Available online at www.sciencedirect.com
ScienceDirect
Bioorganic & Medicinal Chemistry Letters 18 (2008) 350–354
Bioorganic & Medicinal Chemistry Letters

Uncovering false positives on a virtual screening search for cruzain inhibitors

Alberto Malvezzi,[a] Leandro de Rezende,[a] Mario Augusto Izidoro,[b] Maria Helena Sedenho Cezari,[b] Luiz Juliano[b] and Antonia T.-do Amaral[a,*]

[a]*Departamento de Química Fundamental, Instituto de Química, Universidade de São Paulo, C.P. 26077, 05513-970 São Paulo, SP, Brazil*
[b]*Departamento de Biofísica, Escola Paulista de Medicina, Universidade Federal de São Paulo, Rua Três de Maio, 100, 04044020 São Paulo, SP, Brazil*

Received 6 September 2007; revised 16 October 2007; accepted 19 October 2007
Available online 24 October 2007

Abstract—Some unexpected promiscuous inhibitors were observed in a virtual screening protocol applied to select cruzain inhibitors from the ZINC database. Physical–chemical and pharmacophore model filters were used to reduce the database size. The selected compounds were docked into the cruzain active site. Six hit compounds were tested as inhibitors. Although the compounds were designed to be nucleophilically attacked by the catalytic cysteine of cruzain, three of them showed typical promiscuous behavior, revealing that false positives are a prevalent concern in VS programs.

Identification and avoidance of potential false positive results, such as screening hits that are unsuitable as lead structures, have been recognized as a big concern in every virtual screening (VS) or high throughput screening (HTS) program.[1,2] These undesired results could be due to many reasons,[3] being one of them the presence of promiscuous inhibitors among the tested compounds.[4] Organic molecules able to form micromolar or submicromolar aggregates in aqueous buffers are classified as promiscuous inhibitors[5] due to the capacity of these aggregates to inhibit nonspecifically enzymes at in vitro assays. For this reason, these molecules are rarely suitable for drug development and their early detection can avoid worthless work.[2]

Promiscuous inhibitor prediction studies, based on chemical properties, have been carried out with some degree of success.[6,7] Despite these early positive results, the widespread application validity of these empirical models is still being discussed considering that they have been applied only to a limited size test series, and also they seem to be still unreliable for filtering databases with up to a million compounds. Additionally, it has been stressed that more studies should address the issue of false positives and the appearance of promiscuous inhibitors in HTS/VS programs.[2,3,6,7] Therefore, a demand for higher quality data to feed knowledge-based tools which will, one day, be able to reliably identify these undesirable compounds is quite clear.

It has been observed, in this attempt to find specific inhibitors for cruzain (the recombinant form of cruzipaine, a major *Trypanosoma cruzi* protease present in every stage of the parasite life cycle[8]), that some compounds showed typical promiscuous properties, such as poor specificity, micromolar activity, and no structure activity relationship. The procedures applied to select these potential cruzain inhibitors and the methods used for identification and confirmation of promiscuous inhibitory activity have been described in the present work.

Cruzain crystal structures have been elucidated in at least 13 PDB entries (1AIM, 1EWL, 1EWM, 1EWO, 1EWP, 1F29, 1F2A, 1F2B, 1F2C, 1ME3, 1ME4, 1U9Q, and 2AIM, whose resolutions range from 1.2 Å to 2.2 Å). All of them were covalently bonded to a peptide-like inhibitor. Cruzain belongs to the papaine-like family of cysteine proteases and binds peptide substrates

Keywords: Virtual screening; Promiscuous mechanism; Cruzain inhibitors.
* Corresponding author. Tel.: +55 11 30913846; fax: +55 11 38155579; e-mail: atdamara@iq.usp.br

0960-894X/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmcl.2007.10.068

Table 1. Physical–chemical properties applied as filters

Property	Values
Atoms types	H; C; N; O; F; Cl; Br; I; P; S
Molecular weight	250–600
Partition coefficient (log *P*)	−2 to 5
Hydrogen bond donors	2–5
Hydrogen bond acceptors	2–10
Sum of N + O	3–8
Number of halogens	0–2
Number of rings	0–4
Max ring members	3–8
Rotatable bonds	0–6
$CH_3(CH_2)_{n-}$ chain length	0–2

in its active site cleft through backbone and side chain interactions, which lie around several defined pockets in the enzyme. Using the structural information available from the 13 cruzain complex PDB entries, strategies including structure–ligand-based design procedures have been applied to filter the ZINC[9] compound database and to select candidates to be submitted to in vitro enzymatic assay.

Initially, the ZINC compound database (3,294,741 compounds) has been searched by means of DBFilter[10] program and the application of a set of physical–chemical properties filters, in order to select drug-like compounds. Table 1 presents the physical–chemical properties (with their corresponding range values) used to filter the database. These physical–chemical filters and their corresponding values have been derived from the Lipinski's 'rule of five',[11] which has been modified considering the known cruzain inhibitors and their active site structural information. The size of the ZINC dataset has been reduced to 296,700 compounds, approximately 10% of the initial value, through these filters' application.

The resulting filtered compounds were then searched with a 3D pharmacophore model (using the Catalyst[12] program), built from the 13 inhibitor structures extracted from the cruzain complexes. The resulting pharmacophore (Fig. 1) model comprised four different chemical features: hydrophobic aromatic, H-bond acceptor, H-bond donor, and the electrophilic. The hydrophobic aromatic feature corresponds to the occupation of the S2 pocket by a hydrophobic group. The H-bond donor feature corresponds to the hydrogen bonding with cruzain's backbone oxygen of Asp158, and the H-bond acceptor feature corresponds to a hydrogen bond with the enzyme backbone –NH of Gly66. The last feature corresponds to the presence of an electrophilic atom of a chemical group (Fig. 2) susceptible to a nucleophilic attack by the activated catalytic cysteine. The application of the 3D pharmacophore filter excluded approximately 99.9% of the previously selected compounds, being only 308 compounds left.

The remaining compounds were then docked into the cruzain binding cavity by using the Gold[13] program. In this procedure, each compound was docked 10 times

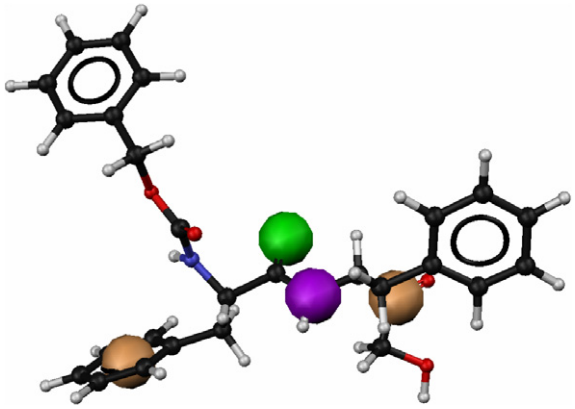


Figure 1. Aligned cruzain inhibitor (PDB ID 1ME4) superimposed over the generated pharmacophore model (magenta sphere: H-bond donor; green sphere: H-bond acceptor; left beige sphere: hydrophobic aromatic, and right beige sphere: electrophilic group).

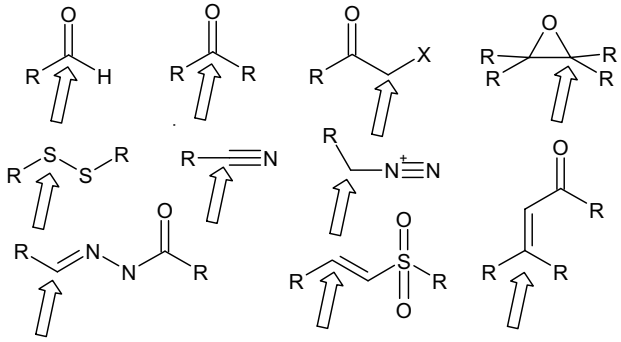
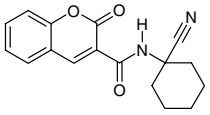
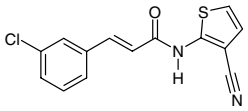
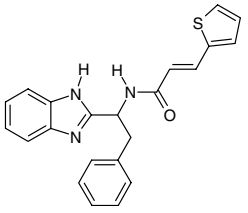
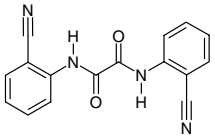
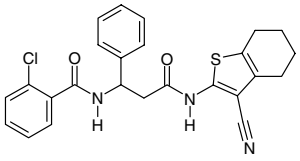
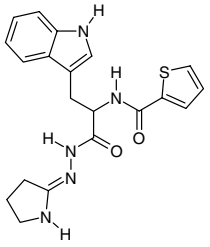


Figure 2. Chemical groups considered in the electrophilic feature engineering. Each arrow points to the corresponding electrophilic atom subject to a nucleophilic attack by the activated catalytic cysteine (X = F, Cl, Br, I).

over the cruzain binding site structure, extracted from the 1ME4 complex (resolution 1.2 Å). After the docking runs, the binding fitness values were estimated by the scoring function and the top 50 scored compounds were visually inspected in their docked position inside the active site. In order to select potential hits, through visual inspection the following criteria were considered: (1) degree of occupancy of S2 and S1 enzyme pockets, in particular related to the achieved protein–ligand surface complementarity; (2) distance and orientation of the electrophilic group in relation to the activated catalytic cysteine; (3) formation of hydrogen bonds with Gly66 and with Asp158, and (4) quality of the overall binding conformation, from which six compounds have been selected. The structures of the selected compounds are shown in Table 2, together with their corresponding molecular weight, calculated log *P*[14] values, and the vending company from which the compounds had been purchased. The selected compounds were then tested as cruzain inhibitors. All assays were performed at 37 °C, pH 6.5 (using 1 mL of disodium phosphate (50 mM) and EDTA (2.5 mM) buffer). Cruzain[15] (0.18 nM) and DTT (500 mM) were pre-incubated for 5 min. The reaction was initiated with the addition of the substrate *Bz-FR-MCA* (845 μM). Finally, increasing aliquots out of each of the six inhibitors were added until either total

Table 2. Hit compounds selected by virtual screening and their corresponding molecular weight, calculated log *P* values, and the vending company

Number	Compound	M_w	log P[a]	Vendor
1		296.3	2.9	Specs
2		288.7	3.4	Life chemicals
3		373.5	3.7	Enamine Ltd
4		290.3	1.3	Asinex Ltd
5		463.99	4.6	Enamine Ltd
6		411.5	2.5	Enamine Ltd

[a] Values calculated using the XLOGP[14] program.

inhibition of the enzyme or lack of solubility of the inhibitor was observed. Stock solutions of each inhibitor were prepared in DMSO. The effect of DMSO on the results was controlled and, in all assays, the maximum concentration of DMSO was less than 10% (v/v). All reactions were recorded on an Hitachi 4500 spectrophotometer (λ_{ex} = 380 nm and λ_{em} = 460 nm). As suggested by the literature,[4] the promiscuous mechanism was analyzed through the observation of the inhibitor potency after the addition of 0.1% of Triton X-100 as well as after a 10-fold increase in the enzyme concentration (from 0.18 nM to 1.80 nM) using the same experimental conditions as described before. It is assumed that the promiscuous mechanism occurs, if a compound loses its potency after the addition of Triton X-100 and after the increase of the enzyme concentration, respectively.[4]

Three out of the six tested compounds, **1**, **2**, and **6**, showed no cruzain inhibition activity up to the tested concentrations of 7.22 mM, 38.71 mM, and 0.67 mM, respectively. At higher concentrations, compounds were insoluble. The other three, compounds **3**, **4**, and **5**, showed some degree of enzyme inhibition at micro molar concentrations, respectively at 44.91 μM, 17.26 μM, and at 4.17 μM. However, after the addition of Triton (0.1%, v/v), the inhibition activity was lost and the enzyme returned to its previously uninhibited value. Figure 3 presents the inhibition assay plots showing substrate consumption by cruzain in the presence of each of the six tested compounds and the effect of the addition of Triton X-100 on the assays where inhibitory activity was observed.

Confirmation of promiscuous mechanism of compounds **3**, **4**, and **5** was verified by analyzing the effect of enzyme concentration (increased by 10-fold) on the cruzain inhibition potency for each compound, due to the fact that, a well behaved μM competitive inhibitor would not be

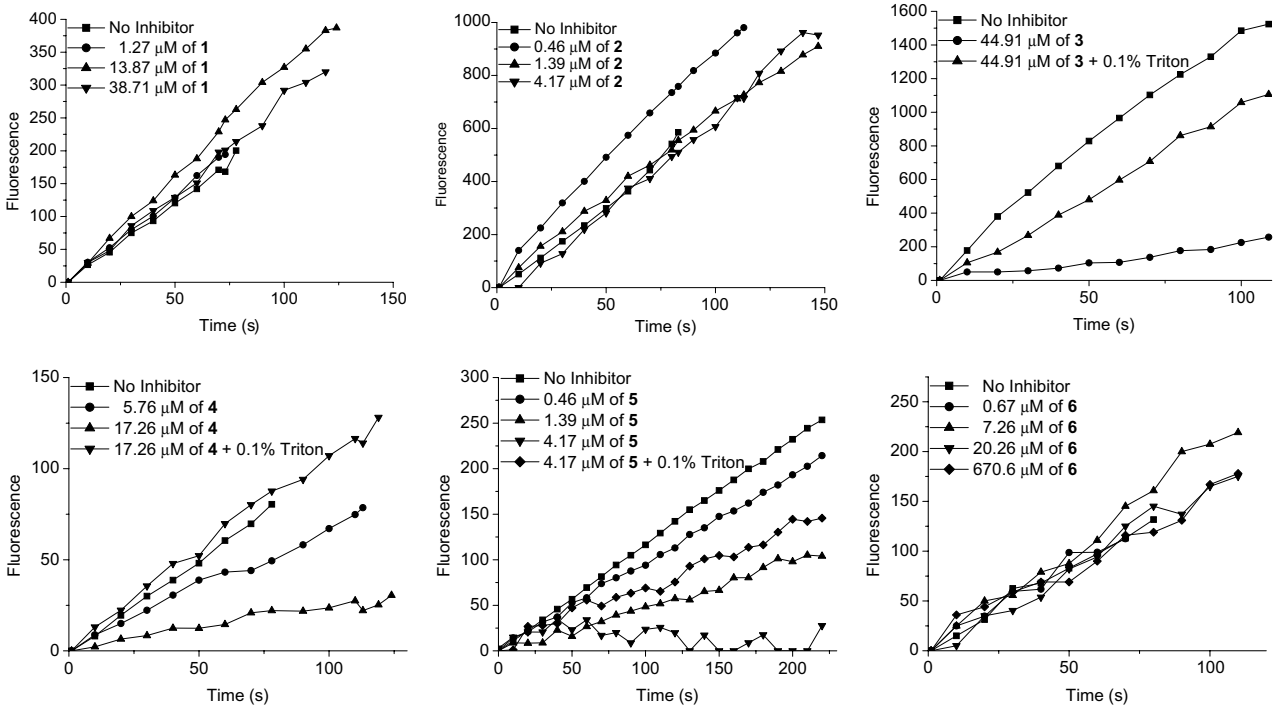


Figure 3. Inhibition assays. Plots showing substrate consumption (fluorescence units per second) by cruzain in the presence of each of the six tested compounds (using at least two different concentrations) and the effect of the addition of (0.1%) Triton X-100 in the assays, where inhibitory activity has been observed (compounds **3**, **4**, and **5**, respectively).

significantly affected by a nM change in the concentration of the enzyme, but instead an aggregating promiscuous inhibitor would show a dramatic decrease in potency.[4] It has been observed, in the present work, that all the three compounds were not only sensitive to the enzyme concentration but they also lost their activity when cruzain concentration was increased by 10-fold. Figure 4 presents the inhibition assay plots showing

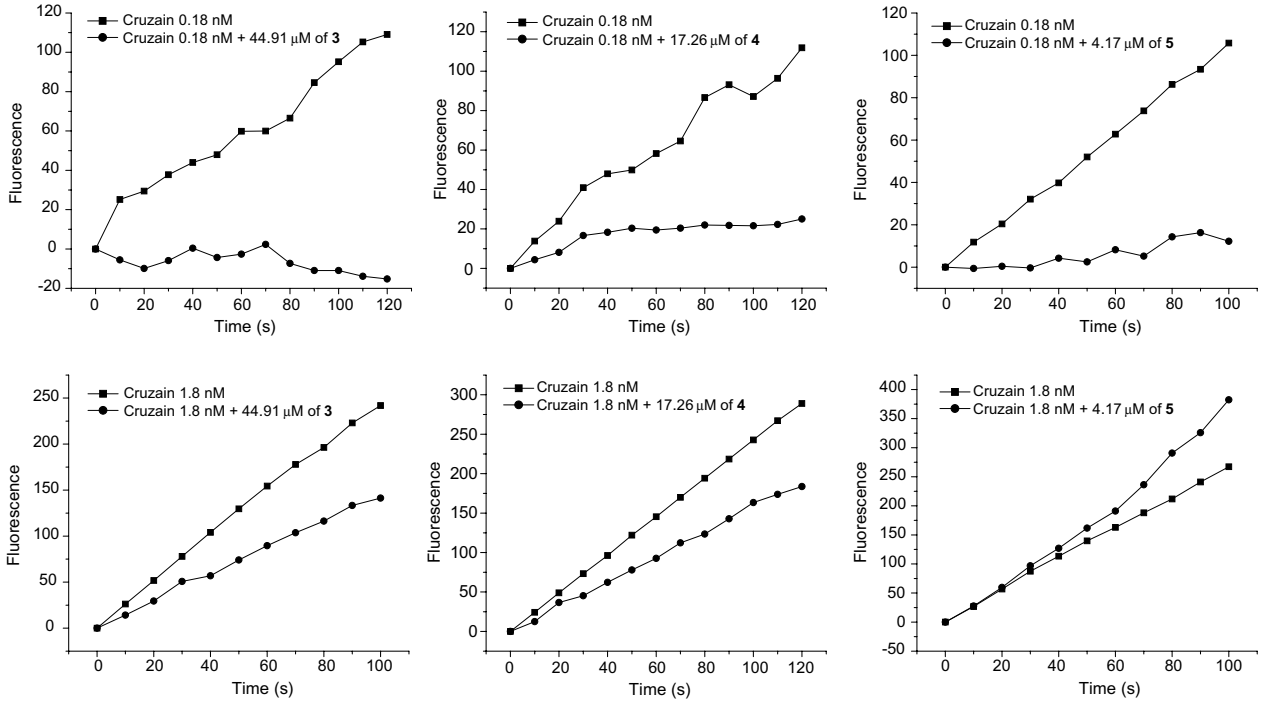


Figure 4. Inhibition assay plots of substrate consumption (fluorescence units per second) by cruzain in the presence of compounds **3**, **4**, and **5**. Enzyme concentration values were increased 10-fold, from 0.18 nM (top rows) to 1.8 nM (bottom rows), to confirm the promiscuous mechanism of compounds **3**, **4**, and **5**, respectively.

substrate consumption by cruzain in the presence of compounds **3**, **4**, and **5** at two different enzyme concentrations (0.18 nM and 1.8 nM). The plots show that, in every case, the selected inhibitors were less potent at higher enzyme concentrations.

In conclusion, although the six compounds have been engineered to be nucleophilically attacked by the catalytic cysteine of cruzain through the inclusion of chemically reactive groups (ex. nitriles, semicarbazones, and unsaturated ketones) into the pharmacophoric features used to select compounds, three of them were, in fact, promiscuous inhibitors, regarding cruzain. This high proportion (50%) of promiscuous acting compounds shows that this kind of artifact cannot only be prevalent at in vitro assays but also present a real concern in both HTS and VS programs.

Acknowledgments

The present study was financially supported by FAPESP (02/09518-8, 01/01192-3, 04/08928-3) PRP-USP Procontes and CAPES PROBRAL (198/05). The authors acknowledge the valuable collaboration of Prof. Hugo Kubinyi on this project.

References and notes

1. Rishton, G. M. *Drug Discovery Today* **1997**, *2*, 382.
2. Shoichet, B. K. *Drug Discovery Today* **2006**, *11*, 607.
3. Rishton, G. M. *Drug Discovery Today* **2003**, *8*, 86.
4. McGovern, S. L.; Caselli, E.; Grigorieff, N.; Shoichet, B. K. *J. Med. Chem.* **2002**, *45*, 1712.
5. McGovern, S. L.; Helfand, B. T.; Feng, B.; Shoichet, B. K. *J. Med. Chem.* **2003**, *46*, 4265.
6. Seidler, J.; McGovern, S. L.; Doman, T. N.; Shoichet, B. K. *J. Med. Chem.* **2003**, *46*, 4477.
7. Feng, B. Y.; Shelat, A.; Doman, T. N.; Guy, R. K.; Shoichet, B. K. *Nat. Chem. Biol.* **2005**, *1*, 146.
8. Chagas, J. R.; Authie, E.; Serveau, C.; Lalmanach, G.; Juliano, L.; Gauthier, F. *Mol. Biochem. Parasit.* **1997**, *88*, 85.
9. Irwin, J. J.; Shoichet, B. K. *J. Chem. Inf. Model.* **2005**, *45*, 177.
10. Wang, S.-H. Personal communication, 2003.
11. Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. *Adv. Drug Deliv. Rev.* **1997**, *23*, 3.
12. CATALYST version 4.8 software; ACCELRYS Inc., 2001.
13. Verdonk, M. L.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Taylor, R. D. *Proteins* **2003**, *52*, 609.
14. Wang, R.; Fu, Y.; Lai, L. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 615.
15. Judice, W. A. S.; Cezari, M. H. S.; Lima, A. P. C. A.; Scharfstein, J.; Chagas, J. R.; Tersariol, I. L. S.; Juliano, M. A.; Juliano, L. *Eur. J. Biochem.* **2001**, *268*, 6578.