Structural bioinformatics

SMARTCyp 3.0: Enhanced cytochrome P450 site-of-metabolism prediction server

Lars Olsen[1,2], Marco Montefiori[1,3], Khanhvi Phuc Tran[1] and Flemming Steen Jørgensen[1,*]

[1]Department of Drug Design and Pharmacology, Faculty of Health and Medical Sciences, University of Copenhagen, Universitetsparken 2, DK-2100, Copenhagen Ø, Denmark, [2]Present address: Protein Engineering, Novozymes A/S, Krogshøjvej 36, DK-2880 Bagsvaerd, Denmark, [3]Present address: CNR-ICRM, Via Mario Bianco 9, 20131 Milan, Italy.

*To whom correspondence should be addressed.

Associate Editor: XXXXXXX

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Abstract

Motivation: Cytochromes P450 (CYPs) are the most important class of drug metabolizing enzymes. Prediction of drug metabolism is important in development of new drugs, to understand and reduce adverse drug reactions (ADRs) and to reduce animal testing.

Results: SMARTCyp 3.0 is an updated version of our previous web-server for prediction of site-of-metabolism (SOM) for CYP-mediated metabolism, now in Python 3 with increased structural coverage and new features. The SMARTCyp program is a first principle-based method using density functional theory (DFT) determined activation energies for more than 250 molecules to identify the most likely SOM. New features include a similarity measure between the query molecule and the model fragment, a new graphical interface and additional parameters expanding the structural coverage of the SMARTCyp program.

Availability: The SMARTCyp server is freely available for use on the web at smartcyp.sund.ku.dk

Contact: fsj@sund.ku.dk

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Metabolism of foreign compounds, xenobiotics, comprising drug compounds, food ingredients and chemicals in the environment is an important part of the human organisms defense system. A large number of enzymes attack and convert foreign compounds to more soluble metabolites in order to facilitate excretion. Some of the metabolites are harmless, whereas others are reactive and potentially dangerous for the human organism (Thompson *et al.*, 2016). The foreign compounds, especially drugs, as well as their metabolites, may also affect the bioavailability of other drugs and, thereby, lead to drug-drug interactions (DDI). The interplay between different drugs, or drugs and metabolites, may lead to adverse drug effects (ADRs), which are causing thousands of hospitalizations and even deaths worldwide.

Fragment-based methods are methods based on identifying specific molecular fragments present in individual compounds, associate an pre-calculated value for each fragment, and subsequently derive a molecular property from the values of the fragments. We have previously used this approach in the development of the SMARTCyp method for prediction of cytochrome p450 mediated metabolism (Rydberg *et al.*, 2010; Rydberg *et al.*, 2010). The SMARTCyp method is a first principle-based method, i.e. transferable and not limited by access to experimental data, diversity of training sets and parameterization.

2 Methods

The development of the SMARTS rules is based on a dataset comprising a total of 475 CYP substrates from the literature. The procedure for the density functional theory (DFT) determinations of the activation energies, the energy difference between the transition state and the reactant complex, have already been described elsewhere (Leth *et al.*, 2015; Olsen *et al.*, 2015; Rydberg *et al.*, 2014).

3 Results

Whereas the original SMARTCyp program was Java-based using the CDK library, SMARTCyp 3.0 is Python-based using the RDKit library (cf. Supplementary Material, Table S1). CDK and RDKit handle aromaticity different and, accordingly, there is a difference in which atoms the SMARTS patterns match, e.g. due to different atom typing. In order to secure backward compatibility, the differences in the SMARTS rules detected by CDK and RDKit for all sites were identified for a test set of 475 3A4 substrates (Zaretzki *et al.*, 2012). Each divergent SMARTS rule was individually analyzed and compared with the matched molecule and the substructure it was created from, and corrected when necessary. See Supplementary Material Table S2 and S3 for examples. SMARTCyp 3.0 using the updated SMARTS rules (Supplementary Material, Table S4) has a performance equal to the previous version of the program (Supplementary Material, Table S5).

One of the new features implemented in SMARTCyp 3.0 is the *Similarity* feature, which based on Morgan fingerprints compares the similarity of the matched substructure to the full molecule fragment for which the DFT calculation was made upon. A score of 1.0 indicates a perfect match, whereas a score of 0.0 indicates no matching fragment, which means the atom is either not considered reactive, or the assigned reactivity is not based on the calculated data, and therefore not as reliable (cf. Fig. 1).

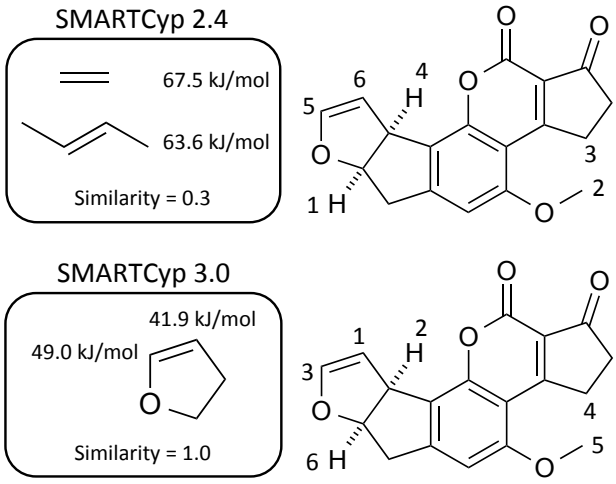

Fig. 1. SMARTCyp 2.4 and 3.0 predicted SOMs for aflatoxin B1. Numbers on the structures indicate reactivity order with 1 being the most reactive site. In boxes the fragments and energies used for assignment of SOMs in version 2.4 and 3.0, respectively. The prediction in version 3.0 is in agreement with the experimentally observed metabolism.

A number of new SMARTS rules were implemented in SMARTCyp 3.0 to expand its applicability domain (See Supplementary Material for details). SMARTCyp not only predicts 3A4 metabolism, but by applying some geometrically derived isoform specific corrections to the DFT determined activation barriers, it also predicts 2C9 and 2D6 metabolism (Rydberg *et al.*, 2012; Rydberg *et al.*, 2012; Rydberg *et al.*, 2013).

4 Implementation

SMARTCyp 3.0 is a program developed in Python 3.0 and using the RDKit library while the webserver is based on Flask. It is available as a web server at smartcyp.sund.ku.dk (cf. Fig. 2).

5 Conclusion

SMARTCyp 3.0 is a completely unsupervised, objective, and extremely fast fragment-based method based on first principles for prediction of SOM in CYP substrates.

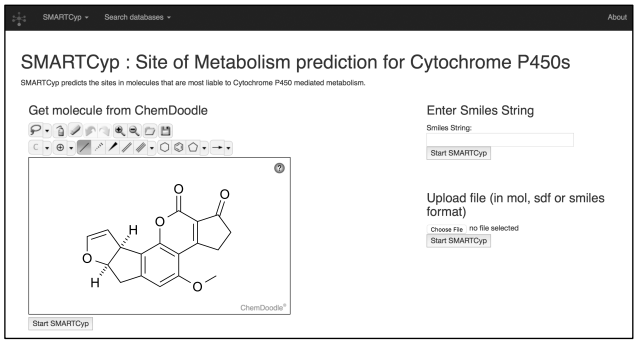

Fig. 2. The new SMARTCyp main web page with the three input modes: draw a molecule, enter a SMILES string or upload a SDF file, the latter containing multiple molecules.

Funding

EU supported this work via the ARIADME project (607517).

Conflict of Interest: none declared.

References

Leth, R. *et al.* (2015) Density functional theory study on the formation of reactive benzoquinone imines by hydrogen abstraction. *J Chem Inf Model,* **55**, 660-666.

Olsen, L. *et al.* (2015) Prediction of cytochrome P450 mediated metabolism. *Adv Drug Deliv Rev,* **86**, 61-71.

Rydberg, P. *et al.* (2010) The SMARTCyp cytochrome P450 metabolism prediction server. *Bioinformatics,* **26**, 2988-2989.

Rydberg, P. *et al.* (2010) SMARTCyp: A 2D Method for Prediction of Cytochrome P450-Mediated Drug Metabolism. *ACS Medicinal Chemistry Letters,* **1**, 96-100.

Rydberg, P. *et al.* (2014) Use of density functional theory in drug metabolism studies. *Expert Opinion on Drug Metabolism & Toxicology,* **10**, 215-227.

Rydberg, P. *et al.* (2012) Ligand-Based Site of Metabolism Prediction for Cytochrome P450 2D6. *ACS Medicinal Chemistry Letters,* **3**, 69-73.

Rydberg, P. *et al.* (2012) Predicting Drug Metabolism by Cytochrome P450 2C9: Comparison with the 2D6 and 3A4 Isoforms. *ChemMedChem,* **7**, 1202-1209.

Rydberg, P. *et al.* (2013) The Contribution of Atom Accessibility to Site of Metabolism Models for Cytochromes P450. *Molecular Pharmaceutics,* **10**, 1216-1223.

Thompson, R.A. *et al.* (2016) Reactive Metabolites: Current and Emerging Risk and Hazard Assessments. *Chem Res Toxicol,* **29**, 505-533.

Zaretzki, J. *et al.* (2012) RS-Predictor Models Augmented with SMARTCyp Reactivities: Robust Metabolic Regioselectivity Predictions for Nine CYP Isozymes. *J Chem Inform Model,* **52**, 1637-1659.