WhichP450: Predicting Which Cytochrome P450 Isoforms are involved in the Metabolism of a Xenobiotic

Jonathan Tyzack, Nicholas Foster, Peter Hunt, Matthew Segall
Optibrium Ltd, Cambridge, UK

Introduction

Optibrium™, as part of the European HeCaToS project, has developed models to predict which Cytochrome P450 isoforms are involved in the metabolism of a xenobiotic, an important consideration when assessing its metabolic fate *in vivo* and the subject of recent studies [1,2]. The various P450 isoforms have active sites of different shapes, sizes and characters [3], favouring different binding pharmacophores which can lead to metabolism at different sites within the molecule. Therefore, predicting the P450 isoforms likely to be involved in metabolism is a useful precursor to predicting the metabolites that might be formed from the application of isoform-specific predictive models [4]. 'WhichP450' models also have application in assessing the risk of drug-drug interactions and the impact of genetic polymorphisms. A molecule reliant on a single isoform for metabolic clearance is at an increased risk of exhibiting drug-drug interactions or of genetic polymorphisms affecting its pharmacokinetics.

Here we present QSAR models to predict which P450 isoforms are likely to be involved in the metabolism of particular molecules. We have compared a variety of molecular descriptors [5] and demonstrate that a major metabolising P450 isoform can be identified in the top 2 predictions for over 90% of the independent test set. Furthermore, these models can predict when multiple isoforms may contribute to the metabolism of a compound and, combined with models of regioselectivity of metabolism, estimate the resulting metabolite profile.

Method and Results

We have performed a detailed and comprehensive review of the literature to create a data set consisting of 484 molecules where the major and minor metabolising isoforms for each molecule and site of metabolism have been identified. An analysis of the data sets showing the occurrence frequency by major and minor is shown in **Figure 1**.

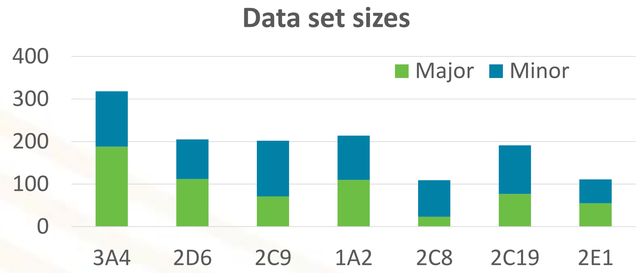


Figure 1: Chart to show the size of the data sets used in this work. A molecule, and site, can be a substrate (major or minor) for more than one isoform, the total number of molecules used in this work is 484.

QSAR models of the major isoform data, and of the major plus minor data, have been built using a multi-class SVM methodology [6,7] based on a variety of molecular fingerprints [5]. A test set, containing 30% of the data, was randomly selected and not used in the model building process enabling us to evaluate the predictive performance in identifying a major metabolising isoform in the top-k predictions, as shown in **Figure 2**. The impact of incorporating the minor isoform data (but giving the major data greater emphasis through over-sampling) was investigated for all the fingerprints but only the results for the MACCS keys are shown as the performance is illustrative of the other fingerprints.

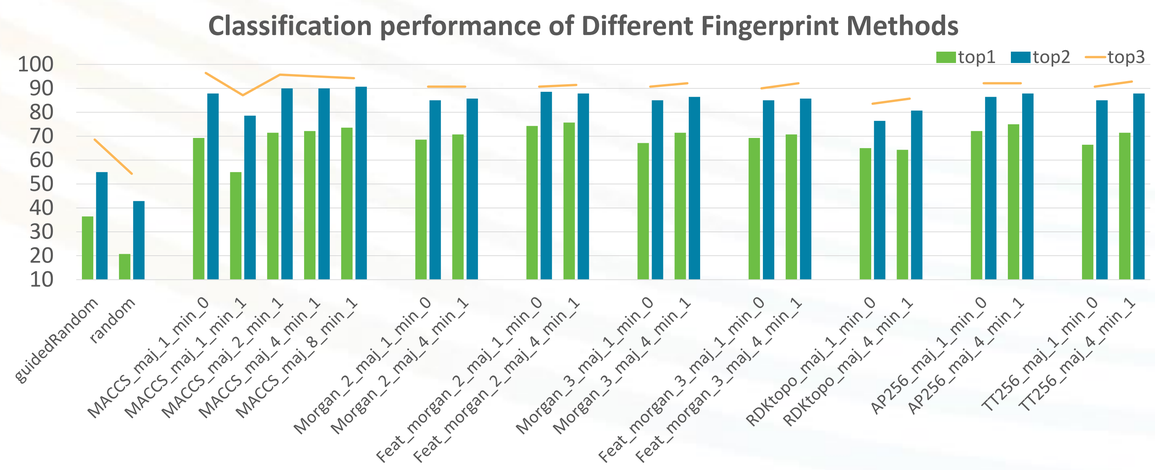
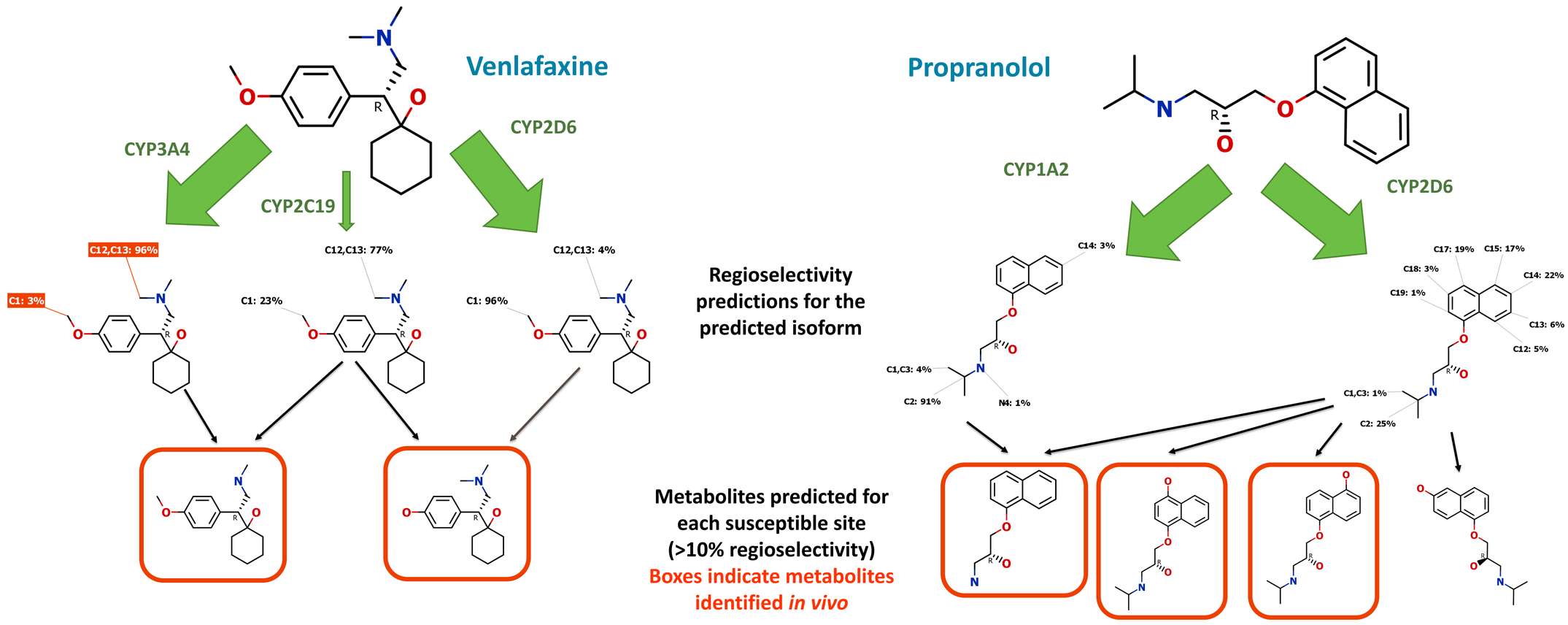


Figure 2: Predictive performance for identifying a major metabolising isoform for an independent test set of molecules using a variety of over-sampling criteria and different fingerprint methodologies [5]. The axis labels indicate the fingerprint type and the number of times the major and minor data were sampled.

The predictions from the Atom Pair AP256_maj_4_min_1 model are shown below with the figures next to each isoform being the chance that that isoform is a metaboliser of the test molecule. The predictions for two of the test compounds, Venlafaxine and Propranolol are shown to illustrate how the WhichP450 model fits into a workflow of human metabolite prediction with the P450 module incorporated in the StarDrop software[4].

Venlafaxine_R --- (observed major: 2D6, minor: 3A4 2C9 2C19)[8]
Predicted probabilities (using AP256bit fingerprint)
2D6: 0.34, 3A4: 0.26, 2C19: 0.14, 2C9: 0.10, 1A2: 0.09, 2C8: 0.04, 2E1: 0.02

Propranolol_R --- (observed major: 2D6 1A2, minor: 1A2)[9,10]
Predicted probabilities (using AP256bit fingerprint)
2D6: 0.46, 1A2: 0.23, 2C19: 0.09, 3A4: 0.09, 2C9: 0.06, 2E1: 0.03, 2C8: 0.03



Discussion

The QSAR methodology presented, using either MACCS keys or AtomPair descriptors and a multi-class SVM classifier, can identify a ***major*** metabolising isoform as the top prediction for 75% of an independent test set, and in the top 2 predictions for 90% of that set. The equivalent figures for identifying ***any*** metabolising isoform (i.e. either major or minor) are 86% and 95% respectively, (data not shown). We have found that the incorporation of the minor metabolising isoform data into the model training, but placing greater emphasis on the major data through oversampling, provides an advantage to simply taking the major isoform alone and also expands the available data for modelling. The combination of these 'WhichP450' models with the regioselectivity models currently implemented in the StarDrop software [4] provides a powerful guide as to the most relevant isoforms for the metabolism of compounds and to which isoform-specific metabolites to investigate. The extension of this methodology to multiple steps (i.e. metabolites of metabolites) and other metabolic pathways is ongoing.

References

[1] P. Rydberg et al. (2013) Bioinformatics. 29(16): 2051-2
[2] G.P.S. Raghava et al. (2010) BMC Pharmacology. 10(1): 8
[3] M.B.G. Cruciani et al. (2006) Molecular Interaction Fields, ch.12, 273-290, Wiley-VCH
[4] StarDrop, version 6.1, Optibrium Ltd, Cambridge, UK
[5] RDKit, Open-Source Cheminformatics. http://www.rdkit.org
[6] R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.
[7] D. Meyer et al. (2014) e1071: Misc Functions of the Department of Statistics (e1071)
[8] S. Fogelman et al. (1999) Neuropsychopharmacology. 20: 480-90
[9] K. Yoshimoto et al (1995) Br. J. Clin. Pharmacol. 39: 421-31
[10] M. Ching (1996) Drug Metabolism & Disposition. 24(6): 692-4

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under the grant agreement nº 602156