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

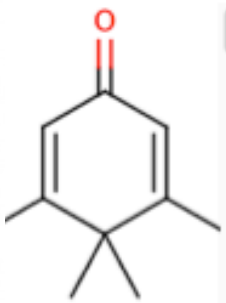
  => More known knows

- 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.

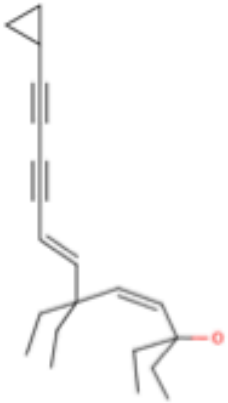
    => More known unknows

HOW CAN I RAPIDLY IDENTIFY SIGNIFICANT HAZARD RISKS?

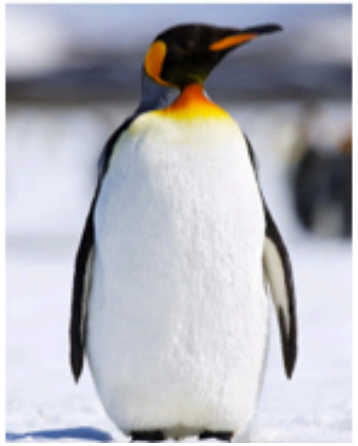
ORGANIZING THE DATA



Kingdom: Organic compounds
Superclass: Organooxygen compounds
Class: Carbonyl compounds
Subclass: Ketones
Level 5: Cyclic ketones



Kingdom: Organic compounds
Superclass: Organooxygen compounds
Class: Alcohols and polyols
Subclass: Tertiary alcohols

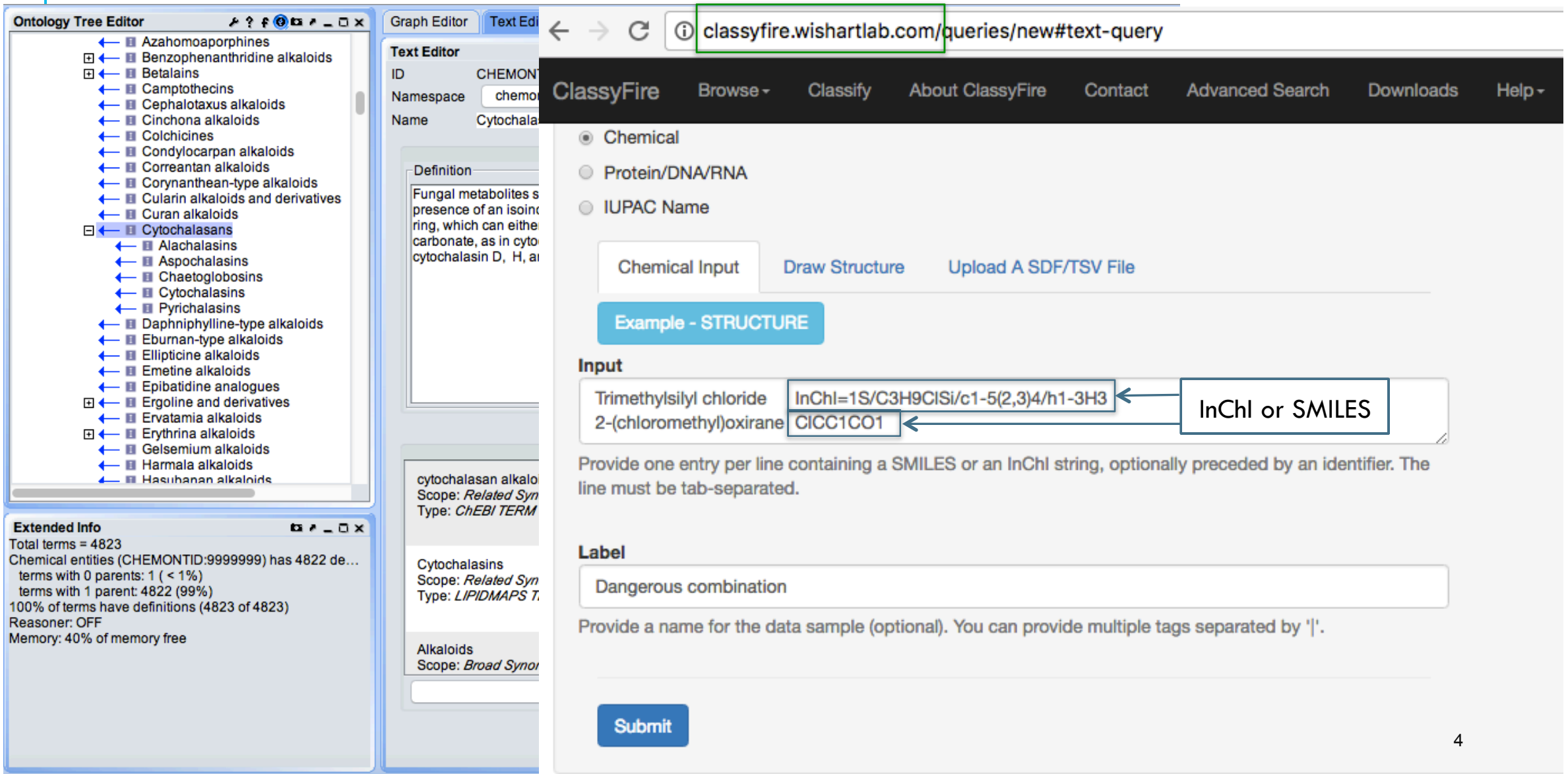


Kingdom: Animalia
Phylum: Chordata
Class: Aves
Order: Sphenisciformes
Family: Spheniscidae
Genus: Aptenodytes
Species: A. patagonicus

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

ChemOnt (ClassyFire) and The Linnean Taxonomy

CLASSYFIRE & CHEMONT

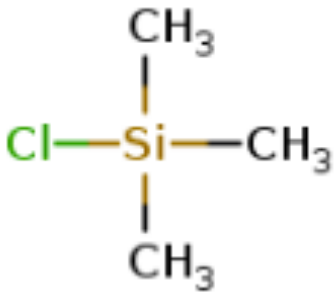


classyfire.wishartlab.com/entities/IJOOHPMOJXWVHK-UHFFFAOYSA-N ← InChIKey

Taxonomic Classification

Taxonomy Tree

- Kingdom: Chemical entities
  - Superclass: Inorganic compounds
    - Class: Mixed metal/non-metal compounds
      - Subclass: Metalloid salts
        - Level 5: Metalloid chlorides



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

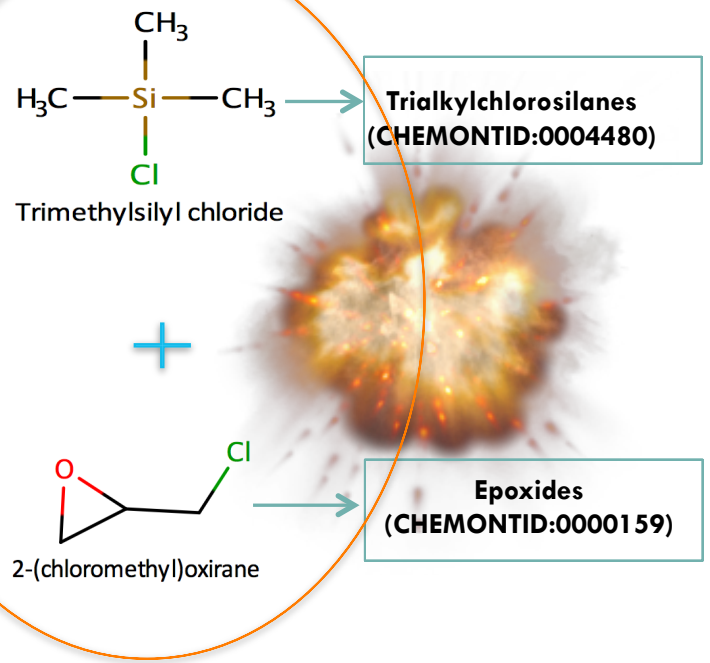
Substituents

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

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 SACTIVITYR – SREACTIVITYR - SHAZARDR



Prediction is often a bit more complex than this

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

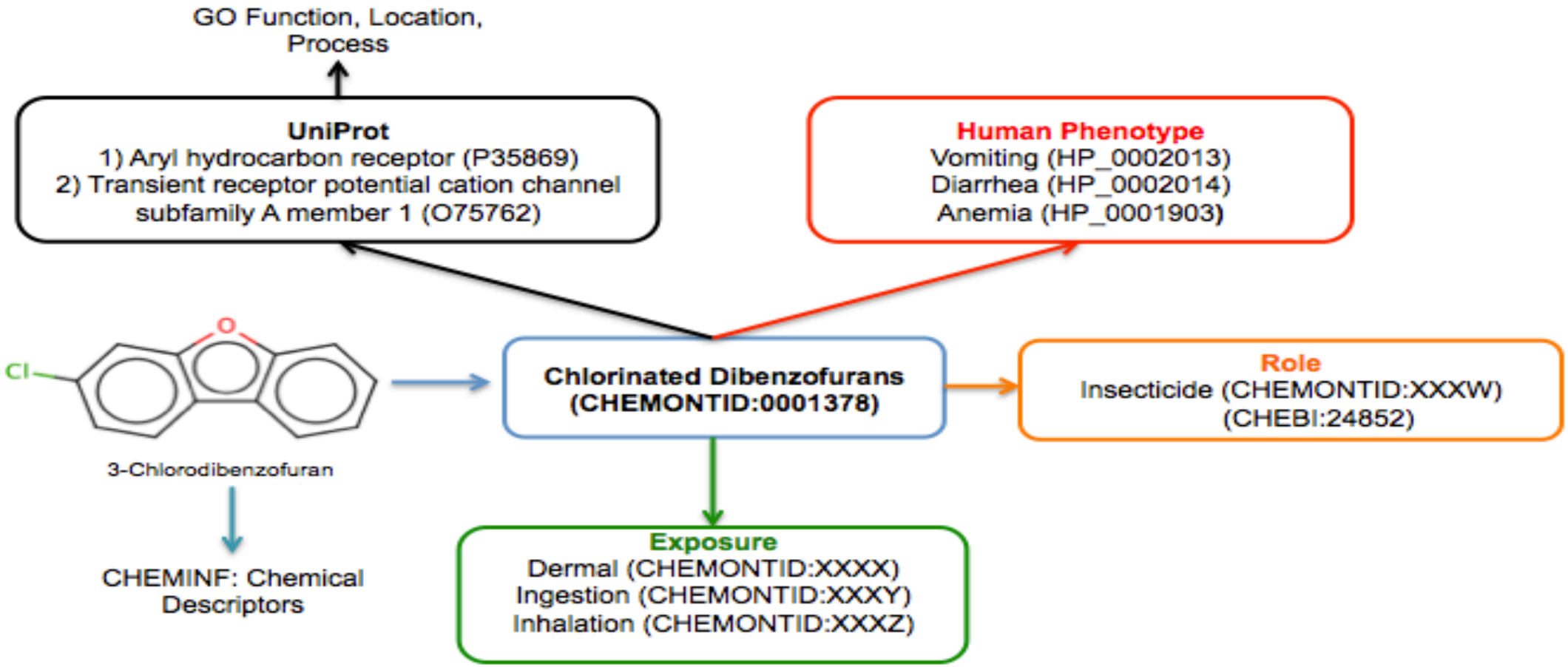
Chlorosilanes may react with epoxides to liberate toxic halocarbon and HX gases (J. Org. Chem., 1998, 63 (8), pp 2428-2429).

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).

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

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

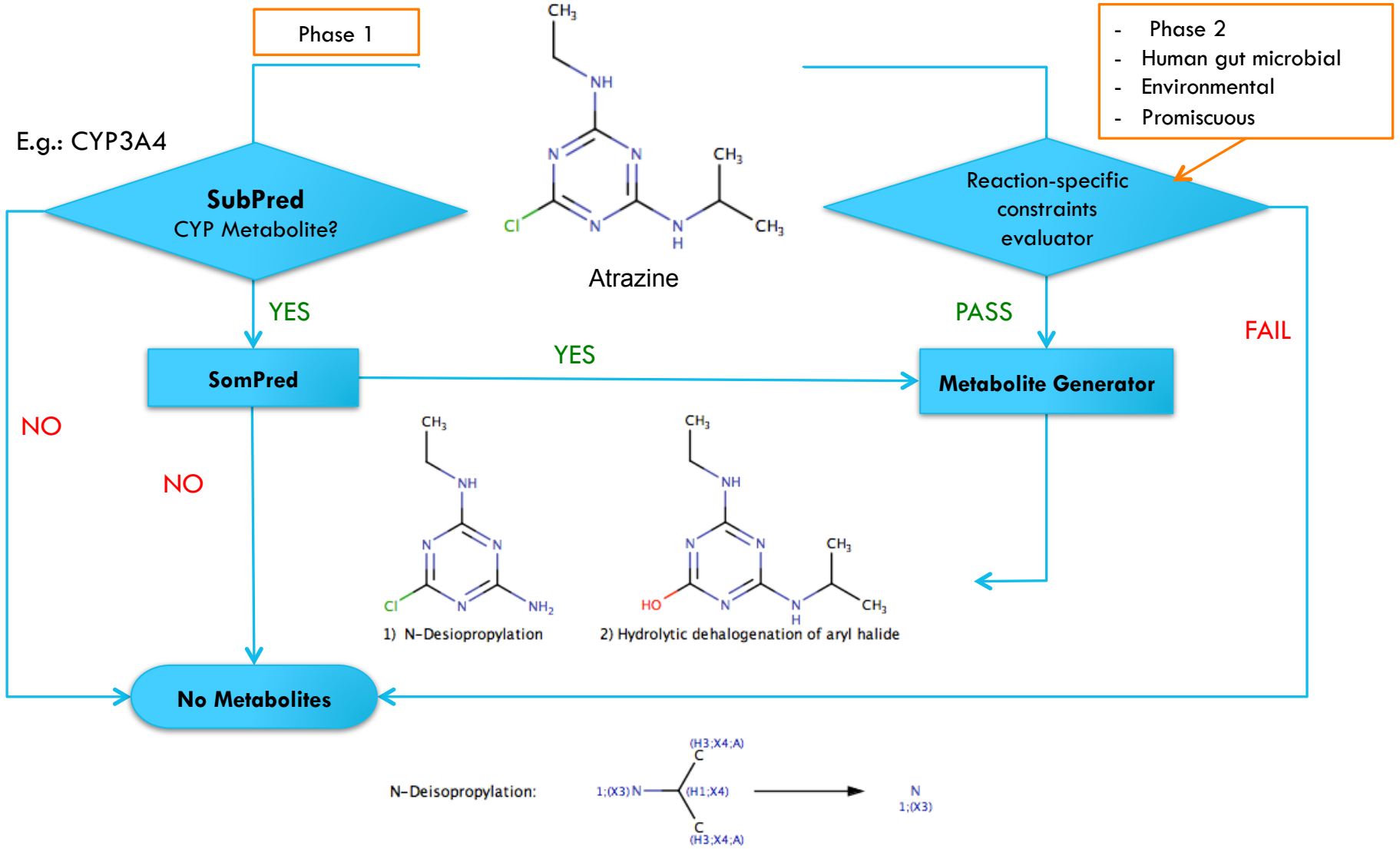
CHEMONT FOR DATA INFERENCE

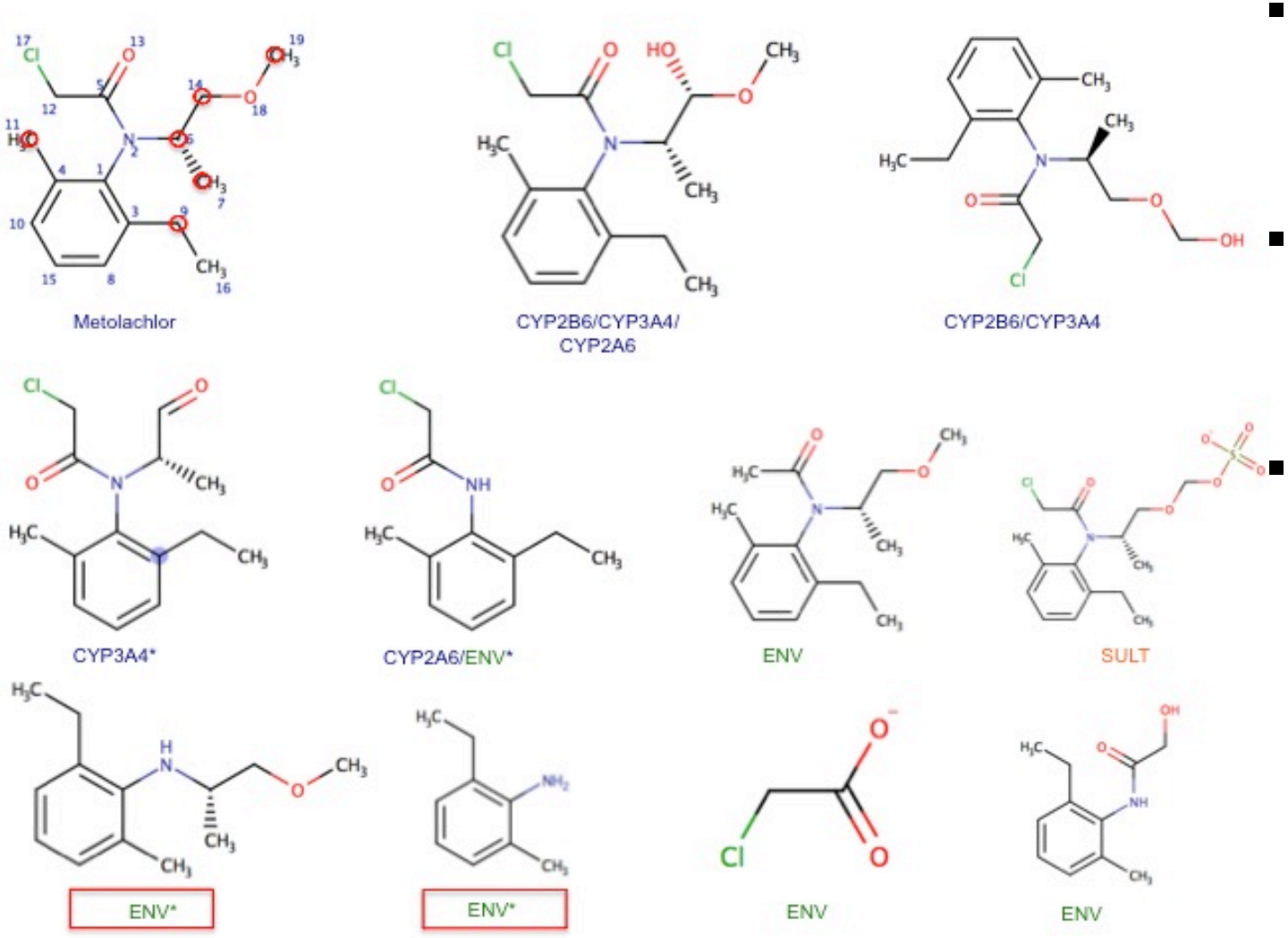


- Compounds targeting AhR that might induce vomiting upon inhalation

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



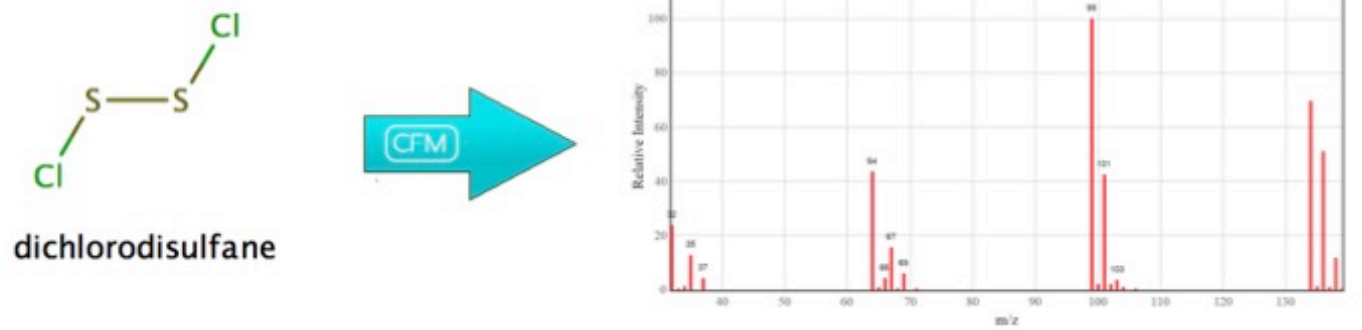


Examples of predicted Metolachlor metabolites

- 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

CFMID: MS SPECTRA PREDICTION

- Competitive Fragmentation Modeling (CFM):
  - In-silico mass spectrum prediction



dichlorodisulfane

  - 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

Courtesy of Felicity Allen *et al.*, Metabolomics, 11:1, pp. 98-110, 2015

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

CHEMOSUMMARIZER

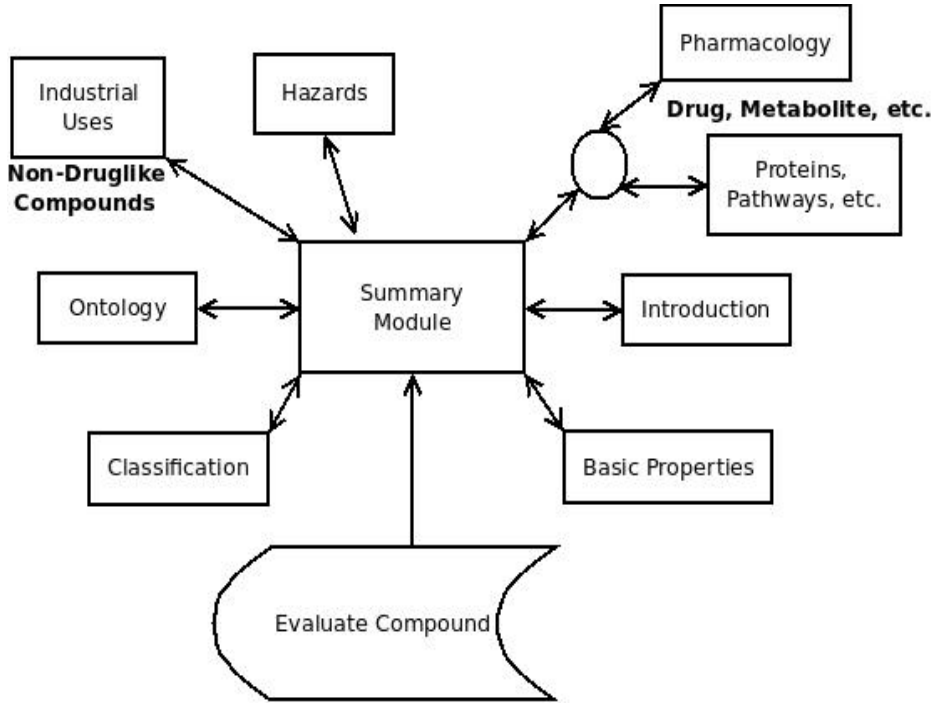
Goal: To provide detailed summaries of compounds

- 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