National Institute of Environmental Health Sciences
Division of Translational Toxicology

Structure-based Chemical Taxonomy to Focus Chemical Queries

Tripp LaPratt[1], Victoria Hull[1], Emily Reinke[1], Nicole Kleinstreuer[2]

[1]Inotiv, Research Triangle Park, NC; [2]NIH/NIEHS/DTT/NICEATM, Research Triangle Park, NC

Introduction

- Characterizing and navigating chemical space is an ongoing challenge, particularly with large queries that can be quite diverse and have a variety of chemical structures.
- Identifying substructures and chemical groups related to chemical activity to distinguish chemical space by chemical structure can focus these queries.
- ClassyFire (Djoumbou et al. 2016) is a tool that can fulfill this purpose.
  - ClassyFire is an automated, structure-based chemical taxonomy tool created by the Wishart Research Group with 4,285 unique classifications across 11 hierarchical levels.
- We conducted a case study using chemicals tested in the U.S. Environmental Protection Agency's Tox21 program to demonstrate how ClassyFire classifications can be used to focus and contextualize chemical queries.
- To explore toxicologically-relevant data, we used the Integrated Chemical Environment (ICE; https://ice.ntp.niehs.nih.gov/; Daniel et al. 2022). ICE is an open-access resource developed by NICEATM that contains highly curated data and computational tools that allow users to explore chemical space.
  - We focused queries of bioactivity and chemical use data in ICE using the Search and Chemical Characterization tools.
  - We demonstrate potential connections between chemical groups and specific use cases, modes of action, and mechanistic targets.

Computational Tools

ClassyFire

- The ClassyFire hierarchical tree has a depth of 11 levels with a total of 4,825 unique classifications.
  - Levels include, in order: Kingdom, Superclass, Class, Subclass, and Parent Level 1-7 (Fig 1A).
- The ClassyFire tool accepts inputs as SMILES, InChi, InChikey, IUPAC name, and FASTA
  - ClassyFire is accessible through a free web interface (Fig 1B), a batch search tool, or an API.

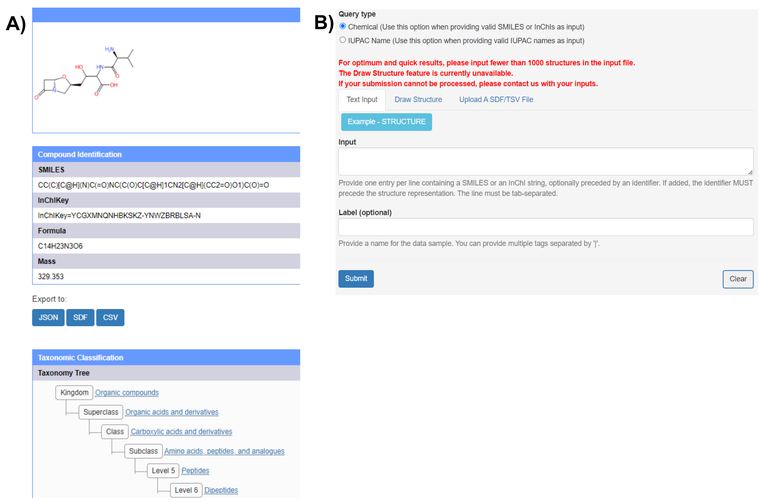


Figure 1: A) An example of a ClassyFire output for the chemical valclavam. B) ClassyFire user input web interface.

ICE

- ICE is an open-access resource developed by NICEATM. It contains toxicologically-relevant data and computational tools as shown in Figure 2 below.

Figure 2: ICE homepage

Visit ICE
https://ice.ntp.niehs.nih.gov/

Applying ClassyFire to Case Study Chemicals

- To demonstrate the use of ClassyFire in guiding and informing chemical queries in ICE, we used the ClassyFire tool to classify the 9329 chemicals in the federal Tox21 program.
- There were 8,583 Tox21 chemicals with ClassyFire classifications spanning 8 levels of hierarchy (8,477 organic; 106 inorganic).
  - "Benzenoids" and "organoheterocyclic" were the most abundant Superclasses with 2,744 and 1,959 chemicals, respectively.

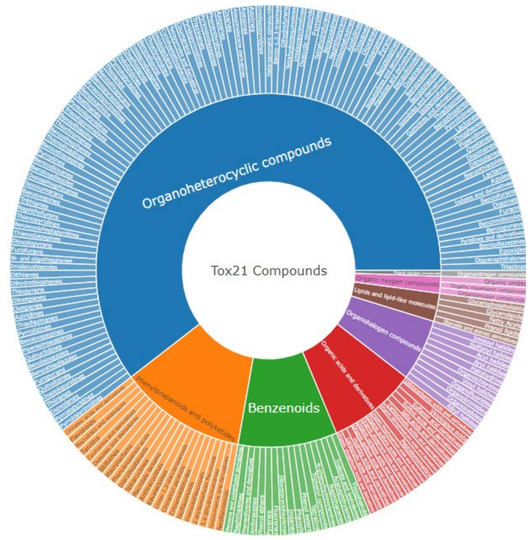


Figure 3: Sunburst plot for Tox21 Superclasses containing 100 or more chemicals.

Case Study Chemical Set: Bioactivity from ICE

- We ran the 8,583 Tox21 chemicals with ClassyFire classifications through the ICE Search tool to see available bioactivity data.
  - We focused the search on curated high-throughput screening (cHTS) data that have been annotated to "Cellular Stress Response" mechanistic targets.

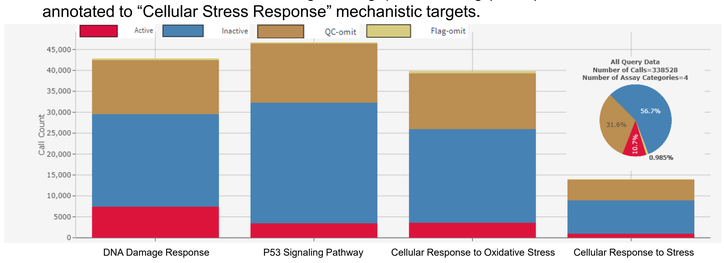


Figure 4: Query results for available ICE bioactivity data for Tox21 chemicals.

- 3,215 chemicals are bioactive for at least one assay mapped to "Cellular Stress Response."
  - The must abundant Superclass for these active chemicals is "Benzenoids" (1,232 chemicals).
    - The most active Class of Benzenoids was "Benzene and Substituted Derivatives" (918 chemicals).
      - The most abundant Subclass for "Benzene and Substituted Derivatives" was "Diphenylmethanes" (138 chemicals).
- Using the ICE filter chain, we examined AC50s for Diphenylmethanes.

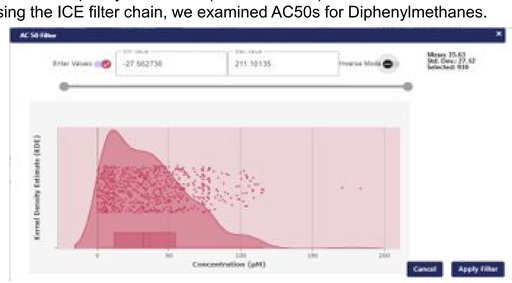


Figure 5: Distribution of AC50 results for the Subclass of Diphenylmethanes

- We retrieved additional information for the 138 Diphenylmethanes with the ICE Search tool.
  - These chemicals have other toxicologically relevant data in ICE, including acute lethality data, endocrine data, cancer data, and cardiotoxicity data.
  - This allows us to further explore bioactivity of Diphenylmethanes in in vivo, in vitro, and in silico data sets.

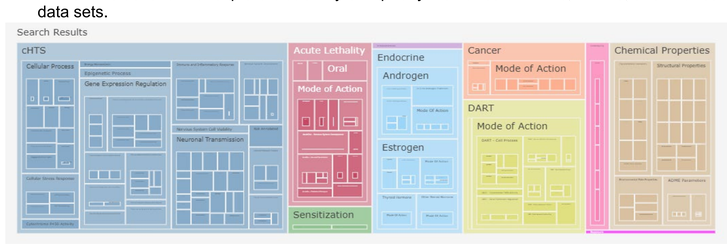
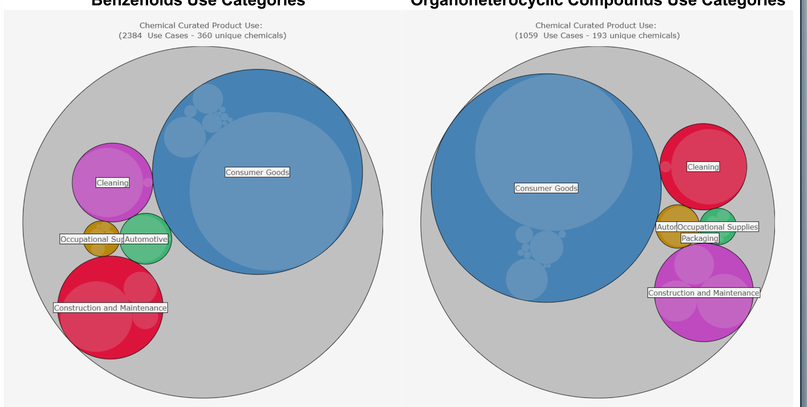


Figure 6: Additional available data in ICE for Diphenylmethanes.

Use Categories

- We ran chemicals classified by ClassyFire as benzenoids and organoheterocyclic compounds through the ICE Chemical Characterization tool to examine potential links between chemical class and product use category. Use categories are derived from EPA's Chemicals and Product Database (Dionosio et al. 2018).
- Use data were available for 360 benzenoids and 193 organoheterocyclic compounds.
  - For both chemical Superclasses, "consumer goods" was the most abundant use followed by "construction/maintenance" and "cleaning."

Benzenoids Use Categories | **Organoheterocyclic Compounds Use Categories**



Future Directions

- We are expanding ClassyFire classifications to the entire DSSTox chemical database of over 1.2 million chemicals.
- ICE does not currently provide a chemical taxonomy. To address this, we are considering including ClassyFire in a future update of ICE, where it would be incorporated into existing tools and represented chemical spaces.
  - **Chemical Characterization:** Visualizations and download files would connect chemical structures with chemical properties and chemical use cases.
  - **Chemical Quest:** ClassyFire is a structure-based taxonomy and thus could assist in narrowing searches by common substructures or levels it has identified.
  - **Search:** With the ability to directly filter by taxonomic classification, users would be able to quickly select chemicals and assay data based on these criteria.

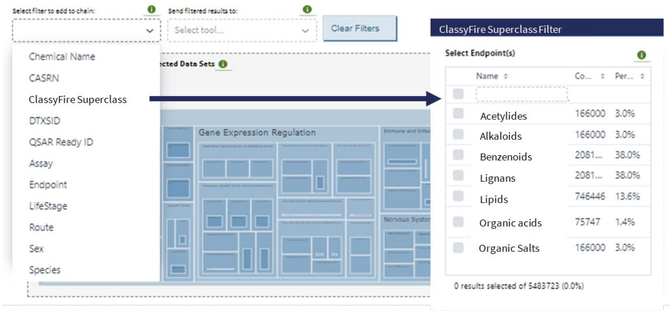


Figure 7: Mockup illustration implementation of ClassyFire in Search.

Summary

- By linking taxonomy to bioactivity, use, and toxicological endpoints, we can explore chemical structures and use cases that potentially contribute to these endpoints.
- We demonstrated the utility of using ClassyFire to focus chemical queries with Tox21 chemicals:
  - We used ICE Search to examine bioactivity data for a specific mechanistic target, then focused the exploration of AC50s by the most abundant Subclasses.
  - We used ICE Chemical Characterization to see the most abundant Superclasses and Classes with product use data and compared use distribution between Superclasses.
- ClassyFire can help focus chemical selection, aid in selection of alternative chemicals, and potentially identify chemical substructures that are correlated with chemical activity and other endpoints.
  - Additionally, taxonomies may be easier for naïve users to understand in comparison to other, more complex structure-based descriptions.
- Classifications will be pulled for all of EPA's DSSTox database and incorporated into future releases of ICE, allowing users to access these taxonomies during their own chemical explorations.

References

Daniel et al. 2022. Front Toxicol. 4:987848. doi: 10.3389/ftox.2022.987848.
Dionisio et al. 2018. Sci Data. 10;5:180125. https://doi.org/10.1038/sdata.2018.125.
Djoumbou et al. 2016. J Cheminfo. 8:16. doi: 10.1186/s13321-016-0174-y.

Acknowledgments

The Intramural Research Program of the National Institute of Environmental Health Sciences (NIEHS) supported this poster. Technical support was provided by Inotiv under NIEHS contract HHSN273201500010C. Since the poster was written as part of the official duties of the authors, it can be freely copied. We thank Elizabeth Farley-Dawson and Catherine Sprankle, Inotiv, for editorial input.