A comparative *in silico* linear B-cell epitope prediction for South American and African *Trypanosoma vivax* strains

Rafael Lucas Muniz Guedes[1], Carla Monadeli Filgueira Rodrigues[2], Luiz Gonzaga Paula de Almeida[1], Paola Minoprio[3], Marta M.G. Teixeira[2], Ana Tereza Ribeiro de Vasconcelos[1]

[1]*Laboratório Nacional de Computação Científica (LNCC), Av. Getúlio Vargas, 333, Petrópolis, RJ, Brazil.*

[2]*Department of Parasitology, University of São Paulo, Av. Prof. Lineu Prestes 1374, São Paulo, SP, Brazil .*

[3]*Department of Infection and Epidemiology, Institut Pasteur, Paris, France*

Single-celled parasitic protists from the Kinetoplastida order are the etiological agents of trypanosomiasis, an important neglected disease that affects humans, domestic and wild animals world-wide. Animal tripanosomiasis, also known as *Nagana*, are caused by different species, i.e. *Trypanosoma vivax*, *Trypanosoma congolense* and *Trypanosoma brucei brucei,* being the former the most prevalent in livestocks in West Africa. *T. vivax* is disseminated across Africa and South America, infecting diverse mammals like cattle, sheep and goats leading to a profound economical impact in agriculture. Linear B-cell epitopes are predictable specific peptides recognized by host's antibodies triggering immune response. Due to the difficulty of maintaining *T. vivax* under laboratory conditions, little is known about its immunogenicity. Here we have combined several bioinformatic tools in order to select the best *in silico* linear B-cell epitope candidates for improving serodiagnosis and serotyping. A representative dataset of *T. vivax* strains was prepared with transcriptomic data from bloodstream forms from the Western African (Til: *T. vivax* IL1392, from Kenya) and two South American (Tsp: *T. vivax* Lins, from São Paulo state, Brazil, sequenced in the present work and Tvv: *T. vivax* LIEM-176, from Venezuela) isolates. The annotated genome of another Western African strain (Tvi: *T. vivax* Y486, from Nigeria) was also included. Transcriptomes were assembled with Trinity and proteins predicted with Transdecoder. Gene expression was estimated with FPKM mapping reads to assembled sequences, using Bowtie2 and HTSeq-count. An in-house pipeline (SignalP, TargetP, PredGPI, ProtComp, WoLFPSort and PROSITE) was used to select predicted cell-membrane and secreted proteins. The selected sequences were screened for the presence of linear B-cell epitopes with BepiPred, LBTope and IEDB tools and also scanned for intrinsically unstructured/disordered regions with IUPred. Possible cross-reactivity with other trypanosomatids was filtered with tBLASTn (>= 70% identity over 15mer). A total of 23, 24, 3 and 68 epitopes at 18, 19, 3 and 57 proteins were selected for Til, Tsp, Tvv and Tvi, respectively. Most of the identified epitopes are present in proteins annotated as hypothetical proteins, with variable FPKM values (from only 3.4 to 611.1). Selected epitopes metrics for BepiPred, LBTope, IEDB and IUpred tools presented similar results for those observed in a control set for experimentally validated epitopes from *T. cruzi.* Epitopes clustering revealed the presence of common and exclusive sequences between the four strains. These are *in-silico* candidates with better probabilities of positive results for future experimental specie-specific serotyping and serodiagnostic tests for *T. vivax*.

Funding support:

CAPES, FAPESP, Campus France