

# Discovery of new *Schistosoma mansoni* aspartyl protease inhibitors by structure-based virtual screening

Bárbara Figueira Gomes<sup>1</sup>, Mario Roberto Senger<sup>1</sup>, José Teófilo Moreira-Filho<sup>2</sup>, Fabio Jorge de Vasconcellos Junior<sup>1</sup>, Rafael Ferreira Dantas<sup>1</sup>, Raymond Owens<sup>3</sup>, Carolina Horta Andrade<sup>2</sup>, Bruno Junior Neves<sup>4</sup>, Floriano Paes Silva-Junior<sup>1/+</sup>

<sup>1</sup>Fundação Oswaldo Cruz-Fiocruz, Instituto Oswaldo Cruz, Laboratório de Bioquímica Experimental e Computacional de Fármacos, Rio de Janeiro, RJ, Brasil

<sup>2</sup>Universidade Federal de Goiás, Faculdade de Farmácia, Laboratório de Planejamento de Fármacos e Modelagem Molecular, Goiânia, GO, Brasil

<sup>3</sup>University of Oxford and Rosalind Franklin Institute, Oxfordshire, UK

<sup>4</sup>Universidade Federal de Goiás, Faculdade de Farmácia, Laboratório de Quimioinformática, Goiânia, GO, Brasil

**BACKGROUND** Schistosomiasis is a neglected tropical disease caused by trematodes of the genus *Schistosoma*, with a limited treatment, mainly based on the use of praziquantel (PZQ). Currently, several aspartic proteases genes have already been identified within the genome of *Schistosoma* species. At least one enzyme encoded from this gene family (*SmAP*), named *SmCD1*, has been validated for the development of schistosomicidal drugs, since it has a key role in haemoglobin digestion by worms.

**OBJECTIVE** In this work, we integrated a structure-based virtual screening campaign, enzymatic assays and adult worms *ex vivo* experiments aiming to discover the first classes of *SmCD1* inhibitors.

**METHODS** Initially, the 3D-structures of *SmCD1*, *SmCD2* and *SmCD3* were generated using homology modelling approach. Using these models, we prioritised 50 compounds from 20,000 compounds from ChemBridge database for further testing in adult worm aqueous extract (AWAE) and recombinant *SmCD1* using enzymatic assays.

**FINDINGS** Seven compounds were confirmed as hits and among them, two compounds representing new chemical scaffolds, named **5** and **19**, had  $IC_{50}$  values against *SmCD1* close to 100  $\mu$ M while presenting binding efficiency indexes comparable to or even higher than pepstatin, a classical tight-binding peptide inhibitor of aspartyl proteases. Upon activity comparison against mammalian enzymes, compound **50** was selective and the most potent against the AWAE aspartic protease activity ( $IC_{50} = 77.7$   $\mu$ M). Combination of computational and experimental results indicate that compound **50** is a selective inhibitor of *SmCD2*. Compounds **5**, **19** and **50** tested at low concentrations (10  $\mu$ M) were neither cytotoxic against WSS-1 cells (48 h) nor could kill adult worms *ex-vivo*, although compounds **5** and **50** presented a slight decrease on female worms motility on late incubations times (48 or 72 h).

**MAIN CONCLUSION** Overall, the inhibitors identified in this work represent promising hits for further hit-to-lead optimisation.

Key words: *Schistosoma mansoni* - drug discovery - hit identification - aspartic protease - virtual screening

Schistosomiasis is a parasitic disease caused by trematodes of the genus *Schistosoma*, which has several species, including *S. mansoni*. More than 236.6 million people needed treatment for schistosomiasis in 2019, according to the World Health Organization (WHO).<sup>(1)</sup> Praziquantel (PZQ) has been the drug of choice for the treatment of schistosomiasis for more than 40 years.<sup>(2)</sup> Park et al.<sup>(3)</sup> demonstrated that PZQ activates a transient

ion channel receptor called TRPM<sub>PZQ</sub>. This receptor promotes calcium influx and parasite paralysis. Despite a known mode of action, PZQ does not act on immature forms (eggs, larvae and juvenile worms) of *S. mansoni*.<sup>(4,5)</sup> Moreover, the bitter taste and size of PZQ tablets are also limitations, mainly for children's treatment. Finally, it has already been observed that the use of large-scale chemotherapy through successive treatments may result in the emergence of resistant isolates, through selective drug pressure.<sup>(6,7)</sup> Thus, the discovery of new drugs with new mechanisms of action would be of great value to the therapeutic arsenal against this important neglected tropical disease.

The development of the larval form (schistosomula) up to adulthood, as well as nutrition, reproduction, and growth of worms, requires proteolytic enzymes, which digest the proteins obtained from the host.<sup>(8,9,10)</sup> Parasites such as *Schistosoma* spp. produce a cathepsin D-like aspartyl protease 1 (*SmCD1*)<sup>(11)</sup> that, along with cysteinyl proteases, have a pivotal role in the early stages of the *S. mansoni* digestion cascade.<sup>(10,12)</sup> Studies from our group showed that *S. mansoni* presents a family of aspartyl

doi: 10.1590/0074-02760230031

Financial support: CAPES (Finance Code 001), CNPq, FAPEG (grants #20171026700006 and #202010267000272), FAPERJ, FIOCRUZ.

FPSJr is a CNPq productivity fellow and FAPERJ CNE fellowship recipient; CHA is a CNPq, FAPEG and "L'Oréal-UNESCO-ABC Para Mulheres na Ciência" and "L'Oréal-UNESCO International Rising Talents" fellowship recipient.

+ Corresponding author: floriano@ioc.fiocruz.br

✉ <https://orcid.org/0000-0003-4560-1291>

Received 15 February 2023

Accepted 01 August 2023



proteases genes (collectively called *SmAPs*)<sup>(13)</sup> Although only one of these genes, named *SmCD1*, has been formally described in the literature,<sup>(11,12,14)</sup> two new *SmCDs* were identified and named *SmCD2* and *SmCD3*.<sup>(13)</sup> Araujo-Montoya et al.<sup>(14)</sup> developed an expression system in HEK 293T mammalian cells and performed a biochemical and biophysical characterisation of the recombinant, r*SmCD1* enzyme. The purified enzyme showed activity in degrading human haemoglobin and cleaving a commercial aspartyl protease peptide substrate, as well being inhibited by pepstatin, a classic inhibitor of aspartyl proteases.

Previously, it has been observed that the RNAi-mediated reduction in transcriptional levels of *SmCD1* leads to phenotypic changes in the parasite, including significant *in vitro* growth retardation.<sup>(11,15)</sup> In addition, the dark pigment of hemozoin, which is the product of haemoglobin proteolysis, did not accumulate in the gastrodermis of the treated parasites. Thus, these studies indicate that *SmCD1* is an essential enzyme in the stages of schistosomes parasitising the mammalian host. As such, *SmCD1* can be considered a validated and exceptionally promising target for the development of drugs against schistosomiasis. Until this moment, known inhibitors of *SmCD1* are scarce. Only one HIV-1 protease inhibitor and few statin-based peptides are known to inhibit this enzyme or its orthologue in *S. japonicum*.<sup>(13,15,16,17)</sup>

The technological advances in the computational field have contributed to increase the efficiency of the drug discovery process.<sup>(18,19)</sup> In this context, along with advances in structural biology, the virtual screening (VS) technique emerged as a strategy for the identification of new bioactive substances that is widely used in the early stages of the drug discovery pipeline.<sup>(20)</sup> The main reason behind the success of this computational approach is the possibility of prioritising compounds from a large virtual library based on their predicted biological activity. Thus, there is a significant decrease in the number of experimental trials to be performed and consequently in the total cost and time of the whole process.<sup>(21)</sup> Several active compounds against *S. mansoni* have already been identified by VS, including protease inhibitors.<sup>(19,22,23)</sup>

In this work, we performed a structure-based VS campaign to identify the first classes of small molecule inhibitors of *SmCD1*. To achieve this goal, we pursued the following steps: (i) to build the 3D-structures of *SmCD1*, *SmCD2* and *SmCD3* using homology modelling; (ii) to perform a docking-based VS with 20,000 compounds available on MicroFormat and DIVERSet-EXP datasets; (iii) to perform experimental validation of the best scored computational hits against the adult worm aqueous extract (AWAE) and recombinant *SmCD1*; (iv) to assess inhibitor selectivity by comparing to inhibition of mammalian homologous enzymes; and (v) to evaluate the *in vitro* effect of hit compounds on *ex vivo* adult *S. mansoni* worms and cultured WSS-1 cells.

## MATERIALS AND METHODS

### Computational

**Homology modelling** - The amino acid sequences of *SmCD1* (ID: P91802), *SmCD2* (ID: P91802), and *SmCD3* (ID: P91802) were retrieved from the UniProt

database<sup>(24,25)</sup> and used as targets for homology modelling. Templates for modeling were selected from the Protein Data Bank (PDB)<sup>(26)</sup> using as criteria the highest sequence identity with the targets, best resolution of the available experimental structure and biological function relevance. Initially, a target-template alignment was calculated by T-Coffee server.<sup>(27)</sup> Subsequently, 50 homology models were built for each protein using “satisfaction of spatial restraints” method implemented in MODELLER v.8.<sup>(28)</sup> In this approach, a 3D model is obtained by satisfying main-chain and side-chain dihedral angles of all non-hydrogen atoms derived from the alignment and expressed as probability density functions (PDFs). The 3D models were obtained by optimisation of the molecular PDF such that the model violates the input restraints as little as possible.<sup>(28)</sup> Finally, the generated models were exported to SAVES server (<http://services.mbi.ucla.edu/SAVES/>) and their overall stereochemical and structural quality were checked according to PROCHECK,<sup>(29,30)</sup> VERIFY,<sup>(31,32)</sup> and PROVE<sup>(33)</sup> scores.

**Ligand preparation** - The 3D structures of the small molecules were imported to MarvinSketch v.6.3.1 (ChemAxon, Budapest, Hungary, <http://www.chemaxon.com>) and protonated at pH = 3.5 ± 0.5. Subsequently, up to 1,000 conformers were generated for each structure using OMEGA v.3.0.0.,<sup>(34,35)</sup> and the AM1-BCC charges<sup>(36)</sup> were added using QUACPAC v. 1.7.0.2.<sup>(37)</sup>

**Protein preparation** - The Protein Preparation Wizard program, which is available on Maestro workspace v.9.3 (Schrodinger, LCC), was employed to process the 3D structures of the *SmCDs*. This procedure involves the addition of hydrogen atoms to the proteins and the adjustment of bond orders and formal charges. Additionally, the protonation states of polar amino acid residues were forecasted at pH = 3.5 ± 0.5 using PROPKA v.3.1. In contrast, Asp219 and Asp33's protonation states, which are part of the catalytic triad, were manually adjusted to their neutral form based on experimental findings for other aspartic proteases within the same family<sup>(38,39,40,41)</sup> and following the catalytic mechanism proposed by Veenrpadian et al.<sup>(42)</sup>

**Molecular docking** - The 3D structures of *SmCD1*, *SmCD2*, and *SmCD3* were subjected to the grid-generation in OEDocking suite v.3.2.0.2<sup>(43,44,45)</sup> by fixing a box on the geometrical center of catalytic triads (Asp33, Thr34, and Gly35). Details of grid boxes can be found in Supplementary data (Table I). Finally, molecular docking calculations were performed in FRED program<sup>(43,44,45)</sup> using standard and high-resolution protocols as well as the ChemGauss4 score function.<sup>(44,45,46)</sup> High resolution protocol improves every possible rotation and translation of each conformer of the ligand being docked within a box enclosing the active site.<sup>(43,44,45)</sup> During the docking, any pose examined by the exhaustive search that does not fit within the outer shape (~ 5 Å of the triad) of the corresponding grid box was rejected.

**Virtual screening (VS)** - After defining the molecular docking protocols for each target, they were employed as filters for the virtual screening of 20,000

compounds available in the MicroFormat and DIVERSet-EXP collections of the ChemBridge database (<http://www.chembridge.com/>). Following that, the KNIME workspace v.1.3.5 was utilised to import the top 2.5% virtual hits, which were then assessed for their metabolic stability against nine cytochrome P450 enzymes (CYP3A4, CYP2D6, CYP2C9, CYP2C19, and CYP1A2) using our in-house machine learning models. Meanwhile, the hERG blockage potential was predicted using the Pred-hERG web server.<sup>(47,48)</sup> To develop the cytochrome P450 machine learning models, we used the Random Forest algorithm and ECFP4 fingerprints, which were implemented in KNIME v.1.3.5. Then, a medicinal chemistry inspection filter based on using Tanimoto coefficient and MACCS keys was employed to ensure the structural novelty of virtual hits in relation to known anti-schistosomal compounds available on ChEMBL database (ID: ChEMBL612893). At the end of this process, the docking poses were visually inspected, and the best virtual hits were selected based on interactions with the catalytic triad.

## Experimental

**Chemicals** - The reagents bovine cathepsin D (*BtCD*) (C3138) and pepstatin A (P5318) were obtained from Sigma-Aldrich, porcine pepsin (0151-17 / 215110) from Difco, Abz-AIAF/FSRQ-EDDnp fluorescent substrate (102023-1) from GenOne and E-64 (78434) from Thermo Scientific. The other reagents used, as salts, buffers, 3-[(3-cholamidopropyl)dimethylammonio]-1-propane-sulfonate (CHAPS), dimethyl sulfoxide (DMSO), culture medium were of the PA grade or the highest level of purity available on the market.

**Parasites** - Swiss Webster mice were used to obtain *S. mansoni* adult worms. The animal infection was described previously by our group.<sup>(49)</sup> All experiments were carried out in accordance with the Oswaldo Cruz Foundation Institutional Ethics Committee for Laboratory Animals (CEUA / FIOCRUZ, Brazil, L-044/15).

**Preparation of adult worm aqueous extract (AWAE)** - The extract was prepared from the homogenisation of 1 mL of worms, containing both male and female, in 5 mL of phosphate buffered saline (PBS) + 1% CHAPS, at pH 7.4, using a Turrax disperser. This volume of parasites is roughly equivalent to 20 adult parasites. The homogenate was centrifuged at 20,000 g for 1 hour at 4°C, and the supernatant stored at -80°C. An aliquot was used for the protein dosing by the Lowry method (DC<sup>TM</sup> protein assay - BIO-RAD).

**SmCD1 cloning, expression and purification** - Expression of the recombinant enzyme was previously performed by our group<sup>(14)</sup> by transfection of the vector with recombinant proSmCD1 gene (proSmCD1r) into culture bottles containing human embryonic kidney (HEK293) cells. After an incubation period of three to five days, the culture medium of the HEK293 cells was collected, filtered, and used for the purification of the recombinant protein. A HisTrap FF column in-tandem with a HiLoad 16/60 Superdex column, connected to an Äkta Xpress system were used for the purification of the protein.

**SmCD1 enzymatic assay** - Initially, 2.5 µg of SmCD1 was added to a reaction medium containing 50 mM sodium acetate buffer with 50 mM NaCl at pH 3.5. Next, this medium was incubated with either test compounds (10 µM), pepstatin (1 - 500 µM), as a positive control, or DMSO (0,1%), as negative control, for 10 minutes, at 37°C. After incubation, to start the reaction, 10 µM of the substrate octapeptide fluorescent (FRET) Abz-AIAF/FSRQ-EDDnp (102023-1, GenOne) was added to the medium. The substrate hydrolysis reaction product was continuously detected ( $\lambda_{ex}$  = 310 nm and  $\lambda_{em}$  = 420 nm), using a Flex Station III microplate reader (Molecular Devices), during the initial 10 min of the reaction.

All enzymatic assays in this work were performed as described above, and repeated at least three times, each condition in triplicate. In tests involving screening of compounds, the percentage of inhibition of enzymatic activity was calculated considering the mean of the negative control as 100% activity.

**AWAE aspartic protease activity assay** - For assaying the aspartyl protease activity present in the AWAE, the same protocol described above for SmCD1 was followed, except for 20 µg of total proteins present in AWAE were added instead of the recombinant SmCD1 enzyme. Additionally, 5 µM of E-64, an inhibitor of cysteinyl proteases, was added to prevent enzymes of this class present in the AWAE<sup>(15)</sup> to also contribute to substrate hydrolysis on the acidic pH of the assay.

**Specificity assays toward mammalian proteases** - For recombinant *BtCD* and porcine pepsin, the same protocol described above for SmCD1 was followed, except for 10 µg of *BtCD* or 5 µg of pepsin were added.

**Determination of IC<sub>50</sub> values** - For IC<sub>50</sub> (inhibitor concentration in which the enzymatic activity is reduced by 50%) determination assays, serial dilutions of the inhibitors were performed, ranging from 1 µM to 500 µM. The IC<sub>50</sub> values were determined by fitting SmCD1 activity data into the 4-parameter logistic equation.

**Cytotoxicity assays on human WSS-1 cells** - The cytotoxicity of compounds 5, 19 and 50 was evaluated against normal epithelial cells derived from human kidney (WSS-1, ATCC<sup>TM</sup> CRL-2029). Cells were seeded in 96-well plates, at 5x10<sup>4</sup> cells/mL, in DMEM high glucose medium supplemented with 10% foetal bovine serum (FBS). After adherence period of 24 h, at 37°C and 5% CO<sub>2</sub>, the cells were treated with compounds at 10 µM or 100 µM concentrations and incubated for 44 h. After this time, resazurin solution was added to cells culture, at a final concentration of 0.01 mg/mL. The cells returned to the incubator for another 4 h to complete 48 h total incubation time.<sup>(50)</sup>

Fluorescence readings of resorufin ( $\lambda_{ex}$  = 560 nm,  $\lambda_{em}$  = 590 nm) were taken immediately (F0) and 4 hours (F4) after resazurin addition, using Flex Station III. The percentage of viable cells at the end of the experiment, that is, after compounds exposure for 48h, was calculated as follows:

$$\text{Viable cells} = \frac{F4-F0 \text{ treated}}{F4-F0 \text{ control}} \times 100$$

**Schistosoma mansoni adult worms ex vivo assay** - The effect of 10  $\mu$ M of the test compounds or PZQ (positive control) on adult *S. mansoni* worms motility for up to 72 h of incubation at 37°C and 5% CO<sub>2</sub> was evaluated by a high content assay previously developed by our group.<sup>(50)</sup> Image analysis was carried out in a customised pipeline of the open-source CellProfiler software v. 4.2.1<sup>(51)</sup> and consisted of quantifying parasites area (Area measurement of MeasureObjectSizeShape module) and its displacement over pairs of consecutive images (false negative rate measurement of MeasureObjectOverlap module). These data were then modeled by a multilevel generalised linear statistical model. The effects of the experimental treatments on parasites motility were estimated from this model by the integrated nested laplace approximation (INLA) method available in INLA package of R (<https://www.r-inla.org>). The results were expressed as percentual relative motility considering the estimated effect of the negative control group (0.2% DMSO) for each incubation time as 100%.

**Graphics and statistical analysis** - GraphPad Prism version 5.04 (GraphPad Software Inc., San Diego, USA) was used for statistical analysis. The results were expressed as mean  $\pm$  standard error. Differences were considered significant when  $p < 0.05$ .

## RESULTS AND DISCUSSION

**Homology modelling** - The 3D structures of *SmCD1*, *SmCD2*, and *SmCD3* were not available on the PDB at the time this work was conducted. Therefore, homology models were built using the crystal structure of human cathepsin D (PDB ID: 1LYB<sup>(52)</sup>) as template, since it shared identity of 57% and the highly conserved catalytic triad (Asp33, Thr34, and Gly35). Subsequently, the quality of obtained 3D models (Fig. 1) was analysed for various levels of structural information.

Initially, PROCHECK analysis was performed to determine stereochemical quality of models by analysing residue by residue geometry and overall structure geometry.<sup>(29,30)</sup> Analyses of dihedral angles *phi* against *psi* of amino acid residues (see Ramachandran plots in Fig. 1) indicated that 92.3-89.9% of residues of 3D models are within the most favoured regions, 9.7-7.0% of residues are within the additional allowed regions, only 1.0-0.4% of residues are within the disallowed regions.

Subsequently, PROVE analysis calculated the atom volumes of each *SmCD* structure using an algorithm that treats the atoms like hard spheres and considered a statistical Z-score deviation in relation to high resolution structures (2.0 Å or better) deposited in PDB.<sup>(33)</sup> As shown in Fig. 1, all models had Z-score values close to 0,

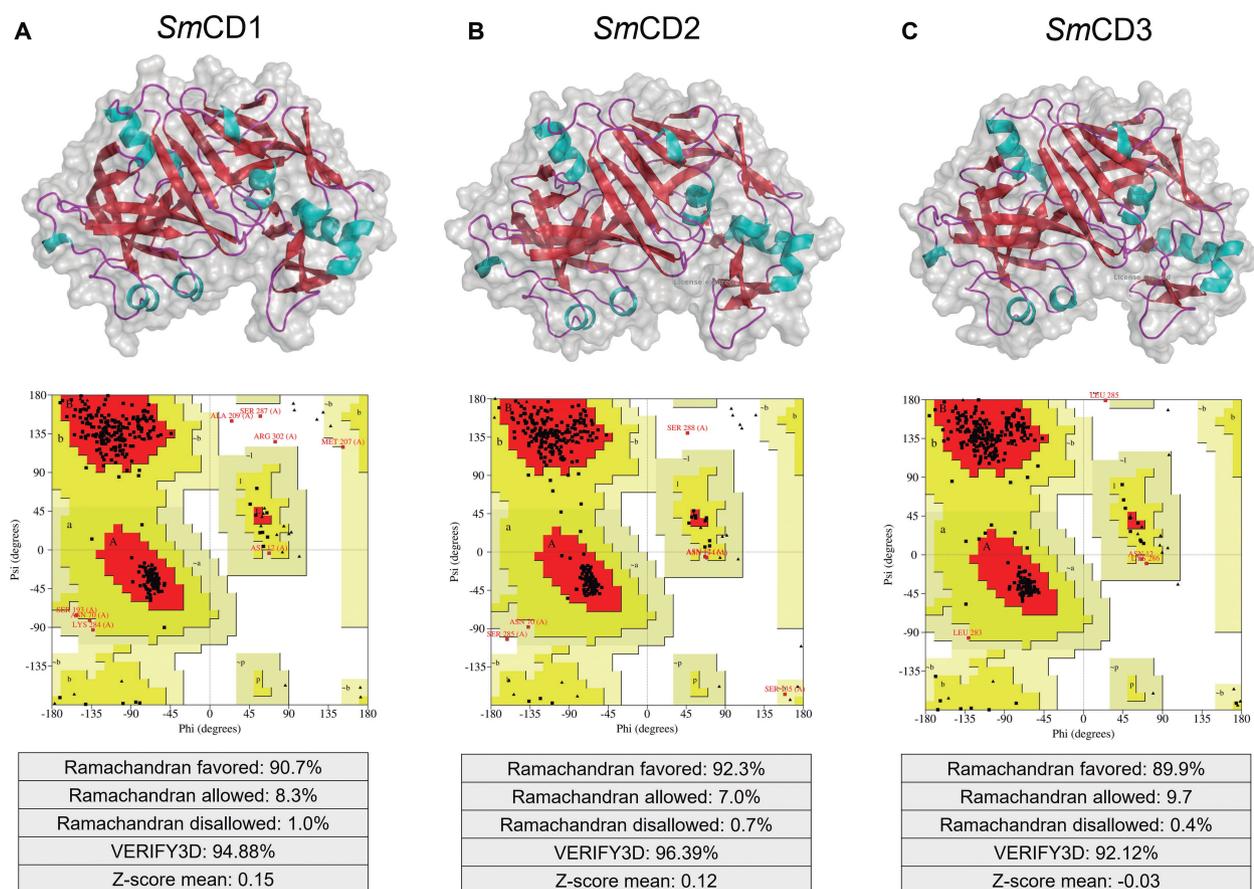


Fig. 1: predicted 3D structures of *Schistosoma mansoni* cathepsin D-like aspartyl protease 1 (*SmCD1*, (A), 2 (*SmCD2*, (B) and 3 (*SmCD3*, (C), generated by SWISS-MODEL. Ramachandran plots for each homology model were obtained using PROCHECK and show the dihedral angles Psi and Phi of amino acid residues. Red represents most favoured regions in plot; yellow represents additional allowed regions; beige represents generously allowed regions; and white areas are disallowed regions.

reflecting the accuracy of homology modelling protocol used in this work.

Further, VERIFY-3D was used to determine the compatibility of 3D models with its own amino acid sequence by assigning a structural class based on its location and environment (alpha, beta, loop, polar, nonpolar, etc.) and comparing the results to good structures.<sup>(31,32)</sup> As a result, 96.39-92.12% of the residues had average 3D-ID scores  $\geq 0.2$ . This suggests that the 3D models investigated in this work have overall self-consistency in terms of sequence-structure compatibility.

As depicted in Fig. 2A, the active site of *Sm*CDs consists of adjoining pockets denoted as -S4'-S3'-S2'-S1'-S1-S2-S3-S4-. These pockets are chemical and spatial complements of the substrate amino acid residues, consecutively numbered as -P<sub>4'</sub>-P<sub>3'</sub>-P<sub>2'</sub>-P<sub>1'</sub>-P<sub>1</sub>-P<sub>2</sub>-P<sub>3</sub>-P<sub>4</sub>-.<sup>(16,53)</sup> The enzyme catalyses hydrolysis of the substrate scissile

bond located between the P<sub>1'</sub> and P<sub>1</sub>. Usually, aspartic protease inhibitors that bind to the S1-S4 pockets are frequently encountered, while the S1'-S4' pockets typically remain unoccupied.

To explore the feasibility of the suggested binding site, active site similarity profiles were validated by 3D alignment of the *Sm*CDs (Fig. 2B). The alignment of *Sm*CD1 and *Sm*CD2 residues in Fig. 2C highlights a significant degree of preservation of the general topology of S1'-S1-S2-S3-S4 residues. Conversely, the S1 (Leu79 and Asp80), S1' (Ser297 and Met194), S3 (Tyr115, Asp114, Asn117, Tyr120, and Cys118), and S4 (Phe13) pockets of *Sm*CD3 exhibit marked physicochemical disparities when contrasted with *Sm*CD1 and *Sm*CD2 (Fig. 2C). Based on these dissimilarities, it appears that the active sites of *Sm*CD1 and *Sm*CD2 enzymes are comparatively more solvent-accessible than that of *Sm*CD3.

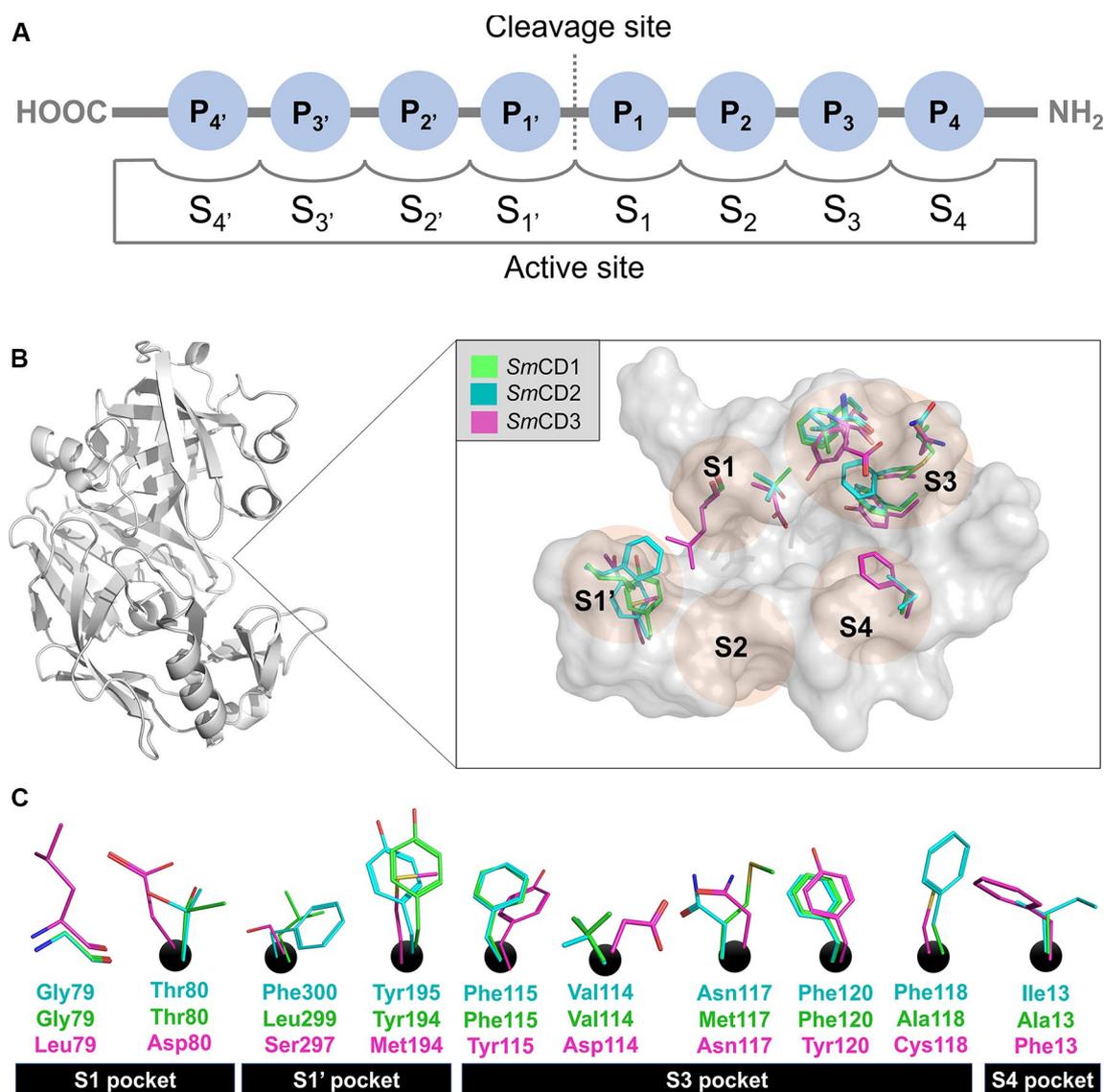


Fig. 2: (A) A diagram illustrating the binding of a substrate to the active site of *Schistosoma mansoni* cathepsin D-like aspartyl proteases (*Sm*CDs), wherein the positions are named based on the cleavage site's location.<sup>(16)</sup> (B-C) A representation depicting the primary variations among the active site residues of the enzymes *Sm*CD1, *Sm*CD2, and *Sm*CD3.

**Structure-based virtual screening** - The VS of new *SmCDs* inhibitors was carried out following the workflow presented in Fig. 3.

Initially, 20,000 compounds with lead-like properties available on MicroFormat and DIVERSet-EXP collections of ChemBridge were downloaded and standardised for VS. Then, the standard and high-resolution docking protocols were used as filters to prioritise compounds with potential *SmCD1*, *SmCD2*, and/or *SmCD3* inhibitory activity. Docking assays against these three enzymes were performed with the rationale that a compound inhibiting multiple *SmAP* targets should maximise the potential schistosomicidal activity of putative hits. The top 2.5% ranked hits (ChemGauss4 scores ranging between -7.00 and -14.00) were selected for further filtering. The compounds were then evaluated by predicting a panel of pharmacokinetic and toxicity properties including CYP450 metabolic stability and inhibition and hERG blockage.<sup>(47,48)</sup> At the end of the VS workflow, 50 putative hits representing new chemical scaffolds for schistosomiasis were visually inspected and purchased for experimental validation.

**Experimental validation** - The inhibitory activity of the 50 putative hits, at 10  $\mu\text{M}$ , was experimentally investigated in porcine pepsin, *SmCD1*, and AWAE aspartic protease activity assay, using pepstatin as a positive control [Supplementary data (Table II)]. The AWAE assay captures the combined activities of all *SmAPs* present in schistosome adult worms, while excluding the concurrent cysteinyl protease activity, due the addition of the broad-spectrum E-64 inhibitor. Among the 50 compounds tested, seven compounds (**1**, **5**, **12**, **19**, **22**, **32** and **50**, Fig. 4) were considered active based on the  $\geq 25\%$  inhibition of any of the aspartic protease activity [Supplementary data (Table II)].

Following the initial screening, the seven primary hits were selected for determining  $\text{IC}_{50}$  against pepsin, *SmCD1*, and AWAE (Table).

Among primary hits, compounds **5** (2,3-dioxo-N-(m-tolyl)-1,2,3,4-tetrahydroquinoxaline-6-sulfonamide) and **19** (2-((7-methyl-5,6,7,8-tetrahydrobenzo[4,5]thieno[2,3-d]pyrimidin-4-yl)thio)-N-(m-tolyl)acetamide) showed the best potencies against *SmCD1*, with  $\text{IC}_{50}$  values of 83.6  $\mu\text{M}$  and 102.5  $\mu\text{M}$ , respectively (Fig. 5A-B). To es-

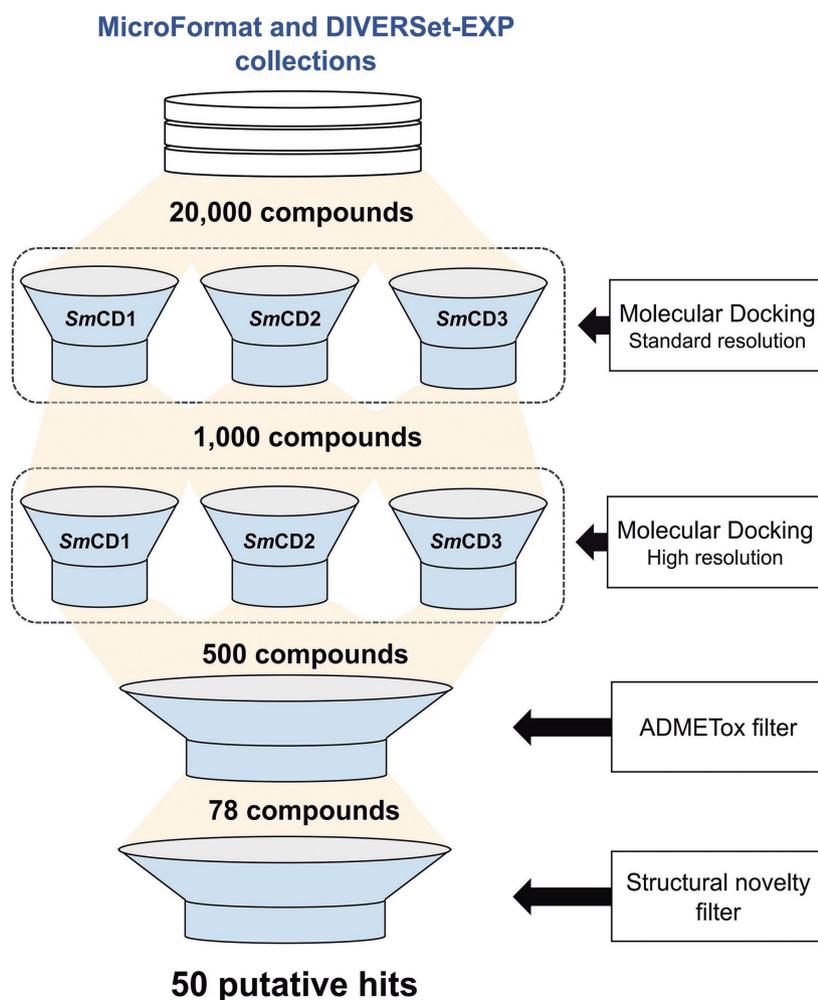


Fig. 3: virtual screening workflow used for identifying potential *Schistosoma mansoni* cathepsin D-like aspartyl protease (*SmCD*) inhibitors.

cape the affinity-biased selection and optimisation towards larger ligands, we calculated the binding efficiency index defined as  $BEI = pIC_{50}/MW$  (BEI)<sup>(54,55)</sup> of our most potent hits. Prioritising hits according to their BEI values allows smaller low affinity compounds to be attractive for further hit-to-lead optimisation. Compounds **5** and **19** presented BEI values (12.3 and 10.4, respectively) comparable to positive control pepstatin (BEI = 11.9). However, our hits have lower molecular weights (compound **5** = 331 Da, compound **19** = 384 Da) than pepstatin (685.9 Da), and therefore have greater potential for prospective hit-to-lead optimisation. Since BEI rank compounds on a negative logarithmic scale ( $pIC_{50}$ ), an increase or decrease of one unit implies a corresponding 10-fold change of potency per MW.<sup>(54,55)</sup>

*Rationalising SmCD1 inhibition by compounds 5 and 19* - Understanding the interaction pattern of our most promising hits *i.e.*, compounds **5** and **19**, into the SmCD1 active site is essential for designing more potent analogs.<sup>(16,53)</sup>

As shown in Fig. 6A for human pepsin, it is common to find aspartic protease inhibitors that bind to S<sub>1</sub>-S<sub>4</sub> pockets (*e.g.*, phosphonate inhibitor IVA-Val-Val-Leu(P)-(O)Phe-Ala-Ala-Ome<sup>(56)</sup>), PDB ID: 1QRP, whereas the pockets S<sub>1'</sub>-S<sub>4'</sub> remain empty. According to the docking poses, compounds **5** and **19** bind to the pockets S1-S3 and S1'-S1 (see Fig. 6B-C, respectively).

The binding mode of compound **5** in active site of SmCD1 (Fig. 6B) can be generalised as follows: the quinoxaline-2,3-dione moiety can form hydrogen bonds (represented as blue lines) with the carbonyl/amine back-

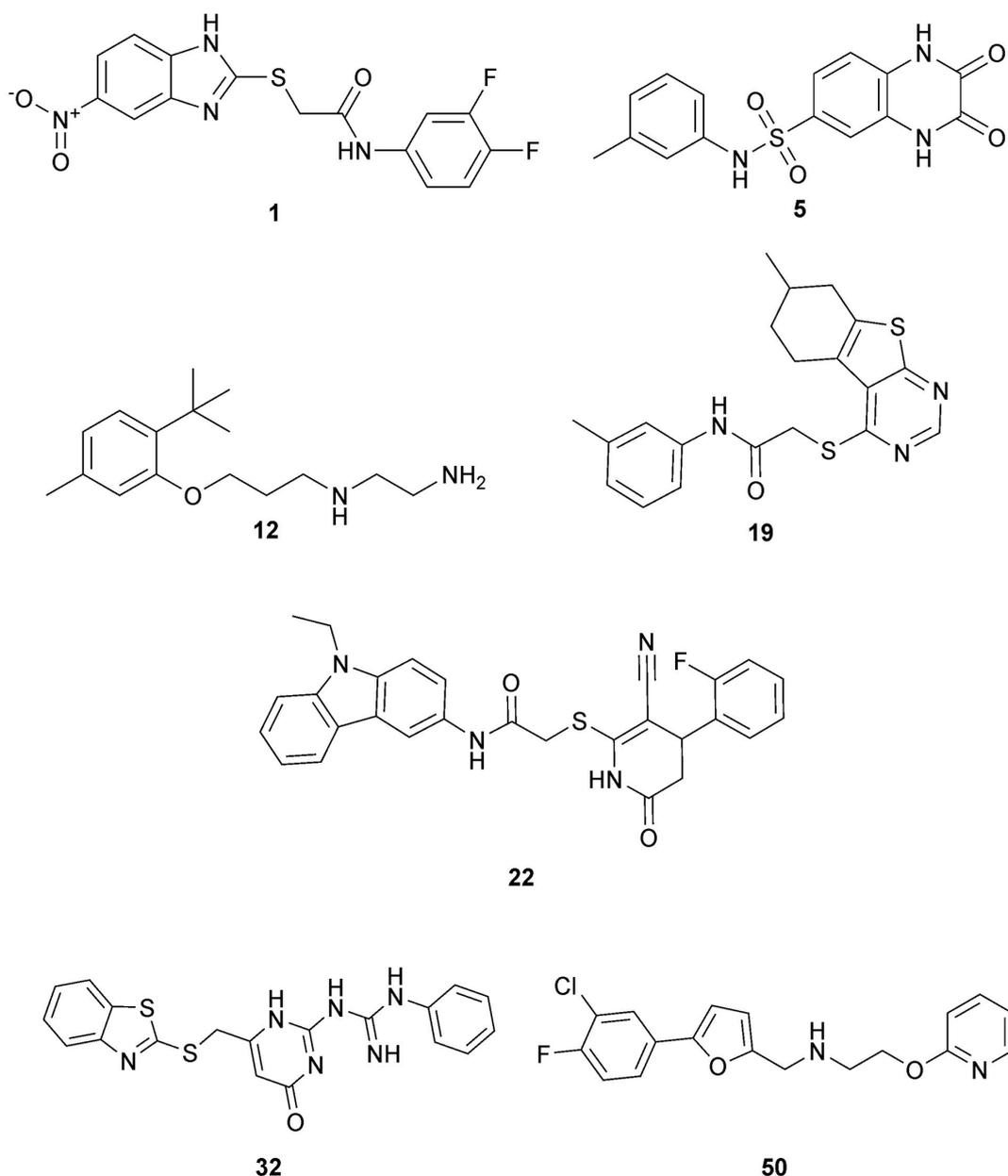


Fig. 4: chemical structures of active compounds on aspartic protease inhibition screening.

TABLE

Inhibitory potencies of seven primary hits against the adult worm aqueous extract (AWAE), *Schistosoma mansoni* cathepsin D-like aspartyl protease 1 (*SmCD1*) and pepsin. Confidence intervals (95% level) are shown in parentheses

Compound	IC <sub>50</sub> (μM)		
	Pepsin	AWAE	<i>SmCD1</i>
1	> 500	> 500	> 500
5	92.6 (81.6 to 108)	135.1 (*to 240)	83.6 (75.1 to 98.6)
12	231 (215 to 248)	315 (300 to 338)	> 500
19	23.1 (14.8 to 36.6)	> 500	102.5 (77.7 to 138)
22	38.0 (20.7 to 59.4)	> 500	> 500
32	72.5 (35.3 to 309.5)	> 500	> 500
50	> 500	77.7 (63.9 to 104)	> 500

\*The lower limit could not be determined.

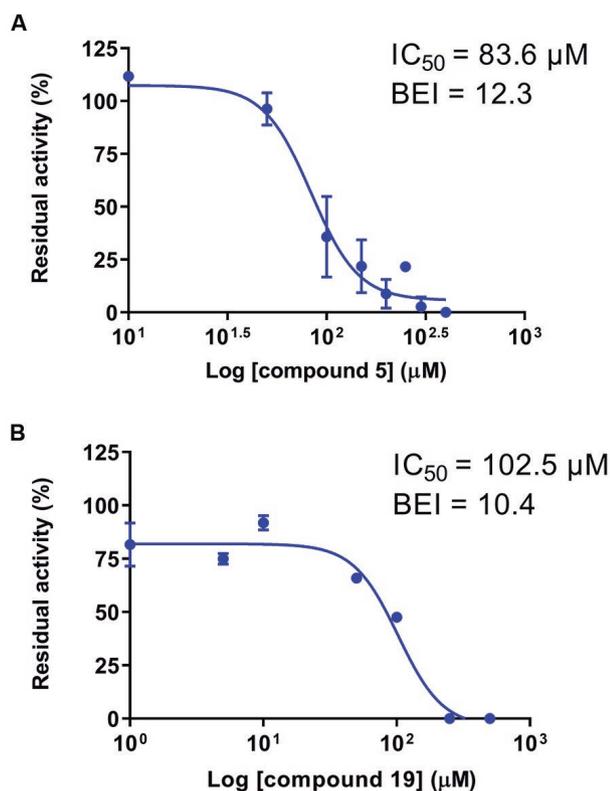


Fig. 5: concentration-response curves for *Schistosoma mansoni* cathepsin D-like aspartyl protease 1 (*SmCD1*) inhibition by compounds 5 (A) and 19 (B). Error bars: mean ± standard deviation (SD) of three independent replicates. Binding efficiency index (BEI) = pIC<sub>50</sub>/MW(kDa).

bone of the amino acid Thr80, Asp219, and Gly221. Thr80 is part of the “Flap”, a flexible region of the enzyme, which can be in different positions, according to the state of the protein. The possible function of this structure is to facilitate the binding of the substrate or inhibitor to the enzyme,<sup>(57)</sup> with a key role in this enzyme-binding interaction.<sup>(58)</sup> In addition, the phenyl ring in compound 5 forms van der Waals interactions with pocket formed by Val114 and Met117, whereas nitrogen of sulfonamide group forms a hydrogen bond with Ala113.

On the other hand, compound 19 is predicted to interact through the substrate-binding cleft (Fig. 6C) in a different way, as follows: the amide group can form hydrogen bond interactions with Asp33 and the backbone of Thr80 whereas aromatic and aliphatic rings form van der Waals interactions with Ile123, Val31, Phe120, Leu297, Leu299, Tyr194, and Ile217.

Considering that our hit compounds do not completely occupy the S<sub>1</sub>-S<sub>4</sub> pockets, we believe that their chemical structures still need to be optimised to generate larger molecular entities with improved potency and BEI values against *SmCD1*. Since both hits partially occupy the S<sub>1</sub> pocket, we emphasise the possibility of carrying out a molecular hybridisation through the fusion of quinoxaline-2,3-dione moiety of compound 5 and the phenyl ring of compound 19. Another relevant possibility for structural optimisation involves the growth of our hits, in order to increase the size and occupation of the active site toward the S<sub>4</sub>, S<sub>2</sub>, and S<sub>1</sub> regions (Fig. 6D).

**Selectivity towards parasite proteases** - Compound 5 could not inhibit more than 45% of the *BtCD* activity at the highest concentration evaluated (500 μM due to solubility limitation), indicating its IC<sub>50</sub> value for *BtCD* will be over 500 μM. However, Table shows that while compounds 5 and 19 were the most potent against recombinant *SmCD1*, these compounds were not selective, being able to inhibit porcine pepsin with similar or even higher potencies. Compound 22 was a potent inhibitor of porcine pepsin but not against *SmAPs* in the AWAE or recombinant *SmCD1*. On the other hand, compound 50 was the most potent against the AWAE aspartic protease activity (IC<sub>50</sub> = 77.7 μM) while not showing significant inhibition against pepsin or *SmCD1* up to tested concentration 500 μM.

Although compound 50 has been prioritised as a potential dual inhibitor of *SmCD1* and *SmCD2*, our experimental validation indicates that this compound has inhibitory activity on the AWAE assay but not on *SmCD1*. Therefore, we hypothesise that the observed inhibition of the AWAE activity by compound 50 is due to inhibition of *SmCD2* present in this extract. In view of this, differences in the binding modes of compound 50 at the active sites of the three *SmCDs* were exploited to rationalise its preference for *SmCD2*. According to the docking calculations, compound 50 binds to the pockets S1'-S3 of *SmCD1* (Fig. 7A) and *SmCD2* (Fig. 7B). Although the binding modes are quite similar, the pyridine and aromatic rings of compound 50 make π-stacking interactions with the Tyr195, Phe300 and Phe118 residues of *SmCD2* (Fig. 7B). The planar binding profile does not

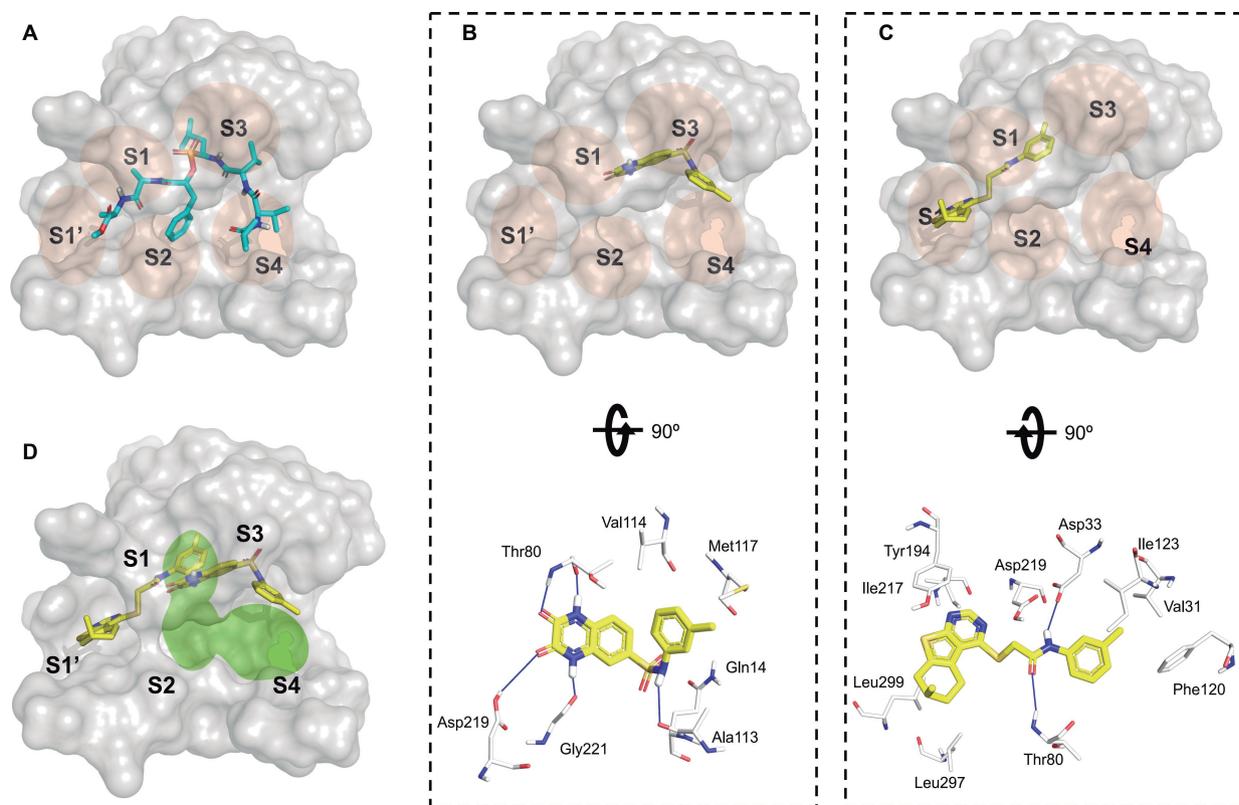


Fig. 6: rationalising *Schistosoma mansoni* cathepsin D-like aspartyl protease 1 (*SmCD1*) inhibition.<sup>(16)</sup> (A) phosphonate inhibitor IVA-Val-Val-Leu(P)-(O)Phe-Ala-Ala-Ome bound to human pepsin. (B-C) Binding interactions of *SmCD1* with compounds 5 and 19 predicted by docking. (D) Superimposition of docked poses for compounds 5 and 19, highlighting in green regions for exploring structural optimisation by fusion or growth. The amino acids involved were represented as CPK coloured sticks. The hydrogen bond interactions with protein backbone and side-chain atoms are shown in blue lines.

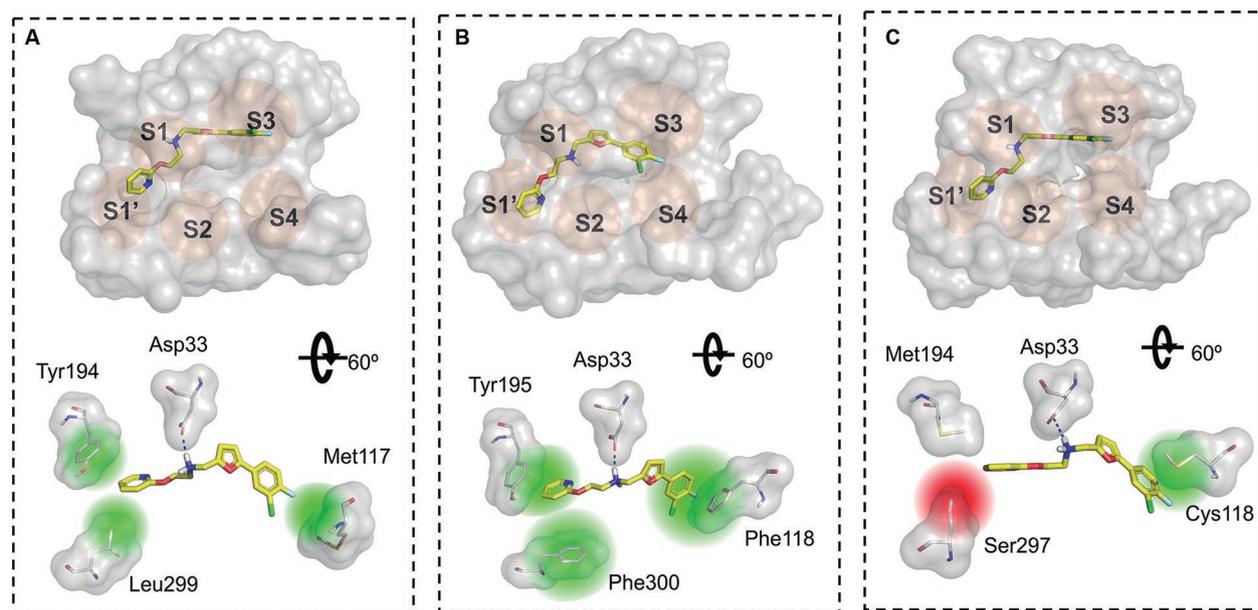


Fig. 7: binding modes of compound 50 into the active sites of *Schistosoma mansoni* cathepsin D-like aspartyl proteases (*SmCDs*). (A) *SmCD1*, (B) *SmCD2*, and (C) *SmCD3* predicted by docking. Amino acid residues are shown as gray colored sticks. The hydrogen bonding interactions with protein backbone and side-chain atoms are shown in dashed blue lines. Green regions indicate van der Waals or  $\pi$ -stacking interactions. The red sphere indicates a region where steric hindrance may occur.

occur in the active site of *SmCD1* (Fig. 7A), since the corresponding residues in this region (*i.e.*, Leu299 and Met117) do not have an aromatic character. On the other hand, compound 50 was probably not ranked as a putative hit for *SmCD3* because it is sterically hindered by Ser297 in the active site (Fig. 7C).

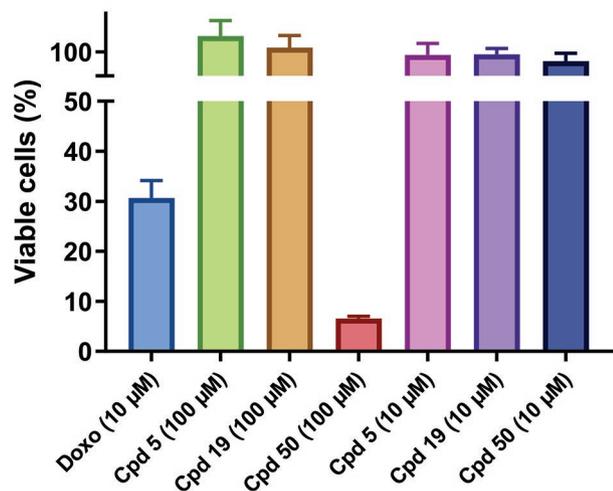


Fig. 8: effect on viability of WSS-1 cells after treatment with hit compounds 5, 19 and 50, at concentrations of 10 and 100 µM, for 48 hours. Doxo = doxorubicin; Cpd = compound. Bars represent the percentage of viable cells, calculated as the mean  $\pm$  standard error.

**Cytotoxicity and *S. mansoni* adult worms *ex vivo* assay** - In order to evaluate the antischistosomal effect and safety at the cellular level, hit compounds 5, 19 and 50 were tested on phenotypic assays against human WSS-1 cells and adult *S. mansoni* worms. Compounds 5 and 19 were non-toxic against WSS-1 cells for up to 100 µM. On the other hand, compound 50 was safe only at 10 µM, having decreased cell viability by more than 90% at 100 µM (Fig. 8).

In antischistosomal *ex vivo* assays, using female and male adult worms, compounds were evaluated at 10 µM for incubation times varying from 0 (just after compound addition) to 72 h incubation. Unfortunately, due to compound solubility limitations, we could only test the compounds at 10 µM concentration, which is the highest possible due to the carried over DMSO concentration (0.1%) from compound stock solutions, above which is toxic to the parasite. The results are shown in Fig. 9.

As shown in Fig. 9, distinctly from the reference drug PZQ, none of the compounds could kill adult worms as judged by lack of substantial decrease in worm motility. Compounds 5 and 50 presented a slight decrease on female worms motility on late incubations times (48 or 72 h). This apparent selective effect is consistent with the higher host protein digestion activity on female worms.<sup>(8,9,10)</sup> It is important to note that, the compounds were evaluated *ex vivo* at a concentration about 10 times lower (10 µM) than their respective  $IC_{50}$  values for EAV and/

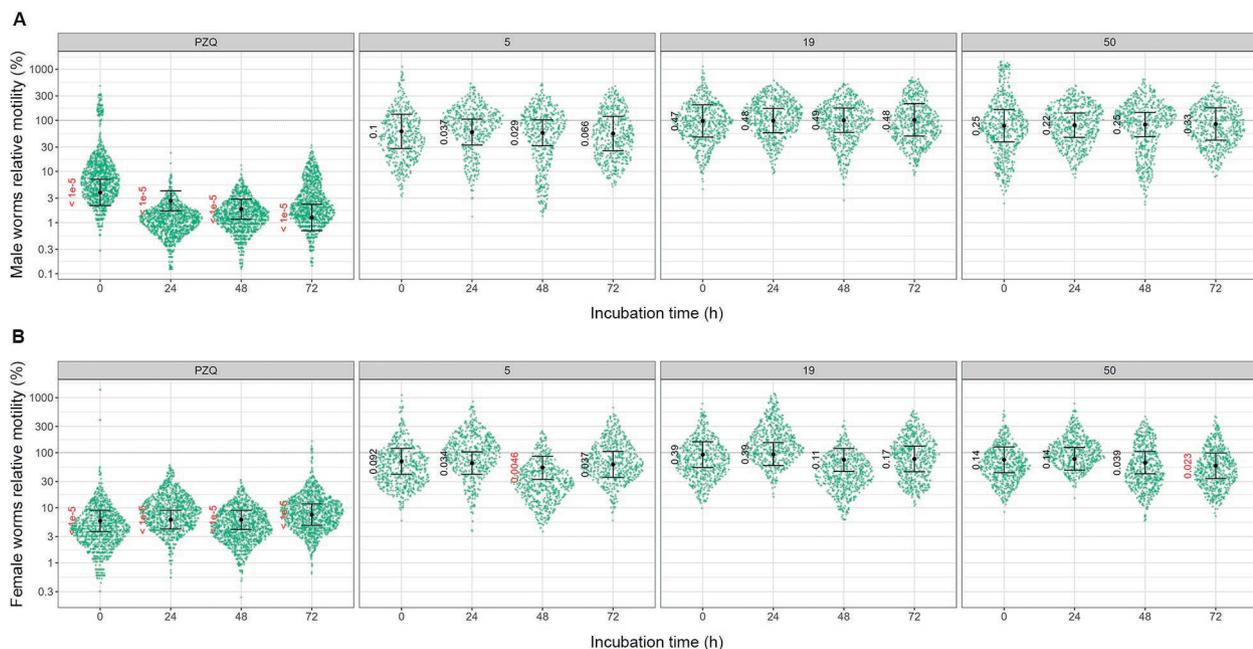


Fig. 9: relative motility of male (A) and female (B) adult *Schistosoma mansoni* worms incubated for up to 72 h with 10 µM of compounds 5, 19, 50 and praziquantel (PZQ). The x-axis of each graph represents the incubation time (immediately and after 24, 48 and 72 h of compound addition) with the test compounds while the y-axis represents the motility of the treated parasites relative to the control group (0.1 % DMSO) in the logarithmic scale. Each green dot represents single motility measurement of a worm. The estimated effects of compounds at each incubation time and their confidence intervals (95%) are represented by black dots and error bars, respectively. Motility estimates of treated worms whose confidence intervals do not touch the dashed line show statistically significant difference in relation to those of controls. The values shown next to the estimates represent the posteriori probability of the effect of a compound being null or opposite to the estimated effect at that time of incubation. In red, the effects considered statistically different from the effect of the control group and, in black, those that are not statistically significant (cutoff of 0.025). The plots are representative of two independent experiments and were generated with the R ggforce package (<https://cran.r-project.org/web/packages/ggforce/index.html>).

or SmCD1 (Table). At 10  $\mu$ M, a maximum inhibition of 35% of SmAPs activity was expected [Supplementary data (Table II)]. Thus, considering that statistically different effects on motility would result from the inhibition of SmAPs, the *ex vivo* results (Fig. 9) are consistent with the results of enzymatic activity.

*In conclusion* - The inhibitory capacity of compounds selected by virtual screening was tested on aspartyl protease activities of the recombinant enzyme SmCD1, aqueous extract of *S. mansoni* adult parasites and the homologous mammalian enzyme, porcine pepsin. Of a total of 20,000 compounds submitted to the virtual screening, the best ranked 50 were selected and tested on such enzymes. Of these, three compounds, **5**, **19** and **50** had the most interesting experimental enzyme inhibition results. Despite the lack of potent anti-parasite effect, under the *ex-vivo* assay condition, the results reported herein allow us to conclude that compounds **5** and **19** are relatively safe new chemical scaffolds that can be optimised into more potent and selective inhibitors of SmCD1 for the development of new antischistosomal drugs.

#### ACKNOWLEDGEMENTS

To the “Bioassay and Drug Screening Platform” from IOC-FIOCRUZ (RPT-111); to Walter CG Valente, João M Rezende-Neto, Giuliana V Schirato and Jordano F Reis for technical support.

#### AUTHORS' CONTRIBUTION

BFG prepared the original draft; FPSJr conceptualised the work. All authors contributed to editing and writing the final version of the manuscript. The authors report no conflicts of interest.

#### REFERENCES

- WHO - World Health Organization. Schistosomiasis. 2023. Available from: <https://www.who.int/news-room/fact-sheets/detail/schistosomiasis>.
- Park SK, Marchant JS. The journey to discovering a flatworm target of praziquantel: a long TRP. *Trends Parasitol.* 2020; 36(2): 182-94.
- Park SK, Friedrich L, Yahya NA, Rohr CM, Chulkov EG, Maillard D, et al. Mechanism of praziquantel action at a parasitic flatworm ion channel. *Sci Transl Med.* 2021; 13(625): 5832.
- Le Clec'h W, Chevalier FD, Mattos ACA, Strickland A, Diaz R, McDew-White M, et al. Genetic analysis of praziquantel response in schistosome parasites implicates a transient receptor potential channel. *Sci Transl Med.* 2021; 13(625): 9114.
- Pica-Mattoccia L, Cioli D. Sex- and stage-related sensitivity of *Schistosoma mansoni* to *in vivo* and *in vitro* praziquantel treatment. *Int J Parasitol.* 2004; 34(4): 527-33.
- Blanton RE. Population structure and dynamics of helminthic infection: schistosomiasis. *Microbiol Spectr.* 2019; 7(4): AME-0009-2019.
- Vianney TJ, Berger DJ, Doyle SR, Sankaranarayanan G, Serubanja J, Nakawungu PK, et al. Genome-wide analysis of *Schistosoma mansoni* reveals limited population structure and possible praziquantel drug selection pressure within Ugandan hot-spot communities. *PLoS Negl Trop Dis.* 2022; 16(8): e0010188.
- Caffrey CR, McKerrow JH, Salter JP, Sajid M. Blood “n” guts: an update on schistosome digestive peptidases. *Trends Parasitol.* 2004; 20(5): 241-8.
- Leontovyč A, Ulrychová L, O'Donoghue AJ, Vondrášek J, Marešová L, Hubálek M, et al. SmSP2: a serine protease secreted by the blood fluke pathogen *Schistosoma mansoni* with anti-hemostatic properties. *PLoS Negl Trop Dis.* 2018; 12(4): e0006446.
- Caffrey CR, Goupil L, Rebello KM, Dalton JP, Smith D. Cysteine proteases as digestive enzymes in parasitic helminths. *PLoS Negl Trop Dis.* 2018; 12(8): e0005840.
- Morales ME, Rinaldi G, Gobert GN, Kines KJ, Tort JF, Brindley PJ. RNA interference of *Schistosoma mansoni* cathepsin D, the apical enzyme of the hemoglobin proteolysis cascade. *Mol Biochem Parasitol.* 2008; 157(2): 160-8.
- Brindley PJ, Kalinna BH, Wong JYM, Bogitsh BJ, King LT, Smyth DJ, et al. Proteolysis of human hemoglobin by schistosome cathepsin D. *Mol Biochem Parasitol.* 2001; 112(1): 103-12.
- Rennó Soderó AC. Modelagem molecular de aspartil proteases de *Schistosoma mansoni* para desenvolvimento de potenciais moléculas esquistossomicidas [PhD Thesis]. Rio de Janeiro: Fundação Oswaldo Cruz; 2011. 144fls. Available from: <https://www.arca.fiocruz.br/handle/icict/5679>.
- Araujo-Montoya BO, Senger MR, Gomes BF, Harris G, Owens RJ, Silva-Jr FP. *Schistosoma mansoni* cathepsin D1: biochemical and biophysical characterization of the recombinant enzyme expressed in HEK293T cells. *Protein Expr Purif.* 2020; 167: 105532.
- Delcroix M, Sajid M, Caffrey CR, Lim KC, Dvořák J, Hsieh I, et al. A multienzyme network functions in intestinal protein digestion by a platyhelminth parasite. *J Biol Chem.* 2006; 281(51): 39316-29.
- Silva FP, Ribeiro F, Katz N, Giovanni-De-Simone S. Exploring the subsite specificity of *Schistosoma mansoni* aspartyl hemoglobinase through comparative molecular modelling. *FEBS Lett.* 2002; 514(2-3): 141-8.
- Caffrey CR, Placha L, Barinka C, Hradílek M, Dostál J, Sajid M, et al. Homology modeling and SAR analysis of *Schistosoma japonicum* cathepsin D (SjCD) with statin inhibitors identify a unique active site steric barrier with potential for the design of specific inhibitors. *Biol Chem.* 2005; 386(4): 339-49.
- Ferreira LLG, de Moraes J, Andricopulo AD. Approaches to advance drug discovery for neglected tropical diseases. *Drug Discov Today.* 2022; 27(8): 2278-87.
- Moreira-Filho JT, Silva AC, Dantas RF, Gomes BF, Souza Neto LR, Brandao-Neto J, et al. Schistosomiasis drug discovery in the era of automation and artificial intelligence. *Front Immunol.* 2021; 12: 1829.
- Maia EHB, Assis LC, de Oliveira TA, da Silva AM, Taranto AG. Structure-based virtual screening: from classical to artificial intelligence. *Front Chem.* 2020; 8: 343.
- Gimeno A, Ojeda-Montes M, Tomás-Hernández S, Cereto-Masagué A, Beltrán-Debón R, Mulero M, et al. The light and dark sides of virtual screening: What is there to Know? *Int J Mol Sci.* 2019; 20(6): 1375.
- Moreira BP, Batista ICA, Tavares NC, Armstrong T, Gava SG, Torres GP, et al. Docking-based virtual screening enables prioritizing protein kinase inhibitors with *in vitro* phenotypic activity against *Schistosoma mansoni*. *Front Cell Infect Microbiol.* 2022; 12: 913301.
- Gallinger TL, Aboagye SY, Obermann W, Weiss M, Grünweller A, Unverzagt C, et al. First *in silico* screening of insect molecules for identification of novel anti-parasitic compounds. *Pharmaceutics.* 2022; 15(2): 119.
- Apweiler R, Bairoch A, Wu CH, Barker WC, Boeckmann B, Ferro S, et al. UniProt: the Universal Protein knowledgebase. *Nucleic Acids Res.* 2004; 32(Database issue): D115-9.

25. Bateman A, Martin MJ, Orchard S, Magrane M, Agivetova R, Ahmad S, et al. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.* 2021; 49(D1): D480-9.
26. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat T, Weissig H, et al. The protein data bank. *Nucleic Acids Res.* 2000; 28(1): 235-42.
27. Notredame C, Higgins DG, Heringa J. T-coffee: a novel method for fast and accurate multiple sequence alignment. *J Mol Biol.* 2000; 302(1): 205-17.
28. Šali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol.* 1993; 234(3): 779-815.
29. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr.* 1993; 26(2): 283-91.
30. Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, Thornton JM. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J Biomol NMR.* 1996; 8(4): 477-86.
31. Bowie JU, Lüthy R, Eisenberg D. A method to identify protein sequences that fold into a known three-dimensional structure. *Science.* 1991; 253(5016): 164-70.
32. Lüthy R, Bowie JU, Eisenberg D. Assessment of protein models with three-dimensional profiles. *Nature.* 1992; 356(6364): 83-5.
33. Pontius J, Richelle J, Wodak SJ. Deviations from standard atomic volumes as a quality measure for protein crystal structures. *J Mol Biol.* 1996; 264(1): 121-36.
34. OMEGA 3.0.0.1: OpenEye Scientific Software, Santa Fe, NM. 2018. Available from: <http://www.eyesopen.com>.
35. 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.* 2010; 50(4): 572-84.
36. Jakalian A, Jack DB, Bayly CI. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J Comput Chem.* 2002; 23(16): 1623-41.
37. QUACPAC 1.7.0.2: OpenEye Scientific Software, Santa Fe, NM. 2018. Available from: <http://www.eyesopen.com>.
38. Coates L, Erskine PT, Mall S, Gill R, Wood SP, Myles DAA, et al. X-ray, neutron and NMR studies of the catalytic mechanism of aspartic proteinases. *Eur Biophys J.* 2006; 35(7): 559-66.
39. Coates L, Erskine PT, Wood SP, Myles DAA, Cooper JB. A neutron laue diffraction study of endothiapepsin: implications for the aspartic proteinase mechanism. *Biochemistry.* 2001; 40(44): 13149-57.
40. Coates L, Erskine PT, Crump MP, Wood SP, Cooper JB. Five atomic resolution structures of endothiapepsin inhibitor complexes: implications for the aspartic proteinase mechanism. *J Mol Biol.* 2002; 318(5): 1405-15.
41. Coates L, Tuan HF, Tomanicek S, Kovalevsky A, Mustyakimov M, Erskine P, et al. The catalytic mechanism of an aspartic proteinase explored with neutron and X-ray diffraction. *J Am Chem Soc.* 2008; 130(23): 7235-7.
42. Veerapandian B, Cooper JB, Šali A, Blundell TL, Rosati RL, Dominy BW, et al. Direct observation by X-ray analysis of the tetrahedral "intermediate" of aspartic proteinases. *Protein Science.* 1992; 1(3): 322-8.
43. OEDOCKING 3.2.0.2: OpenEye Scientific Software, Santa Fe, NM. 2018. Available from: <http://www.eyesopen.com>.
44. McGann M. FRED pose prediction and virtual screening accuracy. *J Chem Inf Model.* 2011; 51(3): 578-96.
45. McGann M. FRED and HYBRID docking performance on standardized datasets. *J Comput Aided Mol Des.* 2012; 26(8): 897-906.
46. McGann MR, Almond HR, Nicholls A, Grant JA, Brown FK. Gaussian docking functions. *Biopolymers.* 2003; 68(1): 76-90.
47. Braga RC, Alves VM, Silva MFB, Muratov E, Fourches D, Lião LM, et al. Pred-HERG: a novel web-accessible computational tool for predicting cardiac toxicity. *Mol Inform.* 2015; 34(10): 698-701.
48. Braga RC, Alves VM, Silva MFB, Muratov E, Fourches D, Tropsha A, et al. Tuning HERG out: antitarget QSAR models for drug development. *Curr Top Med Chem.* 2014; 14(11): 1399-415.
49. Neves BJ, Dantas RF, Senger MR, Valente WCG, Rezende-Neto JDM, Chaves WT, et al. The antidepressant drug paroxetine as a new lead candidate in schistosome drug discovery. *MedChemComm.* 2016; 7(6): 1176-82.
50. Neves BJ, Dantas RF, Senger MR, Melo-Filho CC, Valente WCG, De Almeida ACM, et al. Discovery of new anti-schistosomal hits by integration of QSAR-based virtual screening and high content screening. *J Med Chem.* 2016; 59(15): 7075-88.
51. Lamprecht MR, Sabatini DM, Carpenter AE. CellProfiler™: free, versatile software for automated biological image analysis. *Bio-techniques.* 2007; 42(1): 71-5.
52. Baldwin ET, Bhat TN, Gulnik S, Hosur M V, Sowder RC, Cachau RE, et al. Crystal structures of native and inhibited forms of human cathepsin D: implications for lysosomal targeting and drug design. *Proc Natl Acad Sci.* 1993; 90(14): 6796-800.
53. Schechter I, Berger A. On the size of the active site in proteases. I. Papain. *Biochem Biophys Res Commun.* 1967; 27(2): 157-62.
54. Abadzapatero C, Metz J. Ligand efficiency indices as guideposts for drug discovery. *Drug Discov Today.* 2005; 10(7): 464-9.
55. Schultes S, de Graaf C, Haaksma EEJ, de Esch IJP, Leurs R, Krämer O. Ligand efficiency as a guide in fragment hit selection and optimization. *Drug Discov Today Technol.* 2010; 7(3): e157-62.
56. Fujinaga M, Cherney MM, Tarasova NI, Bartlett PA, Hanson JE, James MNG. Structural study of the complex between human pepsin and a phosphorus-containing peptidic transition-state analog. *Acta Crystallogr D Biol Crystallogr.* 2000; 56(3): 272-9.
57. Gustchina A, Weber IT. Comparison of inhibitor binding in HIV-1 protease and in non-viral aspartic proteases: the role of the flap. *FEBS Lett.* 1990; 269(1): 269-72.
58. Hong L, Tang J. Flap position of free memapsin 2 (beta-secretase), a model for flap opening in aspartic protease catalysis. *Biochemistry.* 2004; 43(16): 4689-95.