National Institute of Environmental Health Sciences
Division of Translational Toxicology

Integrating Structure-based Chemical Taxonomies to Focus Queries of NAMs Data

V.D. Hull[1], B.A. Chambers[1], T. LaPratt[1], E.N. Reinke[1], N.C. Kleinstreuer[2,3], H.T. Hogberg[3]

[1]Inotiv, United States; [2]NIH/OD/DPCPSI, United States; [3]NIH/NIEHS/DTT/NICEATM, United States

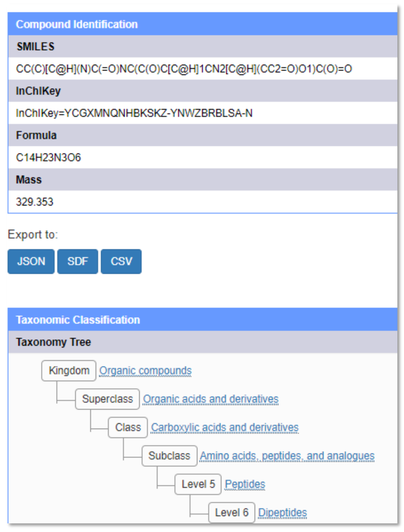
Introduction

- Characterizing and navigating chemical space is a complex challenge. This is especially true for data generated using new approach methodologies (NAMs), as in silico and in vitro approaches typically yield larger volumes of information than traditional animal methods.
  - One way to focus chemical space exploration is through tools such as ClassyFire [1], an automated, structure-based chemical taxonomy tool created by the Wishart Research Group.
- To aid in the exploration of NAMs data, the ClassyFire chemical taxonomy has been incorporated into the Integrated Chemical Environment [2] (ICE; https://ice.ntp.niehs.nih.gov/).
  - ICE is an open-access resource developed by the National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM). It provides toxicologically relevant methods and tools, with a focus on NAMs.
- This poster describes a case study focused on DNA-damaging chemicals that illustrates how ClassyFire classifications and ICE tools can be used to identify groups of chemicals potentially related to adverse effects.

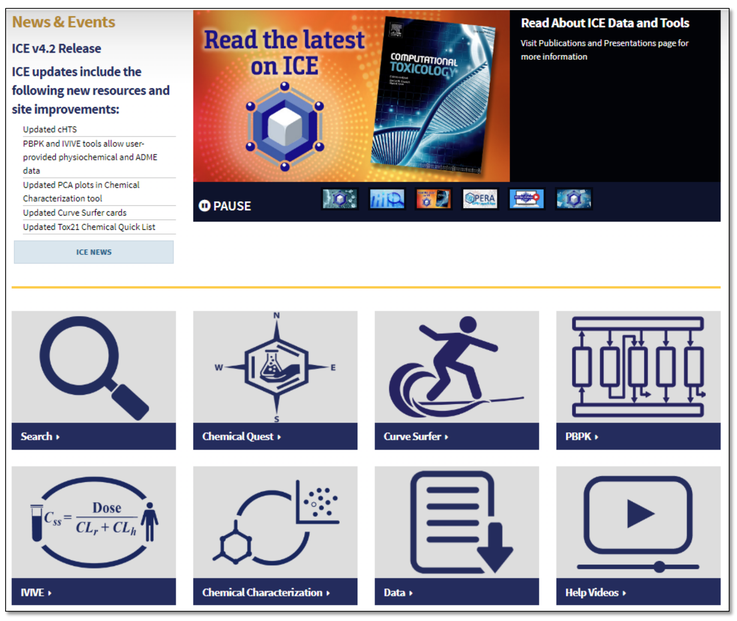
Data Source

- The ClassyFire hierarchical tree has 11 levels with a total of 4,825 unique classifications.
- The ClassyFire tool accepts a variety of chemical identifier inputs, including SMILES, InChi, IUPAC name, and FASTA.
- ClassyFire is accessible through a free web interface, a batch search tool, and an application programming interface (API).

http://classyfire.wishartlab.com



Integrated Chemical Environment



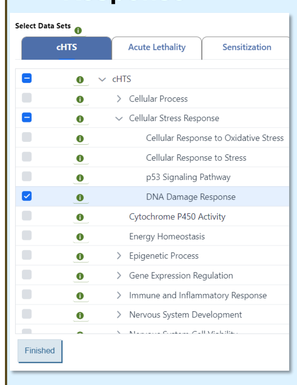
This screenshot shows the ICE home page, which provides easy access to all of its tools and resources.

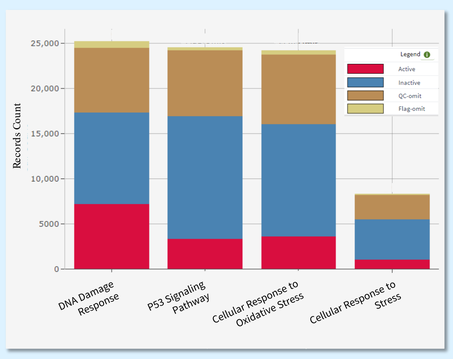
Classifications for DSSTox

- ICE provides access to ClassyFire classifications for all chemicals in the U.S. Environmental Protection Agency (EPA) Distributed Structure-Searchable Toxicity (DSSTox) network [3] (February 2024 release).
- The DSSTox download contains 1,125,991 unique InChIKey chemical identifiers.
  - To generated this download, all InChIKeys were first run through the ClassyFire API, which returned classifications for 1,040,098 InChIKeys. The remaining 84,230 chemicals were run through the ClassyFire web interface using SMILES as the structural identifier.
- Classifications can be downloaded through the ICE Data Sets page (https://ice.ntp.niehs.nih.gov/DATASETDESCRIPTION).

Case Study: Linking DNA Damage Response to Chemical Classifications

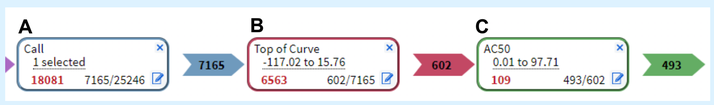
1 Use the ICE Search tool to identify chemicals tested in high-throughput screening assays for "DNA Damage Response"

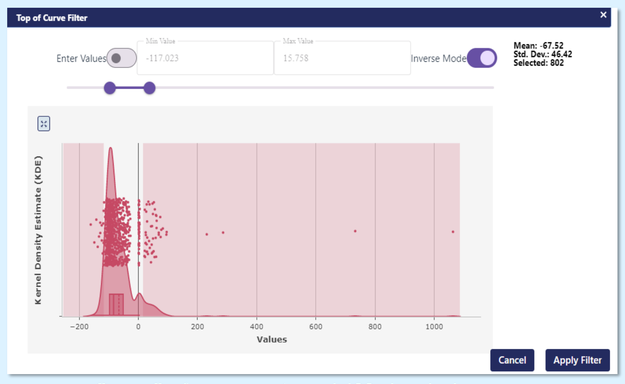




ICE Search query returned 159,444 data records representing 8,107 chemicals and 16,258 assay-chemical pairs.

2 Send ICE search results to the ICE Curve Surfer tool. Use the Curve Surfer filter chain to identify chemicals that are (A) bioactive for DNA damage and (B, C) highly potent.





Curve Surfer query returned 493 chemicals.

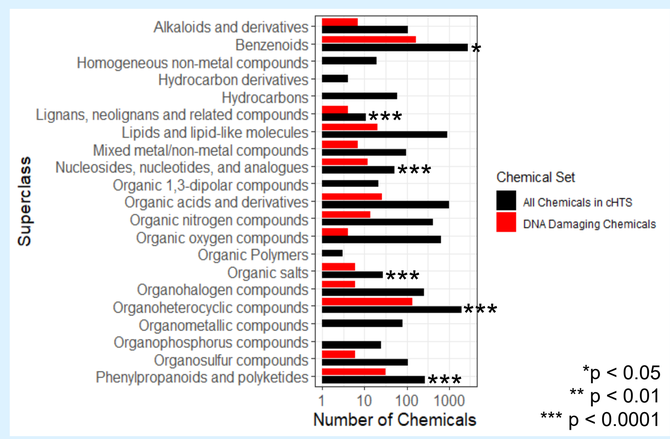
3 Use the ClassyFire download file in ICE to determine the classifications of the bioactive chemicals.

- ClassyFire classifications were available for 492 of the 493 highly potent bioactive chemicals.
- The ClassyFire Kingdom level divides chemicals into organics and inorganics, with organics defined as chemicals with one or more carbon atoms.
  - Most of the 492 chemicals were organic compounds.
- The ClassyFire Superclass level organizes chemicals by general structural identifiers.
  - For our 492 chemicals, the three most abundant superclasses were benzenoids, organoheterocyclic compounds, and phenylpropanoids and polyketides.

Kingdom	Superclass	# of Chemicals
Organic compounds	Benzenoids	162
Organic compounds	Organoheterocyclic compounds	136
Organic compounds	Phenylpropanoids and polyketides	31
Organic compounds	Organic acids and derivatives	26
Organic compounds	Lipids and lipid-like molecules	20
Organic compounds	Organic nitrogen compounds	14
Organic compounds	Nucleosides, nucleotides, and analogues	12
Organic compounds	Alkaloids and derivatives	7
Inorganic compounds	Mixed metal/non-metal compounds	7
Organic compounds	Organic salts	6
Organic compounds	Organohalogen compounds	6
Organic compounds	Organosulfur compounds	6
Organic compounds	Lignans, neolignans and related compounds	4
Organic compounds	Organic oxygen compounds	4
Organic compounds	Organophosphorus compounds	1

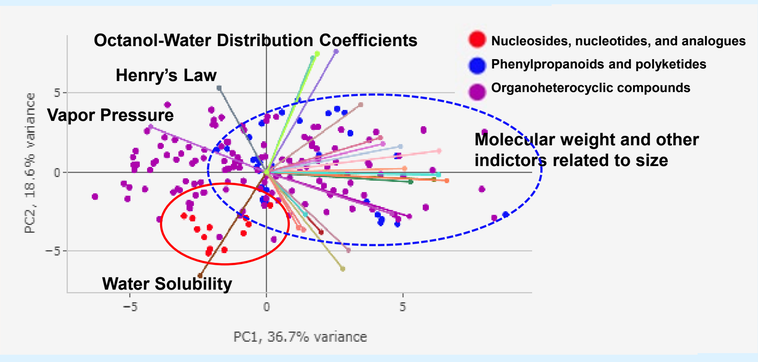
4 Conduct an overrepresentation analysis to see which Superclasses are enriched compared to source data.

- All annotated assays in ICE are derived from the ICE curated high-throughput screening (cHTS) data set.
- We used a hypergeometric test with a Benjamini-Hochberg correction to determine which Superclasses were overrepresented in the DNA Damaging Chemicals compared to all chemicals in cHTS.



For more information about ICE cHTS data, visit Poster 706 (Hull et al.).

5 Conduct principal component analyses of the three most overrepresented chemical classes using the ICE Chemical Characterization tool.



- Nucleosides, nucleotides, and analogues cluster in the bottom left quadrant, which is heavily influenced by water solubility.
- Phenylpropanoids and polyketides cluster along PC1, which is driven by molecular weight and other indictors related to size (number of carbon atoms, number of heavy atoms, number of rotatable bonds, etc.).

6 Link the most overrepresented Superclasses to potential mechanisms for DNA damage.

1) **Nucleosides, nucleotides, and analogues:** these compounds contain a nucleobase linked to a ribose or deoxyribose sugar via a beta-glycosidic linkage. In nucleotides, the ribose or deoxyribose sugar includes at least one phosphate group.
   - Nucleoside analogs are known to affect the structural integrity of DNA [4].
   - Nucleosides and their analogues are incorporated into DNA by polymerases during synthesis, causing replication forks to stall [5].
2) **Phenylpropanoids and polyketides:** these organic compounds are synthesized either from the amino acid phenylalanine (phenylpropanoids) or the decarboxylative condensation of malonyl-CoA (polyketides).
   - Phenylpropanoids and polyketides are often used as anti-cancer agents, as certain phenylpropanoids can inhibit ribonucleotide reductase and DNA polymerase [6].
3) **Organoheterocyclic compounds:** these compounds contain a ring with at least one carbon atom and one non-carbon atom.
   - Certain classes, like quinolines and derivatives, are known to be DNA intercalators and topoisomerase inhibitors [7].
   - Other classes, like carbazoles, can be used as sensitizers of cancer cells to increase DNA damage [8].

Summary and Future Directions

- By linking ClassyFire chemical taxonomies to bioactivity, we can more easily identify groups of chemicals that are potentially related to adverse effects.
- Classifications for all of EPA's DSSTox database are available on ICE as a download file, allowing users to focus their queries of NAMs data.
- The case study above shows how ClassyFire and ICE can be used together to identify chemicals that are highly potent for DNA damage in in vitro systems.
  - Six Superclasses were found to be significantly overrepresented amongst the highly potent chemicals.
  - The three most overrepresented Superclasses have literature evidence of causing DNA damage.
- Use of the ClassyFire taxonomy can help focus chemical selection, aid in selection of substitute chemicals, and potentially identify use-case scenarios for chemicals based on their substructures.
- Future work will aim to tie classifications to genetic target, creating a more in-depth pipeline that links chemical activity and structure to mechanisms for adverse effects.
- ClassyFire classifications will eventually be integrated into other ICE tools, allowing users to better access these taxonomies during chemical explorations.

References and Acknowledgements

[1]Djoumbou et al. 2016. J Cheminfo. doi: 10.1186/s13321-016-0174-y.
[2]Daniel et al. 2022. Front Toxicol. doi: 10.3389/ftox.2022.987848.
[3]CCTE, EPA. 2019. doi: 10.23645/epacomptox.5588566.v7.
[4]Sampath et al. 2003.Oncogene. doi: 10.1038/sj.onc.1207229.
[5]Ewald et al. 2008. Oncogene. doi: 10.1038/onc.2008.316.
[6]Fontecave et al. 1998. LettFEBS. doi: 10.1016/s0014-5793(97)01572-x.
[7]Lauria et al. 2021. EJMECH. doi: 10.1016/j.ejmech.2021.113555.
[8]Yoon et al. European J of Pharmacology. doi: 10.1016/j.ejphar.2012.10.001.

This project was supported by the Intramural Research Program (ES# 103386) at the National Institute of Environmental Health Sciences, National Institutes of Health under contract HHSN273201500010C. Since the poster was written as part of the official duties of the authors, it can be freely copied. We thank Catherine Sprankle, Inotiv, for editorial input.