BIOINFORMATICS APPLICATIONS NOTE Vol. 26 no. 23 2010, pages 2988–2989
doi:10.1093/bioinformatics/btq584

Structural bioinformatics

Advance Access publication October 14, 2010

The SMARTCyp cytochrome P450 metabolism prediction server

Patrik Rydberg, David E. Gloriam and Lars Olsen*
Department of Medicinal Chemistry, Faculty of Pharmaceutical Sciences, University of Copenhagen, Universitetsparken 2, DK-2100 Copenhagen, Denmark

Associate Editor: Anna Tramontano

ABSTRACT

Summary: The SMARTCyp server is the first web application for site of metabolism prediction of cytochrome P450-mediated drug metabolism.

Availability: The SMARTCyp server is freely available for use on the web at www.farma.ku.dk/smartcyp where the SMARTCyp Java program and source code is also available for download.

Contact: smartcyp@farma.ku.dk; lo@farma.ku.dk

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

Received on July 28, 2010; revised on September 27, 2010; accepted on October 12, 2010

1 INTRODUCTION

The most important drug-metabolizing enzymes in humans are cytochromes P450 (CYPs). They oxidize a wide range of substrates and contribute to the metabolism of 90% of all drugs (Lewis, 2008).

While there are methods for site of metabolism prediction which are solely based on semi-empirical calculations of the substrates (Hennemann *et al.*, 2009; Zheng *et al.*, 2009), more accurate results should be achieved by explicitly including the reactivity of each site in a substrate. Hitherto, reactivity models for CYPs have been restricted to semi-empirical calculations of the stability of intermediates (Jones *et al.*, 2002). Later studies have shown that more accurate activation energies for CYPs can be computed by performing very time consuming density functional theory (DFT) transition state calculations on a heme group porphyrin ring model (Olsen *et al.*, 2006; Rydberg *et al.*, 2008a, b; Shaik *et al.*, 2005). SMARTCyp, however, presents the first solution for both fast and exact prediction of CYP reactivity. Accuracy of activation energies is ensured by making very extensive DFT calculations and speed is gained by saving these in a precalculated library so that they need only be looked up at execution. SMARTCyp has been validated for prediction of metabolism by the 3A4 isoform, which contributes to the metabolism of half of all drugs (Guengerich, 1999). SMARTCyp has also been shown to work well for six other CYP isoforms (Rydberg *et al.*, 2010a, b). In this application note, we describe the SMARTCyp server, the first web service for prediction of CYP-mediated metabolism.

2 NEW FUNCTIONALITY AND IMPROVEMENTS TO SMARTCYP

The detailed description of the SMARTCyp algorithm, development and DFT calculations has been published elsewhere (Rydberg *et al.*, 2010c). In this section, we highlight significant improvements as well as new features and functions in the current version 1.5.

2.1 Increased accuracy

For this version, 72 new DFT calculations were performed and added to the 139 previous to refine and create additional SMARTS rules. The library of SMARTS rules is used to look up atom activation energies in molecules by SMARTS patterns. The majority of the new calculations were performed to systematically evaluate and refine the rules for nitrogen dealkylations of peptide and acetamide groups, as well as oxidation of aromatic five-membered rings and heterocycles. The new rules have resulted in significant improvements to the prediction of the activation energies of these fragments. Rules and calculations are presented in Table 1 (overall distribution) and Supplementary Material (DFT results and SMARTS definitions).

We have analyzed the accuracy of SMARTCyp on a set of 361 drug-like CYP3A4 substrates (Table 2). The substrates were taken from the dataset in our previous work (Rydberg *et al.*, 2010c), by extracting all molecules defined as drug-like according to the 'lip_druglike' descriptor in the MOE software (Chemical Computing Group Inc., version 2007.09). The results show an improved accuracy of 2 percentages compared with the original SMARTCyp implementation.

A preliminary validation against data for other CYP isoforms shows that a similar performance can be expected for the isoforms 1A2, 2A6, 2B6, 2C8, 2C19 and 2E1 (Rydberg *et al.*, 2010a, b).

Table 1. Distribution of the transition state calculations and SMARTS rules by atom types

Atom type	sp^3 C	sp^2 C	N	S	P	All
Calculations	77 (26)	87 (41)	24 (5)	19	1	211
Rules	11 (7)	16 (12)	10 (8)	8	1	46

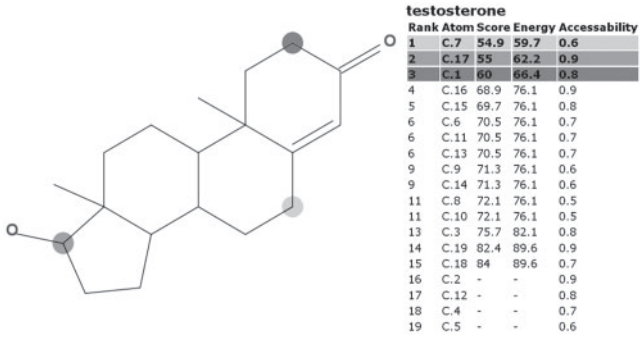
The numbers of new or modified calculations and rules are given within parentheses.

Table 2. Percent correctly predicted substrates

Atom Rank	Any Metabolite (v. 1.0)	Major Metabolite (v. 1.0)
1	70 (68)%	60 (58)%
1–2	81 (79)%	72 (70)%
1–3	86 (84)%	81 (79)%

A molecule is considered correctly predicted if an experimentally determined metabolic site is found among the one, two or three top ranked atoms, respectively.

*To whom correspondence should be addressed.



testosterone

Rank	Atom	Score	Energy	Accessability
1	C.7	54.9	59.7	0.6
2	C.17	55	62.2	0.9
3	C.1	60	66.4	0.8
4	C.16	68.9	76.1	0.9
5	C.15	69.7	76.1	0.8
6	C.6	70.5	76.1	0.7
6	C.11	70.5	76.1	0.7
6	C.13	70.5	76.1	0.7
9	C.9	71.3	76.1	0.6
9	C.14	71.3	76.1	0.6
11	C.8	72.1	76.1	0.5
11	C.10	72.1	76.1	0.5
13	C.3	75.7	82.1	0.8
14	C.19	82.4	89.6	0.9
15	C.18	84	89.6	0.7
16	C.2	-	-	0.9
17	C.12	-	-	0.8
18	C.4	-	-	0.7
19	C.5	-	-	0.6

Fig. 1. SMARTCyp results for testosterone. All five metabolic sites of testosterone are predicted correctly as top ranking atoms and the major metabolite is ranked as number one (Yamazaki *et al.*, 1997).

2.2 Improved detection of symmetric atoms

CDK (Steinbeck *et al.*, 2003, 2006) offers several ways for topological symmetry detection. SMARTCyp 1.0 implemented a combination of Morgan numbers (Morgan, 1965) and SMARTS rules, whereas the current version uses the algorithm by Hu and Xu (1994). The change of method has increased accuracy significantly.

2.3 Java program output options

The command-line SMARTCyp Java program now has three optional flags, which enable the user to suppress HTML or csv output, as well as direct the output to a specific directory. Suppressing HTML and generating only csv output makes the program even faster and allows for integration with other softwares.

3 SMARTCYP WEB SERVER

3.1 Interface features

The SMARTCyp server offers the user three ways to submit molecules. The user can upload a file in any standard format, enter SMILES strings representing molecules or draw a molecule.

The results are displayed directly in the browser and include the molecular structure and an atom ranking table for each molecule as shown in Figure 1. The three top-ranked atoms are highlighted both in the structure and the table. Furthermore, all atom numbers can be displayed by hovering the mouse pointer over the structure.

3.2 Implementation

The SMARTCyp web server uses php code to run SMARTCyp and the interface functionality. To support all standard formats, uploaded files are converted by Open Babel (Guha *et al.*, 2006) when necessary. SMARTCyp is implemented using the CDK Java library (Steinbeck *et al.*, 2003, 2006).

Funding: Alfred Benzon foundation (to P.R. and D.E.G.); Danish Medical Research Council (to L.O.).

Conflict of Interest: none declared.

REFERENCES

Guengerich,F.P. (1999) Cytochrome P-450 3A4: regulation and role in drug metabolism. *Ann. Rev. Pharmacol. Toxicol.*, **39**, 1–17.

Guha,R. *et al.* (2006) Blue Obelisk - interoperability in chemical informatics. *J. Chem. Inf. Model.*, **46**, 991–998.

Hennemann,M. *et al.* (2009) CypScore: quantitative prediction of reactivity toward cytochromes P450 based on semiempirical molecular orbital theory. *ChemMedChem,* **4**, 657–669.

Hu,C.Y. and Xu,L. (1994) Algorithm for computer perception of topological symmetry. *Anal. Chim. Acta*, **295**, 127–134.

Jones,J.P. *et al.* (2002) Computational models for cytochrome P450: a predictive electronic model for aromatic oxidation and hydrogen atom abstraction. *Drug Metab. Dispos.*, **30**, 7–12.

Lewis,D.F. and Ito,Y. (2008) Human cytochromes P450 in the metabolism of drugs: new molecular models of enzyme-substrate interactions. *Expert Opin. Drug Metab. Toxicol.*, **4**, 1181–1186.

Morgan,H.L. (1965) Generation of a unique machine description for chemical structures-a technique developed at chemical abstracts service. *J. Chem. Doc.*, **5**, 107–113.

Olsen,L. *et al.* (2006) Prediction of activation energies for hydrogen abstraction by cytochrome. P450. *J. Med. Chem.*, **49**, 6489–6499.

Rydberg,P. *et al.* (2008a) Sulfoxide, sulfur, and nitrogen oxidation and dealkylation by cytochrome P450. *J. Chem. Theory Comput.*, **4**, 1369–1377.

Rydberg,P. *et al.* (2008b) Prediction of activation energies for aromatic oxidation by cytochrome P450. *J. Phys. Chem. A*, **112**, 13058–13065.

Rydberg,P. *et al.* (2010a) SMARTCyp: a 2D method for prediction of cytochrome P450-mediated drug metabolism. In *5th Joint Sheffield Conference on Chemoinformatics.* Sheffield, UK.

Rydberg,P. *et al.* (2010b) SMARTCyp: drug metabolism prediction based on DFT calculations of transition states. In *8th European Conference on Computational Chemistry*. Lund, Sweden.

Rydberg,P. *et al.* (2010c) SMARTCyp: a 2D method for prediction of cytochrome P450-mediated drug metabolism. *ACS Med. Chem. Lett.*, **1**, 96–100.

Shaik,S. *et al.* (2005) Theoretical perspective on the structure and mechanism of cytochrome P450 enzymes. *Chem. Rev.*, **105**, 2279–2328.

Steinbeck,C. *et al.* (2003) The Chemistry Development Kit (CDK): an open-source Java library for chemo- and bioinformatics. *J. Chem. Inf. Comput. Sci.*, **43**, 493–500.

Steinbeck,C. *et al.* (2006) Recent developments of the Chemistry Development Kit (CDK) - an open-source Java library for chemo- and bioinformatics. *Curr. Pharm. Design*, **12**, 2111–2120.

Yamazaki,H. and Shimada,T. (1997) Progesterone and testosterone hydroxylation by cytochromes P450 2C19, 2C9, and 3A4 in human liver microsomes. *Arch. Biochem. Biophys.*, **346**, 161–169.

Zheng,M. *et al.* (2009) Site of metabolism prediction for six biotransformations mediated by cytochromes P450. *Bioinformatics*, **25**, 1251–1258.