

Combining an Expert System with Machine Learning to Rank Metabolites

Edward Rosser

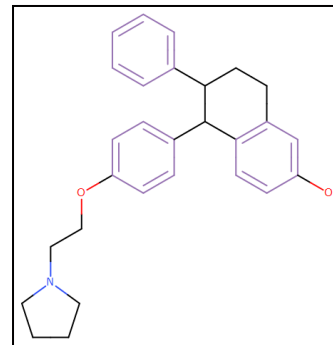


How Meteor Nexus Works

Knowledge base

Dictionary of biotransformations

Rule base



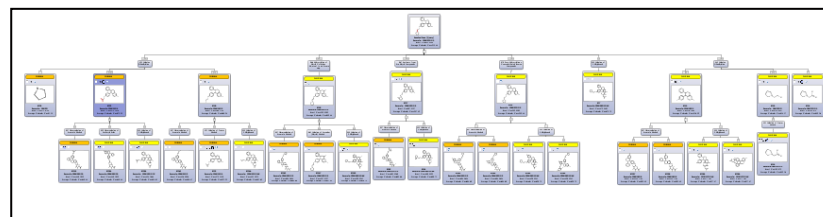
What reactions **could** occur?



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



Processing constraint



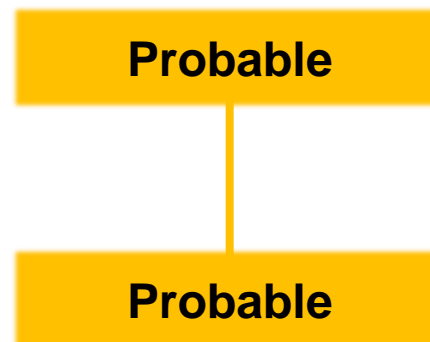
Rule Base

- 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

Occurrence Ratio

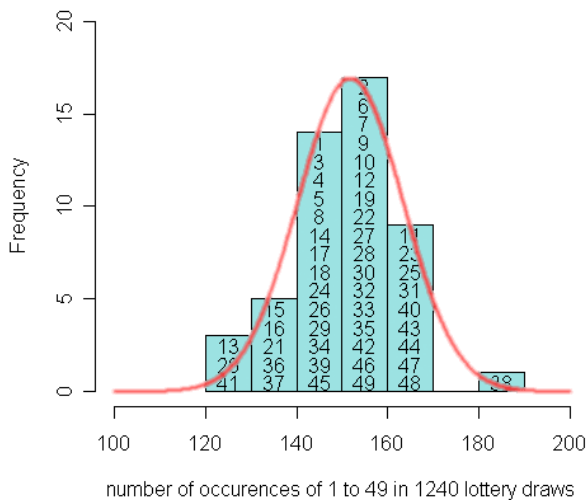
Tonight's
1240th National
Lottery draw



Number 13!
Drawn **125** times as a
main ball. Last outing
was two weeks ago
and this is its first
appearance this year..

$$\text{Occurrence Ratio} = \frac{\text{No. of times ball selected}}{\text{Total number of draws}} = \frac{125}{1240} = 0.10$$

Occurrence Ratio Examples

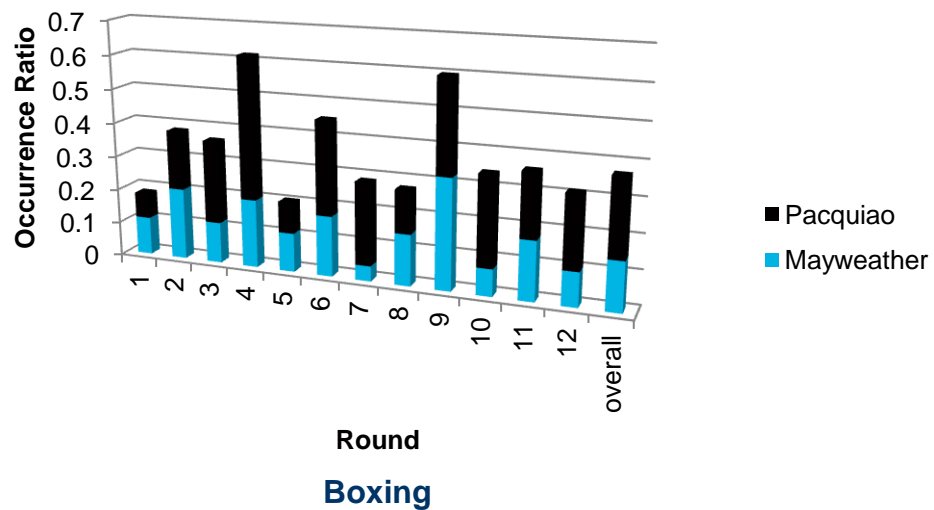


National Lottery

Player	Balls	Wkts	Strike Rate
SP Jones (Eng)	612	18	34
RT Ponting (Aus)	36	1	36
SK Warne (Aus)	1517	40	37.9
GD McGrath (Aus)	804	19	42.3
MJ Hoggard (Eng)	733	16	45.8
A Flintoff (Eng)	1164	24	48.5
SJ Harmison (Eng)	967	17	56.8
B Lee (Aus)	1147	20	57.3

2005 Ashes Bowling Figures

**Mayweather vs Pacquiao
Successfully Landed Punches**



Other Statistical Approaches to Metabolite Ranking

- MetaPrint2D-React

- ⊙ SE Adams, Molecular Similarity and Xenobiotic Metabolism, PhD Thesis, University of Cambridge (2010)

- 1. Detect and fingerprint reaction centres
 2. Generate occurrence ratio for each reaction centre.

- SyGMa

- ⊙ L Ridder & M Wagener, ChemMedChem 3 821-832 (2008)

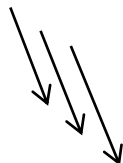
- 1. Generate library of biotransformations
 2. Generate occurrence ratio for each biotransformation.

Generating Occurrence Ratios

Meteor KB



BTs



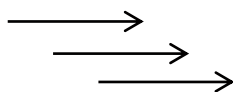
Occurrence
Ratio =

Number of times the BT is
experimentally observed

Number of times the BT
is activated in Meteor



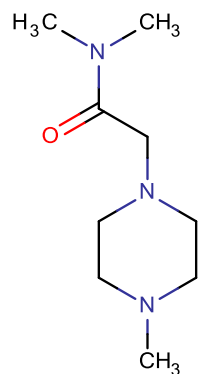
Metabolism DB



**Meteor
Matcher**

Biotransformation	Name	Activated	Observed	Ratio
267	Dihydrodiols from Arene Oxides	17	15	0.88
486	Deamination of Cytosine Nucleosides and Analo...	23	20	0.87
144	Hydrolysis of Acyclic Carboxylic Esters	1364	1129	0.83
123	Aromatisation of 1,4-Dihydropyridines	52	43	0.83
147	Hydrolysis of Phosphate Esters	190	152	0.80
453	S-Oxidation of Dibenzothiophenes and Heteroc...	5	4	0.80
134	Reduction of Triazeno Compounds	4	3	0.75
348	3-Hydroxylation of Pyridazines	4	3	0.75
459	Ring Contraction of 2,2,6,6-Tetramethylpiperid...	4	3	0.75
148	Hydrolysis of Phosphorothionate Ester	23	17	0.74
363	Lactonisation of delta-Hydroxy Carboxylic Acids	15	11	0.73
297	Hydrolysis of Cyclic Thioesters	7	5	0.71
062	Oxidation of Aldehydes	230	161	0.70
089	Vicinal Diols from Epoxides	164	114	0.70
003	O-Methylation of Pyrogallols	19	13	0.68
455	Hydrolysis of Carbonate Esters	59	40	0.68
107	Sulphone Formation from Sulphoxides	152	102	0.67

Occurrence Ratios – Making a Prediction



Meteor
Matcher



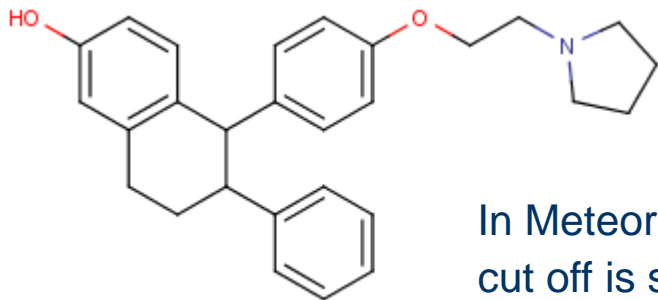
067
242 243 034
285 392 445
249 100



Rank using
Occurrence Ratios

Alert ID	Biotransformation	Score (%)
242	Oxidative N-Demethylation	50.7
285	Amine Oxides from Tertiary Alicyclic Amines	35.4
243	Oxidative N-Dealkylation	29.3
392	Hydrolysis of Acyclic Carboxylic Amides	28.0
249	Oxidative Ring Opening of Piperazines	18.1
067	Lactams from Aza-Alicyclic Compounds	9.7
100	Amine Oxides from Tertiary Alicyclic Amines	6.8
445	Glutathione Conjugation of Tertiary Aza-Alicyclic Compou...	5.7
034	Glucuronidation of Tertiary Aliphatic Amines	2.7

Lasofoxifene: Meteor Nexus Prediction



In Meteor Nexus the biotransformation cut off is set using the reasoning level.

BT	Name	Reasoning Level
243	Oxidative N-Dealkylation	PROBABLE
66	Hydroxylation of Alicyclic Methylene Adjacent to Aromatic Ring	PLAUSABLE
222	Oxidation of 4-Alkylphenols	PLAUSABLE
253	Oxidative O-Dealkylation	PLAUSABLE
245	Oxidative N-Dealkylation	PLAUSABLE
67	Lactams from Aza-Alicyclic Compounds	PLAUSABLE
78	Para Hydroxylations of Monosubstituted Benzene Compounds	PLAUSABLE
76	Ortho Hydroxylations of Monosubstituted Benzene Compounds	EQUIVOCAL
27	Glucuronidation of Aromatic Alcohols	EQUIVOCAL
20	O-Sulphation of Aromatic Alcohols	EQUIVOCAL
234	6-Hydroxylation of 1,2,4-Trisubstituted Benzenes	EQUIVOCAL
235	5-Hydroxylation of 1,2,4-Trisubstituted Benzenes	EQUIVOCAL
34	Glucuronidation of Tertiary Aliphatic Amines	EQUIVOCAL
69	Hydroxylation of Unfunctionalised Alicyclic Methylene	EQUIVOCAL
233	3-Hydroxylation of 1,2,4-Trisubstituted Benzenes	DOUBTED
77	Meta Hydroxylation of Monosubstituted Benzene Compounds	DOUBTED

Lasofoxifene: Relative Threshold

Relative Threshold

Take the occurrence ratio for the highest ranked biotransformation in the prediction and discard the biotransformation's which fall below n % of this value.

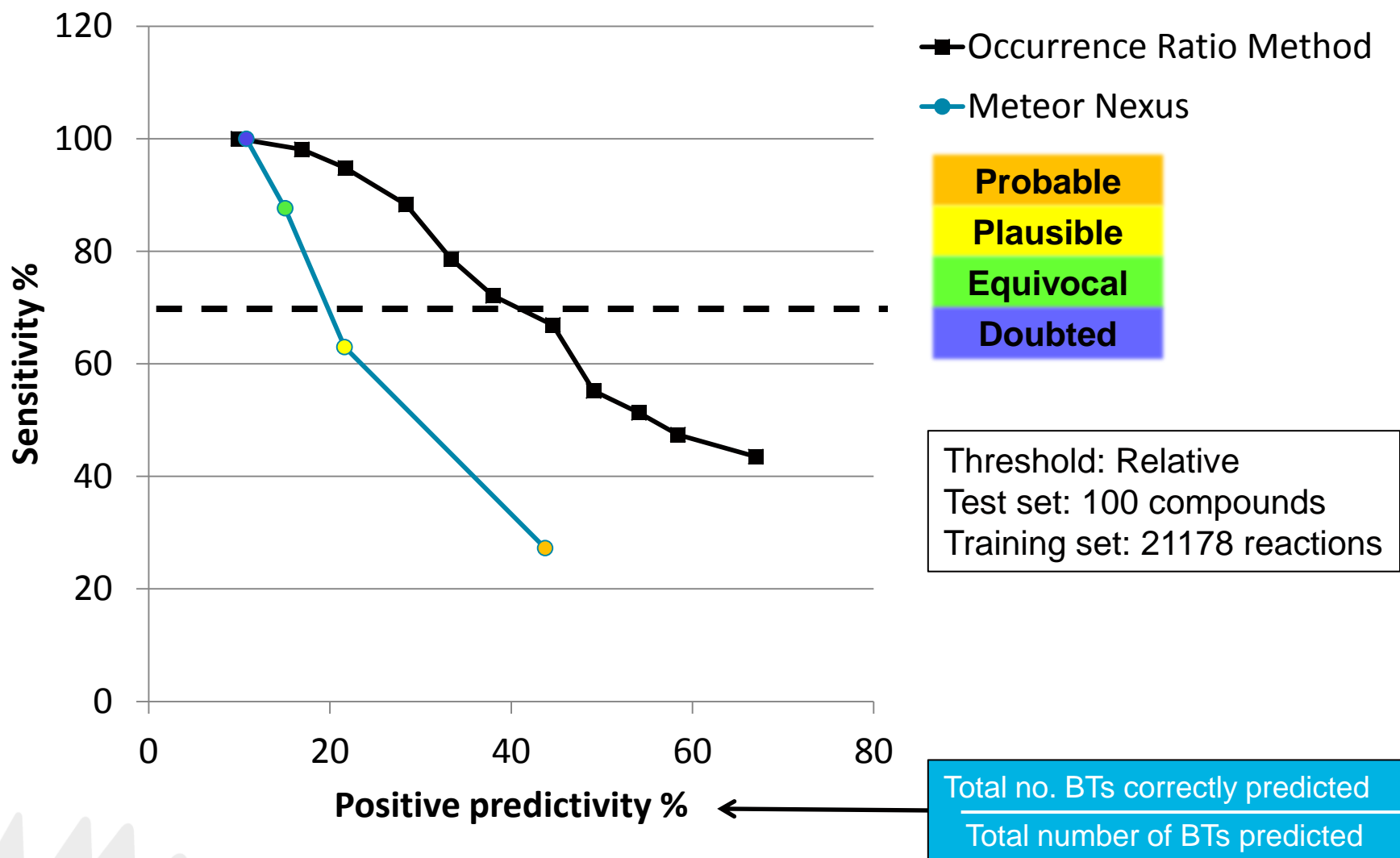
Alert ID	Biotransformation	Score (%)
027	Glucuronidation of Aromatic Alcohols	43.0
243	Oxidative N-Dealkylation	32.7
253	Oxidative O-Dealkylation	24.2
066	Hydroxylation of Alicyclic Methylene Adjacent to an Arom...	24.0
020	O-Sulphation of Aromatic Alcohols	22.3
078	Para Hydroxylation of Monosubstituted Benzene Compou...	16.1
081	Aromatic Ring Hydroxylation via Arene Oxides	12.6
100	Amine Oxides from Tertiary Alicyclic Amines	8.1
069	Hydroxylation of Unfunctionalised Alicyclic Methylene	6.6
245	Oxidative N-Dealkylation	6.1
235	5-Hydroxylation of 1,2,4-Trisubstituted Benzenes	5.6
433	Hydroxylation of Aromatic Methine	5.3
067	Lactams from Aza-Alicyclic Compounds	4.9
234	6-Hydroxylation of 1,2,4-Trisubstituted Benzenes	3.8
034	Glucuronidation of Tertiary Aliphatic Amines	3.2
077	Meta Hydroxylation of Monosubstituted Benzene Compou...	2.1
233	3-Hydroxylation of 1,2,4-Trisubstituted Benzenes	2.0
076	Ortho Hydroxylation of Monosubstituted Benzene Compou...	1.7
445	Glutathione Conjugation of Tertiary Aza-Alicyclic Compou...	0.9
439	Glucosidation of Alcohols	0.7
540	PROTOTYPE - Benzylic Glutathione Conjugation with orth...	0.5
082	Dihydrodiols via Arene Oxides	0.4
557	2-Hydroxylation of 1,4-Disubstituted Benzenes	0.2
084	Mercapturic Acids via Arene Oxides	0.2
083	Premercapturic Acids via Arene Oxides	0.0

RT = 30%

RT = 15%

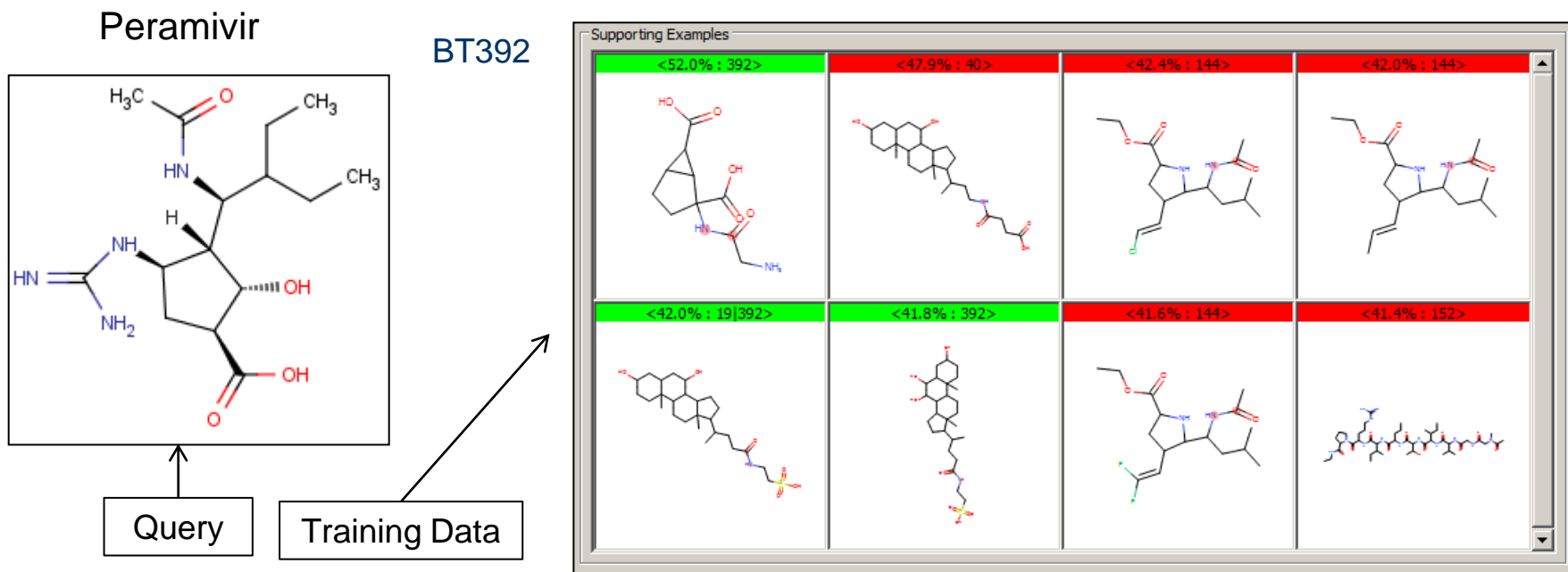
RT = 10%

Occurrence Ratio Method Vs Meteor Nexus



Local Occurrence Ratio

The supporting examples which were used to generate the occurrence ratio for each biotransformation are sorted in order of similarity with the query structure using structural fingerprints & the Tanimoto coefficient.



- Local occurrence ratio for a biotransformation is determined:

$$\text{Local occurrence ratio} = \frac{\sum \sqrt{\text{exp observed similarity}}}{\sum \sqrt{\text{similarity}}}$$

Local Occurrence Ratio

Results

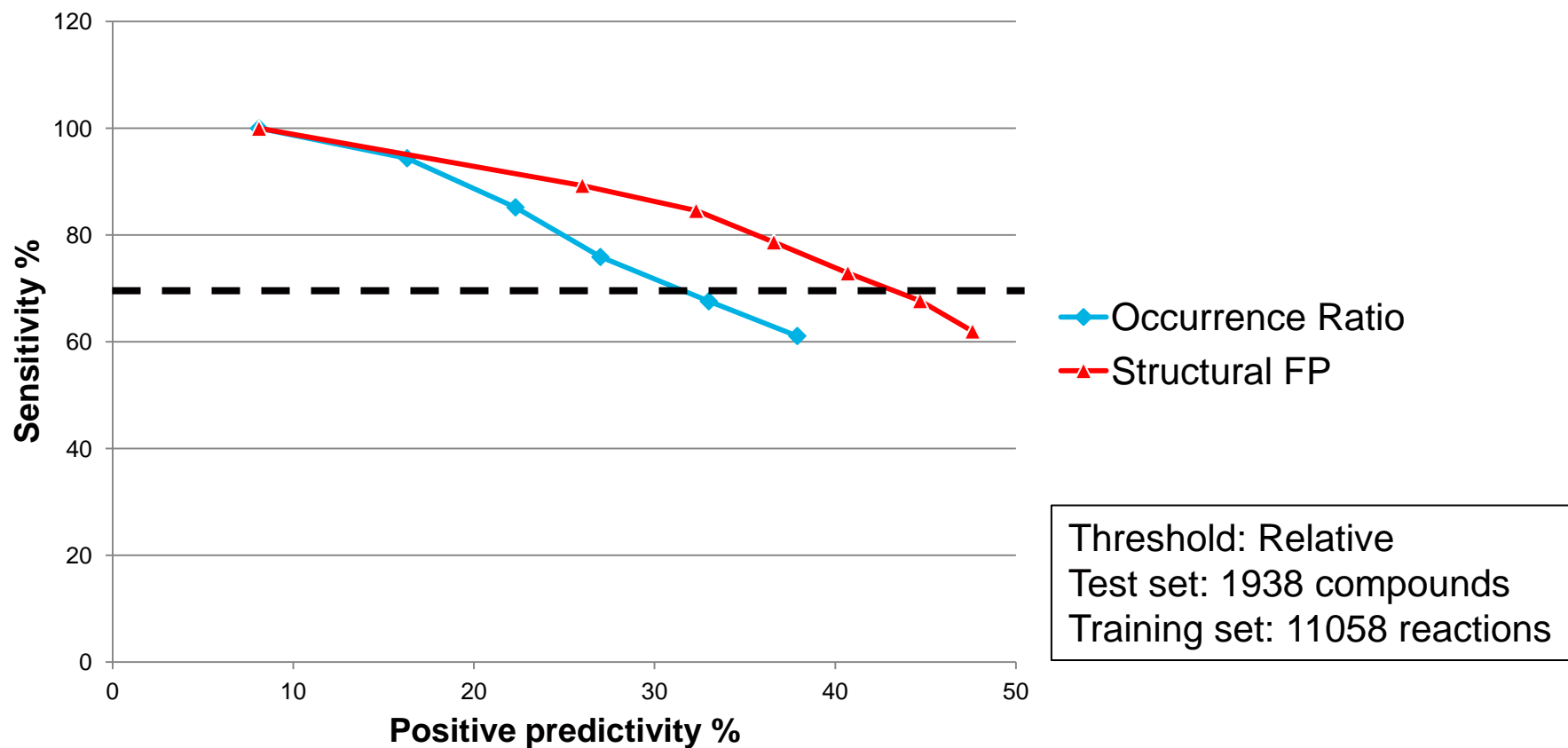
Alert ID	Biotransformation	Score (%)
392	Hydrolysis of Acyclic Carboxylic Amides	38.1
023	Glucuronidation of Primary and Secondary Aliphatic and Benzylic Alcohols	12.5
040	Conjugation of Carboxylic Acids with Taurine	12.3
019	O-Sulphation of Aliphatic Alcohols	12.0

Supporting Examples

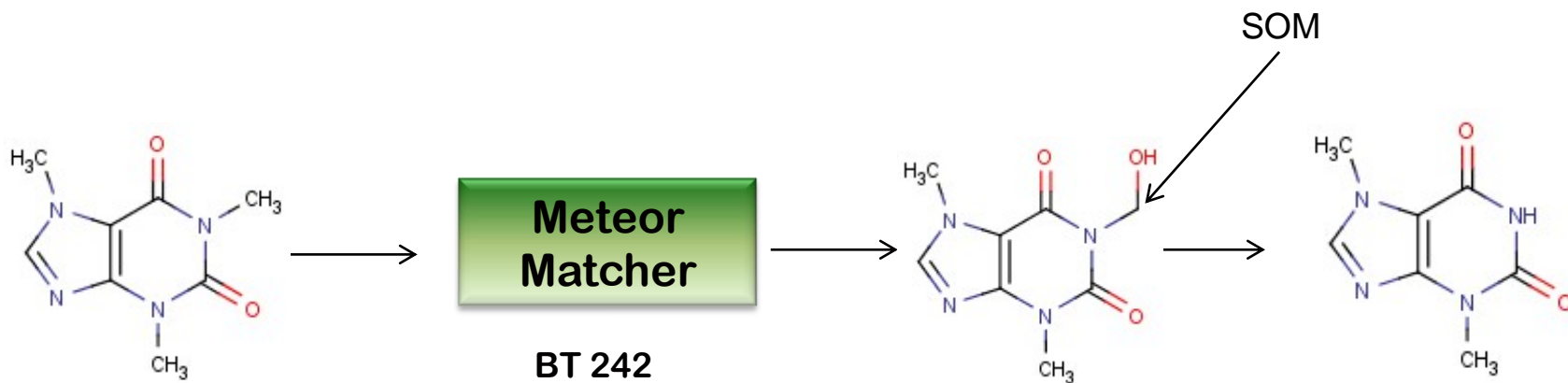
<52.0% : 392>	<47.9% : 40>	<42.4% : 144>	<42.0% : 144>
<42.0% : 191392>	<41.8% : 392>	<41.6% : 144>	<41.4% : 152>

The scores for each biotransformation are ranked to give the final prediction.

Local Occurrence Ratio Validation



SOM Based Predictions



The Meteor matching algorithm has been modified to record the atoms in the query structure where bonds are either formed or broken in the process of forming the **FIRST** intermediate.

SOM Based Predictions

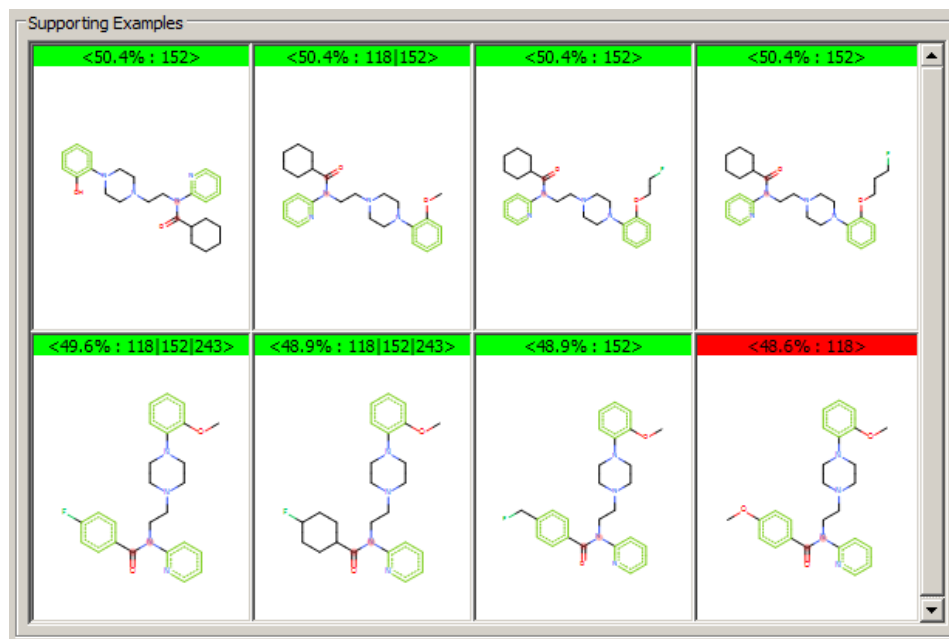
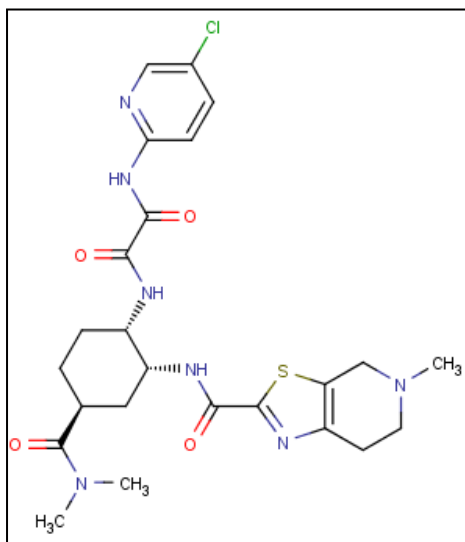
The SOM is retrieved for the **query** compound for the given biotransformation (i.e. BT 152)
The atom fingerprint is generated for the SOM

For each supporting example

The SOM is retrieved for the matched biotransformation (i.e. BT 152)

The atom fingerprint is then retrieved for the SOM

The supporting examples SOMs are sorted in order of similarity with the query structures SOM using the SOM fingerprints & Tanimoto coefficient.



(If the SOM consists of > 1 atom, all the atom fingerprints are retrieved and the logical OR of the fingerprints is generated.)

SOM Based Predictions

Chemical structure of the input molecule:

CN(C)C(=O)N1CC[C@H](NC(=O)N2C=NC=C(Cl)N2)C1C(=O)N3C=NC=C(N4CCN(C)CC4)S3

Results

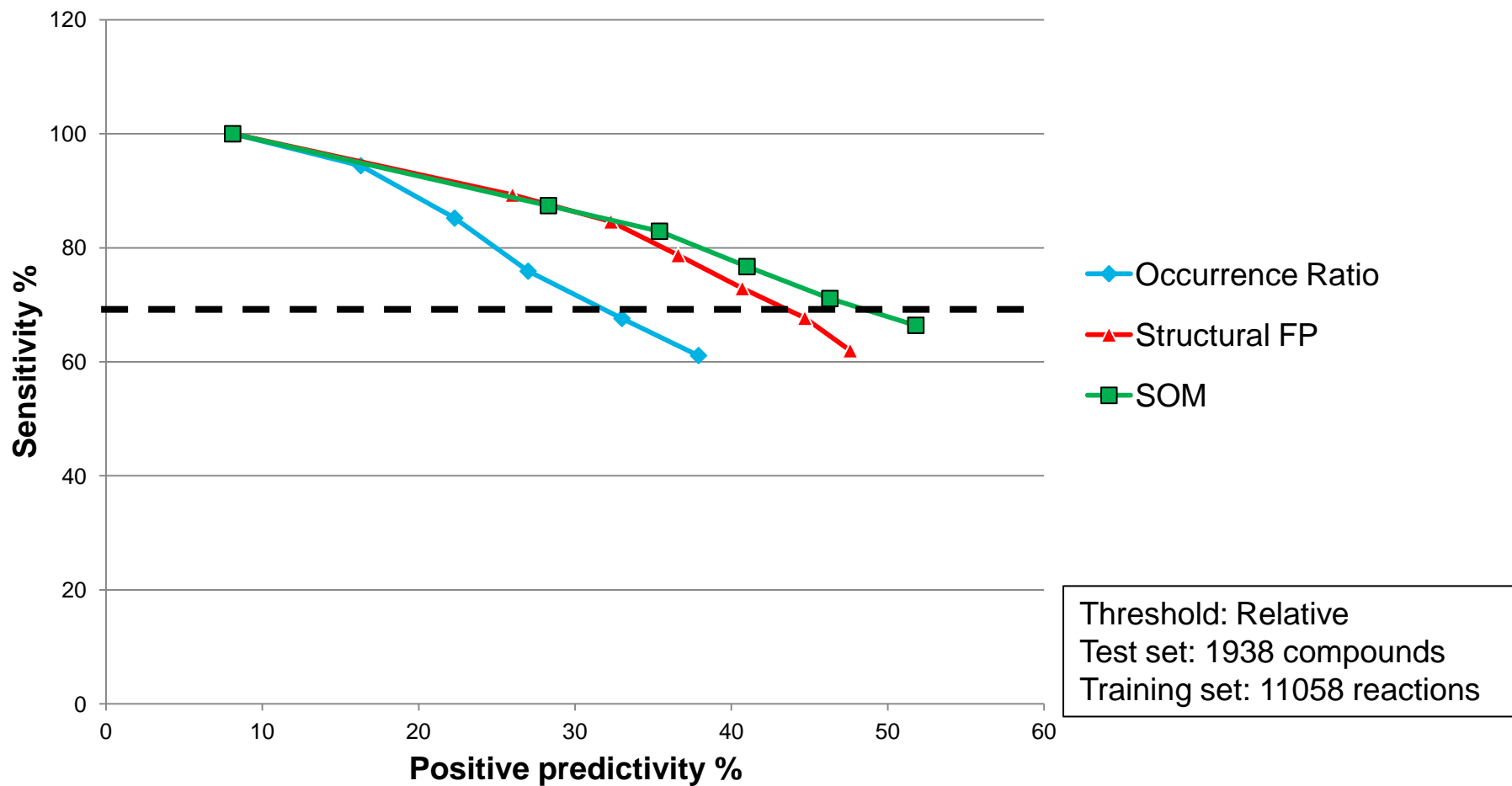
Alert ID	Biotransformation	Score (%)
152	Hydrolysis of Acyclic Carboxylic Amides	87.6
390	Pyridinium Salts from N-Substituted Piperidines	37.1
242	Oxidative N-Demethylation	36.5
152	Hydrolysis of Acyclic Carboxylic Amides	25.0
152	Hydrolysis of Acyclic Carboxylic Amides	24.7
242	Oxidative N-Demethylation	24.4
285	Amine Oxides from Tertiary Alicyclic Amines	24.2

Supporting Examples

<58.6% : 152 563>	<56.7% : 152 451>	<55.6% : 144>	<51.7% : 242>
<51.7% : 106>	<48.4% : 20 23 27...>	<48.4% : 23 69 99...>	<48.4% : 144>

Bathala BS, Masumoto H, Oquma T, He L, Lowrie C, Mendell J (2012) Pharmacokinetics, Biotransformation, and Mass Balance of Edoxaban, a Selective, Direct Factor Xa Inhibitor, in Humans, *J. Drug Metab Dispos.* **40**(12), 2250-5

Site of Metabolism Validation



Extending Predictions To Multiple Generations

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

078 : 49.4%
233 : 18.5%
027 : 18.5%
235 : 18.3%
233 : 37.5%
027 : 37.4%
235 : 37.1%
067 : 25.0%
234 : 25.0%
020 : 24.9%
081 : 24.2%
081 : 24.2%

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%
Second generation: 37.5%
Overall = 49.4% x 37.5% = 18.5%
30% Relative threshold 49.4% x 0.3 = 14.8%

<63.4% : 168>
<63.4% : 27|81|235>

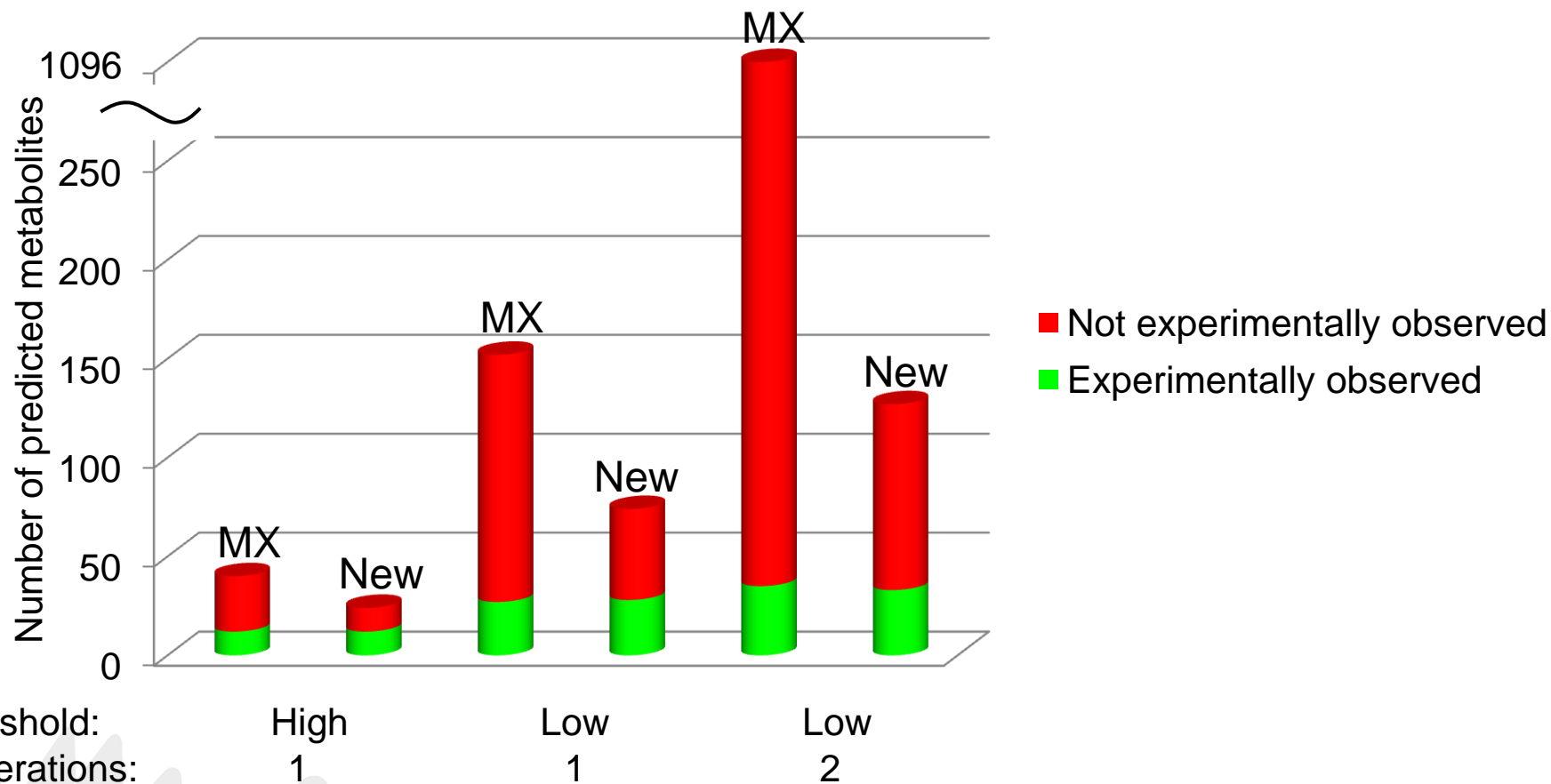
Performance

Comparison of Meteor Nexus 2.1.0 with our new approach.

Trained on 8380 unique metabolic reactions from Lhasa's metabolism database.


Test set contains test set of 10 unseen compounds.

Predictions made using Lhasa atom fingerprints and relative threshold.






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
- 



Acknowledgements

- Carol Marchant
 - Jonathan Vessey
 - Thierry Hanser
 - Stephane Werner
 - Chris Barber
- 



Work in progress disclaimer

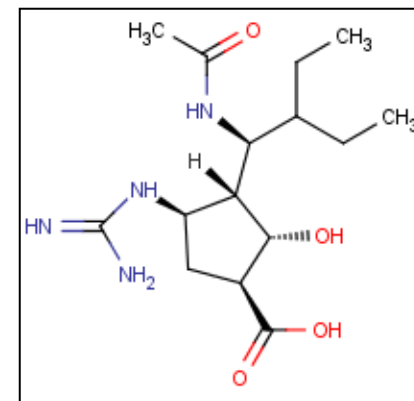
This document is intended to outline our general product direction and is for information purposes only, and may not be incorporated into any contract. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon. The development, release, and timing of any features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



Out Takes



SOM Based Predictions

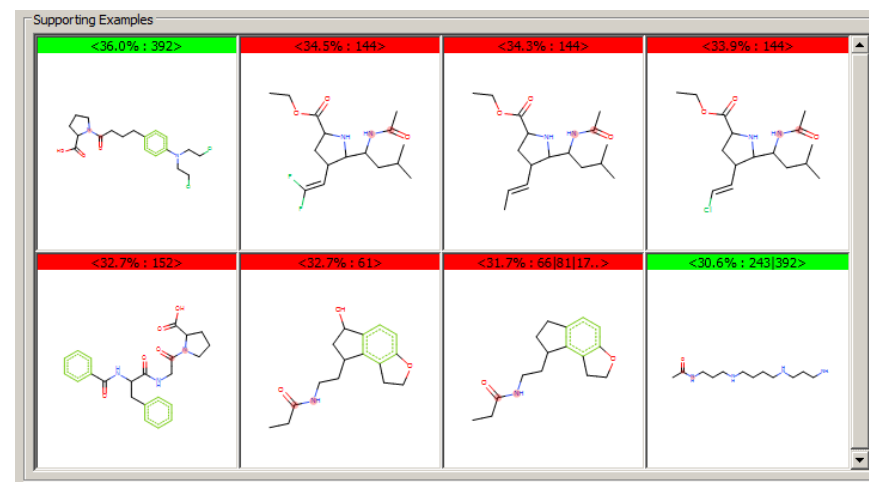
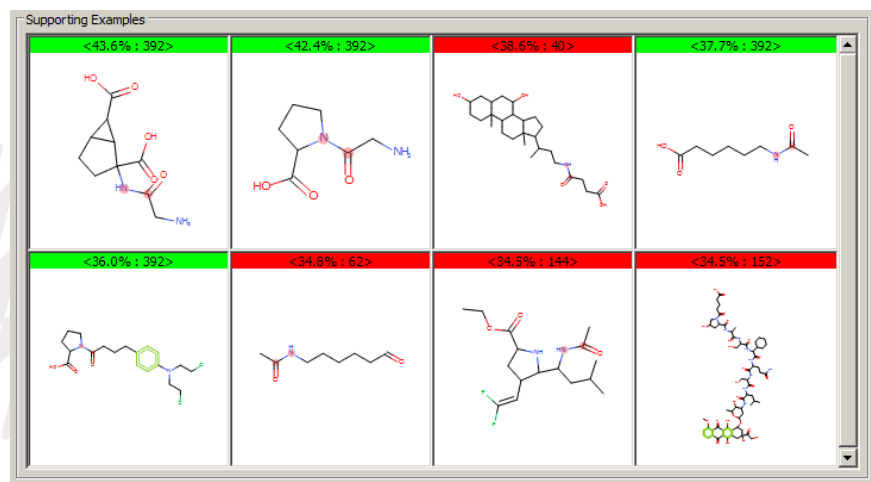


SOM Based Fingerprints

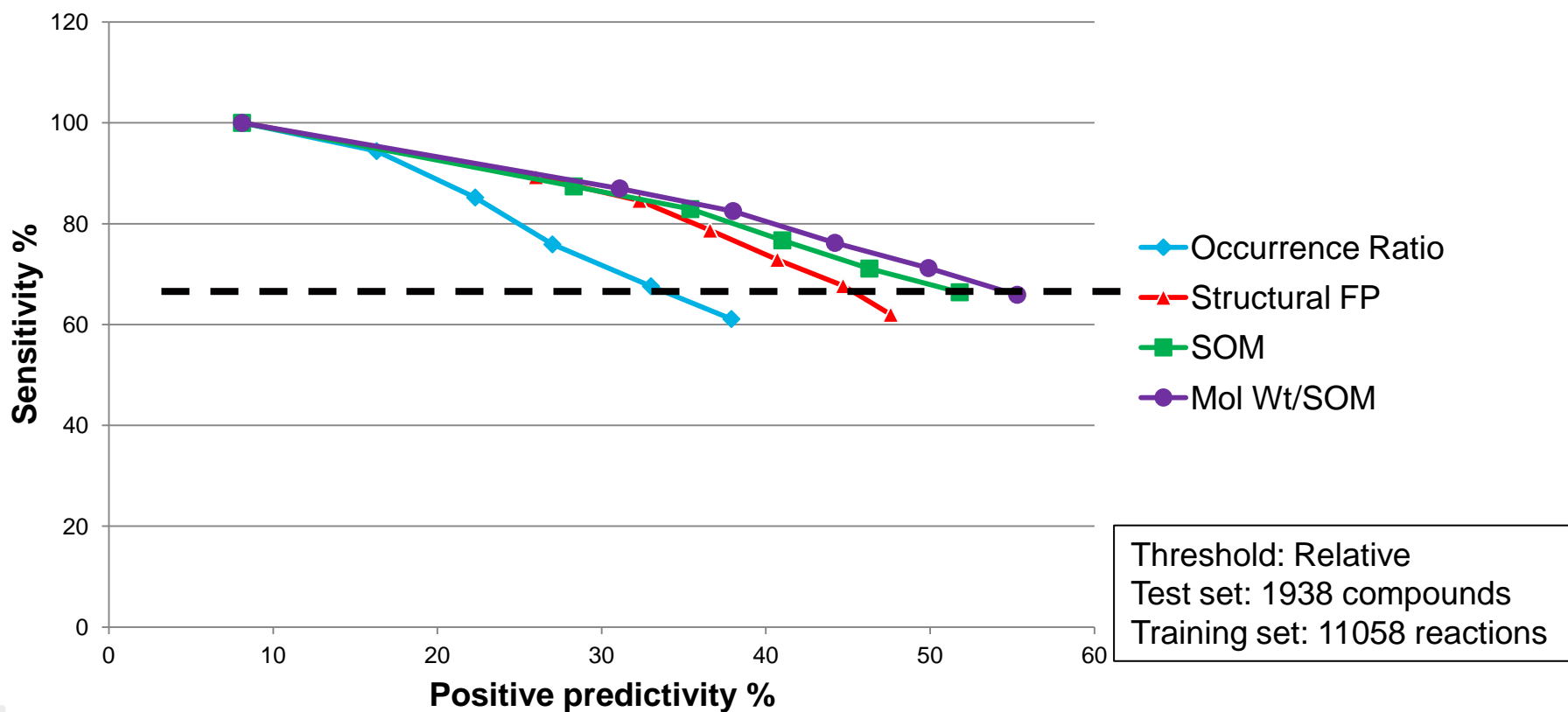
Alert ID	Biotransformation	Score (%)
392	Hydrolysis of Acyclic Carboxylic Amides	51.4
029	Glucuronidation of Carboxylic Acids	12.9
178	Hydroxylation alpha to a Carbonyl Group	12.4

SOM Based Fingerprints with Mol Wt. Filter

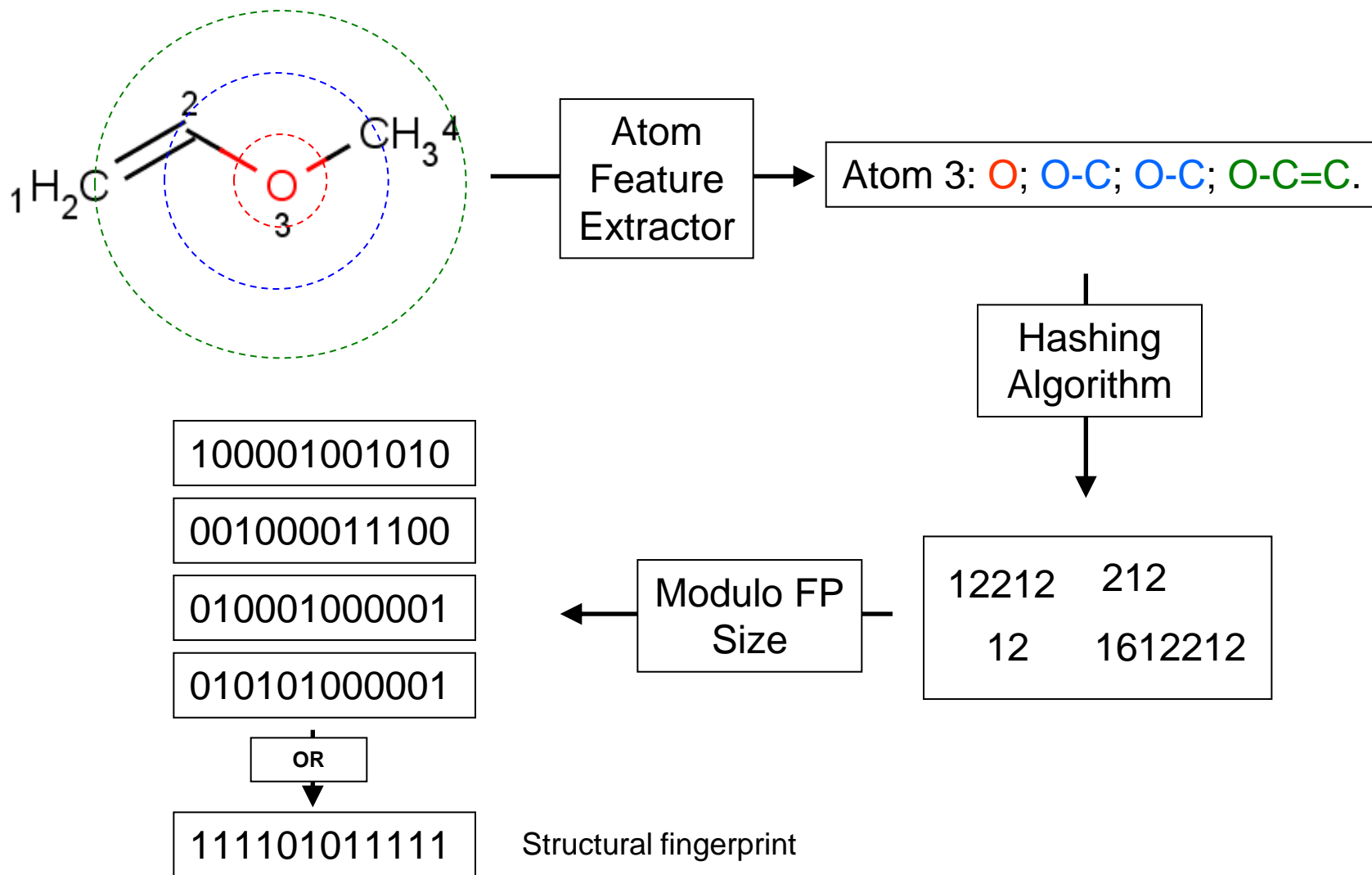
Alert ID	Biotransformation	Score (%)
392	Hydrolysis of Acyclic Carboxylic Amides	25.0
029	Glucuronidation of Carboxylic Acids	12.9
178	Hydroxylation alpha to a Carbonyl Group	12.5



Site of Metabolism Validation



Ceres Fingerprints

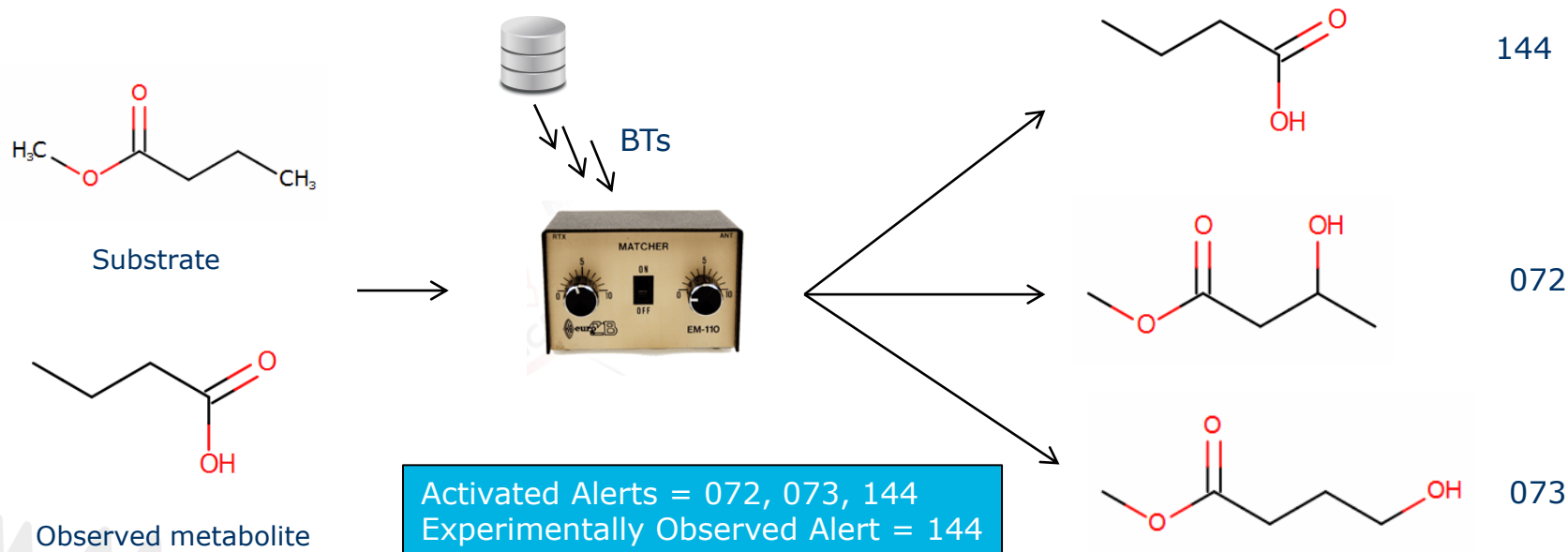


Generating Occurrence Ratios

Retrieve all substrates (structures) from the metabolite database and record which metabolites were experimentally observed.

Process each substrate through the Meteor matcher and record the BTs that were activated by the substrate.

Compare each of the metabolites generated in Meteor with the experimentally observed metabolite to determine the experimentally observed alert.



Performance

Comparison of Meteor Nexus 2.1.0 with our new approach.

Trained on 2819 unique metabolic reactions from Lhasa's metabolism database.

Test set contains test set of 10 unseen compounds.

Predictions made using Lhasa atom fingerprints and relative threshold.

System	Metabolites	Exp. Observed	Threshold	Generation
MX	40	12	Plausible	1
Next Gen.	24	12	Rel. 40%	1
MX	152	27	Equivocal	1
Next Gen.	74	28	Rel. 15%	1
MX	1096	35	Equivocal	2
Next Gen.	127	32	Rel. 15%	2