Mammalian Drug Metabolism In Silico
Strategies for Increasing Specificity of Prediction

Anthony Long
Lhasa Limited
22 - 23 Blenheim Terrace, Woodhouse Lane, Leeds, LS2 9HD
Direct Tel: +44 (0)113 394 6045   Fax: +44 (0)113 394 6099
Email: tony.long@lhasalimited.org   Web: www.lhasalimited.org
**Expert Systems**

**Definition**
“A program in which a reasoning (or inference) engine solves problems (or makes predictions) by applying rules from a knowledge base in response to single or multiple queries (or hypotheses)”

Such rules may be a combination of two types: factual and heuristic, and these are, in the main part (but not exclusively), non-numerical.

**An Example - A Visit to the Doctor:**

**Symptoms (Factual Knowledge):**
- Runny Nose
- Sore Throat
- Headache
- Congestion

**Diagnosis**
- A Common Cold

**Inferences (Heuristic Knowledge):**
- “It’s December and there’s a lot of it about”
- “You’re the 27th case I’ve seen this morning!”
“A program in which a reasoning (or inference) engine solves problems (or makes predictions) by applying rules from a knowledge base in response to single or multiple queries (or hypotheses)”

Goals:

1. Generation of all reasonable Phase I and Phase II metabolites
2. Organisation of metabolites into a metabolic tree (simulation of metabolic profile)
3. Identification of potentially reactive, adduct-forming metabolites
4. Assessment of likelihoods of formation of metabolites and pathways
5. Likelihoods of formation of metabolites and pathways under different biological conditions

Inference Engine: Graphical user interface, chemistry engine, etc.

Knowledge Base: Biotransformation dictionary
Rules expressing likelihood of biotransformation under different conditions

Query: A drawn chemical structure
Mammalian Drug Metabolism In Silico
Strategies for Increasing Specificity of Prediction

Knowledge Base: Biotransformation dictionary
Collection of generalised reactions (simple or complex):

\[ \text{CH}_3 \rightarrow \text{CH}_2\text{OH} \]

- **“N-Oxidation at Aromatic Nitrogen”**
  - \( R_1 = \text{aromatic carbon} \)

- **“Hydroxylation at Benzylic Methyl”**
  - \( R_1 = \text{H, F, CH}_3, \text{NHR}_3 \)
  - \( R_2 = \text{H, F, NHR}_4, \text{CH}_2\text{R}_5, \text{CONHR}_6 \)
  - \( R_3, R_4, R_6 = \text{H, C} \)
  - \( R_5 = \text{any atom} \)

- \( R_7 = \text{H, F} \)
  - \( R_8 = \text{H, F, CH}_2\text{R}_9, \text{NHR}_10, \text{OR}_11 \)
  - \( R_9, R_{10}, R_{11} = \text{H, C} \)

- \( R_{12} \equiv \text{O, S (as part of a five membered aromatic ring)} \)

- \( R_{13} = \text{H, F, CH}_2\text{R}_{14}, \text{NHR}_{15}, \text{OR}_{16} \)
  - \( R_{14}, R_{15}, R_{16} = \text{H, C} \)

- \( R_{17} = \text{H, F} \)

- \( R_{18} = \text{H, F} \)
  - \( R_{19}, R_{20} = \text{O, S (as part of a five membered aromatic ring)} \)
Mammalian Drug Metabolism In Silico
Strategies for Increasing Specificity of Prediction

Knowledge Base: **Biotransformation dictionary**
Collection of generalised reactions (simple or complex):

**“Amine Oxides from Tertiary Amines”**

H\(_3\)C-N-CH\(_3\)  \(\rightarrow\)  H\(_3\)C-N-CH\(_3\)
R1 = hydrogen or carbon

H\(_3\)C-N-CH\(_3\)  \(\rightarrow\)  H\(_3\)C-N-CH\(_3\)
R1 = aromatic carbon

**“Amine Oxides from Tertiary Amines”**

R\(_1\)N-R\(_2\)  \(\rightarrow\)  R\(_1\)-N\(^+\)-R\(_2\)
R\(_3\)
R\(_1\)-R\(_3\) = aliphatic carbon (no extra heteroatom attachments allowed)
can not be -CH\(_2\)N(CH\(_3\))\(_2\)
Nitrogen atom cannot be in a ring

R\(_4\)N-R\(_5\)  \(\rightarrow\)  R\(_4\)-N\(^+\)-R\(_5\)
R\(_6\)
R\(_4\) = aromatic carbon
R\(_5\), R\(_6\) = aliphatic carbon (no extra heteroatom attachments allowed)
can not be ArN(CH\(_3\))\(_2\)
Nitrogen atom cannot be in a ring
Each application of the biotransformation may be supported by:

**Comments**
- Mechanism (enzymology)
- Scope and diversity (Structure-Metabolism Relationships)
- Species difference/similarities

**References**
- Books, reviews, articles

**Example Reactions**
- With biodata from primary literature
## Knowledge Base:
Rules expressing likelihood of biotransformation under different conditions

### The Need for Prioritisation:

**THE COMBINATORIAL DATA EXPLOSION**

Numbers of metabolites being generated in 12-step sequences if each parent metabolite generates a further 20 child metabolites.

<table>
<thead>
<tr>
<th>Step</th>
<th>Metabolites</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>20</td>
</tr>
<tr>
<td>3</td>
<td>400</td>
</tr>
<tr>
<td>4</td>
<td>8,000</td>
</tr>
<tr>
<td>5</td>
<td>160,000</td>
</tr>
<tr>
<td>6</td>
<td>3,200,000</td>
</tr>
<tr>
<td>7</td>
<td>64,000,000</td>
</tr>
<tr>
<td>8</td>
<td>1,280,000,000</td>
</tr>
<tr>
<td>9</td>
<td>25,600,000,000</td>
</tr>
<tr>
<td>10</td>
<td>512,000,000,000</td>
</tr>
<tr>
<td>11</td>
<td>10,240,000,000,000</td>
</tr>
<tr>
<td>12</td>
<td>204,800,000,000,000</td>
</tr>
<tr>
<td>13</td>
<td>4,096,000,000,000,000</td>
</tr>
</tbody>
</table>
Mammalian Drug Metabolism In Silico Strategies for Increasing Specificity of Prediction

Knowledge Base: Rules expressing likelihood of biotransformation under different conditions

Absolute Reasoning

Influenced by:
- Frequency of occurrence
- Yields
- Scope of applicability
- Physicochemistry: lipophilicity/molecular weight

Probable Plausible Equivocal Doubted Improbable

only the very most likely metabolites are generated

more speculative metabolites are generated

all possible metabolites are generated
Mammalian Drug Metabolism In Silico
Strategies for Increasing Specificity of Prediction

Knowledge Base: Rules expressing likelihood of biotransformation under different conditions

Relative Reasoning
Expresses: Preferences for a single activating biophore (REGIOSELECTIVITY)

para-Hydroxylation

ortho-Hydroxylation

meta-Hydroxylation

Number of Examples

para-Hydroxylation
ortho-Hydroxylation
meta-Hydroxylation
Knowledge Base: Rules expressing likelihood of biotransformation under different conditions

Relative Reasoning
Expresses: Preferences for a similar functional groups (CHEMOSELECTIVITY)

- **N-Demethylation**
  - Methyl amines and related compounds

- **N-Dealkylation**
  - Branched and non-branched alkyl amines

- **Oxidative ring-opening**
  - Alicyclic amines
Mammalian Drug Metabolism In Silico Strategies for Increasing Specificity of Prediction

Knowledge Base: Rules expressing likelihood of biotransformation under different conditions

Relative Reasoning
Expresses: Preferences for a similar functional groups (CHEMOSELECTIVITY)

Sulphoridazine

FIRST Prediction

SECOND Prediction

THIRD Prediction
Mammalian Drug Metabolism In Silico
Strategies for Increasing Specificity of Prediction

Knowledge Base: Rules expressing likelihood of biotransformation under different conditions

Relative Reasoning Expresses: Preferences for a similar functional groups (CHEMOSELECTIVITY)

Venflaxazine

Mianserin

Naltrexone

Verapamil
Mammalian Drug Metabolism In Silico
Strategies for Increasing Specificity of Prediction

Biotransformation Analysis for 18 Compounds

**METHOD:**

**Absolute Reasoning:**

**Relative Reasoning:**

**EQUIVOCAL**

**OFF**

**First Generation Biotransformations Only**

<table>
<thead>
<tr>
<th>Compound</th>
<th>Observed Biotransformations</th>
<th>Predicted Biotransformations</th>
<th>Unconfirmed Positive Biotransformations</th>
<th>Congruent Biotransformations</th>
<th>False Negative Biotransformations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Probable</td>
<td>Plausible</td>
<td>Equivocal</td>
</tr>
<tr>
<td>CP-122,721</td>
<td>4</td>
<td>25</td>
<td>2</td>
<td>4</td>
<td>15</td>
</tr>
<tr>
<td>Reparixin</td>
<td>4</td>
<td>8</td>
<td>0</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Imrecoxib</td>
<td>1</td>
<td>11</td>
<td>0</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>Tramadol</td>
<td>6</td>
<td>14</td>
<td>0</td>
<td>3</td>
<td>5</td>
</tr>
</tbody>
</table>
Biotransformation Analysis for 18 Compounds

- **Confirmed Positive**
  - Probable: 30%
  - Plausible: 50%
  - Equivocal: 20%

- **Unconfirmed Positive**
  - Probable: 49%
  - Plausible: 46%
  - Equivocal: 5%

**Probable**

**Plausible**

**Equivocal**
Mammalian Drug Metabolism In Silico
Strategies for Increasing Specificity of Prediction

Biotransformation Analysis for 18 Compounds

Unconfirmed Positive : Confirmed Positive = 3 : 1
Confirmed Positive : False Negative = 22 : 1
Mammalian Drug Metabolism In Silico
Strategies for Increasing Specificity of Prediction

The “acceptance” of UNCONFIRMED POSITIVE Predictions (I)
SMALL UNOBSERVABLE FRAGMENTS

\[
\begin{align*}
\text{R-N} & \rightarrow \left( \begin{array}{c}
\text{R-N} \\
\text{OH}
\end{array} \right) \\
\text{HO-} & \leftrightarrow \left( \begin{array}{c}
\text{H+} \\
\text{O}
\end{array} \right) \\
\text{HO-} & \rightarrow \left( \begin{array}{c}
\text{HO-} \\
\text{O}
\end{array} \right) \\
\text{HO-} & \leftrightarrow \left( \begin{array}{c}
\text{HO-} \\
\text{N}
\end{array} \right)
\end{align*}
\]
Mammalian Drug Metabolism In Silico
Strategies for Increasing Specificity of Prediction

The “acceptance” of UNCONFIRMED POSITIVE Predictions (II)
INTERMEDIATE METABOLITES
The “acceptance” of UNCONFIRMED POSITIVE Predictions (III) MULTIPLE BIOTRANSFORMATION PATHWAYS
The “acceptance” of UNCONFIRMED POSITIVE Predictions (IV)

EXPERIMENTAL LIMITATIONS

Pathway Characterisation
Metabolite Characterisation
Mass Balance Analysis
Assay Type, Cross Referencing and Extrapolation
Mammalian Drug Metabolism In Silico Strategies for Increasing Specificity of Prediction

Analysis of Unconfirmed Positive Biotransformations
Mammalian Drug Metabolism In Silico
Strategies for Increasing Specificity of Prediction

Analysis of Unconfirmed Positive Biotransformations

Biotransformation 70
Benzylic Hydroxylation

R1 = aromatic carbon
R2 = aliphatic carbon but cannot be the following group:

\[ \text{O} \quad \text{R3} \quad \text{R1 and R2 cannot be joined in a common ring} \]

Alicyclic Methylene Hydroxylation
Terminal Methyl Hydroxylation
Allylic Hydroxylation
Benzylic Methylene Hydroxylation
alpha-to-Carbonyl Hydroxylation
Alkyl Methine Hydroxylation
Benzylic Methine Hydroxylation
Penultimate Methylene Hydroxylation
Mammalian Drug Metabolism In Silico Strategies for Increasing Specificity of Prediction

Biotransformation 73
Hydroxylation of Terminal Methyl

R1 = acyclic aliphatic carbon
(not multiply-bonded and attached to
no more than one heteroatom)
The following exclusion applies:

R2 = aromatic carbon

R3, R4 = any atom in a ring
R5 = any atom
The atom (*) cannot be in a ring

Analysis of Unconfirmed Positive Biotransformations
Mammalian Drug Metabolism In Silico
Strategies for Increasing Specificity of Prediction
Mammalian Drug Metabolism In Silico
Strategies for Increasing Specificity of Prediction

Case Study: CITALOPRAM

N-Oxidation
N-Demethylation
N-Dealkylation
N-Glucuronidation

Carboaromatic Hydroxylation

Hydrolytic Ring-Opening
Oxidative Ring-Opening
Mammalian Drug Metabolism In Silico Strategies for Increasing Specificity of Prediction

Case Study: CITALOPRAM

Hydrolytic Ring-Opening

"Hydrolysis of Benzylic Ethers"
Hydrolytic Ring-Opening

"Hydrolysis of Benzylic Ethers"

R₁, R₂, R₃, R₄, R₅, R₆, R₇, R₈, R₉, R₁₀, R₁₁, R₁₂, R₁₃, R₁₄, R₁₅

R₁ = aromatic carbon
R₂ - R₆ = hydrogen, aliphatic or aromatic carbon
The ether can be cyclic but not an oxirane or oxetane

R₄ =

R₅ - R₈, R₁₁, R₁₂, R₁₄, R₁₅ =
aliphatic carbon (bearing at least two hydrogen atoms and not attached to a further heteroatom)
R₉, R₁₀, R₁₃, R₁₆ =
aliphatic carbon (bearing at least two hydrogen atoms and not attached to a heteroatom) or hydrogen
Mammalian Drug Metabolism In Silico
Strategies for Increasing Specificity of Prediction

Case Study: CITALOPRAM

Hydrolytic Ring-Opening
"Hydrolysis of Benzylic Ethers"

Chemical structures of diphenhydramine, orphenadrine, chlorphenoxamine, clemastine, and adamon.
Mammalian Drug Metabolism In Silico
Strategies for Increasing Specificity of Prediction

Case Study: CITALOPRAM

Oxidative Ring-Opening
"Oxidative O-Dealkylation"
Case Study: CITALOPRAM

Oxidative Ring-Opening
"Oxidative O-Dealkylation"

R1, R2 = aliphatic or aromatic carbon
The oxygen atom must be in a ring of size five or greater

The following exclusion also applies:

bond (*) is aromatic
Case Study: CITALOPRAM

Oxidative Ring-Opening
“Oxidative O-Dealkylation”

FURAMETPYR
Mammalian Drug Metabolism In Silico Strategies for Increasing Specificity of Prediction

Case Study: CITALOPRAM

**Oxidative Ring-Opening**

“Oxidative O-Dealkylation”

\[
\begin{align*}
\text{O} & \quad \text{H} \\
\text{O} & \quad \text{R1} \\
\text{R2} & \quad \text{R3} \\
\text{R3} & \quad \text{OH} \\
\text{R1} & \quad \text{R2}
\end{align*}
\]

R1, R2, R3 = aliphatic or aromatic carbon
The oxygen atom must be in a ring of size five or greater

The following exclusion also applies:

\[
\begin{align*}
\text{R4} & \quad \text{R4} = \text{any atom except hydrogen} \\
\text{H} & \quad \text{Bond (*) is aromatic}
\end{align*}
\]
Mammalian Drug Metabolism In Silico
Strategies for Increasing Specificity of Prediction

Case Study: CITALOPRAM

Oxidative Ring-Opening
“Oxidative O-Dealkylation”

DARIFENACIN

RAMELTEON
Mammalian Drug Metabolism In Silico
Strategies for Increasing Specificity of Prediction

Case Study: CITALOPRAM

Oxidative Ring-Opening
“Oxidative O-Dealkylation”

Molecules:
- Codeine
- Morphine
- Thebaine
- Galantamine
Mammalian Drug Metabolism In Silico
Strategies for Increasing Specificity of Prediction

Strengths/Weakness of Prediction System Types

HIGHEST PREDICTIVITY

- Statistical Methods
  - Neural Networks
  - Genetic Algorithms
  - Support Vector Machines
  - 2D QSAR
  - Decision Trees
  - Random Forest
  - 3D QSAR
  - Molecular Modelling
  - Quantum Mechanical
  - Systems Platforms
  - Databases
  - Knowledge-based Expert System

HIGHEST INTERPRETABILITY

Sites of METABOLISM
Prediction of METABOLITES & INTERMEDIATES
Generation of METABOLIC TREES (Mechanistic Richness)
Mammalian Drug Metabolism In Silico
Strategies for Increasing Specificity of Prediction

Complementary Systems: Conjoint Approach

Tramadol

Acenocoumarol

M445526

SMARTCyp Calculated Scores

SMARTCyp Calculated Scores

SMARTCyp Calculated Scores

Atom 1  Atom 2  Atom 3  Atom 4

Atom 1     Atom 2     Atom 3     Atom 4

Atom 1     Atom 2     Atom 3
“There are known knowns - these are things we know that we know. There are known unknowns - that is to say, there are things that we now know we don't know. But, there are also unknown unknowns - there are things we do not know we don't know!”

Donald Rumsfeld
Summary and Conclusions

CURRENT STATUS

- There are many different tools available -
  - use the right one(s) to answer the appropriate question(s)
- Many tools are complementary -
  - use the right combination of tools
- The best combination of tools may not answer all the questions –
  - use “smartly” in conjunction with expertise and experimental evidence
- Gaps in data and knowledge CAN be addressed

FUTURE PERSPECTIVES AND CHALLENGES

- Validation methods -
  - statistical methods
  - “fit-for-purpose”
- Biological differences –
  - reaction rates
  - quantification of metabolites
  - differences between species (what’s important in human?)
  - age, sex, disease state etc.
  - enzyme isoforms and genetic polymorphisms
- Higher level integration of complementary systems