Novel chemical scaffolds against *Schistosoma mansoni* discovered by combi-QSAR, virtual screening, and confirmed by experimental evaluation with automated imaging

Cleber C. Melo-Filho[1,§], Rafael F. Dantas[2,§], Rodolpho C. Braga[1], Bruno J. Neves[1], Mario R. Senger[2], Walter C.G. Valente[2], João M. Rezende-Neto[2], Willian T. Chaves[2], Eugene N. Muratov[3], Ross A. Paveley[4], Nicholas Furnham[4], Lee Kamentsky[5], Anne E. Carpenter[5], Floriano P. Silva-Junior[2] & Carolina H. Andrade*[1]*

[1] LabMol - Laboratory for Molecular Modeling and Design, Faculdade de Farmácia, Universidade Federal de Goiás, Rua 240, Qd.87, Goiânia, GO 74605-510, Brasil.

[2] Laboratório de Bioquímica Experimental e Computacional de Fármacos, Fundação Oswaldo Cruz, Instituto Oswaldo Cruz, Av. Brasil, 4365, Rio de Janeiro, RJ 21040-900, Brazil.

[3] Laboratory for Molecular Modeling, Division of Chemical Biology and Medicinal Chemistry, Eshelman School of Pharmacy, University of North Carolina, Chapel Hill, NC 27599, USA.

[4] Department of Pathogen Molecular Biology & Department of Infection and Immunity, London School of Hygiene and Tropical Medicine, London, United Kingdom.

[5] Imaging Platform, Broad Institute of Massachusetts Institute of Technology and Harvard, Cambridge, Massachusetts, USA.

[§] These authors contributed equally to this work

**Authors for correspondence: CHA: Tel: + 55 62 3209 6451; Fax: +55 62 3209 6037. E-mail: carolina@ufg.br; FPSJ: Tel: + 55 21 3865 8248; Fax: +55 21 2590 3495. E-mail: floriano@ioc.fiocruz.br*

Abstract

Background: Schistosomiasis is a neglected tropical disease that affects millions of people worldwide. Thioredoxin glutathione reductase of *Schistosoma mansoni* (*Sm*TGR) is a validated drug target that plays a crucial role in the redox homeostasis of the parasite. We report the discovery of new chemical scaffolds against *S. mansoni* using a combi-QSAR approach followed by virtual screening of a commercial database and confirmation of top ranking compounds by *in vitro* experimental evaluation with automated imaging of schistosomula and adult worms.

Results: We constructed 2D and 3D quantitative structure–activity relationship (QSAR) models using a series of oxadiazoles-2-oxides reported in the literature as *Sm*TGR inhibitors and combined the best models in a consensus QSAR model. This model was used in a virtual screen of a commercial database and allowed the identification of ten new potential *Sm*TGR inhibitors. The latter were screened on schistosomula and two active compounds were further evaluated on adult worms. **Conclusions:** We succeed to develop predictive virtual screening tool based on 2D and 3D QSAR models. After prioritizing virtual screening hits, high activity of two compounds representing new chemical scaffolds, 4-nitro-3,5-bis(1-nitro-1*H*-pyrazol-4-yl)-1*H*-pyrazole and 3-nitro-4-{[(4-nitro-1,2,5-oxadiazol-3-yl)oxy]methyl}-1,2,5-oxadiazole (LabMol-17 and LabMol-19, respectively) was experimentally validated. These compounds will be subjects for additional testing and, if necessary, modification to serve as new schistosomicidal agents.

Defined key terms

Hologram quantitative structure-activity relationship (HQSAR): 2D quantitative structure-activity relationship (QSAR) approach based on a special type of structural fingerprints, also known as a molecular hologram, where information is used as an independent variable during calculations.

Comparative molecular field analysis (CoMFA): 3D QSAR approach based on the calculation of molecular interaction fields between aligned molecules and a defined probe atom placed in different intersections of a lattice box. It uses the principle that most ligand-target interactions are noncovalent and can be described by steric and electrostatic fields.

Comparative molecular similarity indices analysis (CoMSIA): 3D QSAR approach based on the calculation of molecular similarity indices between aligned molecules and a probe atom placed in different intersections of a lattice box. In addition to steric and electrostatic fields, it calculates hydrophobic, H-bond donor and acceptor fields.

High content screening: A high-throughput phenotypic screening approach relying on automated microscopy followed by multiparametric computational image analysis.

Introduction

Schistosomiasis is one of the major neglected tropical diseases (NTDs) that affects millions of people worldwide [1]. Recent estimates suggest that at least 261 million people required preventive treatment for this disease in 2013. This parasitosis is reported in 78 countries located in sub-Saharan Africa, the Middle East, the Caribbean, and South America, resulting in 20,000 to 200,000 deaths annually [2]. The disease is caused by flatworms of the genus *Schistosoma* (*S. mansoni*, *S. japonicum*, *S. haematobium*, *S. intercalatum*, and *S. mekongi*) [3,4]. In the absence of a vaccine, praziquantel (PZQ) has been the drug of choice recommended by the World Health Organization for the treatment and control of all the major *Schistosoma* species in mass drug administration programs for almost three decades [5]. However, the disseminated and repeated use of this drug in endemic areas, because of the high incidence of reinfection, brings concerns about the development of parasitic resistance [6,7]. This problem is further emphasized by the known lack of efficacy of PZQ against juvenile worms [8], which is a potential cause of treatment failure in endemic areas. For these reasons, the development of new schistosomicidal drugs is urgently required [6–10].

The complete genome sequencing of *S. mansoni* has brought the possibility of exploring a great variety of biological targets in the search for new drugs against this parasite [11,12]. Thioredoxin glutathione reductase of *S. mansoni* (*Sm*TGR, E.C. 1.8.1.9) plays a crucial role in the redox homeostasis of the parasite [13]. *Sm*TGR is a multifunctional enzyme that acts in the detoxification of reactive oxygen species (ROS) present in the blood vessels of the mammalian host, thus allowing parasite survival. While mammalian cells use several enzymes of the glutathione and thioredoxin systems, *S. mansoni* ROS detoxification relies only on *Sm*TGR [13,14]. Moreover, it has been validated as a potential drug target as demonstrated in studies silencing *Sm*TGR expression using RNA interference [13]. Validation studies have also been performed in *S. japonicum*, confirming the importance of TGR in parasite survival [15,16].

Advances in computational hardware and software over the last few decades have enabled the development of new strategies for computer-aided/assisted drug design (CADD), which has the advantages of reducing the time and costs in the identification of new drug candidates [17–20].

Quantitative structure-activity relationship (QSAR) has been widely-used as a lead optimization tools as well as for pharmacokinetics property optimization and in virtual screening campaigns [21,22]. Our group has been working on developing and applying many CADD strategies aiming at discovering new drug candidates for neglected tropical diseases [23–35]. Another important technological advance impacting drug discovery was the introduction of automated microscopes along with powerful image analysis software enabling high-throughput phenotypic assays of cells and small organisms, a technique known as high-content screening (HCS), imaging (HCI) or analysis (HCA) [36]. Whole-organism antihelminthic screens employing the HCS approach have already proved useful for the larval stage of *S. mansoni* [37]. In this work, we report the discovery of two new antischistosomal chemical scaffolds by a combi-QSAR approach, using 2D- and 3D-QSAR models, followed by virtual screening of a commercial database and experimental evaluation of the potential *Sm*TGR inhibitors against schistosomula and adult lifecycle stages of *S. mansoni*. For the latter we present a new medium-throughput assay methodology using 96-well microplate and HCI technology. The general workflow of the study is presented in Figure 1.

Materials & Methods

Dataset

The QSAR studies were performed using a series of 35 oxadiazoles-2-oxides reported in the literature as inhibitors of *Sm*TGR whose *in vitro* enzymatic potency values (measured by IC_{50}) were obtained by the same experimental protocol [38,39]. The IC_{50} values were converted to negative logarithmic units, pIC_{50} (-log IC_{50}) and used as dependent variables in the QSAR analysis. The pIC_{50} values span an approximate range of three orders of magnitude. The dataset was manually divided into training and test sets, ensuring a representative coverage across the entire range of pIC_{50} values. The training set was used to generate the QSAR models and the test set was used for external validation. The chemical structures and corresponding IC_{50} and pIC_{50} values are listed in Table 1.

The 3D structures of the compounds were generated using OMEGA software v.2.5.1.4 [40,41]. OMEGA generates various initial conformations for each compound based on a database of pre-calculated

fragments and the structures are optimized by MMFF94 force field. All QSAR models were generated and analyzed using SYBYL-X software v.1.2 [42].

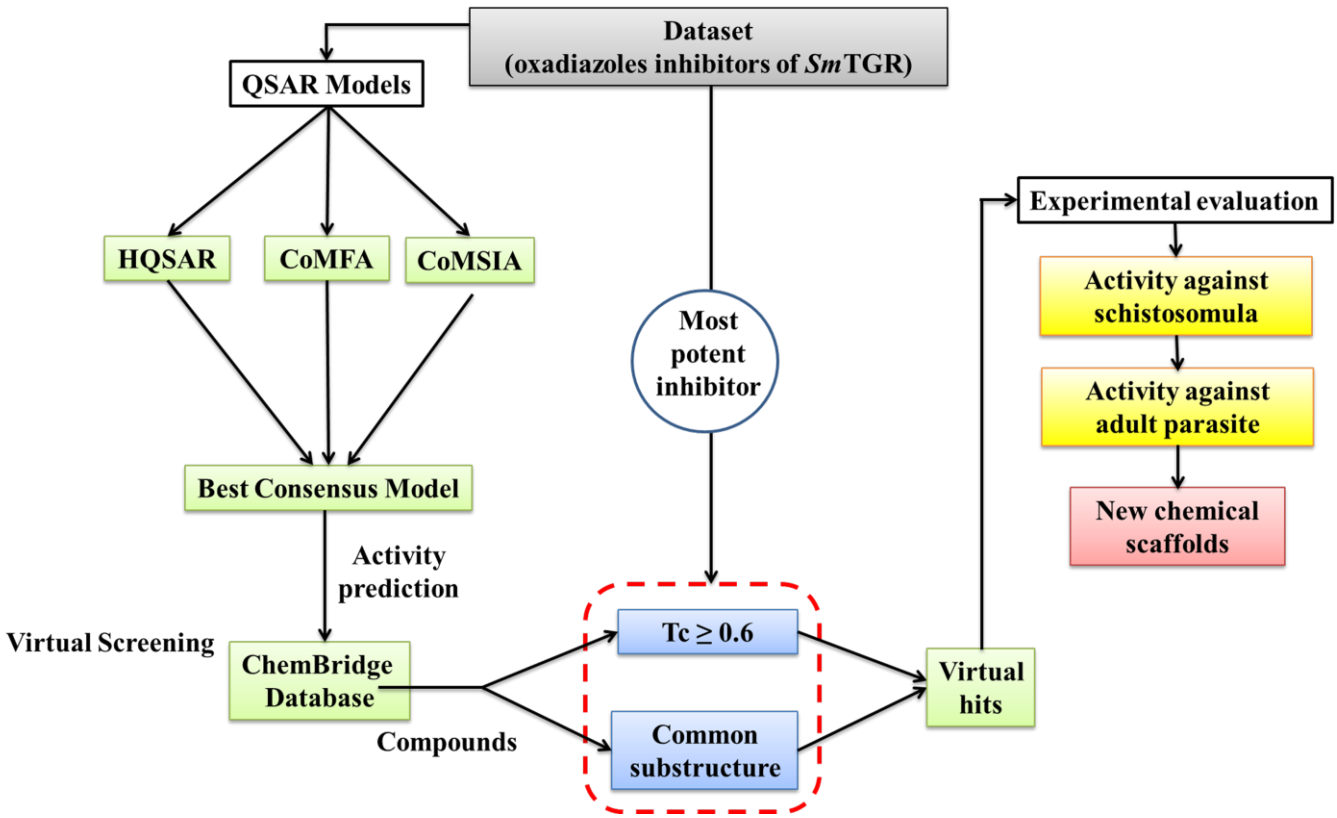


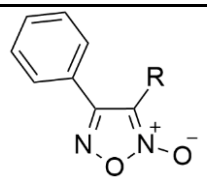
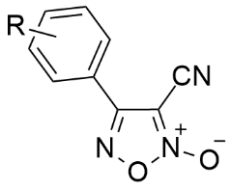
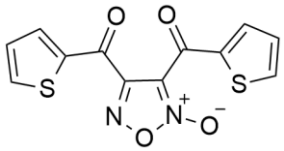
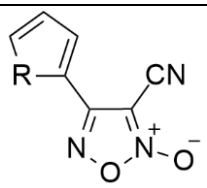
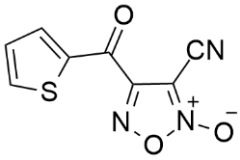
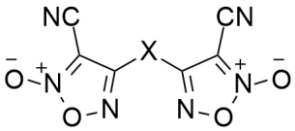
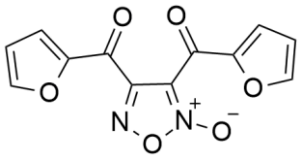
Figure 1. Flowchart summarizing the workflow of this work.

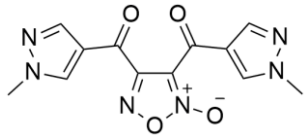
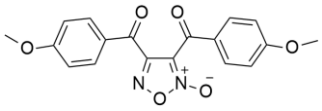
2D-QSAR

HQSAR

The HQSAR module available at SYBYL-X software v.1.2 [42] was used to HQSAR models generation. The initial models were derived using the standard fragment size (4-7 atoms) and various combinations of fragment distinctions. The models with $q^2_{LOO} > 0.5$ were used to predict the activity of all compounds of the dataset.

Table 1. Chemical structures and corresponding IC_{50} and pIC_{50} experimental values of *Sm*TGR inhibitors.

Cpd	Structure	R	X	IC_{50} (μM)	pIC_{50}
1		CN	-	6.30	5.20
2		CH_3	-	50.1	4.30
3		CH_2OH	-	11.2	4.95
4[a]		CHO	-	0.11	6.95
5		COOH	-	0.63	6.20
6[a]		$CONH_2$	-	17.7	4.75
7		3-NO_2	-	2.23	5.65
8		3-CF_3	-	2.51	5.60
9		3-Br, 4-F	-	2.81	5.55
10*		3-Br	-	2.81	5.55
11		3-Cl	-	3.54	5.45
12		4-Br	-	3.54	5.45
13*		4-Cl	-	4.07	5.39
14		4-CF_3	-	7.07	5.15
15		3-OH	-	7.07	5.15
16*		4-F	-	7.94	5.10
17		2-OMe	-	7.94	5.10
18		3,4,5-OMe	-	8.91	5.05
19*		3-OMe	-	8.91	5.05
20		4-OMe	-	10.0	5.00
21		4-Me	-	11.2	4.95
22		4-Ph	-	15.8	4.80
23		4-OH	-	18.1	4.74
24		-	-	0.04	7.39
25*		O	-	2.81	5.55
26		S	-	3.54	5.45
27		-	-	0.063	7.20
28		-	1,3-Ph	3.54	5.45
29		-	1,4-Ph	1.00	6.00
30		-	5-F-1,3-Ph	0.47	6.32
31		-	2,5-tiophene	0.40	6.39
32*		-	2,4-tiophene	0.35	6.45
33		-	-	0.038	7.42

34		-	-	1.28	5.89
35*		-	-	0.10	6.98

Cpd= compound; * Test set compounds; [a] outliers.

The final HQSAR models, after the division into training and test sets, were obtained using combinations of different fragment distinctions (A, atoms; B, bonds; C, connectivity; H, hydrogen; DA, donor and acceptors of hydrogen bonds), fragment size (2-5, 3-6, 4-7, 5-8, 6-9, 7-10 atoms) and hologram length (53-401). The partial least squares regression (PLS) [43–45] was performed to correlate the information coded in the molecular holograms with biological activity.

3D-QSAR

Atomic charges assignment

Two different charge assignment methods were tested, the empirical method Gasteiger-Hückel [46,47], available at SYBYL-X v.1.2 platform [42] and the semi-empirical AM1-BCC charges [48,49] available at QUACPAC software v.1.6.3.1 [50].

Molecular alignment

The shape-based alignment and alignment based on a morphological similarity function were evaluated. The former was performed using ROCS software v.3.2.1.4 [51,52]. First, several conformers were calculated for each molecule in OMEGA v.2.5.1.4 [40,41], which generated various initial conformations obtained from a database of pre-calculated fragments. Then, the conformers were superimposed, using ROCS, with the most potent *Sm*TGR inhibitor (**33**). The conformers were classified according to the TanimotoCombo score function [52]. The alignment based on morphological similarity

function was done in Surflex-Sim, available at SYBYL-X v.1.2 [42]. Two most potent *Sm*TGR inhibitors, compounds **24** and **33**, were chosen for template definition. Remaining compounds were superimposed with this template. The maximal number of poses generated per molecule was 20. The best pose of each compound was chosen based on the calculated similarity to the template.

CoMFA

The aligned training set molecules were placed in a 3D lattice box, with a grid spacing of 0.5, 1.0, 1.5 and 2.0 Å in the x, y, and z directions. CoMFA steric and electrostatic fields were calculated at each grid point with the Tripos force field using a carbon atom probe with sp^3 hybridization (Csp^3) and charge +1.0. The energy cutoff was set to 30 kcal/mol. The standard deviation coefficients (SDC) were used for region focusing, with values ranging from 0.3 to 1.5. The calculated electrostatic and steric fields were correlated with biological activity by PLS regression.

CoMSIA

CoMSIA models were generated using the same molecular alignments used for CoMFA. The aligned compounds were placed in a 3D lattice box, with a grid spacing of 2.0 Å. In addition to the steric and electrostatic fields, hydrophobic, hydrogen bond donor and acceptor descriptors were included in CoMSIA studies. A probe carbon atom with radius of 1.0 Å and charge +1.0 was used for obtaining the similarity indices. A Gaussian function was used to describe the energetic terms in function of the distance between the probe atom and the aligned molecules with an attenuation factor of 0.3.

QSAR model validation

The internal validation of the QSAR models was performed to evaluate their robustness, using the full cross-validation r^2 (q^2), leave-one-out (LOO) and leave-many-out (LMO) methods. The latter was used to evaluate the stability of the best models and was performed using five groups and 25 runs.

The predictive ability of the models was assessed by the correlation coefficient of prediction Q^2_{ext}, using the Q^2_{F3} equation (Eq. 1) and the variants of the modified r^2 (r_m^2): r_m^2(average) and r_m^2(delta) [53–55] (Eq. 2 and 3) .

$$Q^2_{F3} = 1 - \frac{\left[\sum_{i=0}^{n} (\hat{y}_i - y_i)^2\right] / n_{EXT}}{\left[\sum_{j=1}^{n_{TR}} (y_i - \bar{y}_{TR})^2\right] / n_{TR}} \quad (1)$$

$$r^2_{m(average)} = \left(r_m^2 + r_m'^2\right)/2 \quad (2)$$

$$r^2_{m(delta)} = \left|r_m^2 - r_m'^2\right| \quad (3)$$

In Eq. 1, $\hat{y}_i$ represents the predicted biological activity; y_i is the experimental biological activity; $\bar{y}_{TR}$ is the mean of the biological activity for training set; n_{EXT} is the number of compounds in test set; n_{TR} is the number of compounds in training set. In Eqs. 2 and 3, r_m^2 is the calculated modified r^2 and r'^2_m is the modified r^2 after the inversion of *x* and *y* axes of the plot for experimental *versus* predicted biological activities.

Consensus models

Four consensus models were generated using different combinations of the best HQSAR, CoMFA and CoMSIA models. The predicted activity of each compound was the arithmetic mean of individual models predictions. The external validation of these models was done using the same metrics applied in individual models. Moreover, three other metrics were calculated for consensus models: root mean square error of calibration (RMSEC), root mean square error of prediction (RMSEP) and Durbin-Watson test (*d*). These metrics are described elsewhere [56–60].

Virtual screening of new potential *Sm*TGR inhibitors

The virtual screening of potential *Sm*TGR inhibitors was performed based on similarity and common substructure search in the ChemBridge database [61]. First, the most potent inhibitor of the dataset (**33**)

was used as a template for the similarity search. Compounds with Tanimoto coefficient ($T_c \geq 0.6$) were selected. Additionally, the substructure search was applied to find compounds containing the common substructure, i.e., the oxadiazole ring. All compounds were prepared using the same protocol and software used in QSAR dataset preparation, i.e., 3D structure and conformer generation, partial atomic charges calculation and molecular alignment. The chosen method of alignment and partial charges calculation was the same of the best individual CoMFA and CoMSIA models. In the next step, the best consensus model was used to predict the biological activity of the potential *Sm*TGR inhibitors. The most promising compounds, with highest predicted pIC_{50}, were selected for biological evaluation. Furthermore, five highly-predictive in-house models, described elsewhere [26,27], were used to predict some ADME properties of the compounds, such as logP, Caco-2 cell permeability, blood-brain barrier penetration (BBBP), hERG inhibition, CYP3A4 inhibition and water solubility.

Biological evaluation on *S. mansoni*

Compounds and Media

Ten selected compounds were purchased from ChemBridge (San Diego-CA, USA) and given the identifiers LabMol-13 to LabMol-22. Compounds were resuspended in 100% DMSO and used immediately in the assays. DMEM and M169 media were purchased from Vitrocell (Campinas-SP, Brazil). All other reagents were purchased from Sigma-Aldrich (St. Louis-MO, USA).

Larval *Schistosoma mansoni* (schistosomula) *in vitro* assay

Schistosomula were produced by the mechanical method adapted from both Mansour et al. (2010) [62] and Marxer et al. (2012) [63]. The cercarie (*S. mansoni*, BH strain) were vortexed at maximum speed for 5 minutes for tail shedding and cercarie transformation into schistosomula. The schistosomula were resuspended in Medium 169, plated in 384 well plates (120 per well) and maintained in an incubator with 5% CO2 overnight before compound addition. Schistosomula were divided into three groups: negative

control (0.625% DMSO), positive control (10 µM of PZQ or Oltipraz (OLT)) and treated (LabMol compounds at a concentration range of 0.3125-20 µM). The effect of the compounds on schistosomula motility and phenotypes was assessed at 48h after compound addition using an automated analysis method described below.

Automated scoring of schistosomula motility and phenotype

The automated image-based method for scoring schistosomula motility and phenotype was performed as described previously [64]. Bright-field images were collected using an ImageXpressMicro HCS microscope (IXM; Molecular Devices, Wokingham, UK). For motility analysis 5 x 6 sec interval time-lapse images were collected using a 4x objective. For detailed morphology a 10x objective was used to collect 4 adjacent images fields from within a well, which were considered together to maximize larval numbers for phenotype analysis. Analysis of both the larval phenotype and motility was then carried out in Pipeline Pilot 9 as described by Paveley et al. (2012) [64]. Phenotype analysis of individual parasites was carried out by a two class Laplacian-modified Bayesian categorization model analysis of 80 image descriptors which constituted shape, size, image intensity and texture statistics and compared to a training set of data comprising 20,000 parasites. Motility analysis of individual parasites was also analyzed by the average object displacement from the origin point in subsequent 4x image across the time-frame series. Both the Baeysian phenotype and motility scores are subsequently adjusted to the control wells (DMSO treated) on each plate [64].

Adult *Schistosoma mansoni ex vivo* assays

Three- to six-days old Swiss mice were individually infected percutaneously with 150 ± 10 *S. mansoni* cercariae (BH strain). The animals were placed into cylindrical vials under incandescent light with a thin water layer containing the cercariae for a period of 30 min. At 42-49 days after infection (i.e., the time required for *S. mansoni* to reach maturity), the animals were euthanized, and the worms were perfused (with 0.85% sodium chloride and 0.75% sodium citrate solution) from mesenteric and portal hepatic veins.

Worms were rinsed and individually transferred into 96 well plates with complete DMEM media (i.e. DMEM plus 10% fetal calf serum, 2mM L-glutamine, 100 µM/ml penicillin, 100 µg/ml streptomycin). Male and female worms were distributed in three groups of six individuals each: negative control (0.02% DMSO), positive control (10 µM PZQ) and treatment (10 µM LabMol compounds). The plates were maintained at 37 °C in a humidified atmosphere of 5% CO2 throughout the entire experiment. The effect of the compounds on adult worm motility was assessed either immediately or 24, 48 or 72h after compound addition using the automated analysis method described below.

Automated measurement of adult worm mobility

Our strategy was based on sequential pairwise comparison of 100 time-lapse images captured every 250-300 ms using an automated bright-field microscope with a 2x objective lens (ImageXpress Micro XLS, Molecular Devices, CA). Subsequent quantitative image analysis used a custom-developed pipeline for detecting changes in parasite motility using the open-source CellProfiler software version 2.1.2 [65]. The pipeline along with its validation will be thoroughly described in a subsequent publication and the pipeline itself will be made freely available. Briefly, at each cycle of the pipeline, an image captured at a given instant (t_n) is compared with the image captured at the preceding instant (t_{n-1}) and so on until all images are processed. Two different motility measurements were calculated. First, a precursor metric, "AdjustedRandIndex" is calculated by comparing worm objects identified on images captured at times t_n and t_{n-1} with CellProfiler's Overlap module. This measure ranges from 0 to 1, with 1 meaning two objects are perfectly aligned (no movement). Hence, we created an "Overlap" mobility score, which is directly proportional to the amount of movement, by subtracting 1-"AdjustedRandIndex". Another motility measure, "DiffWorms", is the mean pixel intensity of the image calculated from the absolute difference of the parasite images in t_{n-1} and t_n. The higher the DiffWorms score higher is the parasite mobility measured. Both measures are iteratively taken for the 99 image pairs and scores per well are calculated by averaging over all measurements.

Statistical Analysis

All statistical analysis and graphs were performed using GraphPad Prism version 5.00 for Windows (GraphPad Software, La Jolla California USA, www.graphpad.com).

Ethics statement

Animals were maintained and experiments carried out in accordance with the Institutional Ethics Committee for Laboratory Animal Use at the Oswaldo Cruz Foundation (CEUA/FIOCRUZ, Brazil; license number, LW-78/12) or using the NC3Rs and ARRIVE guidelines under the United Kingdom Animal's Scientific Procedures Act 1986 with approval from the London School of Hygiene and Tropical Medicine Ethics committee.

Results and discussion

Hologram length and fragment size and distinction can affect the quality of HQSAR models [43,66]. In this work, various combinations of these parameters were tested. The final HQSAR models were obtained using combinations of different fragment distinctions (A, atoms; B, bonds; C, connectivity; H, hydrogen; DA, donor and acceptors of hydrogen bonds), fragment size (2-5, 3-6, 4-7, 5-8, 6-9, 7-10 atoms) and hologram length (53-401).

Two compounds were identified as outliers (compounds **4** and **6**, Table 1). These compounds presented residual values higher than two times the standard deviation in models containing all compounds of the dataset with the standard fragment size (4-7 atoms) and various combinations of fragment distinctions. Besides presenting high residue values, compounds **4** and **6** have unique structural features that are not observed in the other molecules of the dataset, i.e., the presence of aldehyde and amide functional groups.

Although the measures of internal consistence (q^2 and r^2) are important parameters for the evaluation of QSAR models, in some cases they do not reflect the external predictivity of the models [67]. Thus, in this study, all QSAR models were validated using the external predictivity parameters (Q^2_{ext}, $r_m^2{}_{(average)}$

and r_m^2(delta)), calculated after the prediction of biological activity of the test set compounds, which were completely excluded from models generation.

The three best HQSAR models are presented in Table S1 (Supplementary Material). They have similar statistical characteristics but model 2 showed a slight superiority when evaluated the external set. The best HQSAR model was obtained using the fragment distinction A/DA (Table 2). The predicted activity (pIC_{50}) for the test set compounds using this model (Table S2, Supplementary Material) indicated that only two compounds had their predicted values greater than the standard deviation of residuals, indicating a good predictive capacity of the model. The plot of the experimental versus predicted biological activity of the best HQSAR model is displayed in Supplementary Materials (Figure S1A).

Table 2. Statistical characteristics for the best individual QSAR models obtained.

Model	q^2_{LOO}	q^2_{LMO}	r^2	N	Q^2_{ext}	r_m^2(average)	r_m^2(delta)
HQSAR	0.61	0.57	0.85	4	0.94	0.88	0.02
CoMFA-model I	0.71	0.66	0.99	6	0.90	0.83	0.04
CoMFA-model II	0.72	0.67	0.99	6	0.82	0.69	0.17
CoMSIA-model I	0.51	0.48	0.99	6	0.95	0.88	0.06
CoMSIA-model II	0.60	0.55	0.99	6	0.96	0.90	0.04

q^2_{LOO}, leave-one-out cross-validated correlation coefficient; q^2_{LMO}, leave-many-out cross-validated correlation coefficient; r^2, noncross-validated correlation coefficient; N, optimal number of latent variables in PLS analysis; Q^2_{ext}, correlation coefficient of prediction; r_m^2(average) and r_m^2(delta), variants of the modified r^2.

Besides predicting the biological activity of untested compounds, HQSAR models also give information regarding the relationships between the structural fragments and the biological activity, which can be visualized through the contribution maps. These maps indicate the individual contribution of each atom or fragment for the activity by color-coded schemes. Colors around red spectrum (orange, red orange and red) indicate negative contribution, while colors around green (yellow, green blue, green) indicate positive contribution to biological activity. The contribution maps for the most (**33**) and the less (**2**) potent compounds of our data set are presented in Figure 2.

The contribution map for the most potent inhibitor of the dataset (**33**, Figure 2) suggests that the oxygen atom (O_{11}) from carbonyl group has positive contribution for the biological activity.

Moreover, the furan ring is important for the biological activity since the O_{17} has a positive contribution for activity. The carbon atoms C_{10}, C_{13} and C_{14} also have positive contribution to the biological activity. Despite the fragments in green and yellow, the contribution map for the less potent inhibitor (**2**) suggests that the hydrogen atom attached to the carbon atom C_{12} negatively contribute to biological activity. The absence of the two carbonyl groups and furan rings in compound **2** suggests that these groups may play a critical role in *Sm*TGR inhibition, because the activity decreased three logarithmic units in comparison to compound **33**.

Previous study carried out by Gasco and co-workers [68] indicated that the oxadiazoles are capable of releasing nitric oxide in solution containing thiols by nucleophilic attack in C_3 and C_4 carbon atoms. Because *Sm*TGR has a selenocysteine (Sec) residue in the C-terminal end, the oxadiazoles may undergo nucleophilic attack mainly by this residue due to its superior reactivity in comparison to cysteine. The resulting nitric oxide release may be the reason for the antiparasitic activity described for these compounds [38]. The presence of the carbonyl group in the most potent inhibitors of the dataset suggests an important role of this group, which behaves as a linker that favors the nucleophilic attack in C_3 and C_4 atoms.

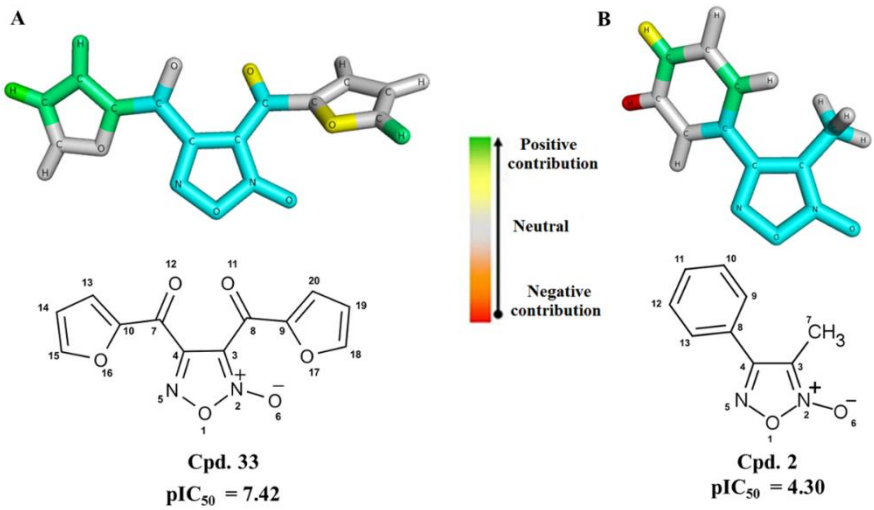


Figure 2. HQSAR contribution maps for the most (**A, 33**) and less (**B, 2**) potent *Sm*TGR inhibitors. The 1,2,5-oxadiazole ring is highlighted in blue, which is the maximal common substructure.

In CoMFA and CoMSIA studies, the calculation of electrostatic descriptors depends on the assignment of partial atomic charges of the compounds. Therefore, the charge assignment method is critical to the success and may affect the quality of the developed models [24]. Furthermore, the contour maps may have some differences depending on the method used for partial atomic charges assignment [69,70]. Two different charge assignment methods were tested in this study, the empirical method Gasteiger-Hückel,[46,47] and the semi-empirical AM1-BCC charges [48,49].

Another crucial aspect in 3D-QSAR studies is the structural alignment, which is used to represent the probable bioactive conformation of the compounds. The quality of the 3D-QSAR models can be directly affected by the structural alignment [71]. Therefore, two ligand-based alignment hypotheses were tested, the shape-based alignment using ROCS software v.3.2.1.4 [51,52] and the morphological similarity alignment, implemented in Surflex-Sim, available at SYBYL-X v.1.2 platform [42]. The data set was aligned using the two schemes, displayed in Figure 3. In alignment 1, the best conformer of each molecule, after superposition, was classified according to *TanimotoCombo* score function. *TanimotoCombo* is a combination of the functions *ShapeTanimoto*, which compares the molecules according to the best molecular volume superposition, and *ColorTanimoto* which is related to the appropriate superposition of groups with certain properties like hydrogen bond donors and acceptors, hydrophobic, cation, anion and rings [52]. In alignment 2, no previous conformers were generated since Surflex-Sim has fast techniques to generate poses. Additionally, a morphological similarity function is used to align the molecules. This function is defined as a Gaussian function of the distances of two molecules to observation points of a grid [72].

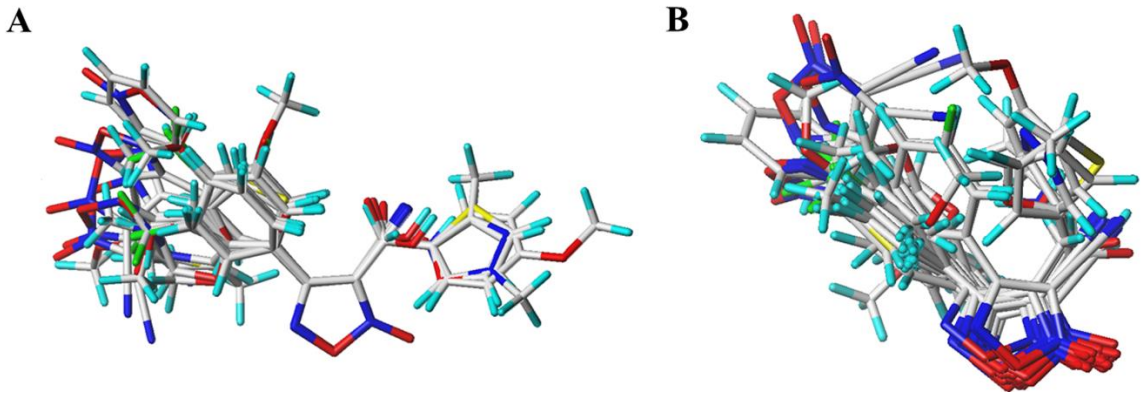


Figure 3. Molecular alignment used in 3D-QSAR studies. (A) Shape-based alignment performed in ROCS; (B) Surflex-Sim alignment which uses a morphological similarity function.

In 3D-QSAR, similarly to HQSAR, compounds **4** and **6** were identified as outliers, the same identified in HQSAR modeling. The outlier detection was performed using a CoMFA model generated with all compounds of the dataset and default grid spacing of 2.0 Å. The compounds **4** and **6** were identified as outliers, since their calculated residues were near or higher than two times the standard deviation of residues.

CoMFA and CoMSIA models were investigated by PLS analysis, using the full cross-validated r^2 (q^2) leave-one-out (LOO) method. The leave-many-out (LMO) method was used to evaluate the stability of the best models. To evaluate the 3D-QSAR models predictive power, the final models were externally validated with the same test set of compounds. The biological activity was predicted for the test set molecules, and the Q^2_{ext}, r_m^2(average) and r_m^2(delta) were calculated. The full results of the best CoMFA and CoMSIA models are displayed in Supplementary Materials (Tables S3 and S4, respectively).

The two best CoMFA models presented good internal consistence and high external predictivity (Table 2). The statistical results for the external validation (Q^2_{ext}, r_m^2(average) and r_m^2(delta)) were within the recommended values for predictive models [73]. The models generated using the AM1-BCC charges and ROCS alignment (Models I and II) showed superior performance in comparison with those generated using Gasteiger-Hückel charges and Surflex-Sim alignment (Table 2 and Table S3, Supplementary Material).

The best CoMSIA models were obtained using the steric, electrostatic, hydrophobic and H-bond acceptor fields (Table S4, Supplementary Material). The two best CoMSIA models were obtained using ROCS alignment, one using Gasteiger-Hückel charges (Model I, Table 2) and the other using AM1-BCC charges (Model II, Table 2). These two models have similar results for internal validation and external predictivity. The differences between the two best CoMSIA models can be better visualized in the prediction of the activity of test set compounds (Table S2, Supplementary Material). The graphic results for the experimental *versus* predicted activities of both compound sets (training and test sets) are displayed in Figure S1B and S1C (Supplementary Material).

The best CoMFA and CoMSIA models were used to generate contour maps. These maps indicate regions where certain types of interactions are favorable or unfavorable for biological activity [74]. The interpretation of contour maps is useful to guide the design of new potent inhibitors of *Sm*TGR. CoMFA and CoMSIA contour maps were generated using the STDEV*COEFF field type and the function *Contour by actual*. Figure 4 shows contour maps obtained from (A) CoMFA steric (green/yellow) and (B) electrostatic (red/blue) fields; (C–F) CoMSIA steric (green/yellow), electrostatic (red/blue), hydrophobic (yellow/gray) and hydrogen bond acceptor (purple/magenta) fields with the most potent compound of the data set **(33)**.

The steric contour map of the best CoMFA model (Figure 4A) shows green contours surrounding the furan ring of the compound (**33**), indicating that bulky substituents could be favorable to biological activity. Figure 4B shows red regions near the oxygen atoms O_{12} and O_{11} of the carbonyl groups, where substitution for electronegative groups can favor the biological activity. Furthermore, this map shows blue regions surrounding the carbon atoms of the furan rings, indicating that electronegative substituents at the furan rings are unfavorable. This map indicates that the carbonyl group represented by C_7 and O_{12} atoms is important for biological activity. For the CoMFA and CoMSIA models, steric and electrostatic contribution maps were similar. The steric contour map of the best CoMSIA (Figure 4C) model also indicates that bulky groups in the region near the furan ring are favorable to biological activity. The electrostatic map (Figure 4D) indicates that electronegative groups in the region of the oxygen atoms of

the carbonyl groups (O_{12} and O_{11}) are favorable. Furthermore, electropositive groups in the region near the carbon atom C7 are favorable. In addition, the electrostatic CoMSIA map indicates that electronegative groups near the oxygen atom O17 of the furan ring are favorable to biological activity. The hydrophobic contour map (Figure 4E) shows two gray contours near the oxygen atoms of the carbonyl groups (O12 and O11), indicating that hydrophobic groups are unfavorable in this region. Figure 4F shows a purple region near the oxygen atoms of the carbonyl groups (O12 and O11), suggesting that hydrogen bond acceptors groups in this region are favorable, whereas in regions near the two furan rings there are magenta contours, indicating that substitution for hydrogen bond acceptors is unfavorable to biological activity.

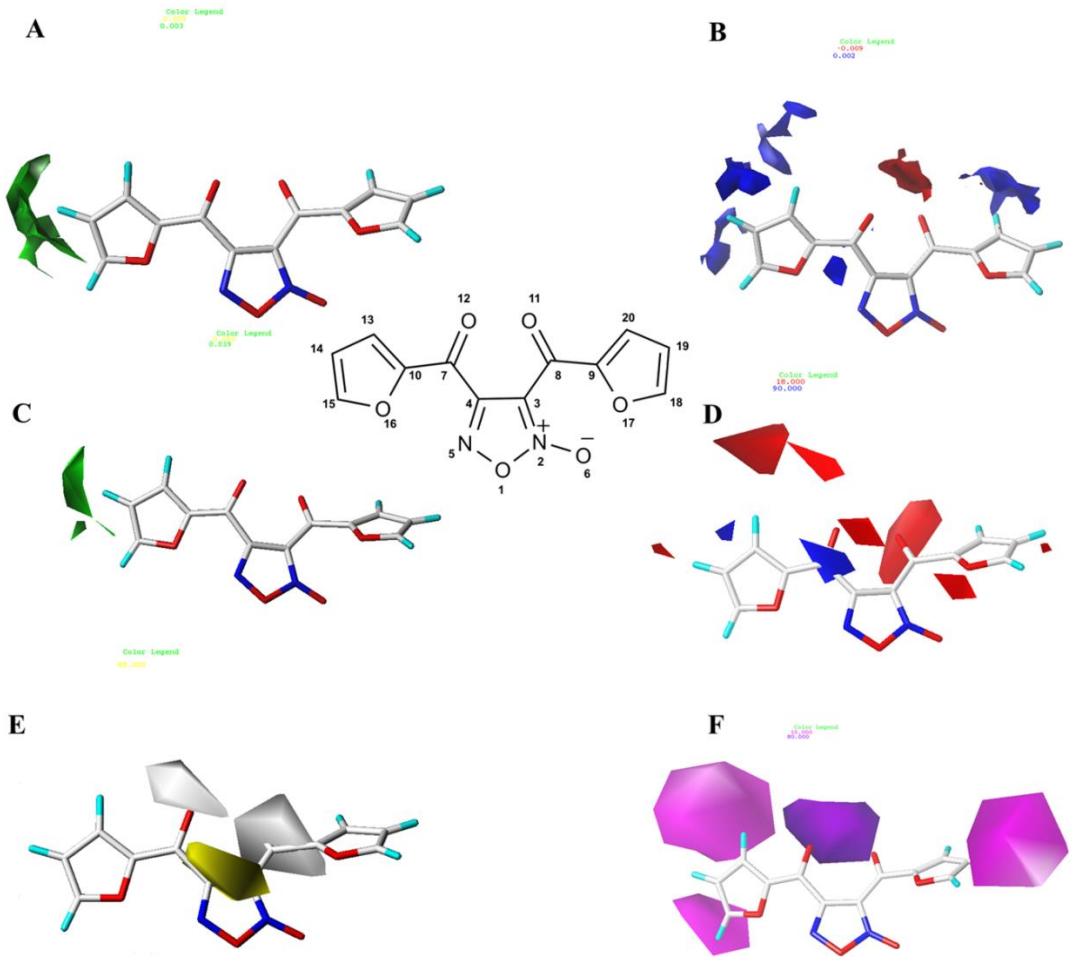


Figure 4. CoMFA and CoMSIA countour maps. CoMFA steric and electrostatic (A-B) and CoMSIA steric, electrostatic, hydrophobic and H-bond acceptor (C-F) contour maps surrounding the most potent inhibitor (33). Steric fields: green contours indicate region where bulky groups are favorable for biological activity; electrostatic fields: red contours indicate regions where electronegative groups are favorable to biological activity and blue contours indicate regions where electronegative groups are unfavorable; Hydrophobic fields: yellow contours indicate regions where hydrophobic groups are favorable and gray contours indicate regions where these groups are unfavorable; H-bond acceptor fields: purple contours indicate regions where

hydrogen bond acceptors are favorable for biological activity and magenta contours indicate regions where hydrogen bond acceptors are unfavorable.

Four different consensus models were obtained using different combinations of the best 2D- and 3D-QSAR models (Table 3). The external validation of consensus models was performed using the same test set and metrics used in individual models. However, as observed in Table 3, the metrics Q^2_{ext}, $r_m^2{}_{(average)}$ and $r_m^2{}_{(delta)}$ presented similar results between models. Thus, three additional metrics were calculated for consensus modeling: root mean square error of calibration (RMSEC), root mean square error of prediction (RMSEP) and Durbin-Watson Test. The RMSEC is a measure of goodness of fit, while RMSEP is used to measure the difference between experimental results and predictions obtained by the model. These metrics are described elsewhere [56,57]. The Durbin-Watson Test (*d*) is used to evaluate the autocorrelation of residues [58–60]. A value of d equal or near 2 indicates no autocorrelation between residues. Values of d < 2 indicate that residues have a negative autocorrelation, i.e. residues have very different values. Values of d > 2 indicate a positive autocorrelation between residues. Model 4 (Table 3) was selected as the best consensus model because it had good performance and unites the characteristics of the three QSAR approaches explored.

Table 3. Statistical characteristics for the consensus QSAR models.

Model	Models	*RMSEC*	*RMSEP*	Q^2_{ext}	$r_m^2{}_{(average)}$	$r_m^2{}_{(delta)}$	*d*
1	HQSAR + CoMFA	0.16	0.16	0.96	0.92	0.01	1.44
2	HQSAR + CoMSIA	0.16	0.12	0.98	0.93	0.03	1.41
3	CoMFA + CoMSIA	0.04	0.18	0.95	0.91	0.05	2.02
4	**HQSAR + CoMFA + CoMSIA**	**0.11**	**0.14**	**0.97**	**0.93**	**0.04**	**1.41**

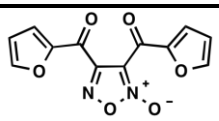
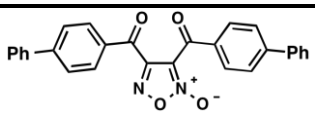
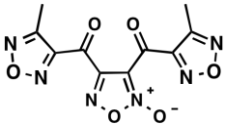
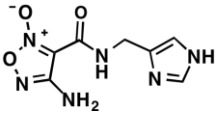
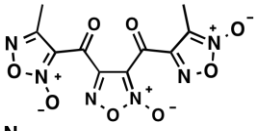
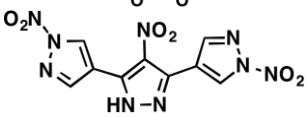
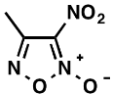
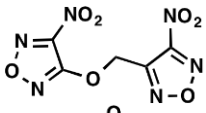
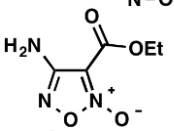
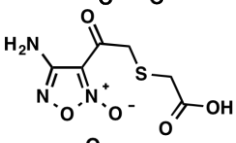
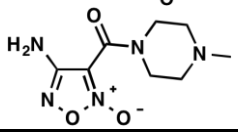
The best consensus model is highlighted in bold font. *RMSEC,* root mean square error of calibration; *RMSEP,* root mean square error of prediction; Q^2_{ext}, correlation coefficient of prediction; $r_m^2{}_{(average)}$ and $r_m^2{}_{(delta)}$, variants of the modified r^2; *d*, Durbin-Watson Test.

The best consensus model was selected to perform a virtual screening of new potential *Sm*TGR inhibitors. Firstly, a similarity search on the ChemBridge database [61] identified 80 compounds with at least $T_C \geq 0.6$ similarity with compound **33**. Additionally, 377 compounds containing the oxadiazole ring (common substructure) were identified. Duplicates and compounds already used to derive the QSAR models were excluded. Then, model **4** was used to predict the *Sm*TGR inhibitory activity of remaining

194 compounds. The consensus prediction of the biological activity was calculated using the arithmetic mean of the predictions from individual CoMFA and CoMSIA models (Table S5, Supplementary Material) shows the pIC_{50} prediction of each individual model as well as the consensus model.

Poor pharmacokinetic properties are important causes of costly late-stage failures in drug development [75]. Our laboratory has been working to overcome or reduce these failures using *in silico* tools for early prediction and optimization of ADME properties, such as Caco-2 cell permeability, blood-brain barrier penetration (BBBP), hERG inhibition, CYP3A4 inhibition and water solubility. Five in-house highly-predictive models were developed using large datasets of diverse compounds to cover the chemical space for the prediction of new compounds and are described elsewhere [26,27]. Table 4 shows the structure, consensus predicted potency against *Sm*TGR (IC_{50} in μM), and some predicted ADME properties of the ten new potential *Sm*TGR inhibitors.

Table 4. Chemical structures, predicted potency against *Sm*TGR and some calculated ADME properties of the lead compound (**33**) and the virtual hits.

Cpd	Structure	$IC_{50\ pred}$ [a] (μM)	clogP [b]	Caco2	BBBP	Water Solubility (logS)	CYP3A4 Inhibition	hERG inhibition
33*		0.04	0.96	Moderate-poor	BBB+	-1.91	Non-Inhibitor	Non-blocker
LabMol-13		0.23	5.10	Moderate-poor	BBB+	-2.20	Non-Inhibitor	Non-blocker
LabMol-14		0.03	-0.85	Moderate-poor	BBB+	-1.86	Non-Inhibitor	Non-blocker
LabMol-15		1.23	-1.41	Moderate-poor	BBB+	-2.47	Non-Inhibitor	Weak blocker
LabMol-16		0.06	-2.37	Moderate-poor	BBB+	-2.07	Non-Inhibitor	Non-blocker
LabMol-17		2.40	-0.20	Moderate-poor	BBB+	-1.75	Non-Inhibitor	Non-blocker
LabMol-18		1.54	-0.58	Moderate-poor	BBB+	-2.94	Non-Inhibitor	Non-blocker
LabMol-19		0.51	-0.37	Moderate-poor	BBB+	-2.37	Non-Inhibitor	Non-blocker
LabMol-20		1.00	-0.93	Moderate-poor	BBB+	-1.98	Non-Inhibitor	Non-blocker
LabMol-21		1.12	-1.11	Moderate-poor	BBB+	-2.17	Non-Inhibitor	Non-blocker
LabMol-22		0.89	-1.72	Moderate-poor	BBB+	-2.47	Non-Inhibitor	Weak blocker

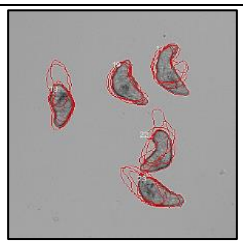
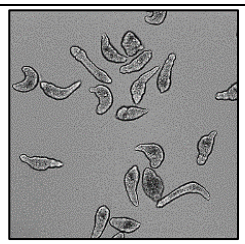
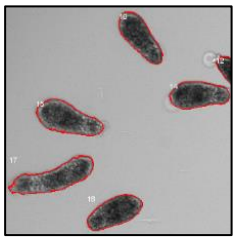
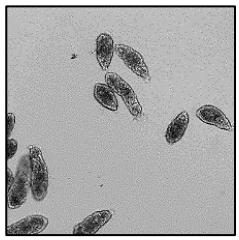
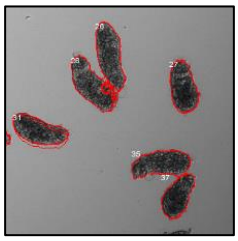
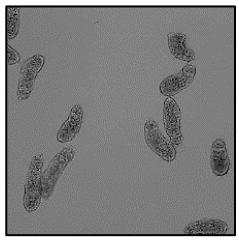
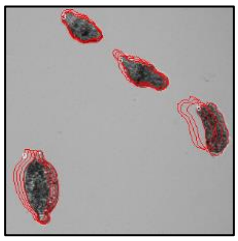
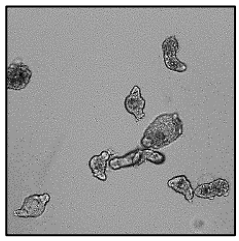
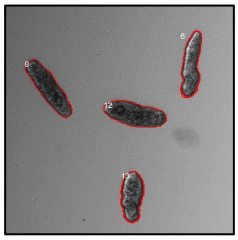
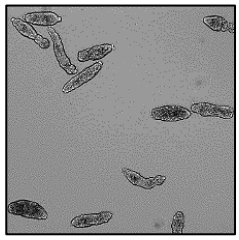
[a] IC_{50} prediction by the best consensus model. [b] Calculated octanol/water partition coefficient using RDKit 2.4.0.

From Table 4 we can see that ten compounds selected as virtual hits presented high predicted potency against *Sm*TGR, using the consensus QSAR model. Moreover, the selected hits were predicted to present favorable ADME properties and did not show any potential of being hERG blocker or CYP3A4 inhibitors. These compounds were selected for subsequent *in vitro* biological evaluation against

schistosomula and adult *S. mansoni* worms. From them, two compounds (***LabMol-17*** **and** ***LabMol-19*****)** were confirmed experimentally as new hits against *Schistosoma mansoni*.

Motility and phenotypic scores of the two hit compounds at 10 μM along with the negative (DMSO 0.625%) and positive controls (PZQ or OLT 10 μM) are shown in Table 5. The phenotypic and mobility scores calculated by the Bayesian model for both hits indicate their effects on schistosomula are equivalent or even more pronounced than produced by the reference drugs PZQ and OLT. EC_{50} values could be calculated from dose-response curves using either motility score (LabMol-19 EC_{50} 1.00 ± 0.11 μM) or phenotype score as the response (LabMol-17 EC_{50}: 4.76 ± 1.15 μM). The motility EC_{50} value determined for LabMol-19 was comparable to PZQ (EC_{50}: 2.2 μM) [76].

Table 5 – Motility and phenotype adjusted index values for *S. mansoni* schistosomula exposed for 48h to hit compounds or standard drugs at 10 µM.

Compound	Motility adjusted index (mean ± SD)	Phenotype adjusted index (mean ± SD)	Motility image	Phenotype image
Control (DMSO 0.625%)				
LabMol-17	-0.95 ± 0.01	-0.43 ± 0.11		
LabMol-19	-0.95 ± 0.01	-0.63 ± 0.02		
PZQ	-0.48 ± 0.04	-0.17 ± 0.02		
OLT	-0.90 ± 0.04	-0.34 ± 0.07		

The outlines showed in motility images represent the position of each parasite over 5 timepoints (11s interval).

Furthermore, the Bayesian model was able to classify the phenotype induced by both hit compounds as OLT-like indicating that the phenotype induced by these compounds was closest to OLT in comparison to the other modelled schistosomacides (PZQ, dihydroartemisinin, methylclonazepam, Ro15-5458, oxamniquine). This result is consistent with the fact that both LabMol-17 and LabMol-19 are potential *Sm*TGR inhibitors, an enzyme involved in ROS detoxification in the *S. mansoni*, similar to the OLT mechanism of action which is also thought to interfere with the parasite's redox defense system [77]. Compounds active on schistosomula were then tested on *S. mansoni* adult worms *in vitro*. Chemicals were assayed by a new methodology that utilizes HCS technology to automatically score changes on parasite motility. Besides avoiding human bias, this quantitative method is more robust and sensitive to subtle changes in parasite movement than the standard assay using manual microscopic visualization [78]. Figures 5 and 6 shows percent motilities of male and female worms, respectively, measured after exposure to drugs at 10 µM for up to 72h incubation time.

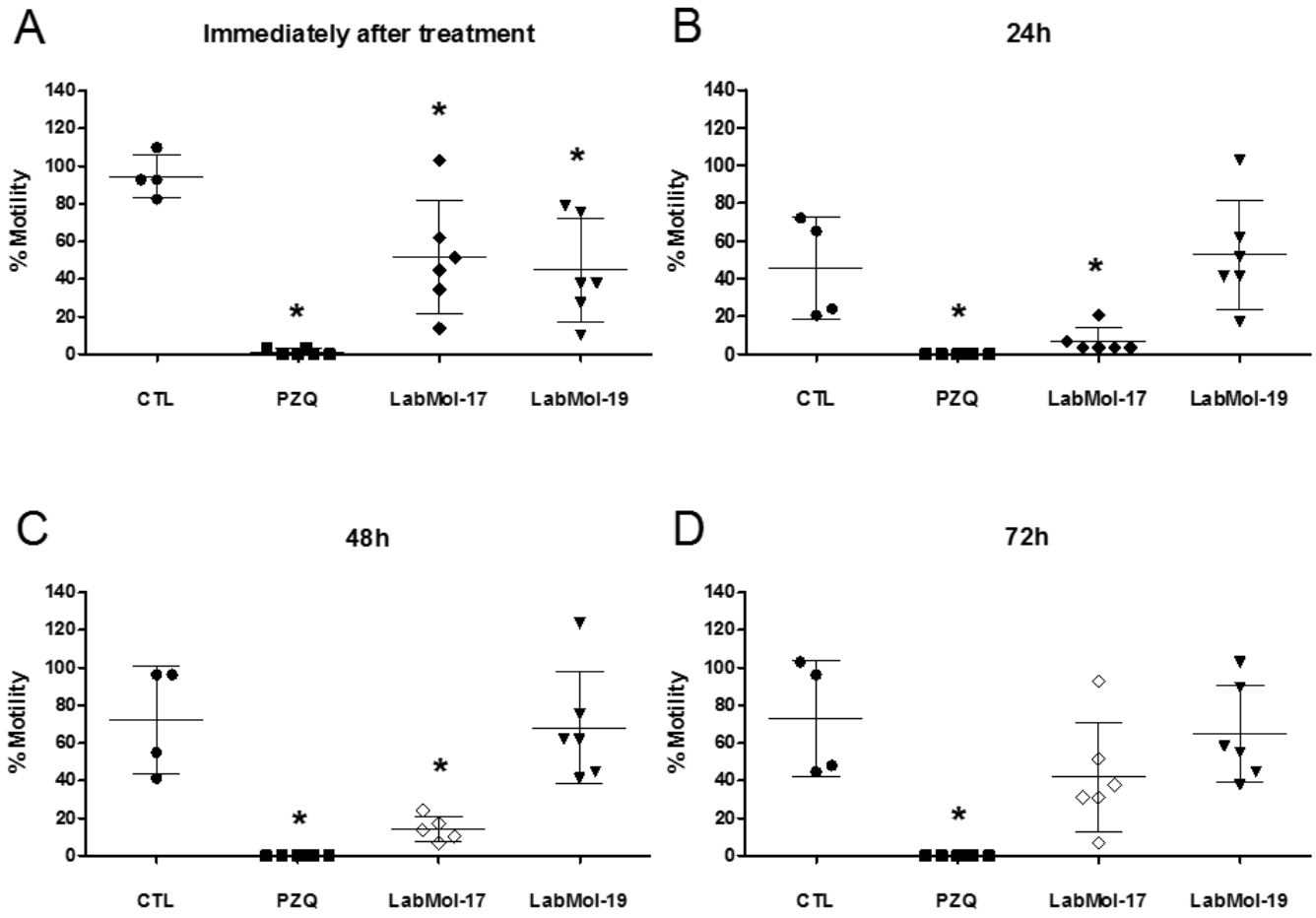


Figure 5. The effect of compounds on male adult worm motility as analyzed by HCS analysis for 0 - 72hrs. PZQ, LabMol-17 and LabMol-19 were screened at 10 µM and DMSO at an equivalent % concentration. The percentage motility values were determined immediately (A), 24hrs (B), 48hrs (C) or at 72 hrs (D) by comparison to the average motility of the worms before compound addition. **CTL** - DMSO 0.02%, **PZQ** - Praziquantel. Statistical significance (* = $p<0.05$) was calculated by One-way ANOVA analysis followed by Tuckey´s post-hoc test.

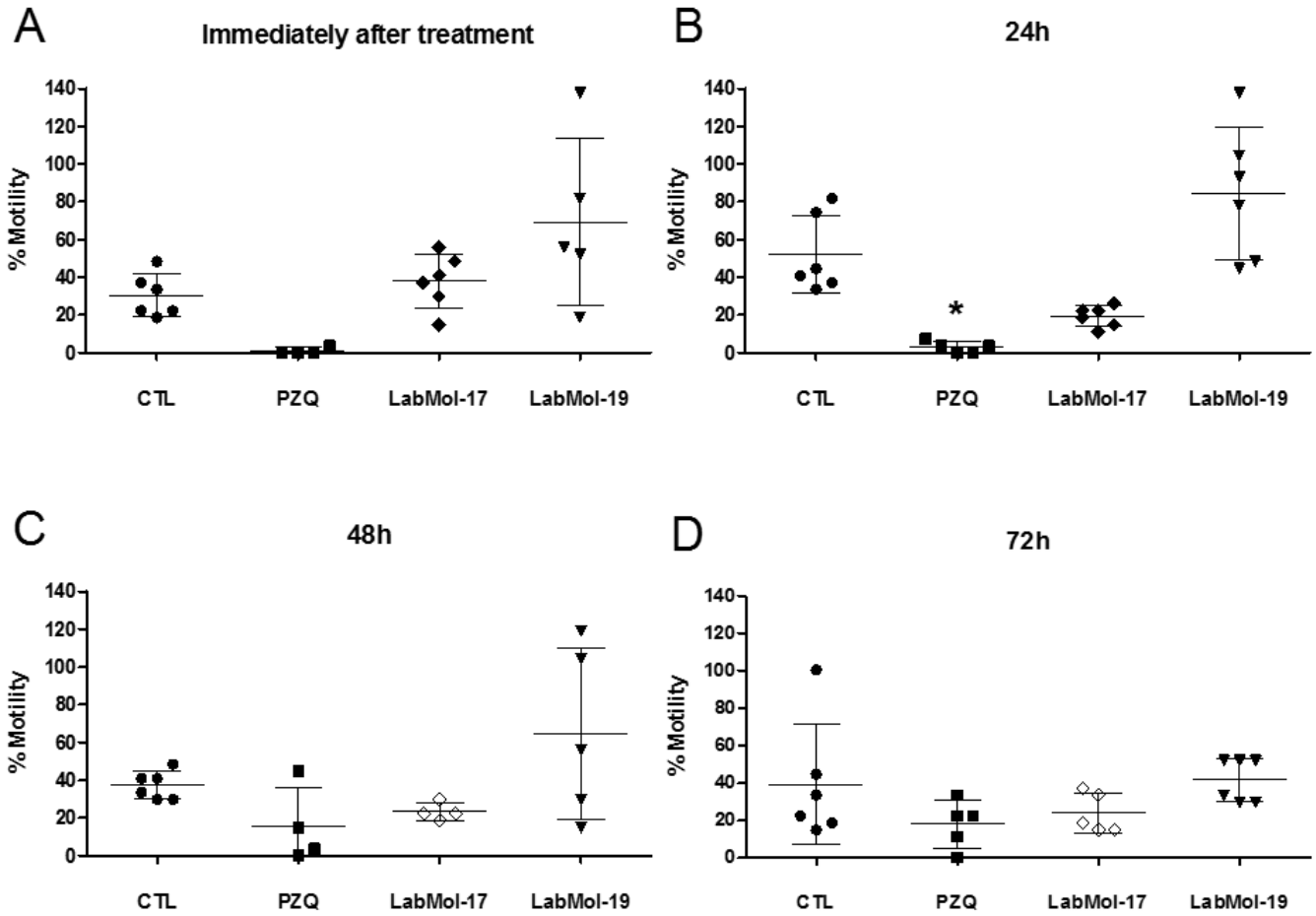


Figure 6. The effect of compound on female adult worm motility as analyzed by HCS analysis for 0 - 72hrs. PZQ, LabMol-17 and LabMol-19 were screened at 10 µM and DMSO at an equivalent % concentration. The percentage motility values were determined immediately (A), 24hrs (B), 48hrs (C) or at 72 hrs (D) by comparison to the average motility of the worms before compound addition. CTL - DMSO 0.02%, PZQ - Praziquantel. Statistical significance (* = $p<0.05$) was calculated by One-way ANOVA analysis followed by Tuckey´s post-hoc test.

Like PZQ, both LabMol-17 and LabMol-19 were more active on male worms. LabMol-17 was the most active hit, reducing male worm motility up to five times compared to untreated control worms. Significant reduction ($p<0.05$) of male worm motility was observed immediately after addition of the drugs to the microplate well and in the case of LabMol-17 peaked after 48h incubation. Effects on female worms were less pronounced. A discernible, although not statistically

significant, trend could still be observed for LabMol-17 after 24h and 48h of drug exposure. Interestingly, although not statistically significant for the whole treatment group, LabMol-19 induced augmented motility in at least half of the female worms in the group on exposure times up to 48h. Further experiments will elucidate if this different behavior on drug sensitivity between male and female worms may be due to different expression patterns of SmTGR or other cause.

It is interesting to note that the two experimentally validated hits, **LabMol-17** and **LabMol-19,** are dissimilar from the most potent compound (**33**) in the training set (T_c of 0.60 and 0.63, respectively (Table S5, Supplementary Material)). More interestingly, the most active hit, **LabMol-17,** is very dissimilar from the current schistosomicidal drug, PZQ (T_c = 0.07) and oltipraz (T_c = 0.08), and also from other known drugs (Figure 7).

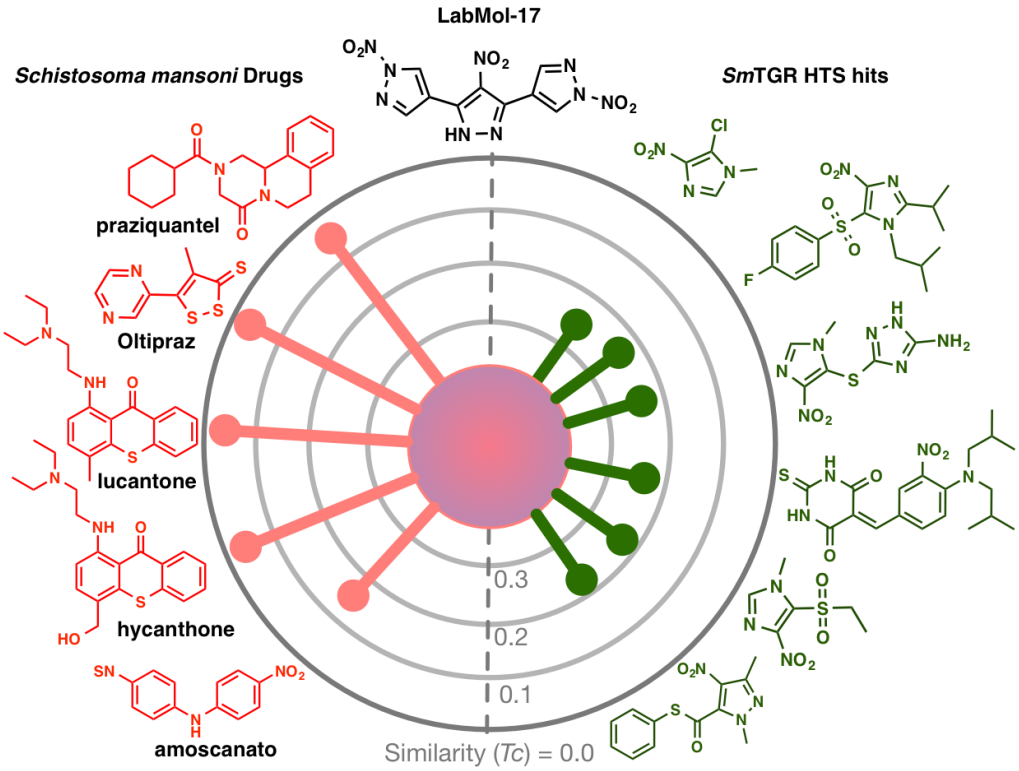


Figure 7. Radial plots for similarity of LabMol-17 with known schistosomicidal drugs and known SmTGR hits. The central node represents the target compound surrounded by known schistosomicidal drugs (left-hand side) and known *Sm*TGR hits from a HTS screen retrieved from PubChem BioAssay (AID: 485364) (right-hand side). The similarity was calculated using Tanimoto coefficient (T_c) using Morgan (ECFP4-like) fingerprints [79].

Moreover, we have also analyzed the similarity between **LabMol-17** and known *Sm*TGR hits discovered in a HTS screening (PubChem BioAssay, AID: 485364). As we can see from Figure 7, the Tanimoto coefficient values are in the range between 0 and 0.3, showing that the compounds are very dissimilar, and therefore, **LabMol-17** has proved to represent a new chemical scaffold against *S. mansoni*.

Because of this structural difference and high activity against schistosomula and adult worms, the two hit compounds (**LabMol-17** and **LabMol-19)** are promising as new potential *Sm*TGR inhibitors representing new not yet explored chemical scaffolds.

Conclusion

We succeed to develop robust and externally predictive consensus model including 2D (HQSAR) and 3D QSAR models (CoMFA and CoMSIA). Developed consensus model was used for virtual screening of ChemBridge database. We identified ten new potential *Sm*TGR inhibitors. High activity of two of these hits (LabMol-17 and LabMol-19) containing new scaffolds and pretty structurally dissimilar from known antishistosomal agents was validated experimentally against *S. mansoni* on both schistosomula and adult worms *in vitro*. We are going to conduct additional experimental validation and structural modification, if needed, of these compounds to develop new schistosomicidal agents on their base.

Future perspective

Despite urgent need in finding a new treatment for schistosomiasis and some success that was achieved recently, there is still a long way to go to defeat this NTD. There are multiple potential ways of battling schistosomiasis. The first one is the development a vaccine. Nowadays, there are ongoing clinical trials for several vaccines against various *Schistosoma* strains [80]. Thus, a vaccine against *Schistosoma haematobium* is under Phase III of clinical trials in Senegal and Niger and two vaccines against

Schistosoma mansoni are under Phase I of clinical trials in US and Brazil [80]. However, despite this, there are numerous challenges in vaccine development related to the antigen discovery, product development, and the preclinical and clinical testing stages. For instance, reverse vaccinology approaches that work well for small bacterial genomes have not yet been successful for eukaryotic parasites because of the greater complexity of the parasite genome and the requirement to employ eukaryotic expression systems to reliably produce soluble and properly folded antigens. However, scientific challenges are not the major obstacle for the development of new vaccines; socioeconomic reasons are the real bottleneck there. For instance, there are no clear ways to make a vaccine, even potent and effective, available to those who need it most – poor people in South America, Africa, and Asia.

Another way of battling schistosomiasis is development of new small molecule drug. Obviously, this drug should be active against all the different strains of *Schistosoma*. Thus, this drug has to be promiscuous enough to cover a variety of strains and selective enough not to hit other targets in the human organism, because the interaction with some of them could lead to strong undesirable side effects. This is still very possible way, although, like in case of vaccines, similar socioeconomic challenges could prevent the development and especially targeted distribution of such drug. In other words, even if developed, such a drug could cost that much that the price could be affordable only by a small richer part of population and still will be unavailable to poorer part, which suffer from schistosomiasis the most. Another possibility is the combination therapy, which could help to overcome solving scientific bottlenecks such as drug resistance phenomenon, toxicity, etc. In the same way, combinatorial therapy could cause undesired drug-drug interactions. All social and economic problems of monotherapy will also remain to combined treatment. One of the solutions to decrease the economical burden is the use of computational techniques for targeted design of new anti-schistosomiasis agents that will decrease both the time and the money spent to develop new drug. Expected future development of combined therapy will promote development of reliable computational tools for handling and analyzing mixtures [81]. In addition, these approaches should take into account both synergistic and antagonistic effects [82]. We strongly believe that in future

battling neglected tropical diseases will include some other approaches unknown and even unpredictable now and that these approaches will help to overcome not only scientific but also socioeconomic challenges.

Concluding this section we would like to point out that our new adult worm HCS platform was instrumental for the reliable quantification of drug-induced alterations in parasite motility, however some work will be done in our group to further extend this technology to score adult worm phenotypes as it is already possible for schistosomula. We are also going to additionally validate the activity of hit compounds identified in this study by different experiments. Structural modification could also be done if needed to improve ADMET properties of designed compounds.

Executive summary

- 2D- and 3D- QSAR methods were used to build highly predictive consensus models.
- A QSAR-based virtual screening identified ten new potential inhibitors of *Sm*TGR.
- A new high-content screening platform for quantification of worm motility was validated.
- *In vitro* activity against schistosomules and adult worms for two compounds was confirmed experimentally.
- LabMol-17 and LabMol 19 will be subjects for additional experimental testing and structural modification to design new schistosomicidal drugs.

Acknowledgments

The authors would like to thank Brazilian funding agencies, CNPq, CAPES, FAPERJ and FAPEG for financial support and fellowships. AEC and LK were supported by the National Institutes of Health (R01 GM095672 to AEC). EM thanks the financial support from NIH (GM 096967 and GM66940) and UNC for Junior Faculty Development Award. We are grateful for Molecular Devices for providing the HCS equipment for this study as well as OpenEye Scientific Software Inc. and ChemAxon for providing us with academic licenses for their software. Authors also thank the Bioassays and Drug Screening Platform (FIOCRUZ RPT11-I subunit) for technological support and the Malacology Laboratory (Dr. Silvana C. Thiengo) from IOC/FIOCRUZ for providing *S. mansoni* cercarie.

Financial & competing interests disclosure

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript.

References

1. Colley DG, Bustinduy AL, Secor WE, King CH. Human schistosomiasis. *Lancet* [Internet]. 383(9936), 2253–64 (2014). Available from: http://www.ncbi.nlm.nih.gov/pubmed/24698483.

2. World Health Organization. Schistosomiasis. Fact sheet N°115. (2015).

3. Gryseels B. Schistosomiasis. *Infect. Dis. Clin. North Am.* [Internet]. 26(2), 383–97 (2012). Available from: http://www.ncbi.nlm.nih.gov/pubmed/22632645.

4. Gryseels B, Polman K, Clerinx J, Kestens L. Human schistosomiasis. *Lancet* [Internet]. 368(9541), 1106–18 (2006). Available from: http://www.ncbi.nlm.nih.gov/pubmed/16997665.

5. World Health Organization. Progress report 2001-2011 and strategic plan 2012-2020. Geneva, Switzerland.

6. Hagan P, Appleton CC, Coles GC, Kusel JR, Tchuem-Tchuenté L-A. Schistosomiasis control: keep taking the tablets. *Trends Parasitol.* 20(2), 92–97 (2004).

7. Loukas A, Bethony JM. New drugs for an ancient parasite. *Nat. Med.* [Internet]. 14(4), 365–7 (2008). Available from: http://www.ncbi.nlm.nih.gov/pubmed/18391931.

8. Wang W, Wang L, Liang Y-S. Susceptibility or resistance of praziquantel in human schistosomiasis: a review. *Parasitol. Res.* [Internet]. 111(5), 1871–7 (2012). Available from: http://www.ncbi.nlm.nih.gov/pubmed/23052781.

9. Caffrey CR. Chemotherapy of schistosomiasis: present and future. *Curr. Opin. Chem. Biol.* 11(4), 433–9 (2007).

10. Neves BJ, Andrade CH, Cravo PVL. Natural Products as Leads in Schistosome Drug Discovery. *Molecules*. 20(2), 1872–1903 (2015).

11. Berriman M, Haas BJ, LoVerde PT, *et al.* The genome of the blood fluke Schistosoma mansoni. *Nature* [Internet]. 460(7253), 352–8 (2009). Available from:

http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2756445&tool=pmcentrez&rendertype=abstract.

12. Protasio A V, Tsai IJ, Babbage A, *et al.* A systematically improved high quality genome and transcriptome of the human blood fluke Schistosoma mansoni. *PLoS Negl. Trop. Dis.* [Internet]. 6(1), e1455 (2012). Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3254664&tool=pmcentrez&rendertype=abstract.

13. Kuntz AN, Davioud-Charvet E, Sayed AA, *et al.* Thioredoxin glutathione reductase from Schistosoma mansoni: an essential parasite enzyme and a key drug target. *PLoS Med.* [Internet]. 4(6), e206 (2007). Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1892040&tool=pmcentrez&rendertype=abstract.

14. Angelucci F, Miele AE, Boumis G, Dimastrogiovanni D, Brunori M, Bellelli A. Glutathione reductase and thioredoxin reductase at the crossroad: the structure of Schistosoma mansoni thioredoxin glutathione reductase. *Proteins* [Internet]. 72(3), 936–45 (2008). Available from: http://www.ncbi.nlm.nih.gov/pubmed/18300227.

15. Song L, Li J, Xie S, *et al.* Thioredoxin Glutathione Reductase as a Novel Drug Target: Evidence from Schistosoma japonicum. *PLoS One* [Internet]. 7(2), e31456 (2012). Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3285170&tool=pmcentrez&rendertype=abstract.

16. Han Y, Fu Z, Hong Y, *et al.* Inhibitory Effects and Analysis of RNA Interference on Thioredoxin Glutathione Reductase Expression in *Schistosoma japonicum*. *J. Parasitol.* 100(4), 463–469 (2014).

17. Iskar M, Zeller G, Zhao X-M, van Noort V, Bork P. Drug discovery in the age of systems biology: the rise of computational approaches for data integration. *Curr. Opin. Biotechnol.* [Internet]. 23(4), 609–16 (2012). Available from: http://www.ncbi.nlm.nih.gov/pubmed/22153034.

18. Bajorath J. Chemoinformatics: recent advances at the interfaces between computer and chemical information sciences, chemistry, and drug discovery. *Bioorg. Med. Chem.* [Internet]. 20(18), 5316 (2012). Available from: http://www.ncbi.nlm.nih.gov/pubmed/22980097.

19. Duffy BC, Zhu L, Decornez H, Kitchen DB. Early phase drug discovery: cheminformatics and computational techniques in identifying lead series. *Bioorg. Med. Chem.* [Internet]. 20(18), 5324–5342 (2012). Available from: http://www.ncbi.nlm.nih.gov/pubmed/22938785.

20. Jorgensen WL. The many roles of computation in drug discovery. *Science* [Internet]. 303(5665), 1813–1818 (2004). Available from: http://www.ncbi.nlm.nih.gov/pubmed/15031495.

21. Jorgensen WL. Challenges for academic drug discovery. *Angew. Chem. Int. Ed. Engl.* [Internet]. 51(47), 11680–4 (2012). Available from: http://www.ncbi.nlm.nih.gov/pubmed/23097176.

22. Andrade CH, Pasqualoto KFM, Ferreira EI, Hopfinger AJ. 3D-Pharmacophore mapping of thymidine-based inhibitors of TMPK as potential antituberculosis agents. *J. Comput. Aided. Mol. Des.* [Internet]. 24(2), 157–72 (2010). Available from: http://www.ncbi.nlm.nih.gov/pubmed/20217185.

23. Melo-Filho CC, Braga RC, Andrade CH. Advances in Methods for Predicting Phase I Metabolism of Polyphenols. *Curr. Drug Metab.* [Internet]. 15(1), 120–126 (2014). Available from: http://www.ncbi.nlm.nih.gov/pubmed/24479689.

24. Melo-Filho CC, Braga RC, Andrade CH. 3D-QSAR Approaches in Drug Design: Perspectives to Generate Reliable CoMFA Models. *Curr. Comput. Aided. Drug Des.* [Internet]. 10(2), 148–59 (2014). Available from: http://www.ncbi.nlm.nih.gov/pubmed/24724896.

25. Bueno R V, Braga RC, Segretti ND, Ferreira EI, Trossini GHG, Andrade CH. New tuberculostatic agents targeting nucleic acid biosynthesis: drug design using QSAR approaches. *Curr. Pharm. Des.* [Internet]. 20(27), 4474–85 (2014). Available from: http://www.ncbi.nlm.nih.gov/pubmed/24245758.

26. Braga RC, Alves VM, Silva AC, Liao LM, Andrade CH. Virtual Screening Strategies in Medicinal Chemistry: The state of the art and current challenges. *Curr. Top. Med. Chem.* 14(16), 1899–1912 (2014).

27. Braga RC, Alves VM, Silva MFB, *et al.* Tuning hERG out: Antitarget QSAR Models for Drug Development. *Curr. Top. Med. Chem.* [Internet]. 14(11), 1399–1415 (2014). Available from: http://www.ncbi.nlm.nih.gov/pubmed/24805060.

28. Andrade CH, Silva DC, Braga RC. In silico Prediction of Drug Metabolism by P450. *Curr. Drug Metab.* [Internet]. 15(5), 514–525 (2014). Available from: http://www.ncbi.nlm.nih.gov/pubmed/25204822.

29. Neves BJ, Bueno R V, Braga RC, Andrade CH. Discovery of new potential hits of Plasmodium falciparum enoyl-ACP reductase through ligand- and structure-based drug design approaches. *Bioorg. Med. Chem. Lett.* [Internet]. 23(8), 2436–41 (2013). Available from: http://linkinghub.elsevier.com/retrieve/pii/S0960894X13001807.

30. Bueno R V, Toledo NR, Neves BJ, Braga RC, Andrade CH. Structural and chemical basis for enhanced affinity to a series of mycobacterial thymidine monophosphate kinase inhibitors: fragment-based QSAR and QM/MM docking studies. *J. Mol. Model.* 19(1), 179–192 (2013).

31. Braga RC, Andrade CH. Assessing the performance of 3D pharmacophore models in virtual screening: how good are they? *Curr. Top. Med. Chem.* 13(9), 1127–1138 (2013).

32. Braga RC, Alves VM, Fraga CAM, Barreiro EJ, de Oliveira V, Andrade CH. Combination of docking, molecular dynamics and quantum mechanical calculations for metabolism prediction of 3, 4-methylenedioxybenzoyl-2-thienylhydrazone. *J. Mol. Model.* 18(5), 2065–2078 (2012).

33. Braga RC, Andrade CH. QSAR and QM/MM approaches applied to drug metabolism prediction. *Mini Rev. Med. Chem.* [Internet]. 12(6), 573–82 (2012). Available from: http://www.ncbi.nlm.nih.gov/pubmed/22587770.

34. Carneiro EO, Andrade CH, Braga RC, *et al.* Structure-based prediction and biosynthesis of the major mammalian metabolite of the cardioactive prototype LASSBio-294. *Bioorg. Med. Chem. Lett.* [Internet]. 20(12), 3734–6 (2010). Available from: http://www.ncbi.nlm.nih.gov/pubmed/20488703.

35. Neves BJ, Braga RC, Bezerra JCB, Cravo PVL, Andrade CH. In Silico Repositioning Chemogenomics strategy identifies New Drugs with potential activity Against Multiple Life Stages of Schistosoma Mansoni. *PLoS Negl. Trop. Dis.* just accep (2014).

36. Zanella F, Lorens JB, Link W. High content screening: seeing is believing. *Trends Biotechnol.* 28(5), 237–245 (2010).

37. Paveley RA, Bickle QD. Automated Imaging and other developments in whole-organism anthelmintic screening. *Parasite Immunol.* [Internet]. 35(9-10), 302–313 (2013). Available from: http://doi.wiley.com/10.1111/pim.12037.

38. Rai G, Sayed AA, Lea WA, *et al.* Structure mechanism insights and the role of nitric oxide donation guide the development of oxadiazole-2-oxides as therapeutic agents against schistosomiasis. *J. Med. Chem.* [Internet]. 52(20), 6474–83 (2009). Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2772170&tool=pmcentrez&rendertype=abstract.

39. Sayed AA, Simeonov A, Thomas CJ, Inglese J, Austin CP, Williams DL. Identification of oxadiazoles as new drug leads for the control of schistosomiasis. *Nat. Med.* [Internet]. 14(4), 407–12 (2008). Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2700043&tool=pmcentrez&rendertype=abstract.

40. OMEGA 2.5.1.4: OpenEye Scientific Software, Santa Fe, NM. http://www.eyesopen.com. .

41. Hawkins PCD, Skillman AG, Warren GL, Ellingson BA, Stahl MT. Conformer generation with OMEGA: algorithm and validation using high quality structures from the Protein Databank and Cambridge Structural Database. *J. Chem. Inf. Model.* [Internet]. 50(4), 572–84 (2010). Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2859685&tool=pmcentrez&rendertyp

e=abstract.

42. SYBYL-X 1.2, Tripos International, 1699 South Hanley Rd., St. Louis, Missouri, 63144, USA. .

43. Seel M, Turner DB, Willett P. Effect of Parameter Variations on the Effectiveness of HQSAR Analyses. *Mol. Inform.* [Internet]. 18(3), 245–252 (1999). Available from: http://dx.doi.org/10.1002/(SICI)1521-3838(199907)18:3<245::AID-QSAR245>3.0.CO.

44. Lindberg W, Persson J-A, Wold S. Partial least-squares method for spectrofluorimetric analysis of mixtures of humic acid and lignin sulfonate. *Anal. Chem.* 55(4), 643–648 (1983).

45. Wold S, Sjöström M, Eriksson L. PLS-regression: a basic tool of chemometrics. *Chemom. Intell. Lab. Syst.* 58(2), 109–130 (2001).

46. Gasteiger J, Marsili M. A new model for calculating atomic charges in molecules. *Tetrahedron Lett.* 19(34), 3181–3184 (1978).

47. Purcell WP, Singer JA. A Brief Review and Table of Semiempirical Parameters Used in the Hückel Molecular Orbital Method. *J. Chem. Eng. Data*. 12(2), 235–246 (1967).

48. Jakalian A, Bush BL, Jack DB, Bayly CI. Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: I. Method. *J. Comput. Chem.* 21(2), 132–146 (1999).

49. Jakalian A, Jack DB, Bayly CI. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J. Comput. Chem.* [Internet]. 23(16), 1623–1641 (2002). Available from: http://www.ncbi.nlm.nih.gov/pubmed/12395429.

50. QUACPAC 1.6.3.1: OpenEye Scientific Software, Santa Fe, NM. http://www.eyesopen.com. .

51. ROCS 3.2.0.4: OpenEye Scientific Software, Santa Fe, NM. http://www.eyesopen.com. .

52. Hawkins PCD, Skillman AG, Nicholls A. Comparison of shape-matching and docking as virtual screening tools. *J. Med. Chem.* [Internet]. 50(1), 74–82 (2007). Available from: http://www.ncbi.nlm.nih.gov/pubmed/17201411.

53. Consonni V, Ballabio D, Todeschini R. Comments on the definition of the Q2 parameter for QSAR validation. *J. Chem. Inf. Model.* [Internet]. 49(7), 1669–78 (2009). Available from: http://www.ncbi.nlm.nih.gov/pubmed/19527034.

54. Pratim Roy P, Paul S, Mitra I, Roy K. On two novel parameters for validation of predictive QSAR models. *Molecules*. 14(5), 1660–701 (2009).

55. Roy K, Chakraborty P, Mitra I, Ojha PK, Kar S, Das RN. Some case studies on application of "r(m)2" metrics for judging quality of quantitative structure-activity relationship predictions: emphasis on scaling of response data. *J. Comput. Chem.* 34(12), 1071–1082 (2013).

56. Aptula AO, Jeliazkova NG, Schultz TW, Cronin MTD. The better predictive model: High q2 for the training set or low root mean square error of prediction for the test set? *QSAR Comb. Sci.* 24(3), 385–396 (2005).

57. Consonni V, Ballabio D, Todeschini R. Evaluation of model predictive ability by external validation techniques. *J. Chemom.* [Internet]. 24(3-4), 194–201 (2010). Available from: http://doi.wiley.com/10.1002/cem.1290.

58. Durbin J, Watson GS. Testing for Serial Correlation in Least Squares Regression: I. *Biometrika*. 37(3/4), 409–428 (1950).

59. Durbin J, Watson GS. Testing for Serial Correlation in Least Squares Regression. II. *Biometrika*. 38(1/2), 159–177 (1951).

60. Savin N, White K. The Durbin-Watson test for serial correlation with extreme sample sizes or many regressors. *Econometrica*. 45(8), 1989–1996 (1977).

61. ChemBridge Online Chemical Store. (2015).

62. Mansour NR, Bickle QD. Comparison of Microscopy and Alamar Blue Reduction in a Larval Based Assay for Schistosome Drug Screening. *PLoS Negl. Trop. Dis.* 4(8), e795 (2010).

63. Marxer M, Ingram K, Keiser J. Development of an in vitro drug screening assay using Schistosoma haematobium schistosomula. *Parasit. Vectors* [Internet]. 5(1), 165 (2012). Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3437206&tool=pmcentrez&rendertype=abstract.

64. Paveley RA, Mansour NR, Hallyburton I, *et al.* Whole organism high-content screening by label-free, image-based bayesian classification for parasitic diseases. *PLoS Negl. Trop. Dis.* 6(7), 1–11 (2012).

65. Kamentsky L, Jones TR, Fraser A, *et al.* Improved structure, function and compatibility for cellprofiler: Modular high-throughput image analysis software. *Bioinformatics*. 27(8), 1179–1180 (2011).

66. Lowis DR. HQSAR: a new, highly predictive QSAR technique. *Tripos Tech. notes*. 1(5), 1–17 (1997).

67. Golbraikh A, Tropsha A. Beware of q2! *J. Mol. Graph. Model.* [Internet]. 20(4), 269–76 (2002). Available from: http://www.ncbi.nlm.nih.gov/pubmed/11858635.

68. Gasco A, Fruttero R, Sorba G, Di Stilo A, Calvino R. NO donors: Focus on furoxan derivatives. *Pure Appl. Chem.* 76(5), 973–981 (2004).

69. Kroemer RT, Hecht P, Guessregen S, Liedl KR. Improving the predictive quality of CoMFA models. In: *3D QSAR in Drug Design Recent Advances*. Kubinyi H, Folkers G, Martin YC (Eds.). . Kluwer Academic Publishers, Dordrecht, 41–56 (1998).

70. Tsai K-C, Chen Y-C, Hsiao N-W, *et al.* A comparison of different electrostatic potentials on prediction accuracy in CoMFA and CoMSIA studies. *Eur. J. Med. Chem.* [Internet]. 45(4), 1544–1551 (2010). Available from: http://www.ncbi.nlm.nih.gov/pubmed/20110138.

71. Doweyko AM. 3D-QSAR illusions. *J. Comput. Aided. Mol. Des.* [Internet]. 18(7-9), 587–596 (2004). Available from: http://www.springerlink.com/index/10.1007/s10822-004-4068-0.

72. Jain AN. Ligand-based structural hypotheses for virtual screening. *J. Med. Chem.* [Internet]. 47(4), 947–61 (2004). Available from: http://www.ncbi.nlm.nih.gov/pubmed/14761196.

73. Ojha PK, Mitra I, Das RN, Roy K. Further exploring rm2 metrics for validation of QSPR models. *Chemom. Intell. Lab. Syst.* [Internet]. 107(1), 194–205 (2011). Available from: http://linkinghub.elsevier.com/retrieve/pii/S016974391100061X.

74. Kubinyi H. QSAR and 3D QSAR in drug design Part 1: methodology. *Drug Discov. Today* [Internet]. 2(11), 457–467 (1997). Available from: http://linkinghub.elsevier.com/retrieve/pii/S1359644697010799.

75. van de Waterbeemd H, Gifford E. ADMET in silico modelling: towards prediction paradise? *Nat. Rev. Drug Discov.* [Internet]. 2(3), 192–204 (2003). Available from: http://www.ncbi.nlm.nih.gov/pubmed/12612645.

76. Ingram-Sieber K, Cowan N, Panic G, *et al.* Orally active antischistosomal early leads identified from the open access malaria box. *PLoS Negl. Trop. Dis.* [Internet]. 8(1), e2610 (2014). Available from: http://www.ncbi.nlm.nih.gov/pubmed/24416463.

77. Nare B, Smith JM, Prichard RK. Oltipraz-induced decrease in the activity of cytosolic glutathione S-Transferase in Schistosoma mansoni. *Int. J. Parasitol.* 21(8), 919–925 (1991).

78. Ramirez B, Bickle Q, Yousif F, Fakorede F, Mouries M-A, Nwaka S. Schistosomes: challenges in compound screening. *Expert Opin. Drug Discov.* 2, S53–S61 (2007).

79. Rogers D, Hahn M. Extended-connectivity fingerprints. *J. Chem. Inf. Model.* 50(5), 742–54

(2010).

80. Hotez PJ, Bottazzi ME, Strych U. New Vaccines for the World's Poorest People. *Annu. Rev. Med.* (New Vaccines for the World's Poorest People.) (2015).

81. Muratov EN, Varlamova E V., Artemenko AG, Polishchuk PG, Kuz'min VE. Existing and Developing Approaches for QSAR Analysis of Mixtures. *Mol. Inform.* [Internet]. 31(3-4), 202–221 (2012). Available from: http://doi.wiley.com/10.1002/minf.201100129.

82. Bulusu KC, Guha R, Mason DJ, *et al.* Modelling of compound combination effects and applications to efficacy and toxicity: state-of-the-art, challenges and perspectives. *Drug Discov. Today* [Internet]. (2015). Available from: http://www.ncbi.nlm.nih.gov/pubmed/26360051.