Nonstandard RNA/RNA interactions likely enhance folding and stability of segmented ribosomes

MARIO RIVAS and GEORGE E. FOX
Department of Biology and Biochemistry, University of Houston, Houston, Texas 77204, USA

ABSTRACT
The ribosome is the molecular factory that catalyzes all coded protein synthesis in extant organisms. Eukaryotic ribosomes are typically assembled out of four rRNAs; namely, 5S, 5.8S, 18S, and 28S. However, the 28S rRNA of some trypanosomatid organisms has been found to be segmented into six independent rRNAs of different sizes. The two largest segments have multiple sites where they jointly form stems comprised of standard base pairs that can hold them together. However, such regions of interaction are not observed among the four smaller RNAs. Early reports suggested that trypanosomatid segmented ribosome assembly was essentially achieved thanks to their association with rProteins. However, examination of cryo-EM ribosomal structures from *Trypanosoma brucei*, *Leishmania donovani*, and *Trypanosoma cruzi* reveals several long-range nonstandard RNA/RNA interactions. Most of these interactions are clusters of individual hydrogen bonds and so are not readily predictable. However, taken as a whole, they represent significant stabilizing energy that likely facilitates rRNA assembly and the overall stability of the segmented ribosomes. In the context of origin of life studies, the current results provide a better understanding of the true nature of RNA sequence space and what might be possible without an RNA replicase.

Keywords: segmented ribosomes; kinetoplastids; trypanosomatid; nonstandard RNA interactions

INTRODUCTION
Ribosomes are responsible for coded protein synthesis in all living organisms. A ribosome is comprised of a small subunit (SSU), which contains the decoding site, and a large subunit (LSU), which contains the catalytic site (Steitz 2008; Fox 2010). Both subunits are composed of ribosomal RNA (rRNA) and ribosomal proteins (rProteins) (Wilson and Doudna Cate 2012). In bacteria and archaea there are three rRNAs, named 5S, 16S, and 23S rRNA, where S refers to Svedberg units. Eukaryotic ribosomes are typically formed by four slightly larger rRNA chains; namely, 5S, 5.8S, 18S, and 26S (or 28S) rRNA (Wilson and Doudna Cate 2012). The 5.8S rRNA is homologous to the 5′ end of the bacterial 23S rRNA.

The ribosomes of many protists contain RNAs of unusual size and numbers. The LSU of *Crithidia oncopelti* was first reported to dissociate nonenzymatically into two RNA components in 1976 (Spencer and Cross 1976). Subsequently, two large rRNAs that together formed the equivalent of most of the 26S rRNA were reported in *Leishmania donovani* (Leon et al. 1978). A similar dissociation phenomenon was reported for the kinetoplast rRNA in *Leishmania tarentolae* (Simpson and Simpson 1978). The occurrence of low molecular weight rRNAs after denaturalization of the LSU was first reported in *Crithidia fasciculata* (Gray 1979). In light of these findings, it was proposed that the rRNA pattern described in *C. fasciculata* was perhaps common to all trypanosomatid hemoflagellates (Gray 1979). Shortly thereafter, the presence of additional unusual small rRNAs in *Trypanosoma brucei* was seen (Cordingley and Turner 1980). These observations were followed by reports that confirmed the presence of low molecular weight rRNA components and their characterization in *C. fasciculata* (Gray 1981; Schnare et al. 1983), and other protists from the Trypanosomatidae family including *Trypanosoma cruzi* (Hernández et al. 1983).

It was concluded that the 26S rRNA from the LSU of several parasitic trypanosomatid protists, like *T. brucei* and *C. fasciculata*, has been replaced by two large rRNAs (α and β), and four small rRNAs (White et al. 1986; Campbell et al. 1987; Spencer et al. 1987). A more extreme case of

Corresponding author: fox@uh.edu

Article is online at http://www.rnajournal.org/cgi/doi/10.1261/rna.079006.121. Freely available online through the RNA Open Access option.

© 2022 Rivas and Fox This article, published in RNA, is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at http://creativecommons.org/licenses/by-nc/4.0/.
ribosome segmentation was found in *Euglena gracilis*, a single-cell alga whose cytoplasmic ribosome is made of 16 discreet rRNAs (Schnare and Gray 1990).

Generation of an active ribosome begins by RNA processing to release the pieces that will comprise each functional rRNA. This begins with the removal of internal transcribed spacers (ITSs), analogous to the editing event from which the 5.8S rRNA is obtained (Spencer et al. 1987). In *E. gracilis*, a single-cell alga, the processing of unmatured rRNA releases the mature 18S and the 5.8S rRNAs. Further cleavage splits the remaining LSU rRNA into 13 pieces by removing several ITSs (Greenwood and Gray 1998). The resulting pieces retain the potential to interact with one another and thereby assemble a functional LSU rRNA (Schnare and Gray 1990; Schnare et al. 1990). In the case of the larger rRNA fragments, standard base pairing is likely involved in the assembly of the active subunit. This is not the case for the many smaller fragments, which do not exhibit obvious sites of standard base pairing among themselves nor with larger segments of the LSU rRNA (Matzov et al. 2020).

The situation is similar in the trypanosomatid segmented ribosomes. In this case, the LSU is segmented into seven rRNAs. They are the product of post-transcription modification of a single unmatured rRNA transcript unit that lacks the 5S rRNA but contains the 18S SSU rRNA, the 5.8S rRNA, and six segments from the LSU; namely, rRNA-α (1847 nt), srRNA-I (210 nt), rRNA-β (1465 nt), srRNA-II (182 nt), srRNA-III (73 nt), and srRNA-IV (125 nt) (Fig. 1; Campbell et al. 1987; Hashem et al. 2013). In general, the LSU rRNA from trypanosomatid resembles the LSU rRNA of Saccharomyces cerevisiae according to cryo-EM structural data (Hashem et al. 2013). The rRNA-α loosely corresponds to domains I and II, srRNA-I to domain III, rRNA-β to domains IV and V, and srRNA-II, srRNA-III, and srRNA-IV correspond to domain VI of the *S. cerevisiae* LSU rRNA (Fig. 1). The 5.8S rRNA and the larger segments α and β exhibit zones of multiple canonical base pairing that act as stabilizing elements for an rRNA scaffold to which the rest of the segments anchor (Liu et al. 2016).

It was recognized that trypanosomatid ribosomes present an uncommon abundance in expansion segments (Gao et al. 2005). This observation, as well as the involvement of trypanosome-specific proteins for ribosome biogenesis (Úmaer et al. 2014), and the identification of

**FIGURE 1.** Segmentation pattern of ribosomes from *Trypanosoma brucei*, *Leishmania donovani*, and *Trypanosoma cruzi*. The LSU from these organisms is well known to be segmented in the characteristic eukaryotic 5.8S and six rRNA fragments; rRNA-α, srRNA-I, rRNA-β, srRNA-II, srRNA-III, and srRNA-IV. (A) 28S rRNA secondary structure from *Saccharomyces cerevisiae*, based on crystallographic structure data (Petrov et al. 2014), with the ribosomal segments from trypanosomatids drawn on top of it according to structural alignments (Supplemental Table S5). Color code is as follows: 5.8S is presented in blue, rRNA-α is in red, srRNA-I is in magenta, rRNA-β is in green, srRNA-II is in cyan, srRNA-III is in orange, and srRNA-IV is in purple; 5S rRNA is depicted in yellow in A and 18S is depicted in gray in B. Numbering system corresponds to *S. cerevisiae*. (B) Transcription unit from which the segmented ribosomes are originated. It starts with the 18S rRNA, followed by the 5.8S and the rest of the fragments in consecutive order when compared against unsegmented eukaryotic ribosomes.
trypanosome-specific RNA/protein interactions in T. cruzi (Umaer and Williams 2015) led to the idea that assembly and stabilization of these segmented ribosomes was mainly a protein-dependent process involving “glue” proteins (Liu et al. 2016). For instance, in T. cruzi, rProteins like eL19 and eL34 have acquired a carboxy-terminal extension and one insertion, respectively. These modifications appear to contribute to their association with srRNA-I. Ribosomal protein uL3 stabilizes srRNA-II forming a C-shaped cavity. Together, L14 and eL33 have a carboxy-terminal extension and an insertion at its globular domain, respectively. Both aid in the association of srRNA-III to rRNA-α. As for srRNA-IV, a positively charged section of amino acids from uL3 as well as an amino-terminal extension of the eL31 seem to contribute to its association to the main scaffold (Liu et al. 2016).

However, this may not be the whole story. It has been shown that the low-molecular-weight rRNAs from C. fasciculata also associate with larger ribosomal segments depending on magnesium concentrations (Gray 1981). Although not explored in detail, magnesium ions, RNA/RNA interactions, and modified nucleotides have been suggested to also contribute to the stabilization of the main scaffold and enhance the association of the smaller rRNAs (Liu et al. 2016). This suggestion agrees with extant knowledge of the ability of RNAs to fold and form compact structures thanks to metallic ion interactions (Pyle 2002, Klein et al. 2004). Also, the ability of long-range nonstandard interactions to stabilize RNA structure and enhance the association of independent RNAs has been observed in studies of the pseudosymmetric region in the peptidyl transferase center of the ribosome (Agmon et al. 2005; Agmon 2009; Rivas and Fox 2020).

In the present study, we take advantage of extant cryo-EM structures from the trypanosomatid ribosomes from T. brucei, L. donovani, and T. cruzi to identify RNA/RNA interactions that might aid in the association of the small rRNAs to the main scaffold. Long-range RNA/RNA interactions were found to cluster into conserved regions that might enhance this association process (Fig. 2).

RESULTS AND DISCUSSION

Structural analysis of the RNA/RNA interactions of the independent segments among the ribosomal cryo-EM structures from T. brucei, L. donovani, and T. cruzi reveal several common regions of interaction. Splayed-apart nucleotides are a widespread feature that aid in the establishment of standard and nonstandard base pairing, base stacking units, and A-minor interactions, as well as standalone hydrogen bonding interactions. Individual hydrogen bonds are the most abundant of all (Tables 1–4). When combined they most likely represent a large stabilizing energy that facilitates RNA assembly and contribute to the stability of the segmented ribosomes. However, due to their nature, individual hydrogen bonds are not easily predictable. Of all the possibilities, the 2′OH from ribose is the most common functional group by which individual hydrogen bonds are established (Supplemental Table S6). The ribose zipper is another type of RNA/RNA interaction that seems to contribute to the stabilization of close contact areas, commonly found between two helices almost touching each other. Together, these interactions are referred to as long-range interactions because they allow distant RNA secondary structure elements to fold into compact and complex tertiary structures. In the context of the segmented ribosomes, they seem to combine and converge into specific areas where two or more segments meet and interact, favoring the stabilization of the segmented ribosome as a whole.

The specific circumstances of each of the small RNAs are described below.

srRNA-I

srRNA-I from T. brucei possesses eight regions of RNA/RNA interaction. These regions facilitate its association with rRNA-α, rRNA-β, and srRNA-II (Table 1). They comprise a variant number of nucleotides that range from one splayed-apart cytosine that creates two hydrogen bonds with rRNA-β, in the case of interacting region 4, to 5 nt that create seven hydrogen bonds with nucleotides that belong to rRNA-α and rRNA-β, respectively, in the case of interacting region 8 (Table 1). Regions 2, 3, 5, and 7 are persistent elements in T. brucei, L. donovani, and T. cruzi cryo-EM ribosome structures (Fig. 2A). These regions vary in the nature of the RNA/RNA interactions that are established in each organism. In region 2, for example, while it is formed by four nucleotides that interact with srRNA-II, those interactions vary in nature depending on the organism. In T. brucei, U52, U54, C55, and U56 from srRNA-I create simple hydrogen bonds with A97, A98, and A99. C55 also establishes a standard base pair with G97, while U56 establishes a nonstandard base pair with G97. In L. donovani, U56, U57, and C58 from srRNA-I establish standard base pairing with A99, A98, and G97 from srRNA-II. The latter is also true for T. cruzi (Table 1; Fig. 3A,B; Supplemental Table S1).

srRNA-II

srRNA-II from T. brucei has 13 regions of interaction that seem to facilitate its association with the 5.8S rRNA and the rest of the segments from the LSU (rRNA-α, rRNA-β, srRNA-I, srRNA-III, and srRNA-IV) (Table 2). The number of nucleotides involved ranges from only two in regions 9 and 12, to seven in regions 2 and 5 (Table 1). Regions 3, 4, 6, 7, 8, 9, 10, 11, and 13 are common features within T. brucei, L. donovani, and T. cruzi cryo-EM ribosomal structures (Fig. 2B). These common interaction regions
also vary in the number of nucleotides and the number of segments involved in each interaction depending on the organism. For instance, region 3 in *T. brucei* has 5 nt that establish contacts with nucleotides from rRNA-β, srRNA-III, and srRNA-IV. In *L. donovani* the same region involves 3 nt and establishes contacts only with srRNA-III and srRNA-IV. This region involves 5 nt in *T. cruzi* that exclusively contact nucleotides from srRNA-IV (Table 2; Fig. 3C,D; Supplemental Table S2).

**srRNA-III**

srRNA-III from *T. brucei* exhibits four regions that allow its interaction with rRNA-α, rRNA-β, and srRNA-II (Table 3). A different number of nucleotides is involved in each case. They range from one nucleotide in region 3 and 4, to four in regions 1 and 2 (Table 3). Regions 1, 2, and 3 seemed to be common to all three structures analyzed here (Fig. 2C). Some regions of interaction involve large numbers of interactions while others have fewer interacting elements. Region 3 is from the latter group. In *T. brucei* it is formed by G23, where it establishes a single hydrogen bond with A1471 from rRNA-β. In *L. donovani* this interaction is maintained by A14 which creates two hydrogen bonds with A1387 and establishes a nonstandard base pair with U1388 from rRNA-β (Fig. 3E,F). In *T. cruzi* this interacting region is barely upheld by A10, which establishes a single hydrogen bond with A1581 from rRNA-β (Table 3; Supplemental Table S3).
srRNA-IV

srRNA-IV from T. brucei possesses seven regions of interaction that connect it with rRNA-α, rRNA-β, and srRNA-II (Table 4). These interactions are formed by nucleotides that range from two, in regions 1, 2, and 6, to four in regions 3 and 5 (Table 4). Regions 1, 3, 5, 6, and 7 are common to T. brucei, L. donovani, and T. cruzi cryo-EM ribosome structures (Fig. 2D). Variation in the number of nucleotides and the nature of the interactions they establish seems to be the dominant theme among all these segments. Region 7 in T. brucei is represented by A101, C102, and C103. Each establishes a single hydrogen bond with U127, C126, and U379, respectively (from rRNA-β). Meanwhile, in L. donovani and T. cruzi, the interaction involves 6 nt in each case, but more important than the number is the nature of their interactions. Although single hydrogen bonds are still present, the establishment of A-minor motifs within this region is worth mentioning (Table 4; Supplemental Table S4). Region 7 in L. donovani exhibits two A-minor motifs, where A69 interacts with the pair C124:A292 and A87 interacts with the pair G123:C293. Region 7 in T. cruzi presents one A-minor motif that involves A92 and the pair G123:C456 (Table 4; Fig. 3G,H; Supplemental Table S4).

The srRNA-II segment exhibits the largest number of interactions. It contacts all the other segments from the LSU; namely, 5.8S, rRNA-α, rRNA-β, srRNA-I, srRNA-III, and srRNA-IV. Due to the latter, many regions of the other small rRNAs have a counterpart depending on which segment was used as reference. srRNA-II regions 1 and 3 include interactions with C11 listed in region 2 of srRNA-III. srRNA-II region 3 shows nonstandard base pairs with nucleotides from srRNA-IV in the case of T. brucei, while it only contacts nucleotides from srRNA-IV in the case of T. cruzi (Table 2; Supplemental Table S2). E and F illustrate differences in srRNA-III region 3 from T. brucei and L. donovani. This region is among those with the least interacting nucleotides but one that remains constant among trypanosomatids (Table 3; Supplemental Table S3). G and H illustrate differences in the nature of interactions with srRNA-IV region 7 from T. brucei and T. cruzi. In T. brucei, 3 nt create single hydrogen bonds (yellow lines) while in T. cruzi an A-minor interaction is created with srRNA-IV A92 and the pair rRNA-β C124:G456, in addition to several hydrogen bonds (yellow lines) (Table 4; Supplemental Table S4). Numbering systems correspond to cryo-EM ribosome structures from Trypanosoma brucei (PDBID 4V8M), Leishmania donovani (PDBID 5T2A), and Trypanosoma cruzi (PDBID ST5H), respectively.

FIGURE 3. RNA/RNA interaction differences between equivalent persistent regions among trypanosomatid species. A and B illustrate differences in the contact nature in srRNA-I region 2 from T. brucei and L. donovani. This area is represented by several hydrogen bonds (yellow lines) and one standard and one nonstandard base pairing in T. brucei, while in L. donovani and T. cruzi this area is represented by three consecutive standard base pairings (Table 1; Supplemental Table S1). C and D illustrate differences in the number of segments interacting with srRNA-II region 3 from T. brucei and T. cruzi. This area associates with several nucleotides from rRNA-β, srRNA-III and srRNA-IV in the case of T. brucei, while it only contacts nucleotides from srRNA-IV in the case of T. cruzi (Table 2; Supplemental Table S2). E and F illustrate differences in srRNA-III region 3 from T. brucei and L. donovani. This region is among those with the least interacting nucleotides but one that remains constant among trypanosomatids (Table 3; Supplemental Table S3). G and H illustrate differences in the nature of interactions with srRNA-IV region 7 from T. brucei and T. cruzi. In T. brucei, 3 nt create single hydrogen bonds (yellow lines) while in T. cruzi an A-minor interaction is created with srRNA-IV A92 and the pair rRNA-β C124:G456, in addition to several hydrogen bonds (yellow lines) (Table 4; Supplemental Table S4). Numbering systems correspond to cryo-EM ribosome structures from Trypanosoma brucei (PDBID 4V8M), Leishmania donovani (PDBID 5T2A), and Trypanosoma cruzi (PDBID ST5H), respectively.
realize that there are additional regions that have the same potential, but whose presence could not be determined due to incomplete/unresolved structures. These include srRNA-I region 4 whose interacting nucleotide is localized in one small expansion segment (sES in Fig. 2A) that was not resolved in the cryo-EM structure from L. donovani nor from T. cruzi. Another example occurred in srRNA-II region 12 whose interacting nucleotides are localized in helix 97 (Figs. 1A, 2B), which was not resolved in the T. cruzi structure. srRNA-III region 4 whose interacting nucleotide is localized in helix 98a (Figs. 1A, 2C) is not resolved in the T. cruzi structure. srRNA-IV region 4 comprises two types of inconsistencies among structures. There, A27 seems to be a splayed-apart nucleotide due to the elongation of the chain by three nucleotides that are localized in helix 98b (Figs. 1B, 2A, 2C). Such an addition of material results in a

| Region | Nucleotide 1 | Segment | Nucleotide 2 | Distance (Å)/interaction | HB/type | Functional groups |
|--------|--------------|---------|--------------|---------------------------|---------|------------------|
| 1      | G10          | rRNA-α  | G1846        | 2.5                       | B–S     | O–OH             |
|        | rRNA-β       | A5      |              | 3.0                       | B–S     | NH$_2$–OH        |
|        | A11          | rRNA-α  | G1846        | 2.7                       | B–S     | O–OH             |
|        | rRNA-β       | A5      |              | 3.2                       | B–B     | NH$_2$–N         |
| 2*     | U52          | srRNA-II| G97          | 3.2                       | sBP     | cWW              |
|        |              |         | A98          | 2.7                       | B–B     | NH$_2$–N         |
|        |              |         | A99          | 3.1                       | B–S     | N–OH             |
|        |              |         |              |                           | B–S     | NH$_2$–OH        |
|        |              |         |              |                           | B–B     | NH$_2$–NH$_2$    |
|        | U56          | srRNA-II| G97          |                           | nsBP    | tWS              |
|        |              |         | A99          |                           | B–B     | NH$_2$–NH$_2$    |
| 3*     | U65          | rRNA-α  | C1813        | 2.7                       | B–S     | O–OH             |
|        |              |         |              |                           | S–S     | OOH–OH           |
|        | A66          | rRNA-α  | A1834        |                           | A-minor | Type X           |
|        |              |         |              |                           |         |                  |
|        | A67          | rRNA-α  | A1835        |                           | S–S     | OOH–OH           |
| 4      | C78          | rRNA-β  | A151         | 3.6                       | B–B     | NH$_2$–N         |
|        |              |         |              |                           | B–B     | NH$_2$–NH$_2$    |
| 5*     | U110         | rRNA-α  | G955         | 2.6                       | S–P     | O–OH             |
|        | C111         | rRNA-α  | G955         | 3.7                       | S–S     | NH–OH            |
|        | G112         | rRNA-α  | G937         | 2.5                       | P–B     | O–NH$_2$         |
|        | C113         | rRNA-α  | G937         | 2.5                       | P–S     | O–OH             |
| 6      | C160         | rRNA-α  | G1817        |                           | P–B     | O–NH$_2$         |
| 7*     | U191         | rRNA-α  | G1807        | 2.9                       | B–S     | O–OH             |
|        | A192+        | rRNA-α  | U1701+       |                           | 2.7     | S–S               | OHS–OH |
|        |              |         | G1807        |                           | B–S     | O–OH             |
|        |              |         | A193+        |                           | 2.9     | S–S               | O–OH–OH |
|        |              |         | C1700+       |                           | B–S     | NH$_2$–OH        |
|        |              |         | A1807        |                           | 2.7     | B–S               | NH$_2$–OH |
|        |              |         | A1808        |                           | A-minor | Type X           |
|        |              |         |              |                           |         |                  |
| 8      | G205         | rRNA-β  | U1           | 2.9                       | S–B     | O–OH             |
|        | G206         | rRNA-α  | C1783        | 3.8                       | P–S     | O–OH             |
|        | G207         | rRNA-α  | C1782        | 3.6                       | B–S     | NH$_2$–NH$_2$    |
|        | G208         | rRNA-α  | C1783        | 3.6                       | B–S     | NH$_2$–NH$_2$    |
|        |              |         |              |                           | B–S     | NH$_2$–NH$_2$    |
|        |              |         |              |                           | B–S     | NH$_2$–NH$_2$    |

Details of the eight interaction regions in the srRNA-I segment from *T. brucei*. These regions allow its interaction with rRNA-α, rRNA-β, and srRNA-II. “Nucleotide 1” column contains the number of the nucleotide from the srRNA-I that establishes interactions with other segments and their respective nucleotides (columns “Segment” and “Nucleotide 2,” respectively). Hydrogen bonding is displayed using its length in Angstroms (“Distance (Å)/interaction” column). Structural elements of the nucleotide that are involved in the hydrogen bonding are shown in the “HB/type” column using capital letters: B for base, S for sugar, and P for phosphate. Functional groups involved in the establishment of the hydrogen bond are shown on the last column. Standard and nonstandard base pairs (sBP and nsBP) are also shown in the “Distance (Å)/interaction” column. Base pair geometry is indicated using Leontis-Westhof nomenclature (Leontis and Westhof 2001) in the “HB/type” column. A-minor motifs and their specific types are also indicated in the “Distance (Å)/interaction” and “HB/type” columns. Members of ribose zipper units are marked with a plus sign (+) in region 7. *T. brucei* cryo-EM structure numbering system is used (PDB ID 4VBM). Regions that are common structural elements in *T. brucei*, *L. donovani*, and *T. cruzi* cryo-EM ribosomes are marked with an asterisk (*).
| Region | Nucleotide 1 | Segment | Nucleotide 2 | Distance (Å)/interaction | HB/type | Functional groups |
|--------|-------------|---------|-------------|--------------------------|---------|------------------|
| 1      | G1          | rRNA-β  | C1461       | 2.7                      | B-B     | O–NH₂          |
|        |             |         |             | 3.0                      | B-S     | NH₂–OH        |
|        |             |         |             | 3.6                      | B-B     | NH₂–O         |
|        | U2          | rRNA-β  | C1461       | 2.6                      | S-B     | OH–O          |
|        | G3          | srRNA-III| C11         | 4.0                      | B-B     | N-NH₂        |
|        |             |         |             | 4.0                      | P-B     | O–NH₂        |
| 2      | G11         | rRNA-α  | G480        | 3.2                      | B-S     | O–OH         |
|        | A12         | rRNA-α  | U231        | 2.5                      | B-P     | NH₂–O        |
|        | A13         | rRNA-α  | C238        | 2.7                      | P-S     | O–OH         |
|        | G14         | rRNA-α  | A237        | 2.6                      | B-B     | NH₂–O        |
|        |             |         |             | 3.1                      | B-S     | O–OH         |
|        | G15         | rRNA-α  | A483        | nsBP                     | tWS     | —              |
|        | G16         | rRNA-α  | A483        | 2.9                      | S-B     | OH–N          |
|        |             |         |             | 3.2                      | S-B     | OH–N          |
|        | A17         | rRNA-α  | A483        | 3.7                      | P-S     | O–OH         |
| 3*     | G22         | srRNA-III| C11         | 2.9                      | B-S     | N-OH          |
|        | C23         | rRNA-β  | G1462       | 3.3                      | P-S     | O–OH          |
|        | G24         | srRNA-IV| U135        | nsBP                     | cHH     | —              |
|        | G25         | rRNA-β  | G1464       | 2.7                      | P-B     | O–NH₂        |
|        | G26         | srRNA-IV| U2          | nsBP                     | tSW     | —              |
| 4*     | A42         | rRNA-β  | U1349       | 3.0                      | P-S     | O–OH          |
|        | C133+       | rRNA-α  | U1306       | 2.5                      | P-B     | O–NH         |
|        |             | rRNA-β  | U1367+      | 2.6                      | S-S     | OH–OH        |
|        | U134+       | rRNA-β  | C1352       | 3.4                      | B-S     | O–OH          |
|        |             |         | C1366+      | 3.1                      | S-S     | OH–OH        |
|        | C135        | rRNA-β  | C1352       | 2.7                      | S-S     | OH–OH        |
|        |             |         |             | 2.6                      | S-S     | OH–OH        |
|        | A157        | rRNA-β  | A1350       | nsBP                     | tW.     | —              |
|        | A158        | rRNA-β  | A1346       | nsBP                     | cSW     | —              |
|        |             |         | A1368       | nsBP                     | G1370   | 3.0          | S-B          | OH–NH₂    |
|        |             |         |             | 3.5                      | S-B     | OH–N          |
| 5      | A52         | rRNA-β  | G1359       | 2.7                      | B-S     | N-OH          |
|        | C53         | rRNA-β  | G1359       | 2.7                      | B-S     | O–OH          |
|        |             |         |             | A-minor                 | Type II | —              |
|        | G54         | rRNA-β  | A1358       | 3.3                      | S-B     | OH–N          |
|        | G56         | rRNA-β  | A1358       | nsBP                     | cWH     | —              |
|        | A57         | rRNA-β  | A1358       | nsBP                     | tWW     | —              |
|        | G58         | rRNA-β  | G1359       | nsBP                     | tSS     | —              |
|        | A60         | rRNA-β  | A1360       | 2.5                      | S-P     | OH–O          |
| 6*     | U82         | rRNA-β  | G609        | 2.7                      | S-S     | OH–OH        |
|        | U83         | rRNA-β  | U79         | 3.2                      | P-S     | O–OH          |
|        | C117        | rRNA-β  | G609        | 3.9                      | S-P     | OH–O          |
|        | U118        | rRNA-β  | A608        | 3.5                      | S-S     | OH–OH        |

Continued
perturbation of the *T. brucei* structure, causing some of the extra bases to be exposed to the solvent at the beginning of helix 101 (Figs. 1A, 2D). In addition, A121 and U122 are localized in a region of helix 101 that was not resolved in either *L. donovani* or *T. cruzi* structures. This suggests that the number of interacting regions common to these srRNA segments could be larger than estimated here.

The ability of RNAs to fold into compact structures thanks to metallic ion interactions has long been recognized (Pyle 2002; Klein et al. 2004). Most likely Mg$^{2+}$ and other divalent and monovalent cations play a role in the stabilization of the srRNAs. Unfortunately, the only cryo-EM structure used in the present study that reports metallic cations was the one from *Trypanosoma cruzi*. As mentioned above, this structure has been resolved lacking many rRNA elements. So far, we have not been able to identify any cation in the nearest area of the interaction regions of the srRNAs.

### TABLE 2. Continued

| Region | Nucleotide 1 | Segment | Nucleotide 2 | Distance (Å)/interaction | HB/type | Functional groups |
|--------|-------------|---------|-------------|--------------------------|---------|------------------|
| 7*     | U86         | srRNA-IV | A116        | 3.6                      | P–B     | O–NH$_2$         |
|        | G87         | srRNA-IV | G113        | 3.3                      | S–S     | OH–OH            |
|        | G88         | srRNA-IV | C40         | 2.7                      | S–S     | OH–OH            |
|        |             |         | A41         | 3.3                      | B–P     | NH$_2$–OH        |
|        | C112        | srRNA-IV | C40         | 3.7                      | S–P     | OH–O             |
| 8*     | G97         | srRNA-I  | U52         | 3.2                      | B–B     | NH–O             |
|        |             |         | C55         | sBP                      | cWW     | —                |
|        |             |         | U56         | nsBP                     | tWS     | —                |
|        | A98         | srRNA-I  | C55         | 2.7                      | B–B     | N–NH$_2$         |
|        |             |         |             | 3.1                      | S–B     | OH–N             |
|        |             |         |             | 3.4                      | S–B     | OH–NH$_2$        |
|        | A99         | srRNA-I  | U54         | 3.6                      | B–B     | NH$_2$–NH        |
|        |             |         | C55         | 3.1                      | B–B     | NH$_2$–NH$_2$    |
| 9*     | G100        | srRNA-IV | A44         | 4.0                      | S–S     | OH–OH            |
|        | G101        | srRNA-IV | G43         | 3.9                      | P–S     | O–OH             |
| 10*    | C103        | rRNA-α   | A1644       | 3.4                      | S–P     | OH–O             |
|        | A104        | rRNA-α   | C1645       | 3.0                      | S–P     | OH–O             |
|        | G106        | rRNA-α   | G1605       | BS                       | —       | —                |
| 11*    | U107        | rRNA-β   | A77         | 3.4                      | B–S     | OH–O             |
|        | G108       | rRNA-β   | C385        | 3.1                      | S–S     | OH–OH            |
|        | C109       | rRNA-β   | A384        | 2.6                      | S–S     | OH–OH            |
| 12     | G144        | rRNA-α   | C1342       | sBP                      | cWW     | —                |
|        | C145        | rRNA-α   | U1345       | 3.8                      | S–P     | OH–O             |
| 13*    | A171        | rRNA-β   | G652        | 3.3                      | S–P     | OH–O             |
|        |             |         | G1452       | nsBP                     | tWW     | —                |
|        | C172        | rRNA-β   | G1453       | 2.7                      | B–S     | NH$_2$–OH        |
|        |             |         | G1454       | 2.8                      | P–B     | O–NH$_2$         |
|        |             |         | G652        | BS                       | —       | —                |
|        | C173        | rRNA-β   | G1454       | 3.3                      | P–B     | O–NH$_2$         |
|        | C174        | rRNA-β   | G1454       | 3.4                      | P–B     | O–NH$_2$         |

Details of the thirteen interaction regions in the srRNA-II segment from *T. brucei*. These regions allow its interaction with the 5.8s as well as the rRNA-α, rRNA-β, srRNA-I, srRNA-III, and srRNA-IV. "Nucleotide 1" column contains the number of the nucleotide from the srRNA-II that establishes interactions with other segments and their respective nucleotides (columns "Segment" and "Nucleotide 2," respectively). Column distribution follows the same pattern as Table 1. Base stacking interactions (BS) are shown in the "Distance (Å)/interaction" column. Members of ribose zipper units are marked with a plus sign (+) in regions 4 and 11. *T. brucei* cryo-EM structure numbering system is used (PDB ID 4V8M). Regions that are common structural elements in *T. brucei*, *L. donovani*, and *T. cruzi* cryo-EM ribosomes are marked with an asterisk (*).
As stated above, the small rRNA segments II, III, and IV loosely correspond to domain VI of the *S. cerevisiae* ribosome. Of these, srRNA-III has the greatest variation in sequence and structure. As a consequence, the structural alignments of this region between trypanosomatid segments and the ribosome crystallographic structures from *S. cerevisiae* and other eukaryotes has proven to be a tough task (Fig. 2C; Supplemental Table S5). It was no surprise that the srRNA-III exhibits such variation and structural versatility since it grossly corresponds to what has been called expansion segment 39. Its sequence and topology are highly variable among eukaryotes, and recently it has been found to be even more variable among the Asgard archaea ribosomes (Penev et al. 2020).

The 2′OH from ribose was found to participate in the largest number of individual hydrogen bond interactions. Overall, more than two-thirds of the individual interactions reported here involved a 2′OH. Their abundance seems to be high regardless of the organism, with >68% in *T. brucei*, 78% in *L. donovani*, and 91% in *T. cruzi* (Supplemental Table S6). Such a pronounced increment in frequency can be attributed in part to unresolved regions in the cryo-EM structures from *L. donovani* and *T. cruzi*. The 2′OH forms hydrogen bonds in two primary ways. It can share a hydrogen with another 2′OH group or act as the donor when in proximity of an oxygen from either one of the bases or the phosphate backbone. The latter case is the highest in frequency according to our observations (Supplemental Table S6). Other less abundant forms of possible hydrogen bonding include the 2′OH being in the proximity of the different nitrogen atoms from the nucleobases.

As previously observed, long-range interactions are a type of nonstandard interactions that occur as an intrinsic characteristic of RNA molecules. Such interactions might have been crucial throughout life’s history and origin (Rivas and Fox 2020). This intrinsic RNA feature seems to be beneficial to organisms whose survival depends on ribosome assembly, stability, and functionality despite the segmentation phenomenon. Several examples of smaller functional noncoding RNAs that fold into a stable form as well as into a functional state due to these types of long-range nonstandard RNA/RNA interactions were described more than 20 yr ago. The contribution of base stacking to an RNA fold was first observed in the crystal structure of tRNA^phe^ (Kim et al. 1974; Robertus et al. 1974). The hammerhead ribozyme (Pley et al. 1994), the hepatitis virus ribozyme (Ferré-D’Amaré et al. 1998). Nonstandard base pairs that group into multiplets have been described as stabilizing elements in tRNAs (Kim et al. 1974; Robertus et al. 1974; Jack et al. 1976) and the *Tetrahymena* intron (Cate et al. 1996), among

### TABLE 3. RNA/RNA interaction regions for srRNA-III fragment from *T. brucei*

| Region | Nucleotide 1 | Segment | Nucleotide 2 | Distance (Å)/interaction | HB/type | Functional groups |
|--------|--------------|---------|--------------|--------------------------|---------|------------------|
| 1*     | C1           | rRNA-α  | G690         | 3.6                      | B–P     | NH₂–O           |
| G2     | rRNA-α       | U516    | G688         | 3.6                      | B–S     | NH₂–OH          |
|        |              |         |              | 4.0                      | S–S     | OH–OH           |
| A3     | rRNA-α       | U514    | C689         | 3.6                      | A-minor | Type X          |
|        |              |         |              | 4.0                      | S–B     | OH–N            |
| 2*     | C11          | srRNA-II| U2           | 3.6                      | B–S     | NH₂–OH          |
|        |              |         | G3           | 4.0                      | B–P     | NH₂–O           |
|        |              |         |              | 4.0                      | B–B     | NH₂–N           |
| U12    | srRNA-II     | G22     | nsBP         | 2.9                      | S–B     | OH–N            |
| C14    | rRNA-β       | U1458   | G22          | 2.9                      | S–B     | OH–N            |
| C16    | rRNA-β       | U1458   |              | 2.9                      | S–B     | OH–N            |
| 3*     | G23          | rRNA-β  | A1471        | 3.1                      | P–B     | O–NH₂           |
| 4      | U40          | rRNA-α  | C552         | 2.9                      | B–P     | NH–O            |
|        |              |         |              | 3.7                      | S–P     | OH–O            |

Details of the four interaction regions in the srRNA-III segment from *T. brucei*. These regions allow its interaction with rRNA-α, rRNA-β, and srRNA-II.

“Nucleotide 1” column contains the number of the nucleotide from the srRNA-III that establishes interactions with other segments and their respective nucleotides (columns “Segment” and “Nucleotide 2,” respectively). Column distribution follows the same pattern as Table 1. *T. brucei* cryo-EM structure numbering system is used (PDB ID 4V8M). Regions that are common structural elements in *T. brucei*, *L. donovani*, and *T. cruzi* cryo-EM ribosomes are marked with an asterisk (*).
The A-minor motif by itself has long been recognized as "an abundant and ubiquitous structural motif that stabilizes RNA tertiary and quaternary structures" and cataloged as the most abundant tertiary structure in the ribosome LSU (Nissen et al. 2001). To our knowledge this is the first time a structural analysis of their prevalence among segmented ribosomes shows their abundance. It is of special interest how these interactions group into clusters strengthening the smaller rRNA segments to the main scaffold, and how variable these regions can be in terms of the tertiary interactions that belong to the same region in different organisms.

**Conclusions**

The role of RNA/protein interactions has been recognized as a fundamental element that enables segmented ribosome assembly. Although RNA/RNA interactions were suggested to contribute to the assembly and stabilization of these ribosomes, they were not extensively explored (Liu et al. 2016). We herein present a comprehensive analysis of potential RNA/RNA interactions in the small rRNA segments that lack large sections of standard base pairing, which might secure their association with larger segments. Based on their persistence among the cryo-EM structures from *T. brucei*, *L. donovani*, and *T. cruzi*, the results presented here suggest that long-range interactions occur between two or more independent segments grouping into regions of interaction that can enhance the association of these small rRNAs with the main scaffold and might also contribute to its stabilization as a cohesive unit. Differences between equivalent regions in different organisms that were described above are examples of the inherent flexibility of RNA. This allows dynamic changes in the nature of RNA's interactions.

### TABLE 4. RNA/RNA interaction regions for srRNA-IV fragment from *T. brucei*

| Region | Nucleotide 1 | Segment | Nucleotide 2 | Distance (Å)/interaction | HB/type | Functional groups |
|--------|--------------|---------|--------------|--------------------------|---------|------------------|
| 1      | U2           | srRNA-II| G26          | nsBP                     | t5W     | —                |
|        | U135         | srRNA-II| A24          | nsBP                     | cH      | —                |
| 2      | C7           | rRNA-α  | A1630        | 3.8                      | S–P     | OH–O             |
|        | G130         | rRNA-α  | A1629        | 2.6                      | S–S     | OH–OH            |
| 3      | A15          | rRNA-β  | A656         | A-minor                  | Type X  | —                |
|        | A16          | rRNA-β  | G1449        | A-minor                  | Type X  | —                |
|        | A17          | rRNA-β  | G1449        | 3.3                      | B–S     | N–OH             |
|        |              |         | G1450        | 3.0                      | S–S     | OH–OH            |
|        | C18          | rRNA-β  | G1450        | 2.6                      | S–P     | OH–O             |
| 4      | A27          | rRNA-β  | U1533        | 2.6                      | B–S     | N–OH             |
|        |              |         | U1534        | 3.1                      | S–B     | OH–O             |
|        |              |         | G1536        | 2.8                      | B–S     | N–OH             |
|        |              |         |              |                          | 2.9     | NH5–N            |
|        |              |         |              |                          | 3.3     | NH5–OH           |
|        | A121         | rRNA-β  | U1534        | 2.9                      | P–B     | O–NH             |
|        |              |         |              |                          | 3.2     | P–B               |
|        | U122         | rRNA-β  | G1535        | 2.8                      | B–B     | O–NH             |
|        |              |         |              |                          | 3.1     | B–B               |
| 5      | C40          | srRNA-II| G88          | 2.7                      | S–S     | OH–OH            |
|        |              |         | C112         | 3.6                      | S–B     | OH–OH3           |
|        |              |         |              |                          | 3.7     | P–S               |
|        | A41          | srRNA-II| G88          | 3.3                      | P–B     | O–NH2            |
|        | G113         | srRNA-II| G87          | 3.3                      | S–S     | OH–OH            |
|        | A116         | srRNA-II| U86          | 3.6                      | B–P     | NH5–O            |
|        |              |         |              |                          | 3.8     | B–P               |
| 6      | G43          | srRNA-II| G101         | 3.9                      | S–P     | O–O              |
|        | A44          | srRNA-II| G100         | 4.0                      | S–S     | OH–OH            |
| 7      | A101         | rRNA-β  | U127         | 3.8                      | S–P     | O–O              |
|        | C102         | rRNA-β  | C126         | 2.2                      | S–S     | OH–OH            |
|        | C103         | rRNA-β  | U379         | 3.8                      | S–S     | OH–OH            |

Details of the seven interaction regions in the srRNA-IV segment from *T. brucei*. These regions allow its interaction with rRNA-α, rRNA-β, and srRNA-II. "Nucleotide 1" column contains the number of the nucleotide from the srRNA-IV that establishes interactions with other segments and their respective nucleotides (columns "Segment" and "Nucleotide 2," respectively). Column distribution follows the same pattern as Table 1. *T. brucei* cryo-EM structure numbering system is used (PDB ID 4V8M). Regions that are common structural elements in *T. brucei*, *L. donovani*, and *T. cruzi* cryo-EM ribosomes are marked with an asterisk (*)..

www.rnajournal.org
of interactions, the number of nucleotides, and the number of segments involved in these regions. Such flexibility might facilitate the preservation of these associations between segments while consolidating the integrity of trypanosomatid ribosomes.

The fact that the 2′OH is one of the features that distinguishes RNA from its molecular cousin DNA, is of course indicative of RNA’s intrinsic ability to form intricate structures and acquire complex shapes. Even further, the high frequency of the 2′OH associated with either another 2′OH or the phosphate backbone implies that this type of interaction is independent of sequence at these interacting regions, although the overall RNA secondary structure is indeed sequence dependent. Hence, it releases selective pressure over sequence variations/mutations at these specific regions while preserving its potential to assemble equivalent structures. The latter is undoubtedly useful for the small fragments of the segmented ribosomes and most likely has been advantageous throughout life’s evolution.

It has been widely speculated that RNA played a key role in the origin of life (Gilbert 1986; Cech 2009, Robertson and Joyce 2012). However, it is not particularly clear how readily RNAs might have associated to produce prebiotically useful RNAs. Here the RNA/RNA interactions that stabilize the small RNAs, which together comprise the segmented ribosomes of trypanosomatids, were examined in detail. The results contribute to previous efforts in understanding the versatility of RNA while stabilizing structures using a large repertoire of innate properties.

MATERIALS AND METHODS
Structure recruitment and selection criteria

Cryo-EM structures from trypanosomatid ribosomes were collected from the protein data bank (Berman et al. 2000) for Trypanosoma brucei (PDBID 4VBM) (Hashem et al. 2013), Leishmania donovani (PDBID 5T2A) (Zhang et al. 2016), and Trypanosoma cruzi (PDBID 5T5H) (Liu et al. 2016). These structures were selected because they were obtained free from interaction with foreign molecules such as antibiotics, thereby minimizing unintended alteration of possible RNA/RNA interactions. At the time of collection, these structures were the ones with highest available resolution. Crystallographic information files (CIF) were edited to extract rRNA chains without losing their relative position with one another using PyMOL (The PyMOL Molecular Graphics System, version 2.4, Schrödinger LLC).

Analysis of RNA chains

The rRNA containing files were subject to an automated search for predefined long-range RNA/RNA interactions using the DSSR software V1.8.9 (Lu et al. 2015). This included the A-minor motifs (Nissen et al. 2001) and ribose zippers (Batey et al. 1999; Tamura and Holbrook 2002) formed between different rRNA chains or single splayed-apart nucleotides that base stack (Batey et al. 1999) or base-pair with distant chains. Results from the latter were used to corroborate RNA/RNA interaction nucleotides detected by visual inspection of several splayed-apart bases that seemed to be establishing hydrogen bonds with nucleotides that belong to other rRNA segments. Visual inspection and measurements of hydrogen bond distances were performed with PyMOL software (The PyMOL Molecular Graphics System, version 2.4, Schrödinger LLC). Visual inspection was carefully done as a strategy to include long-range nonstandard interactions that are not commonly defined in the available software that was used to analyze RNA structures. Hydrogen bonding distances were included according to Jeffrey HB classification where they range from strong (<2.5 Å) to moderate (2.5–3.2 Å) to weak (3.2–4.0 Å) (Jeffrey 1997).

Defining regions of interaction

From the visual inspection of the cryo-EM structures, it became clear that there were long-range interactions that seemed to be concentrated in local regions. Most of these regions have more than one nonstandard RNA/RNA interaction that mediates the interaction between one of the smaller independent RNA fragments and the main scaffold as well as to other small RNA fragments. Proximal interactions that were found to have such distribution in the cryo-EM structures were grouped as single regions. A search was done following the increment in number for each nucleotide along the chain, and regions tend to follow that same order, although some regions can involve distant nucleotides further in the chain. Comprehensive tables of long-range nonstandard RNA/RNA interactions organized by regions are presented as Tables 1–4 and Supplemental Tables S1–S4.

SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

ACKNOWLEDGMENTS

M.R.’s research was supported by an appointment to the NASA Postdoctoral Program at the NASA Astrobiology Institute, administered by Universities Space Research Association under contract with NASA. This work was also supported in part by a subcontract to the University of Houston from NASA, contract 80NSSC18K1139, under the Center for the Origin of Life at the Georgia Institute of Technology. We thank Dr. Quyen Tran for help in implementing and maintaining computational tools.

Author contributions: M.R. and G.E.F. conceived the work. M.R. conducted all the computational work. Results were discussed with G.E.F. Both authors contributed equally to the writing process and preparation of the manuscript.

Received October 3, 2021; accepted November 26, 2021.

REFERENCES

Agmon I. 2009. The dimeric proto-ribosome: structural details and possible implications on the origin of life. Int J Mol Sci 10: 2921–2934. doi:10.3390/ijms10072921
RNA, it is perfectly plausible to think they have played an important role in the early evolution of RNA molecules.

What led you to study RNA or this aspect of RNA science?

Based on the intrinsic properties that allow RNA to function both as a genetic and catalytic biomolecule, there has been speculation of the possibility of an ancestral RNA-based life epoch called “The RNA World” (Gilbert 1986). This idea grants a central role to RNA in the origin and early evolution of life. Therefore, understanding those RNA properties that would drive the early evolution of RNA is undoubtedly a goal worth pursuing.

During the course of these experiments, were there any surprising results or particular difficulties that altered your thinking and subsequent focus?

The prevalence of the 2’OH in the establishment of single hydrogen bonding was a favorable surprise since that is one of the major differences between RNA and DNA. It surely increases our discussion and strengthened our hypothesis.

If you were able to give one piece of advice to your younger self, what would that be?

Work on your communication abilities. Producing robust results is not enough. You must communicate them in simple, jargon-free text and speeches.

Are there specific individuals or groups who have influenced your philosophy or approach to science?

The International Society for the Study of the Origin of Life (ISSOL) is a community that promotes the participation of people from all around the globe with different personal and scientific backgrounds; one that discusses, questions, and critiques ideas with passion and rigor, but always with an open mind.

Meet the First Author(s) is a new editorial feature within RNA, in which the first author(s) of research-based papers in each issue have the opportunity to introduce themselves and their work to readers of RNA and the RNA research community. Mario Rivas is the first author of this paper, “Nonstandard RNA/RNA interactions likely enhance folding and stability of segmented ribosomes.” Mario is a NASA Postdoctoral Fellow funded by the NASA Astrobiology Program. His current mentor is Dr. George E. Fox from the Center for the Origin of Life (COOL). He is working in the Fox laboratory in the Biology and Biochemistry Department of the University of Houston. His research interest focuses on the origin and early evolution of life from the perspective of molecular evolution.

What are the major results described in your paper and how do they impact this branch of the field?

Following the trend of ideas that we proposed in 2020 (Rivas and Fox 2020), we highlight long-range RNA/RNA interactions that favor the association and stability of segmented ribosomes. We speculate that since they are the result of intrinsic properties of RNA, it is perfectly plausible to think they have played an important role in the early evolution of RNA molecules.

White TC, Rudenko G, Borst P. 1986. Three small RNAs within the 10 kb trypanosome rRNA transcription unit are analogous to domain VII of other eukaryotic 28S rRNAs. Nucleic Acids Res 14: 9471–9489. doi:10.1093/nar/14.23.9471

Wilson DN, Doudna Cate JH. 2012. The structure and function of the eukaryotic ribosome. Cold Spring Harb Perspect Biol 4: a011536. doi:10.1101/cshperspect.a011536

Zhang X, Lai M, Chang W, Yu I, Ding K, Mrazek J, Ng HL, Yang OO, Maslov DA, Zhou ZH. 2016. Structures and stabilization of kinetoplastid-specific split rRNAs revealed by comparing leishmanial and human ribosomes. Nat Commun 7: 13223. doi:10.1038/ncomms13223

Meet the First Author

Mario Rivas