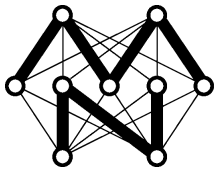


Molecular Networks
Inspiring Chemical Discovery

# Modeling Chemical Reactions in Drug Design and Chemical Risk Assessment


**Johann Gasteiger**

**Computer-Chemie-Centrum**

**University of Erlangen-Nuremberg**

**and**

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

# Chemical Reactions in Drug Design

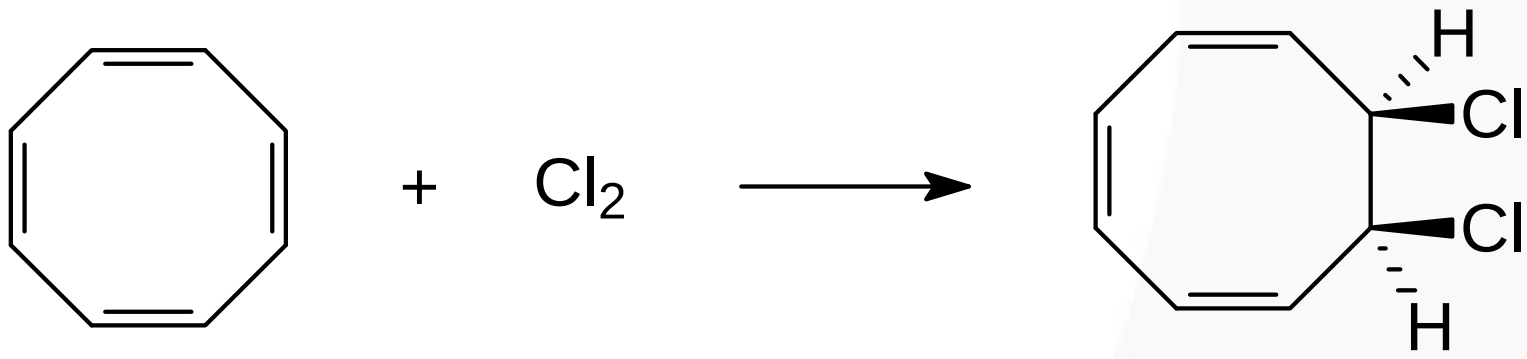
- **Which biochemical pathways are critical for a certain disease?**
- **How can I inhibit a certain biochemical reaction?**
- **How can I synthesize a desired compound?**
- **How stable is my cpd?**
- **What is the $pK_a$ value of my cpd?**

# Chemical Reactions in Risk Assessment

- How will my cpd react with proteins or DNA (reactive toxicity)
- How will my cpd be metabolized?
- Are the metabolites reactive/toxic?

# Reactions of Cyclooctatetraene

- *cis* - halogenation



# Outline

- Modeling physicochemical effects
- Prediction of chemical reactions
- Design of organic syntheses
- Biochemical pathways
- Risk assessment
- Metabolism of xenobiotics

# 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: $q_\sigma$ 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)**

# Calculation of Physicochemical Effects

## Charge distribution

J. Gasteiger, M. Marsili, *Tetrahedron* **36**, 3219 (1980)

## Inductive effect

J. Gasteiger, M. G. Hutchings, *Tetrah. Lett.* **24**, 2541 (1983)

## Resonance effect

J. Gasteiger, H. Saller, *Angew. Chem. Int. Ed. Engl.* **24**, 687 (1985)

## Polarizability effect

J. Gasteiger, M. G. Hutchings, *J. Chem. Soc. Perkin 2*, 559 (1984)

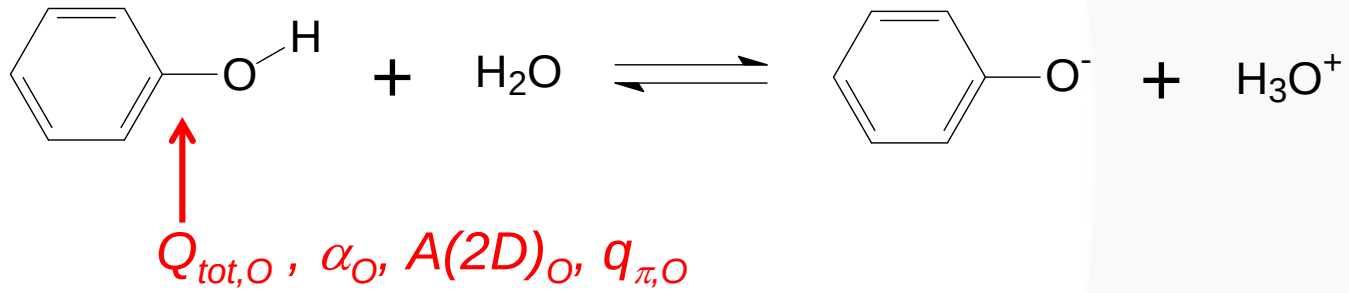
## Bond dissociation energy

J. Gasteiger, *Comp. Chem.* **2**, 85 (1978)

# Prediction of Chemical Reactivities

- **Gas phase reactions (proton affinities, acidities)**
- **p$K_a$ values**
- **General nucleophilicity scale**
- **Hydrolysis of amides**
- **Stability of compounds in DMSO/$H_2O$**

# A $pK_a$ Model for Phenols




$$pK_a = 7.0 - 48.4Q_{tot,O} - 2.0\alpha_O + 1.2A(2D)_O + 154.2q_{\pi,O} + 2.3I_{o-carboxy}$$

n = 452, $r^2$ = 0.81, s = 0.95

# Reactivity Prediction

- **pKa calculation of aliphatic carboxylic acids and alcohols**

J.Zhang, T.Kleinöder, J. Gasteiger**,** *J. Chem. Inf. Model.,***2006,** *46***,** 2256-2266

- **Prediction of life-time of screening compounds in DMSO solution**

E.Zitha-Bovens, P.Maas, D.Wife, J.Tijhuis, Q.-N.Hu, T.Kleinöder, J.Gasteiger, *J. Biomol. Screen.,* **2009,***14,* 557-565

# Design of Organic Syntheses

**The amount of information to be processed and the need to make decisions between many alternatives**

→ **suggests the use of computers**

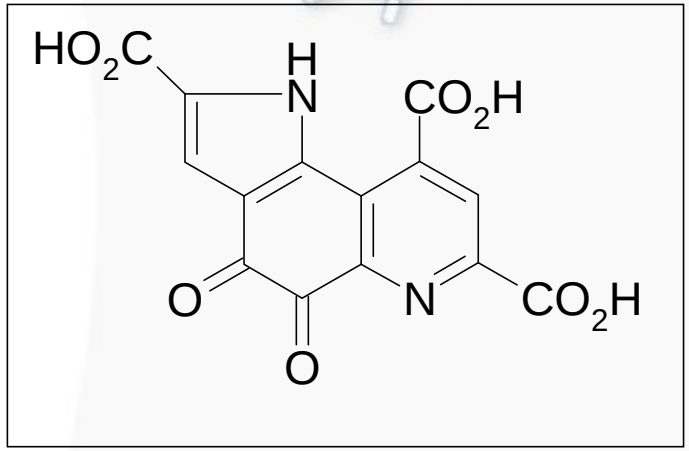
# THERESA – THE REtroSynthesis Analyzer

- Extract reaction types from a reaction database
- Store these reaction types in a separate database
- Apply these reaction types onto a synthesis target
- The retrosynthetic steps are ranked for usefulness by a series of criteria
- Compare the found retrosynthetic steps with known reactions from the database
- Repeat this procedure until all precursors are available starting materials

  ➡ Can generate completely novel synthesis plans (works with reaction types!)

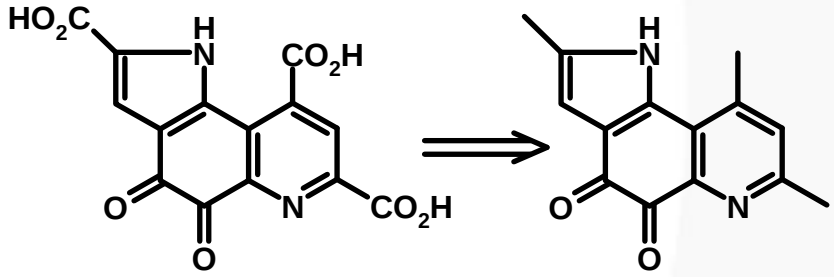
# Example: Methoxatin (coenzyme PQQ)

- **coenzyme for redox catalysis discovered in 1979 in methylotrophic bacteria (methane and methanol oxidation)**
- **present in mammals: growth factor, tissue protective agent, vitamin**
- **eight published syntheses (8 to 13 steps)**

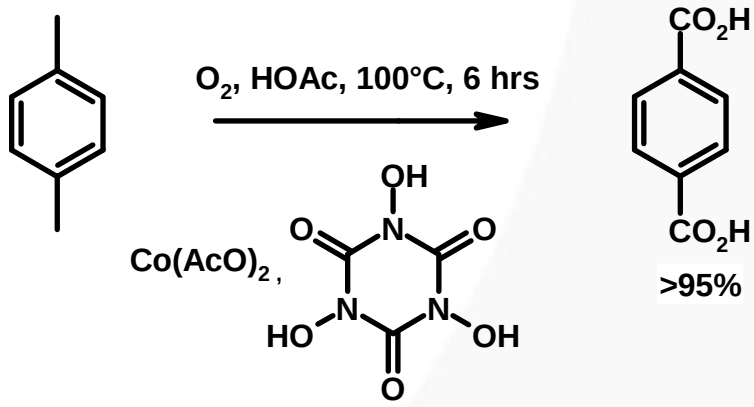


# Oxidation Step (1)

- **Suggested retroreaction:**



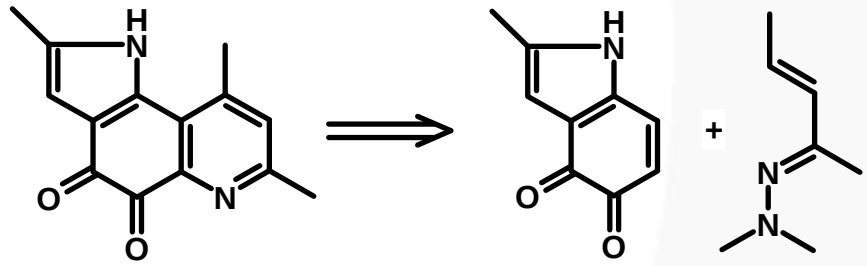

- **Published reaction:**



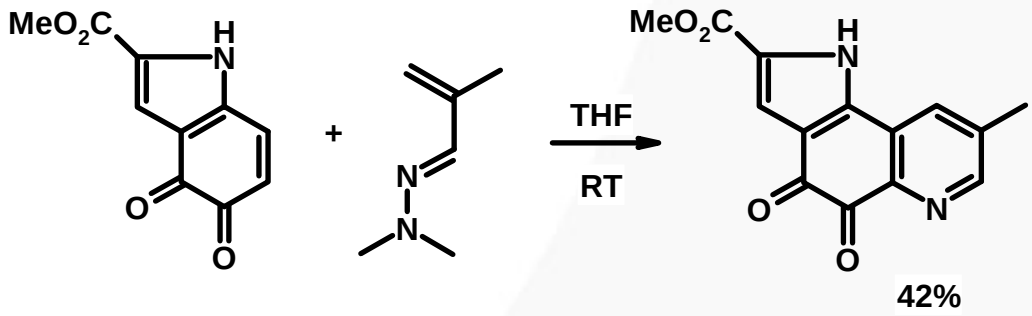

Hirai, N.; Sawatari, N.; Nakamura, N.; Sakaguchi, S.; Ishii, Y. *J. Org. Chem.* **2003**, *68*, 6587-6590

# Diels-Alder Step (2)

- **Suggested retro-reaction:**

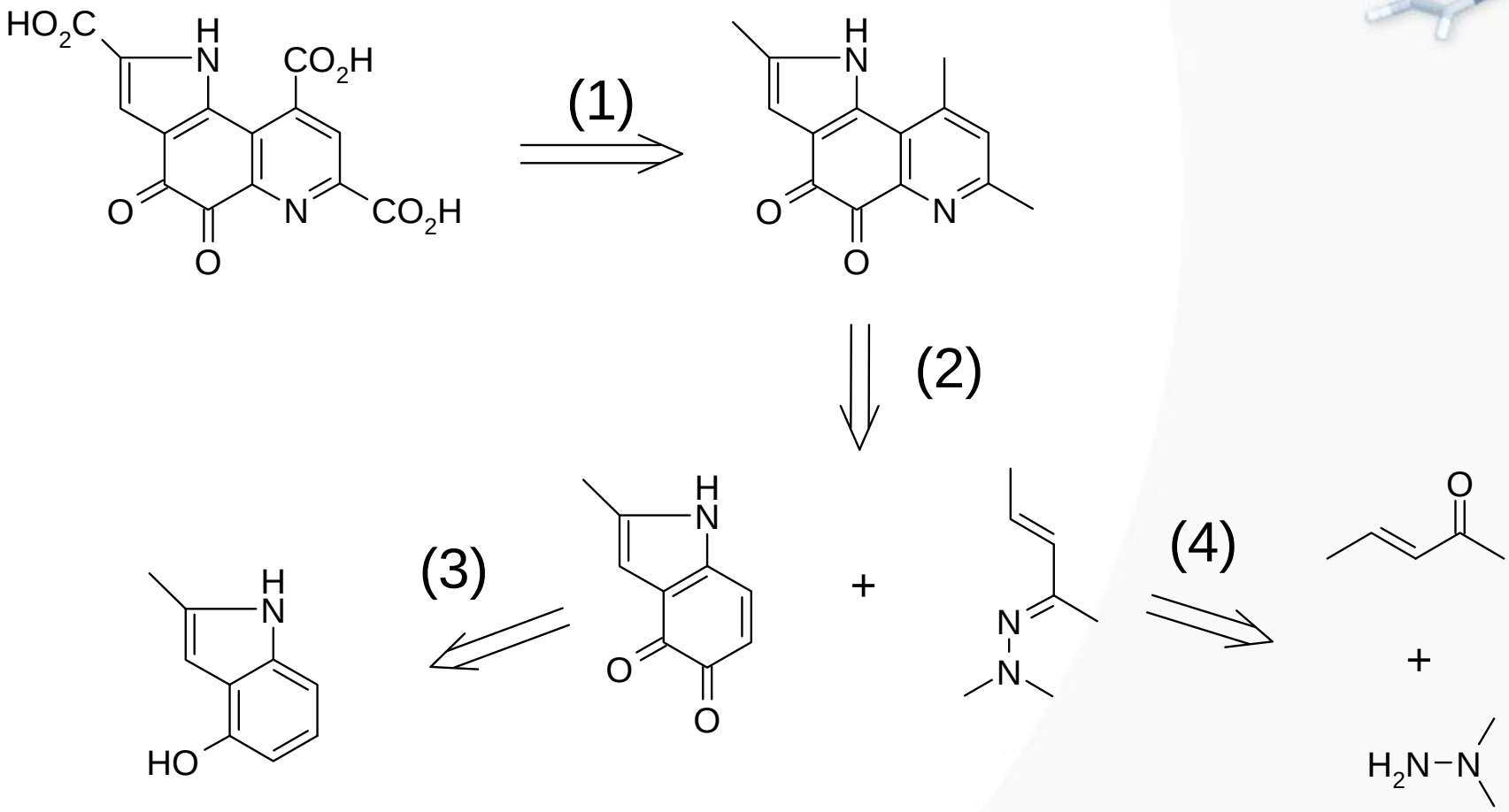



- **Published reaction:**

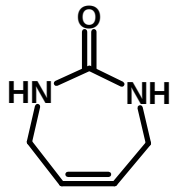



Genisson, V. B.; Nebois, P.; Domard, M.; Fillion, H. *Chem. Pharm. Bull.* **2000**, *48*, 893-894

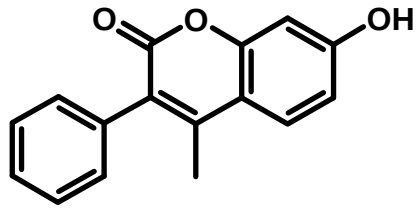
# Methoxatin: Suggested Synthesis with THERESA (4 Steps)



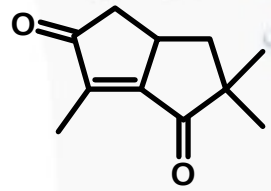
# THERESA Case Studies (1)



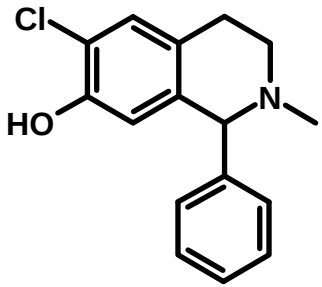

1 step (CCR-2005)



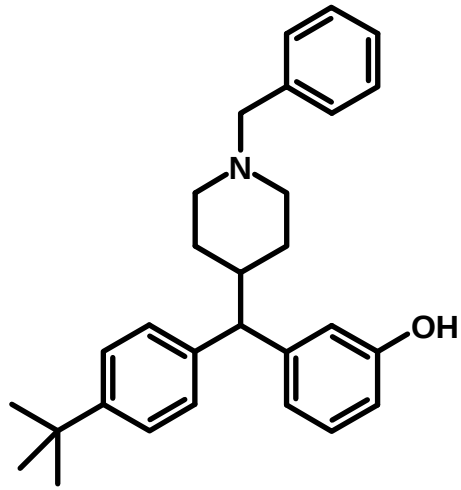

1 step (CCR-2005)



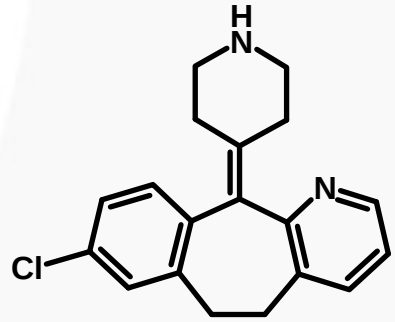

2 steps (CCR-2005)




2 steps (ChemInform)



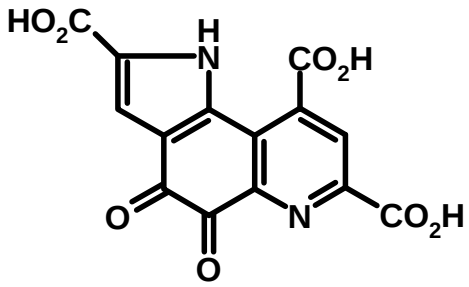

3 steps (ChemInform)



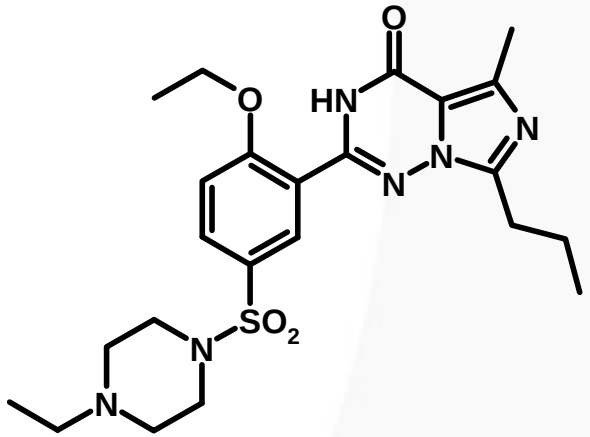

4 steps (Beilstein)

Reaction database used

# THERESA Case Studies (2)




5 steps (ChemInform)




6 steps (ChemInform)

# Advantages of THERESA

- **Driven by reaction types extracted from reaction databases**
  - ***Can handle databases with more than 1 Million reactions***
  - ***Handling of diverse chemistry (heterocycles, rearrangements, ...)***
  - ***Provides literature data***
- **Can design syntheses for novel target structures (also those not in the reaction database!)**
- **Provides ideas to the organic chemist**
- **Easy to use**
  - ***Web-based user interface***

# Endogenous (Central) Metabolism

**Nutrients, food**

# What is a Chemical Reaction?

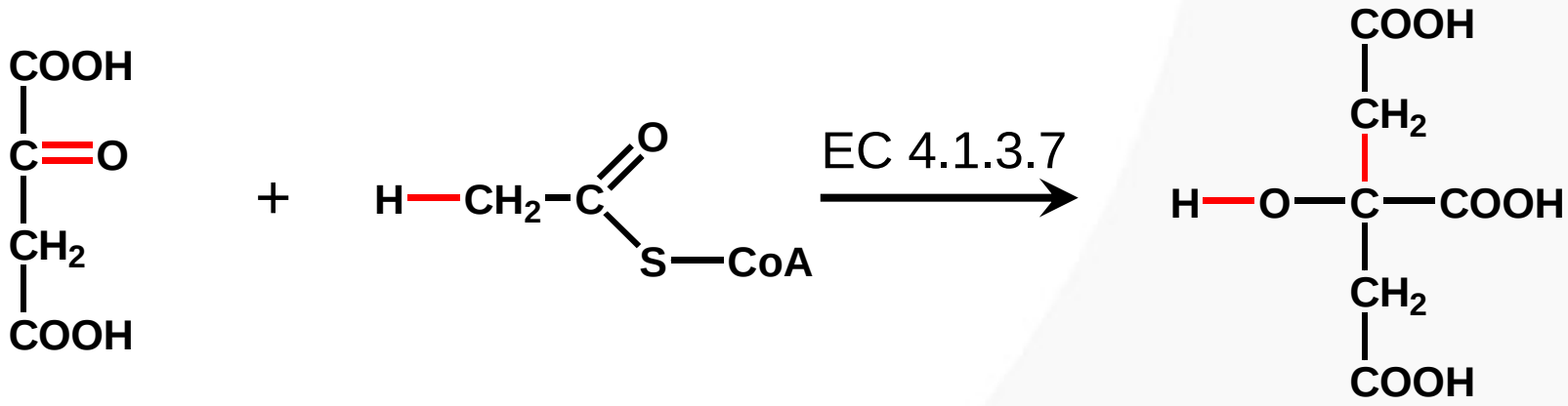
the biologist

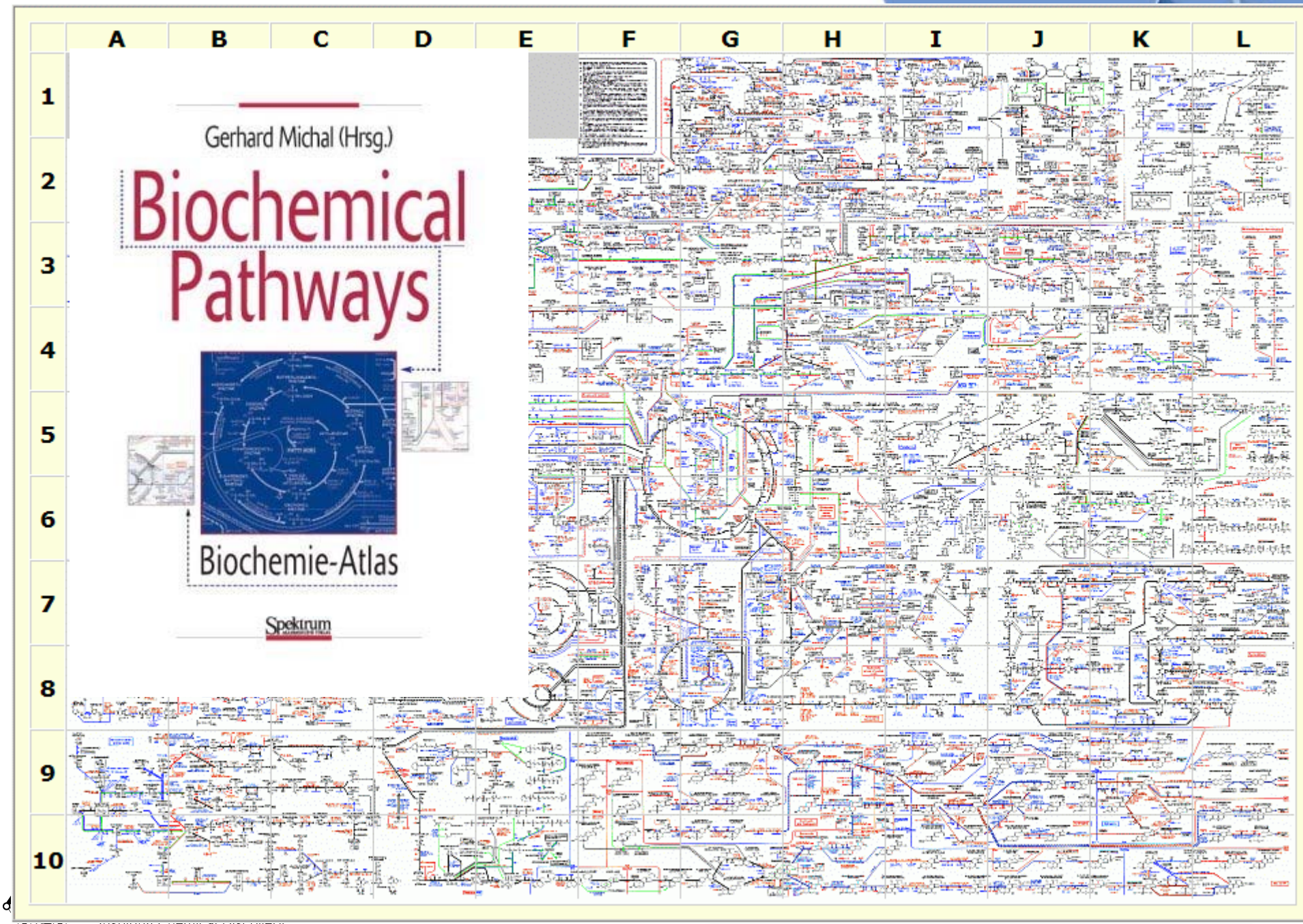
an event influenced by a gene, a protein (→ bioinformatics)

oxalacetate + acetyl-CoA $\xrightarrow{\text{EC 4.1.3.7}}$ citrate

the chemist

an event breaking and making bonds (→ chemoinformatics)




Gerhard Michal (Hrsg.)
Biochemical Pathways
Biochemie-Atlas
Spektrum

# BioPath.*Database*

- **Molecules and reactions are stored with atomic resolution:**
  **- molecules as connection tables**
  **- reactions with reaction center marked**

| | ***version 1*** | ***version 2*** |
|---|---|---|
| ➢ | ***1,175 structures*** | ***2,074 structures*** |
| ➢ | ***1,545 reactions*** | ***2,881 reactions*** |

M.Reitz, O.Sacher, A.Tarkhov, D.Trümbach, J.Gasteiger, *Org. Biomol. Chem.,* **2004,** *2,* 3226-3237.

**http://www.molecular-networks.com/biopath/**

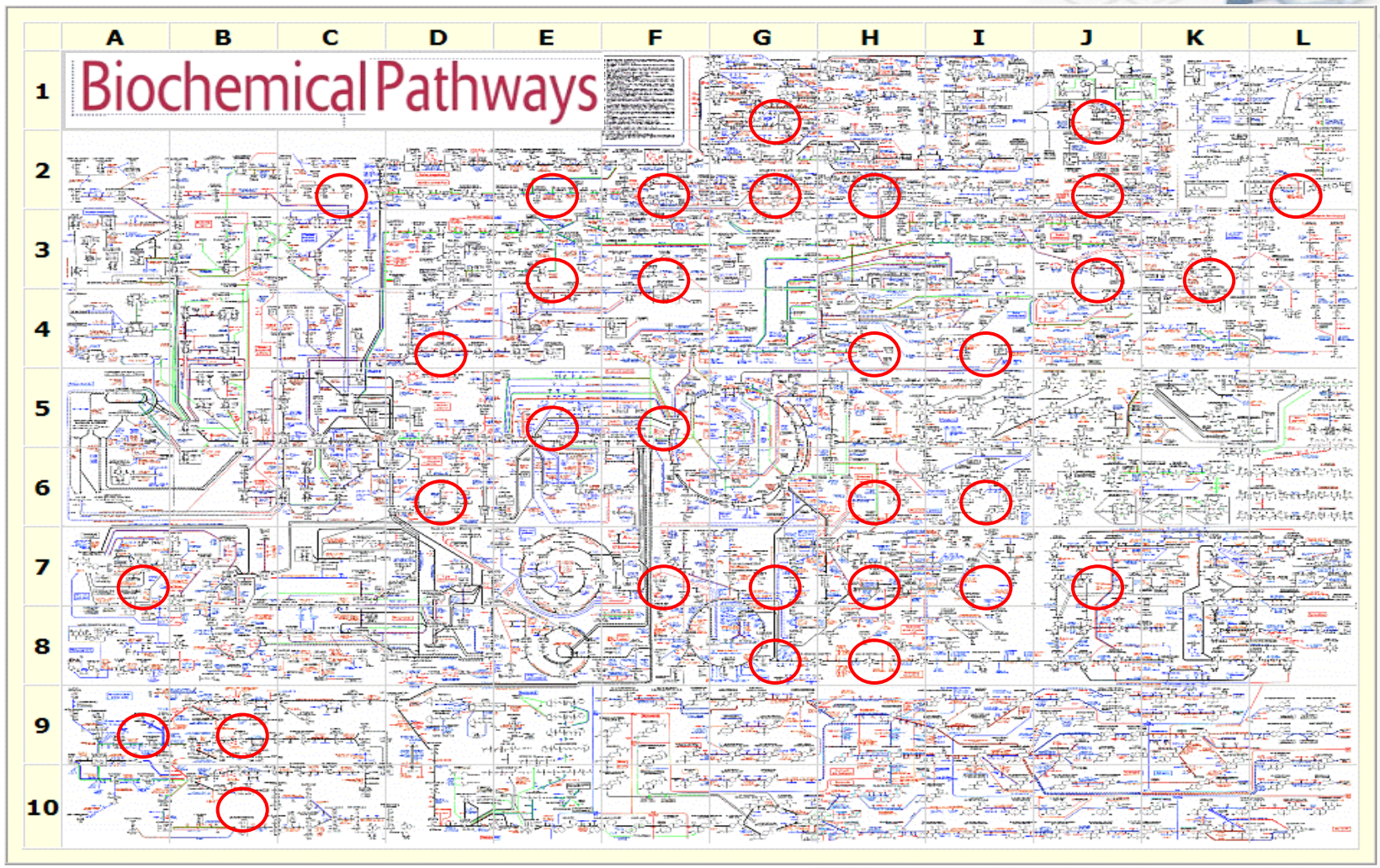
# Search on Wall Chart

- Search on the wall chart of Biochemical Pathways for all positions where

- L-glutamate

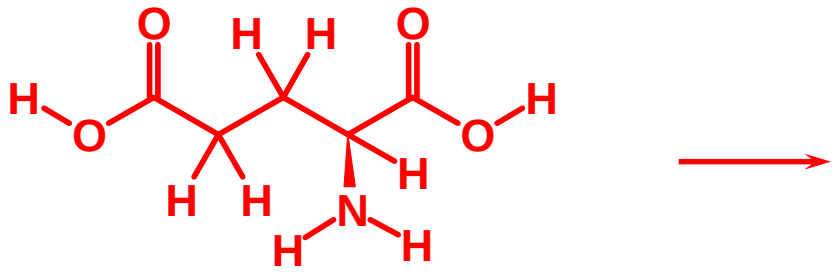
- can be found

# Searching for L-Glutamate on Wall Chart



# Search for L-Glutamate

- **search for biochemical reactions with L-glutamate as reactant**



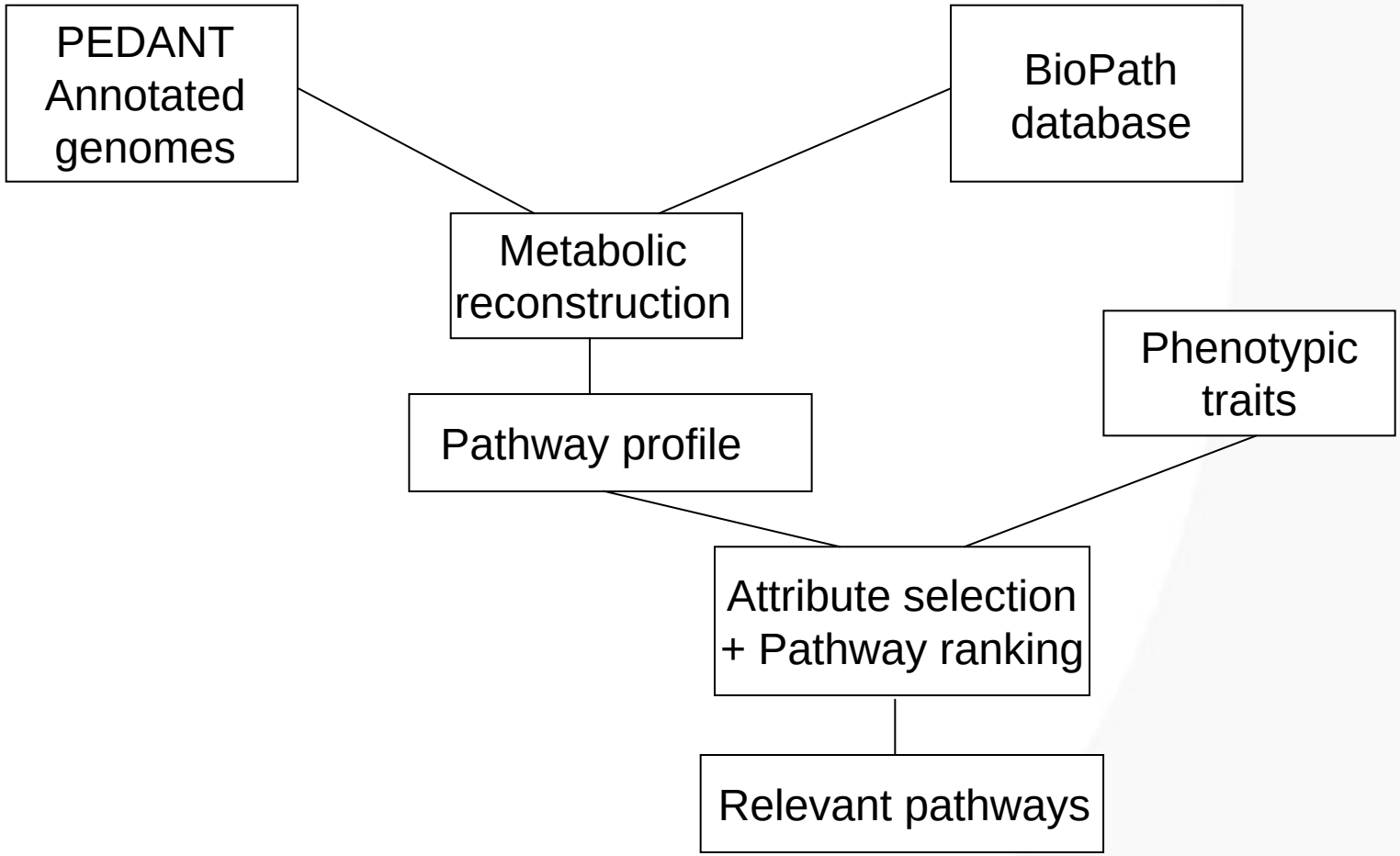

**29 hits**

- **among these hits are biosynthesis of L-valine, L-isoleucine, L-leucine, tetrahydrofolate, ...**

# Applications of BioPath.*Database*

- **Search for enzyme inhibitors**
  - M.Reitz, A.von Homeyer, J.Gasteiger**,** *J.Chem.Inf.Model.,* **2006,**
  - *46,* 2324-2332
- **Search for similar enzymes**
  - O.Sacher, M.Reitz, J.Gasteiger**,** *J.Chem.Inf.Model.,* **2009,** *49,*1525-1534
- **Search for alternative pathways**
- **Essential pathways of diseases**
  - G.Kastenmüller, J.Gasteiger, H.W.Mewes**,** *Bioinformatics***, 2008,** *24,* i56-i62
  - G.Kastenmüller, M.E.Schenk, J.Gasteiger, H.W.Mewes, *Genome Biology*,
  - **2009,***10,* R28

# Uncovering Metabolic Pathways to Phenotypic Traits



# Periodontal Disease

- **Human oral flora:**
  - ***> 700 species***
- **PEDANT:**
  - ***15 fully sequenced oral genomes***
  - ***(incl. 4 of 6 periodontal pathogens)***
- **BioPath**
  - ***68 global pathways***
  - ***306 smaller pathways***

# Relevant Pathways for Phenotype *Periodontal Disease Causing*

- **Glutamate fermentation**
- **Biosynthesis of L-proline**
- **Biosynthesis of 5-formimino-THF**
- **Conversion of l-glutamate to L-proline**
- **Conversion of l-glutamate to l-ornithine**
- **Degradation of l-histidine to l-glutamate**


➢ G. Kastenmüller, M.E. Schenk, J. Gasteiger, H.W.Mewes, Genome Biology, **2009,***10,* R28

# Relevant Pathways for Phenotype *Periodontal Disease Causing*

- **Glutamate fermentation**
- **Biosynthesis of L-proline**
- **Biosynthesis of 5-formimino-THF**
- **Conversion of l-glutamate to L-proline**
- **Conversion of l-glutamate to l-ornithine**
- **Degradation of l-histidine to l-glutamate**

➡ **Several of these pathways produce $NH_3$**

➡ **Cytotoxic $NH_3$ plays a major role in periodontal disease**

- G. Kastenmüller, M.E. Schenk, J. Gasteiger, H.W.Mewes, Genome Biology, **2009,***10,* R28

# BioPath – Summary

- **BioPath is a rich database**
- **BioPath.*Explore* provides powerful search capabilities**
- **Gives insight into the geometric requirements of enzyme reactions can be gained**
- **Electronic effects at the reaction center provide more detailed information for enzyme classification**
- **Combination with information on genotypes gives insight into diseases**
- **Serves as a basis for green chemistry/metabolic engineering (BioPath.*Design*)**

# Risk Assessment of Chemicals

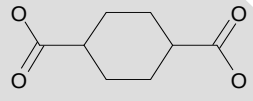
## Legislation in the European Union: REACH; Cosmetics Directive

# Risk Assessment of Chemicals

- **REACH – Registration, Evaluation, Authorization and restriction of CHemicals**
- **Only those chemicals are allowed to be manufactured or imported into the European Union that are proven to be safe**
- **Law since June 1, 2007**
- **Chemicals have to be accepted until Dec 1, 2013**
- **Applies to about 35,000 chemicals**
- **Testing is time-consuming and expensive**

→ **Use computational methods for ranking of chemicals**

# Workflow of Risk Assessment



**Chemical Speciation** → **Collection of Data** → **Categorization Prediction** → **PBT Assessment**

- query
- representation

- **reactivity**
- **degradation**
- **metabolism**

- phys-chem prop
- toxicity
- biological assays

- get data
- read-across
- QSAR prediction

- biodegradation...
- eco-toxicity...
- human health..

Persistence
Bioaccumulation
Toxicity

Molecular Networks
Inspiring Chemical Discovery

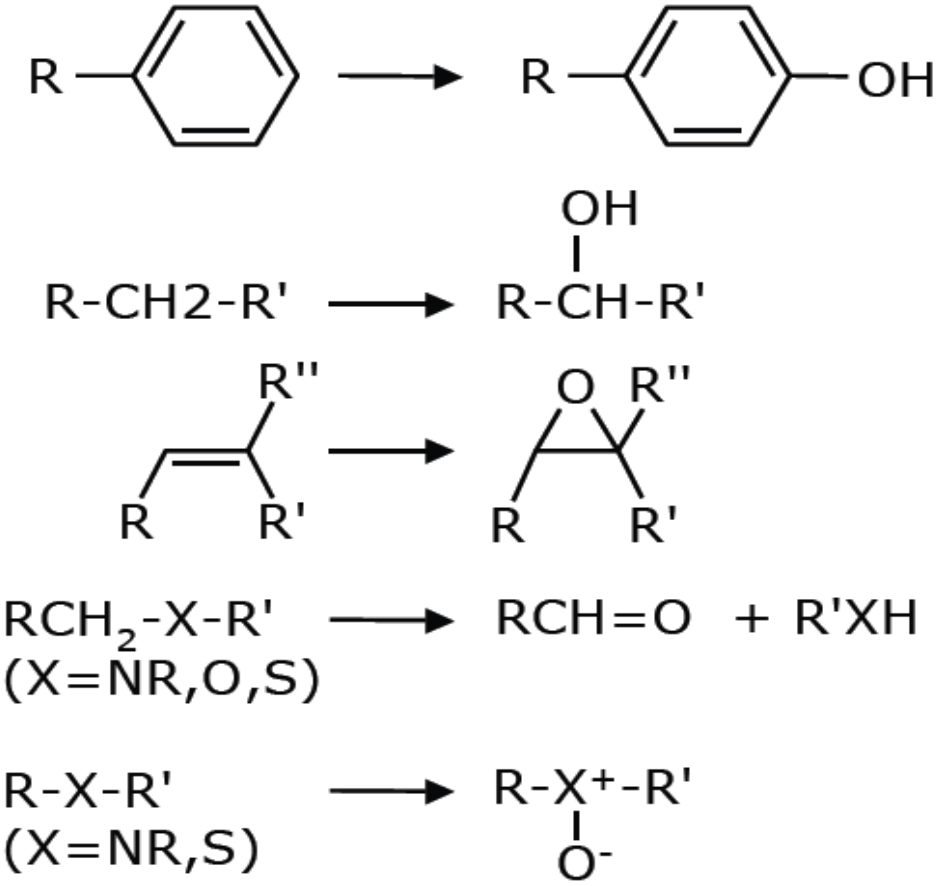
*Slide courtesy Dr. Chihae Yang*

# 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**

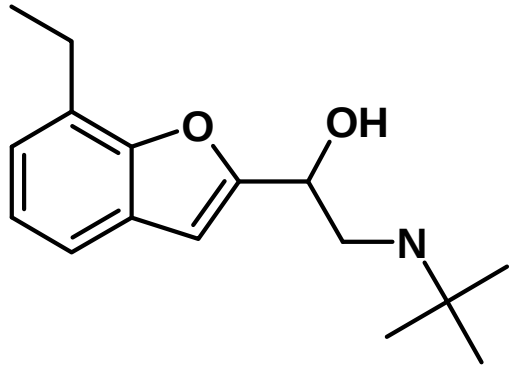


# Different Selectivities

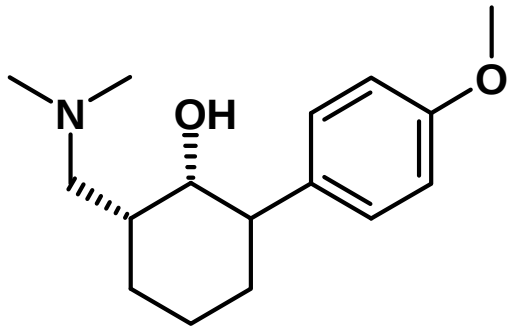
- **Selectivity between different cytochrome P450 isoenzymes**
  - ***in particular 3A4, 2C9, 2C19, 2D6, 1A2***
- **Selectivity between different reaction types**
  - ***chemoselectivity***
- **Selectivity between different reaction sites**
  - ***regioselectivity***

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

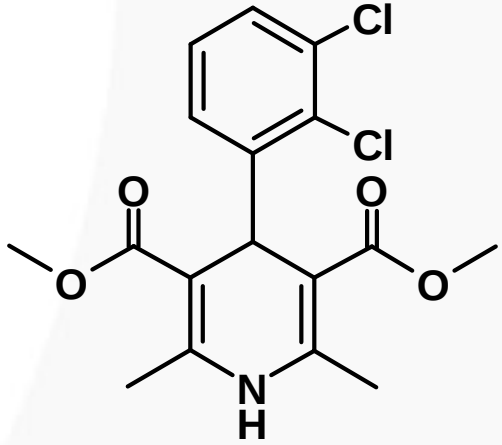
- **Training set: 146 drugs, substrate for 3A4, 2D6 or 2C9* major isoform specified**
- **Validation set: 233 reactants from the Metabolite database**




Bufuralol




Tramadol




Felodipine

*Manga, N. et al. *SAR and QSAR in Env. Res.* 2005, *16*, 43-61.

# 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[Å), n$_{acid_groups}$, n$_{aliphatic_amino}$, n$_{basic_n}$, r$_3$*

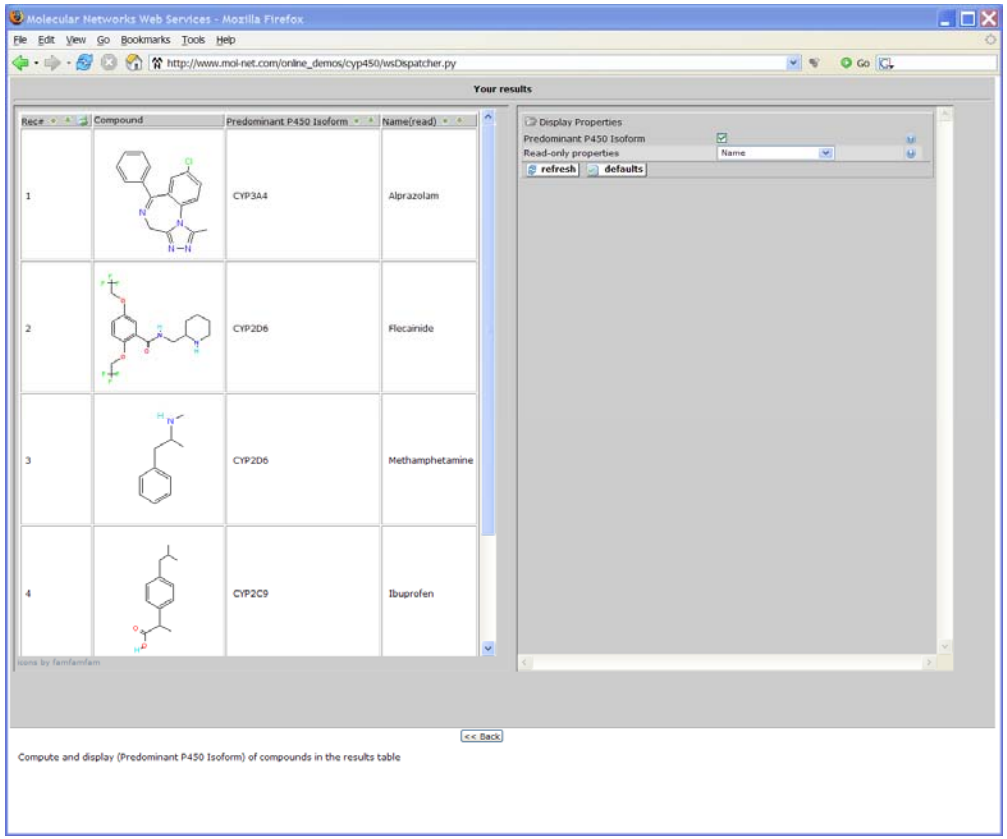
| | **Predictability** |
|---|---|
| **Training:** | **90.4%** |
| **5-fold CV:** | **87.8%** |

# 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




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

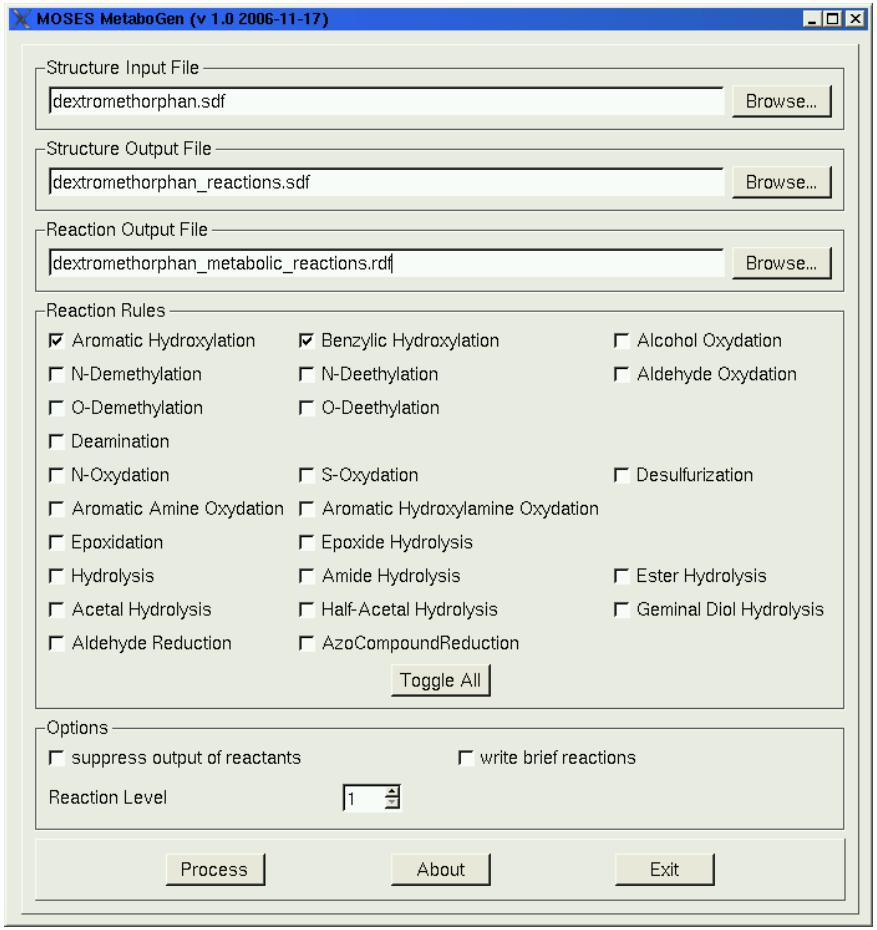
- **http://www.molecular-networks.com/online_services**

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

# 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***

# Reaction Generation with MOSES.Metabolism



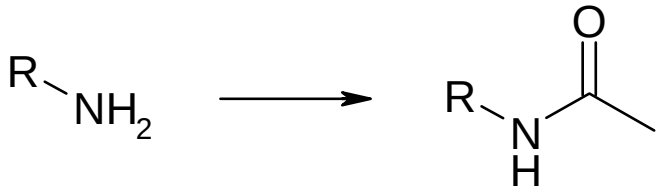

- MetaboGen
- Covers all relevant phase I reactions
- Supports atom-atom mapping
- Provides information on reaction centers
- Is used to generate reactions for XENIA

# 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 likeliness of a reaction based on literature data**

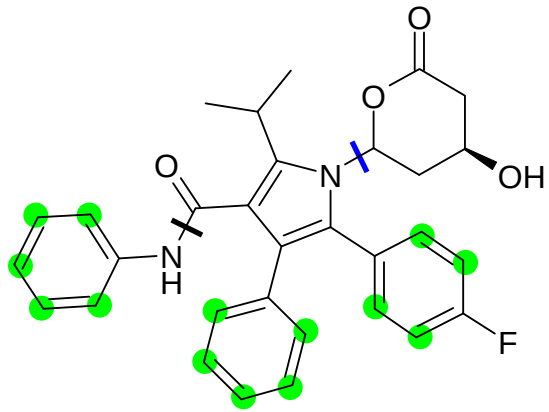
# Derivation of a Rule Base for Metabolite Prediction

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



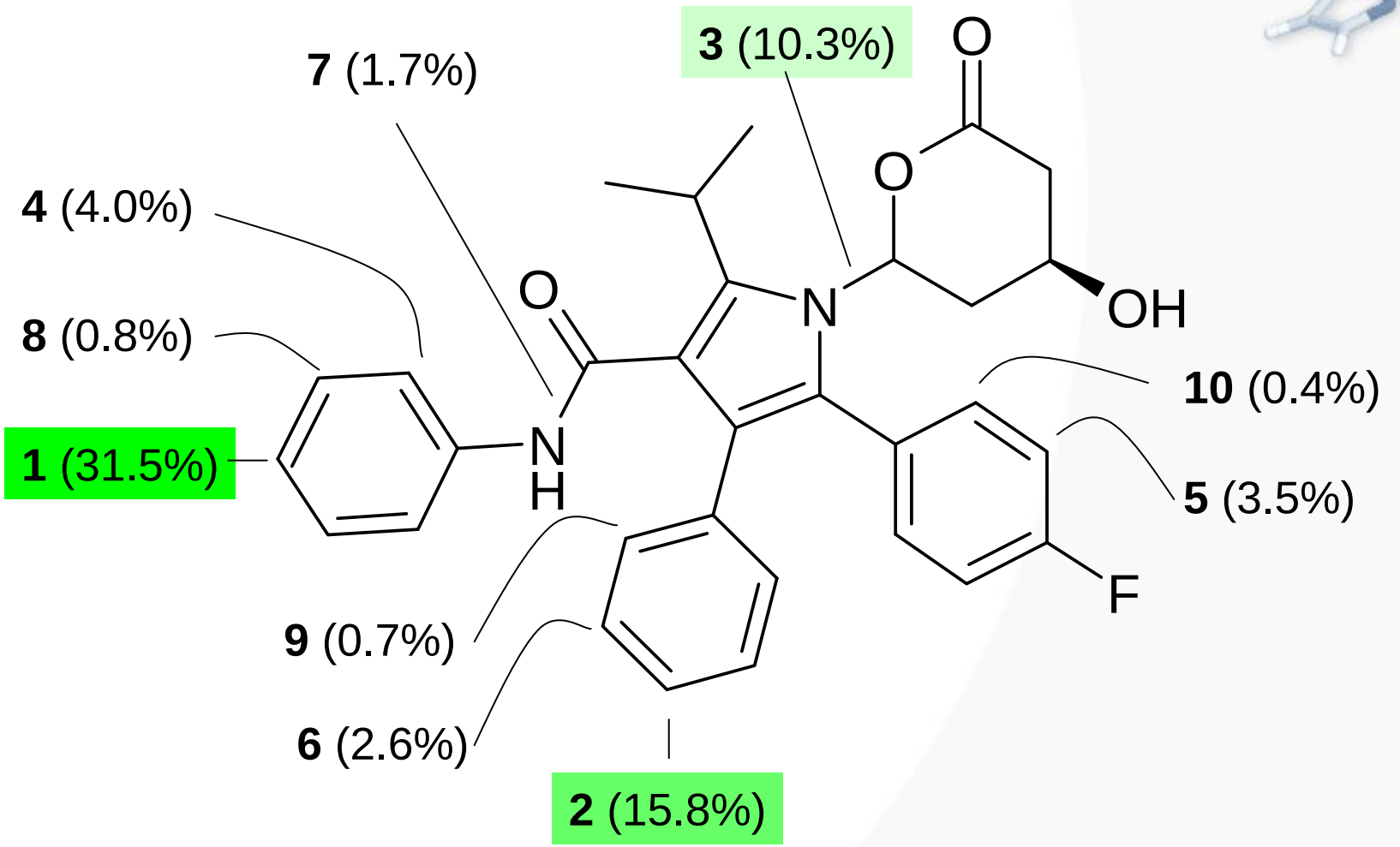

- **Calculate reaction probabilities based on a reaction database (Metabolite, MDL-Symyx)**
  - ***Conceivable metabolites*** ***1223***
  - ***Observed metabolites*** ***122***
  - ***Non-observed metabolites*** ***1101***
  - ***Probability*** ***122/1223 = 0.10***

# Phase I Metabolism of Atorvastatin Lactone

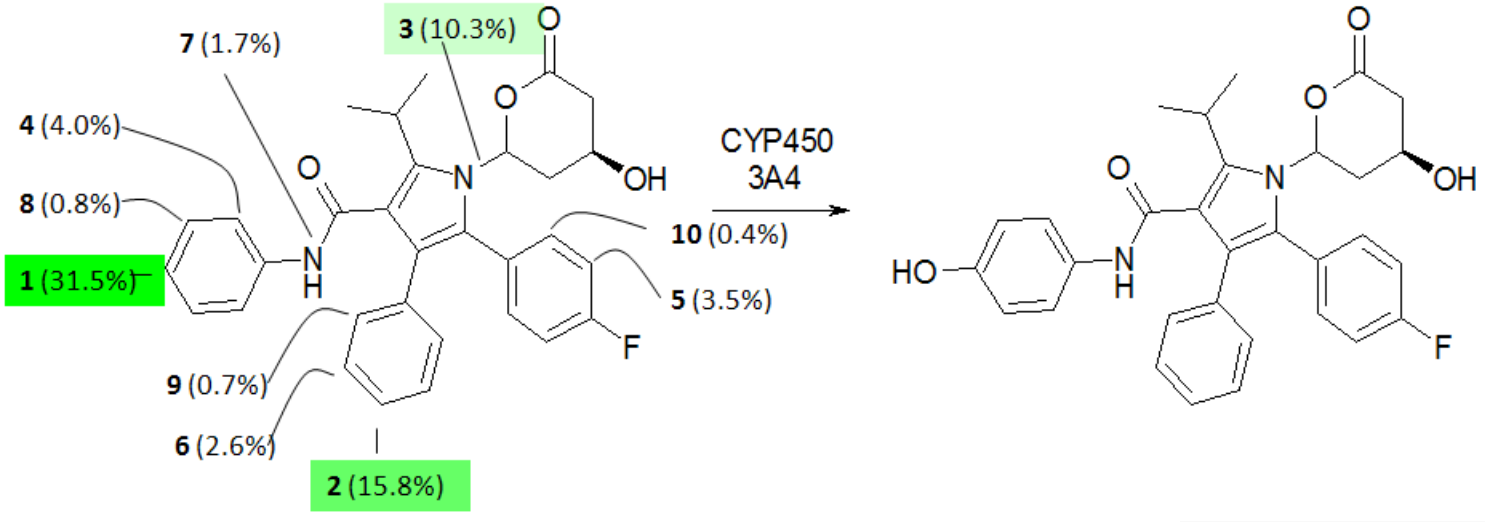



- **Chemoselectivity**
  - ***Aromatic hydroxylation (●)***
  - ***Amide hydrolysis (\)***
  - ***N-Dealkylation ( \ )***

# Ranks and Probabilities of predicted Metabolites of Atorvastatin Lactone



# Experimentally observed Metabolite of Atorvastatin Lactone




- **Metabolite predicted for atorvastatin with highest rank corresponds with experimental observations.**

# MOSES.Metabolism - Summary

- **MOSES.Metabolism is a rule-based system for metabolism prediction**
- **A database of metabolic reactions can be mined for the prediction of metabolism**
- **The rule base is transparent to the user**
- **Metabolism prediction is presently validated**

MOSES.RiskAssessment | Structures - Mozilla Firefox

File Edit View History Delicious Bookmarks Tools Help

MOSES.RiskAssessment | Str...

moses RiskAssessment

New Search | Databases | About

## Enter Query Structure

**Please select a method how to find a chemical record from the database**

By Structure

Enter SMILES string in the text field or sketch the query compound by clicking *Sketch Molecule*

C1=CC(O)=CC(N)=C1 Sketch Molecule

Please select a structure search method.

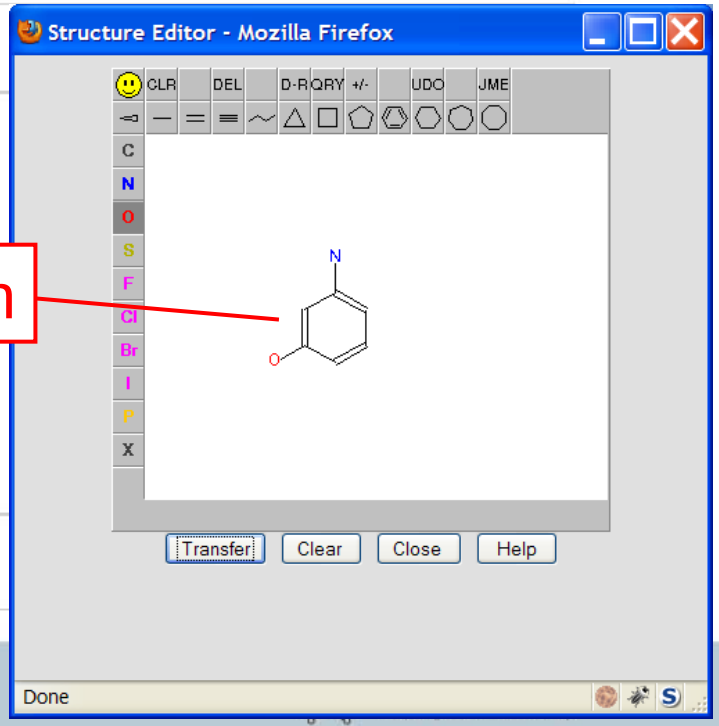
- Exact Structure
- Substructure
- Similar Substructure

CFSAN Sample Database search

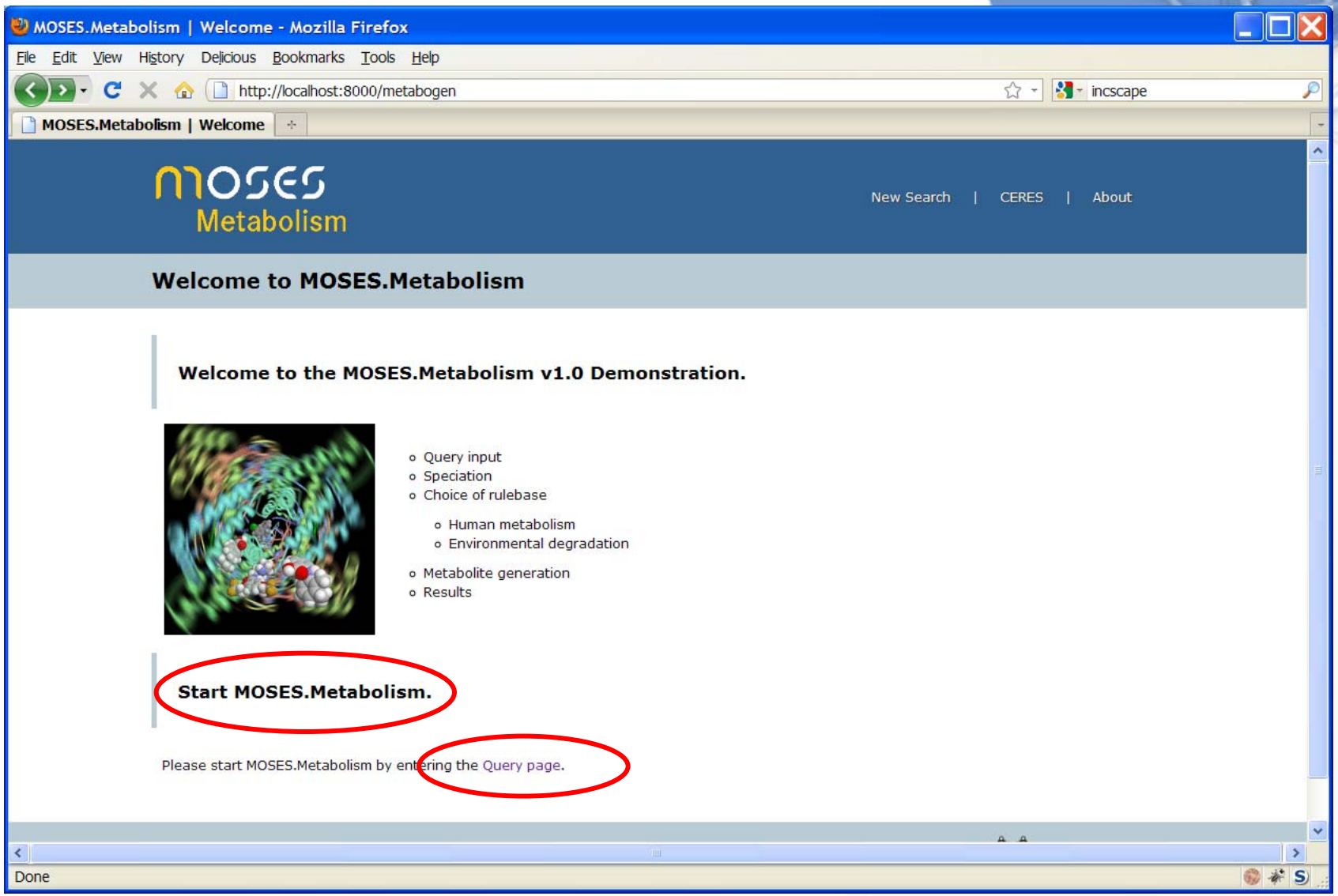
By Structure

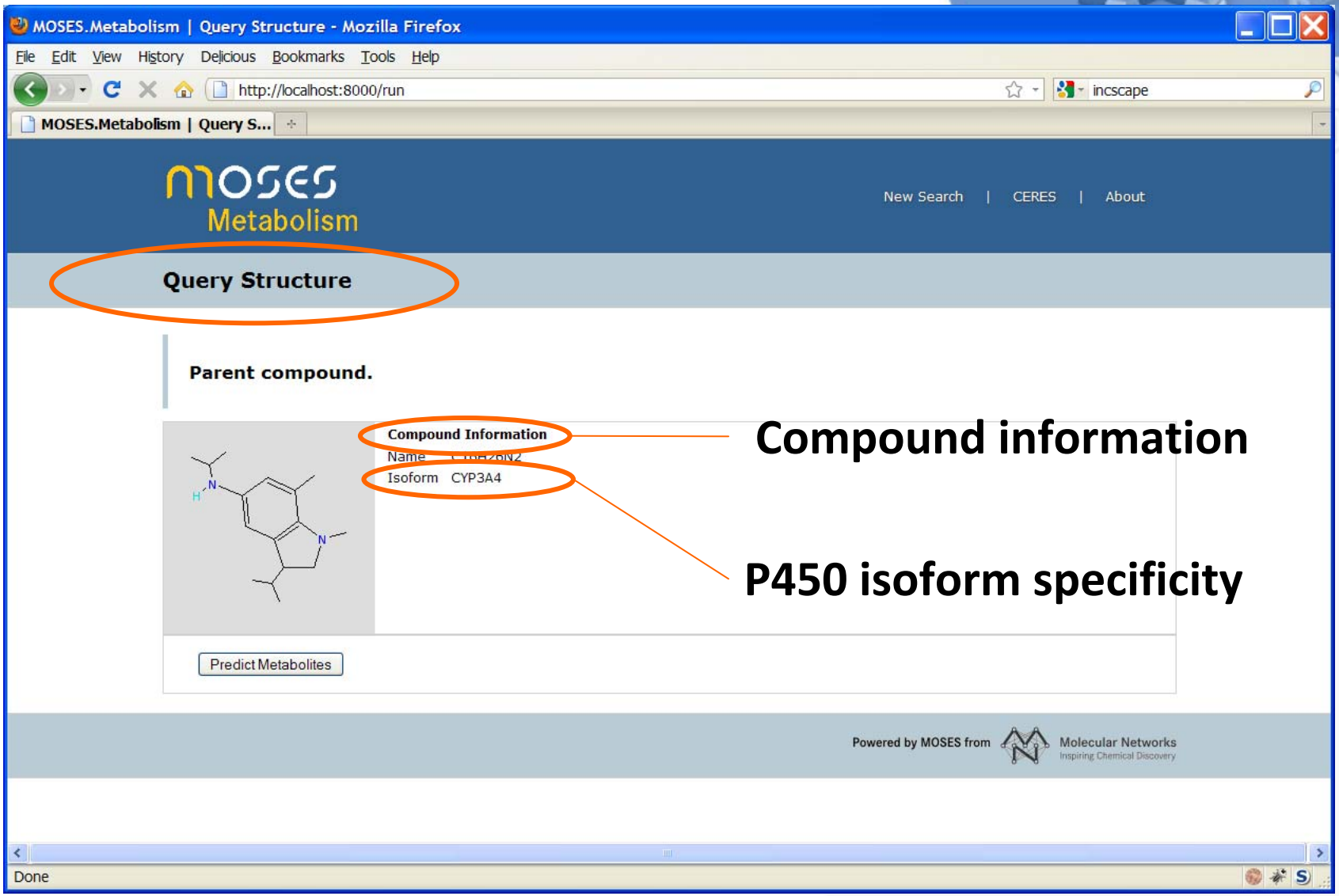
Sketch

Search method



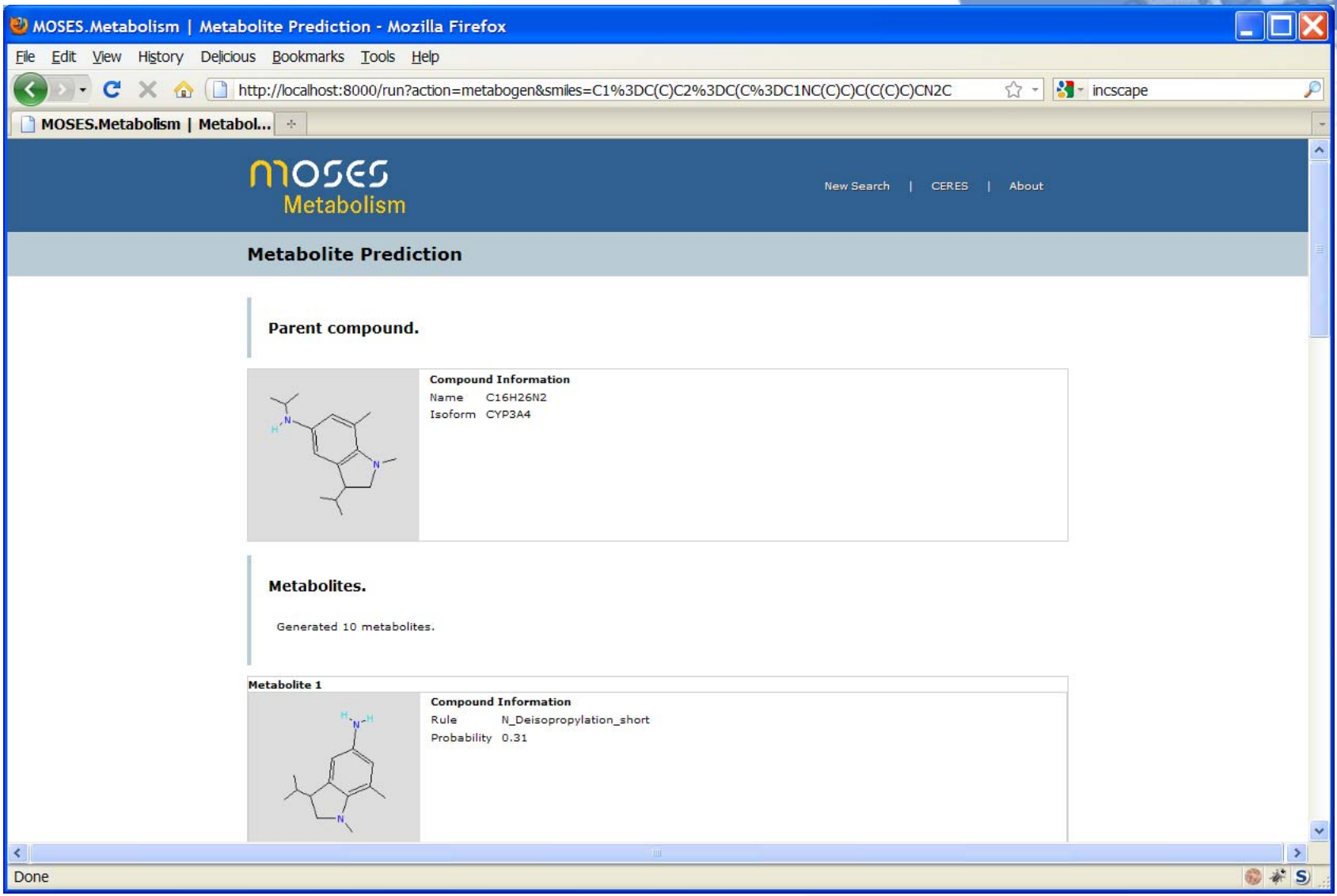

Done


MOSES.Metabolism | Welcome - Mozilla Firefox
File Edit View History Delicious Bookmarks Tools Help
http://localhost:8000/metabogen
incscape
MOSES.Metabolism | Welcome
MOSES Metabolism
New Search | CERES | About
Welcome to MOSES.Metabolism
Welcome to the MOSES.Metabolism v1.0 Demonstration.
Query input
Speciation
Choice of rulebase
Human metabolism
Environmental degradation
Metabolite generation
Results
Start MOSES.Metabolism.
Please start MOSES.Metabolism by entering the Query page.
Done




Compound information

P450 isoform specificity


MOSES.Metabolism | Metabolite Prediction - Mozilla Firefox
File Edit View History Delicious Bookmarks Tools Help
http://localhost:8000/run?action=metabogen&smiles=C1%3DC(C)C2%3DC(C%3DC1NC(C)C)C(C)(C)CN2C
incscape
MOSES.Metabolism | Metabol...
MOSES Metabolism
New Search | CERES | About
Metabolite Prediction
Parent compound.
Compound Information
Name C16H26N2
Isoform CYP3A4
Metabolites.
Generated 10 metabolites.
Metabolite 1
Compound Information
Rule N_Deisopropylation_short
Probability 0.31
Done

# Facts

- **C++ based Chemoinformatics system**
  - *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 applications*
- **under active development since 2001**
  - *Computer-Chemie-Centrum, Universität Erlangen-Nürnberg*
  - *Molecular Networks GmbH*
  - *15 person-years of development*
- **250,000 lines of Code**
  - *well documented and tested*

# Features

*„Integration of chemoinformatics into a single platform“*

- **chemical structure / reaction I/O**
  - ***support for commonly used file formats***
- **substructure / similarity search**
- **2D / 3D display of chemical structures and reaction**
  - ***stand alone GUI (Qt based)***
  - ***web GUI***
- **property calculation**
  - ***278 properties calculated by 80 calculation modules***
- **metabolite generation**
  - ***123 rules relevant to human metabolism***

# Molecular Networks
## http://www.molecular-networks.com

- **ADRIANA.*Code***
  - ***calculation of molecular descriptors on a sound geometric and physicochemical basis***
- **THERESA**
  - ***Design of organic and bioorganic syntheses***
- **SYLVIA**
  - ***estimation of the synthetic accessibility, since 2007***
- **BioPath**
  - ***Database of biochemical pathways in endogenous metabolism***

# MOSES Software Products

- **MOSES.Risk Assessment**
  - ***risk assessment based on TTC alerts and QSAR predictions***
- **MOSES.Metabolism**
  - ***generation of metabolites relevant to human metabolism***
- **MOSES.Editor / MOSES.Browser**
  - ***editing and browsing of chemical structures and reactions***

# Summary

- **Simple models of physicochemical effects can be used for reactivity prediction**
- **Organic syntheses can be designed with THERESA**
- **BioPath is a rich database with many applications**
- **Computational methods are valuable in risk assessment**
- **Metabolism of chemical compounds has to be considered in risk assessment**
- **Models for predicting metabolism of xenobiotics have been developed**

J.Gasteiger, Modeling Chemical Reactions for Drug Design
*J. Comput. Aided Mol. Des*., **2007**, *21*, 33-52

# Acknowledgements


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

**Collaboration on risk assessment**

- Dr. Chihae Yang


http://www.molecular-networks.com