Profiling Cruzain Inhibitors to Improve Leadlikeness. Database, Analyses and Guidelines

Marcelo T. de Oliveira[*1,2] (PG), Adriano D. Andricopulo[1] (PQ).

mtavareso@usp.br

1 – Laboratório de Química Medicinal e Computacional, Centro de Pesquisa e Inovação em Biodiversidade e Fármacos, Instituto de Física de São Carlos, Universidade de São Paulo; 2 – Instituto de Química de São Carlos, Universidade de São Paulo.

keywords: chagas disease, cruzain, drug optimisation, data mining, cheminformatics.

Abstract

Analyses of a comprehensive collection of cruzain inhibitors reveal patterns and relationships of interest for drug design.

Introduction

Chagas' remains in the list of neglected diseases. Once a major health concern restricted to some Latin American countries, numbers of infected patients have grown elsewhere as far as Japan and Australia due to greater human mobility. In this time and age of resurgence of virtually extinct maladies, fears of Chagas disease spiralling out of control have increased building pressure for the development of new therapies.[1] Currently, there are only a couple of drugs available in the market, benznidazole and nufurtimox, which have considerable drawbacks.

Therefore, there is massive scope for contributions towards the development of new drugs for Chagas'. To this end, cruzain[3], the most abundant cysteine protease in *T. cruzi*, has been a major molecular target in drug discovery campaigns.

Building on the recent successful experience of developing the first database for natural products in the country (NuBBE)[2], we set out to chart the chemical and biological space of cruzain inhibitors by mining data from the literature to enable us to establish structure-activity relationships. The results disclosed herein are part of our ongoing efforts towards Chagas disease "big data".

Results and Discussion

A data set comprising of. 500 cruzain inhibitors was assembled from the literature. Molecular structures were acquired and biological activity entered as pIC_{50}. As in Lipinski's rule of 5 and similar approaches, straightforward molecular properties and descriptors were computed. In so doing, correlations between chemical information available in the structure and biological data as the inhibitory concentration were established.

Bearing in mind that lead identification is a critical step in the early stages of drug discovery, we turned our attention to scrutinise the hyperspace of highly active compounds. For this exercise, we considered compounds with $pIC_{50} \geq 7$ as leads. Analyses of the computed properties/descriptors against biological activity provided various correlations. For instance, polar surface area (PSA), a measure of the surface of all polar atoms in the molecule, showed a clear correlation as all compounds showing $pIC_{50} \geq 7$ have PSA ≥ 70 Å^2 (Fig.1A). Regarding partition coefficient, about 90% of "leadlike" compounds have Alog*P* ≥ 3 (Fig.1B).

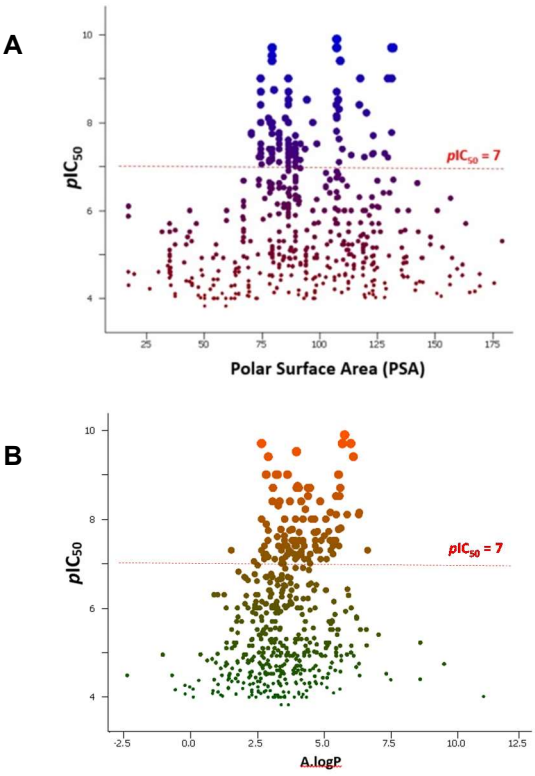

Figure 1. Molecular descriptors plotted against pIC_{50} for a set of cruzain inhibitors. (A) Polar surface area (PSA). (B) Alog*P*. In both cases, a red line indicates pIC_{50} 7 cutt-off.

All in all, the group of findings in the present work can possibly serve as guidelines for the design of cruzain inhibitors.

Acknowledgements

The authors acknowledge CAPES, CNPq and FAPESP.

[1] Maguire, J. H. *New Eng. J. Med.* **2015**, *373*, 1369.
[2] Valli, M.; dos Santos, R. N.; Figueira, L. D.; Nakajima, C. H.; Castro-Gamboa, I.; Andricopulo, A. D.; Bolzani, V. S. *N. Prod. Rep.* **2013**, *76*, 439.
[3] Martinez-Mayorga, K.; Byler, K. G.; Ramirez-Hernandez, A. I.; Terrazas-Alvares, D. E. *Drug Discov. Today* **2015**, *20*, 890-898.