

Atomic resolution snapshot of eukaryotic 18s ribosomal RNA of Neglected Tropical Disease (NTDs) parasites as a therapeutic target.

Mwangi Harrison Ndung'u^{1*}, Edward Kirwa Muge¹, Peter Waiganjo Wagacha², Albert Ndakala³, Francis Jackim Mulaa¹

¹Department of Biochemistry, University of Nairobi, Nairobi, Kenya

²Department of Computing and Informatics, University of Nairobi, Nairobi, Kenya

³Department of Chemistry, University of Nairobi, Nairobi, Kenya

Abstract

Most of the tropical neglected diseases casing agents are single celled eukaryotic parasites that affect millions of global populations physiologically, economically and sociologically. These diseases not only affect humans but also affect the animals that sustain life within these populations completely influencing the societal fabric on sustenance. The current therapies are limited to poorly selected drugs that have a poor efficacy and extreme adverse effects. The current quick growth in sequencing and atomic structure determination of the parasites cellular organelles has accelerated the venture to obtain newer intervenes that would serve to benefit these populations. The rapid in silico determination of the ribosomal structure, specifically the 18S rRNA has revealed the mechanisms by which some therapeutics induces their effect on the process of translation to the parasites. We show some in silico atomic resolution structures of 18S rRNA of neglected tropical diseases. We further show strategies and scenarios that are exploited to interfere with translation, thus highlighting the 18S rRNA as a primary drug target. The results also highlight unique as well as conserved elements in binding pocket that can serve as hotspots for novel therapeutics development.

Keywords: Neglected Tropical Diseases (NTDs), 18S rRNA, Ribosome, Insilico structure determination, Translation.

Accepted on August 12, 2020

Introduction

This research work performed as a collaborative project with an eventual goal of developing newer drugs that offer treatment of parasitic infections with lesser side effects, less susceptibility to the mechanism of resistance [1,2]. The novelty of this work is that it brings to bear the state of art RNA and compound modeling methodologies some of which was developed in the laboratory, to yield high-resolution models of biological molecules and macromolecules. In our case, the smaller subunit ribosomal RNA of the selected parasite was modeled to enable performs virtual screening for new compounds or existing drugs that would allow therapeutic switching. Regions of 18S RNA that have demonstrated the functional significance and highest conservation within the ribosome were identified in this study as targets [3,4].

The essential nucleotide motifs were selected based on a genetic *in vivo* system from high throughput identification [5-8]. From various studies conducted earlier, there are some regions that have been identified to be essential for the facilitation of protein synthesis from the small subunit [9-11]. This includes a region that binds the initiation factor, the tRNA and ribosomal proteins located on helix 23 and helix 24 of on *E.coli* which is essential to protein synthesis [12-14]. There are several structural differences between human and parasitic 18S rRNA that could be exploited to tailor specific anti-infectives that only target the rRNA of this pathogenic parasites that are classified as a major class of tropical neglected diseases.

To realize this goal we split this into two parts one: The determination of the three-dimensional structure of the selected parasites and identification of the important motif which lies adjacent to helix 23 and 24 of the *E.coli* and other 2 motifs. Importantly as shown from various studies human helix 24 has a base modification (Pseudouridine) which is lacking in the prokaryotic homolog [15-17] it is therefore important to see how it appears for the selected parasites. Determining the structural implication of this particular modified residue toward protein synthesis of these eukaryotes by solving the structures of the chosen parasites hence is important. The next part is performing virtual screening for selected lead compounds by performing lead compound-target recognition.

These studies will yield results which will be used to refine further the design strategy of the compound libraries. Identifying interesting structural domains from the 18S RNA of the chosen parasites, such as helix 23 which forms a domain known as the kink-turn motif [18] is important. Studies have shown that this motif is highly dynamic, hence an excellent target for drugs reason being stabilization of one conformer by a target affects the process of protein synthesis [19].

In addition to this no structure that has been solved completely for an isolated kink-turn [20], therefore characterizing the RNA dynamics and ligand recognition in silico and *in vitro* offers some difficulty but this could be helped by this study. In silico structure prediction was used to determine the RNA dynamic

structures for the parasites. These was followed by identification of interesting motifs sites for drug targets, which were virtually screened with compound databases, and finally this motif-compound binding was validated through the *in vitro* and finally *in vivo* techniques [21].

Introduction to the ribosome

The ribosome is a universal organelle that performs the function of protein synthesis. All parasites have ribosomes that do this process of translation and they are highly conserved due to their functional aspect. Ribosomes are composed of primarily a Ribonucleic Acid (RNA) and several proteins complexed together to achieve a functional unit. These components are either indirectly or directly involved in the process of peptide synthesis. Nucleotide monomers combine to form a polymer known as RNA while amino acid monomers combine to form polymers known as proteins. Due to the diverse functions the RNA molecules play, these molecules are central to the biology of the cell. These functions, for example, may include a host of catalytic roles and various aspects of involvement in gene expression. RNA can adopt different three-dimensional structures to concur with their functional diversity while they interact with other molecules such as protein, cofactors or with other RNAs. It has been observed that there is high sequence conservation in many relevant functional motifs, and this evolutionary conservation points to the important roles that these RNAs have performed throughout time. The desire to unravel underlying mechanisms the various roles that RNA(s) play and the RNA functions has sparked immense interest in structural biology. This is especially due to the surge in the availability of complete genomes sequences. The basic primary methods of elucidating the three-dimensional structure of biological molecules and macromolecules are X-ray crystallography and Nuclear Magnetic Resonance Spectroscopy though there are some concerted efforts from the cryo-electron microscopy. NMR for isolated RNA domains still remains the method of choice for solution structure determination. However due to the size limitation of NMR towards solving larger RNA newer ways have to be invented. Newer theoretical methods have been developed that have worked to help minimize the time it takes to obtain the three-dimensional structure of more complex RNA [14,21]. These include both homology and *de novo* modeling. The work presented in this study is based on the theoretical methods taking to account highly conserved and functionally relevant ribosomal motifs from the small subunit of the selected parasites.

Materials and Methods

The atomic level structural crystalline coordinates used in this study are listed from along with details of their construction [5,19]. The sequences used for modeling was conducted using through blasting in the gene bank (NCBI). Further process of verification was required and led us to specialized groups for verification of the 18S rRNA sequences (CRW) site that has a database that shows the completeness of sequences among other analyzed and verified annotation.

Selection of template structures for the parasites rRNA was a rigorous exercise that is described previously. The approach for identification of parasites ribosomal RNA drug targets was to utilize the RNA homology modeling software to develop a high throughput method to identify ribosome RNA motif-small molecule partners, or “paired” motifs. Namely three-Dimensional Homology Model Screening (3DHMS) described earlier [4,19]. Analysis of sensitivity and assays of parasites RNA motif were conducted as described previously. Refinement, evaluation of 18S rRNA, energy minimization on 18S rRNA, and measurements and prediction of binding to 18S rRNA in the reconstituted structure were carried out as reported previously, with the minor modifications [5].

Results and Discussion

Determining the first structure of the *P. falciparum*, *M. tuberculosis*, Trypanosome and Leishmania 18S rRNA has led to better understanding of the structural basis for its protein-synthesizing roles in these cells. This will enable research in the field of drug development run in silico ligand screening experiments using the solved 18S rRNA structure as a target against a library of potential anti-kinetoplastid compounds. As shown in Figure 1-Figure 3 the various 18S rRNA are highly conserved but have a substantial difference on the expansion segments that lender the process of protein synthesis different from one to the other. The tree dimensional structure further offer differing conformations that could be exploited for potential drug screening and repurposing for more effectiveness towards combating the diseases they cause.

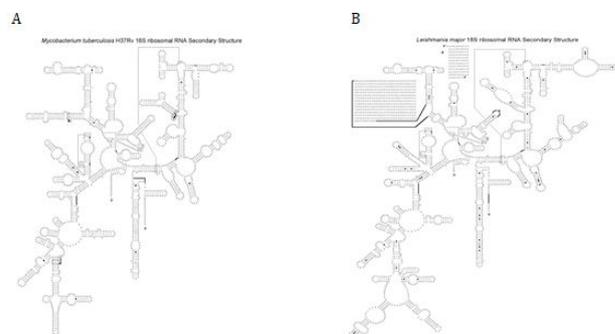


Figure 1. (A) Showing the Secondary structure of *M. tuberculosis* and (B) *Leishmania major* 18S rRNA key to note are the differences in the expansion segments that dictates the process of protein synthesis.

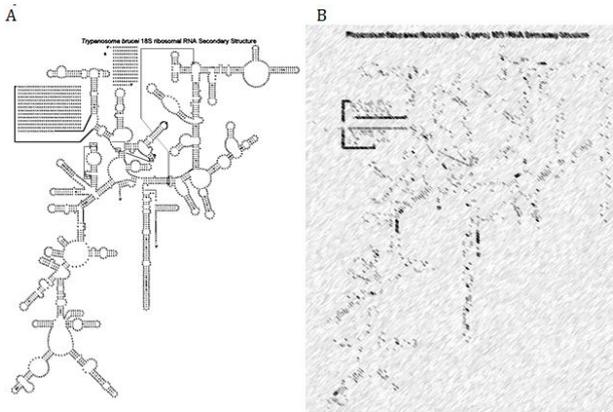


Figure 2. (A) Showing the Secondary structure of *T. brucei* and (B) *P. falciparum* 18S rRNA key to note are the differences in the expansion segments that dictates the process of protein synthesis.

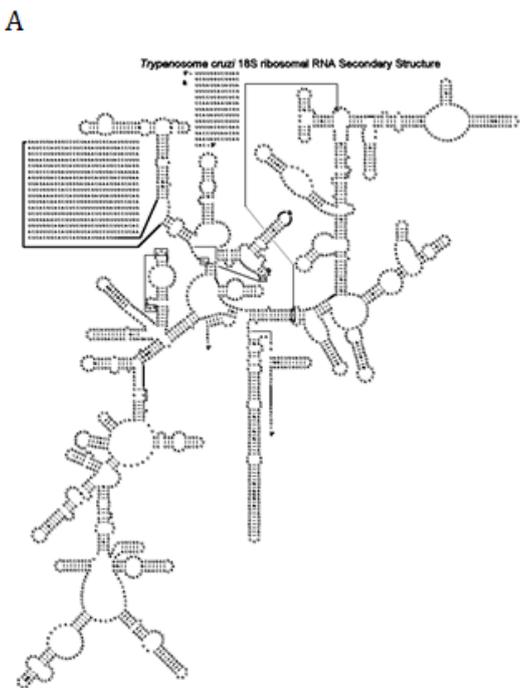


Figure 3. Showing the secondary structure of *T. cruzi*.

The use of structure prediction and modeling technologies in this study will dramatically reduce the time it takes from target identification to drug lead determination. Additionally, millions of compounds that can be tested in silico against the generated structure would have been impossible to test experimentally.

Innovation and output

The approaches used in this study to validate and screen rRNA targets are highly innovative. They are superior to traditional screening approaches because of the likelihood of success in finding translation inhibitors that are highly efficacious, but at the same time unlikely to select for target mutations, is very high. Its innovation is based on conserved ribosomal sequences. The exploitation of these sequences could lead to new anti-protozoan drugs refractory to resistance. Drug

resistance significant is more likely the highest drug-related priority to date. The production of new anti-protozoan compounds refractory to drug resistance would be a major advance in the repertoire of NTDs. Directing this approach towards other organisms including eukaryotic infective agents could make a significant positive worldwide health impact.

This is an application exploiting the conserved regions of ribosomal rRNA that form functional motifs or loops to develop next-generation anti *P. falciparum*, *M. tuberculosis*, trypanosome and leishmania 18S candidate compounds that can be further developed to potent anti trypanosoma and leishmania to combat kinetoplastids family against many protozoans that cause moderate to severe diseases and is applicable to a broad spectrum of apicomplexan protozoan’s diseases. The application is unique as it proposes to use high-resolution macromolecular models that will lead to the identification of superb modeling systems to the identification of conserved functional motifs of kinetoplastids ribosomal RNA that for functional reasons cannot be altered and therefore good targets for the screening of lead anti-protozoan compounds (Tables 1 and 2).

	Orga nism(2)	L(3)	RT(4)	RC	Size	Cmp	Acc	Com mon name	Phyl ogen y (M)
1	<i>Plas modi um falcip arum</i>	N	R	16S	2090	100	M191 72	Malar ia paras ite	Cellul ar organ isms
2	<i>Mycobact erium tuber culosi s</i>	N	R	16S	1537	100	Z838 62		Cellul ar organ isms
3	<i>Trym anom abruc ei</i>	N	R	16S	2251	100	M126 76	Kinet oplast ids	Cellul ar organ isms
4	<i>Trypa noso macr uzi</i>	N	R	16S	2315	100	AF24 5382	Kinet oplast ids	Cellul ar organ isms
5	<i>Leish mani a major</i>	N	R	16S	2203	100	AC00 5806	Kinet oplast ids	Cellul ar organ isms

Table 1. Five selected parasites *P. falciparum*, *M. tuberculosis*, *T. brucei*, *T. cruzi*, and leishmania major sequences selected for this study. L (3) refers to the cell location, which is the nucleus, RT (4) refers to RNA type R:Ribosomal RNA (rRNA); RC refers to the RNA Class 16S, Nucleotide size, Cmp means % Complete, Acc means gene bank accession number, common name and the phylogenetic classification.

Species	18SrRNA.std.energy	18SrRNA.opt.energy
<i>Plasmodium falciparum</i>	2,397,780.107	-46,366.56367
<i>Mycobacterium tuberculosis</i>	199459.580	-37,57278

<i>Trypanomabrucei</i>	2118859.871	-79877.95586
<i>Trypanosomacruzi</i>	7455258.998	-91558.00067
<i>Leishmania major</i>	895477.8943	-96921.59635

Table 2. Total 18S rRNA structure energy optimization table obtained from results of RNA 123, which helps minimize the energy from a large positive figure to a more acceptable negative figure that is biologically functional.

Conclusion

The current quick growth in sequencing and atomic structure determination of the parasites cellular organelles has accelerated the venture to obtain newer intervenes that would serve to benefit these populations. The rapid in silico determination of the ribosomal structure, specifically the 18S rRNA has revealed the mechanisms by which some therapeutics induces their effect on the process of translation to the parasites. We show some in silico atomic resolution structures of 18S rRNA of neglected tropical diseases. We further show strategies and scenarios that are exploited to interfere with translation, thus highlighting the 18S rRNA as a primary drug target. The results also highlight unique as well as conserved elements in binding pocket that can serve as hotspots for novel therapeutics development.

Acknowledgments

We thank Dr. Fredrick Sijenyi and Dr. Solomon Deresse for insights and comparative reviews on theoretical and experimental techniques to obtain structures, and advice Dr. Khan Nelson for helpful discussion regarding molecular dynamics, Prof. Francis Mulaa for valuable advice, and Prof. Dimbuson Wallace Bulimo, Miriam Watetu Muratiri and Dr. Daniel Njau for critical comments. This work was supported by the Prof Mulaa research program of Nairobi university and by HEP bioinformatics consultants LTD.

Conflict of Interest

The authors declare no conflict of interest.

References

1. Knight S, Bisaria N, Yesselman JD, et al. High-throughput investigation of diverse junction elements in RNA tertiary folding. *J cell*. 2018;174(2):377-390.
2. K Lee, S Varma, J SantaLucia Jr, et al. *In vivo determination of RNA structure-function relationships: Analysis of the 790 loop in ribosomal RNA*. *J Mol Biol*. 1997;269(5):732-743.
3. Martina Z, Hargrove AE. Small molecule targeting of biologically relevant RNA tertiary and quaternary structures. *Cell Chem Biol*. 2021;28(5):594-609.
4. Sijenyi F, Saro P, Ouyang Z, et al. The RNA folding problems: different levels of sRNA structure prediction. 2012;27:91-117.
5. Mwangi HN, Muge ER, Wagacha PW, et al. Methods for identifying microbial natural product compounds that target

- kinetoplastid RNA structural motifs by homology and de novo modeled 18S rRNA. *Int J Mol Sci*. 2021;22(9):4493.
6. Schluenzen F, Tocilj A, Zarivach R, et al. Structure of functionally activated small ribosomal subunit at 3.3 Å resolution. *Cell*. 2000;102(5):615-623.
7. Stark H, Orlova EV, Appel JR, et al. Arrangement of tRNAs in pre-and post translocational ribosomes revealed by electron cryomicroscopy. *Cell*. 1997;88(1):19-28.
8. Brown DK, Penkler DL, Amamuddy OS, et al. MD-TASK: A software suite for analyzing molecular dynamics trajectories. *NLM*. 2017;33(17):2768-2771.
9. Charette M, Gray MW. Pseudouridine in RNA: what, where, how, and why. *IUBMB Life*. 2000;49(5):341-351.
10. Ofengand J, Bakin A. Mapping to nucleotide resolution of pseudouridine residues in large subunit ribosomal RNAs from representative eukaryotes, prokaryotes, archaeobacteria, mitochondria and chloroplasts. *J Mol Biol*. 1997;266(2):246-68.
11. Huang L, Lilley DMJ. The kink-turn in the structural biology of RNA. *Q Rev Biophys*. 2018;51.
12. Klein DJ, Schmeing TM, Moore PB, et al. The kink-turn: a new RNA secondary structure motif. *EMBO J*. 2001;20(15):4214-4221.
13. Matsumura S, Ikawa Y, Inoue T. Biochemical characterization of the kink-turn RNA motif. *Nucleic Acids Res*. 2003;31(19):5544-5551.
14. Shi X, Huang L, Lilley DMJ, et al. The solution structural ensembles of RNA kink-turn motifs and their protein complexes. *Nat Chem Biol*. 2016;12:146-152.
15. Mowbray CT, Holter MC, Teague GB et al. Fidelity criteria: Development, measurement and validation. *Am J Eval*. 2003;24(3):315-340.
16. Klein DJ, Schmeing TM, Moore PB et al. The kink-turn: a new RNA secondary structure motif. *EMBO J*. 2001;20(15):4214-4221.
17. Tolbert BS, Miyazaki Y, Barton S, et al. Major groove width variations in RNA structures determined by NMR and impact of 13 C residual chemical shift anisotropy and 1 H-13 C residual dipolar coupling on refinement. *J Biomol NMR*. 2010;47(3):205-219.
18. Mwangi HN, Wagacha P, Mathenge P et al. Structure of the 40S ribosomal subunit of *Plasmodium falciparum* by homology and de novo modelling. *Acta Pharm Sin B*. 2017;7(5):97-105.
19. NCBI resource coordinators. Database resources of the national center for biotechnology information. *Nucleic Acids Res*. 2016;44(D1):D7-D19.
20. Cannone JJ, Subramanian S, Schnare MN. The Comparative RNA Web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinform RNAs*. 2002.
21. Gutell RR, Lee JC, Cannone, JJ. The accuracy of ribosomal RNA comparative structure models. *Curr Opin Struct Biol*. 2002;12(3):301-310.

***Corresponding to:**

Mwangi Harrison Ndung'u

Department of Biochemistry

University of Nairobi

Nairobi

Kenya

E-mail:harryndungu@gmail.com