BIOINFORMATICS APPLICATIONS NOTE

Vol. 24 no. 19 2008, pages 2270–2271
doi:10.1093/bioinformatics/btn415

Databases and ontologies

PK/DB: database for pharmacokinetic properties and predictive *in silico* ADME models

Tiago L. Moda, Leonardo G. Torres, Alexandre E. Carrara and Adriano D. Andricopulo*

Laboratory of Computational and Medicinal Chemistry, Center for Structural Molecular Biotechnology, Institute of Physics of São Carlos, University of São Paulo, São Carlos-SP, 13566-970, Brazil

Received on May 6, 2008; revised on August 1, 2008; accepted on August 3, 2008

Advance Access publication August 6, 2008

Associate Editor: Jonathan Wren

ABSTRACT

Summary: The study of pharmacokinetic properties (PK) is of great importance in drug discovery and development. In the present work, PK/DB (a new freely available database for PK) was designed with the aim of creating robust databases for pharmacokinetic studies and *in silico* absorption, distribution, metabolism and excretion (ADME) prediction. Comprehensive, web-based and easy to access, PK/DB manages 1203 compounds which represent 2973 pharmacokinetic measurements, including five models for *in silico* ADME prediction (human intestinal absorption, human oral bioavailability, plasma protein binding, blood–brain barrier and water solubility).

Availability: http://www.pkdb.ifsc.usp.br

Contact: aandrico@if.sc.usp.br

1 INTRODUCTION

The challenges facing the pharmaceutical industry are tremendous at every step of the drug discovery and development process. Technology-based discovery is certainly an important element to increase R&D productivity (Guido *et al.*, 2008; Jónsdóttir *et al.*, 2005). A drug intended for use in humans should have an ideal balance of efficacy and safety, as well as good pharmacokinetic properties (PK) (Moda *et al.*, 2007a). Problems with drug candidates' absorption, distribution, metabolism and excretion (ADME), however, have been identified as a major cause of drug candidate failure in late stages of the drug development process. Therefore, it is critical to accurately predict these qualities earlier in the investigation of lead candidates (Moda *et al.*, 2007a; Norinder and Bergström, 2006). Computational methods have emerged as a powerful strategy for the prediction of human PK. In this regard, a variety of useful *in silico* ADME models have been developed with different levels of complexity for the screening of large data sets of compounds, creating tools that are faster, simpler and more cost-effective than traditional experimental procedures (Canavan, 2007).

In an effort to make high quality pharmacokinetic data and predictive models available to a worldwide scientific community, PK/DB (a freely available database for PK) was designed by our research group incorporating high quality databases of structurally diverse drug-like and lead-like molecules for a variety of PK.

*To whom correspondence should be addressed.

2 DATABASE CONTENT

The chemical and pharmacokinetic data were collected both from public databases and from the literature (http://www.pkdb.ifsc.usp.br/pkdb/literature_src.php), resulting in a total of 1203 compounds with 2973 property values grouped and organized as shown in Table 1.

3 INTERFACE AND DATA MANAGEMENT

In PK/DB, a web-based query tool incorporating a molecular drawing interface enables the database to be searched by chemical structure or standard name, substructure or molecular fragment, molecular formula or by an exact or range of a specific pharmacokinetic property (Table 1). Also available for searching is the information on human CYP-mediated drug metabolism for a number of compounds. The user can also employ a combination of criteria as a useful way for database searching. To facilitate the analysis, the results of such searches are showed in two steps. In the first step, the user can choose the number of compounds at the top of the search session (e.g. 10, 25, 50, 100 or all) and sort it by molecular weight, compound name, HIA, F, PPB, BBB, Vd, Cl and T1/2. The column field presents the compounds, showing the PK/DB identification (MID), 2D structure, standard name, SMILES, molecular weight and a list of properties available in PK/DB, as shown in Figure 1. The second step allows access to more detailed information using some hyperlinks: 3D structures, PK and pharmacological action extracted from PubChem (http://pubchem.ncbi.nlm.nih.gov/).

Table 1. Number of measurements in PK/DB for a number of pharmacokinetic properties

Acronym	Property	Measurements
HIA	Human intestinal absorption	677
F	Human oral bioavailability	660
PPB	Plasma protein binding	440
BBB	Blood Brain Barrier	200
Vd	Volume of distribution	291
Cl	Renal clearance	360
T1/2	Half-life	355

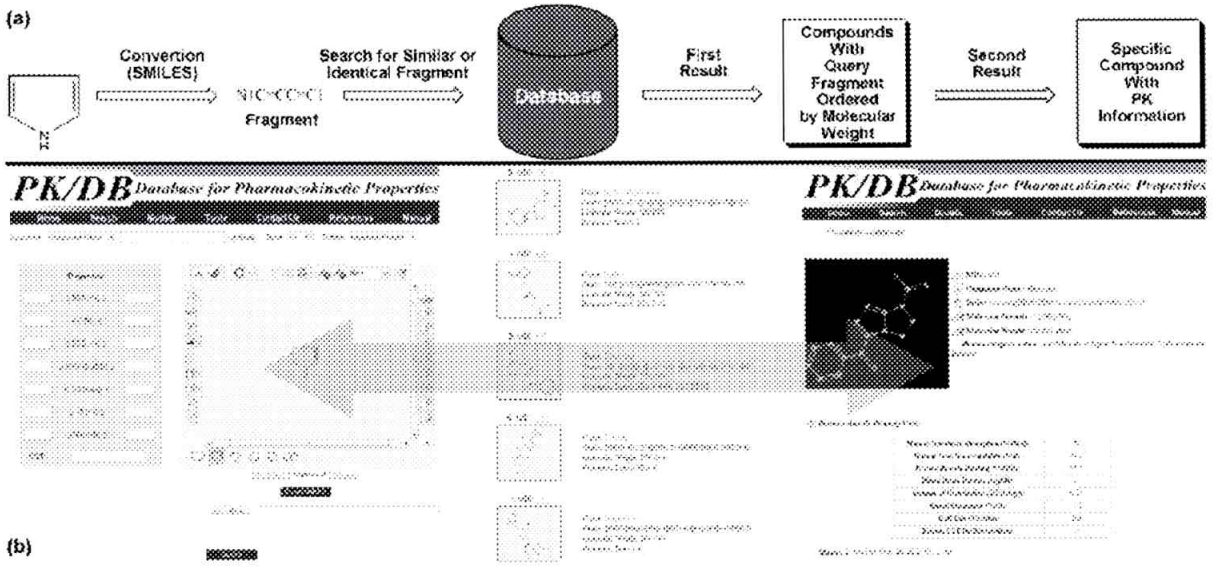

Fig. 1. PK/DB allows searching of PK for a large number of compounds using an easy to access interface. (**a**) Overview of the PK/DB system architecture. (**b**) This screenshot shows the results of a search based on a specific molecular fragment.

4 *IN SILICO* PREDICTIVE ADME MODELS

PK/DB presents five *in silico* predictive models available for the evaluation of ADME properties, including human oral bioavailability, plasma protein binding, human intestinal absorption, blood–brain barrier permeation and water solubility. These predictive models are statistically robust and have both good internal and external consistency. In addition, the models have been validated by external test sets of compounds which were not considered for model generation. The predictive models were developed by our research group based on specialized molecular fragments using the hologram quantitative structure–activity relationships (HQSAR) method (Castilho *et al.*, 2007; Moda *et al.*, 2007a, b).

5 IMPLEMENTATION

An integrated data structure and a variety of querying logics were developed to allow easy and efficient retrieval of pharmacokinetic data. The PK/DB system is installed on Red Hat Enterprise Linux workstations and employs the PostgreSQL v8.2 (http://www.postgresql.org) as a relational database management system (RDBMS) and Apache 2.0 (http://www.apache.org) server as a web server platform. Its web interface is implemented using PHP (http://www.php.net), Javascript and DHTML. For the flexible integration of the information present in PK/DB, the server interface was implemented using PHP and the search logic implemented in C++ using the OEChem TK (OpenEye Inc., Santa Fe, NM, USA), a programming library of functions that properly handle the details of working with molecules. Towards more user-friendly searching and retrieval systems, PK/DB provides interactive web interfaces, including the graphic structure editor MarvinSketch applet (ChemAxon Inc., Budapest, Hungary).

6 CONCLUSIONS

The use of computational models in the prediction of PK of active compounds is growing rapidly in drug discovery due to the benefits they provide in throughput and early application in the design of new drug candidates. PK/DB is a new database that provides useful information on a variety of important PK, as well as access to predictive *in silico* ADME models. The PK/DB suite is designed to be utilized by all researchers in the drug discovery field, and will be continuously updated and upgraded as new information becomes available.

Funding: FAPESP (The State of São Paulo Research Foundation); CNPq (The National Council for Scientific and Technological Development), Brazil.

Conflict of Interest: none declared.

REFERENCES

Canavan,N. (2007) FDA and drug companies alike want ADME-tox testing performed earlier and earlier in a drug's life cycle. *Drug Discov. Dev.*, **10**, 34–36.

Castilho,M.S. *et al.* (2007) Classical and hologram QSAR studies on a series of tacrine derivatives as butyrylcholinesterase inhibitors. *Lett. Drug Des. Discov.*, **4**, 106–113.

Guido,R.V.C. *et al.* (2008) Virtual screening and its integration with modern drug design technologies. *Curr. Med. Chem.*, **15**, 37–46.

Jónsdóttir,S.O. *et al.* (2005) Prediction methods and databases within chemoinformatics: emphasis on drug and candidates. *Bioinformatics*, **21**, 2145–2160.

Moda,T.L. *et al.* (2007a) Hologram QSAR model for the prediction of human oral bioavailability. *Bioorg. Med. Chem.*, **15**, 7738–7745.

Moda,T.L. *et al.* (2007b) *In silico* prediction of human plasma protein binding using hologram QSAR. *Lett. Drug. Des. Discov.*, **4**, 502–509.

Norinder,U. and Bergström,C.A.S. (2006) Prediction of ADMET properties. *ChemMedChem.*, **1**, 920–937.