Article

Virtual Screening Based on Ligand and Structure with *in vitro* Assessment of Neolignans against *Trypanosoma cruzi*

Mayara S. Maia,[a] Rodrigo S. Andrade,[b] Julyanne M. S. Sousa,[c] Natália F. Sousa,[a]
Gabriela C. S. Rodrigues,[a] Renata P. B. Menezes,[a] Marcelo S. Silva,[b] Josean F. Tavares,[b]
Klinger A. F. Rodrigues,[c] Luciana Scotti[a] and Marcus T. Scotti[a] *,[a]

[a]Laboratório de Quimioinformática,
Programa de Pós-Graduação em Produtos Naturais e Sintéticos Bioativos (PgPNSB),
Centro de Ciências da Saúde, Universidade Federal da Paraíba, 58051-900 João Pessoa-PB, Brazil

[b]Núcleo de Caracterização e Análise (NUCAL),
Programa de Pós-Graduação em Produtos Naturais e Sintéticos Bioativos (PgPNSB),
Centro de Ciências da Saúde, Universidade Federal da Paraíba, 58051-900 João Pessoa-PB, Brazil

[c]Laboratório de Doenças Infecciosas, Universidade Federal do Delta do Parnaíba,
São Benedito, 64202-020 Parnaíba-PI, Brazil

Chagas disease, caused by the parasite *Trypanosoma cruzi*, occurs most commonly in Latin America. As the treatment is highly toxic and ineffective in the chronic phase of the disease, alternative treatments are needed. Through quantitative structure-activity relationship modeling (QSAR) analysis using ligand-based and structure-based virtual screening methods, we predicted the trypanocidal potential of 47 neolignans against three targets, the enzymes cruzain, trypanothione reductase, and sterol 14-alpha demethylase. A combined analysis allowed for the selection of potent inhibitors against *Trypanosoma cruzi*. Of these compounds, two were isolated and shown to inhibit the growth of epimastigotes at concentrations of 9.64 and 8.72 μM, and trypomastigote forms at 4.88 and 2.73 μM. Therefore, the compounds (2*R*, 3*R*)-2,3-dihydro-2-(4-methoxyphenyl)-3-methyl-5-(*E*)-propenylbenzofuran (**46**) and ottomentosa (**47**) may be a good option of growth inhibitors for the parasite stages and warrant additional study.

Keywords: Chagas disease, *Trypanosoma cruzi*, neolignane, virtual screening, molecular docking, molecular dynamics simulations

Introduction

Trypanosoma cruzi the protozoan parasite that causes Chagas disease (CD), was initially endemic to Latin America, but has spread to other locations such as Canada, the United States, Europe, Australia, and Japan.[1] It currently affects 6-7 million people worldwide and causes approximately 50,000 deaths *per* year.[1,2] The transmission of *T. cruzi* can occur congenitally, through organ transplantation, blood or, ingestion of food and drinks contaminated by the parasite.[3]

The parasites are transmitted predominantly to humans as metacyclic trypomastigote (MT) forms through the contaminated feces of blood-sucking triatomines at the bite site. After internalization by host cells near the entry site, MTs initially reside in a vesicle containing the parasite, the parasitophore vacuole, from where they escape into the host cell's cytoplasm and differentiate in the proliferative amastigote form. After several cycles of replication, amastigotes differentiate into mobile flagellated trypomastigotes, which are released into the bloodstream, from where they can spread by infecting distant tissues or are captured by the triatomine vector during a blood meal. Ingested blood trypomastigotes become epimastigotes in the vector's midgut, multiplying and then differentiating into infectious metacyclic trypomastigotes. Infectious trypomastigotes and intracellular replicative amastigotes are the parasite's clinically relevant cycle stages to the drug target.[3,4]

Currently, no vaccines prevent diseases caused by trypomastigotas, and although chemotherapeutic drugs have been available for decades, they are highly

toxic and have unpleasant side effects. Benznidazole, a derivative of nitroimidazole, and nifurtimox, a nitrofuran compound, both developed more than 40 years ago, are currently the only drugs available for the treatment of CD.[5] Although benznidazole is the first-line drug due to its better tolerability, both drugs have significant side effects.[6] Therefore, patients should be monitored frequently. Unfortunately, the medications available are effective only in the acute phase and in 20% of cases the treatment must be stopped due to side effects.[7]

As neolignans are a diverse group of chemical structures present in several plant families and responsible for multiple biological activities, we chose to investigate the trypanomicidal potential of a set of neolignans.[8-11] Some studies[12-14] have reported the importance of neolignans as promising compounds in the treatment against *T. cruzi*.

Through a literature review, we were able to find and highlight three important targets (the enzymes cruzain, trypanothione reductase [TR], and sterol 14-alpha demethylase [CYP51]) for the proliferation and survival of the parasite in the parasite's three cellular forms. Cruzain and TR enzymes are found only in tryponomastids, which contributes to the design of selective drugs that do not harm humans. Cruzain is the main cysteine protease of *T. cruzi*, TR is essential for redox metabolism and CYP51 is the key enzyme in ergosterol biosynthesis. Furthermore, the enzymes cruzain and CYP51 are essential for all stages of parasite development.[15-17]

We explored computational and experimental studies to select the most promising trypanomicidal compounds that may be effective in the acute and chronic phases of CD and present low toxicity. Thus, the identified compounds may be effective as insecticidal and drug agents in the acute and chronic phases of the disease.

Results and Discussion

Quantitative structure-activity relationship modeling (QSAR) modeling

Three prediction models were built using the random forest (RF) algorithm to perform ligand-based virtual screening. For the construction of these models, molecular descriptors were calculated for molecules with known activity against cruzain, TR and CYP51 of *T. cruzi* obtained through the ChEMBL database.[18,19]

RF models were evaluated for their predictive powers, using the parameters of specificity, sensitivity, precision, accuracy (AUC), positive predicted value (PPV) and negative predicted value (NPV), in addition to performance and robustness, using the receiver's operating characteristic (ROC) curve and Mathews correlation coefficient (MCC). Table 1 describes the characteristics of the models, in terms of predictive power and robustness, and Figure 1 shows the performance of the models. The results showed that the models provided satisfactory classification, performance, and robustness, except for the CYP51 enzyme, for which the accuracy and specificity values were below 0.6. Therefore, this model was disregarded for the prediction.

After the models were validated, they were used to analyze the set of neolignans for activity against *T. cruzi*. The neolignan bank was then screened to select compounds that are potentially active against cruzain and TR.

The RF model was able to select a compound with active potential, with probabilities from 50% for cruzain (Table 2). The TR model was able to classify all 47 compounds as potentially active, with probabilities ranging between 54 and 84% (Table 2). According to these results, neolignans have a greater active potential for the TR protein than for cruzain.

Docking consensus

The 47 neolignans also underwent a docking consensus assessment to increase the method's reliability and decrease false positives. The enzymes cruzain, TR and CYP51 were used for docking studies. The docking results were generated using three different scoring functions and validated by redocking the PDB ligand (ligand inhibitor crystallized with the protein of the Protein DataBank-PDB)[20] for each enzyme. More negative values indicated better predictions for most scoring functions.

After docking, the results were standardized for each scoring function using the docking probability ($Prob_{Dc}$) formula:[21]

Table 1. Summary of parameters corresponding to the results obtained for all models

Enzyme	Validation	Specificity	Sensitivity	Accuracy	PPV[a]	NPV[b]	MCC[c]	AUC[d]
Cruzain	test	0.84	0.77	0.81	0.81	0.81	0.61	0.86
	cross	0.80	0.68	0.75	0.75	0.75	0.49	0.82
TR	test	0.63	0.95	0.85	0.85	0.87	0.65	0.94
	cross	0.70	0.92	0.85	0.86	0.81	0.65	0.89

[a]Positive predicted value; [b]negative predicted value; [c]Mathews correlation coefficient; [d]receiver's operating characteristic. TR: trypanothione reductase.
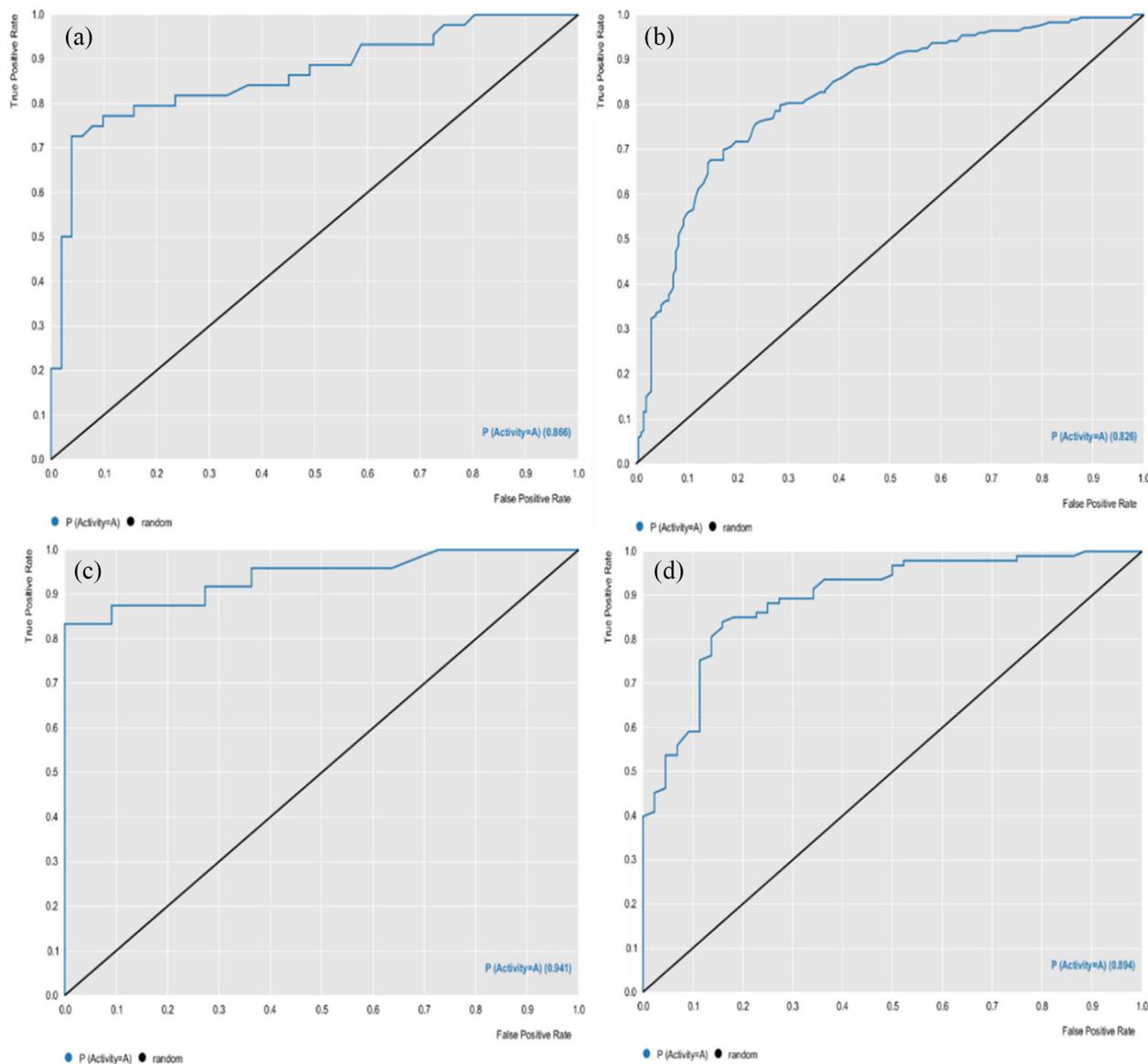
Figure 1. Receiver operating characteristic (ROC) curve-generated random forest (RF) model. (a) Test and (b) cross-validation for the enzyme cruzain and (c) test and (d) cross-validation for the TR.

$$\text{Prob}_{Dc} = \frac{E_{Lig}}{E_{MLig}}, \text{ if } E_{lig} < E_{Inib} \qquad (1)$$

where E_{Lig} is the energy of the ligand, E_{MLig} is the energy of the ligand with the highest score, and E_{Inib} is the energy of the inhibitor obtained from the crystallographic data of the test protein. The highest value in Prob_{Dc} is equal to 1. Thus, only those compounds with energy values equal to or greater than the interaction with energy of the crystallographic inhibitor ligand were considered to be potentially active. Then, an average of the standardized results for each scoring function was determined.

The docking results were validated by re-docking the crystallographic ligand and the root-mean-square deviation (RMSD) of the positions. Re-docking consists

of positioning and predicting the binding affinity of the crystallographic ligand in the region of the enzyme's active site. The RMSD compares and calculates the mean square root deviation of the postures obtained by re-docking and the ligand structure obtained experimentally. For the adjustment to be reliable, the RMSD value must be 2.0 Å or less.[22] The results showed that the targets cruzain, TR and CYP51 obtained 0.77, 0.64 and 0.31 Å, respectively.

Among the 47 neolignans analyzed by molecular docking, two compounds were potentially active against cruzain. The cruzain test inhibitor had a Prob_{Dc} value of 0.88 and only two neolignans showed values higher than the inhibitor, with values of 0.91 (compound **1**) and 0.90 (compound **35**). For the TR enzyme, 31 neoligans were considered active, with Prob_{Dc} values ranging from 0.49

Table 2. Neolignans activity probabilities (pActivity) against cruzain and TR as assessed by the RF model. The compounds considered active in the models are highlighted in bold

ID	pActivity		ID	pActivity		ID	pActivity	
	Cruzain	TR		Cruzain	TR		Cruzain	TR
1	0.37	**0.73**	**17**	0.38	**0.73**	**33**	0.46	**0.83**
2	0.38	**0.67**	**18**	0.27	**0.65**	**34**	0.31	**0.78**
3	0.40	**0.76**	**19**	0.41	**0.75**	**35**	0.48	**0.82**
4	0.39	**0.73**	**20**	0.40	**0.73**	**36**	0.41	**0.74**
5	0.29	**0.75**	**21**	0.31	**0.78**	**37**	0.34	**0.69**
6	0.35	**0.66**	**22**	0.36	**0.75**	**38**	0.35	**0.68**
7	0.31	**0.72**	**23**	0.29	**0.75**	**39**	0.34	**0.72**
8	0.48	**0.79**	**24**	0.29	**0.68**	**40**	0.37	**0.69**
9	0.41	**0.76**	**25**	0.39	**0.78**	**41**	0.37	**0.73**
10	0.37	**0.54**	**26**	0.30	**0.79**	**42**	0.34	**0.78**
11	0.35	**0.64**	**27**	0.48	**0.79**	**43**	0.47	**0.82**
12	**0.50**	**0.78**	**28**	0.29	**0.78**	**44**	0.29	**0.80**
13	0.34	**0.70**	**29**	0.35	**0.76**	**45**	0.35	**0.74**
14	0.37	**0.76**	**30**	0.47	**0.84**	**46**	0.36	**0.62**
15	0.31	**0.75**	**31**	0.28	**0.80**	**47**	0.39	**0.72**
16	0.39	**0.71**	**32**	0.32	**0.60**			

TR: trypanothione reductase.

to 0.81. The TR inhibitor had a $Prob_{Dc}$ value equal to 0.49. For the CYP51 enzyme, the inhibitor obtained a value of 0.56 and 18 neolignans obtained an equal or greater value. These results indicated that neolignans, in general, were more likely to activate TR and CYP51 proteins, and are not selective for the cruzain enzyme. These results can be seen in Tables 3-5.

Structure and ligand-based combined analyses

A second consensus analysis was carried out to identify potential lignans and multitarget, which demonstrate the probabilities of being active potentials for more than one protein, based on the RF and docking model. In this case, we used all the results of the prediction of the biological activity of the neolignans and combined them with the docking results. In addition to selecting the active compounds, this combined analysis allowed the selection of the most potent compounds by combining two important methodologies, based on the ligand and structure. For this analysis, the following formula was used:[21]

$$Prob_{Comb} = \frac{(Prob_{Dc} + (1 + ESP)) \times P_{Activity}}{2 + ESP}, \text{ if } Prob_{Comb} > 0.5 \quad (2)$$

where $Prob_{Comb}$ is the combined probability between the RF model and the docking model, $Prob_{Dc}$ is the probability of a compound being active in the molecular coupling

analysis, ESP is the value of the specificity of the RF model and $P_{Activity}$ is the probability value of a compound to be active in the RF model. The combined probability ($Prob_{comb}$), based on the ligand and structure, can increase the predictive power of the models and decrease the number of false positives. For the molecules to be considered potentially active, the probability values must be equal to or greater than 0.5. The higher the $Prob_{comb}$ value, the greater the potential of the molecule. The combined probability values were calculated for all neolignans and each target enzyme, and we analyzed which molecules were multitargets. Only for the CYP51 enzyme was it impossible to calculate the $Prob_{comb}$ value, as we did not have biological activity data for this enzyme due to the low quality of the generated RF model. Therefore, we considered only the $Prob_{Dc}$ values for this target. In this case, the $Prob_{comb}$ value needed to be higher than the crystallographic ligand, which was 0.56.

After performing the combined analysis, based on the ligand and structure, and using the formula to identify potentially active and multitarget molecules, we identified 29 potentially active molecules for more than two target enzymes from the entire set of neolignans analyzed. In addition, after the combined probability analysis, we selected the multitarget compounds that passed the applicability domain for all enzymes in this study. Using $Prob_{comb}$, we were able to select 22 compounds with a probability of activity ranging from 50-65% for cruzain

Table 3. Improved results in the combined probability between the prediction model and molecular docking analysis ($Prob_{comb}$) for potential activity against cruzain. The compounds shown are the compounds considered active (with $Prob_{comb}$ equal to or above 0.50) with values of binding energy, molecular docking probability ($Prob_{Dc}$) and probability of biological activity ($Prob_{Ac}$)

ID	Moldockscore	$Prob_{Dc}$[a]	Plantscore	$Prob_{Dc}$	Vina	$Prob_{Dc}$	Mean $Prob_{Dc}$	$Prob_{Ac}$[b]	$Prob_{Comb}$[c]
1	−116.3	0.94	−100.3	0.90	−5.9	0.89	0.91	0.37	0.56
2	−89.40	0.72	−65.45	0.58	−5.7	0.86	0.72	0.38	0.50
3	−95.33	0.77	−83.19	0.74	−5.2	0.78	0.76	0.40	0.53
4	−88.76	0.71	−70.87	0.63	−5.3	0.80	0.71	0.39	0.50
8	−103.2	0.83	−74.23	0.66	−5.7	0.86	0.78	0.48	0.59
10	−107.3	0.86	−88.76	0.79	−5.7	0.86	0.84	0.37	0.53
12	−83.28	0.67	−37.54	0.33	−5.6	0.84	0.61	0.5	0.54
13	−109.6	0.88	−82.69	0.74	−5.8	0.87	0.83	0.34	0.51
19	−82.13	0.66	−75.93	0.68	−5.6	0.84	0.73	0.41	0.52
20	−96.09	0.77	−81.80	0.73	−5.3	0.80	0.77	0.40	0.53
21	−99.46	0.80	−77.27	0.69	−5.2	0.78	0.76	0.36	0.50
25	−96.83	0.78	−76.85	0.69	−.53	0.80	0.75	0.39	0.51
37	−94.64	0.76	−57.00	0.51	−5.7	0.86	0.71	0.48	0.56
30	−123.6	1.00	−111.2	1.00	−6.6	1.00	1.00	0.47	0.65
32	−109.5	0.88	−95.16	0.85	−6.4	0.96	0.90	0.32	0.52
33	−106.9	0.86	−76.32	0.68	−5.6	0.84	0.79	0.46	0.57
35	−106.3	0.86	−56.69	0.50	−5.6	0.84	0.73	0.48	0.57
36	−93.47	0.75	−64.61	0.58	−5.6	0.84	0.72	0.41	0.52
40	−94.63	0.76	−83.37	0.74	−5.4	0.81	0.77	0.37	0.51
43	−91.60	0.74	−80.04	0.71	−6.6	1.00	0.82	0.47	0.59
45	−92.24	0.74	−80.47	0.72	−5.8	0.87	0.78	0.35	0.50
47	−83.80	0.67	−77.48	0.69	−5.3	0.80	0.72	0.39	0.51
Inhibitor	−117.3	0.94	−74.33	0.66	−6.9	1.04	0.88		

[a]Probability of a compound being active in the molecular coupling analysis; [b]probability of biological activity; [c]combined probability between the ligand and structure-based.

(Table 3). The combined analysis allowed us to select more active compounds for this target than using the $P_{Activity}$ and $Prob_{Dc}$ results separately. It was also possible to select 46 neolignans potentially active against the TR enzyme with a $Prob_{comb}$ value ranging from 52-82% (Table 4), while 18 neolignans obtained a probability ranging between 56-88% for the CYP51 enzyme (Table 5).

The set of 47 neolignans was submitted to several predictive parameters to identify the compounds with the best pharmacokinetic, pharmaco-chemical and pharmacological profiles. Initially, through physicochemical properties, we sought to verify compounds with good absorption, using the lipid rule as a parameter.

We evaluated the absorption and bioavailability properties using the Lipinski rule,[23] comprising molecules with molecular weights below 500 Da, calculated the partition coefficient (LogP) (cLogP) values less than five, less than five hydrogen bonding donors, no more than ten hydrogen bonding acceptors and ≤ 10 rotating bonds with excellent absorption and bioavailability. Molecules

that violate two or more of these rules do not demonstrate sufficient absorption. We observed that only one neolignan from our set of compounds did not meet this requirement. Therefore, 97.87% of the neolignans showed good absorption and bioavailability.

Factors such as lipophilicity and solubility contribute to the distribution of the drug *in vivo*, which is a requirement for advancing to preclinical and clinical testings. The most common descriptor for lipophilicity is the partition coefficient between *n*-octanol: water (log P). The results showed that all neolignans had ideal logP values below 5.0.

Pharmacokinetics are essential for understanding drug metabolism in the body, half-lives, and toxic metabolites. Unfortunately, many compounds fail in the pre-clinical and clinical testing phase due to the effects of metabolism and malabsorption on the brain. Therefore, an early assessment of this effect is necessary and an *in silico* approach contributes substantially to mitigate adverse reactions that some may experience. The results showed that 47.55% of

Table 4. Improved results in the combined probability between the prediction model and molecular docking analysis (Prob$_{comb}$) for potential activity against TR. The compounds shown are the compounds considered active (with Prob$_{comb}$ equal to or above 0.50) with values of binding energy, molecular docking probability (Prob$_{Dc}$), and probability of biological activity (Prob$_{Ac}$)

ID	Moldockscore	Prob$_{Dc}$[a]	Plantscore	Prob$_{Dc}$	Vina	Prob$_{Dc}$	Mean Prob$_{Dc}$	Prob$_{Ac}$[b]	Prob$_{Comb}$[c]
1	−66.32	0.83	−59.59	1	−5.9	0.57	0.80	0.69	0.72
2	−57.65	0.72	−17.88	0.30	−4.9	0.47	0.49	0.67	0.61
3	−70.54	0.88	−23.93	0.40	−5.2	0.50	0.59	0.76	0.70
4	−49.40	0.61	−37.15	0.62	−5.5	0.53	0.59	0.73	0.68
5	−48.52	0.60	−30.76	0.51	−6.1	0.59	0.57	0.75	0.69
6	−51.82	0.64	−12.09	0.20	−5.9	0.57	0.47	0.66	0.59
7	−55.53	0.69	−42.41	0.71	−5.8	0.56	0.65	0.72	0.70
8	−54.15	0.67	−45.39	0.76	−4.5	0.43	0.62	0.79	0.73
9	−38.40	0.48	−9.34	0.15	−5.1	0.49	0.37	0.76	0.62
10	−45.26	0.56	−16.38	0.27	−6.5	0.63	0.49	0.54	0.52
11	−46.73	0.58	−15.93	0.26	−5.8	0.56	0.47	0.64	0.58
12	−62.18	0.77	−11.01	0.18	−4.3	0.41	0.46	0.78	0.66
13	−46.89	0.58	−44.40	0.74	−5.4	0.52	0.61	0.70	0.67
14	−58.57	0.73	−18.65	0.31	−5.3	0.51	0.52	0.76	0.67
15	−45.45	0.56	−10.53	0.17	−5.3	0.51	0.42	0.75	0.63
16	−40.49	0.50	−13.13	0.22	−5	0.04	0.25	0.71	0.55
17	−36.43	0.45	−12.86	0.21	−5.4	0.52	0.39	0.73	0.61
18	−61.88	0.77	−20.42	0.34	−5.9	0.57	0.56	0.65	0.61
19	−61.43	0.76	−37.00	0.62	−5.4	0.52	0.63	0.75	0.71
20	−49.21	0.61	−20.69	0.34	−5	0.04	0.33	0.73	0.59
21	−44.15	0.55	−23.60	0.39	−5.7	0.55	0.50	0.78	0.68
22	−64.33	0.80	−9.64	0.16	−5.9	0.57	0.51	0.75	0.66
23	−40.16	0.50	−16.93	0.28	−5.3	0.51	0.43	0.75	0.63
24	−40.14	0.50	−11.51	0.19	−5.6	0.54	0.41	0.68	0.58
25	−49.90	0.62	−34.74	0.58	−5.1	0.49	0.56	0.78	0.70
26	−49.24	0.61	−32.83	0.55	−5.8	0.56	0.57	0.79	0.71
27	−68.09	0.85	−4.23	0.07	−4.9	0.47	0.46	0.79	0.67
28	−49.14	0.61	−37.18	0.62	−6.1	0.59	0.61	0.78	0.72
29	−32.22	0.40	0.14	0.00	−5.6	0.54	0.31	0.76	0.60
30	−79.81	1	−35.71	0.54	−6.1	0.59	0.73	0.84	0.80
31	−50.03	0.62	−32.6	0.54	−6.1	0.59	0.58	0.80	0.72
32	−59.18	0.74	−20.43	0.34	−6.4	0.62	0.56	0.66	0.62
33	−77.97	0.97	−58.24	0.97	−5.2	0.50	0.81	0.83	0.82
34	−72.31	0.90	−21.00	0.35	−5.8	0.56	0.60	0.78	0.72
35	−65.48	0.82	−24.32	0.40	−5.2	0.50	0.57	0.82	0.73
36	−48.22	0.60	−4.96	0.08	−5.4	0.52	0.40	0.74	0.62
37	−56.29	0.70	−19.97	0.33	−5.7	0.55	0.53	0.69	0.63
39	−62.93	0.78	−12.24	0.20	−5.2	0.50	0.49	0.72	0.64
40	−52.84	0.66	−20.89	0.35	−5.4	0.52	0.51	0.69	0.63
41	−52.96	0.66	−44.53	0.74	−5.2	0.50	0.63	0.73	0.70
42	−62.18	0.77	−37.79	0.63	−6	0.05	0.49	0.78	0.67
43	−66.39	0.83	−28.39	0.47	−10.3	1	0.76	0.82	0.80
44	−53.40	0.66	−13.27	0.22	−6.3	0.61	0.50	0.80	0.69
45	−50.15	0.62	−21.29	0.35	−5.7	0.55	0.51	0.74	0.66
46	−55.91	0.70	−42.10	0.70	−6	0.05	0.48	0.62	0.57
47	−57.64	0.72	−36.96	0.62	−5.2	0.50	0.61	0.72	0.68
Inhibitor	44.76	0.56	−23.02	0.38	−5.6	0.54	0.49		

[a]Probability of a compound being active in the molecular coupling analysis; [b]probability of biological activity; [c]combined probability between the ligand and structure-based.

Table 5. Improved results in the probability of molecular docking ($Prob_{Dc}$) for potential activity against CYP51. Compounds considered active (with values equal to or greater than 0.56 in the $Prob_{Dc}$) are shown

ID	Moldockscore	$Prob_{Dc}$[a]	Plantscore	$Prob_{Dc}$	Vina	$Prob_{Dc}$	Mean $Prob_{Dc}$
2	−103.5	0.50	−82.78	0.70	−7.3	0.60	0.60
3	−88.14	0.43	−82.24	0.70	−7.2	0.60	0.57
5	−89.88	0.44	−68.66	0.58	−8.1	0.675	0.56
8	−124.1	0.60	−80.00	0.68	−6.2	0.51	0.60
9	−97.84	0.48	−79.03	0.67	−7.4	0.61	0.59
12	−121.2	0.59	−91.58	0.78	−6.6	0.55	0.64
18	−108.3	0.53	−88.22	0.75	−7	0.58	0.62
19	−96.69	0.47	−78.88	0.67	−8.1	0.67	0.60
20	−96.77	0.47	−74.36	0.63	−7.2	0.60	0.56
22	−99.31	0.48	−91.35	0.77	−7.3	0.60	0.62
23	−82.40	0.40	−71.74	0.61	−7.2	0.60	0.53
27	−106.5	0.52	−86.27	0.73	−6.6	0.55	0.60
30	−118.7	0.58	−100.1	0.85	−8.6	0.71	0.71
32	−85.79	0.42	−75.08	0.63	−8.5	0.70	0.58
33	−203.6	1	−117.4	1	−8	0.66	0.88
35	−107.0	0.52	−82.86	0.70	−7.2	0.60	0.61
37	−89.37	0.43	−72.98	0.62	−8.2	0.68	0.58
41	−94.52	0.46	−76.22	0.64	−7.3	0.60	0.57
43	−116.7	0.57	−107.4	0.91	−8.5	0.70	0.73
44	−87.29	0.42	−75.81	0.64	−7.3	0.60	0.56
Inhibitor	−80.67	0.39	−77.65	0.66	−7.5	0.62	0.56

[a]Probability of a compound being active in the molecular coupling analysis.

the neolignans were not a substrate for the CYP enzyme and did not cross the blood-brain barrier.

Toxicity was also evaluated, and we observed that only nine neolignans showed low or high toxicity in at least one parameter evaluated, such as mutagenicity or tumorigenesis, negative effects on the reproductive system, and irritability. Therefore, 80.85% of neolignans were considered to have the best ADMET (absorption, distribution, metabolism, excretion, and toxicity) properties, as they do not present toxicity risks. Tables S1-S4, in Supplementary Information (SI) section, shows the ADMET profile of the entire set of neolignans.

Interaction analysis

Several neolignans in this study obtained promising binding energy values, biological activity, combined analyses, pharmacokinetics and pharmaco-chemical properties, and low toxicity. We chose to analyze the interactions of the two compounds (**33** and **43**) that stood out in all these properties and of the compounds that were isolated (**46** and **47**) that obtained good results in docking (Figures 2 to 4).

Cruzain

Cruzain is the main cysteine protease in *T. cruzi*, and is essential at all stages of the parasite's development. Overexpression of cruzaine increases the transformation of the parasite into an infectious form and is therefore considered an attractive target in drug planning.[15]

Compound **33** formed a hydrophobic interaction with Leu37, a steric interaction with Ala133 and three hydrogen bonds with the amino acids Trg59, Asp60 and Asp158. While compound **43** formed only two hydrogen bonds with the amino acids Ser61 and Asn70, in addition to several hydrophobic interactions with residues Met68, Gly65, Leu67, Gly66 and Ala133. Compound **47** showed a hydrogen bond with the amino acid Gly66, a van der Waals interaction with the amino acid Gly65 and hydrophobic interactions with the amino acids Leu67, Ala133, Asp156 and Leu157. According to Durrant *et al*.,[24] an irreversible cruzain inhibitor, benzoyl-Tyr-Ala, could interact similarly to neolignans with a cruzain active site. According to the authors, interactions with Met68, Leu67 and Glu205 were important to inhibit the active of the enzyme's active.

Figure 2. 3D and 2D interactions of neolignan **33**, **43** and **47** with the cruzain enzyme. Hydrogen bonds are highlighted in green, hydrophobic interactions are highlighted in pink, and electrostatic interactions are highlighted in red.

TR

TR, found in the epimastigote and trypomastigote forms of the parasite, is a key enzyme in redox metabolism and is essential for trypanosomes. This enzyme is absent in humans, replaced by glutathione and glutathione reductase, offering a target for selective inhibition.

Compound **33** formed three strong hydrogen bonds with residues Gly14, Ser15 and Glu19. Hydrophobic interactions were also observed with the amino acids Cys53, Tyr111 and Met114. Compound **43** showed more hydrogen bonds

compared to compound **33**, forming notable bonds with the amino acids Ser15, Glu19, Ile107, Ser110, Tyr111 and Met114. Only a weaker interaction was observed with the amino acid Ile339. Compound **46** formed an interaction with the amino acid Thr335 and several hydrophobic interactions with the amino acids Gly57, Cys58, Lys61, Ile200, Phe204 and Asp327. Compound **47** formed more stable bonds with amino acids Ser110, Tyr111, Met114 and hydrophobic interactions with amino acids Ser15 and Glu19. Compound **47** formed more stable bonds with amino

Figure 3. 3D and 2D interactions between neolignans **33**, **43**, **46** and **47** with the TR enzyme. Hydrogen bonds are highlighted in green, hydrophobic interactions are highlighted in pink, and electrostatic interactions are highlighted in red.
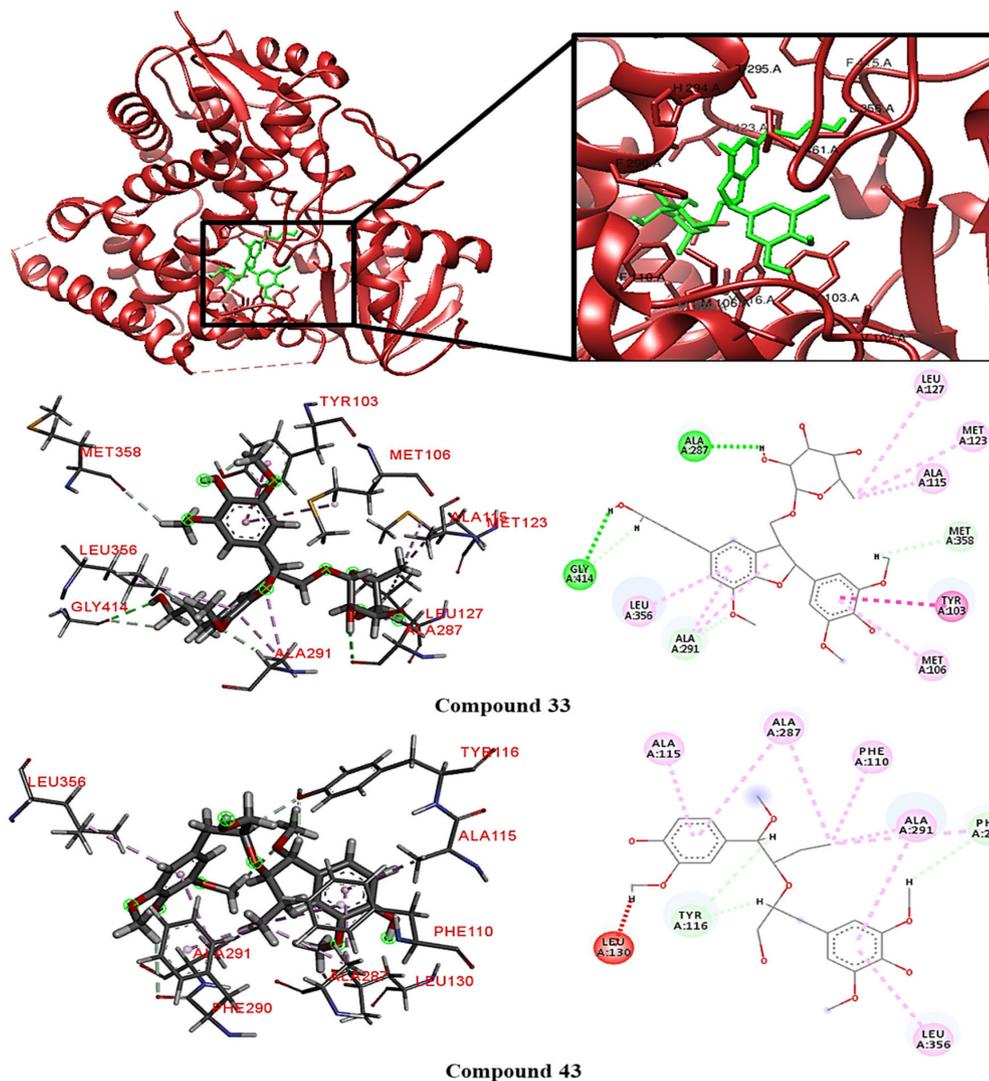
Figure 4. 3D and 2D interactions between neolignan **33** and **43** with the CYP51 enzyme. Hydrogen bonds are highlighted in green, hydrophobic interactions are highlighted in pink, and electrostatic interactions are highlighted in red.

acids Ser110, Tyr111, Met114 and hydrophobic interactions with amino acids Ser15 and Glu19.

A study by Saravanamuthu *et al.*,[16] showed that a *T. cruzi* inhibitor of TR formed several interactions also observed with neolignans and the active site of TR. Notable interactions include those with Glu19, Trp22, Cys53, Ser110, Tyr111, Asp117 and Leu399 stand out.

CYP51

The CYP51 enzyme is involved in a key step in ergosterol biosynthesis, responsible for the oxidized demethylation of intermediate sterols through the heme group. It is essential for the parasite's survival, development and proliferation, which is why it is present in all cellular forms.[17]

Compound **33** formed four strong hydrogen bonds with the amino acids Ala287, Ala291, Met358 and Ala414 from the active site of CYP51. It also showed several

hydrophobic interactions with the residues Tyr103, Met106, Ala115, Met123, Leu127 and Leu356. While compound **43** showed hydrophobic interactions with residues Phe110, Ala115, Ala287 and Leu356. Two hydrogen bonds with the amino acids Tyr116 and Phe290 and a steric interaction with the amino acid Leu130.

Prediction of ADMET properties

Molecular dynamics simulations

After the virtual screening and the analysis of the activity potential of several neolignans against important *T. cruzi* enzymes, we conducted molecular dynamics simulations with the two compounds that we were able to isolate (compounds **46** and **47**) to assess the flexibility and stability of the enzymes and their interactions in the presence of factors such as solvent, ions, pressure,

and temperature. This information is important because it complements the docking results and allows us to evaluate if the compounds remain strongly linked to the studied enzymes in the presence of factors found in the host organism. We chose the TR enzyme for this analysis, because neolignans were more selective for this protein. Then, the RMSD was calculated for the Cα atoms of the complexed enzyme and the structures of each ligand, separately.

The RMSD analysis of the TR enzyme complexed with the crystallographic ligand showed conformations ranging from 0.35 to 0.45 nm in size for 50 ns, with high stability (Figure 5). The same pattern was observed for the enzyme complexed with the neolignans. The stability of this protein is essential to keep compounds bound to the active site.
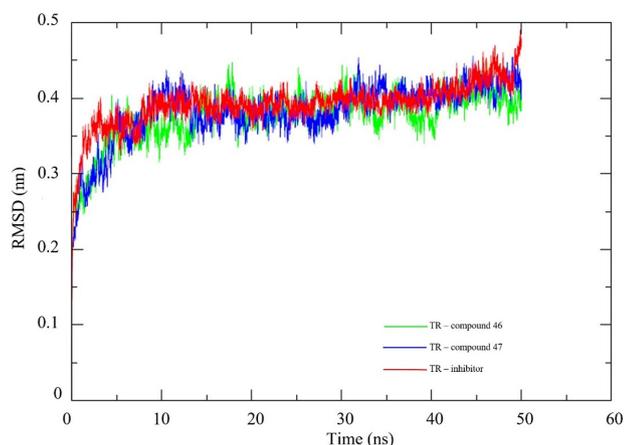
Figure 5. RMSD values for the Cα atoms of the TR enzyme complexed with neolignans and the Protein Data Bank (PDB) ligand.

When we analyzed the flexibility of the ligands, we found that the crystallographic ligand was drastically more unstable than the neolignans during the entire dynamics simulation (Figure 6). Therefore, in the presence of solvents, ions and other factors, neolignans can establish stronger bonds with the active site.

To understand the flexibility of the residues and amino acids that contribute to the conformational changes in the TR enzyme, the mean quadratic fluctuation (RMSF) values were calculated for each amino acid in each enzyme. High RMSF values suggest greater flexibility, while low RMSF values reflect less flexibility. Since amino acids with fluctuations above 0.3 nm contribute to the flexibility of the protein structure, we found that residues at positions 1, 80-90, 460-462, and 486 contribute to conformational changes in the TR enzyme (Figure 7). We also found that none of the amino acids affecting the structural conformations identified in TR were active site components. This helps the neolignans to remain in the active site.
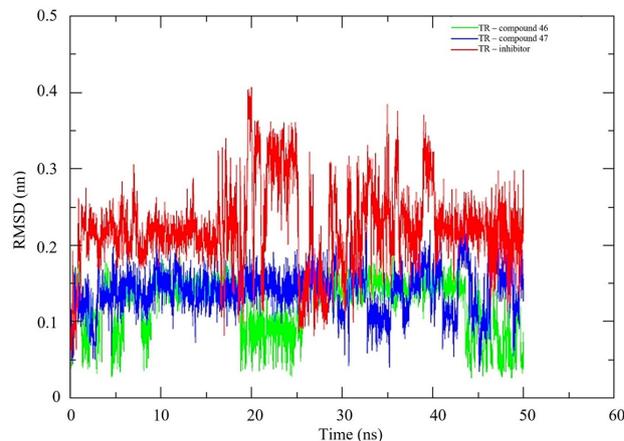
Figure 6. The RMSD values of the Cα atoms of the neolignans and the PDB ligand.

Through graphic programs of molecular modeling, it was possible to analyze 2D interactions at different times during the simulation of molecular dynamics (Figure 8). We noticed that most of the interactions observed in docking, were also observed in the dynamics simulations, that is, even in the presence of solvent and ions. Among the observed interactions, Val58, Ile106, Tyr110 and Met113 are notable.
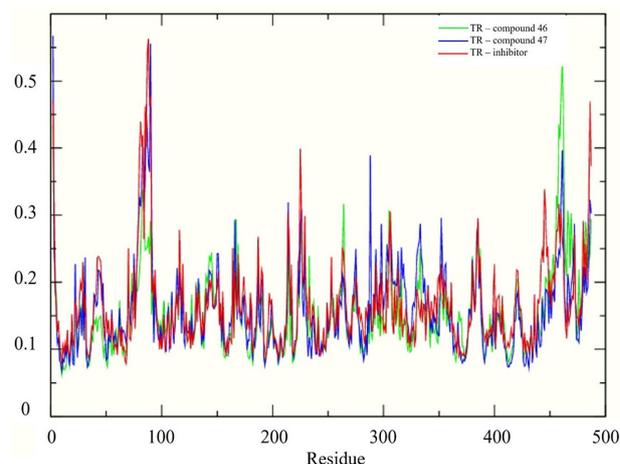
Figure 7. Root-mean-square fluctuation (RMSF) for the Cα atoms of the TR enzyme complexed with the neolignans and the PDB ligand.

Free energy calculations

The molecular mechanics - Poisson Boltzmann surface area approach (MM/PBSA) method was used to explore the arrangement of the interactions further and estimate the free energy after the molecular dynamic (MD) simulation. As seen in Table 6, the crystallographic ligand had superior free binding energy to the neolignans that obtained the best results in docking and predicting of biological activity. Although lignans **46** and **47** showed higher values of binding affinity in the docking than the
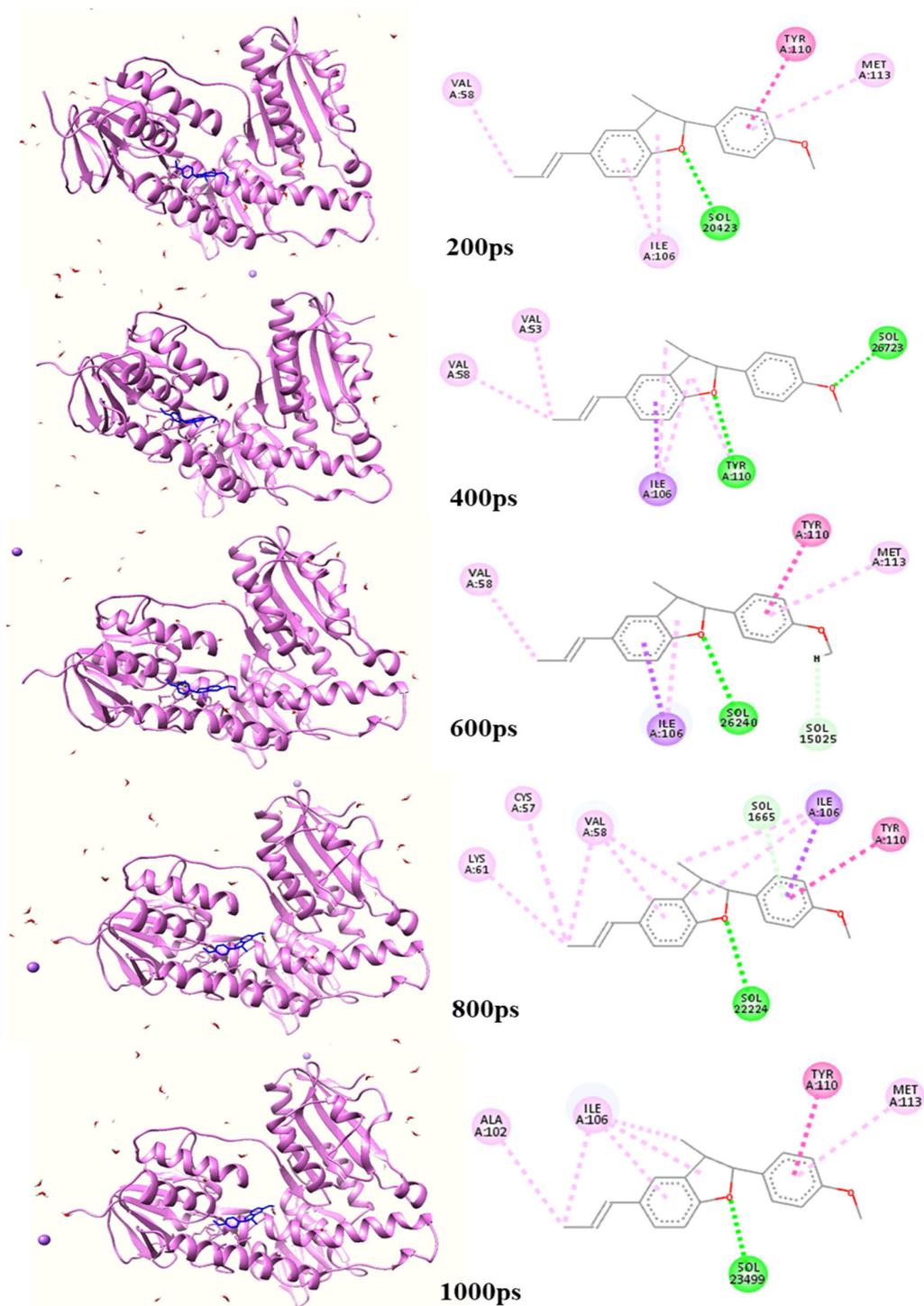
Figure 8. 3D structures and 2D interactions of compound **46** at different times during dynamics simulation.

Table 6. Results of free energy calculations using the MM/PBSA method

ID	ΔE_{wdw} / (kJ mol^{-1})	$\Delta E_{Electrostatic}$ / (kJ mol^{-1})	ΔG_{PB} / (kJ mol^{-1})	ΔG_{SA} / (kJ mol^{-1})	ΔG_{MMPBSA} / (kJ mol^{-1})
46	−99.24 ± 13.60	−9.39 ± 6.847	43.20 ± 15.78	−12.041 ± 1.43	−77.46 ± 14.73
47	−108.26 ± 14.39	−4.73 ± 3.46	42.03 ± 13.70	−13.23 ± 1.61	−84.20 ± 11.98
Inhibitor	−151.46 ± 29.15	−139.33 ± 59.85	112.97 ± 89.26	−14.48 ± 2.92	−192.31 ± 37.04

ΔE_{wdw}: summation of van der Waals interaction energies; $\Delta E_{Electrostatic}$: summation of eletrostatic interaction energies; ΔG_{PB}: polar solvation energy; ΔG_{SA}: nonpolar free solvation energy; ΔG_{MMPBSA}: molecular mechanics - Poisson Boltzmann surface area approach energy.

crystallographic ligand, only with the results obtained by the MM/PBSA calculations it was it possible to verify that in the presence of a solvent, the inhibitor provided better performance than the neolignans. When analyzing energy contributions, we note that the value of electrostatic and van der Waals energies favored the inhibitor's increased free energy. However, although these results are excellent for the inhibitor, they do not indicate that neolignans did not produce activity, as hydrogen bonds are strong and are not evaluated by the MM/PBSA method.

Activity of two neolignans against epimastigotes and trypomastigotes *T. cruzi*

We isolated two neolignans that obtained excellent results during the virtual screening process. The compounds were evaluated for the potential to inhibit the growth of epimastigotes and trypomastigotes. The results showed that the neolignans (2*R*,3*R*)-2,3di-hydro-2-(4-methoxyphenyl) -3-methyl-5-(*E*)-propenylbenzofuran (**46**) and ottomentosa (**47**) were able to potentially inhibit both cellular forms of *T. cruzi*, with inhibitory activity superior to benznidazole (Table 7). The best results were against trypomastigote forms, which correspond to the infectious phase of CD. The ottomentosa neolignan (**47**) proved to be lightly more potent than the neolignan (2*R*,3*R*)-2,3di-hydro-2-(4-methoxyphenyl)-3-methyl-5-(*E*)-propenylbenzofuran (**46**). Therefore, both substances have trypanocidal activity and do not show cytotoxicity to renal cells.

Some previous studies have demonstrated the trypanocidal activity of neolignans, but this is the first report that describes the activity of these two neoligans as trypanocidal. A study by Cabral *et al.*[13] showed that two neolignans, licarin A and burchellin, could inhibit the growth of the epimastigote by 45 and 20%, respectively. The authors also found that lycarin A and burchellin could induce trypomastigote death with IC$_{50}$ (represents the concentration required for 50% inhibition of the parasites)/24 h of 960 and 520 µM, respectively. Pelizzaro-Rocha *et al.*[25] demonstrated that the neolignan eupomatenoid-5 exhibited activity against trypomastigotes, the infective form of *T. cruzi* half maximal effective concentration (EC$_{50}$ 40.5 µM), leading

to ultrastructural alteration and lipoperoxidation in the cell membrane. In addition, they reported that the trypanocidal action of eupomatenoid-5 might be associated with mitochondrial dysfunction and oxidative damage, which can trigger destructive effects on the biological molecules of *T. cruzi*, leading to the death of the parasite. Ferreira *et al.*[26] used a semi-synthetic library of 23 derivatives of the neolignan dehydrodieugenol B that was prepared to explore synthetically accessible activity structure (SAR) relationships against *T. cruzi*. Five compounds demonstrated activity against trypomastigotes (IC$_{50}$ values from 8 to 64 µM) and eight showed activity against intracellular amastigotes (IC$_{50}$ values from 7 to 16 µM).

Conclusions

We used a comprehensive computational studies approach to investigate the potential of neolignans in the treatment of CD, which made possible the isolation and experimental testing of natural products against cellular forms of *T. cruzi*. The predictive models generated from essential enzymes of the parasite obtained satisfactory performance results for the continuity and credibility of this study, with an accuracy greater than 75%, and selected a neolignan with a 50% probability of active potential for the cruzain enzyme. For the TR enzyme, an accuracy of 85% was achieved, and the model selected all neolignans, with activity probabilities between 54 and 84%. Therefore, neolignans were considered selective against the TR enzyme.

For a structure-based investigation, a consensus docking analysis was conducted to ensure the reliability of the RF model and to reduce the number of false positives. Among the 47 neolignans analyzed by molecular docking, two compounds were considered potentially active against cruzain, 31 neolignans active against TR, and 18 against CYP51. These results indicated that neolignans, in general, are more likely to activate TR and CYP51 proteins, and are not selective for the cruzain enzyme.

A structure and ligand-based combined analysis, employed to increase the predictive power was able to identify potentially active molecules, using RF models and molecular docking, resulting in the identification of 22 compounds with a probability of activity ranging from

Table 7. Effect on *Trypanosoma cruzi* and renal cells

ID	Epimastigote IC$_{50}$ / µM	Trypomastigote IC$_{50}$ / µM	Cytotoxicity (cells LLC-MK2) / µM
46	9.64 ± 1.75	4.88 ± 0.61	> 400
47	8.72 ± 1.14	2.73 ± 0.58	> 400
Benznidazole	16.58 ± 2.9	257.5 ± 14.32	> 400

IC$_{50}$: concentration required for 50% inhibition of the parasites.

50-65% for cruzain and 46 neolignans potentially active against the TR enzyme with a probability ranging from 52-82%. We found that the combined analysis expanded the selection of active compounds for cruzain relative to the RF model and molecular docking. We also found that neolignans were more selective for the TR enzyme.

MD simulations revealed that neolignan-complexed RT was stable under several conditions, including solvent, ions, temperature, and pressure, with only small variations observed for some complexed compounds. Therefore, the binding affinity between proteins and ligands is unlikely to be affected by environmental changes. In addition, none of the amino acids responsible for the enzymatic conformational changes were at the active site, allowing the active site to remain stable. In addition, through calculations of free energy using the MM/PBSA method, we found that although the crystallographic ligand presented a higher energy bond value in the presence of factors, it is notable that strong hydrogen bonds also favor the permanence of the neolignans at the active site of TR.

Two neolignans with excellent ADMET profiles demonstrated to be potentially active inhibitors against the enzymes cruzain and TR enzymes by virtual screening, were isolated from *Krameria tomentosa* and subjected to *in vitro* tests. The results showed that the two neolignans (**46** and **47**) could potentially inhibit *T. cruzi*, at concentrations of 9.64 and 8.72 μM for the epimastigote forms and 4.88 and 2.73 μM for the trypomastigote forms, respectively. Therefore, the compounds (2R,3R)-2,3-di-hydro-2-(4-methoxyphenyl)-3-methyl-5-(E)-propenylbenzofuran (**46**) and ottomentosa (**47**) proved to be promising inhibitors of growth for the epimastigote and promastigote stages of the parasite.

We also concluded that the neolignans investigated in this study that were considered active against the enzymes cruzain and CYP51 could be potent inhibitors of these enzymes in amastigotes, since these enzymes are present in all cellular forms of *T. cruzi*. Therefore, the neolignans selected in this study serve as a starting point for the development of new antichagasic compounds and should be investigated for their potential activity against the most relevant parasitic form of the parasite, the amastigote form.

Experimental

Data collection and curation

The biological activity and 3D structure data of the enzymes cruzain, trypanothione reductase (TR), and sterol 14-alpha demethylase (CYP51) were investigated. Datasets with information on compounds and their activity values

for each selected enzyme (Table 8) were downloaded from the ChEMBL database[18,19] with the codes CHEMBL3563 (cruzain), CHEMBL5131 (TR) and CHEMBL1075110 (CYP51). These compounds were used to build predictive models and were classified based on the pIC_{50} ($-\log IC_{50}$). The compounds of each dataset were classified based on the pIC_{50}, this information is described in Table 8. The IC_{50} value represents the concentration required for 50% inhibition of the parasites. The classification of compounds based on pIC_{50} is performed in order to have the necessary number of samples to create a pattern and increase the probability of the model being correct in selecting active molecules, that is, the higher the pIC_{50}, the lower the concentration of the chemical compound and consequently more potent is its activity. However, it is necessary to have a sufficient number of samples so that the model can distinguish active from inactive compounds. Figure 9 summarizes all the procedures performed in this study.
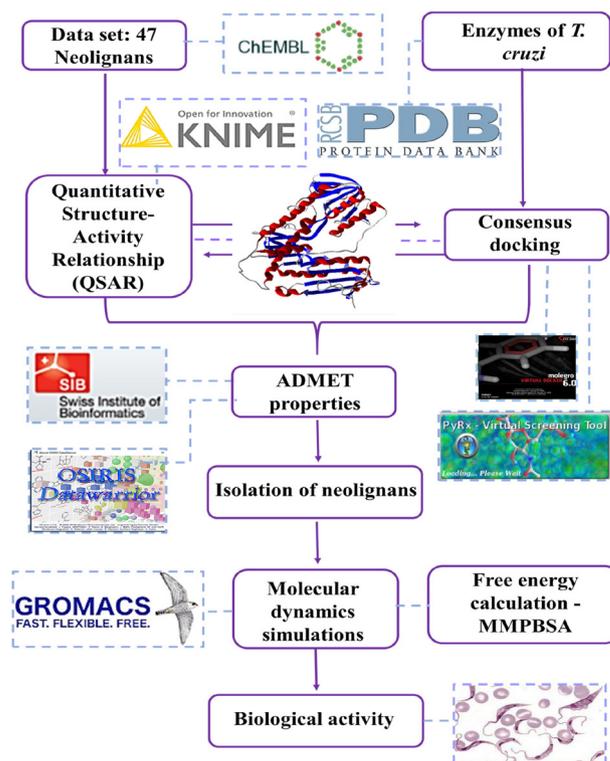
Figure 9. Outline of all study procedures.

In addition, a search was performed in the ChEMBL database for lignans extracted from natural products. A total of 47 lignans were found, therefore, these were evaluated by virtual screening to identify molecules with potential activity against the three main enzymes listed above that are involved in the proliferation and survival of *T. cruzi*, according to the workflows presented by Fourches *et al.*[27] Three-dimensional structures were generated and

standardized using the Standardizer v.18.17.0.[28] This standardization is of paramount importance to create consistent compound libraries and is done through the following steps: addition of hydrogens, aromatization, generation of 3D structure (clean the molecular graph in three dimension) and exporting the compounds in SDF format. For a more detailed description on how the dataset was curated, please refer to the workflows described by Fourches *et al.*[27,29,30] Codes for the structures of neolignans are available in the Table S1, SI section.

Quantitative structure-activity relationship modeling (QSAR)

Knime 3.6.2 software (Knime 3.6.2, Konstanz Information Miner, Zurich, Switzerland)[31,32] was used to perform QSAR modeling. Given the success of our previous studies,[33,34] we opted to perform a QSAR 3D analysis. To generate descriptors, all compounds with their standardized chemical structure were saved in SDF format and imported into Dragon 7.0 software (Kode Chemioinformatics SRL, Pisa, Italy), a total of 5270 molecular descriptors.[35,36] The descriptors generated in Dragon were imported into the Knime software and the random forest algorithm (RF) was used to build prediction models.

Each dataset was divided using the "Partitioning" tool, with the "stratified sample" option, to create a training set and an external test set, which represented 80 and 20% of the compounds, respectively. Although the compounds were selected randomly, the same proportion of active and inactive samples was maintained in both sets.

For external validation, we employed 5-fold cross-validation using randomly selected, stratified groups. The distributions according to activity class variables were found to be maintained in all validation groups and in the training set. Descriptors were selected and modeled following a 5-fold external cross validation procedure using the random forest algorithm (RF).[37-39] The 5-fold external validation technique used in this work is a technique that the entire data set is partitioned five times into a modeling set (training set) including 80% of the compounds the set, and the external cross validation data set, comprising the remaining 20% of the compounds the data set.[37-39] After this, only the modeling set is used to build the models

and then the models are validated with the external cross validation technique.[37-39] There were 200 the total number of trees constructed and 1 seed in the generation of random numbers for the RF for all generated models.

Using Knime nodes the most important descriptors in the generation of each prediction model were evaluated. The external performances of the selected models were analyzed for sensitivity (true positive rate, i.e., active rate), specificity (true negative rate, i.e., inactive rate) and accuracy (overall predictability). The positive (PPV) and negative (NPV) predictive values inform us about the probability of predicted positives (PPV) and negatives (NPV) being the true positives and negatives, respectively. In addition, the sensitivity and specificity of the receiver operating characteristic (ROC) curve were found to describe true performance with more clarity than accuracy.[39]

The model was also analyzed by the Matthews coefficient, a way to evaluate the model globally from the results obtained from the confusion matrix. The Matthews correlation coefficient (MCC) is a correlation coefficient between observed and predictive binary classifications. It results in a value between −1 and +1, where a coefficient of +1 represents a perfect forecast, 0 is nothing more than a random forecast, and −1 indicates total disagreement between forecast and observation.[40]

The MCC can be calculated from the following formula:

$$MCC = \frac{VP \times VN - FP \times FN}{\sqrt{(VP + FP)(VP + FN)(VN + FP)(VN + FN)}} \quad (3)$$

where VP is the value of true positive, VN is the value of true negative, FP is the value of false positives and FN of false negatives.

The applicability domain (APD) was used to analyze the compounds of the test sets to evaluate whether their predictions were reliable. The APD is based on Euclidean distances, and similarity measures between the descriptors of the training set are used to define the applicability domain. Meaning, if a test set compound has distances and similarity beyond this limit, its prediction is not reliable. The APD calculation is performed behind the formula:

$$APD = d + Z\sigma \quad (4)$$

Table 8. Set of molecules from the ChEMBL Databases for each cruzain, TR and CYP51 database of *T. cruzi*

Database	Active molecules	Inactive molecules	Total	ChEMBL ID
Cruzain	217 ($pIC_{50} \geq 5.02$)	255 ($pIC_{50} \leq 5.0$)	472	CHEMBL3563
TR	117 ($pIC_{50} \geq 4.5$)	150 ($pIC_{50} \leq 4.46$)	267	CHEMBL5131
CYP45	271 ($pIC_{50} \geq 4.6$)	317 ($pIC_{50} \leq 4.5$)	588	CHEMBL1075110

TR: trypanothione reductase; IC_{50}: concentration required for 50% inhibition of the parasites.

where d and σ are the Euclidean distances and the standard mean deviation, respectively, of the compounds in the training set. Z is an empirical cut-off value, and in this work the Z value was used as 0.5.[39,41-43]

Study of molecular docking

Molecular docking was used to investigate the mechanism of action of selected compounds against the enzymes cruzain (PDB ID 1AIM),[15] TR (PDB ID 1GXF)[16] and CYP51 (PDB ID 4CK9).[44] The 3D structures of the enzymes were obtained from the Protein Data Bank (PDB).[20] Information about complexed enzymes and inhibitors can be seen in Table 9. Initially, all water molecules were removed from the crystalline structure and the mean square deviation (RMSD) was calculated from the positions, indicating the degree of reliability of the adjustment. To evaluate the docking procedure, we performed the redocking and RMSD calculation. The RMSD provides the connection mode close to the experimental structure and is considered successful if the value is less than 2.0 Å. The inhibitory ligands crystallized in each enzyme were used as templates to signal the active site region of the protein.

Molegro virtual docker (MVD) 6.0

We used the Molegro Virtual Docker v.6.0.1 (MVD) software[45,46] with its predefined parameters. Then, a docking wizard was created in which the enzymes and ligands were inserted to analyze system's stability through the interactions associated with the enzyme's active site, using the energy value of the MolDock Score[45] as a reference. The MolDock SE (Simplex Evolution) algorithm is based on differential evolution and was used with the following parameters: a total of 10 runs with a maximum of 1,500 iterations using a population of 50 individuals, 2,000 minimization steps for each flexible residue and 2,000 steps of global minimization *per* run. The MolDock Score (GRID) and PLANTS score (GRID) scoring function were used to calculate the fit energy values. A GRID was set at 0.3 A and the search sphere radius was set at 15 Å. For the analysis of the ligand energy, electrostatic interactions, hydrogen bonds and sp2-sp2 torsions were evaluated.

AutoDock Vina (Vina)

We used AutoDock Vina[47,48] under the graphical interface of the PyRx Virtual Screening program tool,[49,50] maintaining the default parameters of the software. This

Table 9. Inhibitor-enzyme complex data for *T. cruzi* enzymes cruzain, TR, and CYP51

Enzyme	PDB ID	Ligand	Resolution / Å
Cruzain	1AIM		2.0
TR	1GXF		2.7
CYP51	4CK9		2.7
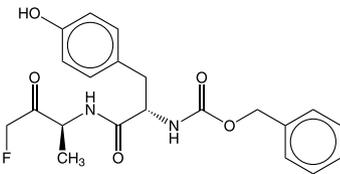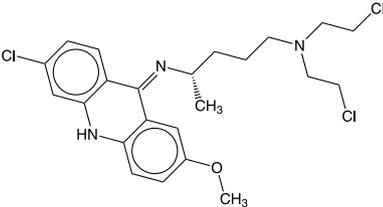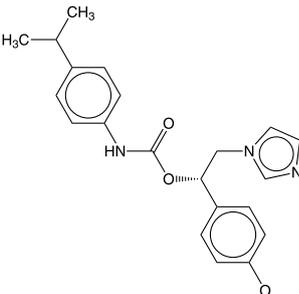
TR: trypanothione reductase; CYP51: sterol 14-alpha demethylase.

program is based on the genetic algorithm and the empirical energy of the field strength. The protein and ligand files were converted to pdbqt format, and the generated GRID was conducted to the active site region. AutoDock Vina generated 10 conformations for each binder which were used for analysis. The binding affinity scoring function (kcal mol^{-1}) corresponds to the sum of intermolecular and intramolecular contributions and the potentials are based on knowledge and empirical scores.

Consensus docking

A consensus analysis using three different scoring functions was used to decrease the number of false positives.

Prediction of absorption, distribution, metabolism, excretion, and toxicity (ADMET) properties

ADME parameters were calculated using the SwissADME open-access web tool,[51,52] which offers a set of rapid predictive models for the assessment of physicochemical, pharmacokinetic and pharmacological properties. The toxicity prediction was performed in the OSIRIS Property Explorer,[53,54] based on the following parameters: mutagenicity, tumorigenicity, reproductive effects and irritability. For absorption, factors included membrane permeability, intestinal absorption and the glycoprotein P substrate or inhibitor. Thus, we investigated compounds that did not exceed more than two violations of the Lipinski rule and for which the logP consensus was not greater than 4.15. In addition, the compounds were not substrates for the permeability glycoprotein enzyme (P-gp). The distribution was evaluated by factors that included the blood-brain barrier (logBB) and the permeability of the central neural system (CNS). Metabolism was predicted based on the CYP substrate or inhibition models (CYP1A2, CYP2C19, CYP2C9, CYP2D6 and CYP3A4).

Molecular dynamics simulations

MD simulations were performed to estimate the flexibility of interactions between proteins and ligands, using GROMACS 5.0 software (European Union Horizon 2020 Programme, Sweden).[55-57] The protein and ligand topologies were also prepared using the GROMOS96 54a7 force field. The MD simulation was performed using the SPC water model of point load, extended in a cubic box.[58] The system was neutralized by adding ions (Cl$^-$ and Na$^+$) and minimized, to remove bad contacts between complex molecules and the solvent. The system was also balanced at 300 K, using the 100 ps V-rescale algorithm, represented by NVT (constant number of particles, volume, and temperature), followed by equilibrium at 1 atm of pressure, using the

Parrinello-Rahman algorithm as the NPT (constant particles pressure and temperature), up to 100 ps. MD simulations were performed in 5,000,000 steps, at 50 ns. To determine the flexibility of the structure and whether the complex is stable and close to the experimental structure, RMSD values of all Cα atoms were calculated relative to the starting structures. RMSF values were also analyzed, to understand the roles played by residues near the receptor binding site. The RMSD and RMSF graphs were generated in Grace software[59] and the protein and ligands were visualized in UCSF Chimera.[60]

Free energy calculations

The Molecular Mechanics-Poisson Boltzmann Surface Area approach (MM/PBSA) was used to calculate the free binding energy of the protein-binding complex in the study of the molecular behavior of the sulfotransferase enzyme and its respective ligands. The GROMACS g_mmpbsa module[61,62] was applied to estimate the bond-free energy of the selected complex using the trajectory files obtained in the molecular dynamics simulation. The GROMACS MM-PBSA calculation consisted of three steps. First, the potential energy in the vacuum was calculated, and then, the energies of polar and, finally, nonpolar solvation were estimated. The non-polar solvation energy was calculated using the solvent accessible surface area model (SASA). The required input files and solvation energy values were then selected to evaluate the following energetic components: van der Waals energy, electrostatic energy, polar energy of solvation, non-polar solvation energy, and free energy of bonding.

Isolation and identification of neolignans

Among the neolignans investigated in this study, two were possible to isolate, (2*R*,3*R*)-2,3di-hydro-2-(4-methoxyphenyl)-3-methyl-5-(*E*)-propenylbenzofuran (compound **46**) and ottomentosa (compound **47**) (Figure 10). The neolignans were isolated from the roots of *Krameria tomentosa* (2.2 kg) collected in the city of Santa Rita-PB, Brazil (7°09'15.4" S, 35°00'14.9" W)

Compound **46**   Compound **47**
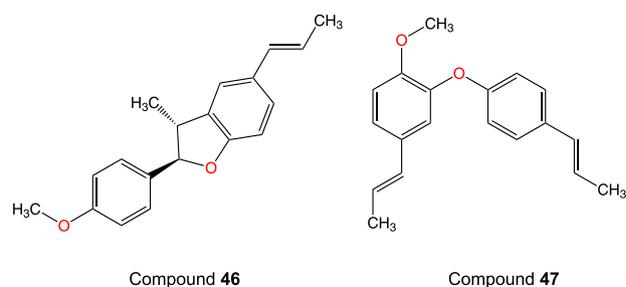
Figure 10. Isolated structures of the roots of *Krameria tomentosa*.

in June 2014. Their access registrations in the National Management System of Genetic Patrimony and Associated Traditinal Knowledge (SisGen) were obtained under number A837A80.

The air-dried and powdered roots of *K. tomentosa* (2.2 kg) were extracted with ethanol 95% for 72 h (3 × 4 L) at room temperature. The extract was concentrated under reduced pressure at 40° C to afford 380 g of crude extract (CE). Part of CE (50.0 g) was suspended in MeOH-H_2O (7:3) and sequentially partitioned with hexane, dichloromethane, ethyl acetate (EtOAc) and *n*-BuOH. An aliquot of the dichloromethane-soluble fraction (2.0 g) was subjected to a medium pressure liquid chromatography (MPLC) with silica gel 60 (40-63 μm, 230-400 mesh, SiliCycle) and eluted with a gradient mixture of hexane, $CHCl_3$, EtOAc and MeOH. The fractions obtained were combined according to the thin layer chromatography (TLC) profile to afford 14 fractions. The fraction 1 (20 mg) was purified by preparative high performance liquid chromatography (HPLC) with the following elution gradient: solvent A = H_2O-formic acid 0.1%; solvent B = MeOH; elution system = 0-80 min (70-100% B). This resulted in compounds **46** (2.1 mg)[63] and **47** (2.4 mg).[64] The equipment used was a Shimadzu apparatus with an SPDM10A VP diode array detector and an ACE C18 column (250 mm × 21.2 mm and 5 μm particles) and the flow rate was 8.0 mL min[-1].

The compounds were identified by 1D- and 2D-nuclear magnetic resonance (NMR), high-resolution mass spectrometry (HRESIMS), optical rotation analysis and by with examination the literature.[63-65]

Activity of selected neolignans against *T.cruzi* epimastigotes and tripomastigotes

Parasites and cells

Epimastigote forms of *T. cruzi*, strain Y, were cultured liver infusion tryptose medium supplemented with 10% fetal bovine serum (FBS) and 1% penicillin-streptomycin 10,000 IU *per* 10 mg and kept at 28 ℃ in a of biochemical oxygen demand incubator. The parasites used in the experiments were aliquoted from cultures in an exponential growth phase, determined using a 10-day growth curve.

Rhesus monkey renal cells LLC-MK2 (ATCC CCL-7) were cultured in Dulbecco's modified eagle medium (DMEM) medium, pH 7.2, supplemented (10% FBS and 1% 100 U mL[-1] penicillin and 100 μg mL[-1] streptomycin) and incubated at 37 °C and 5% CO_2. Maintenance was performed every two days or when the cells reached the confluence.

Trypomastigote forms were obtained by infection of LLC-MK2 cells. The cells were cultured (2 × 10[5]) in DMEM medium supplemented with 10% FBS and 1% penicillin-streptomycin 10,000 IU *per* 10 mg maintained at 37 ℃ and 5% CO_2. After reaching the confluence state, they were infected with a suspension of epimastigotes (1 × 10[7]). The infected cells were kept in DMEM medium with 2% SFB at 37 ℃ and 5% CO_2 for six days. Finally, the trypomastigote forms were removed from the supernatant, centrifuged and available for testing.[66]

Activity assay against epimastigotes and trypomastigotes of *Trypanosoma cruzi*

Epimastigote and trypomastigote forms were seeded in 96-well plates (1 × 10[6]) containing different concentrations of substances (1.56 to 400 μM) and liver infusion tryptose (LIT) medium with 10% FBS for epimastigote forms and DMEM medium with 2% SFB for trypomastigote forms. Then, they were incubated for 48 h at 28 ℃ (epimastigotes) and 37 ℃ (trypomastigotes). Growth inhibition was then assessed by quantification in a Neubauer chamber. The positive control was the standard drug benznidazole (Sigma-Aldrich, St. Louis, MO, USA) at concentrations of 1.56 to 400 μM.[67]

Cytotoxicity

Cytotoxicity was evaluated in 96-well plates using the (3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) colorimetric assay. Approximately 1 × 10[6] LLC-MK2 cells *per* well were incubated in 100 μL of DMEM medium supplemented with FBS at 37 °C and 5% CO_2 for 4 h to determine cell adhesion. First, non-adherent cells were removed by washing with pure DMEM medium. Next, a new supplemented DMEM medium containing increasing concentrations of the substances (1.56 to 400 μM) was added and then incubated at 37 ℃ with 5% CO_2 for 48 h. After incubation, cytotoxicity was assessed by adding 10 μL of MTT (5 mg mL[-1]). Next, the supernatant was discarded, and the formazan crystals were dissolved by adding 100 μL of dimethyl sulfoxide (DMSO). Finally, the absorbance at 540 nm was measured using an ELISA plate reader.[67]

Supplementary Information

Supplementary data (table with the prediction of the ADMET properties of the neolignans) are available free of charge at http://jbcs.sbq.org.br as PDF file.

Acknowledgments

Author Contributions

Mayara S. Maia, Marcus T. Scotti was responsible for conceptualization; Mayara S. Maia, Rodrigo S. Andrade, Julyanne Maria S. de Sousa, Gabriela Cristina S. Rodrigues, Renata Priscila B. de Menezes, Marcelo S. da Silva, Josean F. Tavares, Klinger Antonio F. Rodrigues and Marcus T. Scotti for methodology; Gabriela Cristina S. Rodrigues, Natália F.de Sousa, Renata Priscila B. de Menezes, Luciana Scotti, Marcus T. Scotti for validation; Mayara S. Maia, Rodrigo S. Andrade and Julyanne Maria S. de Sousa for investigation; Marcus T. Scotti, Rodrigo S. Andrade, Marcelo S. da Silva, Josean F. Tavares, Klinger Antonio F. Rodrigues for resources; Mayara S. Maia and Gabriela Cristina S. Rodrigues for data curation; Mayara S. Maia, Natália F. de Sousa, Renata Priscila B. de Menezes, Luciana Scotti, Marcus T.Scotti for writing-review and editing; Mayara S. Maia, Marcus T. Scotti for supervision; Marcus T. Scotti for project administration.

References

1. Lidani, K. C. F.; Andrade, F. A.; Bavia, L.; Damasceno, F. S.; Beltrame, M. H.; Messias-Reason, I. J.; Sandri, T. L.; *Front. Public Heal.* **2019**, *7*, 1. [Crossref]

2. Gasparim, A. Z.; Fontes, C. E. R.; Rossoni, D. F.; Toledo, M. J. O.; *Rev. Soc. Bras. Med. Trop.* **2018**, *51*, 225. [Crossref]

3. Villalta, F.; Rachakonda, G.; *Expert Opin. Drug Discovery* **2019**, *14*, 1161. [Crossref]

4. Ferri, G.; Edreira, M. M.; *Front. Cell. Infect. Microbiol.* **2021**, *11*, 634793. [Crossref]

5. Meymandi, S.; Hernandez, S.; Park, S.; Sanchez, D. R.; Forsyth, C.; *Curr. Treat. Options Infect. Dis.* **2018**, *10*, 373. [Crossref]

6. Rajão, M. A.; Furtado, C.; Alves, C. L.; Passos-Silva, D. G.; de Moura, M. B.; Schamber-Reis, B. L.; Kunrath-Lima, M.; Zuma, A. A.; Vieira-da-Rocha, J. P.; Garcia, J. B. F.; Mendes, I. C.; Pena, S. D. J.; Macedo, A. M.; Franco, G. R.; de Souza-Pinto, N. C.; de Medeiros, M. H. G.; Cruz, A. K.; Motta, M. C. M.; Teixeira, S. M. R.; Machado, C. R.; *Environ. Mol. Mutagen.* **2014**, *55*, 309. [Crossref]

7. Vinuesa, T.; Herráez, R.; Oliver, L.; Elizondo, E.; Acarregui, A.; Esquisabel, A.; Pedraz, J. L.; Ventosa, N.; Veciana, J.; Viñas, M.; *Am. J. Trop. Med. Hyg.* **2017**, *97*, 1469. [Crossref]

8. Zhao, C.; Chen, J.; Shao, J.; Shen, J.; Li, K.; Gu, W.; Li, S.; Fan, J.; *J. Agric. Food Chem.* **2018**, *66*, 10421. [Crossref]

9. de Arruda, C. C. P.; Hardoim, D. J.; Rizk, Y. S.; de Souza, C. S. F.; do Valle, T. Z.; Carvalho, D. B.; Taniwaki, N. N.; Baroni, A. C. M.; Calabrese, K. S.; *Molecules* **2020**, *25*, 37. [Crossref]

10. Salleh, W. M. N. H. W.; Tawang, A.; Jauri, M. H.; *Biointerface Res. Appl. Chem.* **2022**, *12*, 1791. [Crossref]

11. Kordbacheh, F.; Carruthers, T. J.; Bezos, A.; Oakes, M.; Du Fall, L.; Hocart, C. H.; Parish, C. R.; Djordjevic, M. A.; *PLoS One* **2018**, *13*, e0196843. [Crossref]

12. Sear, C. E.; Pieper, P.; Amaral, M.; Romanelli, M. M.; Costa-Silva, T. A.; Haugland, M. M.; Tate, J. A.; Lago, J. H. G.; Tempone, A. G.; Anderson, E. A.; *ACS Infect. Dis.* **2020**, *6*, 2872. [Crossref]

13. Cabral, M. M. O.; Barbosa-Filho, J. M.; Maia, G. L. A.; Chaves, M. C. O.; Braga, M. V.; de Souza, W.; Soares, R. O. A.; *Exp. Parasitol.* **2010**, *124*, 319. [Crossref]

14. Trefzger, O. S.; das Neves, A. R.; Barbosa, N. V.; Carvalho, D. B.; Pereira, I. C.; Perdomo, R. T.; Matos, M. F. C.; Yoshida, N. C.; Kato, M. J.; de Albuquerque, S.; Arruda, C. C. P.; Baroni, A. C. M.; *Chem. Biol. Drug Des.* **2019**, *93*, 313. [Crossref]

15. Gillmor, S. A.; Craik, C. S.; Fletterick, R. J.; *Protein Sci.* **1997**, *6*, 1603. [Crossref]

16. Saravanamuthu, A.; Vickers, T. J.; Bond, C. S.; Peterson, M. R.; Hunter, W. N.; Fairlamb, A. H.; *J. Biol. Chem.* **2004**, *279*, 29493. [Crossref]

17. França, R. R. F.; de Carvalho, A. S.; Branco, F. S. C.; Pinto, A. C.; Boechat, N.; *Rev. Virtual Quim.* **2014**, *6*, 1483. [Crossref]

18. Gaulton, A.; Bellis, L. J.; Bento, A. P.; Chambers, J.; Davies, M.; Hersey, A.; Light, Y.; McGlinchey, S.; Michalovich, D.; Al-Lazikani, B.; Overington, J. P.; *Nucleic Acids Res.* **2012**, *40*, 1100. [Crossref]

19. Bento, A. P.; Gaulton, A.; Hersey, A.; Bellis, L. J.; Chambers, J.; Davies, M.; Krüger, F. A.; Light, Y.; Mak, L.; McGlinchey, S.; Nowotka, M.; Papadatos, G.; Santos, R.; Overington, J. P.; *Nucleic Acids Res.* **2014**, *42*, 1083. [Crossref]

20. Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer Jr., E. F.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M.; *Arch. Biochem. Biophys.* **1978**, *185*, 584. [Crossref]

21. Barros, R. P. C.; Scotti, L.; Scotti, M. T.; *Curr. Top. Med. Chem.* **2019**, *19*, 900. [Crossref]

22. Plewczynski, D.; Łaźniewski, M.; Augustyniak, R.; Ginalski, K.; *J. Comput. Chem.* **2011**, *32*, 742. [Crossref]

23. Mandal, S.; Moudgil, M.; Mandal, S. K.; *Eur. J. Pharmacol.* **2009**, *625*, 90. [Crossref]

24. Durrant, J. D.; Keränen, H.; Wilson, B. A.; McCammon, J. A.; *PLoS Negl. Trop. Dis.* **2010**, *4*, e676. [Crossref]

25. Pelizzaro-Rocha, K. J.; Veiga-Santos, P.; Lazarin-Bidóia, D.; Ueda-Nakamura, T.; Dias Filho, B. P.; Ximenes, V. F.; Silva, S. O.; Nakamura, C. V.; *Microbes Infect.* **2011**, *13*, 1018. [Crossref]

26. Ferreira, D. D.; Sousa, F. S.; Costa-Silva, T. A.; Reimão, J. Q.; Torrecilhas, A. C.; Johns, D. M.; Sear, C. E.; Honorio, K. M.; Lago, J. H. G.; Anderson, E. A.; Tempone, A. G.; *Eur. J. Med. Chem.* **2019**, *176*, 162. [Crossref]

27. Fourches, D.; Muratov, E.; Tropsha, A.; *J. Chem. Inf. Model.* **2010**, *50*, 1189. [Crossref]

28. *ChemAxon*, Standardizer software, version 18.17.0; ChemAxon Ltd., 1998-2021.

29. Fourches, D.; Muratov, E.; Tropsha, A.; *Nat. Chem. Biol.* **2015**, *11*, 535. [Link]

30. Fourches, D.; Muratov, E.; Tropsha, A.; *J. Chem. Inf. Model.* **2016**, *56*, 1243. [Crossref]

31. Berthold, M. R.; Cebron, N.; Dill, F.; Gabriel, T. R.; Kötter, T.; Meinl, T.; Ohl, P.; Thiel, K.; Wiswedel, B.; *ACM SIGKDD Explor. Newsl.* **2009**, *11*, 26. [Crossref]

32. *Knime-Konstanz Information Miner*, v. 3.6.2; Zurich, Switzerland, 2019.

33. Maia, M. S.; e Silva, J. P. R.; de Lima Nunes, T. A.; de Sousa, J. M. S.; Rodrigues, G. C. S.; Monteiro, A. F. M.; Tavares, J. F.; da Franca Rodrigues, K. A.; Mendonça-Junior, F. J. B.; Scotti, L.; Scotti, M. T.; *Molecules* **2020**, *25*, 2281. [Crossref]

34. Maia, M. S.; de Sousa, N. F.; Rodrigues, G. C. S.; Monteiro, A. F. M.; Scotti, M. T.; Scotti, L.; *Comb. Chem. High Throughput Screen.* **2020**, *23*, 504. [Crossref]

35. Mauri, A.; Consonni, V.; Pavan, M.; Todeschini, R.; *Match Commun. Math. Comput. Chem.* **2006**, *56*, 237.

36. *Dragon - Software for Molecular Descriptor Calculation*, v. 7.0; Kode srl, Italy, 2006.

37. Fourches, D.; Pu, D.; Tassa, C.; Weissleder, R.; Shaw, S. Y.; Mumper, R. J.; Tropsha, A.; *ACS Nano* **2010**, *4*, 5703. [Crossref]

38. Cherkasov, A.; Muratov, E. N.; Fourches, D.; Varnek, A.; Baskin, I. I.; Cronin, M.; Dearden, J.; Gramatica, P.; Martin, Y. C.; Todeschini, R.; Consonni, V.; Kuz'min, V. E.; Cramer, R.; Benigni, R.; Yang, C.; Rathman, J.; Terfloth, L.; Gasteiger, J.; Richard, A.; Tropsha, A.; *J. Med. Chem.* **2014**, *57*, 4977. [Crossref]

39. de Menezes, R. P. B.; Viana, J. O.; Muratov, E.; Scotti, L.; Scotti, M. T.; *Curr. Issues Mol. Biol.* **2022**, *44*, 383. [Crossref]

40. Matthews, B. W.; *Biochim. Biophys. Acta, Protein Struct.* **1975**, *405*, 442. [Crossref]

41. Scotti, M. T.; Scotti, L.; Ishiki, H. M.; Peron, L. M.; de Rezende, L.; do Amaral, A. T.; *Chemom. Intell. Lab. Syst.* **2016**, *154*, 137. [Crossref]

42. Aptula, A. O.; Roberts, D. W.; *Chem. Res. Toxicol.* **2006**, *19*, 1097. [Crossref]

43. Gramatica, P.; *Int. J. Quant. Struct.-Prop. Relat.* **2020**, *5*, 61. [Crossref]

44. Friggeri, L.; Hargrove, T. Y.; Rachakonda, G.; Williams, A. D.; Wawrzak, Z.; Di Santo, R.; De Vita, D.; Waterman, M. R.; Tortorella, S.; Villalta, F.; Lepesheva, G. I.; *J. Med. Chem.* **2014**, *57*, 6704. [Crossref]

45. Thomsen, R.; Christensen, M. H.; *J. Med. Chem.* **2006**, *49*, 3315. [Crossref]

46. *Molegro Virtual Docker*, v. 6.0.1; CLC Bio Company, Denmark, 2013.

47. Trott, O.; Olson, A. J.; *J. Comput. Chem.* **2009**, *31*, 455. [Crossref]

48. *AutoDock Vina*, v. 1.2.0; Molecular Graphics Lab, The Scripps Research Institute, USA, 2009.

49. Dallakyan, S.; Olson, A. J. In *Chemical Biology*; Hempel, J. E.; Williams, C. H.; Hong, C. C., eds.; Humana Press: New York, NY, 2015, p. 243-250.

50. *PyRx - Virtual Screening Tool*, v. 0.9.7; Sarkis Dallakyan, USA, 2015.

51. Daina, A.; Michielin, O.; Zoete, V.; *Sci. Rep.* **2017**, *7*, 42717. [Crossref]

52. *SwissADME*; Molecular Modeling Goup, Swiss Institute of Bioinformatics, Switzerland, 2017.

53. Sander, T.; Freyss, J.; von Korff, M.; Reich, J. R.; Rufener, C.; *J. Chem. Inf. Model.* **2009**, *49*, 232. [Crossref]

54. *OSIRIS Property Explorer*; Thomas Sander, Idorsia Pharmaceuticals Ltd, Switzerland, 2009.

55. Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E.; *SoftwareX* **2015**, *1-2*, 19. [Crossref]

56. Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R.; *Comput. Phys. Commun.* **1995**, *91*, 43. [Crossref]

57. *Gromacs*, v. 5.0; Royal Institute of Technology and Uppsala University, Sweden, 2022.

58. Bondi, A.; *J. Phys. Chem.* **1964**, *68*, 441. [Crossref]

59. *Grace software*; Grace Development Team, 1996-2008.

60. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E.; *J. Comput. Chem.* **2004**, *25*, 1605. [Crossref]

61. Kumari, R.; Kumar, R.; Lynn, A.; *J. Chem. Inf. Model.* **2014**, *54*, 1951. [Crossref]

62. Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A.; *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 10037. [Crossref]

63. Achenbach, H.; Utz, W.; Usubillaga, A.; Rodriguez, H. A.; *Phytochemistry* **1991**, *30*, 3753. [Crossref]

64. Madeiro, S. A. L.; de Lucena, H. F. S.; Siqueira, C. D.; Duarte, M. C.; Braz-Filho, R.; Barbosa Filho, J. M.; da Silva, M. S.; Tavares, J. F.; *J. Braz. Chem. Soc.* **2012**, *23*, 2021. [Crossref]

65. Sawasdee, K.; Chaowasku, T.; Likhitwitayawuid, K.; *Molecules* **2010**, *15*, 639. [Crossref]

66. Aparicio, I. M.; Scharfstein, J.; Lima, A. P. C. A.; *Infect. Immun.* **2004**, *72*, 5892. [Crossref]

67. Lima, D. B.; Sousa, P. L.; Torres, A. F. C.; Rodrigues, K. A. F.; Mello, C. P.; de Menezes, R. R. P. P. B.; Tessarolo, L. D.; Quinet, Y. P.; de Oliveira, M. R.; Martins, A. M. C.; *Toxicon* **2016**, *120*, 128. [Crossref]