

The receptor-dependent LQTA-QSAR: application to a set of trypanothione reductase inhibitors

Euzébio G. Barbosa · Kerly Fernanda M. Pasqualoto ·
Márcia M. C. Ferreira

Received: 29 November 2011 / Accepted: 29 August 2012
© Springer Science+Business Media B.V. 2012

Abstract A new Receptor-Dependent LQTA-QSAR approach, RD-LQTA-QSAR, is proposed as a new 4D-QSAR method. It is an evolution of receptor independent LQTA-QSAR. This approach uses the free GROMACS package to carry out molecular dynamics simulations and generates a conformational ensemble profile for each compound. Such an ensemble is used to build molecular interaction field-based QSAR models, as in CoMFA. To show the potential of this methodology, a set of 38 phenothiazine derivatives that are specific competitive *T. cruzi* trypanothione reductase inhibitors, was chosen. Using a combination of molecular docking and molecular dynamics simulations, the binding mode of the phenothiazine derivatives was evaluated in a simulated induced fit approach. The ligands alignments were performed using both ligand and binding site atoms, enabling unbiased alignment. The models obtained were extensively validated by leave-*N*-out cross-validation and *y*-randomization techniques to test for their robustness and absence of chance correlation. The final model presented Q^2 LOO of 0.87 and R^2 of 0.92 and a suitable external prediction of $Q_{ext}^2 = 0.78$. The adapted binding site obtained is useful to perform virtual screening and ligand structure-based design and the descriptors in the final model can aid in the design new inhibitors.

Keywords LQTA-QSAR · *T. cruzi* · Docking · Molecular dynamics simulations

E. G. Barbosa · M. M. C. Ferreira (✉)
Chemistry Institute, University of Campinas (UNICAMP), POB
6154, Campinas, SP 13083-970, Brazil
e-mail: marcia@iqm.unicamp.br

K. F. M. Pasqualoto
Butantan Institute, São Paulo, SP, Brazil

Introduction

American trypanosomiasis or Chagas' disease is one of the most serious protozoan diseases. It occurs throughout Latin America, particularly in South America. It is mainly transmitted to man by the infected feces of a blood-sucking triatomine bug through the insect's sting [1]. Coura and Viñas [2] have recently reported that the *T. cruzi* parasite can travel with population movements from endemic to non-endemic countries such as North America, the western Pacific region, and also Europe. There are no prophylactic drugs to prevent infections with *T. cruzi*. The current chemotherapy of Chagas' disease is based on the nitroaromatic compounds benznidazole and nifurtimox. Compounds with low toxicity and increased efficacy during the indeterminate and chronic phases, are still needed [3].

The parasites belonging to the Trypanosomatidae family have a unique thiol-dependent redox metabolism based upon the substrate trypanothione [T(SH)₂], which is reduced to trypanothione disulfide [T(S)₂], and specific enzymes including a trypanothione reductase (TR). This system is absent in the mammalian host, being replaced by the omnipresent glutathione reductase (GR), which has glutathione (GSH) as substrate [3]. Host GR is not able to reduce TR substrate, trypanothione disulfide (T[S]₂) and this mutually exclusive recognition of substrates between host and parasite enzymes suggests that a selective inhibitor design approach could be achieved, since the reduction of (T[S]₂) is crucial for maintaining the reducing intracellular environment of these parasites [4].

X-Ray crystallographic data for TR from *T. cruzi* is available in its free form and complexed with its natural substrates [5–7]. Rational drug design of potent inhibitors requires accurate structures of enzyme-inhibitor complexes, and a *T. cruzi* TR complexed with a competitive

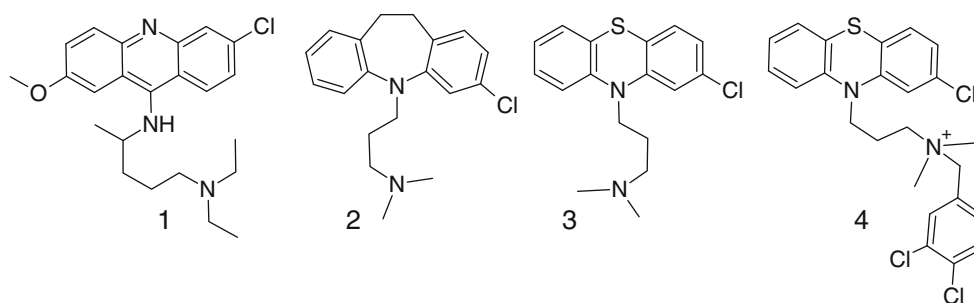


Fig. 1 Chemical structures of mepacrine (1), chlomipramine (2), chlorpromazine (3), and the inhibitor A6 (4)

inhibitor (Fig. 1, Compound 1) is available [4]. When *T. cruzi* TR is compared to human GR, four out of five residues inside the active site, where mepacrine is bound, are not conserved. That provides selective inhibition of the parasite enzyme, which can be explored for drugs against *T. cruzi* [8]. Thus, different classes of tricyclic compounds have been shown to be competitive TR inhibitors [9].

Molecular modeling led to the early discovery of imipramine, phenothiazine, and other tricyclic antidepressants as specific competitive inhibitors for *T. cruzi* TR over GR [10]. The dibenzazepine drug chlomipramine (Fig. 1, compound 2) inhibits *T. cruzi* TR with an inhibition constant (K_i) value of 6.5 μM . Studies on phenothiazines, containing (dimethylamino) propyl and (dimethylamino) ethyl substituents, underlined the role of a hydrophobic core and positively charged side chain for ligand binding at the active site of *T. cruzi* TR [11].

Chlorpromazine (Fig. 1, compound 3) was found to be a much more efficient inhibitor of *T. cruzi* TR than human GR (K_i values of 10 μM and 762 μM , respectively). In contrast, phenothiazines with quaternary positive nitrogen replaced by a negative carboxyethyl side chain showed a nine-fold selectivity for GR over *T. cruzi* TR inhibition, demonstrating the importance of a positive moiety in the inhibitors' structures [11]. The most potent inhibitor (A6) (Fig. 1, compound 4) presented a K_i value of 120 nM, which is about two orders of magnitude lower than the parent chlorpromazine [12].

Building a 3D-QSAR (quantitative structure–activity relationship) model would not be an easy task due to the lack of crystallographic data regarding the phenothiazine inhibitors complexed to *T. cruzi* TR. In order to estimate correctly the binding modes, one can make use of molecular docking. However, the enzyme TR from *T. cruzi*, which has many degrees of freedom, would certainly adopt an appropriate conformational arrangement to accommodate the phenothiazine inhibitors and, unfortunately, the motions involved in this enzyme-inhibitor fitting process are not yet properly considered in the available molecular docking approaches. Even small motions, such as the local rearrangement of the amino acid side chains and the loops,

have an undesirable effect on the docking results. One strategy to overcome these docking issues, for instance, is taking into account protein flexibility employing molecular dynamics (MD) simulations with explicit solvent [13].

MD is a powerful tool, now routinely applied to simulate complex dynamic processes that occur in biological systems, such as those involved in the molecular recognition of a certain drug by its target (receptor, enzyme, DNA, etc.) [14]. Despite all MD capabilities, there are a very few studies reporting the combination of QSAR and MD simulations. Thus, the building of receptor and time-dependent QSAR models [*receptor-dependent (RD) 4D-QSAR*] has some advantages when compared to a mono configurational 3D-QSAR analysis [15]. In a *RD 4D-QSAR*, protein flexibility and induced fit can be explored and the ligand alignment bias can be minimized by performing MD simulations of the complex ligand-receptor.

A new 4D-QSAR approach, named LQTA-QSAR (LQTA, *Laboratório de Quimiometria Teórica e Aplicada*), was recently reported [16]. This methodology simultaneously explores the main features of CoMFA and 4D-QSAR [17] paradigms. LQTA-QSAR makes use of the free GROMACS package [18] to carry out the MD simulations and to generate the conformational ensemble profile (CEP) for each compound or ligand of the investigated set. MD simulations can be carried out in a system bearing any desirable degree of complexity, as in this case, where the enzyme *T. cruzi* TR and its inhibitors are surrounded by bulky aqueous media and counterions. The ligands' MD frames from this temporal representation of the binding phenomenon, when aligned to the remaining ligands, can be used to calculate 3D molecular properties and, subsequently, build *RD 4D-QSAR* models.

In the present study, molecular docking and MD simulations were combined to simulate the induced fit of 38 phenothiazine derivatives as *T. cruzi* TR inhibitors [12, 19, 20]. Additionally, a *RD 4D-QSAR* approach was performed to verify that the binding mode suggested for the phenothiazine derivatives by the combination of molecular docking and MD simulations generated reliable models with chemical and biological meaning and also having a

good external predictability (test set). The ligands' alignments were explored based upon both ligand and binding site atoms, which is capable of providing unbiased CEP alignment.

Methodology

A three-dimensional structure of inhibitor (mepacrine) complexed to *T. cruzi* TR was retrieved from PDB [21] under the entry code 1GXF, [22] in which the inhibitor binds to the receptor as a dimer. The binding pocket shape carved by this mepacrine pair is significantly different from any other chlorpromazine derivative binding mode. The later compound contains a bent tricyclic ring and might bind in different ways in the same pocket. Therefore, the binding site must be adjusted in order to properly fit the phenothiazine derivatives (induced fit). These structural changes cannot be achieved by molecular docking itself, but through the combination of molecular docking, to bind the ligands, and MD simulations, to simulate the induced fit. This combined approach can account for a more realistic representation of the complex *T. cruzi* TR-phenothiazine derivatives and, consequently, to more reliable *RD-LQTA-QSAR* models. The most active compound (A6) (Fig. 1, compound 4) was the first used for this purpose and a detailed stepwise procedure is reported in the section named Receptor molecular model.

Ligand molecular models

Ligands' molecular structures were built using Gauss View [23]. The structures were energy-minimized by applying the DFT/M052x [24] level and the cc-pVDZ basis set employing Gaussian'03, [25] and a CHELPG [26] population analysis was carried out. The optimized structures were used to produce the gromos96 ligand's topologies at the PRODRG [27] server. The estimation of the protonation state at pH 7 for all ligands complexed with the *T. cruzi* TR model was performed using the PROPKA [28] web server. Table 1 presents the chemical structures of the 38 phenothiazine derivatives used as ligands in this *RD-LQTA-QSAR* analysis.

The biological activities were expressed as the free binding energies (ΔG , kcal/mol), which were calculated from the inhibitory binding constant (K_i , mM) values using the equation $\Delta G = -RT \ln K_i$, where T is the absolute temperature and R is the gas constant.

Receptor molecular model

The complex structure 1GXF [22] was used to build up the receptor model. Mepacrine covalently bound to *T. cruzi* TR

was eliminated, the aminoacid residues were properly rebuilt, and the final structure was refined using DeepView [29]. The protonation state of the aminoacid residues at pH 7 was estimated employing the PROPKA [28] web server. The final protein model was used to perform the molecular docking of ligand A6. The GROMOS force field was employed to represent the protein structure in the MD simulations performed by GROMACS 4 [18]. After the MD simulations with ligand A6 bound to *T. cruzi* TR, the modified protein structure was used to dock the remaining ligands.

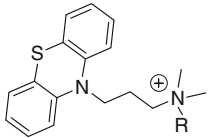
Binding site adaptation

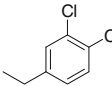
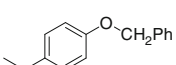
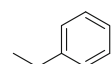
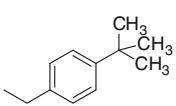
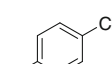

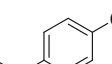
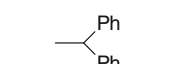
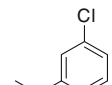
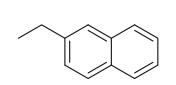
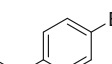
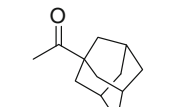
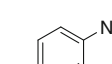
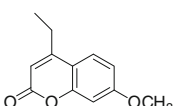
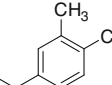
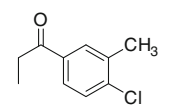
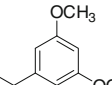
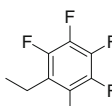
Autodock [30] was used to generate several poses of the A6 inhibitor, selected because of its higher activity and its suitable size. Several distinct docking poses were used as starting points for MD simulation of 1 ns. At the end of each simulation, the energy of the interaction ligand-surroundings was evaluated to decide which complex was the most favorable. The binding pocket of the best binding pose was selected to be used to re-dock A6 in order to obtain important interactions within the binding site. Then, like in the previous docking, a few poses were selected and resubmitted to MD simulations. When inspecting the obtained results, if eventually no further improvements could be made, the binding pocket was considered to be adapted to A6. The optimum binding pocket was then used to dock the remaining analogues with a higher consistency when compared to the original mepacrine pocket. Fig. 2 gives a general representation of the desired procedure. A similar procedure was performed for every ligand in order to obtain the most favorable binding mode.

Molecular dynamics simulations

Each complex obtained from the docking procedure was submitted to a MD simulation. The complex was first placed in a cubic box with a minimum 10 Å distance from the protein surface and the box border, which was then filled with SPC/E water molecules. Counterions were added, so that the system had zero net charge. Periodic boundary conditions were observed in the cubic box when submitted to the simulations. The initial complexes had their geometry optimized employing the steepest descent and conjugated gradients minimization algorithms. As convergence criterion, the system energy was minimized until the maximum force acting on the atoms was not higher than 50 N. If this criterion was not achieved, a second order algorithm (LBFGS) [31] would be used to reach system convergence.

Water positions were relaxed using protein and ligand positions restrained in MD simulations at 298 K. Then, the

Table 1 Chemical structures of phenothiazine derivatives and their respective biological data, expressed as the free binding energies (ΔG , kcal/mol)


Code	R	ΔG (kcal/mol)	K_i (μm)	Code	R	ΔG (kcal/mol)	K_i (μm)
A6		9.50	0.12	10a		8.68	0.47
1a		8.08	1.3	11a		8.46	0.68
2a		7.88	1.8	12a		7.74	2.3
3a		7.92	1.7	13a		8.44	0.71
4a		7.71	2.4	14a		7.97	1.55
5a		7.99	1.5	15a		8.11	1.23
6a		7.58	2.98	16a		8.58	0.56
7a		8.68	0.47	17a		8.02	1.43
8a		8.39	0.77	18a	$-\text{CH}_3$	8.18	1.1
9a		7.93	1.67				

whole system was warmed up by performing 50 ps simulations at 50, 100, 200 and 350 K. Subsequently, the system was cooled down to 300 K and simulated during 1000 ps. Long-range electrostatic interactions were calculated using the particle-mesh Ewald method. Lennard-Jones and short-range Coulombic interactions were cut off at 1.1 nm. Pressure was kept constant by a Parrinello-Rahman barostat and the temperature was kept constant using a Berendsen thermostat. Integration was carried out

every one fs time step. Backbone atomic positions were restrained during the simulations. The stabilization of the root mean square deviation (RMSd) values of the binding site relative to the first frame of MD simulation at 300 K was used as criterion to select the time interval to provide the conformations for the *RD-LQTA-QSAR* analysis. If one nanosecond simulation was not enough for RMSd stabilization a further one ns simulation would be employed. The “modified protein structure” corresponds to

Table 1 continued

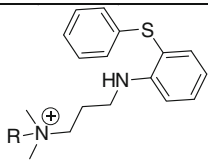
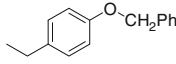
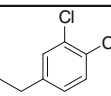
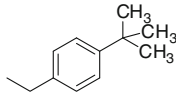

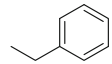
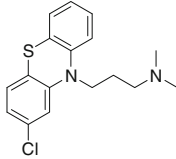
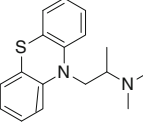
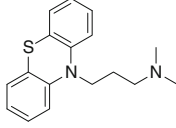
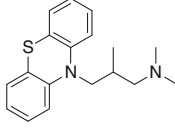
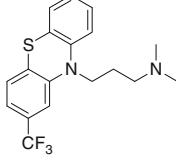
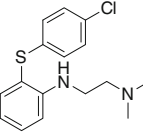
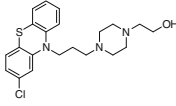
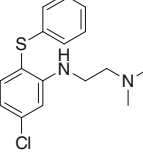
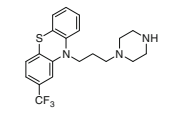
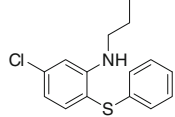
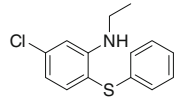
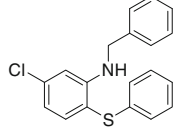
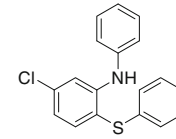
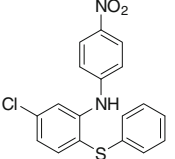
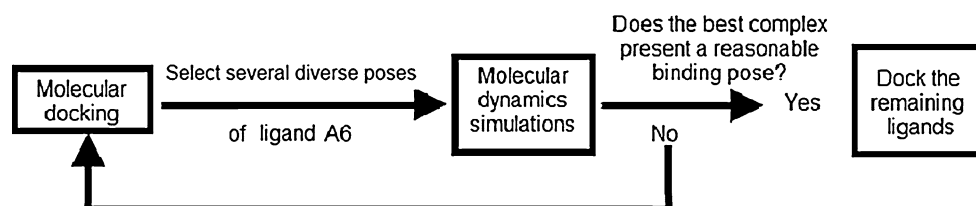
							
1b		6.36	6.6	4b		5.35	1.69
2b		5.03	6.5	5b		7.79	5.3
3b		6.41	14.2				
6		6.81	10.8	13		7.12	216
7		5.44	108	14		7.24	127
8		5.43	110	15		6.33	280
9		6.41	21.2	16		5.99	90
10		6.36	23.0	17		6.20	11.3
11		6.45	20	18		5.84	24.7
12		6.33	24.5	19		7.12	42.8

Fig. 2 Iterative scheme used to adapt the *T. cruzi* TR binding site for the phenothiazine derivatives



the average of conformations regarding the time span where the RMSD values were stable.

Receptor Dependent LQTA-QSAR alignment, model building and validation

The conformations required to create the time dependent molecular interaction field descriptors to build the LQTA-QSAR models were obtained from the MD simulations. Every simulation uses initial coordinates resulting from the docking of each ligand to the binding site, which was adapted as previously described. The aligned conformations from the MD simulations of each one of the 38 ligands (see Table 1) provide the conformational ensemble profile (CEP) of each ligand. The molecular alignment of the MD frames was performed using the GROMACS program *g_confrms*. A novel procedure of molecular alignment was adopted in this work where, instead of using only the ligand atoms, a few atoms from the binding pocket were also selected (see Fig. 3). Such receptor-based alignments proved to be advantageous compared to a ligand-based alignment. The atoms were chosen to comprise the whole extent of the ligand and adjacent ones in the binding site. Since the ligands are confined inside the binding pocket no significant differences on the resulting

CEPs are observed independent of the atoms chosen. This strategy enabled the maintenance of each ligand in a particular binding mode.

The aligned CEP of each ligand was the input to the LQTAgrid [16] module to generate the interaction descriptors. These descriptors are the interaction energy with the probe for every conformation divided by the number of conformations. A box size of $22 \times 22 \times 18 \text{ \AA}$, with 1 \AA resolution, was used to compute interaction energy descriptors, producing a total of 17,424 descriptors. The probe used was $-\text{NH}_3^+$, representing the *N*-terminus moiety of a protein, to generate the Lennard-Jones interaction descriptors (LJ). The probe had a +1 point charge to produce Coulomb interaction descriptors (QQ).

LJ descriptors were truncated according to Eq. 1 to avoid positive values with high orders of magnitude and, at the same time, to keep information in the region close to the atoms of the molecule.

$$\begin{aligned}
 \text{if } LJ_{x,y,z} \leq 30 \text{ kcal mol}^{-1} \text{ then } LJ'_{x,y,z} &= LJ_{x,y,z} \\
 \text{if } LJ_{x,y,z} > 30 \text{ kcal mol}^{-1} \text{ then } LJ'_{x,y,z} &= 30 \\
 &+ \log_{10} \left(\frac{LJ_{x,y,z}}{\text{kcal mol}^{-1}} - 29 \right) \quad (1)
 \end{aligned}$$

The descriptors were arranged in a matrix *X* of dimensions $(38 \times 17,424)$. Descriptors associated with the probe positions excessively distant from the CEP, and difficult to interpret, were eliminated a priori [32]. In order to do so, variables with variance lower than 0.01 kcal/mol were excluded. Descriptors having absolute values of Pearson correlation coefficients with the biological activity lower than 0.3, expressed as ΔG , were eliminated. In addition, descriptors whose scatter plots versus the dependent variable ΔG showed non-uniform dispersion were also filtered using the CDDA digital filter approach [33]. After filtering, those descriptors which presented PLS regression coefficient signs different from the sign of their correlation coefficient with the biological activity were manually eliminated [34].

Internal cross-validation was carried out using the leave-one-out (LOO) procedure to define the optimum number of factors in PLS. The final model was validated applying the leave-*N*-out (LNO) cross-validation and *y*-randomization tests, which are highly recommended to check model robustness and the presence of chance correlations, respectively [35].

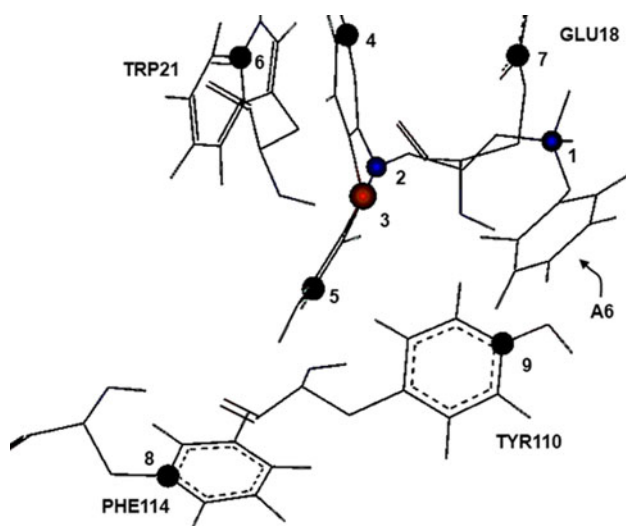


Fig. 3 Selected atoms from the ligand and binding pocket to be used in the molecular alignment employed in the LQTAQSAR methodology

In the LNO test, the training set of I samples is divided into consecutive blocks of N samples. Each block is excluded once, a new model is built without it, and the values of the dependent variable are predicted for the block in question. LNO is performed for $N = 2, 3$, etc., and the leave- N -out cross-validated correlation coefficients (Q^2 LNO, Table 1) are calculated. The model is considered robust if the deviation of the Q^2 LNO values from the Q^2 LOO reference value are smaller than 0.05 for at least 25 % of data set samples. The y-randomization test consists of several runs for which the original descriptors from matrix \mathbf{X} are kept fixed, and only the vector containing the activities, \mathbf{y} , is randomized (scrambled). The models obtained under such conditions should be of poor quality and without real meaning. A model is not considered statistically relevant if the Q^2 LOO for the scrambled \mathbf{y} is higher than 0.2 and the correlation coefficient of multiple determination (R^2) of the scrambled \mathbf{y} is higher than 0.4 [35].

A PLS model was constructed for a training set containing 32 phenothiazine derivatives, and the model's quality was evaluated regarding the Q^2 and R^2 statistical parameters, LNO and y-randomization tests [35]. The optimum or the best model was used to predict the ΔG value of six selected ligands based upon hierarchical cluster analysis to be a representative subset [34], the test set (external validation). The external predictability was evaluated by the Q_{ext}^2 statistical parameter. The descriptors of the best QSAR model were illustrated in the 3D space employing ViewerLite [36] software and a detailed interpretation was provided.

Results and discussion

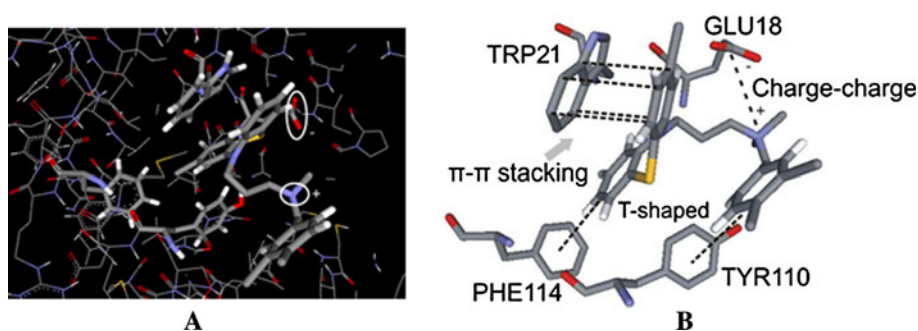
Regarding the binding site adaptation procedure, the molecular docking of A6 provided four conformations which fit in the mepacrine binding site. As already mentioned, the four selected poses for the A6 inhibitor were submitted to a MD simulation of 1 ns. The best binding pocket revealed the strong tendency of the A6 inhibitor to

be buried into the active site closer to the amino acid residue TRP21 (see Fig. 4 a). Also, the quaternary nitrogen was not placed near the Glu18 residue, so the docking procedure had to be reapplied to favor this interaction. Again, a few distinct conformations which likely interact with Glu18 were selected and submitted to MD simulations.

After the second iteration, which successfully optimized the binding site, an ionic interaction between the quaternary nitrogen of the inhibitor and the Glu18 residue of the active site was established (Fig. 4b). Other interaction features considering the inhibitor and the *T. cruzi* TR active site were observed, where reasonable π - π stacking interactions involving the Trp21 residue were refined. Additionally, an arene edge-to-face interaction could be formed involving the Phe114 and Tyr110 residues. The tricyclic antidepressants have been postulated to bind *T. cruzi* TR with the ring system lodged against the hydrophobic wall formed by Trp21 and an adjacent Met113 [19, 22]. A few docking approaches [13, 37] include a possible interaction with the Glu18 residue as observed in the *T. cruzi* TR-mepacrine complex. Additional interaction points, such as the inhibitor's quaternary nitrogen and the Glu18 residue, as well as the ring attached to the inhibitor's quaternary nitrogen and the Tyr110 residue, could explain almost 100-fold higher *T. cruzi* TR affinity found for the phenothiazine derivatives, as reported by Khan *et al.* ($A6$, K_i 0.12 ± 0.01 μM), than for chlorpromazine ($K_i = 10.8 \pm 1.1$ μM) [12].

A recent study also using molecular docking and MD simulations proposed a similar binding mode [13]. The authors suggested that phenothiazine *T. cruzi* TR inhibitors were in the vicinity of the natural substrate T[S]₂, near the enzyme disulfide bridge. In that place, derivatives would be in contact with a hydrophobic pocket surrounded by the residues Glu18, Trp21, Leu17, Tyr110, Met113, and Phe114. This proposed binding mode allows the establishment of a strong interaction between the charged amino group of the inhibitor and the carboxylate group of the Glu18 residue. This binding mode is in agreement with the hypothetical binding mode obtained in this work (see Fig 4b). However, the main difference concerns the role of

Fig. 4 **a** The most energetically favorable binding mode from the first MD simulation showing the inhibitor buried in the active site and closer to the residue Trp 21, but still far from Glu18; **b** the postulated binding mode for phenothiazine derivatives at the active site of *T. cruzi* TR, showing some key interactions with the binding site



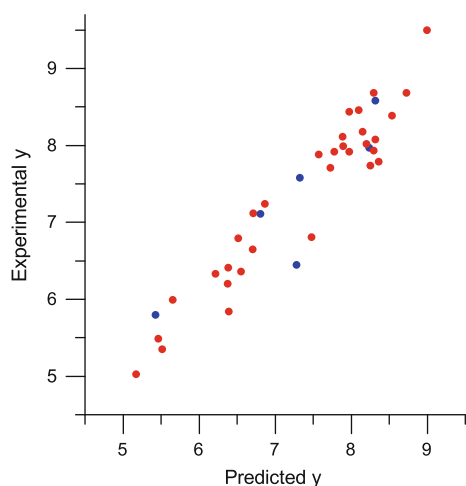


Fig. 5 Scatterplots of predicted versus experimental values for the training set (red dots) and external validation (blue dots)

the Trp21 residue, which might work as an anchor to the bent tricyclic phenothiazine ring system. All these findings were considered reliable enough to represent the binding mode of the investigated phenothiazine derivatives and provide quite good conformation sampling profiles to develop a *RD-LQTA-QSAR* approach.

Each particular phenothiazine analogue tends to bind somewhat differently inside the binding site. For this reason, the alignment procedure has to be able to maintain

every particular ligands binding model. By considering atoms from both, the ligands and binding site, it was possible to retain a common ligands binding mode, providing good CEPs that could generate reliable descriptors and ultimately 4D-QSAR models of good quality.

After the descriptors filtering and selection, accomplished by the ordered prediction algorithm (OPS) [38], a model having good statistical parameters ($Q^2_{LOO} = 0.87$, $R^2 = 0.92$) and suitable external prediction ($Q^2_{ext} = 0.78$) was obtained. These findings underline the significance of using alignments based upon particular ligand binding modes. A PLS model containing 3 LV was built using the 7 selected descriptors. Figure 5 shows scatterplots of predicted versus experimental and also the residual for the training set, the results are shown in detail in Table 2. The external data set was small due to the reduced number of compounds that contain reliable experimental data.

The leave-*N*-out validation test provided a proper indication regarding robustness. The best model maintained its predictability up to $N = 10$ samples out (32 % samples out) when the deviation values $Q^2_{LNO} - Q^2_{LOO}$ became higher than 0.05 and had elevated standard deviations (Fig. 6A). The *y*-randomization test also presented satisfactory results. The procedure was run 100 times. The random models had R^2 values lower than 0.4 and Q^2 values lower than 0.2 (Fig. 6b). The intercept test for *y*-randomization as described by Eriksson et al. [39] was carried out.

Table 2 Experimental and predicted ΔG values for training and test sets

Mol	Predicted	Experimental	Percent error (%)	Mol	Predicted	Experimental	Percent error (%)
16a	8.32 ^e	8.58 ^e	-3	15a	7.89	8.11	-3
14a	8.24 ^e	7.97 ^e	3	16a	8.20	8.02	2
6a	7.33 ^e	7.58 ^e	-3	18a	8.15	8.18	0
13	6.81 ^e	7.11 ^e	-4	2b	5.17	5.03	3
9	7.28 ^e	6.45 ^e	11	4b	5.51	5.35	3
11	5.43 ^e	5.80 ^e	-7	3b	6.38	6.41	0
1a	8.32	8.08	3	1b	6.55	6.36	3
2a	7.57	7.88	-4	5b	8.36	7.79	7
3a	7.97	7.92	1	7	7.78	7.92	-2
4a	7.73	7.71	0	8	6.70	6.65	1
5a	7.89	7.99	-1	14	6.87	7.24	-5
A6	9.00	9.50	-6	10	6.51	6.79	-4
7a	8.30	8.68	-5	15	6.21	6.33	-2
8a	8.54	8.39	2	16	5.66	5.99	-6
9a	8.30	7.93	4	6	7.48	6.81	9
10a	8.73	8.68	1	17	6.38	6.20	3
11a	8.10	8.46	-5	18	6.39	5.84	9
12a	8.25	7.74	6	12	5.47	5.49	0
13a	7.97	8.44	-6	19	6.71	7.12	-6

Values are in kcal, ^e test set

Fig. 6 **a** LNO cross-validation results obtained for the final QSAR model. The LNO results are plotted as the difference $Q^2_{LNO} - Q^2_{LOO}$ and the related standard deviation for 20 repetitions. **b** y-randomization results plotted as R^2 versus Q^2 values

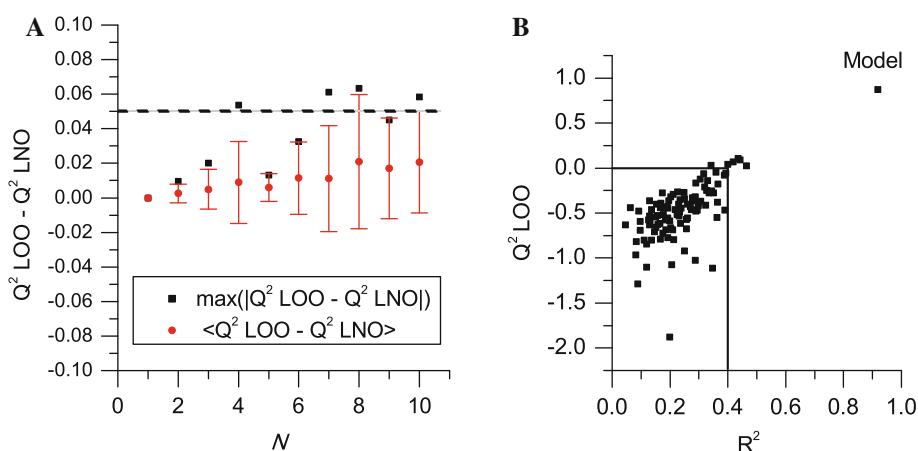
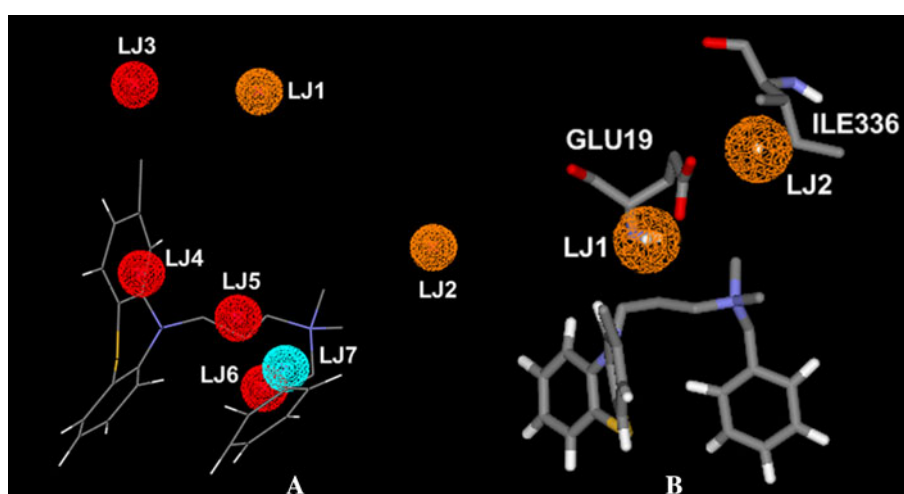


Fig. 7 **a** Descriptors (blue, orange and red spheres) of the best model; **b** Descriptors related to the amino acid residues within the binding pocket (orange spheres). Descriptors in red and orange have a negative regression coefficient sign; **LJ7** (blue) is the only descriptor which has a positive regression coefficient sign



The intercept obtained for R^2 was 0.18, which is lower than the maximum allowed value of 0.3. For Q^2_{LOO} the intercept value was -0.58, which is also lower than the 0.05. These results also attest to the model's absence of chance correlation.

The descriptors of the best QSAR model are graphically represented in two views (Figs. 7a, b). A few descriptors were sufficient to describe the ligand free energy of interaction. None of the QQ descriptors appeared in the final model since augmented ligand positive charge does not necessarily enhance affinity towards *T. cruzi* TR. LJ descriptors are depicted in different colors in Figs. 7a, b, where each one represents its peculiar contribution to the model. The descriptors **LJ1** and **LJ2** are associated to the probe placed far enough from the CEPs, and they represent interactions with the binding residues. Red **LJ3-6** and blue **LJ7** descriptors (Fig. 7a) are related to the probe closer to the CEPs and describe structural information related to the conformational flexibility and shape of the ligands.

LJ1 and **LJ2** can be better interpreted when the closer residues in the binding site are plotted simultaneously

(Fig. 7b). **LJ1** can be related to the ligand's ability to get closer to the Glu18 residue and provide more efficiently a charge-charge interaction type. Another interesting descriptor is **LJ2**, which points out a hydrophobic interaction that can be further explored when dealing with molecular modifications. Increasing the length of the quaternary nitrogen substituent (methyl group to ethyl or longer) can improve the interaction with the Ile335 residue resulting in better ligands. The more negative **LJ1** and **LJ2** descriptors values are, the more effective is the interaction with these regions in the active site, thus the free energy of binding is higher.

The descriptors **LJ3** to **LJ7** are associated to molecular modification of the various ligands and their ability to occupy certain positions in the *T. cruzi* TR binding pocket. **LJ3** is related to chlorine, $-\text{CF}_3$ or hydrogen, for example, in the phenothiazine ring. **LJ4** can be associated to some analogues where the tricyclic ring system is opened (less potent molecules). **LJ5** and **LJ6** can be related to the number of carbon atoms in the connecting chain or to the volume around the quaternary nitrogen. **LJ7** is positively

associated to an aromatic ring as substituent at the quaternary nitrogen, which improves the free energy of binding through an additional interaction to the Tyr110 residue.

Conclusions

A receptor dependent, time dependent LQTA-QSAR methodology (*RD-LQTA-QSAR*) was introduced in this work. The *RD-LQTA-QSAR* is an improvement of the receptor independent LQTA-QSAR methodology, but it can only be applied when the target structural information or, at least, a good homology model is available, associated to reliable experimental interaction data of the respective ligands. The results can be used directly for structure-based drug design, being a promising tool for drug discovery.

A novel alignment procedure for the investigated ligands was proposed to provide unbiased conformational ensemble profiles and to be used for calculating the LQTA-QSAR descriptors employed in the construction of a robust QSAR model with suitable predictive power. The importance of optimizing the binding pocket in the complex *T. cruzi* TR-mepacrine to accommodate structurally different ligands was explored and found to be crucial.

The binding mode of the bent tricyclic inhibitors of TR was postulated. The interactions of the residues present only in TR, rather than GR, was explored to provide a reasonable binding mode that is able to explain the selectivity, as well as the potency of the studied dataset.

The findings reported in this study can be quite useful for the rational design and screening of new *T. cruzi* TR inhibitors with better pharmacokinetic and lower toxicity properties. Since weak correlations of in vivo activities against *T. cruzi* and in vitro inhibition of *T. cruzi* TR are found, molecular modifications might be pursued to improve such drug delivery aspects as cell penetration. Improving ligand hydrophobicity can be used to solve such issues by exploring *RD-LQTA-QSAR* interaction descriptor with the nearby isoleucine.

Acknowledgments The authors thank the Brazilian scientific funding agencies, CAPES and FAPESP, for financial support and Prof. Dr. Carol H. Collins for English revision.

References

- World Health Organization (2010) First WHO report on neglected tropical diseases: working to overcome the global impact of neglected tropical diseases. WHO Press 75–81
- Coura JR, Vinas PA (2010) Chagas disease: a new worldwide challenge. *Nature* 465:S6–S7
- Duschak VG, Couto A (2007) An insight on targets and patented drugs for chemotherapy of Chagas disease. *Recent Pat Anti-Infect Drug Discov* 2:19–51
- Schirmer RH, Joachim GM, Krauth-Siegel RL (1995) Disulfide-reductase inhibitors as chemotherapeutic agents: the design of drugs for trypanosomiasis and malaria. *Angew Chem Int Edit* 34: 141–154
- Christina BL, Ilme S, Wolfgang K, Emil FP, Krauth-Siegel RL (1994) The structure of *Trypanosoma cruzi* trypanothione reductase in the oxidized and NADPH reduced state. *Proteins* 18: 161–173
- Bond CS, Zhang Y, Berriman M, Cunningham ML, Fairlamb AH, Hunter WN (1999) Crystal structure of *Trypanosoma cruzi* trypanothione reductase in complex with trypanothione, and the structure-based discovery of new natural product inhibitors. *Struct Fold Des* 7:81–89
- Yihong Z, Charles SB, Susan B, Mark LC, Alan HF, William NH (1996) The crystal structure of trypanothione reductase from the human pathogen *Trypanosoma cruzi* at 2.3 Å resolution. *Protein Sci* 5:52–61
- Bonse S, Santelli-Rouvier C, Barbe J, Krauth-Siegel RL (1999) Inhibition of trypanosoma cruzi trypanothione reductase by acridines: kinetic studies and structure-activity relationships. *J Med Chem* 42:5448–5454
- Krauth-Siegel RL, Inhoff O (2003) Parasite-specific trypanothione reductase as a drug target molecule. *Parasit Res* 90:S77–S85
- Benson TJ, McKie JH, Garforth J, Borges A, Fairlamb AH, Douglas KT (1992) Rationally designed selective inhibitors of trypanothione reductase. Phenothiazines and related tricyclics as lead structures. *Biochem J* 286:9–11
- Faerman CH, Savvides SN, Strickland C, Breidenbach MA, Ponasik JA, Ganem B, Ripoll D, Luise Krauth-Siegel R, Andrew Karplus P (1996) Charge is the major discriminating factor for glutathione reductase versus trypanothione reductase inhibitors. *Bioorg Med Chem* 4:1247–1253
- Khan MOF, Austin SE, Chan C, Yin H, Marks D, Vaghjiani SN, Kendrick H, Yardley V, Croft SL, Douglas KT (2000) Use of an additional hydrophobic binding site, the Z site, in the rational drug design of a new class of stronger trypanothione reductase inhibitor, quaternary alkyl ammonium phenothiazines. *J Med Chem* 43:3148–3156
- Iribarne F, Paulino M, Aguilera S, Tapia O (2009) Assaying phenothiazine derivatives as trypanothione reductase and glutathione reductase inhibitors by theoretical docking and molecular dynamics studies. *J Mol Graph Model* 28:371–381
- Galeazzi R (2009) Molecular dynamics as a tool in rational drug design: current status and some major applications. *Curr Comput-Aided Drug Des* 5:225–240
- Osvaldo AS-F, Anton JH (2002) The 4D-QSAR paradigm: application to a novel set of non-peptidic HIV protease inhibitors. *Quant Struct-Act Relat* 21:369–381
- Martins JPA, Barbosa EG, Pasqualoto KFM, Ferreira MMC (2009) LQTA-QSAR: a new 4D-QSAR methodology. *J Chem Inf Model* 49:1428–1436
- Hopfinger AJ, Wang S, Tokarski JS, Jin B, Albuquerque M, Madhav PJ, Duraiswami C (1997) Construction of 3D-QSAR models using the 4D-QSAR analysis formalism. *J Am Chem Soc* 119:10509–10524
- Van Der David S, Erik L, Berk H, Gerrit G, Alan EM, Herman JCB (2005) GROMACS: fast, flexible, and free. *J Comput Chem* 26:1701–1718
- Chan C, Yin H, Garforth J, McKie JH, Jaouhari R, Speers P, Douglas KT, Rock PJ, Yardley V, Croft SL, Fairlamb AH (1998) Phenothiazine inhibitors of trypanothione reductase as potential antitrypanosomal and antileishmanial drugs. *J Med Chem* 41: 148–156
- Parveen S, Khan MOF, Austin SE, Croft SL, Yardley V, Rock P, Douglas KT (2005) Antitrypanosomal, antileishmanial, and anti-malarial activities of quaternary arylalkylammonium 2-amino-4-

- chlorophenyl phenyl sulfides, a new class of trypanothione reductase inhibitor, and of N-acyl derivatives of 2-amino-4-chlorophenyl phenyl sulfide. *J Med Chem* 48:8087–8097
21. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The protein data bank. *Nucleic Acids Res* 28:235–242
 22. Saravanamuthu A, Vickers TJ, Bond CS, Peterson MR, Hunter WN, Fairlamb AH (2004) Two interacting binding sites for quinacrine derivatives in the active site of trypanothione reductase: a template for drug design. *J Biol Chem* 279:29493–29500
 23. Dennington IR, Keith T, Millam J, Eppinnett K, Hovell WL, Gilliland G (2003) GaussView. Version 3.09 ed. Semichem Inc., Shawnee Mission, KS
 24. Wodrich MD, Corminboeuf C, Schreiner PR, Fokin AA, PvR Schleyer (2007) How accurate are DFT treatments of organic energies? *Org Lett* 9:1851–1854
 25. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Jr. JAM, Vreven T, Kudin KN, Burant JC, Millam JM, Iyengar SS, Tomasi J, Barone V, Mennucci B, Cossi M, Scalmani G, Rega N, Petersson GA, Nakatsuji H, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Klene M, Li X, Knox JE, Hratchian HP, Cross JB, Bakken V, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Ayala PY, Morokuma K, Voth GA, Salvador P, Dannenberg JJ, Zakrzewski VG, Dapprich S, Daniels AD, Strain MC, Farkas O, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Ortiz JV, Cui Q, Baboul AG, Clifford S, Cioslowski J, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Martin RL, Keith DJF, Al-Laham MA, Peng CY, Nanayakkara A, Challacombe M, Gill PMW, Johnson B, Chen W, Wong MW, Gonzalez C, Pople JA (2004) Gaussian 03, Revision C.02. Gaussian, Inc., Wallingford CT
 26. Breneman CM, Wiberg KB (1990) Determining atom-centered monopoles from molecular electrostatic potentials. The need for high sampling density in formamide conformational analysis. *J Comput Chem* 11:361–373
 27. Schüttelkopf AW, van Aalten DM (2004) PRODRG: a tool for high-throughput crystallography of proteinligand complexes. *Acta Crystallogr D* 60:1355–1363
 28. Delphine CB, David MR, Jan HJ (2008) Very fast prediction and rationalization of pKa values for protein-ligand complexes. *Proteins* 73:765–783
 29. Guex N, Peitsch MC (1997) SWISS-MODEL and the Swiss-Pdb viewer: an environment for comparative protein modeling. *Electrophoresis* 18:2714–2723
 30. Garrett MM, David SG, Robert SH, Ruth H, William EH, Richard KB, Arthur JO (1998) Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J Comput Chem* 19:1639–1662
 31. Liu DC, Nocedal J (1989) On the limited memory BFGS method for large scale optimization. *Math Program* 45:503–528
 32. Kubinyi H (1997) QSAR and 3D QSAR in drug design Part 1: methodology. *Drug Discov Today* 2:457–467
 33. Barbosa EG, Ferreira MMC (2012) Digital filters for molecular interaction field descriptors. *Mol Inf* 31:75–84
 34. Kiralj R, Ferreira MMC (2010) Is your QSAR/QSPR descriptor real or trash? *J Chemom* 24:681–693
 35. Kiralj R, Ferreira MMC (2009) Basic validation procedures for regression models in QSAR and QSPR studies: theory and application. *J Braz Chem Soc* 20:770–787
 36. Accelrys (2002) ViewerLite, 5.0th edn. San Diego, CA
 37. Horvath D (1997) A virtual screening approach applied to the search for trypanothione reductase inhibitors. *J Med Chem* 40:2412–2423
 38. Teófilo RF, Martins JPA, Ferreira MMC (2009) Sorting variables by using informative vectors as a strategy for feature selection in multivariate regression. *J Chemometr* 23:32–48
 39. Eriksson L, Jaworska J, Worth AP, Cronin MT, McDowell RM, Gramatica P (2003) Methods for reliability and uncertainty assessment and for applicability evaluations of classification- and regression-based QSARs. *Environ Health Perspect* 111:1361–1375