Bioinformatics, 31(7), 2015, 1136–1137
doi: 10.1093/bioinformatics/btu761
Advance Access Publication Date: 18 November 2014
Applications Note

OXFORD

Structural bioinformatics

XenoSite server: a web-available site of metabolism prediction tool

Matthew K. Matlock[1], Tyler B. Hughes[1] and S. Joshua Swamidass[2,*]

[1]Division of Biology and Biological Sciences and [2]Department of Pathology and Immunology, Washington University School of Medicine, St. Louis, MO 63130, USA

*To whom correspondence should be addressed.

Associate Editor: Anna Tramontano

Received and revised on September 3, 2014; accepted on November 9, 2014

Abstract

Summary: Cytochrome P450 enzymes (P450s) are metabolic enzymes that process the majority of FDA-approved, small-molecule drugs. Understanding how these enzymes modify molecule structure is key to the development of safe, effective drugs. XenoSite server is an online implementation of the XenoSite, a recently published computational model for P450 metabolism. XenoSite predicts which atomic sites of a molecule—sites of metabolism (SOMs)—are modified by P450s. XenoSite server accepts input in common chemical file formats including SDF and SMILES and provides tools for visualizing the likelihood that each atomic site is a site of metabolism for a variety of important P450s, as well as a flat file download of SOM predictions.
Availability and implementation: XenoSite server is available at http://swami.wustl.edu/xenosite.
Contact: swamidass@wustl.edu

1 Introduction

Cytochrome P450 enzymes (P450s) are a family of heme-thiolated enzymes that catalyze a variety of oxidation reactions. They oxidatively metabolize the majority of FDA-approved, small-molecule drugs (Nebert and Russell, 2002). P450s metabolize individual molecules at specific atoms: sites of metabolism (SOMs). The specific SOMs associated with a molecule's P450 metabolism are important to know because they affect its bioavailability, efficacy and toxicity. For example, acetominophen is metabolized by P450s at a specific site into NAPQI, which is toxic to hepatocytes, which can acute liver failure in some patients. Accurately predicting these SOMs enables medicinal chemists to modify molecule to avoid toxic metabolites and maximize bioavailability.

There are *in vitro* assays capable of determining a molecule's SOMs, but computational methods for predicting SOMs are quicker, less expensive and frequently used in drug development. In addition to commercial software packages for predicting SOMs, there are at least two software packages produced by academic groups, RS predictor and SMARTCyp (Rydberg *et al.*, 2010; Zaretzki *et al.*, 2012). Recently, we reported a new method for predicting P450 SOMs, XenoSite, that is more accurate than other approaches (Zaretzki *et al.*, 2013). This article reports a public web server that makes XenoSite available to predict the SOMs of user-submitted molecules.

XenoSite is a machine-learning approach to modeling P450 metabolism. Using a neural network, XenoSite models are trained on the largest set of publicly available P450 metabolism data consisting of over 680 compounds annotated with SOMs for the P450 isozymes 1A2, 2A6, 2B6, 2C8, 2C9, 2C19, 2D6, 2E1 and 3A4, as well as human liver microsomes (HLM), which combines the individual P450 models to represent the expected metabolism of compounds *in vivo*. In cross-validation experiments, XenoSite identifies experimentally observed SOMs within the top two rank positions for substrate sets of each P450 isozyme with a high level of accuracy: 87.1%, 85.7%, 83.4%, 88.7%, 86.7%, 89.0%, 88.5%, 83.5%, 87.6% and 89.4%, respectively, on these isozymes (Zaretzki *et al.*, 2013).

2 Features and implementation

XenoSite server is an online implementation of the XenoSite Cytochrome P450 metabolism model (Zaretzki *et al.*, 2013) for

predicting which atomic sites of a molecule are modified by P450s. Key features of this web server are as follows:

1. The server computes XenoSite predictions for nine P450 isozymes including 1A2, 2A6, 2B6, 2C8, 2C9, 2C19, 2D6, 2E1 and 3A4, as well as HLMs.
2. As a standard for comparison, the server computes SMARTCyp predictions for P450 metabolism Rydberg *et al.* (2010).
3. Predictions are visualized in shaded images of the input molecules (Fig. 1) and are available for download in a flat file.
4. Prediction images are available using several scaling strategies to aid interpretability and are also available for download in a zipped archive.
5. Molecules uploaded to the server that are also part of the training set are displayed with known SOMs circled.
6. XenoSite models in the server will be updated as the training dataset grows and modeling approaches improve.

Complete documentation and a comprehensive performance comparison are available online.

2.1 Implementation

XenoSite server is implemented in python, using the Pyramid web framework (http://www.pylonsproject.org/) as well as SciPy and NumPy for numerical computations, MOPAC (Stewart, 1990) (http://openmopac.net) to compute quantum mechanical descriptions of potential SOMs, OpenBabel (O'Boyle *et al.*, 2011) (http://openbabel.org) to generate topological descriptions of molecules and compute useful measures such as LogP and SMARTP450 (Rydberg *et al.*, 2010) to estimate the reactivity of potential SOMs. It also utilizes the OpenEye OEChem software (http://eyesopen.com) to generate images of SOM predictions.

2.2 Inputs

XenoSite server allows the submission of molecular structures in SMILES or SDF format, either via a file upload or by typing or pasting molecules in either format into the web site. SMILES and SDF format are common chemical formats for which several publicly available libraries exist.

2.3 Outputs

XenoSite provides prediction output in a tab-delimited table with seven columns. The first column is the molecule title or a number corresponding to the order in which the molecules were input if titles are not provided in the SDF or SMILES. The second column is the number of the atom in the input molecule. The third column contains numerical labels that group atoms by common topology or are part of the same reaction. The fourth column gives the probability of metabolism as predicted by XenoSite. The fifth column contains the atom type. The sixth column is a Boolean value indicating whether the atomic site is annotated as a known SOM from the literature sources used in training the XenoSite predictive models. The final column is the background probability of observing a site of metabolism given the model, which can be used to interpret or rescale the prediction.

XenoSite also provides visual output for each molecule and each P450 isozyme. Molecules' SOMs are labeled by a color gradient (as in Fig. 1), with the colors scaled according to the chosen scaling mode. In the default mode 'background', blue corresponds to a zero probability of metabolism, white corresponds to a probability equal to the background probability of observing a SOM at random and red corresponds to a probability of 1.0. Probability values can be scaled to aid interpretation by using other user-selectable scaling algorithms including unscaled, normalized or polynomial scaling. Images of individual molecules with SOM predictions can be downloaded in PNG or PDF format. If a molecule is annotated with known SOMs, these sites are circled.

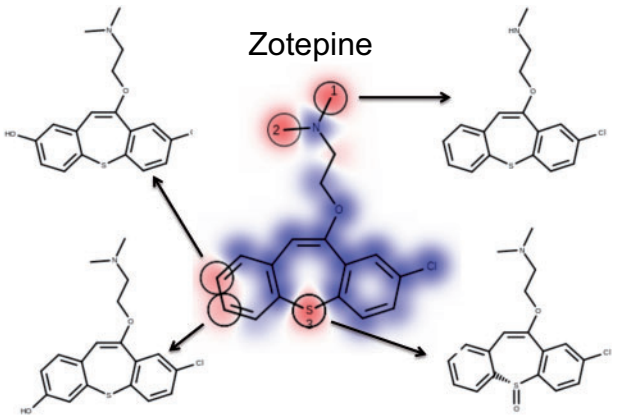

Fig. 1. XenoSite server's HLM metabolism predictions on zotepine (center molecule). Zotepine metabolism by CYPs includes nitrogen dealkylation, carbon hydroxylation and sulfur oxidation. In the figure, sites of metabolism are correctly identified by Xenosite. Zotepine is in the training set, so known SOMs are circled. Each site of metabolism corresponds to specific metabolite molecules (unshaded molecules). Predictions closely match the known sites

3 Conclusion

In this article, we describe a publicly available, web-based tool for predicting P450-mediated xenobiotic metabolism of user-submitted molecules, based on the XenoSite Cytochrome P450 metabolism prediction models (http://swami.wustl.edu/xenosite). XenoSite server provides a simple interface for generating metabolism predictions, as well as tools for visualizing and comparing potential SOMs across a variety P450 isozymes.

Acknowledgements

The authors would like to extend special thanks to Brooke Hauser for her help with website graphics and layout. Figure 1 and molecule images on the web server are created with OEDepict, version 1.7.4.5, OpenEye Scientific Software, Inc., Santa Fe, NM, www.eyesopen.com, 2014. M.K.M. wrote the manuscript and implemented XenoSite server, T.B.H. implemented the molecule visualization tools and S.J.S. provided project oversight and substantive editing.

Funding

This study was funded by the Pathology and Immunology Department of Washington University in Saint Louis.

Conflict of Interest: none declared.

References

Nebert,D.W. and Russell,D.W. (2002) Clinical importance of the cytochromes p450. *Lancet,* **360**, 1155–1162.

O'Boyle,N.M. *et al.* (2011) Open Babel: an open chemical toolbox. *J. Cheminform.*, **3**, 33.

Rydberg,P. *et al.* (2010) The smartcyp cytochrome p450 metabolism prediction server. *Bioinformatics,* **26**, 2988–2989.

Stewart,J. J. (1990) Mopac: a semiempirical molecular orbital program. *J. Comput. Aided Mol. Des.*, **4**, 1–103.

Zaretzki,J. *et al.* (2012) Rs-predictor models augmented with smartcyp reactivities: robust metabolic regioselectivity predictions for nine cyp isozymes. *J. Chem. Inf. Model,* **52**, 1637–1659.

Zaretzki,J. *et al.* (2013). XenoSite: Accurately predicting cyp-mediated sites of metabolism with neural networks. *J. Chem. Inf. Model,* **53**, 3373–3383.