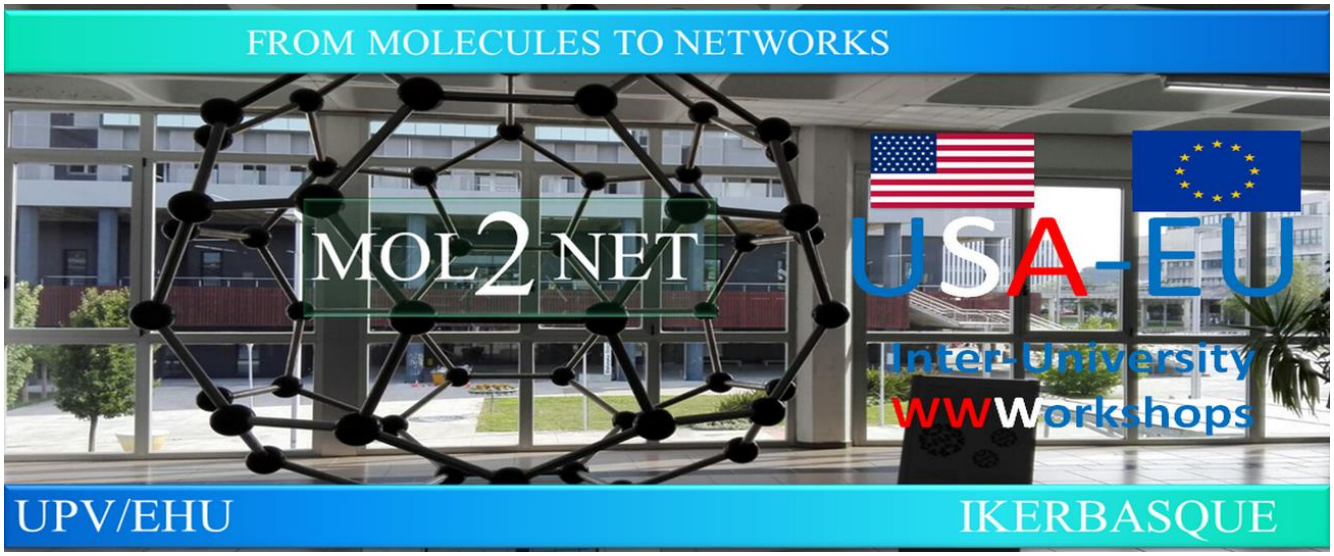
MDPI sciforum

MOL2NET, International Conference Series on Multidisciplinary Sciences



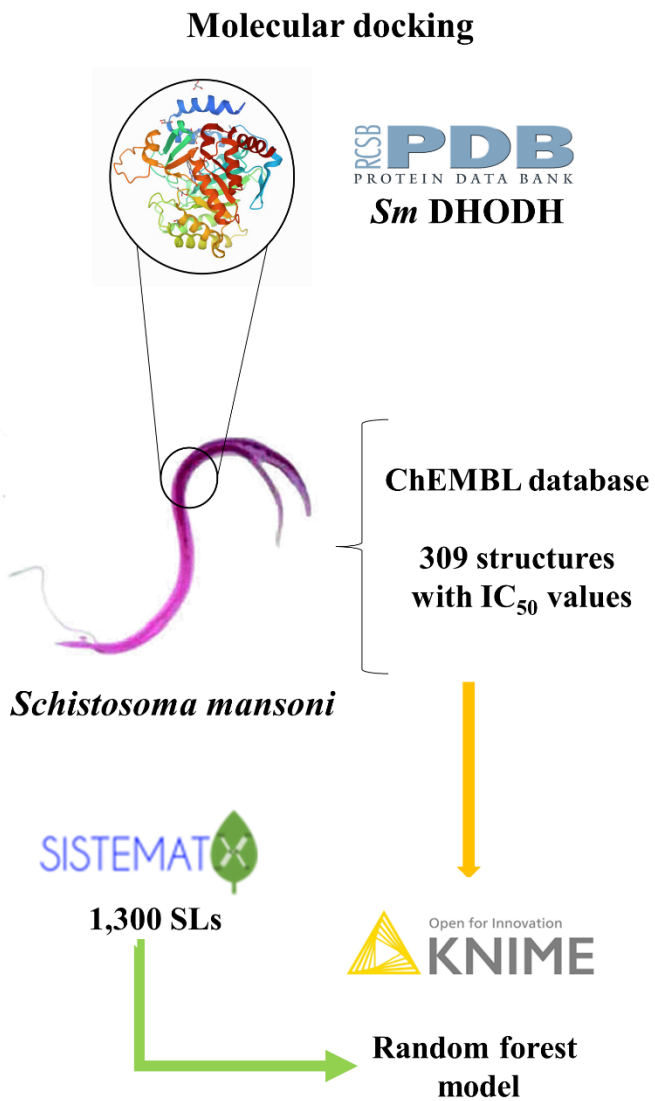
In silico Identification of Potential Sesquiterpene Lactones from SistematX Database against *Schistosoma mansoni*

Chonny Herrera-Acevedo [a,b], Renata Priscila Barros de Menezes [a], Luciana Scotti [a], Ericsson Coy-Barrera [b], Marcus Tullius Scotti [a].*

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

[b] Bioorganic Chemistry Laboratory, Facultad de Ciencias Básicas y Aplicadas, Universidad Militar Nueva Granada, Cajicá 250247, Colombia.

Graphical Abstract	**Abstract.**
	Schistosomiasis is an acute and chronic parasitic disease, caused by blood flukes (trematode worms) of the genus *Schistosoma*. For 2019, the World Health Organization estimated that close to 240 million people required preventive treatment against this disease, mainly poor communities without access to safe drinking water and adequate sanitation. Similarly, to others Neglected Tropical Diseases (NTDs) the treatments against this disease are limited, and new chemotherapies are necessary to the control



and elimination of the Schistosomiasis. Natural products are an interesting alternative in the development of new treatments against NTDs. Interestingly, several studies have demonstrated the great potential of some sesquiterpene lactones (SLs) as potential therapeutic agents for some NTDs and the relationship between the biological activities with their chemical structure. In this study, using a machine learning model with more than 77% accuracy in both the cross-validation and test sets for *Schistosoma mansoni*, a ligand-based virtual screening was performed, using 1,300 SLs registered in SistematX database. The results show that close to 42% of the SLs tested reached active probability values above of 0.5, being identifies some common structural features in the best-ranked molecules. Finally, to explore the mechanism of action of these SLs against *Schistosoma mansoni*, molecular docking calculations were performed for the five best-ranked SLs using the crystal structure of *Sm* dihydroorotate dehydrogenase (DHODH), with 2-[(4-fluorophenyl)amino]-3-hydroxynaphthalene-1,4-dione as inhibitor. Through these calculations, some critical interactions between the tested SLs and the active site of DHODH were identified.

Introduction

Schistosomiasis is an acute and chronic parasitic disease, caused by blood flukes (trematode worms) of the genus *Schistosoma*. For 2019, the World Health Organization (WHO) estimated that close to 240 million people required preventive treatment against this disease, mainly poor communities without access to safe drinking water and adequate sanitation [1].

One of the most interesting targets of *Schistosoma* for the development of new treatments is the dihydroorotate dehydrogenase (DHODH), a flavoenzyme that catalyzes the stereospecific oxidation of (S)-dihydroorotate (DHO) to orotate during the fourth and only redox step of the de novo pyrimidine nucleotide biosynthetic pathway [2]. Atovaquone (an antimalarial treatment) was identified by de Mori et al, as a selective inhibitor against *Schistosoma mansoni* DHODH [3].

Sesquiterpene lactones are chemotaxonomic markers of Asteraceae which have been successfully used in humans against parasite diseases, such artemisinin, an antimalarial sesquiterpene lactone which

discovery led to the 2015 Nobel Prize for Medicine and Physiology [4]. In the case of schistosomiasis, where there have been no in silico studies with SLs [4].

In this study, a machine learning model was built, and a ligand-based virtual screening was performed, using 1,300 SLs registered in SistematX database. The best- ranked molecules were used to explore the mechanism of action of these SLs against *Schistosoma mansoni*, through molecular docking calculations which were performed in the crystal structure of *Sm*DHODH in complex with an atovaquone analogue inhibitor.

Materials and Methods

From the ChEMBL database (https: //www.ebi.ac.uk/chembl/), we selected a diverse set of 309 structures that were initially classified according to their predicted activity against *Schistosoma mansoni*. These compounds were classified according to pIC_{50} values [$-\log IC_{50}$ (mol/L)]; therefore, we stratified them into active ($pIC_{50} \geq 6.0$) and inactive ($pIC_{50} < 6.0$) structures. Data curation of the datasets was performed, according to the suggested procedures in the literature. The 3D structures of the identified molecules, in special data file (SDF) format, were used as input data in Volsurf+, v.1.0.7 resulting in a total of 128 molecular descriptors [5, 6].

The sesquiterpene lactones dataset was obtained from the SistematX database (https ://sistematx.ufpb.br), and a total of 1300 molecules were used in this study. For all structures, SMILES codes were used as the input data in Marvin [ChemAxon, version 20.8.0 (2020), a calculation module developed by ChemAxon, https ://www.chemaxon.com/] and the standardizer software [JChem, version 20.8.0 (2020), a calculation module developed by ChemAxon, https ://www.chemaxon.com/] was used to canonize all SMILES codes.

Knime 4.4.2 software (KNIME 4.4.2 the Konstanz Information Miner Copyright, 2003–2014, www.knime.org) was used to perform all of the machine learning analyses [7]. Initially, the descriptors calculated in the Volsurf+ program were imported, in comma-separated value (CSV) format, and the "Partitioning" node in the stratified sampling option was used to classify 80% of the initial dataset as the training set and the remaining 20% as the test set. The model was generated by employing the modeling set and the RF algorithm, with a "fivefold external validation" procedure, using WEKA nodes. In the fivefold cross-validation procedure, the dataset is divided five times into a modeling set (80–20%). After this modeling set (which was used to build and validate models) is divided additionally into multiple training (80%) and test sets (20%) [8].

In addition, the structure of the *Schistosoma mansoni* dihydroorotate dehydrogenase, DHODH, (PDB ID: 6UY4) in complex with its respective inhibitor: 2-[(4-fluorophenyl)amino]-3-hydroxynaphthalene-1,4-dione (PDB ID: QLA) was downloaded from PDB [2]. Using Molegro 6.0.1, molecular docking calculations were performed using a grid, with a 15-Å radius and a 0.30-Å resolution, to cover the ligand-binding site for the DHODH structure.

Results and Discussion

For the training set used in the Random Forest (RF) model, the match percentage values approached 100%. For the cross-validation (CV) and test sets, values above 78.1% and 77.6% were obtained,

respectively. Sensitivity, which is defined as the true-positive rate, (78.3%) was greater than the specificity rate (77.9%), which is defined as the true-negative rate, in the five-fold CV.

The area under ROC curve (AUC), a quality parameter that plots the sensitivity against (1—specificity) was calculated for both CV and test sets [9]. AUC values of 0.868 and 0.877 were achieved for five-fold CV and external test sets, respectively, demonstrating a high rate of sensitivity and a low false-positive rate.

Additionally, Matthews's correlation coefficient (MCC),which is determined from all of the values obtained from the confusion matrix, also was calculated [10]. MCC values of 0.562 and 0.552 for five-fold CV and external test sets, respectively were obtained, demonstrates a high degree of differentiation between the active and inactive compounds identified in the ChEMBL dataset of *S. mansoni.*

Finally, the applicability domain (APD) was used to assess the reliability of the predictions for the samples in the external test and SLs sets, and the calculation of the APD is based on the molecular interactions determined by the VolSurf+ descriptors [5, 6]. All structures for the external test set and 97.3% for the SistematX SLs set were classified as reliable.

Table 1: Summary of internal cross-validation and external test results obtained using the RF algorithm for *S. mansoni*

	Five-fold CV	**External test**
Active	78.3	77.1
Inactive	77.9	78.0
Overall	78.1	77.6
MCC	0.562	0.552
ROC	0.868	0.877

After, using this RF algorithm, a ligand-based virtual screening was performed on a set including 1,300 molecules obtained from SistematX. For *S. mansoni* 557 SLs (42.8%) showed active probability values above of 0.5. Structurally, some common features were found among the best-ranked structures. Structure **1** and **3** contain a germacranolide skeleton, bound to stearic fatty acid ester. Additionally, the group α-methylene-γ-lactone was observed in four of the five best-ranked structures (structures **1, 3-5)**, the presence of this moiety has been associated with interactions between this type of metabolite and the sulfhydryl group of cysteine, through a Michael addition (Figure 1) [11].

The structure **1** (9α-stearoyloxy-8β-(2-methylbutyryloxy)-15-hydroxy-14-oxo-acanthospermolide), is a secondary metabolite of *Acanthospermun hispidium*, a plant native to Central and South America (Table 2) [12]. Some metabolites structurally related with structure **1,** 9α-linoloyloxy-8β-(2-methylbutyryloxy)-15-hydroxy-14-oxo-acanthospermolide and 9a-linolenoyloxy-15-hydroxy-8b-2-methylbutyryloxy-14-oxo-acanthospermolide have been identified as potential antileishmanial and antichagasic SLs, two of the main NTDs that affect the American continent [8, 13].

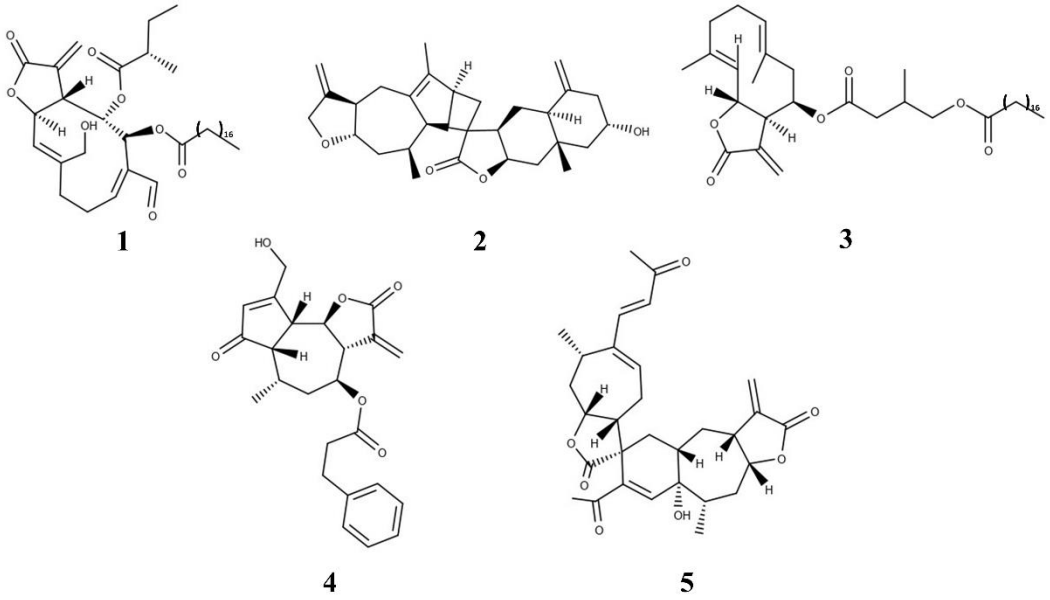


Figure 1: Potentially active SLs, identified using a machine learning model (RF algorithm), for *Schistosoma mansoni*

Table 2: Structure name, botanical species, and ligand-based active probability for the five best-ranked SLs

SL	Structure name	Species	*p(A)*
1	9α-stearoyloxy-8β-(2-methylbutyryloxy)-15-hydroxy-14-oxo-acanthospermolide	*Acanthospermum hispidum*	**0.89**
2	Carpedilactone F	*Carpesium macrocephalum*	**0.87**
3	8β-4-stearoyloxy-isovaleroyloxy-9β-hydroxycostunolide	*Hypochaeris radicata*	**0.86**
4	Hyporadiolide-8-O-cinnamate	*Xanthum brasilicum*	**0.84**
5	Pungiolide A	*Grazielia dimorpholepis*	**0.83**

Molecular docking calculations were performed to explore the mechanism of action of the five best-ranked SLs in the active site of *Schistosoma mansoni* dihydroorotate dehydrogenase (*Sm*DHODH), a flavoenzyme that catalyzes the stereospecific oxidation of (S)-dihydroorotate (DHO) to orotate during the fourth and only redox step of the de novo pyrimidine nucleotide biosynthetic pathway [2]. The crystal structure of *Sm*DHODH in complex with the inhibitor 2-((4-fluorophenyl)amino)-3-hydroxynaphthalene-1,4-dione (QLA) was obtained from the Protein Data Bank (PDB).

The molecular docking calculations were initially validated through redocking procedure, using QLA in the active site of the protein. The results showed a docking score of -114.82 kJ/mol with a Root-Mean-Square-Deviation (RMSD) of 0.249 Angstroms.

Respect the SLs, only three of the five structures showed lower docking scores respect QLA, the structures **2** and **5**, exhibited similar scores respect the inhibitor reported in the PDB (Table 3). The docking results are not influenced by the molecular weight of the structures, being the structure **2**, one of the two highest structures in the tested SLs, but the docking score is higher respect QLA.

Table 3: Docking scores and hydrogen bond interactions of structure **1-5** and QLA with amino acid residues of *Sm*DHODH. Critical interactions are highlighted in bold font

Structure	Docking score (kJ/mol)	Hydrogen bond interactions
1	-199.67	G361
2	-102.68	-
3	-162.92	**S53**
4	-141.38	**S53, R130**
5	-110.63	**S53**
QLA	-114.82	H50, **S53, R130**

The residues S53 appears as critical aminoacid for binding to *Sm*DHODH, being observed in the Structure **3**,**4** and **5.** Interestingly, for the structure **4**, a hydrogen bond interaction with R130 is observed. This interaction is also observed in the inhibitor QLA (Figure 2) . The Structure **4** showed a lower binding docking score (-141,38 kJ/mol) and exhibited a similar pattern of interaction in the active site of *Sm*DHODH respect the inhibitor reported in the PDB, being one of the most promissory SL against *S. mansoni*.

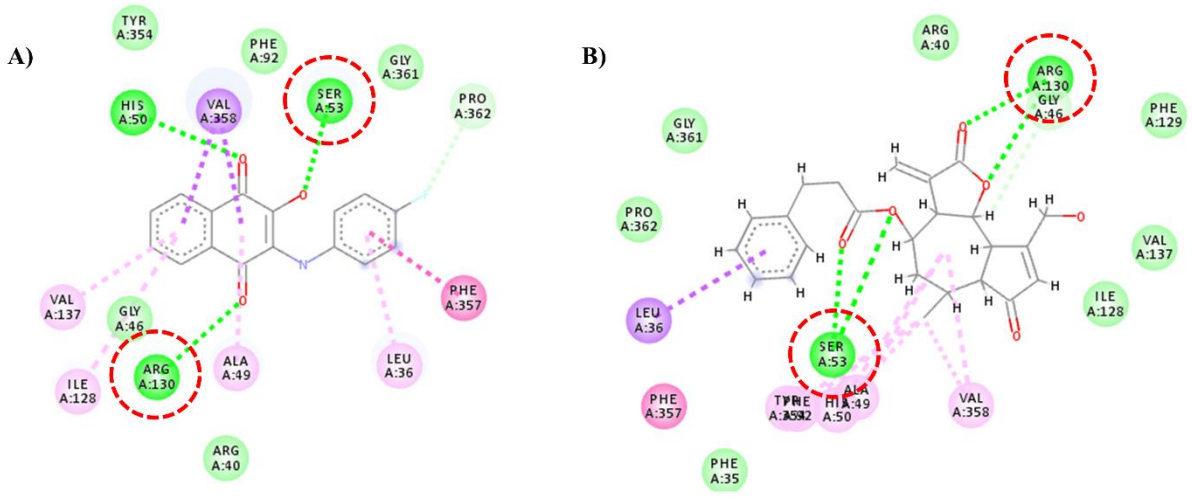


Figure 2: Docking conformation of A) QLA and B) Structure **4**, in the active site of *Sm*DHODH. The interacting residues are shown as colored circles, and interactions are indicated as colored dashed lines: H-bond (lime), van der Waals (green), π–π (purple) and π–alkyl (pink), unfavorable (red) and carbon–H-bond (teal). Red dotted circles indicate the critical H-bond interactions.

Conclusions

Using computational techniques, a preliminary evaluation of a set of 1300 SLs looking for potential agents against *S. mansoni*. The ligand-based virtual screening allows to select the most promissory molecules based in the calculated probability in the machine learning RF model. The five selected molecules showed structural features related with the biological activity observed in the SLs, i.e., the presences of the α-methylene-γ-lactone moiety. After, the molecular docking calculations allowed the

identification of three structures with lower docking values respect QLA, being the structure **4** the most promissory inhibitor of the *Sm*DHODH, one of the most important targets in the *Schistosoma* life cycle. This study is only explanatory but might be considered as a starting point for the development of new treatments against this NTD.

References

1. World Health Organization (WHO). Schistosomiasis. https://www.who.int/news-room/fact-sheets/detail/schistosomiasis (2021/11/18),
2. de Mori, R. M.; Aleixo, M. A. A.; Zapata, L. C. C.; Calil, F. A.; Emery, F. S.; Nonato, M. C., Structural basis for the function and inhibition of dihydroorotate dehydrogenase from Schistosoma mansoni. *The FEBS Journal* **2021,** 288, (3), 930-944.
3. Calil, F. A.; David, J. S.; Chiappetta, E. R. C.; Fumagalli, F.; Mello, R. B.; Leite, F. H. A.; Castilho, M. S.; Emery, F. S.; Nonato, M. C., Ligand-based design, synthesis and biochemical evaluation of potent and selective inhibitors of Schistosoma mansoni dihydroorotate dehydrogenase. *European journal of medicinal chemistry* **2019,** 167, 357-366.
4. Herrera Acevedo, C.; Scotti, L.; Feitosa Alves, M.; Formiga Melo Diniz, M. D. F.; Scotti, M. T., Computer-aided drug design using sesquiterpene lactones as sources of new structures with potential activity against infectious neglected diseases. *Molecules* **2017,** 22, (1), 79.
5. Cruciani, G.; Crivori, P.; Carrupt, P. A.; Testa, B., Molecular fields in quantitative structure–permeation relationships: the VolSurf approach. *Journal of Molecular Structure: THEOCHEM* **2000,** 503, (1-2), 17-30.
6. Cruciani, G.; Pastor, M.; Guba, W., VolSurf: a new tool for the pharmacokinetic optimization of lead compounds. *European Journal of Pharmaceutical Sciences* **2000,** 11, S29-S39.
7. Berthold, M. R.; Cebron, N.; Dill, F.; Gabriel, T. R.; Kötter, T.; Meinl, T.; Ohl, P.; Thiel, K.; Wiswedel, B., KNIME-the Konstanz information miner: version 2.0 and beyond. *AcM SIGKDD explorations Newsletter* **2009,** 11, (1), 26-31.
8. Herrera-Acevedo, C.; Maia, M. D. S.; Cavalcanti, E. B. V. S.; Coy-Barrera, E.; Scotti, L.; Scotti, M. T., Selection of antileishmanial sesquiterpene lactones from SistematX database using a combined ligand-/structure-based virtual screening approach. *Molecular Diversity* **2020**, 1-17.
9. Hanley, J. A.; McNeil, B. J., The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* **1982,** 143, (1), 29-36.
10. Matthews, B. W., Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochimica et Biophysica Acta (BBA)-Protein Structure* **1975,** 405, (2), 442-451.
11. Schmidt, T. J., Toxic activities of sesquiterpene lactones: structural and biochemical aspects. *Curr. Org. Chem* **1999,** 3, (577-608), 4.
12. Walker, R. H.; Wells, L. W.; McGuire, J. A., Bristly starbur (Acanthospermum hispidum) interference in peanuts (Arachis hypogaea). *Weed Science* **1989,** 37, (2), 196-200.
13. Acevedo, C. H.; Scotti, L.; Scotti, M. T., In Silico Studies Designed to Select Sesquiterpene Lactones with Potential Antichagasic Activity from an In-House Asteraceae Database. *ChemMedChem* **2018,** 13, (6), 634-645.