Predicting Routes, Sites and Products of Drug Metabolism
12th International ISSX Meeting, July 30th 2019

Matthew Segall – matt@optibrium.com
Overview

• Approaches to predicting metabolism
  – Empirical vs mechanistic

• Predicting P450 metabolism
  – P450 regioselectivity
  – Which P450

• Beyond P450s
  – Flavin-containing monooxygenases (FMO)
  – UDP glucuronosyltrasfreases (UGT)

• Conclusions
Approaches to Predicting Drug Metabolism

Empirical
- Statistical Modelling
- Machine Learning

Mechanistic
- Molecular Dynamics
- Quantum Mechanics
Approaches to Predicting Drug Metabolism

**Empirical**

**Pros**
- Fast
- Easy to set up

**Cons**
- Needs lots of data
- Non-transferable
- Qualitative

**Mechanistic**

**Pros**
- Can be built on smaller high-quality data sets
- Transferable – based on physical principles
- (Semi) quantitative

**Cons**
- (Very) slow
- Requires detailed understanding
Predicting P450 Metabolism
**Cytochrome P450s**

- Ubiquitous superfamily of haem-containing monooxygenase enzymes

- Responsible for ~70-80% of phase I drug metabolism, leading to:
  - Rapid clearance or low bioavailability
  - Potential for drug-drug interaction
  - Impact of P450 polymorphism
  - Bioactivation to form reactive/toxic metabolites

- Primary isoforms responsible for drug metabolism in humans

---

Zanger and Schwab, Pharmacol. & Therapeut. 138(1) p. 103 (2013)
P450 Catalytic Cycle
Predicting sites of metabolism - regioselectivity

Product formation step

P450 compound I and bound substrate
Predicting Sites of Metabolism
Regioselectivity

Two primary factors determine the sites of metabolism:

• **Electronic properties of substrate – reactivity**
  - $H$-abstraction – aliphatic oxidation, $N$-dealkylation, $O$-dealkylation
  - Direct oxidation – aromatic oxidation, epoxidation, $N$-oxidation, $S$-oxidation
  - Independent of isoform

• **Orientation of substrate in active site**
  - Electrostatic interactions between protein and substrate
  - Freedom to move
  - Steric accessibility
  - Dependent on isoform and substrate
The activation energy ($\Delta H_A$) of the rate-limiting step is a key factor determining the rate of reaction at each site.

- Reaction energetics modelled for $H$-abstraction and direct oxidations using density functional theory.
Quantum Mechanical Models for CYP Reactivity

- Semi-empirical QM methods (AM1) are used for practical calculations
  - Surrogate radical used instead of haem
  - Brønsted relationships used to estimate activation energies
  - Corrections applied based on ab initio QM

- Full substrate included in simulation
  - Not ‘pattern matching’ sites to precalculated energies
  - Includes subtle longer range effects
  - Important when developing a lead series

© 2019 Optibrium Ltd.

Capturing Steric and Orientation Effects

- Corrections to activation energies estimated for each isoform
  - 3A4, 2D6, 2C9, 1A2, 2C8, 2C19, 2E1

- Statistical models using 2D descriptors
  - Distances to charged functionalities, H-bond acceptors/donors, etc.
  - Distances to rings, flexible linkers, ‘bulky’ groups

- Trained and tested using high-quality regioselectivity data sets

<table>
<thead>
<tr>
<th>Isoform</th>
<th>Number of molecules</th>
</tr>
</thead>
<tbody>
<tr>
<td>CYP3A4</td>
<td>305</td>
</tr>
<tr>
<td>CYP2D6</td>
<td>202</td>
</tr>
<tr>
<td>CYP2C9</td>
<td>193</td>
</tr>
<tr>
<td>CYP1A2</td>
<td>201</td>
</tr>
<tr>
<td>CYP2C19</td>
<td>184</td>
</tr>
<tr>
<td>CYP2E1</td>
<td>105</td>
</tr>
<tr>
<td>CYP2C8</td>
<td>106</td>
</tr>
</tbody>
</table>

Validation
Independent test set of 30% of data

Example Regioselectivity Prediction
Venlafaxine

CYP3A4

C12, C13: 96%
C1: 3%

CYP2D6

C12, C13: 4%
C1: 96%

WhichP450

Objectives

• Many isoforms of P450
  – Different active site constraints

• Predictions of regioselectivity for which isoform(s) are most relevant?

• Identify possibilities of DDIs or polymorphic effects

• Compounds may be metabolised by multiple isoforms

Binding sites: CYP3A4 – purple & CYP2E1 – blue
P450 Catalytic Cycle
Predicting which P450 isoform(s)

Substrate binding

Binding sites: CYP3A4 – purple & CYP2E1 – blue
WhichP450

Methods

- **Data set**
  - 465 unique compounds
  - 633 compound/isoform pairs

- **Considers 7 isoforms**
  - 3A4, 2D6, 2C9, 1A2, 2C8, 2C19, 2E1

- **Random forest model**
  - Random forests
  - Whole molecule and 2D descriptors

- **Model rank orders isoforms by probability**
WhichP450
Methods

• Data set
  – 465 unique compounds
  – 633 compound/isoform pairs

• Considers 7 isoforms
  – 3A4, 2D6, 2C9, 1A2, 2C8, 2C19, 2E1

• Random forest model
  – Random forests
  – Whole molecule and 2D descriptors

• Model rank orders isoforms by probability
WhichP450

Results – Top-\(k\)

![Graph showing percent success rate for different models and top-k predictions.]

- **Top-1**
- **Top-2**
- **Top-3**

Putting it Together

Venlafaxine

2C19 is also a minor isoform, but not predicted

CYP3A4

CYP2C9

CYP2D6

Beyond P450s
Flavin-containing Monooxygenase (FMO)

- Phase I enzyme class involved in compound metabolism
  - Found in multiple tissues

- 5 active isoforms (FMO1–5)
  - FMO3 major isoform found in adult liver

- Mechanism involves transfer of Oxygen from FAD–OOH
  - Predominantly N/S-oxidation

© 2019 Optibrium Ltd.
Flavin-containing Monooxygenase (FMO)

- Phase I enzyme class involved in compound metabolism
  - Found in multiple tissues
- 5 active isoforms (FMO1–5)
  - FMO3 major isoform found in adult liver
- Mechanism involves transfer of Oxygen from FAD–OOH
  - Predominantly N/S-oxidation
Modelling the Reaction Mechanism

- QM simulations using DFT to determine reaction mechanism
  - Concerted, $S_{N2}$
- Calculate activation energy, $\Delta H_A$

© 2019 Optibrium Ltd.
Identifying of Sites of FMO Metabolism

Activation Energies

- 121 kJ mol\(^{-1}\)
- 20 kJ mol\(^{-1}\)
- 62 kJ mol\(^{-1}\)
- 70 kJ mol\(^{-1}\)
- 108 kJ mol\(^{-1}\)
- 177 kJ mol\(^{-1}\)
- 69 kJ mol\(^{-1}\)
- 62 kJ mol\(^{-1}\)
- 69 kJ mol\(^{-1}\)
- 133 kJ mol\(^{-1}\)
- 90 kJ mol\(^{-1}\)
Example – Predicting FMO3 Metabolism

- Activation energies calculated with semi-empirical QM model of transition state
- Steric and orientation descriptors included
- Data set
  - 67 molecules
  - 210 potential sites of metabolism
- Gaussian processes machine learning
- Classification of potential sites as metabolised (True) or not (False)
- Results on independent test set
  - Kappa = 0.82
  - Accuracy 92%
UDP-Glucurono-syltransferase (UGT)

- **Major contributors to phase II metabolism**
  - ~40% of all conjugation reactions

- **Conjugation of substrate with glucuronic acid**

- **Several human isoforms implicated in drug metabolism**
  - UGT1A – 1A1, 1A4, 1A9
  - UGT2B – 2B4, 2B7, 2B15
Transition State

- QM simulations to determine reaction mechanism using DFT
- Complex reaction mechanism
  - Proton transfers with active-site histidine residues
- Calculate activation energy, $\Delta H_A$
Transition State

- QM simulations to determine reaction mechanism
- Complex reaction mechanism
  - Proton transfers with active-site histidine residues
- Calculate activation energy, $\Delta H_A$

Glucuronic Acid

Uridine Diphosphate

Proton Donor

Proton Acceptor

Substrate

His372

His39

UMP
Example – Prediction UGT1A1 Metabolism

- Activation energies calculated with semi-empirical QM model of transition state
- Steric and orientation descriptors included
- Data set
  - 79 molecules
  - 242 potential sites of metabolism
- Gaussian processes machine learning
- Classification of potential sites as metabolised (True) or not (False)
- Results on independent test set
  - Kappa = 0.65
  - Accuracy 83%
Conclusions

• Detailed QM simulations enable us to understand the reaction mechanisms for metabolism

• This enables us to predict metabolism with greater accuracy and transferability
  – Reaction energetics are important factor governing metabolism
  – Combined with steric and orientation effects of protein environment

• Combining models of different steps in the catalytic cycle enable us to predict routes, sites and products of metabolism
  – E.g. WhichP450 and regioselectivity

• For more information
  – matt@optibrium.com
Acknowledgements

• P450 Metabolism
  – John Tyzack
  – Rasmus Leth
  – Many former colleagues from Camitro, ArQule, Inpharmatica and BioFocus
  – This research has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under the grant agreement no 602156

• UGT Metabolism
  – Mario Öeren
  – David Ponting
  – Funding from Lhasa Limited

• FMO Metabolism
  – Peter Walton
  – Mario Öeren

• All of the above – Peter Hunt