

CHEMICAL CLASSIFICATION:
CLASSYFIRE'S APPLICATIONS IN
ENVIRONMENTAL HEALTH AND SAFETY

Yannick Djoumbou Feunang
University of Alberta, Canada

08/24/2016
ACS Meeting, Philadelphia (PA)

DATA AND MORE DATA

- Over the years, regulations, standards, and other systems (incl. MSDSs, L-CSSs) have been developed to educate and protect people at risk
- Large amount of data being stored/displayed in repositories, Books (incl. PubChem, T3DB, Bretherick's), Electronic Laboratory Notebooks

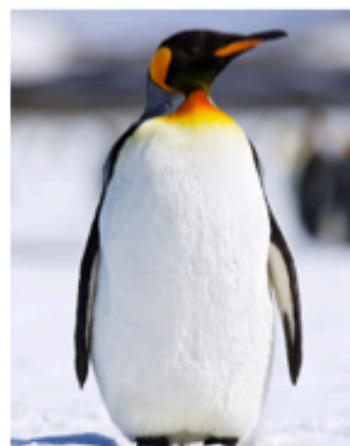
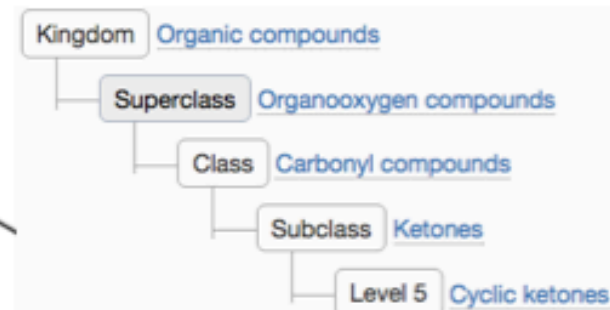
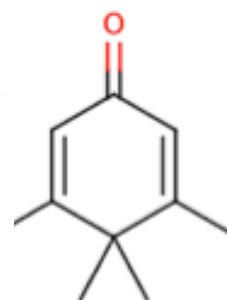
=> More known knows

- Great divide between # of available entities and # annotated entities
 - How can I synthesize my next blockbuster drug without burning the building down ?
 - What would compound X be transformed into, when interacting with Compound Y or Protein P.

=> More known unknowns

HOW CAN I RAPIDLY IDENTIFY SIGNIFICANT HAZARD RISKS?

ORGANIZING THE DATA



Kingdom: Animalia
Phylum: Chordata
Class: Aves
Order: Sphenisciformes
Family: Spheniscidae
Genus: Aptenodytes
Species: *A. patagonicus*



Kingdom: Animalia
Phylum: Chordata
Class: Mammalia
Order: Artiodactyla
Family: Giraffidae
Genus: Giraffa
Species: *G. camelopardalis*

ChemOnt (ClassyFire) and The Linnean Taxonomy

CLASSYFIRE & CHEMONT

The screenshot displays the ClassyFire web interface. On the left, the 'Ontology Tree Editor' shows a hierarchical tree of chemical classes, with 'Cytochalasins' selected. Below it, the 'Extended Info' panel provides statistics for the selected class. The main area features a 'Text Editor' with a 'Definition' field containing text about fungal metabolites. The 'ClassyFire' header includes navigation links like 'Browse', 'Classify', and 'Advanced Search'. The 'Input' section contains a text area with two entries: 'Trimethylsilyl chloride' with InChI string 'InChI=1S/C3H9ClSi/c1-5(2,3)4/h1-3H3' and '2-(chloromethyl)oxirane' with SMILES string 'C1CC1CO1'. A label 'InChI or SMILES' with arrows points to these strings. Below the input is a 'Label' field containing 'Dangerous combination' and a 'Submit' button.

Ontology Tree Editor

- Azahomoaporphines
- Benzophenanthridine alkaloids
- Betalains
- Camptothecins
- Cephalotaxus alkaloids
- Cinchona alkaloids
- Colchicines
- Condylocarpan alkaloids
- Correantane alkaloids
- Corynanthean-type alkaloids
- Cularin alkaloids and derivatives
- Curan alkaloids
- Cytochalasins**
 - Alachalasin
 - Aspochalasin
 - Chaetoglobosin
 - Cytochalasin
 - Pyrichalasin
- Daphniphylline-type alkaloids
- Eburnan-type alkaloids
- Ellipticine alkaloids
- Emetine alkaloids
- Epibatidine analogues
- Ergoline and derivatives
- Ervatamia alkaloids
- Erythrina alkaloids
- Gelsemium alkaloids
- Harmala alkaloids
- Hasubanan alkaloids

Extended Info

Total terms = 4823
Chemical entities (CHEMONTID:9999999) has 4822 de...
terms with 0 parents: 1 (< 1%)
terms with 1 parent: 4822 (99%)
100% of terms have definitions (4823 of 4823)
Reasoner: OFF
Memory: 40% of memory free

Text Editor

ID: CHEMONTID:9999999
Namespace: chemontid
Name: Cytochalasins

Definition

Fungal metabolites s...
presence of an isoinr...
ring, which can eithe...
carbonate, as in cyto...
cytochalasin D, H, a...

cytochalasan alkalo...
Scope: *Related Syn*
Type: *ChEBI TERM*

Cytochalasins
Scope: *Related Syn*
Type: *LIPIDMAPS T...*

Alkaloids
Scope: *Broad Syno...*

classyfire.wishartlab.com/queries/new#text-query

ClassyFire Browse - Classify About ClassyFire Contact Advanced Search Downloads Help -

Chemical
Protein/DNA/RNA
IUPAC Name

Chemical Input Draw Structure Upload A SDF/TSV File

Example - STRUCTURE

Input

Trimethylsilyl chloride InChI=1S/C3H9ClSi/c1-5(2,3)4/h1-3H3
2-(chloromethyl)oxirane C1CC1CO1

InChI or SMILES

Provide one entry per line containing a SMILES or an InChI string, optionally preceded by an identifier. The line must be tab-separated.

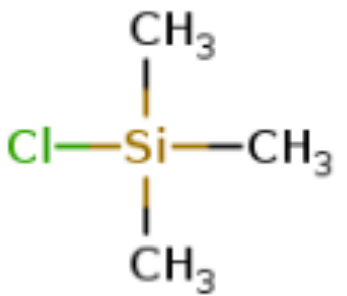
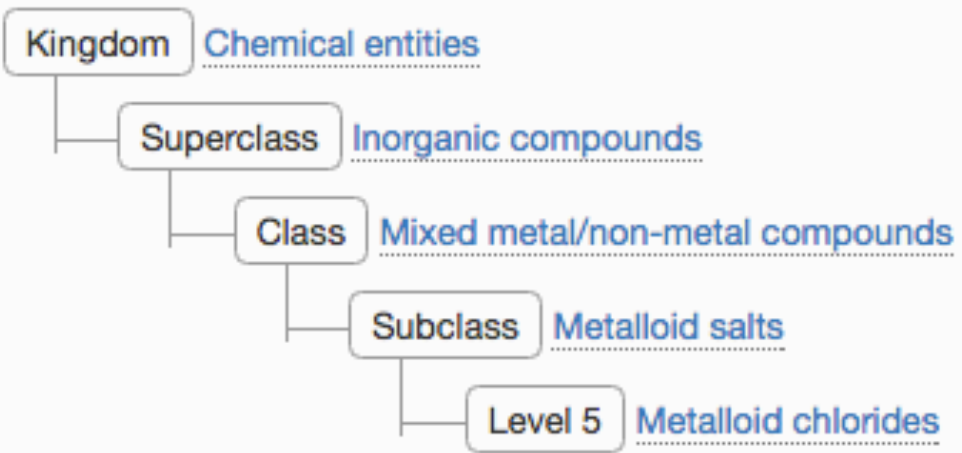
Label

Dangerous combination

Provide a name for the data sample (optional). You can provide multiple tags separated by '|'.
Submit

Taxonomic Classification

Taxonomy Tree



This approach was applied to classify the PubChem database, and a number of other repositories.

Alternative Parents

[Trialkylchlorosilanes](#) [Silyl monohalides](#) [Organochlorosilanes](#) [Alkylhalosilanes](#) [Hydrocarbon derivatives](#)

Molecular Framework

Not available

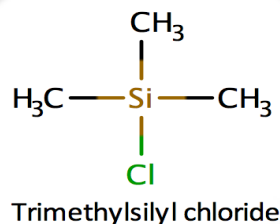
Substituents

Metalloid chloride - Trialkylchlorosilane - Trialkylheterosilane - Silyl monohalide - Organoheterosilane - Organochlorosilane - Alkylhalosilane - Hydrocarbon derivative - Organosilicon compound - Organic metalloid moiety

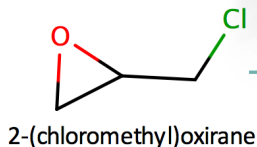
Description

This compound belongs to the class of chemical entities known as metalloid chlorides. These are inorganic compounds in which the largest halogen atom is Chlorine, and the heaviest metal atom is a metalloid.

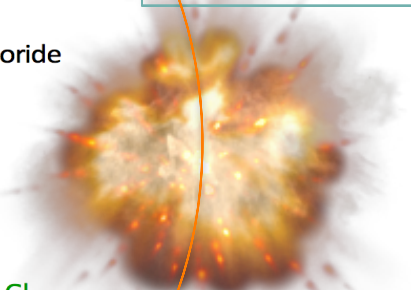
PREDICTING $SA_{ACTIVITY}R - SR_{EACTIVITY}R - SH_{AZARD}R$



Trialkylchlorosilanes
(CHEMONTID:0004480)



Epoxides
(CHEMONTID:0000159)



Prediction is often a bit more complex than this

- 70 classes from the Enhanced NOAA Worksheet and the Bretherick's Handbook were mapped to ClassyFire

Epoxides mixed with Chlorosilanes

CAMEO Chemicals

Hazard Predictions

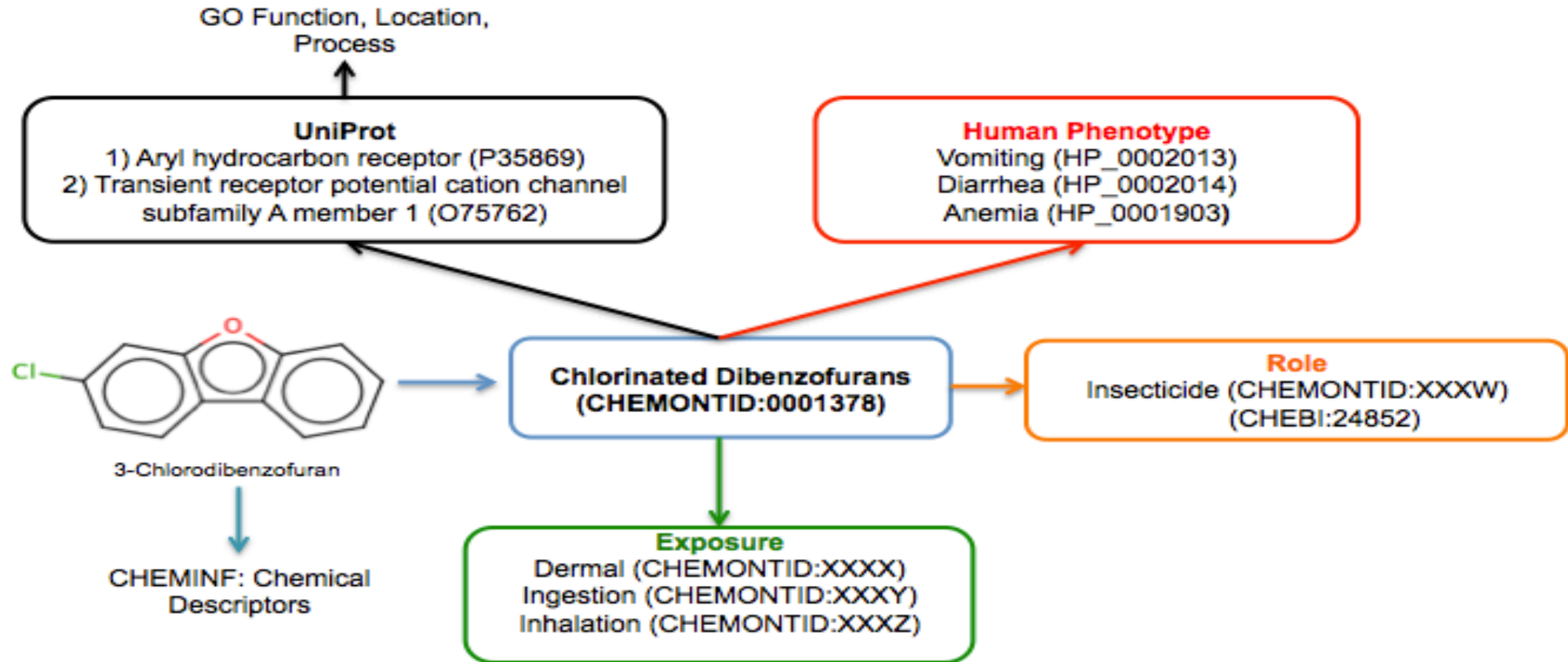
- Flammable:** Reaction products may be flammable
- Generates gas:** Reaction liberates gaseous products and may cause pressurization
- Generates heat:** Exothermic reaction at ambient temperatures (releases heat)
- Intense or explosive reaction:** Reaction may be particularly intense, violent, or explosive
- Toxic:** Reaction products may be toxic

Chlorosilanes may react with epoxides to liberate toxic halocarbon and HX gases (J. Org. Chem., 1998, 63 (8), pp 2428-2429).

Vinyl trichlorosilane is incompatible with alkylene oxides and epichlorohydrin (Pohanish, Richard P. (2004). HazMat Data - For First Response, Transportation, Storage, and Security (2nd Edition). John Wiley & Sons).

ClassyFire	CAMEO/Bretherick's
Organic peroxides	Peroxides, Organic
Carboxylic acids	Acids, Carboxylic
Epoxides	Carbamates
Trialkylchlorosilanes	Chlorosilanes
Metal p-nitrophenoxides	Metal nitrophenoxides

CHEMONT FOR DATA INFERENCE

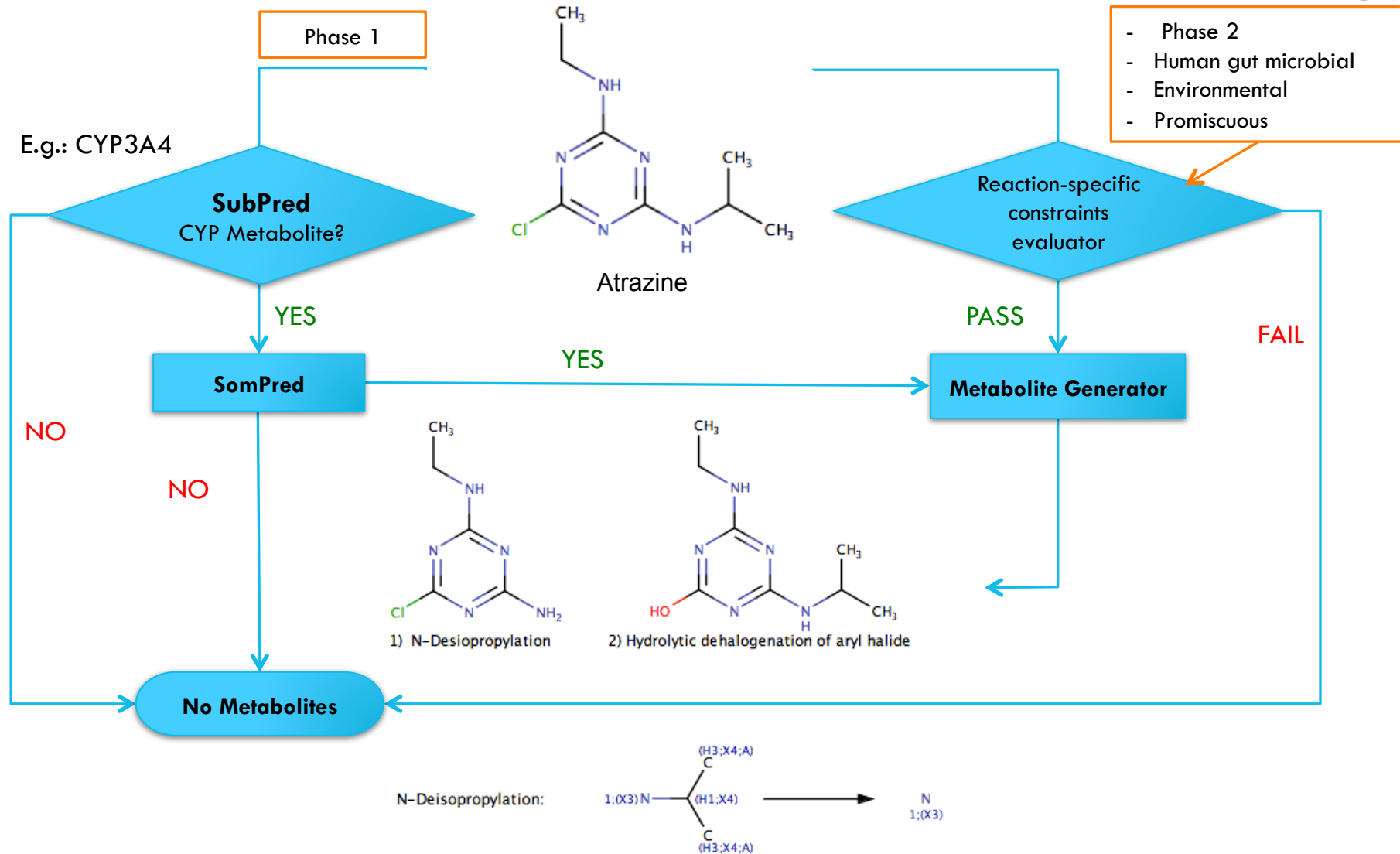


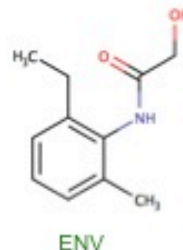
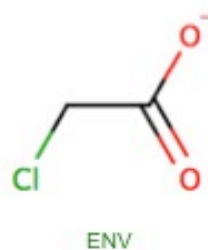
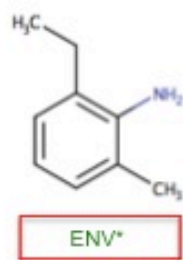
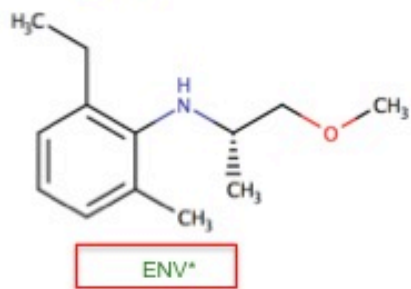
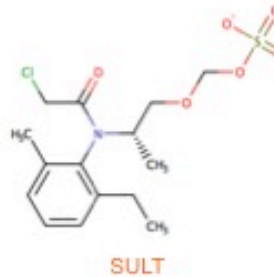
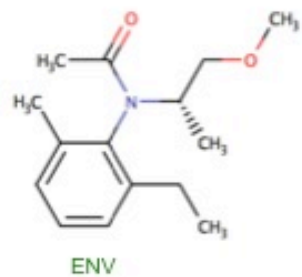
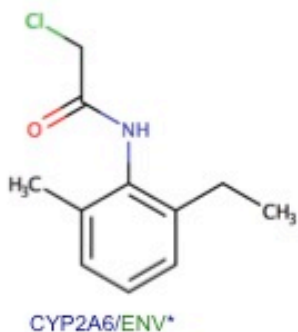
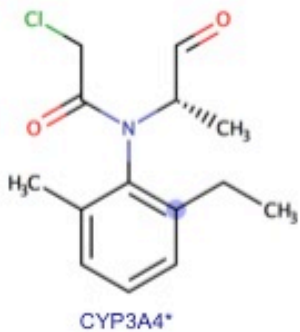
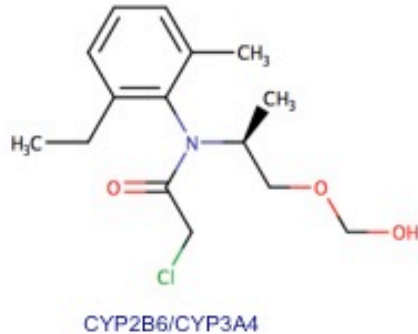
- Compounds **targeting AhR** that might **induce vomiting** upon **inhalation**

BIOTRANSFORMER



- Given a compound, predict its metabolic fate
- Accept SMILES, INCHIs, MOL, or SDF format
- Combine Machine Learning & Rule-based approaches
- Make use of ClassyFire fingerprints to predict metabolism





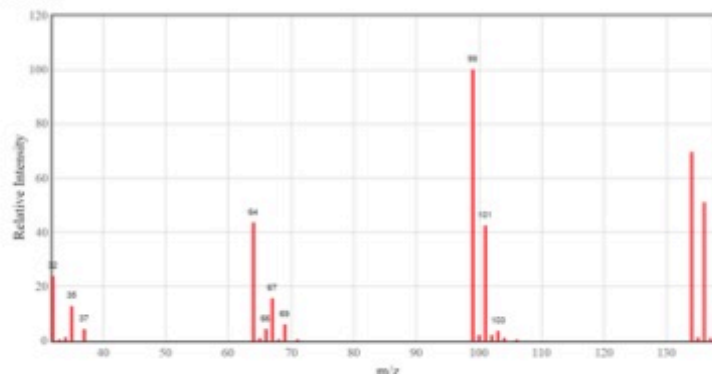
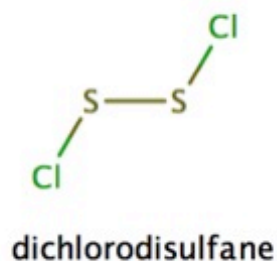
- Identification of putative metabolites upon absorption of chemicals by humans (e.g. in the lab)
- Chemicals absorbed /produced by humans are deposited into the environment, and possibly transformed by plants, other (micro-)organisms
- BioTransformer could be helpful in the assessment environmental toxicity

Examples of predicted Metolachlor metabolites

CFMID: MS SPECTRA PREDICTION

- Competitive Fragmentation Modeling (CFM):

- In-silico mass spectrum prediction



- ClassyFire: Grouped test compounds by chemical class
- Looked for systematic errors in CFM output

WeightedR,	WeightedP,			
100.00000,	100.00000,	Halogen organides,	1,	[NIST2011_1512]
94.99930,	98.77140,	Isocyanates,	1,	[NIST2011_21462]
99.77530,	88.36015,	Propargyl-type 1,3-dipolar organic compounds,	2,	[NIST2011_103606,NIST2011_6342]
98.75780,	78.89390,	Organic oxides,	1,	[NIST2011_20873]
99.36208,	86.06613,	Thiocarboxylic acids and derivatives,	11,	[NIST2011_10834,NIST2011_11702,NIST2011_228
87.85070,	99.34810,	Trithianes,	1,	[NIST2011_110785]
97.95420,	78.24545,	Dihydrothiophenes,	2,	[NIST2011_17386,NIST2011_30186]
96.03668,	81.92703,	Acyclic alkanes,	38,	[NIST2011_22415,NIST2011_22420,NIST2011_22538,NIST2011_225
99.54790,	96.39170,	Imidothioic acids and derivatives,	1,	[NIST2011_22908]
97.02516,	85.31640,	Cycloalkenes,	29,	[NIST2011_105571,NIST2011_122015,NIST2011_134728,NIST2011_17471,NIS
97.57768,	86.49050,	Dialkylamines,	24,	[NIST2011_100340,NIST2011_1063,NIST2011_115238,NIST2011_125850,NIS

- e.g. Halogenated compounds were initially poorly predicted due to lack of isotope modeling.....so fixed it!

- Spectra-based search can enhance compound identification
- Only. ~20,000 experimentally determined MS spectra
- Prediction of MS spectra can be improved by chemical classification
- Newly synthesized compounds can be searched within a reference spectra database
- This could assist in hazard assessment

THANKS

David Wishart, UofA

Russ Greiner, UofA

Craig Knox, UofA

Mike Wilson, UofA

Zheng Shi, UofA

Evan Bolton, NIH

Paul Thiessen, NIH

Christoph Steinbeck, EBI

Gareth Owen, EBI

Venkatesh Muthukrishnan, EBI

Janna Hastings, EBI

Shankar Subramanian, UCSD

Eoin Fahy, UCSD

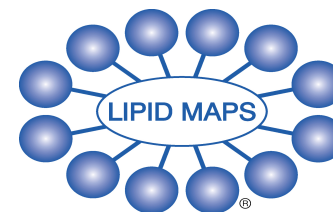
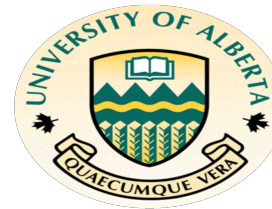
Leonid Chepelev, University of Ottawa/Ottawa Hospital

Stephen Boyer, IBM

Felicity Allen, EBI

Lutz Weber, OntoChem

Leah McEwen, Cornell



EXPOSE OR/AND BE EXPOSED

- Through our daily routines, we expose ourselves or the environment to chemicals (from lab experiments, dust, pollutants, cleaning products, etc.)
- Several types of hazards can be associated with these interactions
 - Explosions, skin corrosion, aquatic toxicity, etc.



CHEMOSUMMARIZER

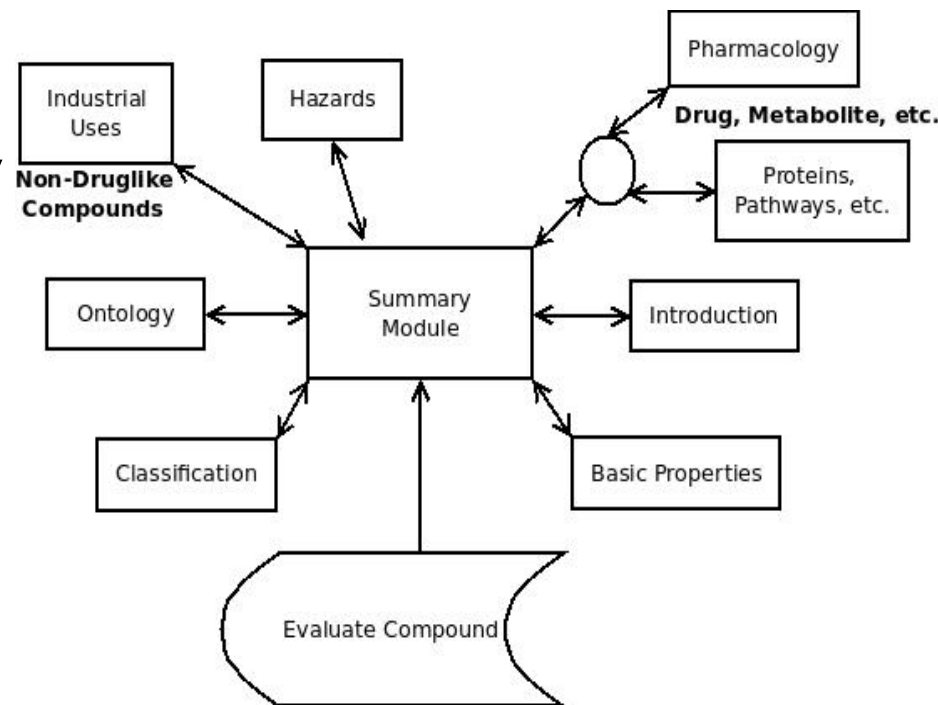
Goal: To provide detailed summaries of compounds

- Get properties info from DBs, pathways/ pharmacology from DBs (DrugBank, HMDB, ChEBI, etc.)

ClassyFire input serves to provide:

- Structural description
- Hazard information
- Class-generic metabolism biotransformation profile

Could we describe all PubChem compounds?



Courtesy of Zachary Budinski, Wishart Lab, 2016