Article

Virtual Screening for the Selection of New Candidates to *Trypanosoma cruzi* Farnesyl Pyrophosphate Synthase Inhibitors

Larissa M. Oliveira,[a,b] Janay S. C. Araújo,[a,b] David B. Costa Júnior,[b] Manoelito C. dos Santos Junior,[a,b] Aníbal F. Santos Júnior*,[c] and Franco H. A. Leite[a,b]

[a]*Programa de Pós-Graduação em Ciências Farmacêuticas, Departamento de Saúde, Universidade Estadual de Feira de Santana (UEFS), 44036-900 Feira de Santana-BA, Brazil*

[b]*Laboratório de Modelagem Molecular, Departamento de Saúde, Universidade Estadual de Feira de Santana (UEFS), 44036-900 Feira de Santana-BA, Brazil*

[c]*Programa de Pós-Graduação em Ciências Farmacêuticas, Laboratório de Biofarmácia, Departamento de Ciências da Vida, Universidade do Estado da Bahia (UNEB), 41195-000 Salvador-BA, Brazil*

Chagas disease is a tropical parasitic disease that is caused by *Trypanosoma cruzi* and causes 12,000 deaths *per* year, mainly in Latin America. The available drugs for treating have severe limitations, including poor efficacy and high toxicity. One way to overcome these limitations is targeting priority molecules with computational tools to direct *in vitro* assays against validated targets. Farnesyl pyrophosphate synthase (E.C. 2.5.1.10) is an enzyme that participates in the initial stage of sterol biosynthesis, and its inhibition causes damage to membrane integrity, leading to parasite death. With the aim to identify potential inhibitors against this target from *T. cruzi*, hierarchical virtual screening approaches were performed through a combination of ligand-based pharmacophore models and molecular docking. First, pharmacophore model filtering resulted in 15,154 molecules that had the minimum structural requirements for inhibition (QFIT > 0). These molecules were subsequently submitted to molecular docking routine, which resulted in 11,762 molecules (Grid Score between –232.74 to –0.96 kcal mol^{-1}). The top 30 ranked molecules in these approaches were grouped in self-organizing maps. These analyses showed four promising compounds from natural products that mimic the major interactions present in the substrate/inhibitor, which indicates that these molecules can be assayed by *in vitro* experiments.

Keywords: farnesyl pyrophosphate synthase, molecular docking, pharmacophore model, *Trypanosoma cruzi*, virtual screening

Introduction

Chagas disease is a tropical disease that is caused by the protozoan *Trypanosoma cruzi*, which is transmitted to humans by a blood-sucking insect vector (triatomine bugs) commonly known as the "kissing bug". Other forms of transmission for this disease occur by blood transfusion and congenital transmission.[1]

This disease presents an initial acute phase, which may be characterized by fever, myalgia, headache, generalized lymphadenopathy, oedema, hepatosplenomegaly, cardiac inflammation, and in severe cases, meningoencephalitis.[2,3] This phase is followed by a chronic stage that is characterized by cardiac, digestive, nervous or other clinical changes and, in some cases, it is only observed years after *T. cruzi* infection.[4,5]

According to the World Health Organization,[6] Chagas disease is categorized among the neglected diseases, which are strongly associated with a population's poverty status and limited access to health services, and these diseases are not seen as needing investment in research and development by the pharmaceutical industry. Considering the global picture, there are approximately 6 to 7 million people infected with Chagas disease and approximately 12,000 deaths, primarily in Latin America.[6]

In Brazil, it is estimated that approximately 2 to 3 million people suffer from chronic disease, which is a major cause of heart pacemaker implants and heart

transplants as well as death among adults 30 to 60 years old.[7] There was an increase of approximately 47% in the number of heart transplants in Brazil between 2011 and 2014.[8] Chagas disease is the leading cause of heart failure in endemic areas and its third-highest leading cause in Brazil.[9]

The only drug available in Brazil is benznidazole (Rochagan®, Roche), although nifurtimox (Lampit®, Bayer AG) can be employed in other countries. However, these therapies are not effective at chronic stage and they have several adverse effects, such as skin rashes, paresthesia, asthenia, and skin peeling, which can hinder patient adherence to the treatment.[10,11] For this reason, it is necessary to develop new selective drugs with low rates of adverse effects.

One of the strategies used for the control of parasitic infections is the search for new validated targets. Among them, the sterol biosynthesis pathway is considered very promising for *T. cruzi* as well as other tripanossomatids.[12-14] The farnesyl pyrophosphate synthase (FPPS) is noteworthy for being a rate limiting step of the isoprenoid synthesis, which is responsible for the condensation of isopentenyl and dimethylallyl diphosphates to form geranyl pyrophosphate and farnesyl pyrophosphate.[15] These precursors are essential for the formation of most isoprenoids, including sterols, and the inhibition of which results in changes in the integrity of the lipid bilayer of cells, leading to parasite death.[16]

For the search of new chemical structures, potent inhibitors (e.g. the pharmacophore model) and/or tridimensional (3D) structures of the biological target (e.g. molecular docking) can help identify compounds with chemical, electronic and steric properties that are essential for biological activity.[17] Due to chemical and structural diversity, the natural products have been widely recognized as an important source of new hit/lead through virtual screening.[18] This study aimed to prioritize promising molecules derived from natural products for the *T. cruzi* FPPS (TcFPPS) inhibition assays by pharmacophore models and docking approaches.

Experimental

Generation of pharmacophore models from TcFPPS inhibitors

A set of 25 TcFPPS inhibitors that were previously described in the literature according to their biological activity information (half maximal inhibitory concentration (IC_{50})) was selected for the generation and evaluation of pharmacophore model.[14,19,20] Thus, five molecules with $IC_{50} < 1$ µM were selected for pharmacophore models generation (training set; Figure 1) and the other compounds (n = 20) were employed in the evaluation steps (test set; Table S1, Supplementary Information (SI) section).

All molecules were drawn in MarvinSketch software 5.12.1[21] and then converted to 3D format in SYBYL-X 2.0 software.[22] Next, using SYBYL-X 2.0 software, the Gasteiger-Huckel atomic partial charges were added and the structures were energy minimized using Tripos Force Field (ε = 80.0; maximum of iterations = 50,000) with 1×10^{-3} kcal mol^{-1} convergence criterion.
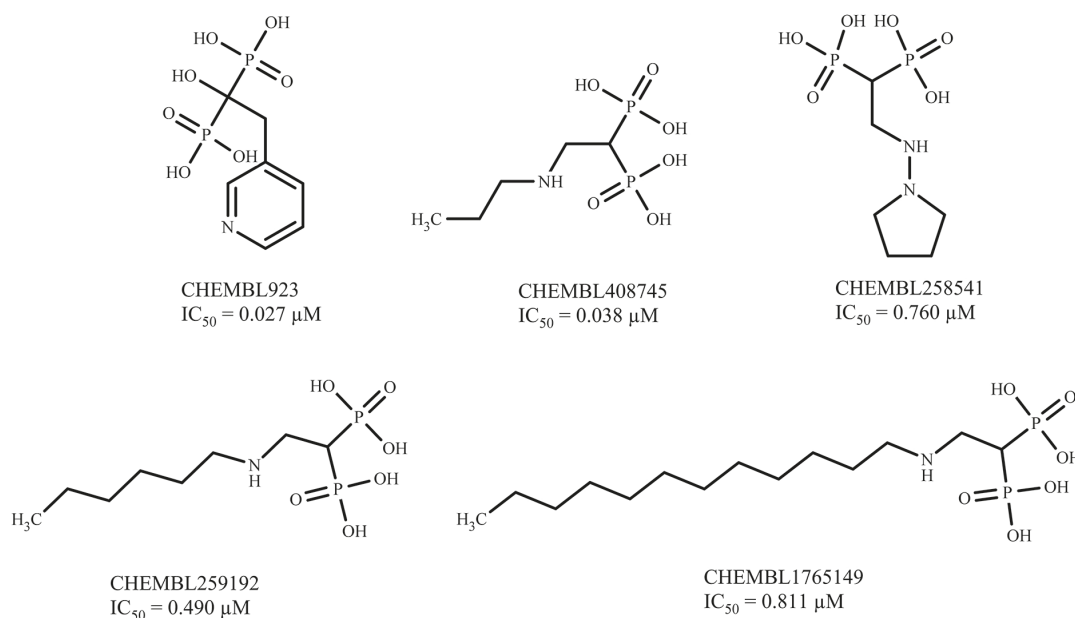

Figure 1. Chemical structures of training set with their respective biological activity (IC_{50}).

The GALAHAD module in the SYBYL 2.0 X platform[23] was used to construct 10 pharmacophore models. For this reason, the population size and the maximum number of generations for the FPPS inhibitors were adjusted to 35 and 80, respectively. The other parameters (mutation rate = 0.4, mutation decay = 1.0 and crossover rate = 1.0) were kept at their default values.

Pharmacophore models evaluation

First, the pharmacophore models with energy levels that were more than two orders of magnitude greater than the others were discarded. The remaining pharmacophore models were evaluated by Pareto score, by using the scoring components of the genetic algorithm (STERICS, HBOND and MOL_QRY). The pharmacophore models with Pareto values of zero were evaluated by their ability to differentiate true inhibitors from decoys, which were built with DUD-E server.[24]

Accordingly, the full database (one inhibitor: 50 decoys) was aligned to each pharmacophore model (template), using GALAHAD default parameters and ranked according to their QFIT values. Receiver operating characteristic (ROC) curves, Boltzmann-enhanced discrimination of ROC (BEDROC, $\alpha = 20$) and early enrichment upon the top 1% ($EF_{1\%}$, enrichment factor) were employed to validate their efficiency in discriminating true inhibitors from decoys.

The best pharmacophore model selected was investigated regarding their ability to explain the structure-activity relationships for congeneric molecules not employed in the generation of pharmacophore models.[25] Thus, 20 positive TcFPPS inhibitors that were previously described in the literature ($IC_{50} = 0.013\text{-}6.386\ \mu M$) were subjected to the same minimization protocol employed for the inhibitors that were used to generate the models.[14,19,20] Then, molecules were aligned individually onto pharmacophore models using the flexible alignment mode available in GALAHAD. The QFIT value was related to the biological activity values expressed in pIC_{50} and the bar graph plotted using SigmaPlot 12.0 software.[26]

Pharmacophore based virtual screening

The best pharmacophore model was used as template for 3D flexible alignment, through UNITY module available in SYBYL-X 2.0, of approximately 99,000 natural products available from ZINC database.[27] The quality of alignment of each molecule was expressed by QFIT value ranging from 0 to 100. Compounds with QFIT > 0 were selected for docking routine.

Molecular docking

Initially, 3D structure of the TcFPPS (PDB: 1YHL) was obtained from Protein Data Bank (PDB)[28] and prepared using DockPrep module in Chimera 1.10.1 software.[29] The selection was based on the resolution (1.95 Å) and was made in the presence of the dimethylallyl diphosphate (DMAPP) substrate. Artifacts water were removed and polar hydrogen atoms were added to the structure to optimize its hydrogen bonds. The protonation state was defined according to the optimum pH of catalysis (pH = 8.5)[12] using H ++ 3.0 web server to determine pK_a value.[30]

The docking routine was performed using the DOCK 6.5 program.[31] The delimitation of the search space was defined by DOCK 6.5 accessory programs. The molecular surface of the receptor was generated by Display Midas System (DMS) accessory program,[32] the negative image of the molecular surface of the orthosteric site (risedronate binding region) was constructed with an additional 8 Å radius with SPHGEN and SPHERE_SELECTOR accessory programs.[33] The molecular properties for that region were calculated using the GRID program in its default setting.[34,35] The docking was performed using Grid Score scoring function.[35]

To evaluate DOCK 6.5 parameters, an initial analysis was performed using root-mean-square deviation (RMSD) calculated between the coordinates of the crystallographic ligand and the best pose obtained by DOCK 6.5. An RMSD of less than 2 Å indicates good docking performance.[36] The following deviations were evaluated: RMSDs, interatomic distance between atoms; RMSDm, minimum distance between the primary atoms; and RMSDh, correction of symmetry between the poses of the ligands.[31]

For the evaluation on the discriminatory power of the DOCK 6.5 to recognize true inhibitors, it was used the same strategy adopted for the pharmacophore models stage. For the construction of the ROC curve, energy data that were obtained through the scoring function were used.

Prioritization of molecules through pharmacophore and energy criteria

Top 500 ranked molecules among pharmacophore and docking screening were directed to the cluster analysis by self-organizing maps to select molecules with the higher QFIT (group A) and Grid Score (group B) values.

Virtual screening by self-organizing maps

The type of intermolecular interaction to the organization (hydrogen, electrostatic, lipophilic and all previous

intermolecular interactions) in the AuPosSOM 2.1[37,38] was selected by area under the curve (AUC) of an ROC curve (data not shown), using the same number of false positives and true inhibitors employed during the evaluation stages of the pharmacophore models. The type of intermolecular interactions with higher AUC value were selected.

After the generation of self-organizing maps, a representative of the group from each cluster was assessed individually in relation to intermolecular interactions. The choice of these representatives was based on equation 1 as follows:

$$BR = QFIT - GS \tag{1}$$

where BR is the best representative; QFIT is the setting of the value of the ligand atoms in relation to the coordinates of the pharmacophore point; and GS is the Grid Score.

Evaluation of intermolecular interaction

The analysis of the intermolecular interactions between the molecules that were selected through the self-organizing maps and the orthosteric site residues of TcFPPS were carried out by PoseView web 1.97.0.[39] and the 3D maps constructed with PyMOL 1.7.4.[40]

Results and Discussion

Generation of pharmacophore models from TcFPPS inhibitors

The set of molecules used (IC$_{50}$ < 1 µM) for the generation of pharmacophore models was selected because these structures possess physicochemical and stereoelectronic features that modulate biological activity towards TcFPPS, that allow the identification of new compounds with similar features.

The GALAHAD program allows to identify molecules with partial overlapping in pharmacophore features,[41] and then it is possible to select molecules with partial stereo-electronic requirements that can be optimized, thereby avoiding the exclusion of molecules without the full alignment of the characteristics in the pharmacophore model.[42] For this reason, the program ensures greater flexibility in selecting the potentially active molecules that are present in compound libraries.[43]

Based on the GALAHAD advantages, this program was employed in the generation of pharmacophore models for TcFPPS inhibitors. Because of the stochastic nature of the genetic algorithm, 10 pharmacophore models were generated and evaluated (Table 1).

Table 1. Statistical parameters of pharmacophore models for TcFPPS inhibitors

Model	Energy / (kcal mol^{-1})	STERICS	HBOND	MOL_QRY
1	9.98	96.10	559.50	74.16
2	5.92	90.50	570.10	59.85
3	8.17	93.10	539.50	85.97
4	7.51	98.90	586.40	17.67
5	482,850.20	77.60	592.30	92.96
6	3,678,012.00	118.20	595.70	15.00
7	9,154,801.00	80.90	637.20	52.49
8	2.27	65.60	554.90	34.62
9	6,670,649.00	62.20	623.10	91.49
10	5.57	74.90	531.50	18.67

To select a pharmacophore model that can be used in the virtual screening for prioritizing potential TcFPPS inhibitors, a set of criteria described in the methodology were employed. The use of conformers with high energy values reflects distortions in molecules (torsional angles) for the atoms to approach the center of pharmacophore points.[42,44] With the analysis of the energy values (> 100 kcal mol^{-1}), the pharmacophore models 5, 6, 7 and 9 were excluded.

All the pharmacophore models have a Pareto = 0, indicating that they are statistically equivalent when compared to each other.[45] Thus, to select a pharmacophore model for the virtual screening stage, enrichment metrics were applied to choose the best pharmacophore model.

Evaluation of pharmacophore models

The evaluation steps are crucial to ensure that they represent not only the chemical space of the molecules used in the pharmacophore generation, but they must also be able to identify different chemical classes of molecules with the same pharmacophoric points.[46]

In general, the remaining pharmacophore models showed acceptable GALAHAD statistical parameters, but they are not sufficient to estimate its discriminatory power. Thus, enrichment metrics such as the ROC curve and the area under the ROC curve (AUC) were employed for the evaluation and selection of a pharmacophore model for the virtual screening step.[47] Due to the lack of experimental data for a set of inactive TcFPPS molecules, it was necessary to use the DUD-E server to generate molecules with physicochemical characteristics (charge, molecular weight, number of rotatable bonds, number of Hbond donors/acceptors and log P) similar to positive inhibitors, but dissimilar based on some 2D fingerprint.[24]

To evaluate the pharmacophore models that were previously selected, a set of 20 positive inhibitors and

1,937 false positives (some true inhibitors showed more than one state of protonation) was aligned with the models, and the set value was used to ROC/AUC, EF1% and BEDROC analysis (Figure 2 and Table 2).
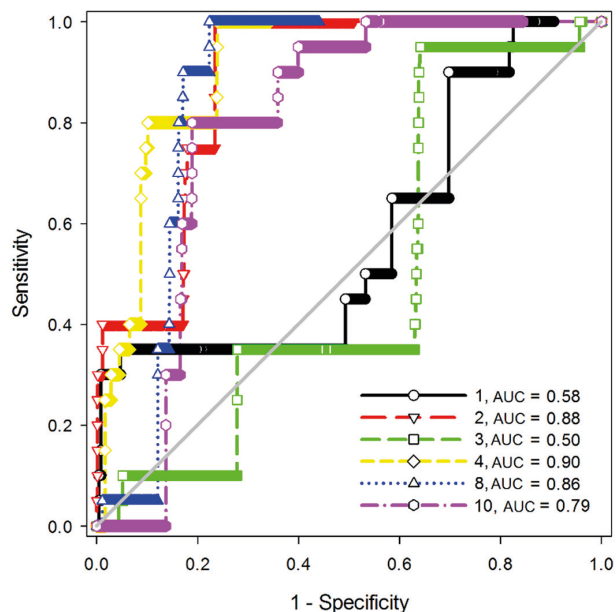

Figure 2. ROC curves of pharmacophore models for TcFPPS inhibitors.

Table 2. Evaluation of the pharmacophore models for TcFPPS inhibitors

Model	AUC	$EF_{1\%}$	BEDROC
1	0.58	19.55	0.30
2	0.88	29.33	0.41
3	0.50	0.00	0.05
4	0.90	0.00	0.33
8	0.86	4.88	0.10
10	0.79	0.00	0.04

AUC: area under the curve; EF: enrichment factor; BEDROC: Boltzmann-enhanced discrimination of receiver operating characteristic.

The evaluation of AUC shows that for an ideal method AUC should be equal to 1.0, whereas methods with AUC ≤ 0.5 would be associated as poor.[48,49] Although the AUC is a metric that allows the overall evaluation of the classification of the data, it is not a metric sensitive to the early recognition. BEDROC is a metric that focuses on giving greater value to the true positives in the first positions of the ordered list and does not present the same limitations regarding the number of molecules and the ratio between active and inactive, characteristics of a random organization method, such as the ROC/AUC analysis.[48] The enrichment factor provides useful information on the early recognition of active molecules. For this reason, the AUC, $EF_{1\%}$ and BEDROC values were evaluated to aid in

the selection of a pharmacophoric model to be applied in virtual screening (Table 2).

Among pharmacophore models for TcFPPS inhibitors, the pharmacophore model 2 (AUC = 0.88; $EF_{1\%}$ = 29.33 and BEDROC = 0.41) was selected for virtual screening. Although the metrics used guarantee a statistical validation of the best model, they are not able to explain the recognition by the activity structure relation for molecules not used in the pharmacophore model generation. For this reason, an analysis was made through the relation between the potency values (pIC_{50}) and the values obtained by the alignment (QFIT) of inhibitors known in the pharmacophore model 2 (Figure 3).
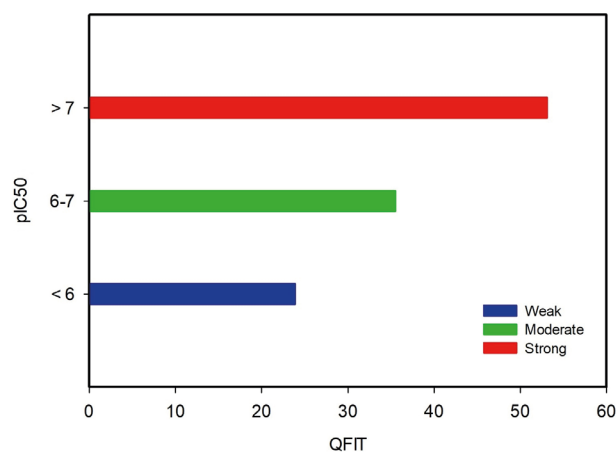
Figure 3. Relationship between biological activity (pIC_{50}) and QFIT for known inhibitors not employed in model training.

Based on the results, only pharmacophore model 2 has a relation between QFIT and pIC_{50} because the potent molecules scored better and the weak inhibitors showed lower QFIT values. This model shows nine pharmacophore features, with six hydrogen bond donor (HBD), one hydrogen bond acceptor (HBA) and two negative centers (NC) (Figure 4). These features are found in bisphosphonates, that have *in vitro* and *in vivo* efficacy against trypanosomes (*T. cruzi* and *T. brucei*), making them promising therapeutic agents against trypanosomal infections.[12,50] Mg^{2+} ion interactions occur with charged phosphonate group (NC), similar to *T. brucei* pharmacophore models previously described on literature.[51] In addition, HBA and HBD can be responsible for the polar interactions to stabilize the complex. Therefore, the pharmacophore model 2 was selected as the most suitable for virtual screening.

For the set used for this analysis, the top ranked molecule (CHEMBL261432, IC_{50} = 0.013 µM) show QFIT = 75.31, and one of the less potent molecules (CHEMBL407221, IC_{50} = 1.399 µM) resulted in QFIT = 10.58.
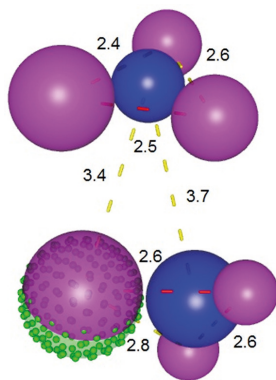
Figure 4. Representation of the best pharmacophore model for TcFPPS. All distances are shown in angstroms. Blue: negative centers; green: hydrogen bond acceptors; pink: hydrogen bond donors. The size of the spheres varies according to the radius of tolerance calculated by GALAHAD.

Pharmacophore-based virtual screening

During the filtering step using pharmacophore model 2, 98,379 structures in the natural products subset from ZINC database were aligned in a flexible manner to assess the degree of accuracy for the atoms with features. After this alignment, approximately 15% of the initial database (n = 15,154) had QFIT values ranging from 0.27 to 88.07. This new set of the molecules was submitted to molecular docking, due to all selected molecules having minimum stereoelectronic requirements for biological activity.

Docking-based virtual screening

The docking based virtual screening have the ability to identify essential features for biological activity; however, the

prediction of the binding modes and affinity of compounds that interact with the binding site are the limitations of the technique. In order to circumvent these limitations, the docking consists of an approach to identify different binding modes by considering the stereoelectronic complementarity between the molecule and target. Thus, the techniques that use the 3D structure of the macromolecular target are complementary strategies to studies that are based on previous knowledge of ligands to allow the assessment of the modes of interaction of potential compounds, that are selected for the pharmacophore model in the active site, respecting the limitations of the molecular volume for docking.[17]

The combination of pharmacophore models with docking has been widely exploited in drug design.[52] Typically, the techniques that have a lower computational demand, as pharmacophore model, are initially applied, and the filtered molecules are consecutively directed to other strategies such as docking.

To evaluate docking search parameters, some metrics evaluations were used, the results are shown in Figure 5. The best crystallographic pose of the ligand generated by DOCK 6.5 was evaluated using the RMSD value, in relation to the crystallographic ligand pose obtained in PDB. This result shows that the RMSDs found was 1.43 Å, and the RMSDh and RMSDm were 1.12 and 0.37 Å, respectively. The literature[36] indicates that the results in which RMSD ≤ 2 Å are considered acceptable. Based on this data, the use of this docking search parameters can result in reliable poses.

The use of GS scoring function is based on molecular mechanics force fields (van der Waals and Coulomb interaction energy).[36] For the evaluation of this function, ROC curve was built with a set of false positives and
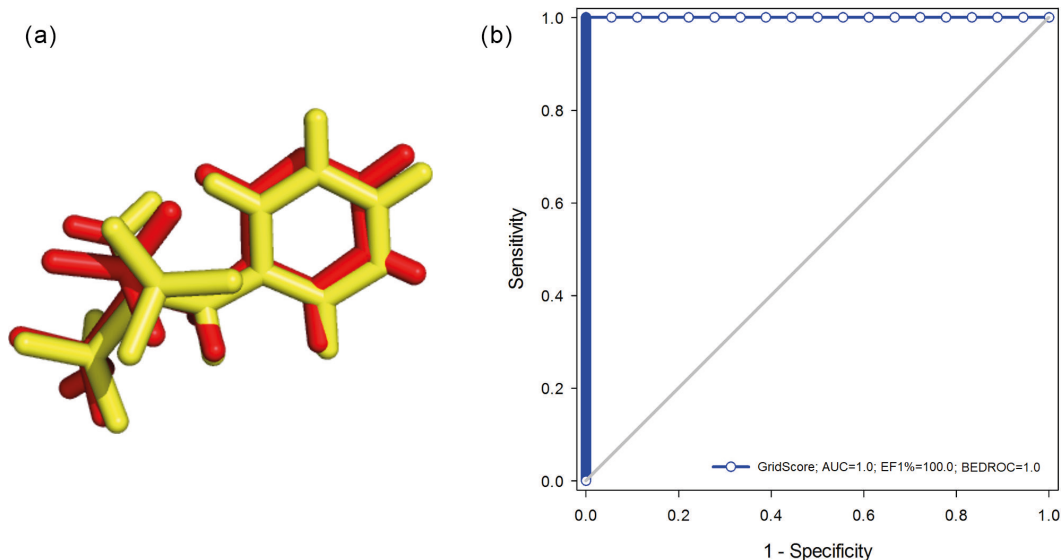

Figure 5. Docking evaluation: (a) re-docking analysis (red: crystallographic pose risedronate; yellow: DOCK 6.5 best pose generated) and (b) ROC curve for evaluation of Grid Score.

positive inhibitors to evaluate the pharmacophore models. In the evaluation, the values of AUC, $EF_{1\%}$ and BEDROC indicate that the score function was able to recover all true positive inhibitor and not decoys,[25] thus the specificity and sensitivity of this function that should be set is 100%. The function was shown to be predictive for selecting natural products filtered by the pharmacophore model. After the docking, a set of 11,762 molecules was selected with GS values ranging from –232.74 to –0.96.

Prioritization molecules through pharmacophore and energy criteria

The hierarchical virtual screening allowed to reduce and to select the universe of molecules contained in the natural products database.[17] A prioritization of the compounds was performed according to ranking generated in both pharmacophore and docking strategies. The top 500 ranked structures in the pharmacophore models screening were compared with the top 500 ranked structures in the molecular docking. The top 30 ranked structures in both strategies were subjected to analyses of intermolecular interactions.

The top 500 findings sorted by QFIT structures and GS were used to identify the best molecules classified for both methods (Figure 6).
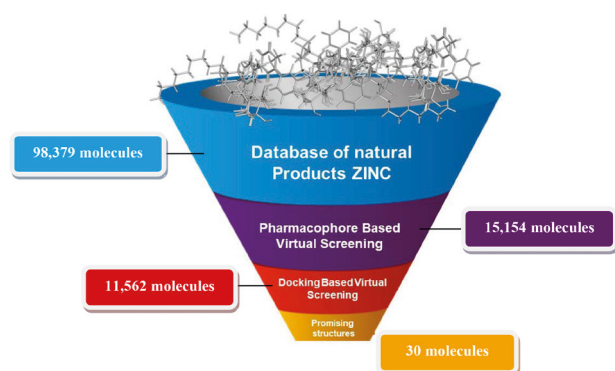
Figure 6. Hierarchical virtual screening scheme (pharmacophore and molecular docking).

It was possible to select 30 structures with better QFIT and GS values. The top 30 QFIT values ranged from 85.18 to 54.53, while GS values ranged from –176.92 and –110.16 kcal mol^{-1}. Thus, this strategy resulted in a set of molecules that was best scored by the pharmacophore model and docking simultaneously.

Virtual screening by self-organizing maps

The results obtained from the hierarchical virtual screening were ranked by affinity energy (GS), and the top 30 ranked structures were used for the clustering analysis by grouping compounds with similar contacts with the receptor.[38] The categorization of molecules through intermolecular interactions generated by molecular docking poses allows to prioritize the compounds in a group on the basis of the number/pattern of contacts.

The evaluation results on the technical efficiency from the ROC curves for each of the four types of interactions evaluated by AuPosSOM showed that of the four types of selection contacts, hydrogen bonds (AUC = 0.93) and electrostatic interactions (AUC = 0.88) represent the best types of interactions for identifying active compounds, and the other contacts proved to be somewhat predictive, with AUC < 0.7.

The best AUC values in the clustering for the hydrogen and electrostatic contacts are due to the characteristics of the amino acids present in the orthosteric site (Asp98, Asp102, Arg107, Lys207, Thr208, Tyr211, Gln247, Lys264, Asp250, Asp251, Asp254, Arg51 and Lys362). Those interactions are also responsible for stabilizing the substrate/inhibitor complex.[50] Thus, lipophilic interactions (AUC < 0.5) are insufficient for discriminating among molecules that are active against TcFPPS, affecting negatively the analysis of all interactions (AUC = 0.65).

The results of clustering analysis using hydrogen bond interactions (best AUC value) for the top 30 ranked molecules are shown in Figure 7.

In the clustering analysis, it is possible to observe the grouping of molecules in four clusters, each cluster containing molecules that show similar footprint contacts. One representative molecule in each cluster, with the best QFIT and GS scores (BR), was selected for individual analysis of intermolecular interactions (Table 3 and Figure 8).

Analysis of intermolecular interactions

The analysis of the binding mode is used to identify and optimize the interactions between ligands and their macromolecular targets, and it can be useful for the design of more potent and selective drugs.[53] For this purpose, the interactions of the four best molecules selected after the construction of self-organizing maps in the TcFPPS active site were analyzed (Figure 9).

In the TcFPPS orthosteric site residues, there is a predominance of those with side chains containing carboxylic group (Asp92, Asp102, Asp170, Asp250, Asp360) and amino group (Arg51 and Lys362),[50] which confirms the possible formation of hydrogen bonds or electrostatic interaction as suggested by the AuPosSOM 2.1.
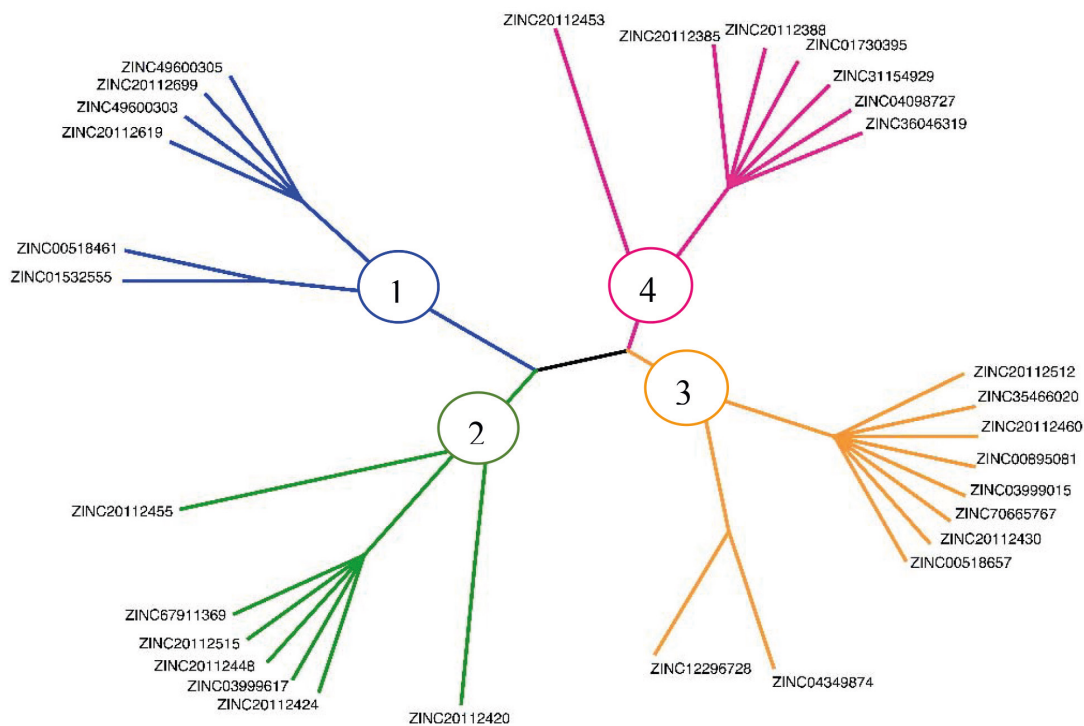
Figure 7. Clustering tree (Newick representation) with the distribution of the top 30 ranked molecules, each cluster (1-4) signifies a different ligand footprint on the protein/ligand complex.

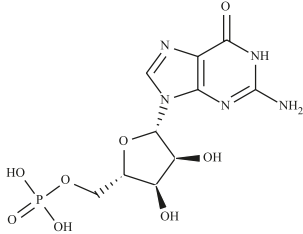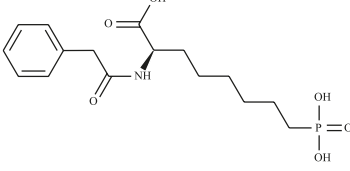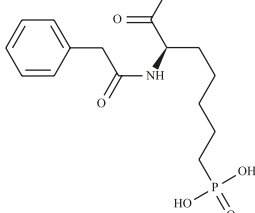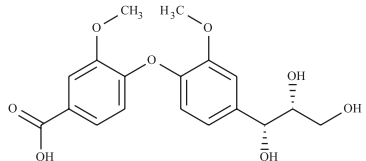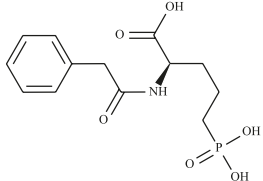**Table 3.** QFIT and GS values of the molecules grouped by AuPosSOM 2.1

Cluster	Molecule	Structure	QFIT	GS / (kcal mol⁻¹)	BR
	ZINC01532555		62.67	−176.92	239.59
	ZINC20112619		84.21	−151.37	235.58
1	ZINC20112699		60.05	−145.40	205.45
	ZINC49600303		61.12	−120.73	181.85

Table 3. QFIT and GS values of the molecules grouped by AuPosSOM 2.1 (cont.)
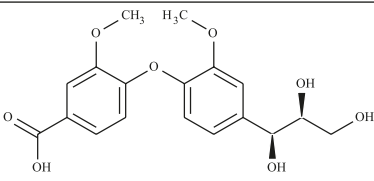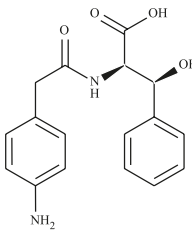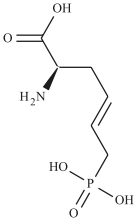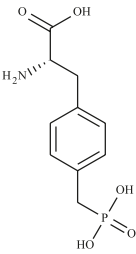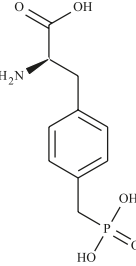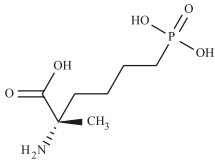
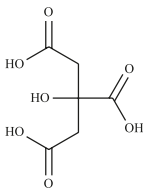| Cluster	Molecule	Structure	QFIT	GS / (kcal mol⁻¹)	BR
1	ZINC49600305		56.06	−120.00	176.06
	ZINC00518461		54.71	−116.44	171.15
2	ZINC20112455		85.18	−126.34	211.52
	ZINC20112448		60.28	−145.85	206.13
	ZINC20112420		79.49	−121.48	200.97
	ZINC20112424		79.49	−115.02	194.51
	ZINC20112515		57.88	−128.71	186.59

Table 3. QFIT and GS values of the molecules grouped by AuPosSOM 2.1 (cont.)
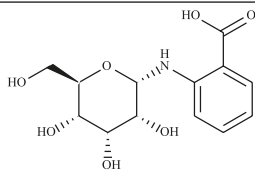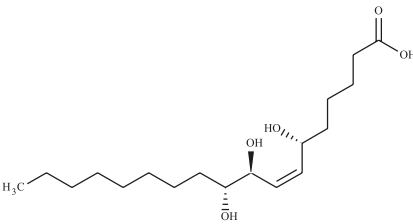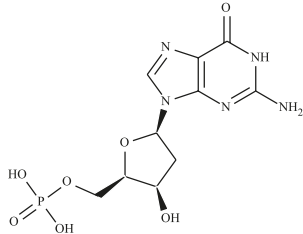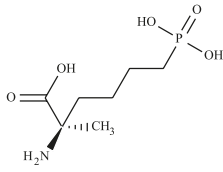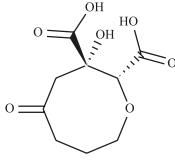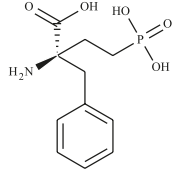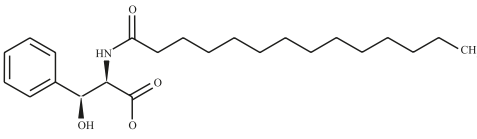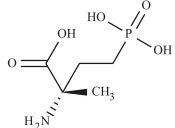
Cluster	Molecule	Structure	QFIT	GS / (kcal mol^{-1})	BR
	ZINC03999617		61.65	−115.01	176.66
2	ZINC67911369		57.80	−115.09	172.89
	ZINC12296728		62.67	−172.72	235.39
	ZINC00895081		57.14	−175.35	232.49
	ZINC20112512		77.32	−135.32	212.64
3	ZINC04349874		63.87	−132.47	196.34
	ZINC20112460		61.23	−132.94	194.17
	ZINC70665767		55.11	−135.29	190.40
	ZINC20112430		54.90	−130.86	185.76

Table 3. QFIT and GS values of the molecules grouped by AuPosSOM 2.1 (cont.)
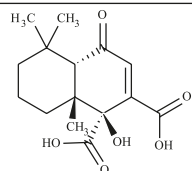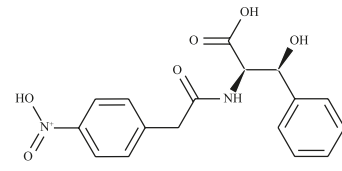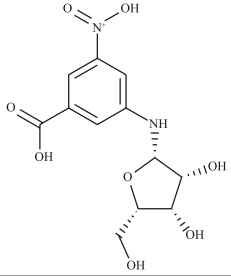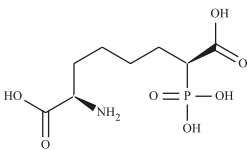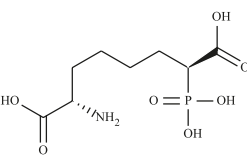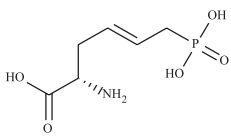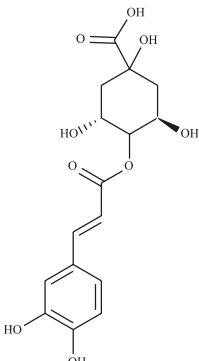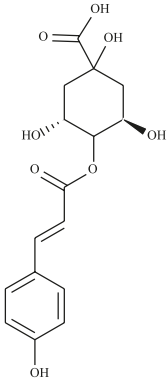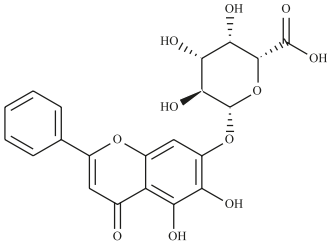
Cluster	Molecule	Structure	QFIT	GS / (kcal mol^{-1})	BR
	ZINC35466020		61.25	−121.18	182.43
3	ZINC00518657		54.67	−117.59	172.26
	ZINC03999015		55.20	−115.90	171.10
	ZINC01730395		62.67	−172.91	235.58
	ZINC20112388		66.45	−161.24	227.69
4	ZINC20112385		66.45	−160.07	226.52
	ZINC20112453		85.18	−126.26	211.44
	ZINC31154929		57.86	−120.11	177.97

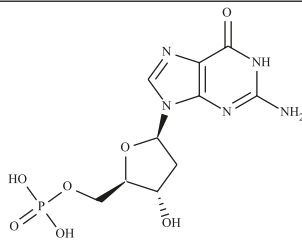**Table 3.** QFIT and GS values of the molecules grouped by AuPosSOM 2.1 (cont.)

Cluster	Molecule	Structure	QFIT	GS / (kcal mol⁻¹)	BR
4	ZINC04098727		57.86	−116.84	174.70
	ZINC36046319		54.53	−110.16	164.69

GS: Grid Score; BR: best representative.
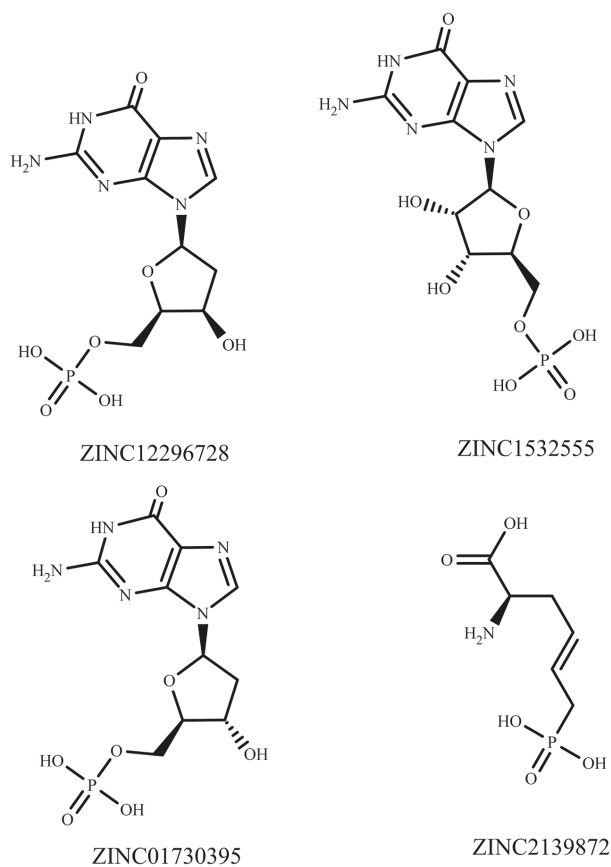
Figure 8. Two-dimensional (2D) chemical structures of top ranked compounds in each cluster.

The ZINC12296728 phosphate group interacts with three Mg^{2+} ions (Figure 9a) that are present at the binding site, which interacts with Asp250, Asp98 and Asp102. In addition, cation-π interactions can be observed between the aromatic ring of the ligand and the amino group of the Lys264 and Lys362 sidechain.

The ZINC1532555 phosphate group (Figure 9b) forms interactions with the three Mg^{2+} ions of TcFPPS. In the dihydropurin ring of ligand forms a hydrogen bond with Lys362, other interaction of the same type was observed between the hydroxyl at tetrahydrofuran ring of the ligand with hydroxyl sidechain of Tyr211. The other hydroxyl tetrahydrofuran ring is coordinated by a hydrogen bond donor interaction with Tyr211. The dihydropurin ring of the ligand forms π-cation interaction with Lys254 protonated amino group.

The complex ZINC01730395-TcFPPS (Figure 9c), in turn, presents the same type of interaction pattern between its phosphate group and Mg^{2+} ions. The apolar side chain of Leu95 interacts with the ligand by hydrophobic contact. The hydrogen bond acceptor occurs with Lys264 and Lys362 residues. Furthermore, dihydropurin ring at ligand form π-cation interaction with Lys362 residue.

For ZINC20112455, interactions are observed between the phosphates and Mg^{2+} ions. Hydrogen bond acceptors are also shown between ligand carbonyl groups with residues Ser104 and Arg107 (Figure 9d).

(a) ZINC12296728

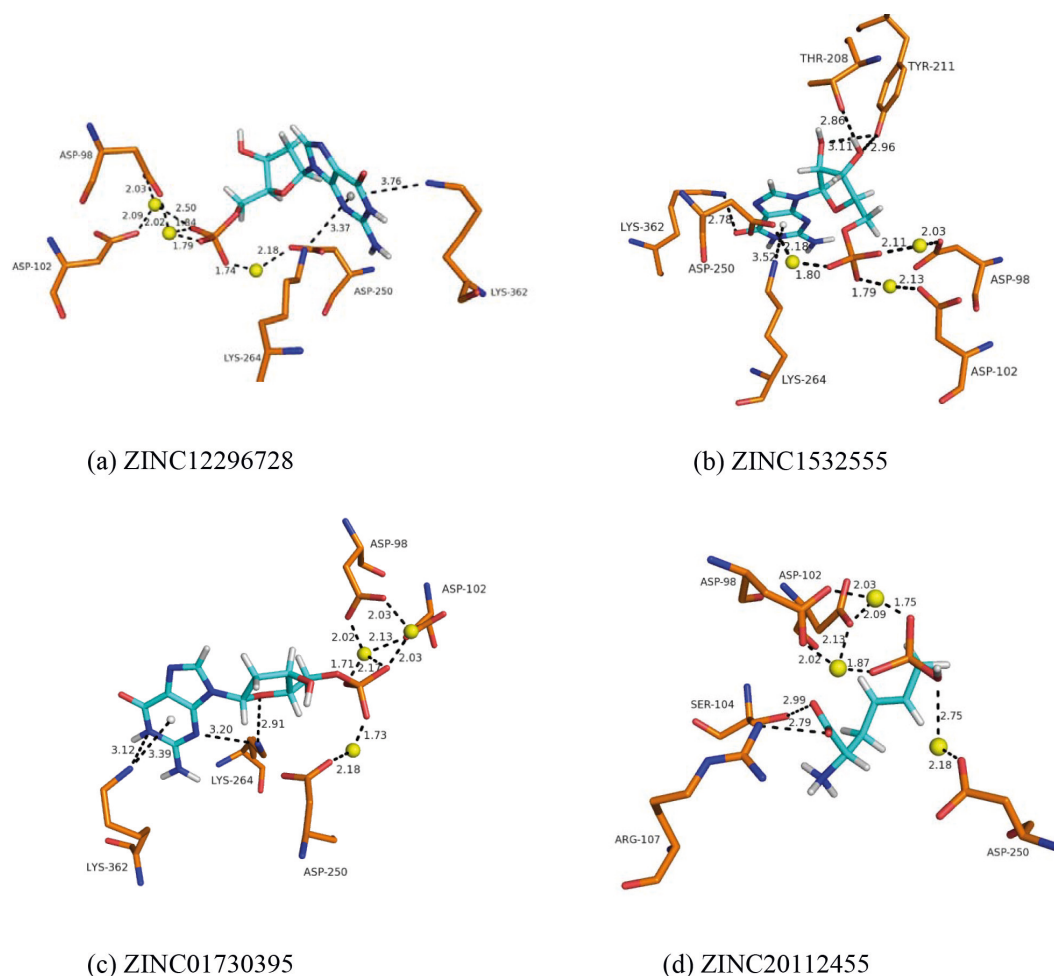(b) ZINC1532555

(c) ZINC01730395

(d) ZINC20112455

Figure 9. Analysis of intermolecular interactions between TcFPPS and selected molecules through self-organizing maps. All distances are shown in angstroms.

Based on the observations on the intermolecular interactions present in both compounds, it can be concluded that the main types observed are the electrostatic interactions and hydrogen bonds. Similar interactions were shown for the crystallographic inhibitor (risedronate), in which the mean distance between the hydrogen bonding pairs and the inhibitor was 3.17 Å and for Mg^{2+} bonds occurred at a distance of 1.95 to 2.24 Å.[50] In other complexes between bisphosphonates and TcFPPS,[53] three Mg^{2+} ions bridge the phosphates of the inhibitor bisphosphonate. The conserved aspartate residues of the two-aspartate rich motifs DDXXD (Asp98-102 and Asp250-254) bind three divalent cations (Mg^{2+}) that are in turn coordinated by the phosphate backbone of the bisphosphonates and are crucial for the activity. Thus, the interactions observed for both compounds mimic those found for risedronate and bisphosphonates and may be recognized in the orthosteric site of TcFPPS. Interactions with Mg^{2+} ions are critical in catalysis and were observed in four complexes.[54]

Conclusions

The use of the GALAHAD program allowed the generation and evaluation of pharmacophore models for TcFPPS inhibitors, resulting in the selection of a pharmacophore model with good discriminatory power that was able to help explain the relations between the setting value of fitting inhibitors and biological activity.

Pharmacophore model 2 has six hydrogen bonding donor, one hydrogen bonding acceptor and two negative centers that are essential for inhibitory activity against the TcFPPS and important to modulate potency of this class.

The selection of a pharmacophore model that is acceptable for prioritizing potential TcFPPS inhibitors and their subsequent application in virtual screening resulted in the identification of 15,154 molecules with minimum requirements to recognize onto binding site.

For the docking, GS score function was used to select molecules that were filtered by the pharmacophore model, which was shown to be predictive for the set of true and

false-positive inhibitors. Furthermore, the DOCK 6.5 search parameters showed good performance in positioning the crystallographic ligand and reproducing the spatial conformation of the crystallographic binder because the RMSD values were less than 2 Å.

Self-organizing maps prioritized four compounds, which showed binding pattern between oxygen atoms, magnesium ions and aspartic acid residues from conserved regions of the catalytic site, which is crucial for their activity.

This study identified four promising compounds from natural products database (ZINC1532555, ZINC12296728, ZINC20112455 and ZINC01730395) that mimic the major interactions present in the substrate/inhibitor (Asp98, Asp102, Asp250 and Mg^{2+} ions), which indicates that these molecules are candidates for *in vitro* assays.

Supplementary Information

Supplementary data (dataset 2-20 true inhibitors) are available free of charge at http://jbcs.sbq.org.br as PDF file.

Acknowledgments

References

1. Nunes, M. C.; Dones, W.; Morillo, C. A.; Encina, J. J.; Ribeiro, A. L.; *J. Am. Coll. Cardiol.* **2013**, *62*, 767.

2. Coura, J. R.; *Mem. Inst. Oswaldo Cruz* **2007**, *102*, 113.

3. Pinto, A. Y. N.; Valente, A. S.; Valente, V. C.; Ferreira Junior, A. G.; Coura, J. R.; *Rev. Soc. Bras. Med. Trop.* **2008**, *41*, 602.

4. Kirchhoff, L. V.; *Adv. Parasitol.* **2011**, *75*, 1.

5. World Health Organization (WHO); *Working to Overcome the Global Impact of Neglected Tropical Diseases*; WHO Press: Geneva, 2010. Available at apps.who.int/iris/bitstream/handle/10665/44440/9789241564090_eng.pdf?sequence=1, accessed in March 2018.

6. World Health Organization (WHO); *Investing to Overcome the Global Impact of Neglected Tropical Diseases*; WHO Press: Geneva, 2015. Available at apps.who.int/iris/bitstream/handle/10665/152781/9789241564861_eng.pdf?sequence=1, accessed in March 2018.

7. Costa, S. A.; Rassi, S.; Freitas, E. M. M.; Gutierrez, N. S.; Boaventura, F. M.; Sampaio, L. P. C.; Silva, J. B. M.; *Arq. Bras. Cardiol.* **2017**, *108*, 246.

8. Mangini, S.; Alves, B. R.; Silvestre, O. M.; Pires, P. V.; Pires, L. J. T.; Curiati, M. N. C.; Bacal, F.; *Einstein* **2015**, *13*, 310.

9. Bocchi, E. A.; Marcondes-Braga, F. G.; Bacal, F.; Ferraz, A. S.; Albuquerque, D.; Rodrigues, D. A.; Mesquita, E. T.; Vilas-Boas, F.; Cruz, F.; Ramires, F.; Villacorta Junior, H.; Souza Neto, J. D.; Rossi Neto, J. M.; Moura, L. Z.; Beck-da-Silva, L.; Moreira, L. F.; Rohde, L. E. P.; Montera, M. W.; Simões, M. V.; Moreira, M. C.; Clausell, N.; Bestetti, R.; Mourilhe-Rocha, R.; Mangini, S.; Rassi, S.; Ayub-Ferreira, S. M.; Martins, S. M.; Bordignon, S.; Issa, V. S.; *Arq. Bras. Cardiol.* **2012**, *98*, 1.

10. Pontes, V. M. O.; Souza Júnior, A. S.; Cruz, F. M. T.; Coelho, H. L. L.; Dias, A. T. N.; Coêlho, I. C. B.; Oliveira, M. F.; *Rev. Soc. Bras. Med. Trop.* **2010**, *43*, 182.

11. Dias, L. C.; Dessoy, M. A.; Silva, J. J. N.; Thiemann, O. H.; Oliva, G.; Andricopulo, A. D.; *Quim. Nova* **2009**, *32*, 2444.

12. Montalvetti, A.; Bailei, B. N.; Martin, M. B.; Severin, G. W.; Oldfield, E.; Docampo, R.; *J. Biol. Chem.* **2001**, *276*, 33930.

13. Garzoni, L. R.; Waghabi, M. C.; Baptista, M. M.; Castro, S. L.; Meirelles, M. N.; Britto, C. C.; Docampo, R.; Oldfield, E.; Urbina, J. A.; *Int. J. Antimicrob. Agents* **2004**, *23*, 286.

14. Recher, M.; Barboza, A. P.; Li, Z. H.; Galizzi, M.; Ferrer-Casal, M.; Szajnman, S. H.; Docampo, R.; Moreno, S. N.; Rodriguez, J. B.; *Eur. J. Med. Chem.* **2013**, *60*, 431.

15. Dhar, M. K.; Koul, A.; Kaul, S.; *New Biotechnol.* **2013**, *30*, 114.

16. Souza, W.; Rodrigues, J. C. F.; *Interdiscip. Perspect. Infect. Dis.* **2009**, *2009*, article ID 642502.

17. Rodrigues, R. P.; Mantoani, S. P.; Almeida, J. R.; Pinsetta, F. R.; Semighini, E. P.; Silva, V. B.; Silva, C. H. P.; *Rev. Virtual Quim.* **2012**, *4*, 739.

18. Almeida, L. C.; Santos, E.; Sampaio, C.; Taranto, A. G.; Leite, F. H. A.; *Textura* **2012**, *5*, 43.

19. Rosso, V. S.; Szajnman, S. H.; Malayil, L.; Galizzi, M.; Moreno, S. N.; Docampo, R.; Rodriguez, J. B.; *Bioorg. Med. Chem.* **2011**, *19*, 2211.

20. Szajnman, S. H.; Liñares, G. E. G.; Li, Z. H.; Jiang, C.; Galizzi, M.; Bontempi, E. J.; Ferella, M.; Moreno, S. N.; Docampo, R.; Rodriguez, J. B.; *Bioorg. Med. Chem.* **2008**, *16*, 3283.

21. ChemAxon; *MarvinSketch*, version 5.12.1; ChemAxon, Cambridge, 2013.

22. *SYBYL-X 2.0*, version 2.0; Tripos International, St. Louis, 2012.

23. *GALAHAD Manual*, *SYBYL 2.0*; Tripos, St. Louis, MO, USA, 2011, p. 1-90.

24. Mysinger, M. M.; Carchia, M.; Irwin, J. J.; Shoichet, B. K.; *J. Med. Chem.* **2012**, *55*, 6582.

25. Braga, R. C.; Andrade, C. H.; *Curr. Top. Med. Chem.* **2013**, *13*, 1127.

26. *SigmaPlot 12.0*; Systat Software: San Jose, CA, 2012.

27. Irwin, J. J.; Shoichet, B.; *J. Chem. Inf. Model.* **2005**, *45*, 177; http://zinc.docking.org, accessed on March 3, 2015.

28. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E.; *Nucleic Acids Res.* **2000**, *28*, 235.

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

30. Anandakrishnan, R.; Aguilar, B.; Onufriev, A. V.; *Nucleic Acids Res*. **2012**, *40*, W537.

31. Allen, W. J.; Balius, T. E.; Mukherjee, S.; Brozell, S. R.; Moustakas, D. T.; Lang, P. T.; Case, D. A.; Kuntz, I. D.; Rizzo, R. C.; *J. Comput. Chem.* **2015**, *36*, 1132.

32. Ferrin, T. E.; Huang, C. C.; Jarvis, L. E.; Langridge, R.; *J. Mol. Graphics* **1988**, *6*, 13.

33. Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E.; *J. Mol. Biol.* **1982**, *161*, 269.

34. Shoichet, B. K.; Kuntz, I. D.; Bodian, D. L.; *J. Comput. Chem.* **1992**, *13*, 380.

35. Meng, E. C.; Shoichet, B. K.; Kuntz, I. D.; *J. Comput. Chem.* **1992**, *13*, 505.

36. Brozell, S. R.; Mukherjee, S.; Balius, T. E.; Roe, D. R.; Case, D. A.; Rizzo, R. C.; *J. Comput.-Aided Mol. Des.* **2012**, *26*, 749.

37. Bouvier, G.; Evrard-Todeschi, N.; Girault, J. P.; Bertho, G.; *Bioinformatics* **2010**, *26*, 53.

38. Mantsyzov, A. B.; Bouvier, G.; Evrard-Todeschi, N.; Bertho, G.; *Adv. Appl. Bioinf. Chem.* **2012**, *5*, 61.

39. Stierand, K.; Maas, P. C.; Rarey, M.; *Bioinformatics* **2006**, *22*, 1710.

40. DeLano Scientific LLC; *The PyMOL Molecular Graphics System*, version 1.3; DeLano Scientific LLC: New York, 2010.

41. Clark, R. D.; Abrahamian, A. E. E.; *J. Comput.-Aided Mol. Des.* **2009**, *23*, 765.

42. Liu, M.; Sun, Z.; Hu, W.; *Neural Regener. Res.* **2012**, *7*, 1398.

43. Richmond, N. J.; Abrams, C. A.; Wolohan, P. R.; Abrahamian, E.; Willett, P.; Clark, R. D.; *J. Comput.-Aided Mol. Des.* **2006**, *20*, 567.

44. Caballero, J.; *J. Mol. Graph. Model.* **2010**, *29*, 363.

45. Athri, P.; Wenzler, T.; Tidwell, R.; Bakunova, S. M.; Wilson, W. D.; *Eur. J. Med. Chem.* **2010**, *45*, 6147.

46. Tropsha, A.; *Mol. Inf.* **2010**, *29*, 476.

47. Shepphird, J. K.; Clark, A. E. R. D.; *J. Comput.-Aided Mol. Des.* **2006**, *20*, 763.

48. Langer, T.; Hoffman, R. D.; *Pharmacophores and Pharmacophore Searches*, 1st ed.; Wiley-VCH: Weinheim, Germany, 2006.

49. Vanagas, G.; *Interact. Cardiovasc. Thorac. Surg.* **2004**, *3*, 319.

50. Gabelli, S. B.; McLellan, J. S.; Montalvetti, A.; Oldfield, E.; Docampo, R.; Amzel, L. M.; *Proteins: Struct., Funct., Bioinf.* **2006**, *62*, 80.

51. Montalvetti, A.; Fernandez, A.; Sanders, J. M.; Ghosh, S.; Brussel, E. V.; Oldfield, E.; Docampo, R.; *J. Biol. Chem.* **2003**, *278*, 17075.

52. Kumar, A.; Zhang, K. Y. J.; *Methods* **2015**, *71*, 26.

53. Bissantz, C.; Kuhn, B.; Stahl, M.; *J. Med. Chem.* **2010**, *53*, 5061.

54. Aripirala, S.; Szajnman, S. H.; Jakoncic, J.; Rodriguez, J. B.; Docampo, R.; Gabelli, S. B.; Amzel, L. M.; *J. Med. Chem.* **2012**, *55*, 6445.