Predicting Metabolites

Enhancing An Expert System With Machine Learning

Chris Barber
Director Of Science
chris.barber@lhasalimited.org

Boston, May 2015
Lhasa Limited is a not-for-profit organisation and educational charity that promotes knowledge and data sharing in chemistry and the life sciences. It is controlled by its members (more than 200 organisations worldwide)

Derek Nexus
An expert system for the assessment of toxicity

Meteor Nexus
An expert system for the assessment of xenobiotic metabolism

Sarah Nexus
A (Q)SAR tool for the assessment of mutagenicity

Vitic Nexus
A structure-searchable toxicity database

Zeneth
An expert system for the assessment of chemical degradation
Why Predict The Structure Of Metabolites?

- Support identification of...

- Metabolites formed in analytical studies

- Sites of metabolism driving high metabolic clearance

- Potentially toxic metabolites
  - Some *in silico* toxicity models implicitly include metabolism
    - …but they may miss unusual metabolic precursors
    - …specific off-target pharmacology, or modelling through AOP’s

- Metabolites that may not translate between assays
  - Some *in vitro / in vivo* assays may not translate
Meteor Nexus

[Chemical structure diagram]

[Software interface with 'Prediction' and 'Structure' tabs]

[Table with columns 'Source', 'Problem', 'Severity']
Meteor Nexus

Tabular summary of metabolic tree
Metabolic path for selected metabolite including identification of potentially adduct-forming and other intermediates
Meteor Nexus

Options to filter by mass, molecular formula etc
Meteor Nexus

Biotransformation scope

R1 = aromatic carbon
saturated, aliphatic or aliphatic carbon
quaternary centres (except trifluoromethyl) are not allowed
Biotransformation supporting information

Oxidative N-dealkylation (sometimes called deamination) is an important biotransformation in mammalian xenobiotic metabolism. The reaction is of wide scope and has been demonstrated for secondary and tertiary amines both aliphatic and aromatic. Examples include bepridil, diethyipropion, beclomethasone, and glibopamid. The rate of N-dealkylation seems to be directly related to the lipid solubility of the substrate. The reaction is nearly always catalysed by cytochrome P450s. The mechanism involves hydrogen abstraction and oxidation addition (hydroxylation) at a carbon atom alpha to the nitrogen atom. This first step may involve the intermediacy of an iminium cation. Bond scission results from hydrolysis of the initially formed carbamoylamine intermediate. Carbamoylamines are sometimes stable enough to be conjugated and detected in urine. Occasionally, oxidation of the carbamoylamine to the amide is a competitive pathway, dealkylation also being effected overall by hydrolysis of the amide.

For tertiary amines, a second competitive pathway has been postulated which involves initial formation of an N-oxide which rearranges to the carbamoylamine. For primary amines the mechanisms may be more complex and, as well as the carbamoylamine, involve the intermediacy of the hydroxylamine, alpha-hydroxylamine, primary imine, and oxime.

The carbon atom alpha to the reaction centre may be aromatic or multiply bonded (vinyl amine) but if it is fully saturated then quaternary...
Dictionary Of Biotransformations

• Dictionary of 500 biotransformations
  • Covering both phase I and phase II reactions

Oxidative N-dealkylation (sometimes called deamination) is an important biotransformation in mammalian xenobiotic metabolism [Testa]. The reaction is of wide scope and has been demonstrated for secondary and tertiary amines both aliphatic and aromatic. Examples include bepridil [Wu et al., diethylpropion [Beckett and Stanopic] and gallopamil [Mullib and Nelson]. The rate of N-dealkylation seems to be directly related to the lipid solubility of the substrate. The reaction is nearly always catalysed by cytochrome P450s. The mechanism involves hydrogen abstraction and oxidation addition (hydroxylation) at a carbon atom alpha to the nitrogen atom. The first step may involve the intermediacy of an iminium cation. Bond scission results from hydrolysis of the initially formed carbilolamine intermediate. Carbilamines are sometimes stable enough to be conjugated and detected in urine. Occasionally, oxidation of the carbilolamine to the amine is a competitive pathway. Dealkylation also being effected overall by hydrolysis of the amide. For tertiary amines, a second competitive pathway has been postulated which involves initial formation...
How Meteor Nexus Works

Knowledge base
- Dictionary of biotransformations
- Rule base

What reactions could occur?

How likely that each reaction will occur?

Processing constraint
Rule Base

- Biotransformation ranking is determined by a reasoning-based interpretation of two types of rules describing

  Absolute likelihood of a single biotransformation

  Relative likelihood of a pair of biotransformations

Meteor Nexus Performance

• T’jollyn et al, Drug Metab Dispos 39, 2066-2075 (2011)
  • Comparative study of Meteor, MetaSite and StarDrop
  • Meteor has higher sensitivity but lower precision
  • High sensitivity is good for metabolite identification but high precision is of more value in a discovery setting

• Research objective
  • Develop methodology to better rank-order metabolite likelihoods
Lasofoxifene: Meteor Nexus Prediction

- Man|rat|monkey, first generation metabolites

Prakash et al, Drug Metab Dispos 36 1218-1226, 1753-1769 (2008)
How Meteor Nexus Could Work

Knowledge base
- Dictionary of biotransformations

Expert system
- What reactions could occur?

Database
- Experimental reactions

Machine learning
- How likely that each reaction will occur?

Processing constraint
Other statistical approaches to metabolite ranking

- SyGMA
- MetaPrint2D-React
Occurrence Ratio Method

How often does a reaction actually occur?

Large metabolism database

Occurrence Ratio

How often could a reaction occur?
Occurrence Ratio Method: Biotransformation 243

How often could a reaction occur?
1946

How often does a reaction actually occur?
636

Occurrence Ratio
32.7%

How often could a reaction occur?
1946
Occurrence Ratio Method

Red bar: the biotransformation has NOT been experimentally observed for this substrate.

Green bar: the biotransformation has been experimentally observed for this substrate.

Selected supporting examples containing the biophore for the biotransformation.
Occurrence Ratio Method Versus Meteor Nexus

At any given sensitivity, the Occurrence Ratio method gives higher precision than Meteor Nexus.

Test set: 100 compounds
Biotransformation counts
Relative threshold
Lasofoxifene: Occurrence Ratio Prediction

- Man|rat|monkey, first generation metabolites

<table>
<thead>
<tr>
<th>Metabolite</th>
<th>Observed</th>
<th>Not Observed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predicted</td>
<td>8</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>3</td>
</tr>
<tr>
<td>Not Predicted</td>
<td>2</td>
<td>7</td>
</tr>
</tbody>
</table>

Prakash et al, Drug Metab Dispos 36 1218-1226, 1753-1769 (2008)
Ways to Calculate the Occurrence Ratios

• How often is a predicted transformation observed?
  • Ratio of observed / predicted across all data

• If 2 transformations could occur, which will win?
  • Relative ranking of each pair of transformations

• How often is a predicted transformation observed…
  …. for compounds like mine?
Similarity-based Occurrence Ratios

- Meteor biotransformation structural key

  Biotransformation: 1 2 3 4 5 6 3
  
  Atom 3: O; O-C; O-C; O-C=C.

- Ceres fingerprint (whole structure)

- Ceres fingerprint (site of metabolism)
Site of Metabolism-driven Occurrence Ratios

- Occurrence ratio for a biotransformation determined by
  - signal of nearest neighbours
  - weighted by similarity around the site of metabolism
Site Of Metabolism vs. Occurrence Ratio Method

At any given sensitivity, the Site Of Metabolism method gives higher precision than the Occurrence Ratio method.

Test set: 1938 compounds
Site of metabolism counts
Relative threshold
Lasofoxifene: Site Of Metabolism Prediction

• Man\rat\monkey, first generation metabolites

Prakash et al, Drug Metab Dispos 36 1218-1226, 1753-1769 (2008)
Extending Predictions To Multiple Generations

- Propagate occurrence ratios down branches of metabolic tree
- Apply threshold constraint to the overall metabolic tree

First generation: $49.4\% \times$
Second generation: $7.5\% = 18.5\%
30\%$ Relative threshold $49.4\% \times 0.3 = 14.8\%$
Summary

- Developed transparent statistical approach to rank expert system-generated metabolites
  - More granularity over previous rule-based approach
  - Leads to increased positive predictivity
  - Allows Meteor Nexus to support a wider range of use cases
Summary

• Developed transparent statistical approach to rank expert system-generated metabolites
  • More granularity over previous rule-based approach
  • Leads to increased positive predictivity
  • Allows Meteor Nexus to support a wider range of use cases

• Future Plans
  • Continue collection of metabolism reactions
    • Have been collecting data for a year (4 student interns)
    • Currently ~1,370 parent compounds (>10K reactions)
  • Test performance against member proprietary data
  • Implement into Meteor Nexus
Questions

Acknowledgements

• Carol Marchant
• Ed Rosser
• Jonathan Vessey
Distribution Of Training Set Occurrence Ratios

Omits 145 biotransformations with rare biophores
Meteor Nexus Data And Knowledge Sharing

Knowledge base
- Dictionary of biotransformations

Database
- Experimental reactions

Knowledge from public domain data
- Knowledge from member data
- Member knowledge

Public domain data
- Member data
- Consortium-shared data
Metabolite Toxicity View
Threshold Definitions

- **Top N threshold**
  - Only display biotransformations with the top N scores

- **Absolute threshold**
  - Only display biotransformations with scores at or above some absolute value

- **Relative threshold**
  - Only display biotransformations with scores at or above some percentage of the maximum score (e.g., 60% of 49.4% = 29.6%)
Query-specific Occurrence Ratios

k-Nearest neighbour methodology ($k = 8$)
Query-specific Occurrence Ratios

Similarity between query and example substrates determined by Tanimoto index
Scores are weighted according to the (non-)observation of the biotransformation according to $\sqrt{\text{similarity}}$.
Summary

- Developed machine-learnt approach to rank expert system-generated metabolites
  - More granularity over previous rule-based approach
  - Leads to increased positive predictivity
  - Allows Meteor Nexus to support a wider range of use cases

- Dependent upon database of metabolic reactions
  - Have been collecting data for a year (4 student interns)
  - Currently ~1,370 parent compounds (>10K reactions)
Performance With Training Data Set Size

Vertical line shows size of data gathering efforts using equivalent data preparation to the original training set.

Test set: 1938 compounds
Site of metabolism counts
Relative threshold
Work in progress disclaimer

This document is intended to outline our general product direction and is for information purposes only, and may not be incorporated into any contract. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon. The development, release, and timing of any features or functionality described for Lhasa Limited’s products remains at the sole discretion of Lhasa Limited.