



Indicate the format in which you wish to present your work: Poster Oral Presentation

TITLE

ADVANCING DRUG DISCOVERY FOR SCHISTOSSOMA MANSONI WITH SIAMESE NETWORKS AND ATTENTION-BASED DEEP LEARNING MODELS

AUTHORS

Sanches, I.H.^{*1,2}; Alves, K.M.^{1,2}; Teixeira, T.R.³; Moraes, J.^{3,4}; Emery, F.S.²; Andadre, C.H.^{1,2}

AFFILIATIONS

¹ Laboratory for Molecular Modeling and Drug Design, Faculty of Pharmacy, UFG, Goiânia, GO, Brazil.

² Center for the Research and Advancement in Fragments and Molecular Targets (CRAFT), School of Pharmaceutical Sciences at Ribeirão Preto, Ribeirão Preto, SP, Brazil.

³ Research Center on Neglected Diseases, Guarulhos University, Guarulhos, SP, Brazil.

⁴ Research Center on Neglected Diseases, Scientific and Technological Institute, SP, Brazil.

ABSTRACT

Schistosoma mansoni is a parasitic worm responsible for schistosomiasis, a neglected tropical disease impacting millions in regions with inadequate sanitation. Infection occurs through contaminated water, resulting in symptoms like fever, abdominal pain, diarrhea, and serious liver damage. The primary treatment, praziquantel, is ineffective against young schistosomes and newly transformed schistosomula (NTS), and rising drug resistance underscores the critical need for new treatment options.

Our goal was to train a deep learning model to identify drug candidates against both adult *S. mansoni* and NTS. We collected compounds with phenotypic *in vitro* IC₅₀ data from the literature, removed duplicates, standardized the structures, and a randomly stratified split of the data into training and test sets were performed, where 80% of the data were dedicated to training and 20% to testing. Using a Siamese network architecture, we trained the model to recognize structural similarities between compounds. To enhance performance, we incorporated ChemBERTa, a pre-trained transformer model from ChEMBL, as the initial layer to leverage its understanding of chemical structures. Additionally, we implemented a multi-head attention mechanism to capture a wide range of molecular interactions and fine-tuned the model using Optuna to achieve better performance.

Two models were developed—one for adult *S. mansoni* worms and another for NTS—and validated using an independent external dataset of 263 compounds provided by Prof. de Morais' group. The adult worm model achieved an accuracy of 0.85, a Matthew's correlation coefficient (MCC) of 0.73, a sensitivity of 0.83, and a specificity of 0.86. In comparison, the NTS model attained an accuracy of 0.82, an MCC of 0.71, a sensitivity of 0.81, and a specificity of 0.84. These strong results are impressive considering the validation dataset included chemically distinct compounds, as revealed by t-SNE analysis indicating that these compounds were located in different regions of chemical space. Despite this challenge, both models demonstrated robust performance and generalizability. Additionally, Monte Carlo uncertainty was calculated for all predictions, providing an extra layer of confidence in the model's outputs.

In conclusion, the Siamese AI models developed for adult worms and NTS demonstrate robustness and the ability to generalize to unseen and chemically diverse data. Moving forward, we will conduct an interpretative analysis of the model predictions utilizing techniques such as SHapley Additive exPlanations (SHAP) and uncertainty estimation. Subsequently, we will perform virtual screening of CRAFT's and commercial databases using these models, incorporating uncertainty values to prioritize the most promising candidates. The top compounds identified will be subjected to phenotypic *in vitro* testing, thereby contributing to accelerating the discovery of new candidates for schistosomiasis.

KEYWORDS

Schistosomiasis; Drug Discovery; Siamese Networks; Phenotypic Screening; Uncertainty Estimation.

FINANCIAL SUPPORT

CNPq, CAPES, FAPESP and FAPEG.