MOL2NET, International Conference Series on Multidisciplinary Sciences

Virtual Screening of 1,2-Diazoles Compounds for Chagas Disease: A Prediction Model

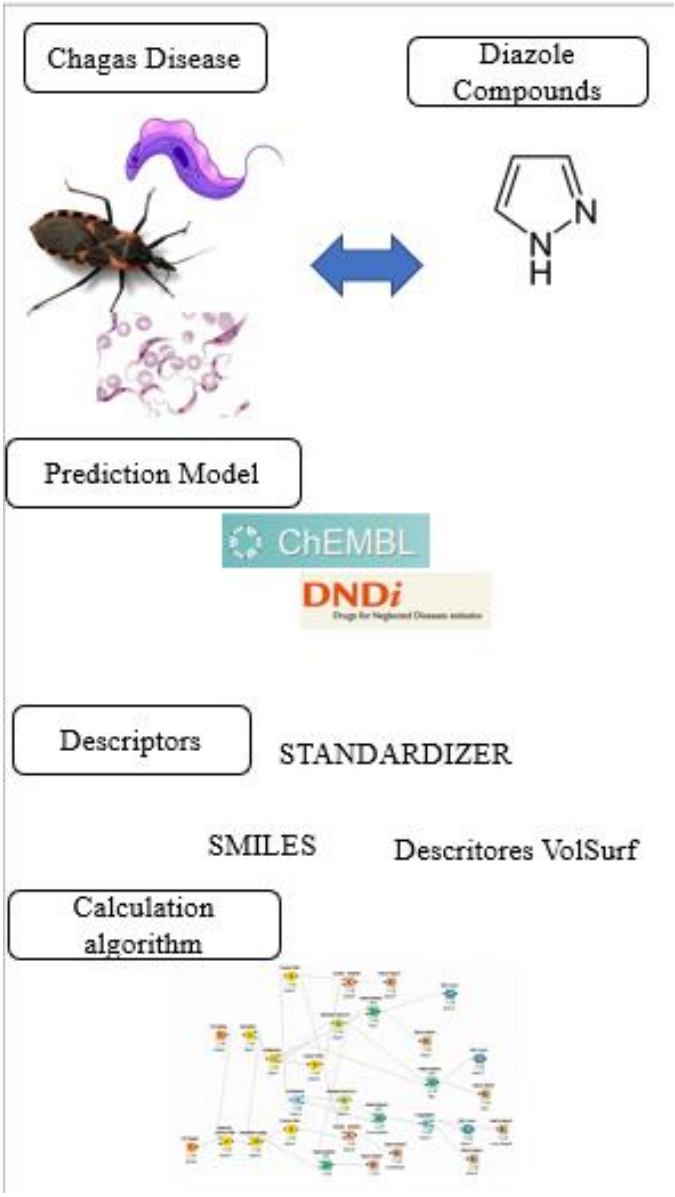
SOUSA, N.F[a]; BARROS, R.P.C[a]; LUÍS, J.A.S[ab]; SCOTTI, L.[a]; SCOTTI, M.T.[a]

[a]Post-Graduate Program in Natural and Synthetic Bioactive Products, Federal University of Paraíba, 58051-900 João Pessoa, PB, Brazil;

[b]Center of Education and Health, Federal University of Campina Grande. 58175-000 Cuité, PB, Brazil.

e-mail: nataliafsousa@ltf.ufpb.br

Graphical Abstract



Abstract.

Introduction: Chagas' disease is a parasitic, chronic and emergency infection caused by the protozoan Trypanosoma cruzi[1]. Because of this, we are constantly seeking new therapeutic alternatives and methods of study. **Objectives:** To perform a virtual screening of 1,2-diazoles compounds as potential therapeutic agents for Chagas' disease through the elaboration of a predictive model. **Methods:** A CHEMBL database[2], composed of 661 chemical structures with activity potential against *Trypanosoma cruzi*, was selected. The prediction set is composed of 31 unpublished 1,2-diazole compounds. The SMILES codes were the input data for all structures. The model was generated by *KNIME* 3.1.0 software, using the Random Forest (RF) calculation algorithm. **Results and Discussion:** In the analysis of the model, the hit rates obtained in the test and cross-validation were higher than 73%. The graph Receiver Operating Characteristic, corresponding in the test set to 0.843, indicating a high classification rate. The Matthews Correlation Coefficient, resulting in 0.63 in the test and 0.61 in the cross validation, indicating that the model has a good prediction. The RF model demonstrated that 13 molecules studied showed a percentage of potential activity above 65%. **Conclusion:** The model presented accuracy, reproducibility and distinguished the probability of potential activity of the molecules under study.

Introduction

Chagas' disease is a parasitic infection caused by the protozoan hemoflagellate Trypanosoma cruzi, transmitted by insect bite vectors, as well as by blood transfusion and congenital transmission[1]. This disease is classified as a Neglected Disease, because it is endemic in low-income populations and is caused by infectious or parasitic agents[2].

The disease, to date, does not present effective drug treatment for its chronic stage[3,4]. Due to this, there is a constantly search for new compounds, among which are the diazoles derivatives, which are heterocyclic compounds, containing two nitrogen atoms in a five-membered ring exhibit remarkable reactivity as well as are involved in a large number of chemical reactions and possible activities[5].

One way of achieving these results is to perform studies of the Quantitative Relationship between Chemical Structure and Biological Activity (QSAR)[6], which use information on compounds with known activity values for the construction of predictive models, as well as conducting similarity research chemistry or based on receptor structure[7]. The objective of this study was to perform a virtual screening of 1,2-diazo compounds as potential therapeutic agents for Chagas' disease through the construction of a prediction model.

Materials and Methods

Database

The elaborated prediction model was based on the classification of the data, using nominal variables and was constructed from two sets of structures.

The first set was obtained from the ChEMBL database[8] and was composed of 661 structures with activity potential against Trypanosoma cruzi, belonging to the research organization and development of drugs DNDi (Drugs for Neglected Diseases Initiative)[9].

The compounds were classified according to the pIC50 value (-log of IC50 (mol / L)), with 332 active (pIC50 $\geq$ 7.1) and 328 inactive (pIC50 <7.1). It is worth mentioning that IC50 is the concentration required to inhibit 50% of Trypanosoma cruzi activity, being randomly selected, maintaining the same proportion of active compounds and inactive compounds, in a training set containing 265 active molecules and 263 inactive and set with 67 active and 66 inactive samples.

The prediction set consists of 31 unpublished 1,2-diazoles compounds with potential for synthesis.

Standardization of Chemical Structures

The chemical structures were converted into SMILES, these being the input data for Marvin[10]. Standardizer software[11] was used to standardize the chemical structures, with the addition of hydrogen atoms (H), aromatic ring and 3D structure generation.

Descriptors

The biological and physicochemical properties of the molecules were generated by the VolSurf software (Volume and Surface)[12], which encode different spatial and geometric dimensions and are generated from the 3D structure[13].

Statistical Analyzes

The model was generated by the statistical software KNIME 3.1.0[14], using Random Forest (RF) as calculation algorithm. (RF) is a supervised algorithm that is based on the combination of prediction trees, so that each tree depends on the values of a randomly sampled vector and the same distribution for all the trees in the forest.

Results and Discussion

Hit Rates

In the analysis of the model, the hit rates obtained in the test and cross-validation were higher than 73%. In both models analyzed, this parameter for active compounds was higher than the inactive ones, being 80% and 80% for the test , 78% and 73% in the cross validation respectively. The training set obtained almost perfect performance, presenting a 99% hit rate.

Rating Rate

The classification rate of the model was evaluated by the receiver operating characteristic (ROC) graph, corresponding in the test set to 0.843, in Cross Validation obtained a value of 0.836. It should be noted that a perfect model presents area under the curve equal to 1, in this way, it is possible to state that the model is capable of performing a high classification rate for the RF method.

Prediction Assessment

The Matthews Correlation Coefficient (MCC) was used to evaluate the prediction of the model, resulting in 0.63 for the test set and 0.61 for Cross Validation, respectively, indicating that the model has a good prediction.

Activity Probability

By means of the probability, the elaborated model was used to triage the possible activity of 1,2-Diazois derivatives against Trypanosoma Cruzi. The molecules that reached a probability of being active greater than 50%, totaled the 31 molecules components of the prediction series, 7 with probability of activity between 50 and 59%; 23 with activity probability between 60 and 69%; 13 molecules with probability above 65% and one molecule presented probability above 70%. Recalling that the applicability domain was reliable for 30 molecules, thus being only one molecule component of the prediction series classified as unreliable.

Conclusions

The model presented significative accuracy and reproducibility, even more, it was able to distinguish the probability of potential activity of the molecules that will be synthesized for further studies.

References

1. WORLD HEALTH ORGANIZATION, Chagas Disease (American Trypanosomiasis). Geneva: 2018. Disponível em: http://www.who.int/news-room/fact-sheets/detail/chagas-disease-(american-trypanosomiasis). Acesso em: 10 de Julho de 2018.
2. FOUNDATION OSWALDO CRUZ, Neglected Diseases. Rio de Janeiro: 2018. Disponível em: https://agencia.fiocruz.br/doen%C3%A7as-negligenciadas. Acesso em: 12 de Setembro de 2018.
3. CASTRO, J.A.; DE MECCA, M.M.; BARTEL, L.C. Toxic side effects of drugs used to treat Chagas disease (American trypanosomiasis). **Hum Exp Toxicol.** v.25, n.8, p.471–479, 2006.
4. CAMPOS, M.C.; et al (2014) Benznidazole-resistance in Trypano- soma cruzi: evidence that distinct mechanism can act in concert. **Mol Biochem Parasitol.** v.193, n.1, p.17–19, 2014.
5. CHEN, C.H.; TSAI, W.Y.; LUO, Z.H. Highly fluorescent Conjugated Copolymers Constructed by Alternating Heterocyclic Diazoles and Alkoxy Benzene. **J Chin Chem Soc.** n.65, v.01, p.74-80, 2018.
6. SARAIVA, A.P.B.; MIRANDA, R.M.; VALENTE, R.P.P.; ARAÚJO, J.O.; COSTA, H.S.; OLIVEIRA, A.R.S.; ALMEIDA, M.O.; FIGUEIREDO, A.F.; FERREIRA, E.V.; ALVES, C.N.;

HONORIO, K.M. Molecular description of α-keto-based inhibitors of cruzain with activity against Chagas Disease combinig 3D-QSAR studies and molecular dynamics. **Chem Biol Drug Des.** v.92, n.1, p.1475-1487, 2018.
7. ABDOLMALEKI, A.; B GHASEMI, J.; GHASEMI, F. Computer Aided Drug Design for Multi-Target Drug Design: SAR/QSAR, Molecular Docking and Pharmacophore Methods. **Current Drug Targets,** v. 18, n. 5, p. 556-575, 2017.
8. EUROPEAN BIOINFORMATICS INSTITUTE. CHEMBL Database. Available in: https://www.ebi.ac.uk/chembl/. Access in: 10 july of 2018.
9. DRUGS FOR NEGLECTED DISEASES INITIATIVE. DNDi Latin American. Available in: https://www.dndial.org/dndi-america-latina/quem-somos/. Access in: 10 july of 2018.
10. Marvin 14.9.1.0, 2014, Chem Axon (http://www.chemaxon.com).
11. Standardizer (JChem 14.9.1.0, 2014, Chem Axon (http:// http://www.chemaxon.com)).
12. MOLECULAR DISCOVERY. Vol Surf (Volume and Surface Descriptors). Available in: http://www.moldiscovery.com/software/vsplus/. Access in: 10 july of 2018.
13. ROY, K.; NARAYAN DAS, R. A review on principles, theory and practices of 2D-QSAR. **Current drug metabolism**, v. 15, n. 4, p. 346-379, 2014.
14. KNIME 3.1.0 (The Konstanz Information Miner Copyrith, 2003-2014. Disponível em: www.knime.org. Acesso em: 10 de julho de 2018.
15. BREIMAN, L. Random forests. Machine learning, v. 45, n. 1, p. 5-32, 2001.