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

ChemOnt (ClassyFire) and The Linnean Taxonomy

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
CLASSYFIRE & CHEMONT

InChI or SMILES

Example - STRUCTURE

Input
Trimethylsilyl chloride

2-(chloromethyl)oxirane

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 \"\".
This approach was applied to classify the PubChem database, and a number of other repositories.
PREDICTING $SA_{CTIVITY} - SR_{EACTIVITY} - SH_{AZARD}$

- 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.

<table>
<thead>
<tr>
<th>ClassyFire</th>
<th>CAMEO/Bretherick’s</th>
</tr>
</thead>
<tbody>
<tr>
<td>Organic peroxides</td>
<td>Peroxides, Organic</td>
</tr>
<tr>
<td>Carboxylic acids</td>
<td>Acids, Carboxylic</td>
</tr>
<tr>
<td>Epoxides</td>
<td>Carbamates</td>
</tr>
<tr>
<td>Trialkylchlorosilanes</td>
<td>Chlorosilanes</td>
</tr>
<tr>
<td>Metal p-nitrophenoxides</td>
<td>Metal nitrophenoxides</td>
</tr>
</tbody>
</table>
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

E.g.: CYP3A4

SubPred
CYP Metabolite?

YES

SomPred

NO

No Metabolites

Metabolite Generator

PASS

FAIL

Reaction-specific constraints evaluator

- Phase 2
- Human gut microbial
- Environmental
- Promiscuous

Atrazine

1) N-Desisopropylation
2) Hydrolytic dehalogenation of aryl halide

N-Desisopropylation:
- 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

- 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 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