 in bacteria (left) undergoes sequence and structure changes in mmt (right) that transform it into a canonical C-loop with both of the characteristic tertiary interactions, the cWS (also present in bacterial rRNA) and the new tWH base-pair (present only in mmt rRNA). The sequence conservation data support the new tWH pair between C825 and A537 (100% CA). (**B**) An example of isosteric base substitution to replace exposed Gs in bacteria (left) with less easily oxidized bases in mmt (right). The mmt kink-turn has three fewer Gs, yet the same WC and non-WC base-pairs form in mmt as in bacteria. New protein interaction (Tyr196 stacked on C136) in mmt further reduces exposure and stabilizes the tSS basepair formed by C136. (**C**) An example of isosteric base substitutions in the internal loop of h34. G1053 in *E. coli*, which is relatively exposed to the solvent, forms a cWH/cWW base triple with G1057 and C1203 in bacteria. Both Gs are conserved in bacteria but not in mmt-SSU, where the equivalent bases, A613/C617/G688, form an isosteric cWH/cWW base triple that maintains the required 3D structure. Consequently, G688 is one of the new conserved Gs in the mmt-SSU alignment (indicate by the brown color). This isosteric change replaces two conserved Gs with one conserved G and places it in a more protected position in the 3D structure. (**D**) Use of Alternative H-bonding Networks. The first internal loop of h24 forms a 90° bend to align hairpin loop 24 to interact with P-site tRNA. Green G776 (C416 in *S. scrofa*) is highly exposed in bacteria (36% N-SASA); its N2 forms a G-specific H-bond with O6 of yellow G803 to stabilize the 90° bend. The high SASA explains the change from 92% G composition in bacteria to just 5% G in mmt (92% C or U). The loop adopts a different strategy that requires neither G776 nor G803 to stabilize the bend in mmt. In mmt, the WC pair C418/G444 occurs at the position corresponding to G778/U804 in bacteria. G444 in mmt is brown (≥95% G in mmt and <15% G in bacteria), and by WC pairing with position 418 forces this base to be C (*E. coli* G778). This positions the N4 amino group of C418 to H-bond the sugar edge O2 carbonyl of Y416 (*E. coli* 776), stabilizing the bend in the absence of G at position 416 in mmt. All nts in this figure are colored according to the G-conservation categories defined in Figure 3.

the change from 92% G composition in bacteria to just 5% G in mmt (92% C or U). In mmt, the loop adopts a different strategy to stabilize the bend that requires neither G776 nor G803. Further, the Watson–Crick pair C418 = G444 occurs at the position corresponding to cWW G778 U804 in bacteria. G444 in mmt is brown (low→high), and the Watson–Crick pair forces the base at 418 to be a C in mmt (corresponding in *E. coli* G778). This positions the N4 amino group of C418 to H-bond to the sugar edge O2 carbonyl of Y416 (*E. coli* 776), stabilizing the bend in the absence of G at position 416 in mmt. Figure 5D shows how, through subtle changes, an alternative H-bonding network is established in the mmt version of the h24 internal loop to maintain the 3D structure, while eliminating the most exposed Gs.

The blue (high→high) base G774 (*S. scrofa* G414) in h24 is an example of a G that is conserved in both structures because it positions C805 (*S. scrofa* C445) by Watson–Crick pairing so the CN4 amino group can form crucial C-specific ionic interactions with the phosphate of U804 (U443), thus further stabilizing the tight turn in h24. G774 has SASA ∼15% in both *E. coli* and *S. scrofa.* This interaction is similar to one formed by the amino group of C536 (*S. scrofa* C262) with the phosphate of A535 (A261) in the kink-turn of L18. This C is positioned by Watson–Crick pairing to blue G515 (*S. scrofa* G241) to form the Watson–Crick pair flanking the L18 kink-turn. Such tertiary interactions were first seen in tRNA T-loops (43) and are called type II UA-handle motifs (see Figure 5D, (44)). Interestingly, the requirement for G to correctly position C to form these interactions appears sufficient to maintain high-G in mmt at these nt positions.

Interactions substituting for G-specific Base-phosphate interactions. Another strategy that mmt rRNA adopts for reducing the number of Gs, while maintaining important long-range contacts that are mediated in bacteria by G-specific interactions, is to substitute a different type of interaction that does not require G. In fact, we observe that 13 out of 18 G-specific base-phosphate interactions in bacteria that are absent in mmt are replaced by different types of interactions, of which six are RNA–protein interactions and seven are non-Watson–Crick tertiary interactions that do not require G (see Supplementary Table S7). We provide three examples: (i) The base-phosphate interaction of G27 (from the sugar edge) with h12 in *E. coli* 16S is replaced by aromatic stacking of U23 on Trp422 of uS5m, a protein that occupies the vacated region of h12 in mmt 12S rRNA. (ii) The base-phosphate interaction (from the Watson–Crick edge) formed in the h20 loop E motif by *E. coli* G581, which is also forming a *trans* Sugar/Hoogsteen (tSH) pair with G760, is replaced in mmt by a trans Watson–Crick/Hoogsteen (tWH) pair between U302 and A399, the bases that substitute for G581 and G760. (iii) The base-phosphate in bacteria formed by G725 in h23.1 with the phosphate of the intercalating base G666 is replaced in mmt by an ionic Hoogsteen-edge pseudo-pair between the intercalating base, A375 and Arg17 of the new mmt protein, mS37.

Section Eight: Sites with G-favored or G-specific RNA tertiary and RNA-Protein interactions

G-stabilized interactions (including G-favored and G-specific interactions) that are formed by core aligned nts in the *E. coli* and *S. scrofa* structures are tabulated in Table 4, organized by percentage G composition in the two alignments. G-favored interactions are most stable with G but do not require G. G-specific interactions require G.

G-favored interactions include most non-Watson–Crick pairs in the *S. scrofa* structure. We have compiled all the non-Watson–Crick pairs formed by core nucleotides in the *E. coli* and *S. scrofa* SSU structures by base-pair family (34) in Supplementary Table S8. These data clearly show the overwhelming preference, among the 12 families, for 5 types of non-Watson–Crick pairs involving Gs, which make up ∼82% of non-Watson–Crick pairs formed by nucleotide positions with medium to high percentage G in each of the structures. These five types of non-Watson–Crick pairs are those that account for over half of non-Watson–Crick pairs in structured RNAs (45).

G-specific interactions (see legend to Supplementary Table S6) include the WC-edge base-phosphate, the GU 'Packing' or 'P-interactions' and the G-ribose interactions (31,32), and some protein interactions.

G-stabilized interactions predict the G composition in the mmt alignment. In this section we evaluate the degree to which the percentage of G in the mmt alignment is predicted by the types and numbers of G-specific or G-favored RNA–RNA and RNA–protein interactions observed in the *E. coli* and *S. scrofa* structures (see Supplementary Material for detailed descriptions). We expect that for those aligned nucleotide positions where the same G-favored interaction occurs in both 3D structures, higher percentages of G will be observed in the mmt alignment than for those positions where the interaction is not observed in the mmt (*S. scrofa*) 3D structure. Furthermore, we expect to see an even higher percentage of G in the mmt alignments where the same G-specific interaction is present in both the *E. coli* and *S. scrofa* 3D structures (and lower percentage of G when it is absent in the *S. scrofa* structure), than for G-favored interactions that can also form without G, including many non-Watson–Crick pairs and RNA–protein interactions.

In order to test this hypothesis, data from SSU structures and sequences, that include all aligned positions that have high-G in the bacterial alignment (blue, cyan, and yellow positions) and form one or more G-stabilized interaction in the *E. coli* structure are presented in Table 4.

Table 4 shows that for nucleotide positions where the same interaction occurs in the *E. coli and S. scrofa* structures, substantially higher average percentage of G is observed in the mmt alignment compared to positions where the interaction does not occur. The largest difference is observed for G-specific base-phosphate (89% versus 6%) and G-ribose or GU packing interactions (80% versus 6%). Almost all of these positions are blue (high→high) or cyan (high→medium) in the first set of columns of Table 4 and yellow (high→low) in the second set of columns. A smaller but significant difference is observed for G-favored (but not G-specific) stacking (69% versus 5%), non-Watson–

Table 4. The %G in mmt alignments is predicted by the presence of G-specific (first two rows) and G-favored (last two rows) interactions, some of which are in bacterial and mmt 3D structures

	Interactions present in *E. coli* and *S. scrofa* 3D structures						Interactions only present in in *E. coli* 3D structure					
	Number of positions forming the interaction	Number of nucleotides by color			Average %G		Number of positions forming the interaction	Number of nucleotides by color			Average %G	
Type of interaction	(Number of positions forming additional interactions in *E.c./S.sc.*)	Blue (High/ High)	Cyan (High/ Interm)	Yellow (High/ low)	Bact	mmt	(Number of positions forming additional interactions in *E.c./S.sc.*)	Blue (High/ High)	Cyan (High/ Interm)	Yellow (High/ low)	Bact	mmt
Base-phosphate	21 (16/16)	15	6	0	99.6	89.0	18 (12/13)	0	3	15	99.2	6.2
G-Ribose or GU	17 (6/5)	10	5	2	99.5	80.0	8 (2/1)	0	1	7	98.2	6.1
Packing Interactions 3° and Multi-Helix Junction Stacking and Intercalation	38 (25/23)	22	9	7	99.0	69.0	1 (0/0)	0	0	1	99.1	5.1
Non-WC Pair or 3°WC pair	71 (36/30)	40	14	17	99.3	66.4%	19 (9/10)	2	4	13	99.3	20.3
RNA–protein interaction	22 (16/14)	14	0	8	99.2	63.8	14 (9/9)	2	4	8	99.5	32.5
Total number of interactions	**169**	101	34	34			**0**	4	12	44		
Total number of nucleotides	**126**	68	27	31			**50**	4	11	35		

All selected positions have high G composition in bacteria. Rows correspond to different types of G-stabilized interactions; the first two rows are G-specific, and the rest are G-favored interactions. The first set of columns cover cases where the interaction is present in both bacteria and mmt; the second set of columns covers cases where the interaction is lost in mmt. The first set of columns includes all positions for which the corresponding bases in the *E. coli* and *S. scrofa* 3D structures form the same G-stabilized interaction and the second set of columns, those for which only the *E. coli* base does so. The number of nt positions is reported along with the average %G of the respective columns from the two alignments. We note that many nts with high-G make more than one G-stabilized interaction and are included in more than one row. The number of positions that form additional interactions in each structure is shown in parentheses after the nt count. There are also positions where the *S. scrofa* structure lacks G that may be found in other mmt sequences. If the interaction is nonetheless observed in *S. scrofa*, it is included in the first set of columns; if not, it goes to the second set of columns.

Crick pairs (66% versus 20%) and RNA–protein interactions (64% versus 32%). For these interactions, the percentage of G in the mmt alignment is lower in the first set of columns than for the G-specific ones, because many of these positions are cyan or yellow, indicating that for many mmt sequences another base is able to replace the G present in the *E. coli* structure, and still make a sterically acceptable, though generally weaker, interaction. For example, G405 and G951 in *S. scrofa* are replaced by A in *Homo sapiens* (*H.s.*) and in 42% of mmt sequences. In both structures, these bases form cSS pairs with As, but the isosteric GA pair is more stable than the AA cSS pair.

For some nucleotide positions with medium-G in the mmt alignment that make G-specific interactions in bacteria, *S. scrofa* lacks a G that is found in some other mmt sequences. The aligned structures of human and porcine mmt SSU each have Gs at 19 out of 40 cyan (high→medium) positions, very close to the average percentage of G for cyan positions in the entire alignment (50%), see rows for cyan nts in Supplementary Table S4. At these positions, sometimes the human and sometimes the *S. scrofa* structure has a G that preserves the interaction as in bacteria. For example, a G-specific GU packing interaction in the *H.s.* 3D structure does not form in the *S. scrofa* structure because the G in *H.s.* becomes A274 *S. scrofa.* Such positions tend to increase the percentage G in the second set of columns of Table 4. In fact, we observe for cyan positions in *S. scrofa* that Gs are preferentially replaced by As (16 As versus 3 Us and 2 Cs, see Supplementary Table S4). However, even though both species are placental mammals, they have only 10 out of 19 cyan Gs in common, consistent with the high mutation rate in mitochondria, and with the idea that the total number of Gs is constrained in mmt rRNA. Because many blue (high→low) and cyan (high→medium) bases make more than one interaction, loss of an interaction in mmt may not change the percentage of G significantly for those positions where one or more remaining interactions require G. These blue and cyan positions also tend to increase the percentage of G in the second set of columns.

Loss of G due to loss of interacting RNA elements. For G-specific interactions, there are no blue (high→low) and few cyan (high→medium) positions in the second set of columns of Table 4 and the percentage of G in the mmt alignments is low. For most of the positions that have lost G-specific interactions in mmt rRNA, the interacting element has been lost or the interaction has been replaced by an interaction not requiring G, as discussed in section 7. An example is G1108 in the h34/35/38 3-Way-Junction, which is high-G in bacteria (it makes a strong G-specific base-phosphate interaction with U1095 in the hairpin loop of h37 so as to position that element to interact with h40 and to maintain the h36-h35 stacking at the h35/36/37 junction). In mmt, h37 is much reduced, eliminating U1095, so the exposed G1108 in bacteria can be replaced by a less easily oxidized base in mmt. The places where RNA tertiary interactions are observed in bacterial SSU rRNA that are lost in mmt SSU rRNA, because of deletion of an interacting element, are indicated by red circles in Figure 4. There are 31 such positions, of which 20 occur at yellow (high→low), 2 at cyan (high→medium), 5 at green (medium→low), and 4 at pink (medium→medium) positions. Many of these positions form Watson–Crick pairs with tertiary interactions in the minor groove; they attest to the important role of the GN2 amino group, unique to G, in mediating tertiary interactions on the sugar edge ('G-ribo' interactions) (46).

Summary of factors affecting percentage G in core nt positions in the mmt alignments. All the interactions formed by the core aligned nts of *E. coli* and *S. scrofa* structures were compiled in a table with rows marked according to the eight colored nt conservation sets, and are provided in full as Supplemental Data (Supplementary Table S4). These data show that for almost every blue (high→low) position, there is some type of G-stabilizing interaction and this type is identical in the two structures.

The data in Supplementary Table S4 are summarized in Figure 6, where the fractions of aligned nucleotides from each colored set that form different types of G-specific and G-favored interactions are given. Figure 6 also lists nearest neighbor thermodynamic effects (column 3), summary statistics for average number of long-range RNA interactions (column 9) and G-stabilized RNA–RNA interactions per nt (column 10). Colors are used to highlight the differences of the values between the *E. coli* and *S. scrofa* structures (see the table legend). More protein interactions are observed in mmt for almost all categories, consistent with the larger amount of protein (see section 9).

Figure 6 shows that the blue (high→high) nucleotides stand out in almost every category, with an average number of interactions close to equal for *E. coli* (1.49) and *S. scrofa* (1.53). Such nucleotides have an average of 1.5 interactions/base, not counting helical Watson–Crick pairs, much higher than for the other sets of aligned core nucleotides. The positions with higher percentage G in bacteria than in mmt consistently have more interactions in bacteria, as can be seen by comparing the cyan, yellow, and green rows, where light and dark blue shading dominate. The average number of interactions for high-G positions in bacteria is 1.49 for blue (high→high), 1.20 for cyan (high→medium) and 1.28 for yellow (high→low). For mmt, the corresponding numbers are 1.53 for blue (high→high), but 1.05 for cyan (high→medium), and only 0.78 for yellow (high→low) positions.

By contrast, the red (medium→high) and brown (low→high) positions have ∼two times more average numbers of interactions in mmt than in bacteria. Red nts have the largest fraction of base-phosphate interactions of all colored groups in mmt. All brown positions but one form Watson–Crick pairs. The only brown loop nucleotide, G807 (*S. scrofa*) in h43, provides a unique example in which a key G-specific interaction is maintained by 'moving' the G to a nearby, more protected position. The flexible element h31, which undergoes conformational changes during translocation, must maintain contact with tRNA and multiple elements in the SSU head, including hairpin loop 43. In *E. coli,* G976 from h31 (92% G in bacteria) intercalates in hairpin loop 43 and makes a Watson–Crick edge base-phosphate interaction with the phosphate of A1363 (95% A), which corresponds to G807 (*S. scrofa*). However, G976 is looped out and highly exposed. The corresponding position in mmt, A581 (0% G), also intercalates, but cannot

					G-stabilized Interactions										Summary of Interactions			
1	2		3		4		5		6		7		8		9		10	
Color code	G composition		Thermo-dynamic Enhancement in bacteria or mmt		Base-Phosphate		G-Ribose or GU Packing		Multi- Helix Junction and 3° Stacking		Non-WC Pair or 3° WC pair		Base-Protein contact		Long-Range tertiary		Average Number of interactions	
	Bact	mmt	5'-end of helix	Tandem GG or GC	*E.c.*	*S.sc.*	*E.c.*	*S.sc.*	*E.c.*	*S.sc.*	*E.c.*	*S.sc.*	*E.c.*	*S.sc.*	*E.c.*	*S.sc.*	*E.c.*	*S.sc.*
Blue	≥95%	≥95%	**33.8%**	43.2%	20%	20%	14%	14%	30%	30%	57%	54%	23%	30%	62%	62%	1.49	1.53
Cyan		≥15% <95%	27.5%	50.0%	23%	15%	15%	18%	23%	23%	45%	35%	10%	10%	43%	38%	1.20	1.05
Yellow		<15%	15.4%	38.5%	23%	2%	14%	3%	12%	12%	46%	28%	26%	31%	49%	20%	1.28	0.78
Red	≥15% <95%	≥95%	25.0%	60.0%	20%	25%	0%	0%	15%	20%	15%	20%	5%	25%	20%	25%	0.55	1.00
Pink		≥15% <95%	28.6%	67.9%	7%	0%	0%	0%	14%	14%	21%	11%	7%	18%	14%	14%	0.50	0.43
Green		<15%	-	-	6%	5%	4%	0%	6%	4%	25%	21%	7%	11%	20%	12%	0.49	0.41
Brown	<15%	≥95%	22.7%	**77.3%**	0%	5%	0%	18%	5%	14%	23%	27%	9%	5%	23%	23%	0.41	0.73
Orange		≥15% <95%	8.3%	54.2%	4%	4%	4%	4%	0%	4%	33%	17%	21%	17%	21%	17%	0.63	0.46

Figure 6. Summary of structural factors observed in 3D structures of *E. coli* and *S. scrofa* SSU rRNA that favor conservation of G at specific nt positions, organized by correlated G-composition in the bacterial and mmt alignments. Column 3: G-specific thermodynamic stabilization; columns 4–7: G-stabilized RNA–RNA interactions; column 8: RNA-Protein interactions; column 9: all long-range interactions; column 10: average number of G-stabilized RNA–RNA interactions per nt, excluding Watson–Crick basepairs. The numbers in columns 4–8 are the fraction of the nts forming each type of G-stabilized interaction in each set. Five colors are used to highlight the differences between the *E. coli* and *S. scrofa* structures. The cell background is colored in white when the numbers for *E. coli* and *S. scrofa* are within the range of ±10% of each other; cells are colored in shades of blue when the numbers for *E. coli* are greater and shades of red when numbers for *S. scrofa* are greater. In more detail, dark blue: > ±20% in *E. coli*; light blue: ±10–20% larger in *E. coli*; dark red: > ±20% in *S. scrofa*; and light red: ±10–20% larger in *S. scrofa*.

form the strong base-phosphate interaction with hairpin loop h43 and h31. However, G807 (94% G in mmt) in h43 is able to form a Watson–Crick edge base-phosphate with U580 (*E. coli* A975) in h31, thus maintaining a strong G-specific contact between hairpin loops 43 and h31, by using the more protected position in h43 to place the G, rather than the more exposed position in h31.

Section Nine: rProteins in mmt SSU reduce rRNA solvent accessibility and reinforce 3D structure

The mmt SSU structure contains significantly more protein than the bacterial one. Consequently, 1.7 times more rRNA surface area is protected from solvent in the *S. scrofa* versus *E. coli* SSU (9831 Å versus 5694 Å); overall, more than double the fraction of rRNA surface area is protected (24% *in S. scrofa* versus 11% in *E. coli,* see Supplementary Table S9).

Analyzing the residue-level RNA–protein interactions in the *E. coli* (14) and *S. scrofa* (11) SSU we find nearly 1.7 times more in mmt (162 versus 93 in bacteria, see Supplementary Table S10 and Table 5) (22). Both the mmt and bacterial structures contain a variety of aa-base interactions of all types, including pseudo-pairs, which involve two or more H-bonds between aa side chains or backbone atoms and RNA base edges (47,48), parallel stacking of planar aromatic or polar aa side chains on RNA bases, 'bidentate' interactions, in which individual aa H-bond with two stacked RNA bases, hydrophobic minor groove interactions, perpendicular stacking of aromatic aa, as observed in crystal structures of benzene (49), and cation-pi interactions (50). A summary by interaction type appears in Supplementary Table S10, which shows that while there are comparable numbers of aa-RNA base edge interactions in the two structures, there are far more aa-base stacking interactions in the mmt structure. The many non-specific ionic interactions with the RNA backbone, which protect the RNA backbone from solvent, are not included in Supplementary Tables S10 and Table 5.

The observed protein interactions involving nucleobases in mmt and bacterial rRNA are mapped onto their respective 2D diagrams in Figure 7. The different types of RNA–protein interactions show distinct preferences according to the RNA 2D structure context. For example, most stacking interactions involve loop nucleotides (58 out of 65 stacking interactions in mmt and 11 out of 17 in bacteria) while bidentate interactions and pseudo-pairs are mainly formed with helical nucleotides.

RNA–protein stacking interactions in mmt-SSU appear to compensate in part for lost RNA–RNA tertiary interactions, many of which involve hairpin loops. Because truncation of helices exposes their ends to solvent, stacking by aa chains protects the exposed bases from solvent. This is especially common with planar aromatic (Phe, Tyr, Trp), cationic (Arg) and polar (Asn, Gln) side chains belonging to mmt-specific extensions of existing rProteins or entirely new proteins. There are just 17 such stacking interactions with bases in the *E. coli* 16S 3D structure compared to 65 in each of the two available mmt-SSU structures (see Supplementary Table S10 and Table 5). Only three of these stacking interactions are observed in both bacterial and mmt structures (C64, A368, and G791 with *S. scrofa* numbering); most of the stacking interactions in mmt involve proteins unique to mmt-SSU. It is noteworthy that mmt rProteins contain substantially more Tyr and Trp residues than bacterial rProteins, both in absolute numbers and as a percentage of all amino acids.

There are 85 high→medium (cyan) and high→low (yellow) positions that are replaced by another base in the *S. scrofa* structure, of which 14 interact with rProteins. Here, the amino acids appear to protect or bolster the bases with additional protein-RNA interactions (see corresponding rows in Supplementary Table S4). An example is G588 in *E. coli,* which forms a tHS pair with A26. The equivalent position in *S. scrofa* is C284, which stacks on His44 (uS12), thus replacing the G-specific RNA–RNA interaction with an RNA–protein interaction that does not re-

Table 5. RNA–protein interactions in bacterial and porcine mitochondrial SSU ribosome, showing the distribution of different types of specific interactions between RNA bases and amino acid sidechains and peptide backbones. All electrostatic interactions between phosphate backbone of rRNA and positively charged sidechains of amino acids from rProteins have been excluded from this analysis

	Counts of interactions					
	E. coli			*S. scrofa*		
Type of RNA–protein interaction	Edges					
Base-edge Interactions	WC	Sugar	H	WC	Sugar	H
Single H-bonds	0	8	2	6	8	1
Pseudopairs	6	15	16	8	11	18
Bidentate H-bonds	2	4	10	6	7	16
Van der Waals	0	8	5	3	9	4
Sub-total	8	35	33	23	35	39
Base-Stacking Interactions	*E. coli*			*S. scrofa*		
Aromatic	3			24		
Cationic	9			22		
Alkyl	3			11		
Hydrophilic	1			4		
Anionic	0			2		
Perpendicular	1			2		
Sub-total	**17**			**65**		
Total	93			162		

quire G. Another example is G111 in *E. coli* that forms a Watson–Crick/Phosphate ionic interaction with A60. The corresponding nt in *S. scrofa* is C67 and the base-phosphate interaction is replaced by WC pseudo-pair with backbone atoms of Arg27 (mS26).

In summary, mmt has many more RNA–protein interactions, but especially aromatic and ionic stacking with nucleobases, many of which serve to protect exposed residues in truncated helices or reinforce RNA interactions weakened by loss of G. Most of these interactions involve rProteins that are found only in mmt or homologs of bacterial rProteins that have acquired new domains or extensions in mmt.

DISCUSSION

Although Gs thermally stabilize RNA structure, they are also the least stable chemically, being most easily oxidized. The mitochondria of highly active organisms, including mammals, produce large amounts of ROS, creating a challenging environment for ribosomes to function in.

Although we analyzed interactions, contacts and accessibilities for only four experimental structures (*E. coli*, *T.th.* and *S. scrofa* and *H.s.* SSU), these structural data are backed by extensive alignments of bacterial (1228) and mitochondrial (899) sequences covering a large part of phylogeny. We also made a quick survey of the mmt large ribosomal subunit (LSU). The PDB file 4V19 of the porcine mt LSU rRNA 3D structure was analyzed to obtain the composition of LSU rRNA. Mt LSU rRNA has a total of 1570 nts of which 611 are As, 370 Us, 329 Cs and 260 Gs. Overall, the LSU rRNA is truncated even more than that of SSU rRNA with only 1570/2904 = 54% of the bacterial rRNA remaining. As Supplementary Table S11A shows, there is even less G in porcine LSU rRNA than in SSU rRNA. Thus, we see that similar trends can be observed in LSU rRNA. By contrast, the SSU and LSU rRNAs of plant chloroplasts and mitochondria have base composition and size very similar to those of bacteria (Supplementary Tables S11B and S11C).

Thus, overall, the data presented are consistent with the overarching idea that the size reduction in mmt rRNA and the significant loss of G nucleotides result from selected evolution to the highly oxidizing environment present in the matrix of mammalian mitochondria. In the following we list the multiple lines of evidence that support this conclusion:

(i) Gs are primarily lost from highly exposed positions in mmt rRNA.
(ii) Peripheral RNA helices that are difficult to protect from oxidation damage and are not essential for binding substrates or positioning RNA primary elements have been eliminated.
(iii) Gs are retained primarily within or near to functional sites, especially where needed for direct interaction with substrates.
(iv) Gs are largely eliminated from all positions where they are not needed to form a G-specific interaction or to provide thermodynamic stabilization.
(v) Many new rProtein-rRNA interactions significantly reduce the solvent accessibility of the rRNA as a whole and Gs in particular. Such interactions involve rProteins or rProtein extensions found only in mmt.

These changes in turn pose new challenges for the RNA to fold into and maintain the correct 3D structures in order to interact with substrates and translational factors. On the basis of our analysis of the 3D structures and sequence alignments the following 'operational rules' help explain the folding of a complex RNA, not only with fewer Gs, but also to shield the remaining Gs from reactive species prevalent in the surrounding environment:

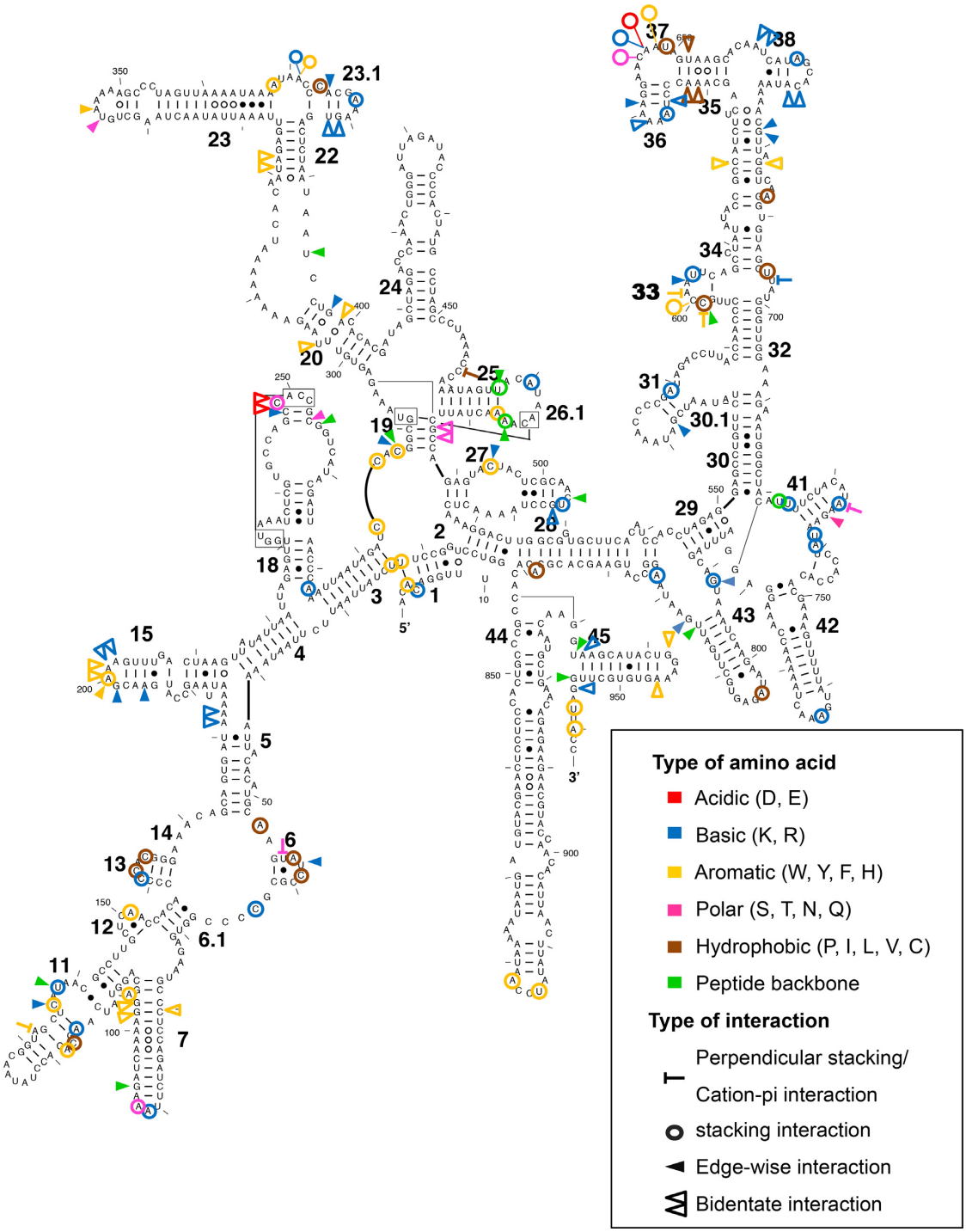

Figure 7. RNA–protein interactions of different types involving RNA bases in the porcine mitochondrial ribosomal SSU. Different modes of interactions are denoted by different symbols as annotated in the legend inset. Interacting amino acids are categorized by the nature of their sidechain.

(i) Gs occur in a smaller number of selected positions, where only G can make the required interaction(s), especially where one G can make multiple stabilizing interactions.
(ii) Gs are replaced whenever another base can mediate the same interaction with nearly the same stabilizing energy, for example A for base-base or base-amino acid stacking interactions.
(iii) Gs are maintained at flanking positions, preferentially at the 5′-ends, rather than in loops or within helices. GU wobble pairs are converted to GC pairs and clusters of GC pairs are consistent with the Turner stability parameters.
(iv) Base-pairs involving G are replaced by isosteric combinations that maintain RNA–RNA interactions.

(v) The weakened RNA–RNA interactions are reinforced by a diverse set of RNA–protein or protein–protein interactions.
(vi) Modules are strengthened by moving Gs to key strategic positions with additional non-Watson–Crick base-pairs.
(vii) Alternative H-bonded networks that minimize the number of Gs are built.

Mmt ribosomes constitute a beautiful illustration of the transition from an RNA-centric world to an RNA–protein world (51), since they are composed of nearly twice as much ribosomal protein (rProtein) as their bacterial counterparts. Here, we could delineate structural elements and interactions that play key roles in the stabilization and folding of the functional RNA elements despite the RNA loss with nucleotide stacking being replaced by amino acid side chain or protein domains replacing entire RNA domains, as in the case of Cyt18 in the group I intron family (52). Under the functional pressure to reduce the number of guanine nucleotides, the ones most susceptible to oxidative damage, mitochondrial rRNA evolved in sequence by exploiting the whole range of molecular interactions accessible to all RNA nucleotides as well as together with amino acid side chains of mmt-specific rProtein elements.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors warmly thank Jamie Cannone and Robin Gutell for sharing and assisting with rRNA sequence alignments. The authors also thank the reviewers for careful and patient reading of the manuscript and constructive suggestions.
Author Contributions: M.H. and P.R. performed most of the calculations and computational analysis. C.L.Z., N.L., E.W. supervised most of the computational work. M.S., E.W. and N.L. designed the project. All authors contributed significantly to the writing of the manuscript.

FUNDING

French National Program Investissement d'Avenir (Labex MitoCross and Labex NetRNA) administered by the Agence National de la Recherche [ANR-11-LABX-0057_MITOCROSS to M.S. and ANR-10-LABX-0036_NETRNA to E.W.]; National Institutes of Health [2R01GM085328-05 to N.B.L. and C.L.Z.]. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Funding for open access charge: National Institutes of Health [2R01GM085328-05].
Conflict of interest statement. E.W. is Executive Editor of *NAR*.

REFERENCES

1. Christian,B.E. and Spremulli,L.L. (2012) Mechanism of protein biosynthesis in mammalian mitochondria. *Biochim. Biophys. Acta*, **1819**, 1035–1054.
2. Zorov,D.B., Juhaszova,M. and Sollott,S.J. (2014) Mitochondrial reactive oxygen species (ROS) and ROS-induced ROS release. *Physiol. Rev.*, **94**, 909–950.
3. Murphy,M.P. (2009) How mitochondria produce reactive oxygen species. *Biochem. J.*, **417**, 1–13.
4. Crease,T.J., Lynch,M. and Spitze,K. (1990) Hierarchical analysis of population genetic variation in mitochondrial and nuclear genes of Daphnia pulex. *Mol. Biol. Evol.*, **7**, 444–458.
5. Gelfand,R. and Attardi,G. (1981) Synthesis and turnover of mitochondrial ribonucleic acid in HeLa cells: the mature ribosomal and messenger ribonucleic acid species are metabolically unstable. *Mol. Cell. Biol.*, **1**, 497–511.
6. Retz,K.C. and Steele,W.J. (1980) Ribosome turnover in rat brain and liver. *Life Sci.*, **27**, 2601–2604.
7. Anderson,S., Bankier,A.T., Barrell,B.G., de Bruijn,M.H., Coulson,A.R., Drouin,J., Eperon,I.C., Nierlich,D.P., Roe,B.A., Sanger,F. *et al.* (1981) Sequence and organization of the human mitochondrial genome. *Nature*, **290**, 457–465.
8. Lewis,F.D., Letsinger,R.L. and Wasielewski,M.R. (2001) Dynamics of photoinduced charge transfer and hole transport in synthetic DNA hairpins. *Acc. Chem. Res.*, **34**, 159–170.
9. Alenko,A., Fleming,A.M. and Burrows,C.J. (2017) Reverse transcription past products of guanine oxidation in RNA leads to insertion of A and C opposite 8-Oxo-7,8-dihydroguanine and A and G opposite 5-Guanidinohydantoin and spiroiminodihydantoin diastereomers. *Biochemistry*, **56**, 5053–5064.
10. Kong,Q. and Lin,C.L. (2010) Oxidative damage to RNA: mechanisms, consequences, and diseases. *Cell. Mol. Life Sci.*, **67**, 1817–1829.
11. Greber,B.J., Bieri,P., Leibundgut,M., Leitner,A., Aebersold,R., Boehringer,D. and Ban,N. (2015) Ribosome. The complete structure of the 55S mammalian mitochondrial ribosome. *Science*, **348**, 303–308.
12. Amunts,A., Brown,A., Toots,J., Scheres,S.H.W. and Ramakrishnan,V. (2015) Ribosome. The structure of the human mitochondrial ribosome. *Science*, **348**, 95–98.
13. Greber,B.J. and Ban,N. (2016) Structure and function of the mitochondrial ribosome. *Annu. Rev. Biochem.*, **85**, 103–132.
14. Noeske,J., Wasserman,M.R., Terry,D.S., Altman,R.B., Blanchard,S.C. and Cate,J.H. (2015) High-resolution structure of the Escherichia coli ribosome. *Nat. Struct. Mol. Biol.*, **22**, 336–341.
15. Polikanov,Y.S., Melnikov,S.V., Söll,D. and Steitz,T.A. (2015) Structural insights into the role of rRNA modifications in protein synthesis and ribosome assembly. *Nat. Struct. Mol. Biol.*, **22**, 342.
16. Coimbatore Narayanan,B., Westbrook,J., Ghosh,S, Petrov,A.I., Sweeney,B., Zirbel,C.L., Leontis,N.B. and Berman,H.M. (2014) The nucleic acid database: new features and capabilities. *Nucleic Acids Res.*, **42**, D114–D122.
17. Cannone,J.J., Sweeney,B.A., Petrov,A.I., Gutell,R.R., Zirbel,C.L. and Leontis,N. (2015) R3D-2-MSA: the RNA 3D structure-to-multiple sequence alignment server. *Nucleic Acids Res.*, **43**, W15–W23.
18. Guex,N. and Peitsch,M.C. (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis*, **18**, 2714–2723.
19. Sweeney,B.A., Roy,P. and Leontis,N.B. (2015) An introduction to recurrent nucleotide interactions in RNA. *Wiley Interdiscip. Rev. RNA*, **6**, 17–45.
20. Gerstein,M. (1992) A resolution-sensitive procedure for comparing protein surfaces and its application to the comparison of antigen-combining sites. *Acta Crystallogr. A: Found. Crystallogr.*, **48**, 271–276.
21. Sarver,M., Zirbel,C.L., Stombaugh,J., Mokdad,A. and Leontis,N.B. (2008) FR3D: finding local and composite recurrent structural motifs in RNA 3D structures. *J. Math. Biol.*, **56**, 215–252.
22. Roy,P. (2017) *Analyzing and classifying bimolecular interactions: I. Effects of metal binding on an iron-sulfur cluster scaffold protein II.* Automatic Annotation of RNA–Protein Interactions for NDB, Bowling Green State University.
23. Gardner,D.P., Xu,W., Miranker,D.P., Ozer,S., Cannone,J.J. and Gutell,R.R. (2012) An accurate scalable template-based alignment algorithm. *Proc. (IEEE Int. Conf. Bioinformatics Biomed.)*, **2012**, 1–7.

24. Shang,L., Gardner,D.P., Xu,W., Cannone,J.J., Miranker,D.P., Ozer,S. and Gutell,R.R. (2013) Two accurate sequence, structure, and phylogenetic template-based RNA alignment systems. *BMC Syst. Biol.*, **7**(Suppl. 4), S13.
25. Cannone,J.J., Sweeney,B.A., Petrov,A.I., Gutell,R.R., Zirbel,C.L. and Leontis,N. (2015) R3D-2-MSA: the RNA 3D structure-to-multiple sequence alignment server. *Nucleic Acids Res.*, **43**, W15–W23.
26. Zirbel,C.L., Šponer,J.E., Šponer,J., Stombaugh,J. and Leontis,N.B. (2009) Classification and energetics of the base-phosphate interactions in RNA. *Nucleic Acids Res.*, **37**, 4898–4918.
27. Saenger,W. (1973) Structure and function of nucleosides and nucleotides. *Angew. Chem. Int. Ed. Engl.*, **12**, 591–601.
28. Jegousse,C., Yang,Y., Zhan,J., Wang,J. and Zhou,Y. (2017) Structural signatures of thermal adaptation of bacterial ribosomal RNA, transfer RNA, and messenger RNA. *PLoS One*, **12**, e0184722.
29. Xia,T., SantaLucia,J., Burkard,M.E., Kierzek,R., Schroeder,S.J., Jiao,X., Cox,C. and Turner,D.H. (1998) Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson–Crick base pairs. *Biochemistry*, **37**, 14719–14735.
30. Varani,G. and McClain,W.H. (2000) The G x U wobble base pair. A fundamental building block of RNA structure crucial to RNA function in diverse biological systems. *EMBO Rep.*, **1**, 18–23.
31. Mokdad,A., Krasovska,M.V., Sponer,J. and Leontis,N.B. (2006) Structural and evolutionary classification of G/U wobble basepairs in the ribosome. *Nucleic Acids Res.*, **34**, 1326–1341.
32. Gagnon,M.G. and Steinberg,S.V. (2002) GU receptors of double helices mediate tRNA movement in the ribosome. *RNA*, **8**, 873–877.
33. Chen,J.L., Dishler,A.L., Kennedy,S.D., Yildirim,I., Liu,B., Turner,D.H. and Serra,M.J. (2012) Testing the nearest neighbor model for canonical RNA base pairs: revision of GU parameters. *Biochemistry*, **51**, 3508–3522.
34. Leontis,N.B. and Westhof,E. (2001) Geometric nomenclature and classification of RNA base pairs. *RNA*, **7**, 499–512.
35. Abu Almakarem,A.S., Petrov,A.I., Stombaugh,J., Zirbel,C.L. and Leontis,N.B. (2012) Comprehensive survey and geometric classification of base triples in RNA structures. *Nucleic Acids Res.*, **40**, 1407–1423.
36. Leontis,N.B. and Westhof,E. (2003) Analysis of RNA motifs. *Curr. Opin. Struct. Biol.*, **13**, 300–308.
37. Leontis,N.B., Lescoute,A. and Westhof,E. (2006) The building blocks and motifs of RNA architecture. *Curr. Opin. Struct. Biol.*, **16**, 279–287.
38. Lescoute,A., Leontis,N.B., Massire,C. and Westhof,E. (2005) Recurrent structural RNA motifs, Isostericity Matrices and sequence alignments. *Nucleic Acids Res.*, **33**, 2395–2409.
39. Parlea,L.G., Sweeney,B.A., Hosseini-Asanjan,M., Zirbel,C.L. and Leontis,N.B. (2016) The RNA 3D motif atlas: computational methods for extraction, organization and evaluation of RNA motifs. *Methods*, **103**, 99–119.
40. Lescoute,A. and Westhof,E. (2006) The interaction networks of structured RNAs. *Nucleic Acids Res.*, **34**, 6587–6604.
41. Lescoute,A. and Westhof,E. (2006) Topology of three-way junctions in folded RNAs. *RNA*, **12**, 83–93.
42. Vénien-Bryan,C., Li,Z., Vuillard,L. and Boutin,J.A. (2017) Cryo-electron microscopy and X-ray crystallography: complementary approaches to structural biology and drug discovery. *Acta Crystallogr. F Struct. Biol. Commun.*, **73**, 174–183.
43. Krasilnikov,A.S. and Mondragón,A. (2003) On the occurrence of the T-loop RNA folding motif in large RNA molecules. *RNA*, **9**, 640–643.
44. Jaeger,L., Verzemnieks,E.J. and Geary,C. (2009) The UA_handle: a versatile submotif in stable RNA architectures. *Nucleic Acids Res.*, **37**, 215–230.
45. Stombaugh,J., Zirbel,C.L., Westhof,E. and Leontis,N.B. (2009) Frequency and isostericity of RNA base pairs. *Nucleic Acids Res.*, **37**, 2294–2312.
46. Ulyanov,N.B. and James,T.L. (2010) RNA structural motifs that entail hydrogen bonds involving sugar-phosphate backbone atoms of RNA. *New J. Chem.*, **34**, 910–917.
47. Kondo,J. and Westhof,E. (2010) Base pairs and pseudo pairs observed in RNA-ligand complexes. *J. Mol. Recognit.*, **23**, 241–252.
48. Cheng,A.C., Chen,W.W., Fuhrmann,C.N. and Frankel,A.D. (2003) Recognition of nucleic acid bases and base-pairs by hydrogen bonding to amino acid side-chains. *J. Mol. Biol.*, **327**, 781–796.
49. Główka,M., Martynowski,D. and Kozłowska,K. (1999) Stacking of six-membered aromatic rings in crystals. *J. Mol. Struct.*, **474**, 81–89.
50. Zhang,H., Li,C., Yang,F., Su,J., Tan,J., Zhang,X. and Wang,C. (2014) Cation-pi interactions at non-redundant protein–RNA interfaces. *Biochemistry (Mosc)*, **79**, 643–652.
51. Cech,T.R. (2009) Crawling out of the RNA world. *Cell*, **136**, 599–602.
52. Paukstelis,P.J., Chen,J.H., Chase,E., Lambowitz,A.M. and Golden,B.L. (2008) Structure of a tyrosyl-tRNA synthetase splicing factor bound to a group I intron RNA. *Nature*, **451**, 94–97.