Artificial intelligence aided discovery of new classes of antiprotozoal agents for Chagas disease

Benjamin Perry[1], Mostafa H. Ahmed[2], Kong T. Nguyen[2]

[1.] Drugs for Neglected Diseases initiative, Geneva, Switzerland. [2.] Atomwise Inc, San Francisco, CA, United States.

Abstract

Chagas disease is caused by the protozoan parasite ***Trypanosoma cruzi*** (***T. cruzi***), which can result in a lifelong infection and can also be life threatening. Unfortunately, currently available therapeutics require prolonged treatment, are often poorly tolerated and have inconsistent efficacy in the chronic phase of the disease. Hence, it is imperative to discover novel classes of therapeutics for this disease. PEX14, an essential protein for glycosomal biogenesis and protein targeting, has emerged as a validated target for antiprotozoal therapy. The inhibition of the PEX14-PEX5 protein-protein interaction has been shown to disrupt glycosomal import leading to ATP depletion, glucose toxicity, and metabolic collapse resulting in trypanosomal death. We have successfully discovered novel classes of protein-protein interaction inhibitors targeting PEX14-PEX5 using the AtomNet® model, a machine learning model trained to predict protein-ligand binding, that display excellent antiprotozoal activities with low toxicity profiles.

Hypothesis

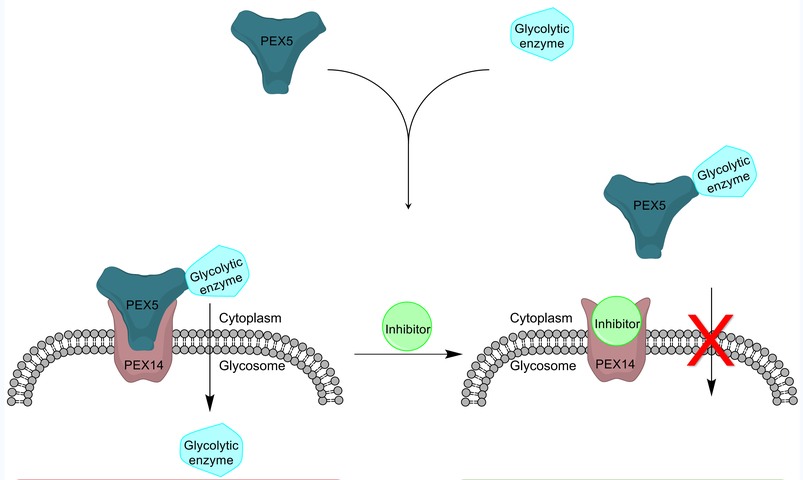

Figure 1. Trypanosomal glycolytic enzymes are transported from the cytoplasm into glycosomes by PEX14 (membrane-bound protein) and PEX5 (transporter protein). An inhibitor of the PEX14/PEX5 interaction prevents the association of cargo-bound PEX5 to PEX14, thereby blocking the transport of glycolytic enzymes into glycosomes and ultimately resulting in trypanosomal death.

References:

- Wallach, I, ***et al.*** Atomnet: A deep convolutional neural network for bioactivity prediction in structure-based drug discovery. arXiv 2015, arXiv:1510.02855.
- Dawidowski, M.; ***et al.*** Inhibitors of PEX14 disrupt protein import into glycosomes and kill Trypanosoma parasites. Science 2017, 10.1126/science.aal1807

Methods

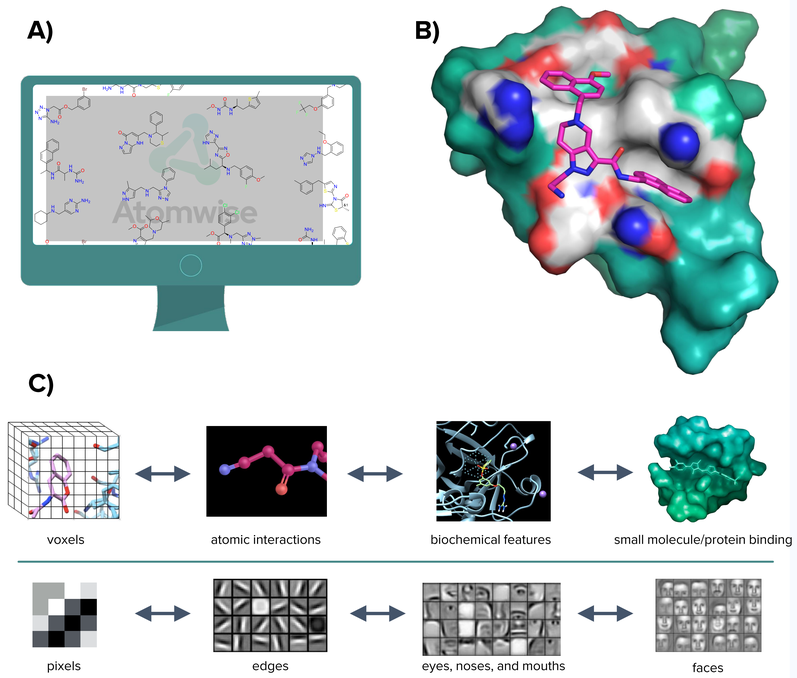

Figure 2. A) Commercially available compound database (7.2m). B) The X-ray crystal structure of the ***T. brucei*** PEX14 co-crystallized with a small molecule inhibitor (the template used for comparative structure modelling; PDB ID: 5N8V). C) The AtomNet® model, which is a deep convolutional neural network. Facial recognition models can learn to identify parts such as noses, ears, and eyes by combining the edges. Finally, the model can learn to recognize faces by combining those parts. Similarly, the AtomNet® model has learned to recognize essential chemical groups as well as fundamental concepts in organic chemistry and can thus provide predictions by combining all of this information.

Here we describe a novel approach for hit identification of novel antiprotozoal agents for Chagas disease.

1. Considering a lack of publicly available structures of ***T. cuzi*** PEX14 protein, we generated a comparative structure model based on the ***Trypanosoma brucei*** (***T. brucei***) PEX14 (PDB ID: 5N8V; Figure 2), which shares a 44% sequence identity with ***Tc***PEX14.
2. We then used the AtomNet® model, which is a deep convolutional neural network, to screen a library of several million readily available commercial compounds.
3. The high-confidence predicted binders were tested ***in vitro*** to prospectively discover novel classes of drug-like small molecule antiprotozoal agents for Chagas disease.

Results

The global AtomNet® model was used to score a library of 7,177,223 molecules on an elastic hybrid cluster consisting of GPU and CPU instances.

The filtering process was a standardized procedure that did not involve a project-specific human bias.

- The top-scoring 20,000 compounds were filtered for molecular weight of >250 Da and then filtered using proprietary set of SMARTS patterns containing moieties generally deemed undesirable.
- The remaining set was clustered for scaffold diversity using chemical fingerprints. Subsequently, drug-like molecular property filters were applied.
- Of the filtered set, 50 predicted high-affinity binder cluster representatives were submitted for purchase, were soluble in DMSO, and also passed an internal quality control procedure. These compounds were made available for experimental validation.
- Compounds were experimentally assayed for their trypanocidal activity as well as host cell toxicity using human lung fibroblasts (MRC5) cells.
- Five compounds exhibited trypanocidal activities with IC_{50} values in the low micromolar range as well as acceptable host cell toxicity profiles (CC_{50} values; Table 1).

Table 1. Experimental results of the artificial intelligence aided discovery of antiprotozoal agents for Chagas disease

Molecule ID	IC_{50} (µM)	CC_{50} (µM)	MW (Da)	Log P	HBD	HBA	RB	PSA (Å)
1	7.95	> 64	463.58	2.57	1	8	8	81.51
2	4.54	20.68	419.51	3.42	1	8	4	117.07
3	7.37	47.55	448.56	3.59	1	8	7	85.17
4	3.66	> 64	399.47	2.16	1	8	5	117.71
5	7.13	36	389.5	2.82	2	6	4	75.6
Lipinski values	**-**	**-**	**≤ 500**	**≤ 5**	**≤ 5**	**≤ 10**	**≤ 10**	**≤ 140**

Conclusion

The PEX14-PEX5 protein-protein interaction was targeted using the artificial intelligence based AtomNet® model for the discovery of new antiprotozoal agents. Five compounds, representing novel chemical matter, were identified and are being pursued as leads for the treatment of Chagas disease.