BMC Bioinformatics

　　　　　　　　　　　　　　Open Access

Designing and implementing chemoinformatic approaches in TDR Targets Database: linking genes to chemical compounds in tropical disease causing pathogens

María Paula Magariños[1*], John Overington[2], Santiago Carmona[1], Dhanasekaran Shanmugam[3], Maria Doyle[4], Stuart Ralph[4], Greg Crowther[5], Christiane Hertz-Fowler[6], Solomon Nwaka[7], Matt Berriman[6], David Roos[3], Wes Van Voorhis[5], Fernán Agüero[1]

Background

Information about chemical compounds and their activity against whole organisms or specific molecular targets is available from the literature or from specialized databases. However, there are few resources that effectively integrate such large chemical datasets with genome data and provide a mechanism to link active compounds to potential target genes. Here, we showcase the integration of chemoinformatic tools for querying chemical datasets and linking chemicals to genes in TDR Targets database (tdrtargets.org), a web accessible resource that integrates a wide range of functional genomic datasets from tropical disease pathogens and provides a ranking mechanism for identifying and prioritising novel therapeutic targets [1].

Figure 1
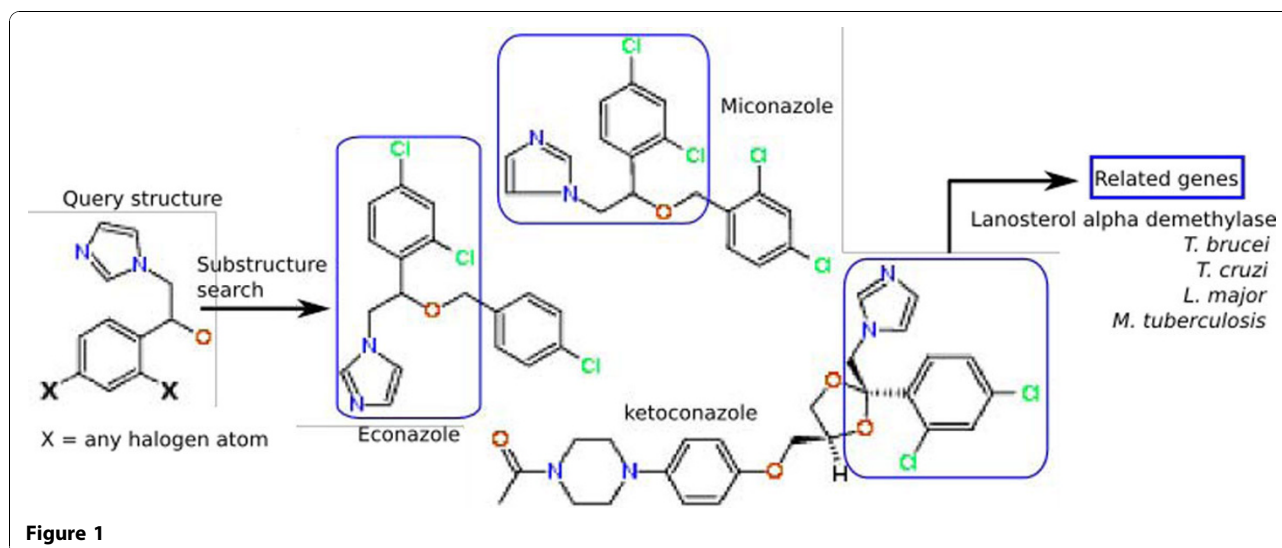
* Correspondence: mpmagarinos@iib.unsam.edu.ar
[1]Instituto de Investigaciones Biotecnológicas, Universidad de San Martín, San Martín, Argentina
Full list of author information is available at the end of the article

Materials and methods

Chemical datasets were obtained from three different resources: DrugBank, PubChem and StARlite (ChEMBL). A pipeline was developed to calculate a number of properties (molecular weight; number of flexible bonds; polar surface area; H bond donors/acceptors; and predicted octanol/water partition coefficient) and descriptors (InChi, IUPAC's standard and open chemical identifiers; SMILES; and molecular formula) for each molecule, to facilitate querying and linking to other databases. We have also calculated a number of binary fingerprints and molecular statistics to accelerate searches.

Results

A dataset of 504,020 chemicals, enriched in drugs and drug-like compounds, integrated into TDRTargets.org can be queried using: a textual search on molecular descriptors or chemical properties; a substructure search to find molecules containing the query structure; and a similarity search to find similar molecules (using Tanimoto distance) (see Figure 1). In the Starlite database 438,791 compounds are associated with 3,512 known druggable targets, and 2,224 of these could be linked to 3,043 pathogen targets based on sequence similarity. These relationships are available at TDRTargets.org.

Conclusions

A comprehensive collection of chemical data can be queried in various ways, including by chemical properties, structure and descriptors in TDRTargets.org. More importantly, one can also link compounds of interest to novel target genes in tropical disease causing parasitic organisms based on sequence similarity to known targets of these compounds.

Author details
[1]Instituto de Investigaciones Biotecnológicas, Universidad de San Martín, San Martín, Argentina. [2]European Bioinformatics: Institute, EBML Outstation, Hinxton, Cambridge, UK. [3]University of Pennsylvania, Philadelphia, PA, USA. [4]University of Melbourne, Victoria, Australia. [5]University of Washington, Seattle WA, USA. [6]Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK. [7]WHO/TDR, Geneva, Switzerland.

Published: 7 December 2010

Reference
1. Agüero F, *et al*: Genomic-scale prioritization of drug targets: the TDR Targets database. *Nat Rev Drug Discov* 2008, **7**(11):900-907.