

Functional analysis of hypothetical proteins unveils putative metabolic pathways, essential genes and Therapeutic drug and vaccine target in *Trypanosoma cruzi*: A Bioinformatics Based Approach

Rodrigo Profeta Silveira Santos¹, Priya Trivedi², Neha Jain², Sandeep Tiwari³, Syed Babar Jamal Bacha⁴, Arun Kumar Jaiswal⁵, Thiago Luiz de Paula Castro⁶, Núbia Seiffert⁶, Siomar de Castro Soares⁷, Artur Silva⁸, Vasco A de C Azevedo¹

1 UFMG

2 DEVI AHILYA UNIVERSITY

3 INSTITUTE OF BIOLOGICAL SCIENCE, UFMG

4 1. INSTITUTE OF BIOLOGICAL SCIENCE, UFMG

5 INSTITUTE OF BIOLOGICAL SCIENCE, UFMG; DEPARTMENT OF IMMUNOLOGY, MICROBIOLOGY AND PARASITOLOGY, INSTITUTE OF BIOLOGICAL SCIENCES AND NATURAL SCIENCES, UFTM

6 UFBA

7 DEPARTMENT OF IMMUNOLOGY, MICROBIOLOGY AND PARASITOLOGY, INSTITUTE OF BIOLOGICAL SCIENCES AND NATURAL SCIENCES, UFTM

8 UFPA

Abstract

The protozoan *Trypanosoma cruzi* is the etiological agent of Chagas disease, a major chronic, systemic, parasitic infection. The disease affects about 8 million people in Latin America, of whom 30-40% either has or will develop cardiomyopathy, digestive mega syndromes, or both. Currently, there are neither effective drugs nor vaccines for the treatment or prevention of the disease. The current synthetic drugs such as nifurtimox (a nitrofur derivative) and benznidazole (a nitroimidazole derivative), are associated to severe side effects, including cardiac and/or renal toxicity and as well not effective, which accounts for the need to search new effective chemotherapeutic and chemo prophylactic agents against *T. cruzi*. Therefore, due to their low efficacies and the resistance developed by the bug to these medications, there is an urgent need to consider newer species-specific targets. Approximately 50% of the predicted protein-coding genes of the *Trypanosoma cruzi* CL Brener strain are annotated as hypothetical or conserved hypothetical proteins. Here in this work, we have attempted to assign probable functions to these hypothetical sequences present in this parasite, to explore their plausible roles as druggable receptors. Thus, putative functions have been defined to 491 hypothetical proteins, which exhibited a GO term correlation and PFAM domain coverage of more than 50% over the query sequence length. We tried to find out if our 491 sequences were showing any similarity

with the already known essential genes by using DEG BLAST, and we were able to predict 114 sequences as essential genes for the protozoan. Pathway analysis was also performed to check if our hypothetical proteins can participate in any metabolic pathway, in which we have got 31 sequences were associated with various metabolic pathways for essential genes while 40 sequences were associated with metabolic pathways for non-essential genes. Furthermore, the localization was predicted to find drug targets and vaccine targets. And Protein-protein Interaction was analyzed using STRING database. We have also tried to find out the orthologous proteins by using Cluster of Orthologous genes (COG's) STRING database. Druggability was also checked by searching against the drug target dataset of DRUGBANK target dataset of DRUGBANK. We got 3 experimentally know drug targets and 07 novel drug targets from this uncharacterized (hypothetical) dataset for *Trypanosoma cruzi*. Data generated by this study might facilitate swift identification of potential therapeutic targets and thereby enabling the search for new inhibitors or vaccines.

Funding: TWAS-CNPq, CNPq, and CAPES