

Predicting Metabolites

Enhancing An Expert System With Machine Learning

Chris Barber

Director Of Science

chris.barber@lhasalimited.org

Boston, May 2015



Leaders in the development of expert chemoinformatic systems
and trusted curators of proprietary data.

Lhasa Limited: What We Do

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

Nexus

File Window Prediction Tools Help

Hide

Predictions

- Study Folder
- Lasofixifene

Jobs

Structure

H C N O S P F Cl Br I

Problems 2/3 Structure Properties

Source	Problem	Severity



Meteor Nexus

The screenshot displays the Meteor Nexus software interface. The main window shows a hierarchical metabolic tree with various chemical structures. A blue callout box highlights the text: "Tabular summary of metabolic tree". Below the tree, a table provides a detailed summary of the metabolites. The table includes columns for 'Choose', 'Structure', 'Name', 'Intermediates', 'Formula', 'Parent For...', 'Query For...', 'Averag...', 'Exact ...', 'Nominal Mass', and 'Query Ma'. The first row is 'Lasofexifene (Query)'. The next three rows are highlighted in yellow: 'M7', 'M17', and 'M54'. The status bar at the bottom indicates 'Metabolites: 30 No filters applied'.

Choose	Structure	Name	Intermediates	Formula	Parent For...	Query For...	Averag...	Exact ...	Nominal Mass	Query Ma
<input type="checkbox"/>		Lasofexifene (Query)		C ₂₈ H ₃₁ N...			413.55	413.2355	413	
<input type="checkbox"/>		M7	I5a, I6o, I7o, I8o, I9o	C ₃₃ H ₃₈ N...	C ₂₈ H ₃₁ N...	-C ₅ H ₇ NO...	574.73	574.2501	574	161.01466
<input type="checkbox"/>		M17	I14o	C ₂₈ H ₂₉ N...	C ₂₈ H ₃₁ N...	-O+H ₂	427.54	427.2147	427	13.97926
<input type="checkbox"/>		M54	I118a, I119o, I120o, I121o, I122o	C ₃₃ H ₃₆ N...	C ₂₈ H ₂₉ N...	-C ₅ H ₅ NO...	588.71	588.2294	588	174.99393



Meteor Nexus

The screenshot displays the Meteor Nexus software interface. The main window shows a hierarchical tree of metabolic predictions. A blue callout box highlights the text: "Metabolic path for selected metabolite including identification of potentially adduct-forming and other intermediates". Below this, the "Metabolite Path" view shows the chemical structure of Lasofoxifene (Query) undergoing "243 Oxidative N-Dealkylation" to form an "Other" intermediate, which then leads to a "Potentially adduct-forming" intermediate. The interface includes a menu bar (File, Window, Prediction, Reports, Tools, Help), a toolbar, and a sidebar with "Predictions" and "Jobs" sections. The status bar at the bottom indicates "Metabolites: 30 No filters applied".

Metabolic path for selected metabolite including identification of potentially adduct-forming and other intermediates

Lasofoxifene (Query) → 243 Oxidative N-Dealkylation → Other → Potentially adduct-forming

Metabolites: 30 No filters applied



Meteor Nexus

The screenshot displays the Meteor Nexus software interface. On the left, there is a sidebar with a 'Predictions' section containing a 'Study Folder' and 'Lasofixifene' with a 'Meteor Prediction' icon. Below it is a 'Jobs' section with a grid. The main window shows a hierarchical tree diagram of metabolites. A large blue-bordered text box is overlaid on the tree, containing the text: 'Options to filter by mass, molecular formula etc'. At the bottom, a table lists various filter options. The 'Formula matches' filter is checked, showing a value of 'C24H22O4' and 2 matches. Other filters include Average Molecular Mass, Biotransformation name, Biotransformation number, Enzyme, Exact mass, Have intermediates, Likelihood, LogKp, LogP, and Metabolite comments. The status bar at the bottom indicates 'Metabolites: 30 Shown after filtering: 2 Filters active: 1'.

Name	Value	Tolera...	# Matched
<input type="checkbox"/> Average Molecular Mass			
<input type="checkbox"/> Biotransformation name			
<input type="checkbox"/> Biotransformation number			
<input type="checkbox"/> Enzyme			
<input type="checkbox"/> Exact mass			
<input checked="" type="checkbox"/> Formula matches	C24H22O4		2
<input type="checkbox"/> Have intermediates			
<input type="checkbox"/> Likelihood			
<input type="checkbox"/> LogKp			
<input type="checkbox"/> LogP			
<input type="checkbox"/> Metabolite comments			
<input type="checkbox"/> Name			

Metabolites: 30 Shown after filtering: 2 Filters active: 1



Meteor Nexus

The screenshot displays the Meteor Nexus software interface. The main window shows a hierarchical tree of biotransformation predictions. A blue callout box highlights the tree with the text "Biotransformation scope". The bottom panel, titled "Biotransformation Details", shows the reaction for "243: Oxidative N-Dealkylation - (Knowledge Base: Meteor KB 2014 1.0.0)".

Biotransformation Details

▼ 243: Oxidative N-Dealkylation - (Knowledge Base: Meteor KB 2014 1.0.0)

R1C(R2)N(R3)C(R1)C → R2N(R3)C(R1)O + R1C(=O)O + R1C(O)C

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

Phase: Phase I
Enzyme: CYP450
Species: Mammals

Metabolites: 30 No filters applied



Meteor Nexus

The screenshot displays the Meteor Nexus software interface. The main window shows a hierarchical metabolite path starting from Lasofoxifene. A blue callout box highlights the text "Biotransformation supporting information". Below this, a "Comments" panel provides detailed information on oxidative N-dealkylation, including its mechanism and examples. The interface also includes a "Jobs" table, a "Table View" button, and a status bar at the bottom indicating "Metabolites: 30 No filters applied".

Biotransformation supporting information

Comments

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 Stanojcic] and gallopamil [Mutlib 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. This first step may involve the intermediacy of an iminium cation. Bond scission results from hydrolysis of the initially formed carbinolamine intermediate. Carbinolamines are sometimes stable enough to be conjugated and detected in urine. Occasionally, oxidation of the carbinolamine 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 carbinolamine. For primary amines the mechanisms may be more complex and, as well as the carbinolamine, involve the intermediacy of the hydroxylamine, alpha-hydroxyhydroxylamine, primary imine, and oxime.

The carbon atom alpha to the reaction centre may be aromatic or multiply bonded (vinylic, propargylic etc.) but if it is fully saturated then quaternary

► References

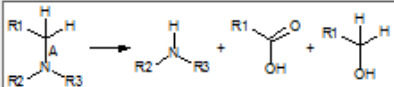
► Examples

Metabolites: 30 No filters applied

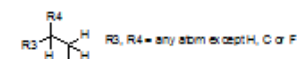
Dictionary Of Biotransformations

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

Biotransformation
✕

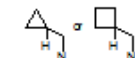


R1 =
aromatic carbon
saturated, aliphatic or allylic carbon
quaternary centres (except fluoromethyl) are not allowed
cyclic carbon
the following exclusion applies:

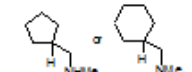


R3, R4 = any atom except H, C or F

Bicyclic carbon
this is allowed for unsubstituted cyclopropyl and cyclobutyl only:



substituted cyclopropyl/cyclobutyl and rings of size five or greater are allowed if the nitrogen atom is part of an methyl aminomethyl or dimethyl aminomethyl group:



R2, R3 = any atom (cannot both be hydrogen)
bond A must not be in a ring
amides, ureas, imides, N-nitrosamines, N-nitrosamides
and related compounds are excluded from this biotransformation

Enabled Biotransformation ID

Created 14/05/02 15:08 Last Modified 16/02/14 12:57

Biotransformation Name

Biotransformation Number

Species

Phase 1

Enzyme 30

Comments
Rule Writer Comments

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 Stanojic] and gallopamil [Mutlib 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. This first step may involve the intermediacy of an iminium cation. Bond scission results from hydrolysis of the initially formed carbinolamine intermediate. Carbinolamines are sometimes stable enough to be conjugated and detected in urine. Occasionally, oxidation of the carbinolamine 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

Go to
◀◀ 199/517 ▶▶

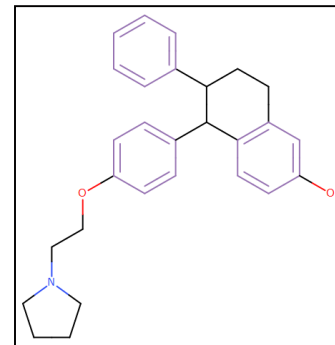


How Meteor Nexus Works

Knowledge base

Dictionary of
biotransformations

Rule base



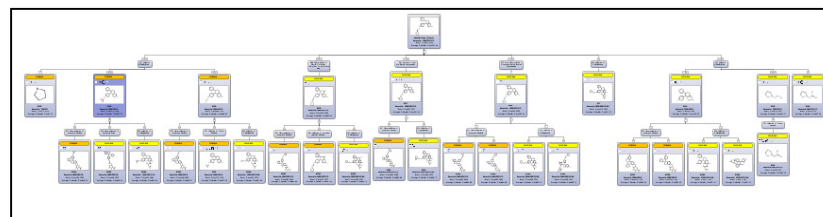
What reactions **could** occur?



How **likely** that each
reaction **will** occur?



Processing constraint





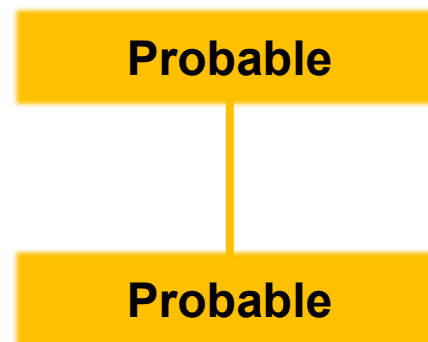
Rule Base

- 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





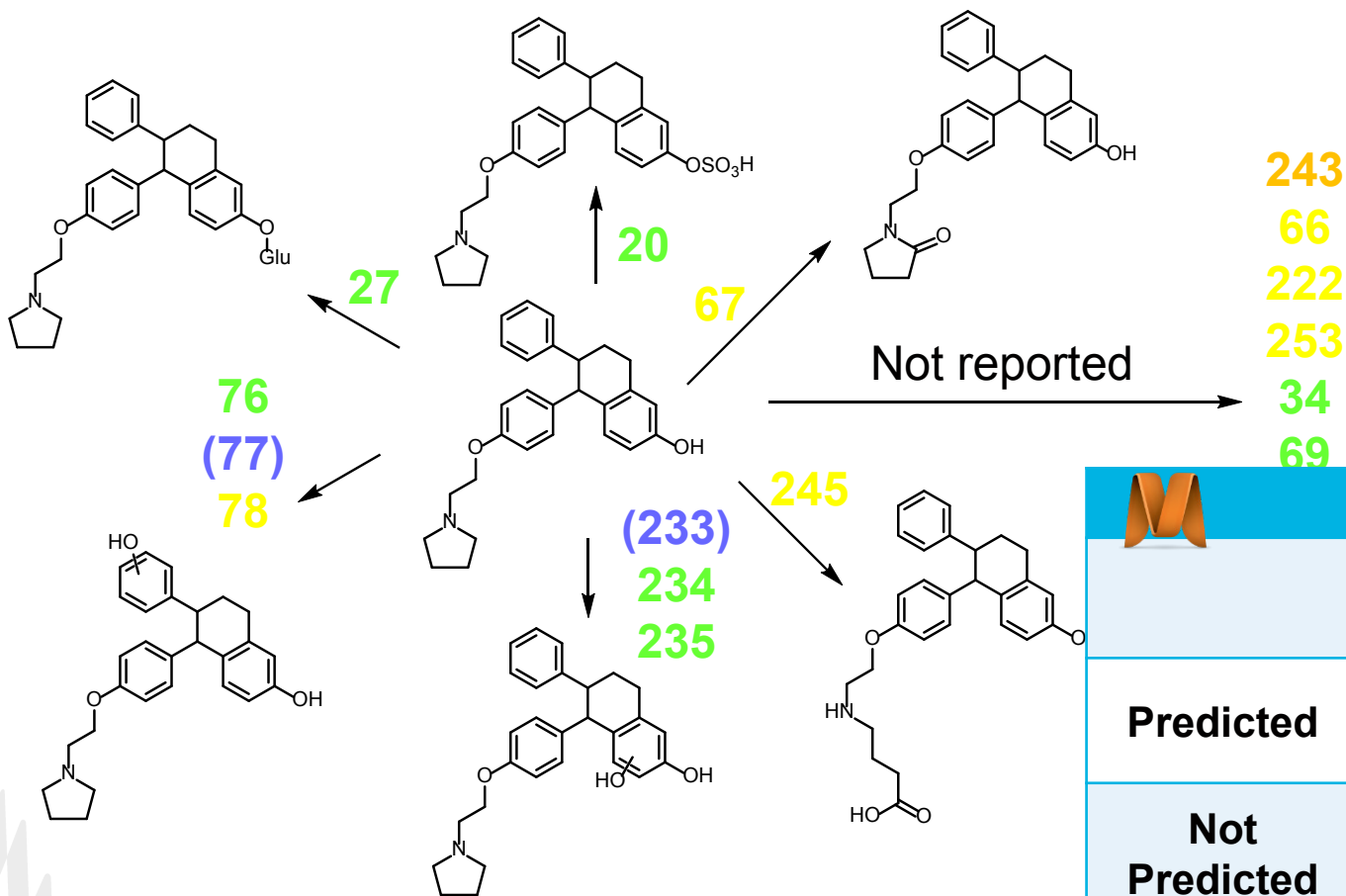
Meteor Nexus Performance

- 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

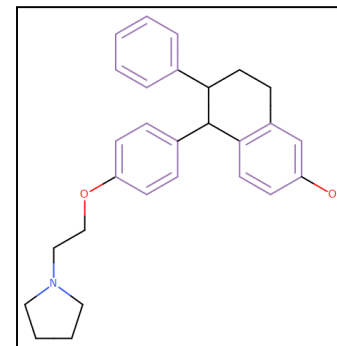
- Man|rat|monkey, first generation metabolites

Probable
Plausible
Equivocal ←
Doubted
Improbable



	Metabolite	
	Observed	Not Observed
Predicted	8	9
Not Predicted	2	

How Meteor Nexus Could Work



Knowledge base

Dictionary of
biotransformations

Expert
system



What reactions **could** occur?



How **likely** that each
reaction **will** occur?

Machine
learning

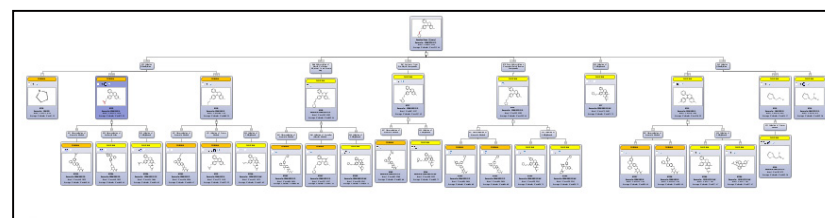


Processing constraint



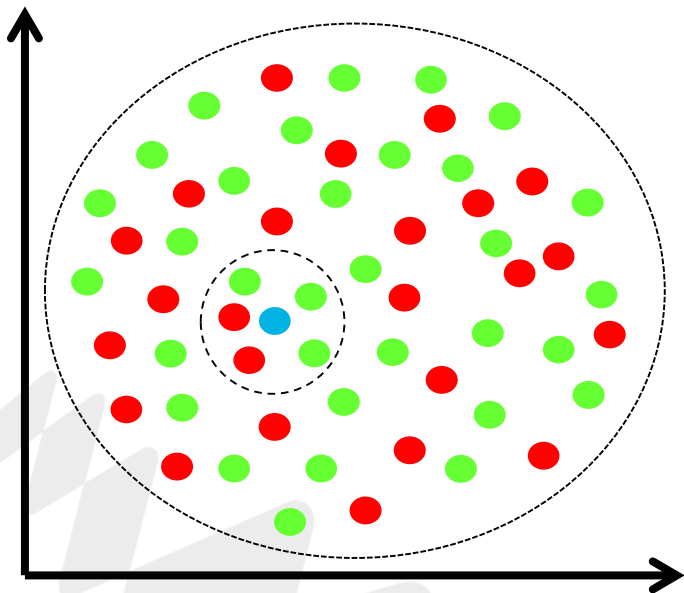
Database

Experimental
reactions



Other statistical approaches to metabolite ranking

- SyGMa
 - L Ridder & M Wagener, ChemMedChem 3 821-832 (2008)
- MetaPrint2D-React
 - SE Adams, Molecular Similarity and Xenobiotic Metabolism, PhD Thesis, University of Cambridge (2010)



UNIVERSITY OF CAMBRIDGE

MetaPrint2D-React metabolic product predictor

University of Cambridge · Department of Chemistry · Centre for Molecular Informatics

Instructions

Click on an atom to display the predicted reactions at that site, and click on one of these reaction types to allow the metabolite formed. Alternatively, select a reaction type from the filter box below the results page, to limit predictions to that reaction type.

The colour highlighting an atom indicates its normalised occurrence ratio (NOR). A high NOR indicates a more frequently reported site of metabolism in the metabolite database.

Note: The normalised occurrence ratio does not indicate how likely a molecule is to be metabolised, but rather the relative likelihood of metabolism occurring at a particular site in the molecule, assuming it is metabolised.

Red	0.88 <= NOR <= 1.00
Orange	0.55 <= NOR < 0.88
Green	0.15 <= NOR < 0.55
White	0.00 <= NOR < 0.15
Grey	Little/no data

Results

Metabolite

Reaction type filter:

Reaction type: Glucuronidation

Links

- Home
- New Query
- Predict sites of metabolism
- About MetaPrint2D
- sourceforge project
- bioRxiv plugin

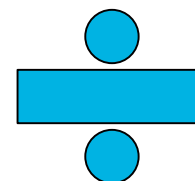
Occurrence Ratio Method



Large
metabolism database



How often does a reaction
actually occur?

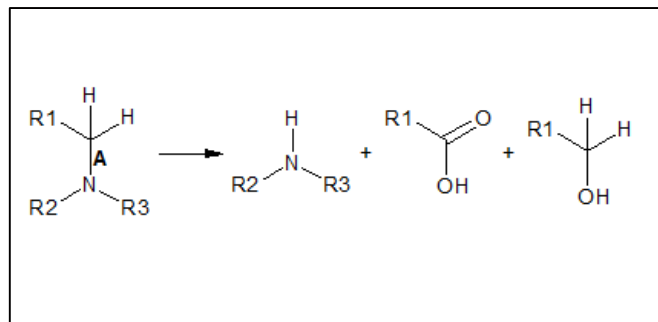


Occurrence Ratio



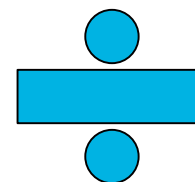
How often could a reaction
occur?

Occurrence Ratio Method: Biotransformation 243



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

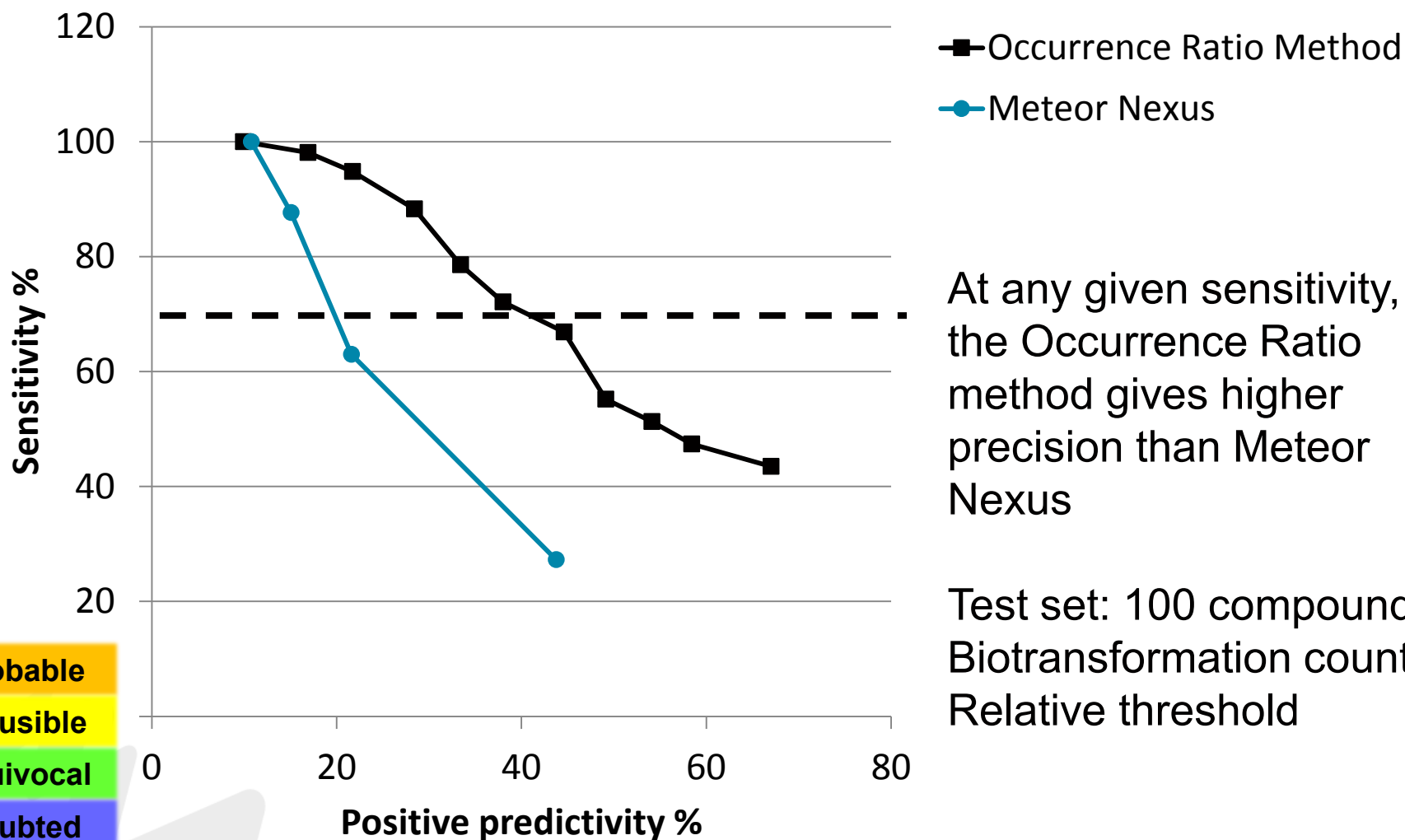
Green bar: the biotransformation has Selected supporting examples containing the biophore for the biotransformation

Supporting Examples

<97.1% : 2>	<94.2% : 2>	<63.7% : 27>	<63.1% : 34 67 111...>
<62.3% : 23>	<59.7% : 67 78 81>	<56.8% : 118>	<56.6% : 243 284>

Overview Search Prediction Validation Structure Descriptor Analyzer Experimental MultiGenerationPrediction

Occurrence Ratio Method Versus Meteor Nexus



Probable

Plausible

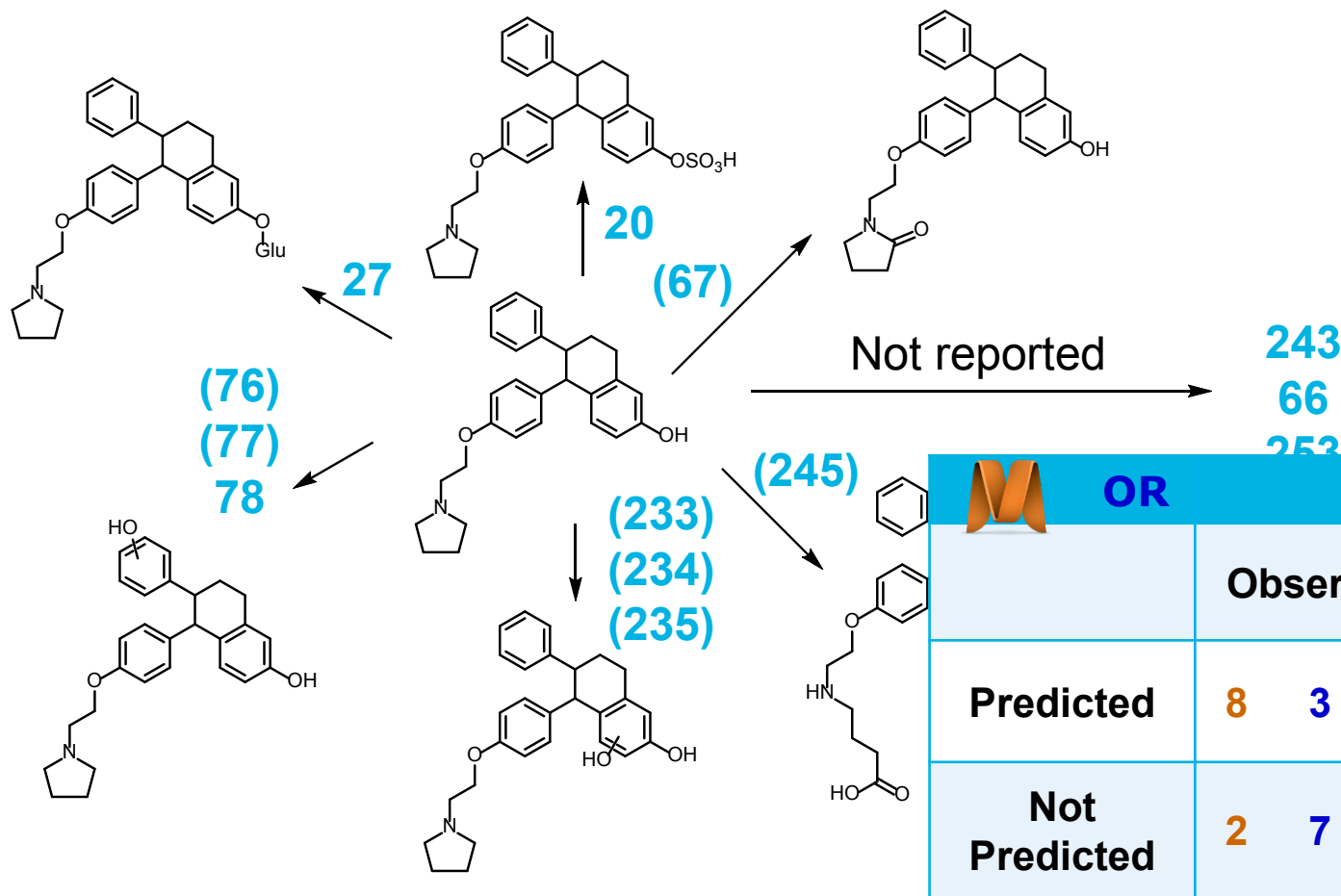
Equivocal

Doubted

Improbable

Lasofoxifene: Occurrence Ratio Prediction

- Man|rat|monkey, first generation metabolites



OR	Metabolite			
	Observed		Not Observed	
Predicted	8	3	9	3
Not Predicted	2	7		

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?

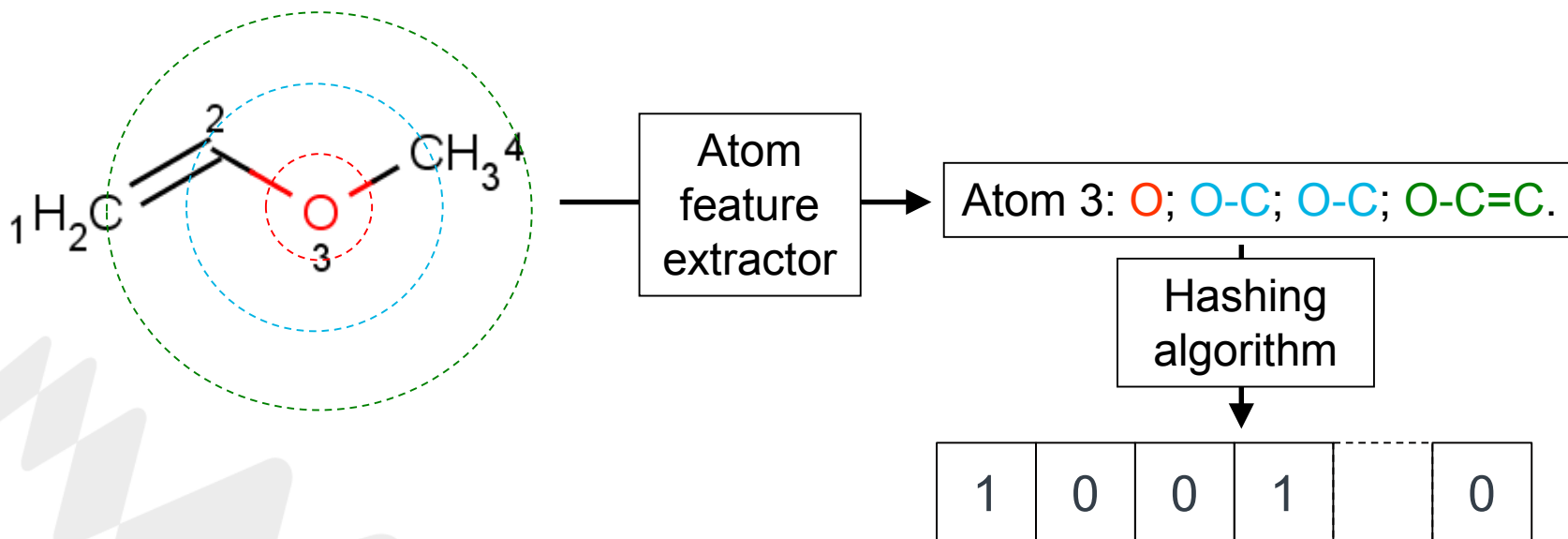
Similarity-based Occurrence Ratios

- Meteor biotransformation structural key

Biotransformation: 1 2 3 4 563

1	0	0	1		0
---	---	---	---	--	---

- Ceres fingerprint (whole structure)
- Ceres fingerprint (site of metabolism)



Site of Metabolism-driven Occurrence Ratios

The screenshot displays the Coralie software interface. The main window shows a chemical structure of a complex molecule with a cyclopentane ring, a benzene ring, and a hydroxyl group. The interface includes a 'Results' panel on the right, a 'Supporting Examples' panel below it, and a 'Configure' and 'Predict' section on the left.

Results

Alert ...	Biotransformation	Score (%)
027	Glucuronidation of Aromatic Alcohols	37.4
235	5-Hydroxylation of 1,2,4-Trisubstituted Benzenes	37.1
067	Lactams from Aza-Alicyclic Compounds	25.0
234	6-Hydroxylation of 1,2,4-Trisubstituted Benzenes	25.0
020	O-Sulphation of Aromatic Alcohols	24.9

Display duplicate SOMs

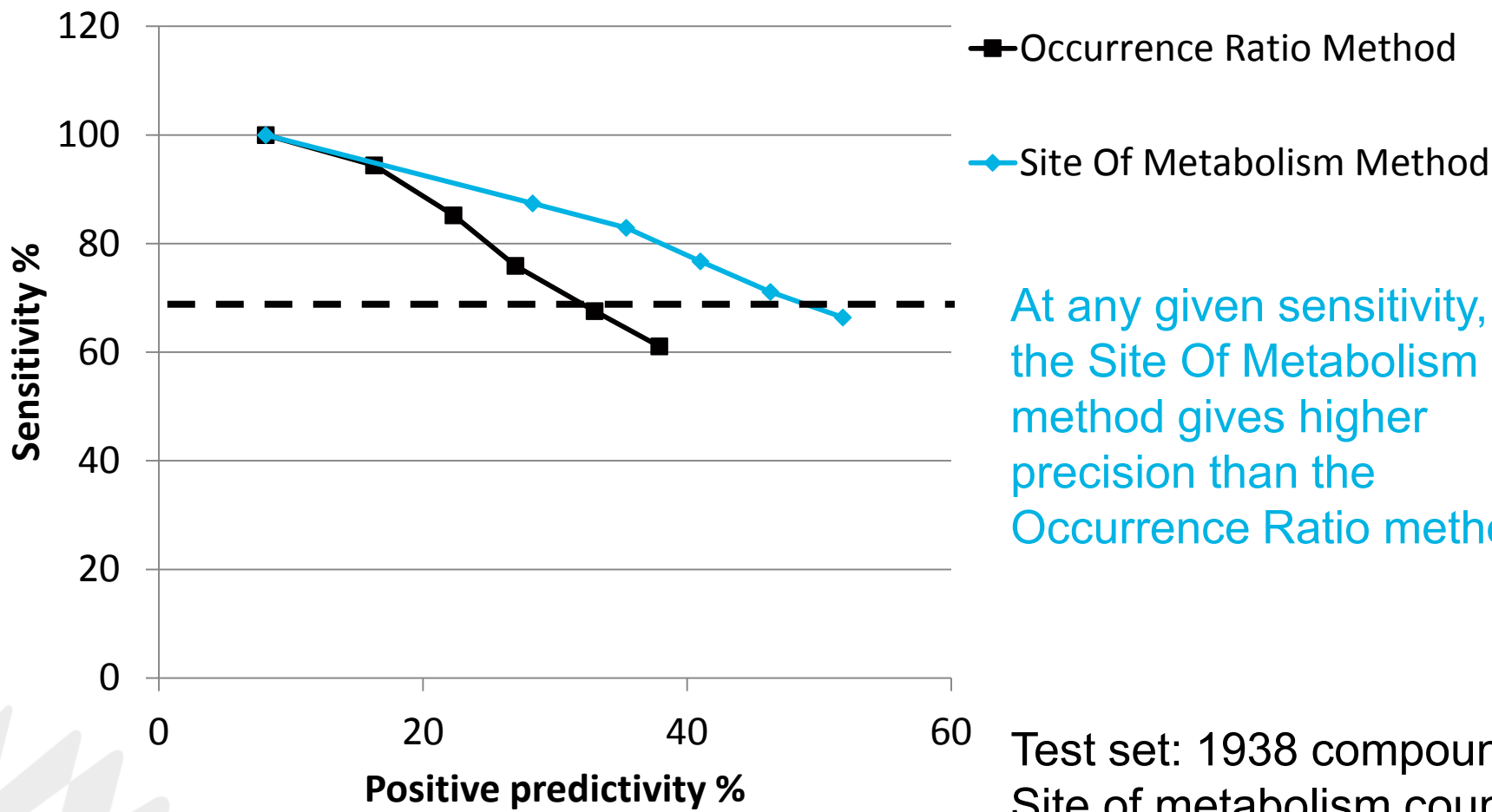
Supporting Examples

<90.9% : 2>	<90.9% : 2>	<88.1% : 89>	<88.1% : 20 27 69..>
<88.1% : 168>	<88.1% : 27 81 235>	<88.1% : 19 20 23..>	<88.1% : 61>

Overview Search Prediction Validation Structure Descriptor Analyzer Experimental MultiGenerationPrediction

- 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

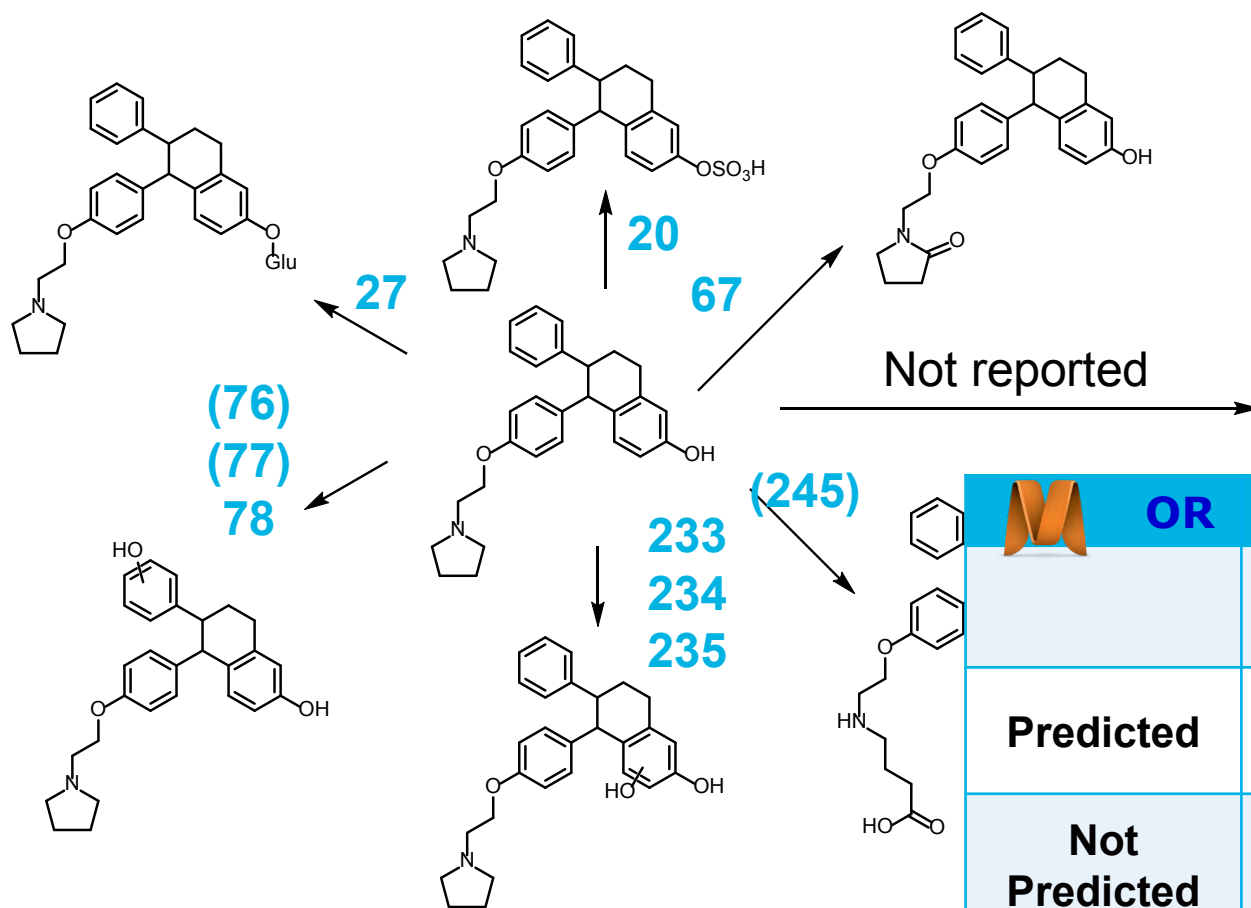


At any given sensitivity, the Site Of Metabolism method gives higher precision than the Occurrence Ratio method

Test set: 1938 compounds
Site of metabolism counts
Relative threshold

Lasofoxifene: Site Of Metabolism Prediction

- Man|rat|monkey, first generation metabolites



	OR SOM Metabolite					
	Observed			Not Observed		
Predicted	8	3	7	9	3	0
Not Predicted	2	7	3			

Extending Predictions To Multiple Generations

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

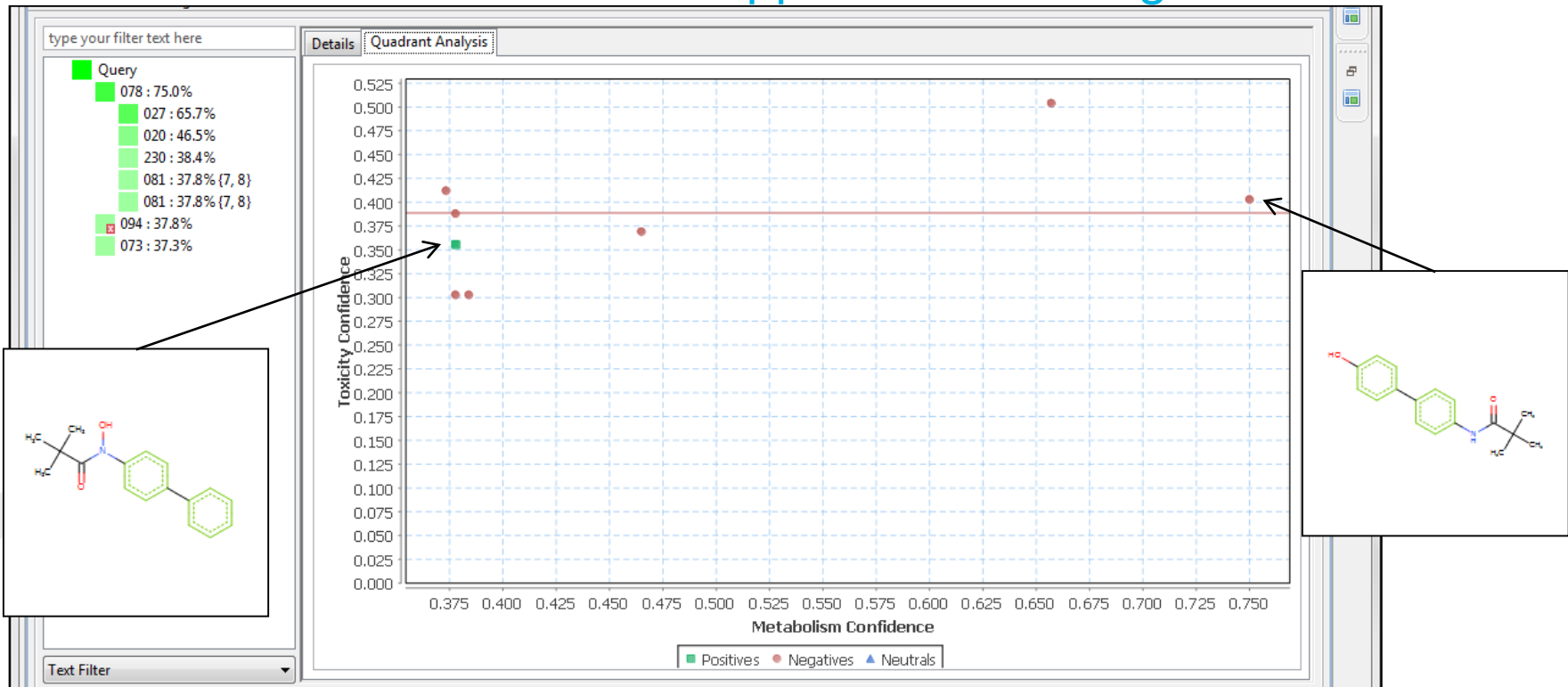
The screenshot shows the Coralie software interface. On the left, a query list is displayed with a blue box highlighting the top two entries: 078 : 49.4% and 233 : 18.5%. In the center, a chemical structure is shown. On the right, a 'Results' table is visible, listing biotransformation alerts and their scores. A blue box highlights the first row of the results table: 233 3-Hydroxylation of 1,2,4-Trisubstituted Benzenes 18.5. Below the results table, two chemical structures are shown with red boxes highlighting their respective scores: <63.4% : 168> and <63.4% : 27|81|235>.

Alert ...	Biotransformation	Score (%)
233	3-Hydroxylation of 1,2,4-Trisubstituted Benzenes	18.5
027	Glucuronidation of Aromatic Alcohols	18.5
235	5-Hydroxylation of 1,2,4-Trisubstituted Benzenes	18.3

First generation: 49.4% x
 Second generation: 7.5% = 18.5%
 30% Relative threshold 49.4% x 0.3 = 14.8%

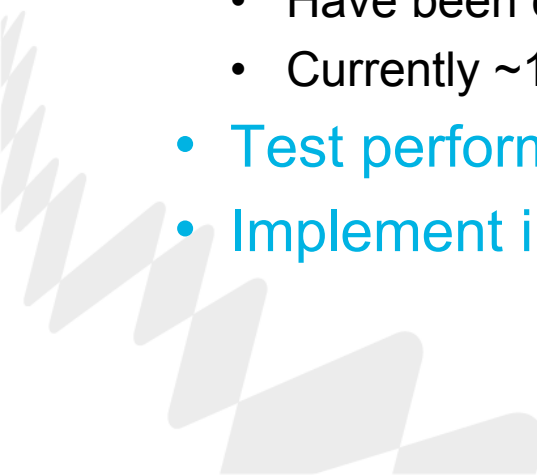
Summary

- 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



shared **knowledge** • shared **progress**

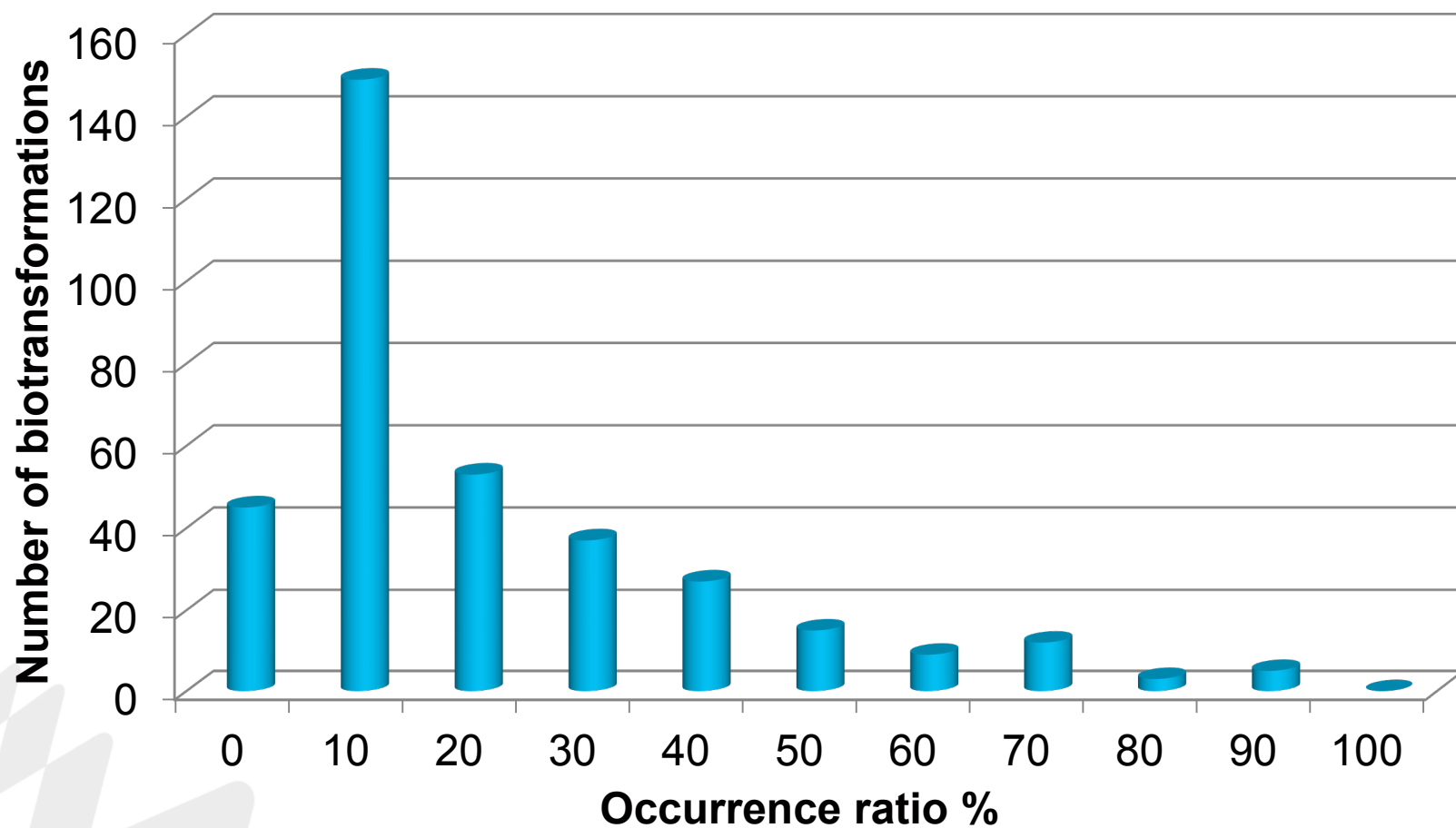
Lhasa Limited Registered Office
Granary Wharf House, 2 Canal Wharf, Leeds LS11 5PS
UK Registered Charity (290866)

+44 (0)113 394 6020
info@lhasalimited.org
www.lhasalimited.org

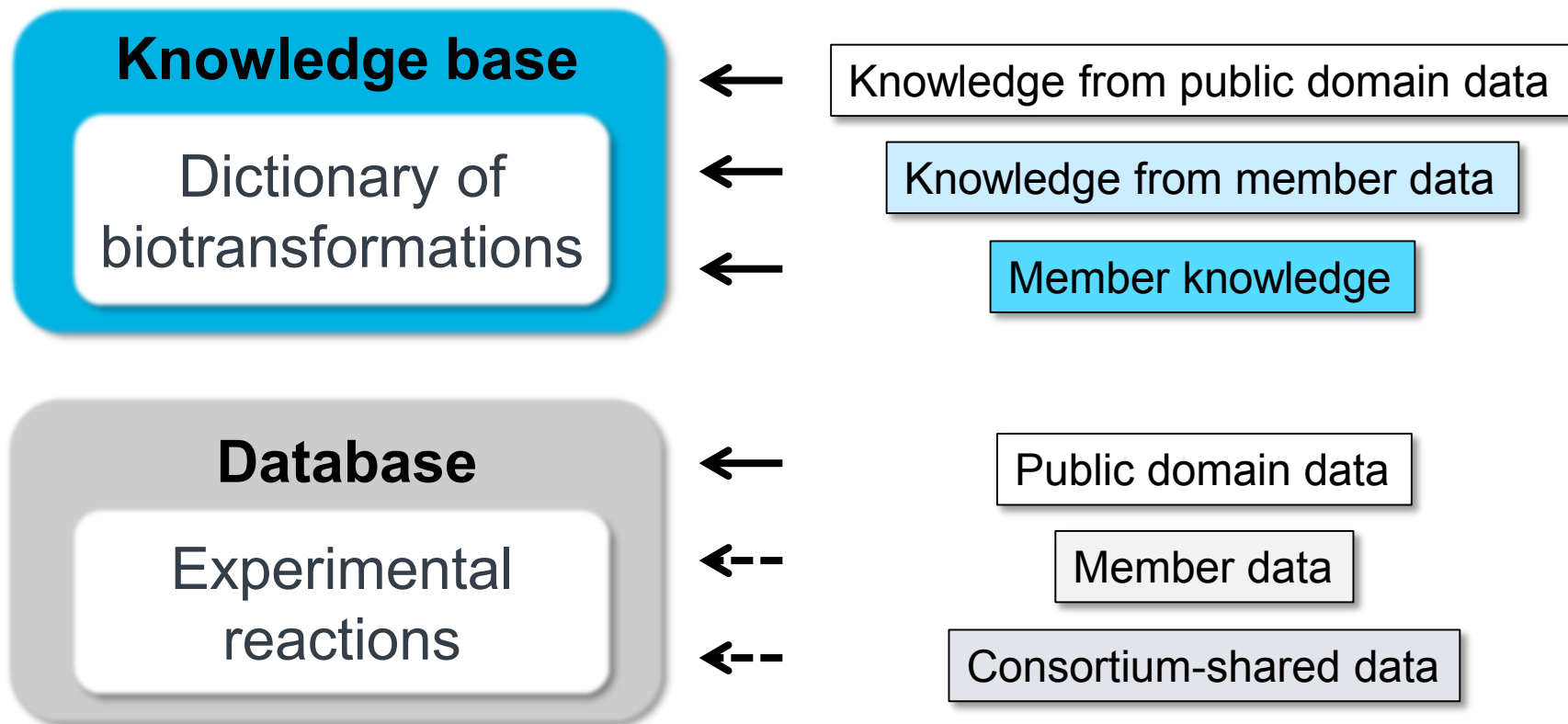


Distribution Of Training Set Occurrence Ratios

Omits 145 biotransformations with rare biophores



Meteor Nexus Data And Knowledge Sharing



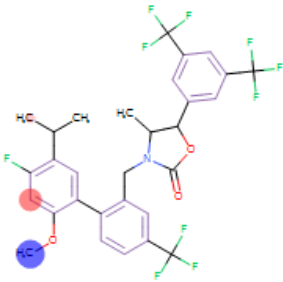
Site Of Metabolism View

Coralie

File Window Help

TrainingData.mdl

/Demonstration/TrainingData.mdl

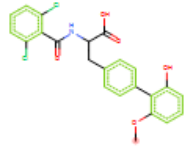
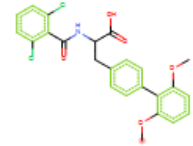
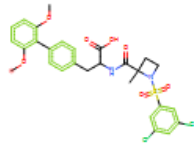
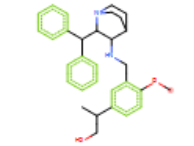
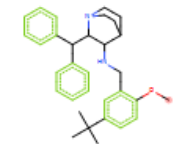
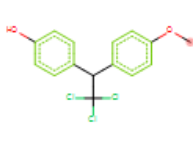
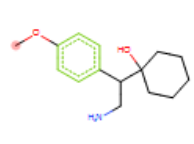
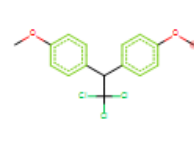


Results

Alert ...	Biotransformation	Score (%)	Toxicity	Tox Score (%)
118	Oxidative O-Demethylation	75.5	NEGATIVE	57.2
240	3-Hydroxylation of 1,2,4,5-Tetrasubstituted...	60.9	NEGATIVE	43.1
073	Hydroxylation of Terminal Methyl	26.5	NEGATIVE	40.6

Display duplicate SOMs

Supporting Examples

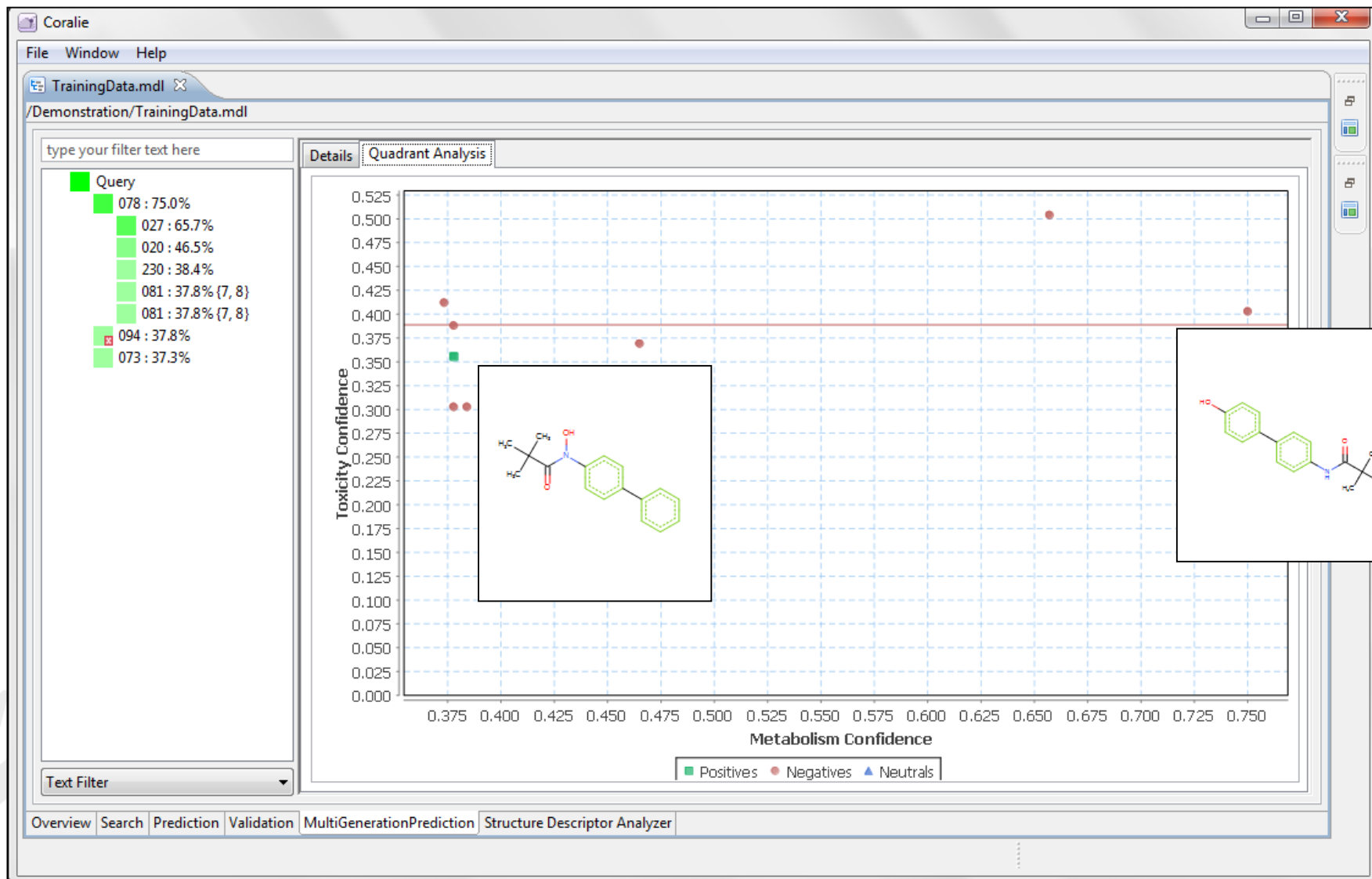
<61.0% : 20 118>	<59.5% : 81 118 231>	<59.5% : 29 118>	<52.4% : 432>
			
<51.2% : 73>	<50.0% : 20 27 81..>	<50.0% : 118 415>	<50.0% : 81 118 1..>
			

Configure Predict

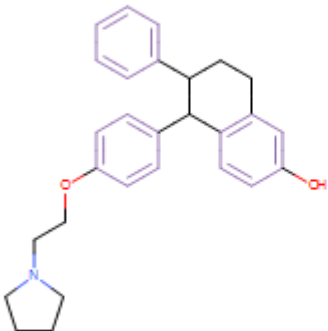
1

Overview Search Prediction Validation MultiGenerationPrediction Structure Descriptor Analyzer

Metabolite Toxicity View



Threshold Definitions



Results

Alert ...	Biotransformation	Score (%)
078	Para Hydroxylation of Monosubstituted Benzene Co...	49.4
233	3-Hydroxylation of 1,2,4-Trisubstituted Benzenes	37.5
027	Glucuronidation of Aromatic Alcohols	37.4
235	5-Hydroxylation of 1,2,4-Trisubstituted Benzenes	37.1
067	Lactams from Aza-Alicyclic Compounds	25.0

Display duplicate SOMs

- 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 (eg 60% of 49.4% = 29.6%)

Query-specific Occurrence Ratios

The screenshot displays the Coralie software interface. The main window shows a query molecule structure on the left and a 'Results' panel on the right. The 'Results' panel lists several alerts, with 'O-Sulphation of Aromatic Alcohols' (Alert 020) highlighted. Below the results is a 'Supporting Examples' section showing a grid of chemical structures with their respective occurrence ratios. A blue box highlights the 'Supporting Examples' section and the 'O-Sulphation of Aromatic Alcohols' result.

Results

Alert ...	Biotransformation
027	Glucuronidation of Aromatic Alcohols
235	5-Hydroxylation of 1,2,4-Trisubstituted Benzene
067	Lactams from Aza-Alicyclic Compounds
234	6-Hydroxylation of 1,2,4-Trisubstituted Benzene
020	O-Sulphation of Aromatic Alcohols

Display duplicate SOMs

Supporting Examples

<90.9% : 2>	<90.9% : 2>	<88.1% : 89>	<88.1% : 20 27 69..>
<88.1% : 168>	<88.1% : 27 81 235>	<88.1% : 19 20 23..>	<88.1% : 61>

Chemical Structures:

- Query molecule: Oc1ccc(cc1C2CCc3ccccc3C2)OCCN4CCCC4
- Supporting Example 1: Oc1ccc(cc1C2CCc3ccccc3C2)OS(=O)(=O)O
- Supporting Example 2: Oc1ccc(cc1C2CCc3ccccc3C2)O
- Supporting Example 3: Oc1ccc(cc1C2CCc3ccccc3C2)O
- Supporting Example 4: Oc1ccc(cc1C2CCc3ccccc3C2)O
- Supporting Example 5: Oc1ccc(cc1C2CCc3ccccc3C2)O
- Supporting Example 6: Oc1ccc(cc1C2CCc3ccccc3C2)O
- Supporting Example 7: Oc1ccc(cc1C2CCc3ccccc3C2)O
- Supporting Example 8: Oc1ccc(cc1C2CCc3ccccc3C2)O

Navigation: Overview | Search | Prediction | Validation | Structure Descriptor Analyzer | Experimental | MultiGenerationPrediction

k-Nearest neighbour methodology (k = 8)

Query-specific Occurrence Ratios

Coralie

File Window Help

TrainingData.mdl X

/Demonstration/TrainingData.mdl

Results

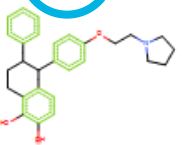
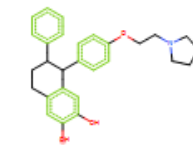
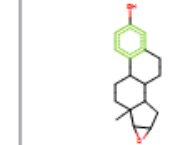
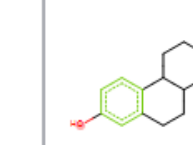
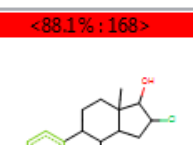
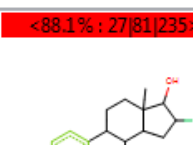
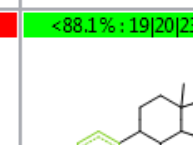
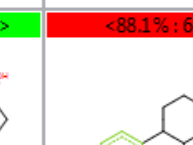
Results

Similarity between query and example substrates determined by Tanimoto index

Display duplicate SOMs

Configure Predict

Supporting Examples

<90.9% : 2>	<90.9% : 2>	<88.1% : 89>	<88.1% : 20 27 69..>
			
<88.1% : 168>	<88.1% : 27 81 235>	<88.1% : 19 20 23..>	<88.1% : 61>
			

Overview Search Prediction Validation Structure Descriptor Analyzer Experimental MultiGenerationPrediction

Detailed description: The image shows a screenshot of the Coralie software interface. The main window displays a chemical structure of a query molecule, which is a complex polycyclic system with a phenyl ring, a piperidine ring, and a hydroxyl group. This structure is circled in blue. Below the query structure are 'Configure' and 'Predict' buttons. To the right, a 'Results' panel shows a text box with the text 'Similarity between query and example substrates determined by Tanimoto index', also circled in blue. Below the results panel is a 'Supporting Examples' section, which is a grid of chemical structures. The first two structures in the first row are circled in blue. Each structure in the grid is accompanied by a similarity ratio and a count in angle brackets. The first two structures in the first row have a similarity of <90.9% and a count of 2. The third structure has a similarity of <88.1% and a count of 89. The fourth structure has a similarity of <88.1% and a count of 20|27|69.. The second row contains four structures with similarity ratios of <88.1% and counts of 168, 27|81|235, 19|20|23.., and 61. The bottom of the interface shows a navigation bar with tabs for Overview, Search, Prediction, Validation, Structure Descriptor Analyzer, Experimental, and MultiGenerationPrediction.

Query-specific Occurrence Ratios

Scores are weighted according to the (non-)observation of the biotransformation according to $\sqrt{\text{similarity}}$


The screenshot shows the Coralie software interface. At the top, there is a menu bar with 'File', 'Window', and 'Help'. Below it, a tab labeled 'TrainingData.mdl' is open, showing the path '/Demonstration/TrainingData.mdl'. A large text box on the left contains the text: 'Scores are weighted according to the (non-)observation of the biotransformation according to $\sqrt{\text{similarity}}$ '. In the center, a list of biotransformation types is displayed, with a table showing the following data:

Bioreaction ID	Bioreaction Name	Score (%)
024	O-Hydroxylation of 1,2,4-Trisubstituted Benzenes	25.0
020	O-Sulphation of Aromatic Alcohols	24.9
004	Aromatic Hydroxylation	21.2

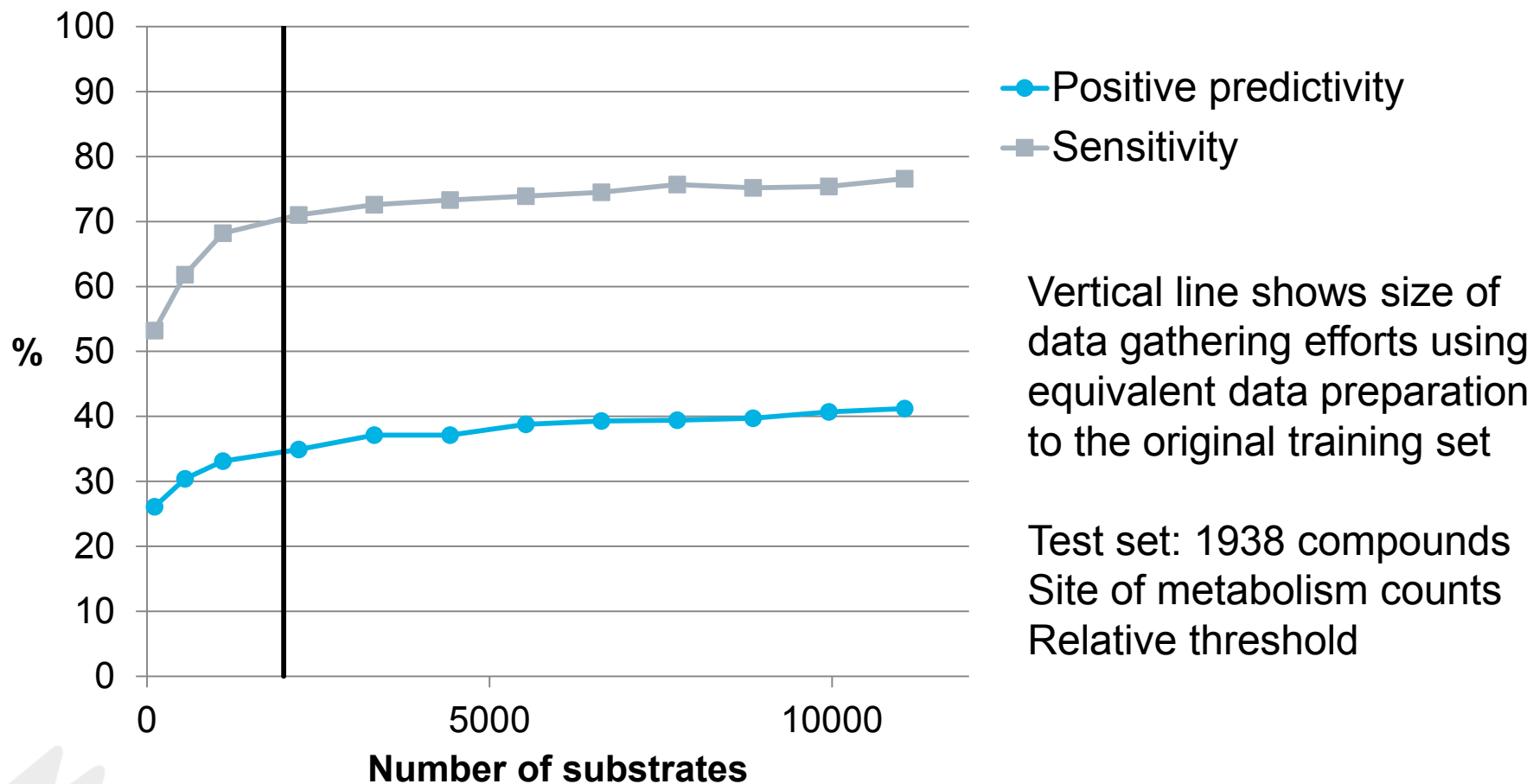
The 'O-Sulphation of Aromatic Alcohols' entry is highlighted in blue, and its score '24.9' is circled in blue. Below the list, there is a checkbox labeled 'Display duplicate SOMs'. To the right, a 'Supporting Examples' section displays a grid of chemical structures. The first row of examples has headers: '<90.9% : 2>', '<90.9% : 2>', '<88.1% : 89>', and '<88.1% : 20|27|69..>'. The first two headers are circled in blue. The second row of examples has headers: '<88.1% : 168>', '<88.1% : 27|81|235>', '<88.1% : 19|20|23..>', and '<88.1% : 61>'. The interface also includes a 'Configure' button, a 'Predict' button, and a large chemical structure viewer on the left showing a complex molecule with a sulfonamide group. At the bottom, there is a navigation bar with tabs for 'Overview', 'Search', 'Prediction', 'Validation', 'Structure Descriptor Analyzer', 'Experimental', and 'MultiGenerationPrediction'.



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





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.

