Risk Assessment of Chemicals and Prediction of Metabolism

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

- MOSES system
- Structure representation
- Toxicity prediction
- MOSES.ChemistryToolbox
- Metabolism of xenobiotics
- MOSES.Metabolism
- MOSES.RiskAssessment
MOlecular Structure Encoding System

- C++ based Chemoinformatics toolkit
  - high performance
  - available for many platforms (Windows, Linux, Unix)

- Python interface available
  - provides easy access to the full functionality of MOSES
  - ideally suited for the development of client / server solutions

- under active development since 2001
  - Computer-Chemie-Centrum, Universität Erlangen-Nürnberg
  - Molecular Networks GmbH

- 300,000 lines of code
  - well documented and tested
User Interfaces

Stand alone

Qt (C++/Python)

Local files

Databases

Web based

Client (Browser)

Server (Python)

Internet

MOSES (C++/Python)
Modeling Chemical Structures and Reactions

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

- **Organic chemist:**
  - Concepts for rationalizing chemical reactivity and reaction mechanisms
  - Partial charges, inductive, resonance, polarizability, steric effect

→ Quantify these physicochemical effects
Calculation of Physicochemical Effects

- Charge calculation: $q_\sigma$ and $q_\pi$
- Inductive effect: $\chi_r$
- Resonance effect: $M^+, M^-$
- Polarizability effect: $\alpha_d$
- Steric accessibility: $A_{access}$
- Heats of formation/heats of reaction

PETRA package

(Parameter Estimation for the Treatment of Reactivity Applications)
Representation of chemical structures

Hierarchy of interpretable structure descriptors
Structure Representation

- Constitution
- 3D model
- Molecular surface

ADRIANA. *Code* – Covered
Descriptor Space

- Structure coding spanned by 3 axes in descriptor space

**Physics**

- *Geometrical resolution*
  - Molecular surface
  - 3D structure
  - 2D structure
  - 1D (global)

**Mathematics**

- *Mathematical transformation*
  - 2D autocorrelation
  - 3D autocorrelation
  - Radial distribution functions
  - Autocorrelation of surface points

**Chemistry**

- *Physicochemical properties*
  - Global properties
  - Atomic properties: charges, polarizabilities, electronegativities
  - MEP, HBP, HPP
ADRIANA. Code – Areas of Application

- Drug design
  - Clustering of compounds according to their biological activity
  - Locating biologically active compounds in sets of diverse chemical compounds
  - Quantitative prediction of biological activities
  - Analysis of results of high-throughput screening

- Prediction of ADME/Tox properties
  - Aqueous solubility of organic compounds
  - pKa values
  - Prediction of major metabolizing CYP450 isoform
  - Classification of toxic mode of action

- Prediction of infrared and $^1$H NMR spectra

- Dye design

For list of publications:
http://www2.chemie.uni-erlangen.de/publications/
Modeling toxicity of compounds

Combination of descriptors
Modeling of Toxicity

- Different data analysis methods

- Representation of chemical structures

- Considering toxicological mechanism
Why Prediction of Toxic Mode of Action (MOA)?

Most QSARs in toxicology focus on a certain class of compounds.

However:

- polar narcotic
- uncoupler of oxidative phosphorylation

Require different QSAR-equations

First classify structures according to their MOA
Dataset: MOA of Phenols

1. Polar narcotics (156 cpds)
2. Uncouplers of oxidative phosphorylation (19 cpds)
3. Precursors to soft electrophiles (24 cpds)
4. Soft electrophiles (22 cpds)

221 cpds

Dataset:

Study:
Predictive Power of Model (Counterpropagation-Neural Network)

estimate of predictive power with 5-fold cross-validation:

$$\text{RDF}(\chi_{LP}, \chi_\sigma), \text{ HBP surface AC}$$

95.9%

(2×8 + 12) descriptors

RDF: radial distribution function
HBP: hydrogen bonding potential

Classification in 5-fold Crossvalidation

polar narcotic

uncoupler of oxidative phosphorylation

Correct classification!
MOSES.ChemistryToolbox

- Program package for the prediction of physical, chemical or biological properties of compounds
- Representation of chemical structures for QSAR studies
- Combining the descriptor calculation of ADRIANA.Code with structural features
MOSES.ChemistryToolbox – Structural Features

- Functional groups, e.g.:
  - aldehyde, aromatic
  - aldehyde, alkenyl
  - aldehyde, alkyl
  - amine, tert-N, alkyl
  - amine, sec-NH, aromatic
  - amine, aromatic, N-hydroxy
  - halide, prim-alkyl
  - silane, trimethyl
  - Michael acceptor
  - urethane derivative

- Structural elements, e.g.:
  - benzofuran
  - imidazole
  - quinoxaline
  - pyrrolidine
  - purine
  - guanidine
  - steroid
  - pyrazine
  - aflatoxin
  - pyrimidine
MOSES.ChemistryToolbox - Functionalities

- Reading of chemical structure files (SDFiles, SMILES, etc.)
- Merging of multiple files into one spreadsheet
- Calculation of physicochemical properties
- Calculation of structural class fingerprints
- Browsing of structures and properties as spreadsheet with database backend
- Output of spreadsheet as structure files
- Output of spreadsheet as table file (compatible with Excel)
- Project management
# MOSES.ChemistryToolbox

## Compounds and Properties

<table>
<thead>
<tr>
<th>Compound</th>
<th>Name</th>
<th>ActivityCategory</th>
<th>ActivityClass</th>
<th>Dipole</th>
<th>LogE</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Compound" /></td>
<td>bda-0001</td>
<td>BOA</td>
<td>1</td>
<td>6.94386</td>
<td>-5.330692</td>
<td>361.852</td>
</tr>
<tr>
<td><img src="image2.png" alt="Compound" /></td>
<td>bda-0002</td>
<td>BOA</td>
<td>1</td>
<td>5.014402</td>
<td>-3.875652</td>
<td>358.397</td>
</tr>
<tr>
<td><img src="image3.png" alt="Compound" /></td>
<td>bda-0003</td>
<td>BOA</td>
<td>1</td>
<td>4.067482</td>
<td>-2.758703</td>
<td>253.226</td>
</tr>
</tbody>
</table>
Classification model for salmonella reverse mutation

Study performed by Dr. Chihae Yang, FDA CFSAN
Modeling process: OECD-compliant

- Training set (salmonella reverse mutation)
  - Transparency
  - biological and chemicals modes of activities
- Interpretable descriptors
  - Structural features
  - Calculated molecular properties
- Statistical algorithms and inference
  - fitting/parameter optimization/cross-validation
- External validation
Descriptors: Structural Features

- FDA Redbook inspired features
  - Generic compound class features
  - Classes defined in Cramer classes
  - Categories for Threshold of toxicological concern
- Known alerts
  - Ashby Tennant genotoxic carcinogen alerts
- Alerts learned from the dataset
ADRIANA.CODE Descriptors

• Whole molecule descriptors
  – xLogP, topological polar surface area, water solubility, molecular weight, (heavy) atom count, hydrogen bond acceptors and donors (N and O specific), Lipinski score violation, rotational bonds, ring complexity

• Electrostatic properties
  – charges (sigma, pi, and lone pair)
  – electronegativity (sigma, pi, and lone pair)

• Surface properties.

• Molecular shape
## Performance Comparisons

<table>
<thead>
<tr>
<th>Model Name</th>
<th>Pos/Neg</th>
<th>Descriptor</th>
<th>PLS factor</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aromatic Amines</td>
<td>321/417</td>
<td>Structure</td>
<td>3</td>
<td>78</td>
<td>62</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ADRIANA</td>
<td>3</td>
<td>82</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Both</td>
<td>4</td>
<td>78</td>
<td>89</td>
</tr>
<tr>
<td>Halides</td>
<td>379/530</td>
<td>Structure</td>
<td>2</td>
<td>77</td>
<td>82</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ADRIANA</td>
<td>8</td>
<td>64</td>
<td>81</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Both</td>
<td>4</td>
<td>77</td>
<td>88</td>
</tr>
<tr>
<td>Global</td>
<td>1807/2490</td>
<td>Structure</td>
<td>10</td>
<td>64</td>
<td>87</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ADRIANA</td>
<td>30</td>
<td>54</td>
<td>85</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Both</td>
<td>19</td>
<td>78</td>
<td>88</td>
</tr>
<tr>
<td>Phenols</td>
<td>283/370</td>
<td>Structure</td>
<td>3</td>
<td>66</td>
<td>88</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ADRIANA</td>
<td>10</td>
<td>68</td>
<td>81</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Both</td>
<td>8</td>
<td>84</td>
<td>93</td>
</tr>
</tbody>
</table>
## Performance Comparisons

<table>
<thead>
<tr>
<th>Model Name</th>
<th>Pos/Neg</th>
<th>Descriptor</th>
<th>PLS factor</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aldehydes</td>
<td>58/34</td>
<td>Structure</td>
<td>4</td>
<td>74</td>
<td>81</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ADRIANA</td>
<td>5</td>
<td>67</td>
<td>78</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Both</td>
<td>2</td>
<td>97</td>
<td>94</td>
</tr>
<tr>
<td>Halides, aliphatic</td>
<td>181/247</td>
<td>Structure</td>
<td>3</td>
<td>81</td>
<td>68</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ADRIANA</td>
<td>7</td>
<td>85</td>
<td>62</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Both</td>
<td>4</td>
<td>80</td>
<td>91</td>
</tr>
<tr>
<td>Azo-Azoxy</td>
<td>64/34</td>
<td>Structure</td>
<td>5</td>
<td>76</td>
<td>62</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ADRIANA</td>
<td>2</td>
<td>82</td>
<td>46</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Both</td>
<td>2</td>
<td>95</td>
<td>94</td>
</tr>
<tr>
<td>Hydrazine</td>
<td>62/34</td>
<td>Structure</td>
<td>2</td>
<td>80</td>
<td>81</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ADRIANA</td>
<td>9</td>
<td>86</td>
<td>77</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Both</td>
<td>2</td>
<td>97</td>
<td>94</td>
</tr>
</tbody>
</table>
Summary

- Both structural features and physicochemical descriptors (ADRIANA.Code) perform equally good

- However, they catch different information

- Therefore, the combined use of structural features and physicochemical descriptors leads to markedly improved models and predictions
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
  - 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**
  - 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%)
Support Vector Machine (SVM) Model

- **Descriptors (242 components)**
- **Automatic variable selection: 12 components**
  - $2D-AC_{\text{idem}}(5)$, $2D-AC_{\varphi}(3)$, $2D-AC_{\varphi}(6)$, $2D-AC_{\chi}(5)$, $2D-AC_{q}(1)$, $2D-AC_{q}(2)$, $2D-AC_{\chi}(6)$, $3D-AC_{\text{idem}}([5.8-5.9]Å)$, $n_{\text{acid_groups}}$, $n_{\text{aliphatic_amino}}$, $n_{\text{basic_n}}$, $r_3$

**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

isoCYP Webservice

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


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
MOSES.Metabolism Reaction Rules

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

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

\[
\begin{array}{c}
R - \text{NH}_2 \\
\rightarrow \\
R - \text{N} \\
\text{H} \\
\text{O}
\end{array}
\]

- Calculate reaction probabilities based on a reaction database (Metabolite, MDL-Symyx)

  - Conceivable metabolites: 1223
  - Observed metabolites: 122
  - Non-observed metabolites: 1101
  - Probability: \(\frac{122}{1223} = 0.10\)
Phase I Metabolism of Atorvastatin Lactone

- **Chemoselectivity**
  - *Aromatic hydroxylation* (●)
  - *Amide hydrolysis* (~)
  - *N-Dealkylation* (\)
Phase I Metabolism of Atorvastatin Lactone

Regioselectivity of aromatic hydroxylation

- Mono substituted ring
  - Ortho hydroxylation (●)
  - Meta hydroxylation (●)
  - Para hydroxylation (●)
- 1,4-substituted ring (●, ortho to first and meta to second substituent)
Predicted Ranks and Probabilities of Atorvastatin Lactone Metabolites

- **1 (31.5%)**
- **2 (15.8%)**
- **3 (10.3%)**
- **4 (4.0%)**
- **5 (3.5%)**
- **6 (2.6%)**
- **7 (1.7%)**
- **8 (0.8%)**
- **9 (0.7%)**
- **10 (0.4%)**
Experimentally observed Metabolite of Atorvastatin Lactone

- Metabolite predicted for atorvastatin with highest rank corresponds to the experimental observations.
Areas of Applications

- Hazard and risk assessment of chemicals
- Product safety of pharmaceuticals, cosmetics, food ingredients and other chemicals
- Computational toxicology
- Registration of chemical substances, e.g., REACH initiative
- Compound profiling
Workflow of Risk Assessment

- get data
- read-across
- QSAR prediction
- phys-chem prop
- toxicity
- biological assays
- reactivity
- degradation
- metabolism
- query
- representation

Persistence
Bioaccumulation
Toxicity

biodegradation
eco-toxicity
human health

Courtesy of Dr. Chihae Yang.
Application Example

- Priority-based Assessment of Food Additives (PAFA) by FDA
- PAFA contains administrative, chemical and toxicological information on over 2,000 substances directly added to food
- Dataset as of July 14, 2010
### Application Example: trans-Anethole

#### Chemical Information

<table>
<thead>
<tr>
<th>CAS #</th>
<th>CAS #</th>
<th>Main Term</th>
<th>Doc Type</th>
<th>CTR REG #</th>
<th>DOC</th>
<th>Comments</th>
<th>Cedi</th>
<th>Chemical Function/Sort Term</th>
</tr>
</thead>
<tbody>
<tr>
<td>0077</td>
<td>0941 02328</td>
<td>TRANS-ANETHOLE</td>
<td>A3P</td>
<td>182.60</td>
<td>0.0</td>
<td>Direct</td>
<td>ANETHOLE</td>
<td></td>
</tr>
<tr>
<td>0077</td>
<td>0941 02328</td>
<td>TRANS-ANETHOLE</td>
<td>A3P</td>
<td>182.60</td>
<td>0.0</td>
<td>Flavour</td>
<td>ANETHOLE</td>
<td></td>
</tr>
<tr>
<td>0077</td>
<td>0941 02328</td>
<td>TRANS-ANETHOLE</td>
<td>A3P</td>
<td>182.60</td>
<td>0.0</td>
<td>Grass</td>
<td>ANETHOLE</td>
<td></td>
</tr>
</tbody>
</table>

#### Toxicological Studies

<table>
<thead>
<tr>
<th>Doc #</th>
<th>Study</th>
<th>Completeness</th>
<th>Effect #1</th>
<th>Source</th>
<th>Let</th>
<th>Unit</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>99</td>
<td>77</td>
<td>25</td>
<td>GENOTOXIC IN PRESENCE OF EXOGENOUS METABOLIC ACTIVATION</td>
<td>MUTAT RES 101:127-140</td>
<td>0.3</td>
<td>mg/plate</td>
<td>1982</td>
</tr>
<tr>
<td>100</td>
<td>77</td>
<td>25.1</td>
<td>NO EFFECTS</td>
<td>MUTAT RES 101:127-140</td>
<td>0.6</td>
<td>mg/plate</td>
<td>1982</td>
</tr>
<tr>
<td>101</td>
<td>77</td>
<td>25.2</td>
<td>GENOTOXIC IN ABSENCE OF EXOGENOUS METABOLIC ACTIVATION</td>
<td>MUTAT RES 101:127-140</td>
<td>0.0</td>
<td>mg/plate</td>
<td>1982</td>
</tr>
<tr>
<td>102</td>
<td>77</td>
<td>26</td>
<td>NO EFFECTS</td>
<td>BULL ENVIRON CONTAM TOXIC</td>
<td>0.06</td>
<td>mg/plate</td>
<td>1982</td>
</tr>
<tr>
<td>103</td>
<td>77</td>
<td>27</td>
<td>NO EFFECTS</td>
<td>ENVIRON MUTAGEN 6 (SUPPL 7)</td>
<td>0.28</td>
<td>mg/plate</td>
<td>1986</td>
</tr>
<tr>
<td>104</td>
<td>77</td>
<td>34</td>
<td>CYTOTOXIC (LD50 IS LESS THAN 5 MILLIMOLAR)</td>
<td>FOOD CHEM TOXICOL 30:4</td>
<td>1</td>
<td>mM</td>
<td>1992</td>
</tr>
<tr>
<td>105</td>
<td>77</td>
<td>40</td>
<td>NO EFFECTS</td>
<td>FOOD CHEM TOXICOL 34:337-3</td>
<td>0.1</td>
<td>mM</td>
<td>1996</td>
</tr>
<tr>
<td>106</td>
<td>77</td>
<td>40.1</td>
<td>NO EFFECTS</td>
<td>FOOD CHEM TOXICOL 34:337-3</td>
<td>0.5</td>
<td>mM</td>
<td>1998</td>
</tr>
<tr>
<td>107</td>
<td>77</td>
<td>41</td>
<td>NO EFFECTS</td>
<td>MUTAT RES 325:129-136</td>
<td>1</td>
<td>mg/kg bw</td>
<td>1994</td>
</tr>
<tr>
<td>108</td>
<td>77</td>
<td>42</td>
<td>NO EFFECTS</td>
<td>MUTAT RES 326:199-209</td>
<td>0.75</td>
<td>mg/plate</td>
<td>1995</td>
</tr>
<tr>
<td>109</td>
<td>77</td>
<td>42.1</td>
<td>NO EFFECTS</td>
<td>MUTAT RES 326:199-209</td>
<td>0.084</td>
<td>mg/plate</td>
<td>1995</td>
</tr>
<tr>
<td>110</td>
<td>77</td>
<td>42.2</td>
<td>NO EFFECTS</td>
<td>MUTAT RES 326:199-209</td>
<td>0.084</td>
<td>mg/plate</td>
<td>1995</td>
</tr>
</tbody>
</table>
An Expert System for Chemical Evaluation and Risk Estimation System

Welcome to MOSES.RiskAssessment

Welcome to the MOSES.RiskAssessment Demonstration

- Structure searching
- Data searching
  - Chemical exposure information
  - Toxicity studies information
- Analysis
  - TTC and QSAR analysis
- Toxicity prediction
- Metabolism prediction

Start the demo.

Please start the demo by entering the Query page.
Please select a method how to find a chemical record from the database:

- By Structure
- Name
- CAS RN
- ID
- Upload a file
- Enter structure

Enter SMILES string in the text field or sketch the structure:

Sketch Molecule
Enter Query Structure

Please select a method how to find a chemical record from structure or text:

- **By structure**
- **Sketch**

Sample Database

Search method:

- Exact Structure
- Substructure
- Similar Substructure
Structure found in the database.

trans-Anethole; Benzene, 1-methoxy-4-(1E)-1-propenyl-; (E)-Anethole; Anisole, p-propiolyl-; (E)- (8CI); Anisole, p-propenyl-, trans-; Benzene, 1-methoxy-4-(1-propenyl)-, (E)-

Compound Information
- ID: 2208
- CAS-RN: 4180-23-8
- Trade Name: trans-Anethole
- Chemical Name: Benzene, 1-methoxy-4-(1E)-1-propenyl-
- Substance Use Types: Flavouring agent

View summarized toxicity data: show/hide

Generate a report for the database record.

Next steps.
Select the next step to be performed.
Structure found in the database.

trans-Anethole; Benzene, 1-methoxy-4-(1E)-1-propenyl; (E)-Anethole; Anisole, p-propenyl; (E)- (8CI); Anisole, p-propenyl, trans-, benzene, 1-methoxy-4-(1-propenyl), (E)-

Compound Information
ID 2208
CAS-RN 4180-23-8
Trade Name
Chemical Name trans-Anethole
Substance Use Types Flavoring agent

Study Information

Study Information 1, Source: cfsan-pafa

<table>
<thead>
<tr>
<th>Study Type</th>
<th>Species</th>
<th>Strain</th>
<th>Metabolic Activation</th>
<th>All Doses</th>
<th>Dose Unit</th>
<th>Test Cytotoxicity</th>
<th>Test Precipitation</th>
<th>Test Call</th>
<th>Study Call</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Bacterial</td>
<td>Salmonella typhimurium</td>
<td>TA98</td>
<td>Absent</td>
<td>280.0 microg/plate</td>
<td>microg/plate</td>
<td>Negative</td>
<td>Negative</td>
<td>Negative</td>
<td>Negative</td>
</tr>
<tr>
<td>2 Bacterial</td>
<td>Salmonella typhimurium</td>
<td>TA100</td>
<td>Absent</td>
<td>280.0 microg/plate</td>
<td>microg/plate</td>
<td>Negative</td>
<td>Negative</td>
<td>Negative</td>
<td>Negative</td>
</tr>
<tr>
<td>3 Bacterial</td>
<td>Salmonella typhimurium</td>
<td>TA1537</td>
<td>Absent</td>
<td>280.0 microg/plate</td>
<td>microg/plate</td>
<td>Negative</td>
<td>Negative</td>
<td>Negative</td>
<td>Negative</td>
</tr>
<tr>
<td>4 Bacterial</td>
<td>Salmonella typhimurium</td>
<td>TA100</td>
<td>Absent</td>
<td>280.0 microg/plate</td>
<td>microg/plate</td>
<td>Negative</td>
<td>Negative</td>
<td>Negative</td>
<td>Negative</td>
</tr>
</tbody>
</table>
Structure found in the database.
trans-Anethole; Benzone, 1-methoxy-4-(1E)-1-propenyl; (E)-Anethole; Anisole, p-propenyl, (E)-8Cl; Anisole, p-propenyl, trans; Benzone, 1-methoxy-4-(1-propenyl), (E)-

**Compound Information**
- **ID**: 2208
- **CAS-RN**: 4180-23-8
- **Trade Name**: 
- **Chemical Name**: trans-Anethole
- **Substance Use Types**: Flavouring agent

**Next steps.**
- Select the next step to be performed.
  - Run analog search
  - Run TTC analysis
  - Run toxicity prediction
  - Run metabolism prediction

**Report generation**
- Generate a report for the database record.
Analogs found in the database.

60 analogs found in the database. (The first 10 most similar analogs are shown below).

List with analogs

Analogs with data
TTC analysis for the query compound.

No TTC alerts for that compound.

Run Models for a toxicity prediction

Important note: the prediction may take some time. Please wait until the results are presented...
Toxicity prediction

Next steps:

- Run analog search
- Run TTC analysis
- Run toxicity prediction
- Run metabolism prediction

Select the next step to be performed.

Run

Powered by MOSES from Molecular Networks
Innovating Chemical Discovery
Results for the toxicity prediction

Toxicity prediction for the query compound.

Endpoint: Genetic toxicity - bacterial mutagenesis

trans-Anethole; Benzene, 1-methoxy-4-(1E)-1-propenyl-; (E)-Anethole; Anisole, p-propenyl-, (E)- (8CI); Anisole, p-propenyl-, trans-; Benzene, 1-methoxy-4-(1 propenyl), (E)

- Probability of being positive: 0.222
- Prediction: negative

Further steps.
Metabolism prediction
List with metabolites

Generated 4 metabolites.
Toxicity prediction for query compound and all metabolites.
Toxicity prediction for the query compound.

Endpoint: Genetic toxicity - bacterial mutagenesis

Trans-Anethole; Benzen, 1-methoxy-4-(1E)-1-propenyl; (E) Anethole; Anisole, p-propenyl, (E)- (8CI); Anisole, p-propenyl, trans-; Benzen, 1-methoxy-4-(1-propenyl), (E)-

Probability of being positive: 0.222
Prediction: negative

Toxicity prediction for the set of 4 metabolites

C9H10O

Probability of being positive: 0.165
Prediction: negative
Toxicity prediction for the set of 4 metabolites

**C9H10O**
- Probability of being positive: 0.165
- Prediction: negative

**C10H12O2**
- Probability of being positive: 0.204
- Prediction: negative

**C10H12O2**

Toxic after metabolic activation?

Further steps.

- Generate a report for the session.
- Run a new search.
Features & Functionality

- Knowledge base for hazard and risk assessment of chemicals
- Database lookup by text-based, analog and similarity searches
- Retrieval of available study information for query compound and analogs
- Generation and evaluation of metabolites of query and analogs (including CYP isoform specificity)
- Analysis tools for query, analogs and their metabolites
- TTC analysis
- QSAR predictions of toxicity endpoints (e.g., Ames mutagenicity)
- Report generation
- Fully web-based, easy-to-use user interface
Metabolism in the ToxCast Dataset
Identification of Parent/Metabolite Pairs in the ToxCast Dataset

**Approach**
- Generate all conceivable metabolites for the compounds in the ToxCast dataset with MOSES.Metabolism
- Determine the intersection of the set of all generated metabolites with the set of compounds in the ToxCast dataset

**Results**
- MOSES.Metabolism generated 1826 metabolites for the 309 unique compounds from the ToxCast dataset (approx. six metabolites per parent compound on average)
- Fourteen parent/metabolite pairs could be identified
Most Frequently Observed Reaction Types in ToxCast

- Aromatic hydroxylation of a phenyl ring: 543
- O-Dealkylation: 99
- Ester hydrolysis: 77
- N-Dealkylation: 71
- Aromatic amine oxidation: 64
- Amide hydrolysis: 63
- Aliphatic hydroxylation of a primary carbon atom next to a secondary carbon atom: 59
- Aromatic hydroxylation of 1,2-substituted aromatic ring in 4 position: 52
- O-Sulphation: 52
**Parent Compounds**

- **Metam-sodium hydrate**
- **Malathion**
- **Dimethylphthalate**
- **Diazinon**
- **Chlorpyrifos**
- **Atrazine**
- **Methoxychlor**
- **Diethylhexylphthalate**
- **Dibutylphthalate**
- **Mancozeb**
- **Maneb**
- **Metiram-zinc**
Extension of Reaction Rules

Reaction rules for oxidative desulfuration were added to MOSES.Metabolism in order to identify the following parent compound metabolite pairs in the ToxCast data set:

- Malathion – Malaoxon
- Diazinon – Diazoxon
- Chlorpyrifos – Chlorpyrifos oxon
- Metam-sodium hydrate – Methylisothiocyanate
Parent/Metabolite Pairs 1 & 2

- **N-Deethylation**
  - Probability: 0.59  
  - Rank: 1

- **N-Deisopropylation**
  - Probability: 0.31  
  - Rank: 2
Parent/Metabolite Pairs 3 & 4

- **Esterhydrolysis**
  - ![Chemical Structure](image1)
  - *Probability: 0.67*  
  - *Rank: 1*

- **Esterhydrolysis**
  - ![Chemical Structure](image2)
  - *Probability: 0.67*  
  - *Rank: 1*
Parent/Metabolite Pairs 5 & 6

- **Esterhydrolysis**
  - Probability: 0.67
  - Rank: 1

- **N-Demethylation of RNMe₂**
  - Probability: 0.41
  - Rank: 1
Parent/Metabolite Pairs 7 & 8

- **N-Demethylation**
  - Probability: 0.17
  - Rank: 3

- **Esterhydrolysis**
  - Probability: 0.67
  - Rank: 1
Parent/Metabolite Pairs 9 & 10

- **N-Acetylation of heterobonded NH$_2$**
  - Probability: 0.08  
  - Rank: 1

- **Esterhydrolysis**
  - Probability: 0.67  
  - Rank: 1
**Parent/Metabolite 11**

- Oxidative desulfuration
  - *Malathion – Malaoxon*

![Chemical structures](image)

[Image of chemical structures]
Oxidative desulfuration

*Diazinon – Diazoxon*
Oxidative desulfuration

*Chlorpyrifos – Chlorpyrifos oxon*
Oxidative desulfuration

Metam-sodium hydrate – Methylisothiocyanate
Missing Pairs

Parent – Metabolite Pairs
- Mancozeb/Maneb/Metiram – Ethylenethiourea
- Methoxychlor – HPTE

Reason
- Missing rule; metal complex
- O-Demethylation in two positions; rules were only applied ones
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.)

- Use of the MOSES.Metabolism rule base for metabolic screening (metabolic profile; metabolic fingerprints)
## 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        | ![Structure 370](image) | 2 0 0 0 0 1 1 0 ... |                      |                                               |                                               |                        |              |                   | ...    |
| 8038       | ![Structure 8038](image) | 6 0 1 1 0 0 0 0 ... |                      |                                               |                                               |                        |              |                   | ...    |
Fingerprint View of Metabolic Reactivity Classes

- Aromatic hydroxylation
- N-acetylation
- O-dealkylation

ToxCast chemicals

- Primary ether cleavage-Br, Cl
- O-sulfation

MetaboGen reactivity classes
Metabolic Reactivity Profile of the ToxCast Dataset

Probability of conceivable metabolic reactions

<table>
<thead>
<tr>
<th>Reactions</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-48</td>
<td>Hydroxylation</td>
</tr>
<tr>
<td>49-62</td>
<td>Hydrolysis</td>
</tr>
<tr>
<td>63-83</td>
<td>N-,O-Dealkylation</td>
</tr>
<tr>
<td>84-109</td>
<td>Conjugation</td>
</tr>
<tr>
<td>110-115</td>
<td>others</td>
</tr>
</tbody>
</table>

309 compounds; 115 reaction rules from MOSES.Metabolism (2009-06-11)
The Metabolic Reactivity Profiles provide an easy method for rapidly screening for potential metabolites in large datasets of compounds.
Molecular Networks

- Innovation company for Chemoinformatics
  - "Chemoinformatics: the processing of chemical information by informatics tools"

- Mission statement
  - Increasing the quality and productivity of discoveries in chemical, pharmaceutical and biotechnology R&D

- Products and services
  - Broad range of scientific software products
  - Consulting and research services
  - Contract development
Molecular Networks Provides Applications for ...

- Drug design and property prediction
- Synthesis design and reaction prediction
- Risk Assessment of Chemicals
- Prediction of metabolism
  - *Endogenous metabolism*
  - *Metabolism of xenobiotics (drugs, agrochemicals, ...)*
- Design of biotechnological processes
- Data warehousing & mining
- Handling, processing and manipulation of chemical structure, reaction and related information
Acknowledgements

- Dr. Aleksey Tarkhov
- Dr. Bruno Bienfait
- Dr. Christof Schwab
- Dr. Jörg Marusczyk
- Dr. Lothar Terfloth
- Dr. Oliver Sacher
- Dr. Thomas Kleinöder

and my former coworkers at Computer-Chemie-Centrum University of Erlangen

Collaboration

- Dr. Chihae Yang