Nature manuscript 2012-10-12908C

High-resolution cryo-electron microscopy structure of the Trypanosoma brucei ribosome

Yaser Hashem[1,*], Amedee des Georges[1,*], Jie Fu[2], Sarah N. Buss[3], Fabrice Jossinet[4], Amy Jobe[5], Qin Zhang[6], Hstau Y. Liao[2], Robert A. Grassucci[1], Chandrajit Bajaj[6], Eric Westhof[4], Susan Madison-Antenucci[3], and Joachim Frank[1,2,5 *]

[*]*These authors contributed equally to this work.*

[1]*HHMI, Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY, USA*

[2]*Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY, USA*

[3]*Division of Infectious Diseases, Wadsworth Center, New York State Department of Health, Albany, NY, USA*

[4]*Architecture et Réactivité de l'ARN, Université de Strasbourg, Institut de Biologie Moléculaire et Cellulaire (CNRS), Strasbourg, France*

[5]*Department of Biological Sciences, Columbia University, New York, NY, USA*

[6]*Department of Computer Science, Institute for Computational Engineering and Sciences, University of Texas, Austin, TX, USA*

[*]*Corresponding author: Joachim Frank, Howard Hughes Medical Institute, Department of Biochemistry and Molecular Biophysics and Department of Biological Sciences, Columbia University, 650 W 168th Street BB2-221, New York, NY 10032, USA.*
Phone: +1 212 305 9510, Fax: +1 212 305 9500, E-mail: jf2192@columbia.edu

Supplementary Information

Map Segmentation

Before modeling the *T. brucei* ribosome the initial analysis of the map was made based on the X-ray structures of yeast ribosome[2] and *T. thermophila* small and large ribosomal subunits[5-6], which were rigid body-fitted into the *T. brucei* cryo-EM map. Based on these structures, a preliminary segmentation of the *T. brucei* map was performed using two different programs:

1- Segger, which exploits the Watershed method[20]. Segger is a module implemented in Chimera[44]. The segments generated by Segger were interpreted based on the above-mentioned X-ray structures.

2- VolRover[45] (also see protocols described in 46-48). The segmentation scheme is based on a multi-seed multi-domain fast-marching method (MDFMM). The method requires

some seed points in each identifiable domain for initiation. These seed points are generated from the fitted X-ray structures.

These preliminary segments were used during the atomic modeling process. After the atomic model was obtained (see RNA, ribosomal proteins modeling and Molecular dynamics flexible fitting below), we used our model to re-segment the *T. brucei* map more accurately as follows: For each chain, RNA or ribosomal protein of the full ribosome, except for the chain being segmented, a simulated cryo-EM map was generated using Chimera[44]. This simulated map was subtracted from the full experimental *T. brucei* cryo-EM map. The resulting difference densities correspond to the chain segment. All the ribosomal proteins displayed in fig. S2D were obtained by this method. Comparison between these segmented maps and simulated density maps of the same proteins, generated from the atomic model (in Chimera[44]) and filtered to 5.0 Å, shows similar features (fig. S2D) and supports the estimated resolution.

Ribosomal RNA Modeling

The rRNA modeling workflow consists on several steps. In a first step, LSU and SSU rRNAs of *T. brucei* have been structurally aligned against homologous sequences from previously solved ribosomal complexes. The 18S rRNA (SSU rRNA) was modeled based on the *T. thermophila* 40S crystal structures[5] because of the similar conformation of this structure to the *T. brucei* SSU, whereas in the yeast ribosome crystal structure[2], although it presents a higher resolution, the SSU appears to be in a different conformation (head is swiveled toward the intersubunit direction). The LSU rRNA was modeled based on the yeast ribosome crystal structure[2]. The structural alignments for LSU and SSU rRNAs have been done using the S2S tool[21]. For each ribosomal subunit, the reference solved-structure (3U5D for LSU and 3XZM for SSU) has been automatically annotated with S2S, producing an extended secondary structure (meaning a secondary structure supplemented with secondary and tertiary interactions described with the Leontis-Westhof classification[64]). This extended secondary structure has been used as a reference and as a structural mask to interactively align *T. brucei* and other related kinetoplastids rRNA sequences. These sequences have been fetched from the SILVA webserver (http://www.arb-silva.de/)[53] and the comparative RNA web site and project (CRW site)(URL: http://www.rna.icmb.utexas.edu/)[54]. The iterative construction of the structural alignments in S2S allowed us to identify the conserved-core structure along with all the expansion segments specific to *T. brucei*. Based on these structural alignments,

we have derived a 3D model for each rRNA core structure of *T. brucei* (the homology approach), followed by several rounds of geometric refinement in S2S.

For each ES, the sub-sequences for *T. brucei* and all other related kinetoplastids have been extracted from the structural alignments. A covariation analysis has been performed on each set of ES sequences, using the mlocarna algorithm[55]. This allowed deriving a consensus secondary structure that has been loaded into the Assemble tool[22] (fig. S9). These secondary structure models were manually modified in order to fit best the density map; especially when the local resolution allowed deriving the rRNA topology at the concerned regions. Based on a library of structural modules extracted from solved atomic structures, and on its ability to generate regular helices from secondary structures, Assemble allowed us to construct ES 3D models directly into the density maps, using an interactive construction process guided by the secondary structure template (the *ab initio* approach)(fig. S9). Single-stranded regions and numerous nucleotides were modeled manually in accordance with the density.

Modeling of Ribosomal Proteins

The workflow consists of five sequential steps adopted for all the ribosomal proteins. The first step is template selection, which is to select the best template(s) from the available homologous protein of known structures. We used Swiss-Model webserver (http://swissmodel.expasy.org/)[56-57] with the Template Identification tool. Based on the homology score and the resolution of the homologous proteins, several templates are selected (we selected three for each protein sequence, fig. S10). The second step of homology modeling is target-template multiple alignment which is to align the sequence of interest with the best three homologous template(s) of known structures. The alignment was performed using T-Coffee (http://tcoffee.crg.cat/)[65]. The third step is model construction, which builds model(s) of the structure of the protein of interest using the known structure(s) of homologous template(s), based on the multiple alignment realized in the previous step using the Alignment Mode tool in Swiss-Model webserver. The fourth step is the models assessment, which is to evaluate the protein structures built from the best three templates for each *T. brucei* ribosomal protein. We use the scores provided by Swiss-Model (fig. S10) and the MolProbity webserver (http://molprobity.biochem.duke.edu/)[59-60] to assess the structures of each model. The fifth and last step is the selection of the best model for each ribosomal protein (fig. S10). For each *T. brucei* ribosomal protein, only one of the three models was selected based on the homology score, the structural assessment score of each one of the three models and the cross-

correlation coefficient between the *T. brucei* ribosomal protein segment and each one of the three models (using Chimera[44]).

Missing residues from some of the *T. brucei* ribosomal protein models, mostly at their tails, as well as the trypanosome- and kinetoplatid-specific protein extensions were modeled *ab initio* wherever the structure of the missing fragments was present in the density map (fig. S10). We used the Phyre and the Phyre2 webservers[58] for the *ab initio* modeling of these missing fragments and extensions into the density map.

Tables S1 and S2 enumerate the *T. brucei* ribosomal proteins, in both nomenclatures used in yeast[2] and *T. thermophila* ribosomes[5-6] crystal structures, along with the templates used for the modeling of each protein.

Molecular Dynamics Flexible Fitting (MDFF)

The initial system was prepared for MDFF[61] using VMD[62]. MDFF is an MD simulation-based fitting procedure, which applies an extra potential to the system related to the gradient of the cryo-EM density map. The system consisted of the atomic model of the *T. brucei* ribosome including the SSU, the LSU and a tRNA at the E site. As the model was built into the EM map directly, no rigid-body fitting was required. The steric intermolecular clashes were fixed manually using PyMOL[66]. A short *in-vacuo* run of MDFF was performed on the initial in order to relax the structure. In NAMD[63], the system was minimized for 1000 steps, followed by MDFF. The run was stopped after 30ps. In order to achieve a better representation of the inter- and intra-molecular interactions, the system was then embedded in a solvent box of TIP3P water molecules, with an extra 12 Å padding in each direction, and neutralized by potassium ions, and an excess of ~ 0.2 M KCl was added. The system was re-minimized for 2000 steps in NAMD[63] followed by MDFF in explicit solvent. The run was stopped at 1.5 ns of simulation time. The simulated systems were prepared using CHARMM force field parameters (Combined CHARMM All-Hydrogen Topology File for CHARMM22 Proteins and CHARMM27 Lipids[67-68].

Rendering of Figures

All figures presenting a density map or an atomic model were rendered using Chimera[44].

References

64. Leontis, N.B., Westhof, E. Geometric nomenclature and classification of RNA base pairs. *RNA* **7**, 499-512 (2001).
65. Notredame, C., Higgins, D., Heringa, J. T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* **302**, 205-217 (2000).
66. The PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC.
67. Brooks, B.R. et al., CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comp. Chem.* **4**, 187-217 (1983).
68. A.D. MacKerell Jr. et al., CHARMM: The Energy Function and Its Parameterization with an Overview of the Program. *The Encyclopedia of Computational Chemistry*, **1**, 271-277, P.v.R. Schleyer *et al.*, editors (John Wiley & Sons: Chichester, 1998).