

Molecular Networks
Inspiring Chemical Discovery



Chemical Reactivity in Metabolism Reactions for the Risk Assessment Workflow

Johann Gasteiger
Molecular Networks GmbH
Henkestraße 91
91052 Erlangen, Germany
www.molecular-networks.com

Outline

- **Modeling physicochemical effects**
- **Representation of chemical structures**
- **Modeling of toxicity**
- **Metabolism of xenobiotics**
- **Reactivity fingerprints in ToxCast**



Modeling Chemical Reactions

- **Theoretical chemist:**

- **Quantum-mechanical** calculations: time-consuming

- **Organic chemist:**

- **Concepts** for rationalizing reaction mechanisms
- **Partial charges, inductive, resonance, polarizability, steric effect**

➔ **Quantify physicochemical effects**



Calculation of Physicochemical Effects

- Charge calculation: q_{σ} and q_{π}
- Inductive effect: χ_r
- Resonance effect: M^+ , M^-
- Polarizability effect: α_d
- Steric accessibility: A_{access}
- Heats of formation/heats of reaction

PETRA package

(Parameter Estimation for the Treatment of Reactivity Applications)



Prediction of Chemical Reactivities

- Gas phase reactions (proton affinities, acidities)
- pK_a values
- General nucleophilicity scale
- Hydrolysis of amides
- Stability of compounds in DMSO/H₂O



ADRIANA. Code – Calculation of Structure Descriptors

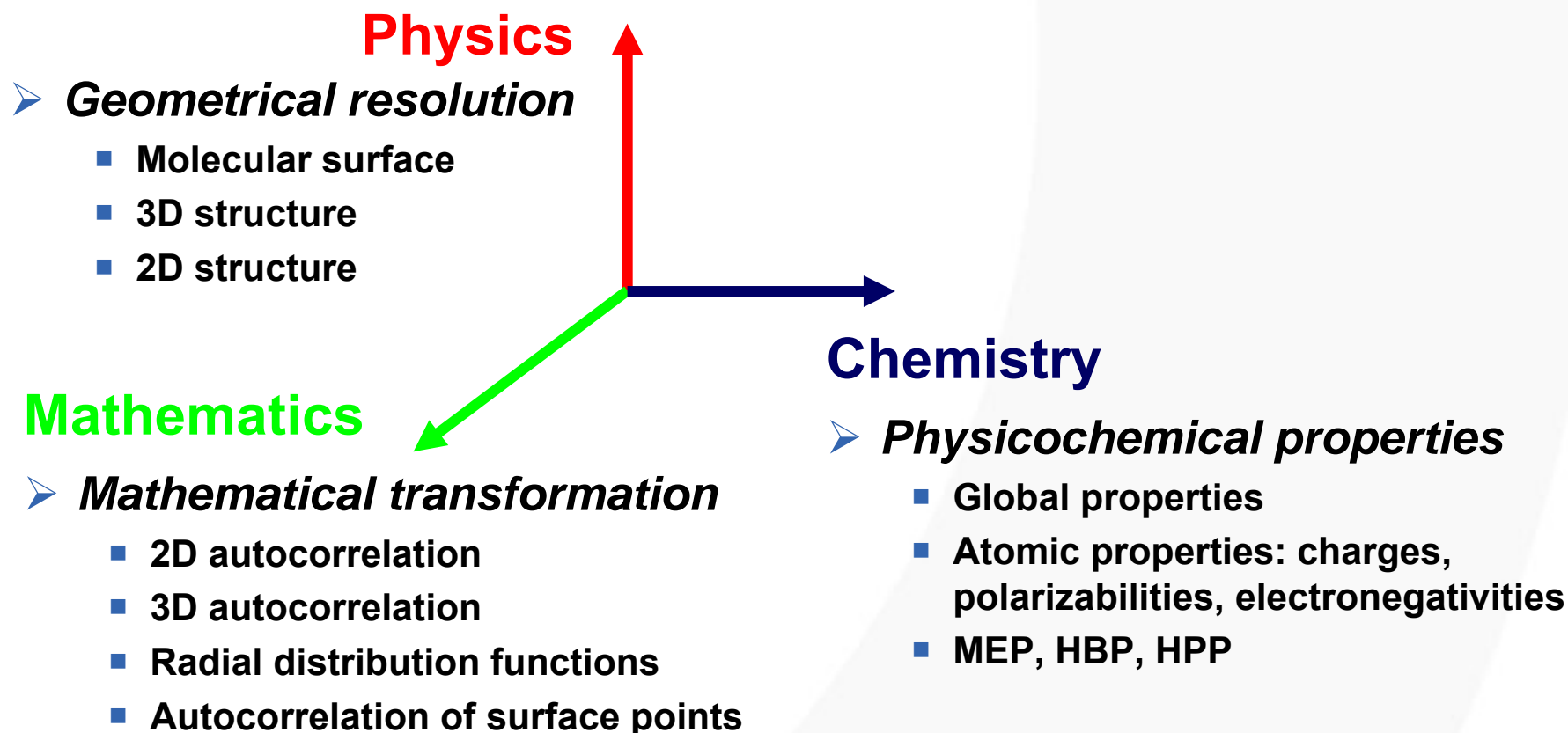


- **Molecular (global) properties**
 - *Number of H bond acceptors and donors, molecular weight, TPSA, dipole moment, molecular polarizability, logP, logS, ...*
- **Molecular shape- and size-related descriptors**
 - *Principal moments of inertia, radius of gyration, span, diameter, ...*
- **Atom properties for 2D and 3D autocorrelation and radial distribution functions**
 - *Charges (sigma, pi and total), electronegativities (sigma, pi and lone pair), effective polarizability*
- **Surface properties for autocorrelation**
 - *Molecular electrostatic potential (MEP), hydrogen bonding potential (HBP) and hydrophobicity potential (HPP)*



ADRIANA.Code – Covered Descriptor Space

- Structure coding spanned by 3 axes in descriptor space



Modeling of Toxicity

■ Data analysis methods

- S.Spycher, M.Nendza, J.Gasteiger, *QSAR Comb. Sci.*, **2004**, 23, 779-791

■ Representation of chemical structures

- S.Spycher, E.Pellegrini, J.Gasteiger, *J.Chem.Inf.Model.*, **2005**, 45, 200-208

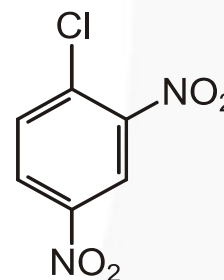
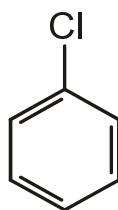
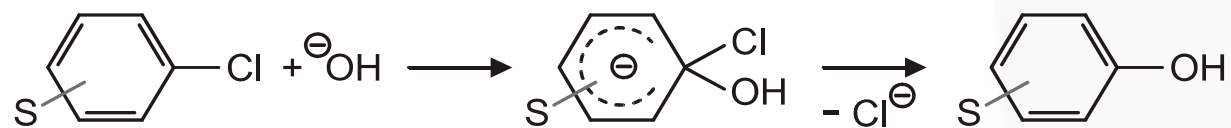
■ Considering toxicological mechanism

- S.Spycher, B.Escher, J.Gasteiger, *Chem.Res.Toxicol.*, **2005**, 18, 1857-1867



Structural Alerts and Chemical Reactivity

■ Aromatic halides



Reaction conditions:

400° C, 300 bar

room temperature

Resonance stabilization:
of intermediate, M_A^-

$\equiv 0.0$

43.5 kJ/mol

➔ Huge difference in reaction conditions

➔ Can be explained by resonance stabilization of intermediate

➔ Modify structural alerts by electronic effects





Metabolism of Xenobiotics

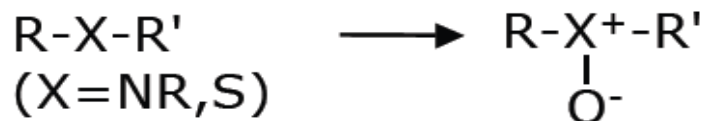
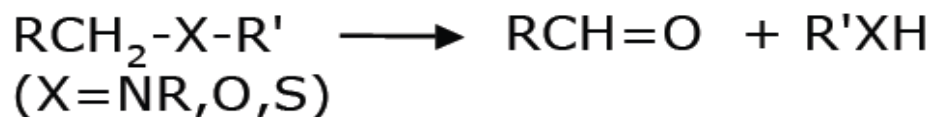
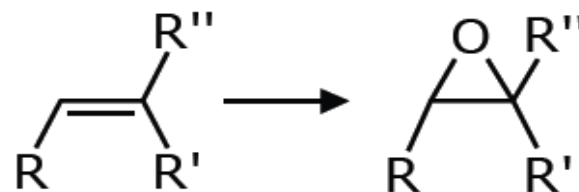
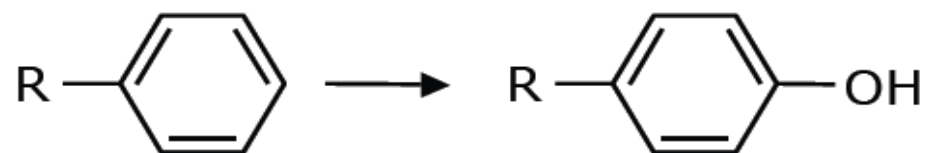
Drugs, agrochemicals, food additives



Oxidations by Cytochrome P450



- **Aromatic hydroxylation**
- **Aliphatic hydroxylation**
- **Epoxidation**
- **N, O, S-dealkylation, oxidative deamination**
- **N,S-oxidation**



Different Selectivities



- **Selectivity between different cytochrome P450 isoenzymes**
 - *in particular 3A4, 2C9, 2C19, 2D6, 1A2*
- **Selectivity between different reaction types**
 - *chemoselectivity*
- **Selectivity between different reaction sites**
 - *regioselectivity*



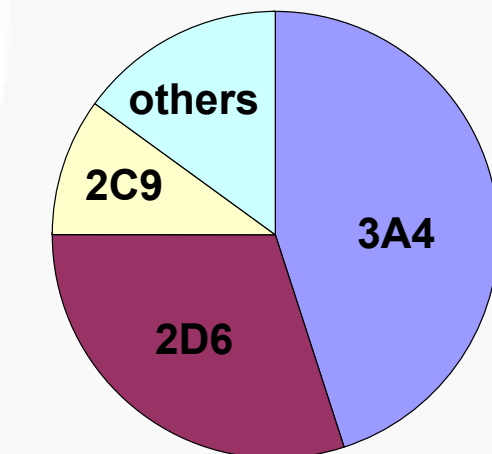
Data Set of 3A4, 2D6, and 2C9 Substrates

■ Training and test data set: 146 compounds

- *Manga et al, SAR and QSAR in Environm. Res. 2005, 16 (1-2), 43-61*
- **80 3A4 substrates (55%)**
- **45 2D6 substrates (31%)**
- **21 2C9 substrates (14%)**

■ Validation data set: 233 compounds

- *Metabolite database*
- **144 3A4 substrates (62%)**
- **69 2D6 substrate (30%)**
- **20 2C9 substrates (8%)**



Relative proportion of drugs metabolized by major P450s



Support Vector Machine (SVM) Model

- Descriptors (242 components)
- Automatic variable selection: 12 components
 - $2D-AC_{identity}(5)$, $2D-AC_{q\pi}(3)$, $2D-AC_{q\pi}(6)$, $2D-AC_{\chi\pi}(5)$, $2D-AC_{q\sigma}(1)$, $2D-AC_{q\sigma}(2)$, $2D-AC_{\chi\sigma}(6)$, $3D-AC_{identity}([5.8-5.9[\text{\AA})$, n_{acid_groups} , $n_{aliphatic_amino}$, n_{basic_n} , r_3

Predictability

- Training: 90.4%
- 5-fold CV: 87.8%



Validation of the Support Vector Machine Model

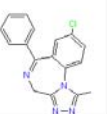
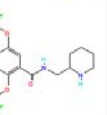
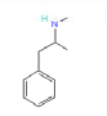
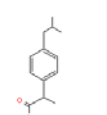
- **Validation set: 233 substrates from the Metabolite database**
- **Predictability: 82.8%**
- **remember: some drugs are metabolized by several isoforms**

L. Terfloth, B. Bienfait, J. Gasteiger, *J. Chem. Inf. Model.* **2007**, *47*, 1688-1710



isoCYP Webservice

The screenshot displays the 'Your results' page of the Molecular Networks Web Services. It features a table with four rows of results:

Rank	Compound	Predominant P450 Isoform	Name(read)
1		CYP3A4	Alprazolam
2		CYP2D6	Flecainide
3		CYP2D6	Methamphetamine
4		CYP2C9	Ibuprofen

Below the table, there is a 'Display Properties' section with a 'Read-only properties' area containing a 'Name' dropdown menu and buttons for 'refresh' and 'defaults'. At the bottom of the page, there is a 'Back' button and a note: 'Compute and display (Predominant P450 Isoform) of compounds in the results table'.

**Prediction of
major
metabolizing
CYP450 isoform
(2D6, 3A4, 2C9)**

■ http://www.molecular-networks.com/online_services

L. Terfloth, B. Bienfait, J. Gasteiger, *J. Chem. Inf. Model.* **2007**, *47*, 1688-1710

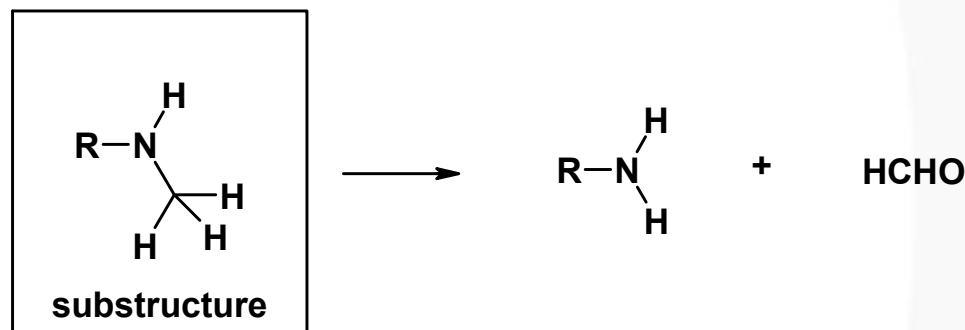


MetaboGen Reaction Rules

- **117 reaction rules**
- **Reaction types covered:**
 - *Aromatic hydroxylation*
 - *Aliphatic hydroxylation*
 - *N- and O-dealkylation*
 - *Hydrolysis (ester, amides)*
 - *Conjugation reactions (glucuronidation, sulphation, glycation, acetylation)*
 - *Oxidation reactions (alcohols, aldehydes, etc.)*
- **Empirical score for likeliness of a reaction based on literature data**



Example for the Definition of a Reaction Rule



N-Demethylation of aromatic N-methylamines

■ Substructure definition

- *R in the substructure is an aromatic carbon atom.*
- *The nitrogen atom in the substructure has exactly three neighbor atoms.*
- *The nitrogen atom in the substructure is not charged.*

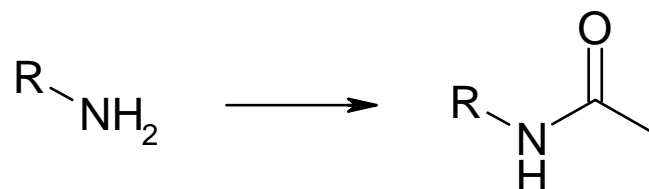
■ (Formal) Definition of the reaction

- *The carbon nitrogen bond is cleaved resulting in a primary amine and formaldehyde.*



Derivation of a Rule Base for Metabolite Prediction

- Define reaction rules, e.g. for an acetylation



- Calculate reaction probabilities based on a reaction database

➤ <i>Conceivable metabolites</i>	1101
➤ <i>Observed metabolite</i>	122
➤ <i>Total number of metabolites generated</i>	1223
➤ <i>Probability</i>	$122/1223 = 0.10$



Reaction Generation with MetaboGen



MOSES MetaboGen (v 1.0 2006-11-17)

Structure Input File
dextromethorphan.sdf

Structure Output File
dextromethorphan_reactions.sdf

Reaction Output File
dextromethorphan_metabolic_reactions.rdf

Reaction Rules

<input checked="" type="checkbox"/> Aromatic Hydroxylation	<input checked="" type="checkbox"/> Benzylic Hydroxylation	<input type="checkbox"/> Alcohol Oxidation
<input type="checkbox"/> N-Demethylation	<input type="checkbox"/> N-Deethylation	<input type="checkbox"/> Aldehyde Oxidation
<input type="checkbox"/> O-Demethylation	<input type="checkbox"/> O-Deethylation	
<input type="checkbox"/> Deamination		
<input type="checkbox"/> N-Oxidation	<input type="checkbox"/> S-Oxidation	<input type="checkbox"/> Desulfurization
<input type="checkbox"/> Aromatic Amine Oxidation	<input type="checkbox"/> Aromatic Hydroxylamine Oxidation	
<input type="checkbox"/> Epoxidation	<input type="checkbox"/> Epoxide Hydrolysis	
<input type="checkbox"/> Hydrolysis	<input type="checkbox"/> Amide Hydrolysis	<input type="checkbox"/> Ester Hydrolysis
<input type="checkbox"/> Acetal Hydrolysis	<input type="checkbox"/> Half-Acetal Hydrolysis	<input type="checkbox"/> Geminal Diol Hydrolysis
<input type="checkbox"/> Aldehyde Reduction	<input type="checkbox"/> AzoCompoundReduction	

Options

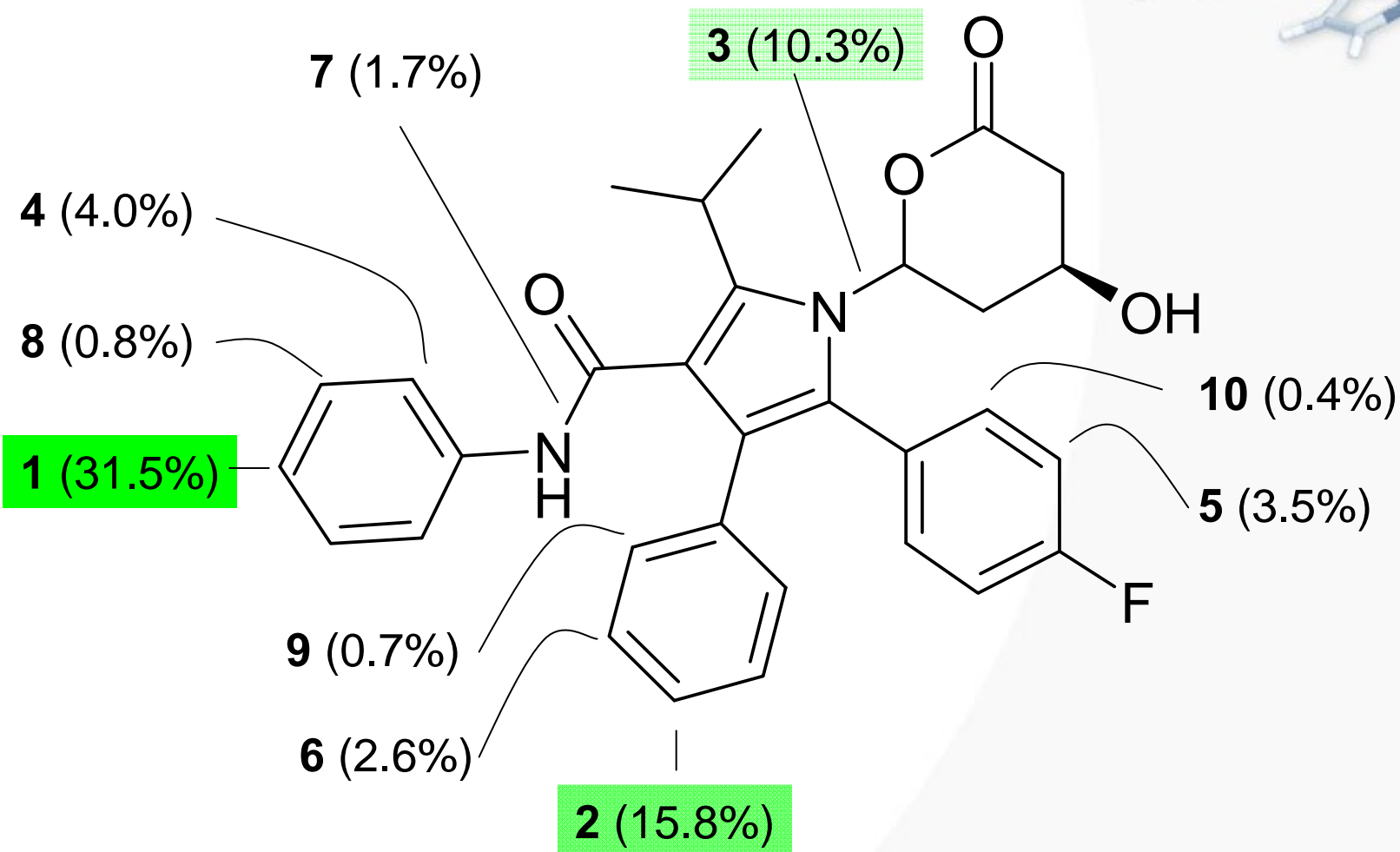
suppress output of reactants write brief reactions

Reaction Level

- **MetaboGen**
- **Covers all relevant phase I reactions**
- **Supports atom-atom mapping**
- **Provides information on reaction centers**
- **Is used to generate reactions for XENIA**



Ranks and Probabilities of Metabolic Reactions Predicted for Atorvastatin



New Descriptors for Metabolic Reactivity

- Describing chemical structures with a priori chemical knowledge on reaction centers and metabolic reactivity
- Metabolic reactivity classes
 - *To describe metabolic fate of chemicals*
 - *Reaction types*
 - aromatic hydroxylation, aliphatic hydroxylation, N- and O-dealkylation, hydrolysis (ester, amide), and conjugation reactions (acetylation, sulfation, etc.)



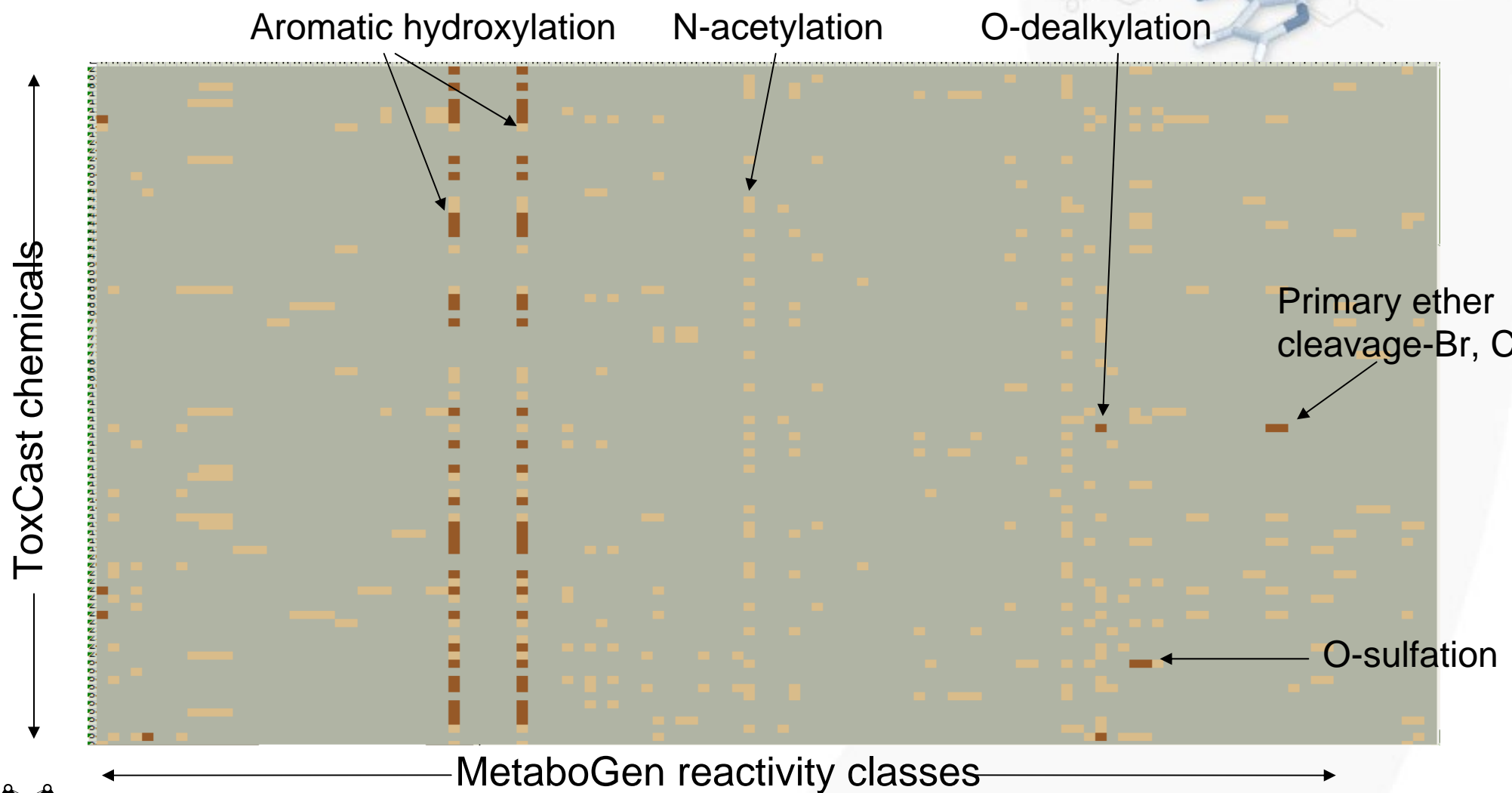
Extract of the Metabolic Reactivity Matrix of the ToxCAST Data Set



DSSTox_CID	Structure	MetaboGen total counts	Aromatic hydroxylation	Aliphatic hydroxylation (prim. C next sec. C)	Aliphatic hydroxylation (sec. C next to CH3)	N-Demethylation R-NMe2	O-Sulphation	N-Acetylation R-NH2	...
370		2	0	0	0	1	1	0	...
8038		6	0	1	1	0	0	0	...



Fingerprint View of Metabolic Reactivity Classes



Acknowledgements

- **Molecular Networks**

- *Bruno Bienfait*
- *Thomas Kleinöder*
- *Aleksey Tarkhov*
- *Lothar Terfloth*

- **Chihae Yang**

