Journal of Molecular Graphics and Modelling 28 (2009) 3–11

Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

journal homepage: www.elsevier.com/locate/JMGM

Quantitative structure–activity relationships for a series of inhibitors of cruzain from *Trypanosoma cruzi*: Molecular modeling, CoMFA and CoMSIA studies

Gustavo H.G. Trossini [a], Rafael V.C. Guido [b], Glaucius Oliva [b], Elizabeth I. Ferreira [a], Adriano D. Andricopulo [b,*]

[a] *Laboratório de Planejamento e Síntese de Quimioterápicos Potenciais Contra Endemias Tropicais, Faculdade de Ciências Farmacêuticas, Universidade de São Paulo, Av. Professor Lineu Prestes 580, 05508-900, São Paulo, SP, Brazil*
[b] *Laboratório de Química Medicinal e Computacional, Centro de Biotecnologia Molecular Estrutural, Instituto de Física de São Carlos, Universidade de São Paulo, Av. Trabalhador São-carlense 400, 13560-970, São Carlos, SP, Brazil*

ARTICLE INFO

Article history:
Received 4 December 2008
Received in revised form 27 February 2009
Accepted 2 March 2009
Available online 14 March 2009

Keywords:
Cruzain
Drug design
Chagas' disease
CoMFA
CoMSIA

ABSTRACT

Human parasitic diseases are the foremost threat to human health and welfare around the world. Trypanosomiasis is a very serious infectious disease against which the currently available drugs are limited and not effective. Therefore, there is an urgent need for new chemotherapeutic agents. One attractive drug target is the major cysteine protease from *Trypanosoma cruzi*, cruzain. In the present work, comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA) studies were conducted on a series of thiosemicarbazone and semicarbazone derivatives as inhibitors of cruzain. Molecular modeling studies were performed in order to identify the preferred binding mode of the inhibitors into the enzyme active site, and to generate structural alignments for the three-dimensional quantitative structure–activity relationship (3D QSAR) investigations. Statistically significant models were obtained (CoMFA, $r^2 = 0.96$ and $q^2 = 0.78$; CoMSIA, $r^2 = 0.91$ and $q^2 = 0.73$), indicating their predictive ability for untested compounds. The models were externally validated employing a test set, and the predicted values were in good agreement with the experimental results. The final QSAR models and the information gathered from the 3D CoMFA and CoMSIA contour maps provided important insights into the chemical and structural basis involved in the molecular recognition process of this family of cruzain inhibitors, and should be useful for the design of new structurally related analogs with improved potency.

1. Introduction

Parasitic diseases are a major global cause of illness, long-term disability, and death, with severe socio-economic consequences for millions of people worldwide. According to the World Health Organization, Chagas' disease is a serious and life threatening disease, affecting 10–14 million people [1]. This problem is aggravated by the rate at which populations are growing, especially in the developing world, where there is a lack of adequate sanitation facilities and poor hygiene practices which present breeding grounds for disease, illness and suffering [2–4]. In spite of the alarming health, economic and social consequences of these parasitic infections, the limited existing drug therapy (the nitroheterocyclic compounds nifurtimox and benznidazole) suffers from a combination of drawbacks including poor efficacy, resistance and serious side effects. Therefore, there is an urgent need for new drugs that can overcome resistance and are safe and effective for use in human [5–8].

Cysteine proteases are widespread in nature [9]. Their implication in numerous vital processes of several parasitic protozoa makes them highly attractive targets for drug design [10–12]. Cruzain (EC 3.4.22), the major cysteine protease of the parasite *Trypanosoma cruzi*, plays a pivotal role during the infection of host cells, replication, and metabolism [13], and has been considered an important target for the development of new antitrypanosomal agents [14–17].

Structure- and ligand-based drug design approaches have become fundamental components of modern drug discovery [18–20]. Quantitative structure–activity relationship (QSAR) methods have been successfully employed to assist the design of new small molecule drug candidates, ranging from enzyme inhibitors to receptor ligands [21–25]. In the present work, molecular docking, comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA) have been performed to investigate a series of chemically diverse thiosemicarbazone and semicarbazone derivatives as inhibitors of cruzain

* Corresponding author. Tel.: +55 16 3373 8095; fax: +55 16 3373 9881.
E-mail address: aandrico@if.sc.usp.br (A.D. Andricopulo).

1093-3263/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2009.03.001

from *T. cruzi*. The predictive 3D QSAR models along with the information gathered from 3D contour maps provided important insights into the structural and chemical basis for potent cruzain inhibition within this series of (thio)semicarbazone derivatives. The identification of key intermolecular features associated potency and selectivity should be a valuable tool for the design of new inhibitors within this structural diversity with enhanced properties.

2. Materials and methods

2.1. Data set

The data set used for the QSAR studies contains 55 thiosemicarbazone and semicarbazone derivatives as inhibitors of the *T. cruzi* cysteine protease cruzain [26]. Representative examples of the chemical structures and corresponding pIC_{50} values ($-\log IC_{50}$, where IC_{50} is the concentration of compound required for 50% inhibition of cruzain activity) employed in this work are shown in Table 1. The complete version of Table 1, including the chemical structures and corresponding biological data, is provided as supplementary information, as well as the chemical structures in SMILES format for the complete data set.

2.2. Computational approach

The QSAR modeling analyses, calculations, and visualizations for CoMFA and CoMSIA were performed using the SYBYL 8.0 package (Tripos Inc., St. Louis, MO, USA) running on Red Hat Enterprise Linux workstations. The 3D structures of the cruzain inhibitors were constructed using standard geometric parameters of the molecular modeling software package SYBYL 8.0. Each single optimized conformation of each molecule in the data set was energetically minimized employing the Tripos force field [27] and the Powell conjugate gradient algorithm [28] with a convergence criterion of 0.05 kcal/mol Å and Gasteiger–Hückel charges [29]. A statistical cluster analysis was carried out with Tsar 3D version 3.3 (Accelrys, San Diego, USA) using the complete linkage clustering method (Euclidean distances) with no data standardization.

2.3. Molecular modeling and structural alignment

Molecular docking and scoring protocols as implemented in FlexX (BioSolveIT GmbH, Sankt Augustin, Germany) were used to investigate the possible binding conformations of the ligands within the cruzain binding pocket. The X-ray crystallographic data for cruzain in complex with the peptide-like inhibitor 1-(1-benzyl-3-hydroxy-2-oxo-propylcarbamoyl)-2-phenyl-ethyl-carbamic acid benzyl ester determined at 1.2 Å (PDB ID 1ME4) used in the docking simulations were retrieved from the Protein Data Bank (PDB) [30]. The ligand and water molecules were removed from the binding pocket and hydrogen atoms were added in standard geometry using the Biopolymer module implemented in SYBYL 8.0. Histidines, glutamines, and asparagines residues within the binding site were manually checked for possible flipped orientation, protonation, and tautomeric states with Pymol 0.99 (DeLano Scientific, San Carlos, USA) side-chain wizard script. The binding site was defined as all the amino acid residues encompassed within an 8.0-Å radius sphere centered on the bound ligand.

The structural alignment protocol was applied on the basis of the inhibitors docked conformation within the cruzain binding site generated by FlexX with default parameters [31]. The docking procedures were repeated 30 times for each data set compound. The implemented FlexX scoring function was employed to select the representative conformation for each data set compound. Subsequently, each of the 30 conformations of each compound was submitted to the DrugScore^{ONLINE} website to rescore the proposed binding modes [32,33]. Finally, the individual ranks obtained from the scoring functions were added to give a rank order list, and then only the top-ranked poses were employed to produce the structural alignment for the 3D QSAR studies. The aligned data set is depicted in Fig. 2.

2.4. 3D QSAR: CoMFA and CoMSIA models

CoMFA [34,35] and CoMSIA [36,37] analyses were carried out on the basis of the interaction energies between a suitable probe atom and the aligned ligand atoms calculated within a cubic lattice of 2 Å spacing. The grid box embedded all ligands with a margin of at least 4 Å in each direction. A positively charged sp^3 carbon atom was used as the probe atom for calculating steric and electrostatic CoMFA fields applying SYBYL standard parameters (TRIPOS standard field, dielectric constant $1/r$, cutoff 30 kcal/mol). CoMSIA fields were computed for steric, electrostatic and hydrophobic properties, using a probe of charge +1, radius of 1, hydrophobicity of +1, and an attenuation factor of 0.3 for the Gaussian distance-dependent function.

Regression analysis was carried out by SAMPLS [38] using the partial least-squares (PLS) to derive a linear correlation between the CoMFA or CoMSIA fields and the biological property of the cruzain inhibitors. Leave-one-out (LOO) cross-validation analysis was applied to determine the number of components that yield optimally predictive models. Minimum-sigma (column filtering) was set at 2.0 kcal/mol to improve the signal-to-noise ratio by omitting those lattice points where energy variation is below the threshold. Subsequently, a PLS analysis was performed without cross-validation, applying no column filtering. Leave-many-out (10 groups, LMO_{10}; and five groups, LMO_5) procedures were used as a more rigorous test to assess model stability and statistical significance. Each random cross-validation run was repeated 25 times to obtain mean values for q^2 and the corresponding standard deviation of error of prediction (SDEP) values. The region focusing method was applied to increase the resolution of the CoMFA and CoMSIA models. Progressive scrambling method was applied to determine the sensitivity of the QSAR models to chance correlations.

The derived 3D QSAR models were externally validated with a test set of compounds, which were not considered for QSAR model generation. After generation of the PLS training set models, the dependent variables (pIC_{50}) were predicted for the test set compounds, allowing predictive r^2 values to be determined for individual 3D QSAR models [39,40]. The r^2_{pred} is a procedure analogous to the cross-validated q^2, and is calculated by using the formula:

$$r^2_{pred} = \frac{SD - PRESS}{SD} \quad (1)$$

where SD is the sum of squared deviation between the biological activities of the test set molecule and the mean activity of the training set molecules and PRESS is the sum of squared deviations between the observed and the predicted activities of the test molecules. Finally, CoMFA and CoMSIA contour maps were generated by interpolation of the pairwise products between the PLS coefficients and the standard deviations of the corresponding descriptors values.

3. Results and discussion

3.1. Chemical and biological data

The quality of the biological data under investigation as well as the structural diversity of the data set is important foundations for