Revisiting and updating chemical groupings with new approach methodologies

US EPA in collaboration with Health Canada,
Environment Climate Change Canada

**The views expressed in this presentation do not represent US EPA policy or endorsement. Mention of trade names of commercial products should not be interpreted as an endorsement.*

Team Members

Accelerating the Pace of Chemical Risk Assessment (APCRA)

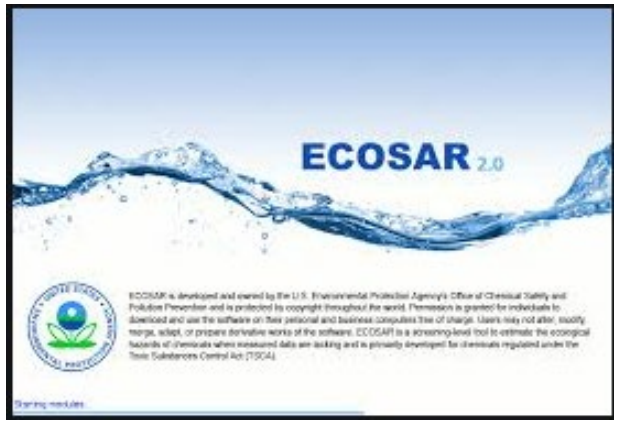
- US EPA
  - Dan Chang
  - Kellie Fay
  - Kristan Markey
  - Martin Phillips
  - Grace Patlewicz
  - Ann Richard
  - Gino Scarano
  - Mahmoud Shobair
  - Haseeb Baha (summer intern)
  - Ellery Saluck (summer intern)
- Environment & Climate Change Canada (ECCC)
  - John Prindiville
- Health Canada
  - Mark Lewis
- ILS
  - Kamel Mansouri

Overview

A chemical category is a group of chemicals whose physicochemical and human health and/or ecotoxicological properties and/or environmental fate properties are likely to be similar or follow a regular pattern, usually as a result of structural similarity. - OECD

- Applications of chemical categorization include first tier assessment efforts and read across from structurally similar analogs:
  - Toxic Substances Control Act (TSCA) New Chemical Program Chemical Categories (NCC; US EPA)
  - ECOSAR (focus of presented work)

US EPA ECOSAR Chemical Classifications



- Class-based SAR to predict aquatic toxicity
- Classification scheme identifies excess toxicity
- Estimates **acute** and **chronic toxicity** based on accumulated data and past decisional precedents

Acute Effects:	Chronic Effects:
Fish 96-hr LC_{50}	Fish ChV
Daphnid 48-hr EC_{50}	Daphnid ChV
Algae 72/96-hr EC_{50}	Algae ChV

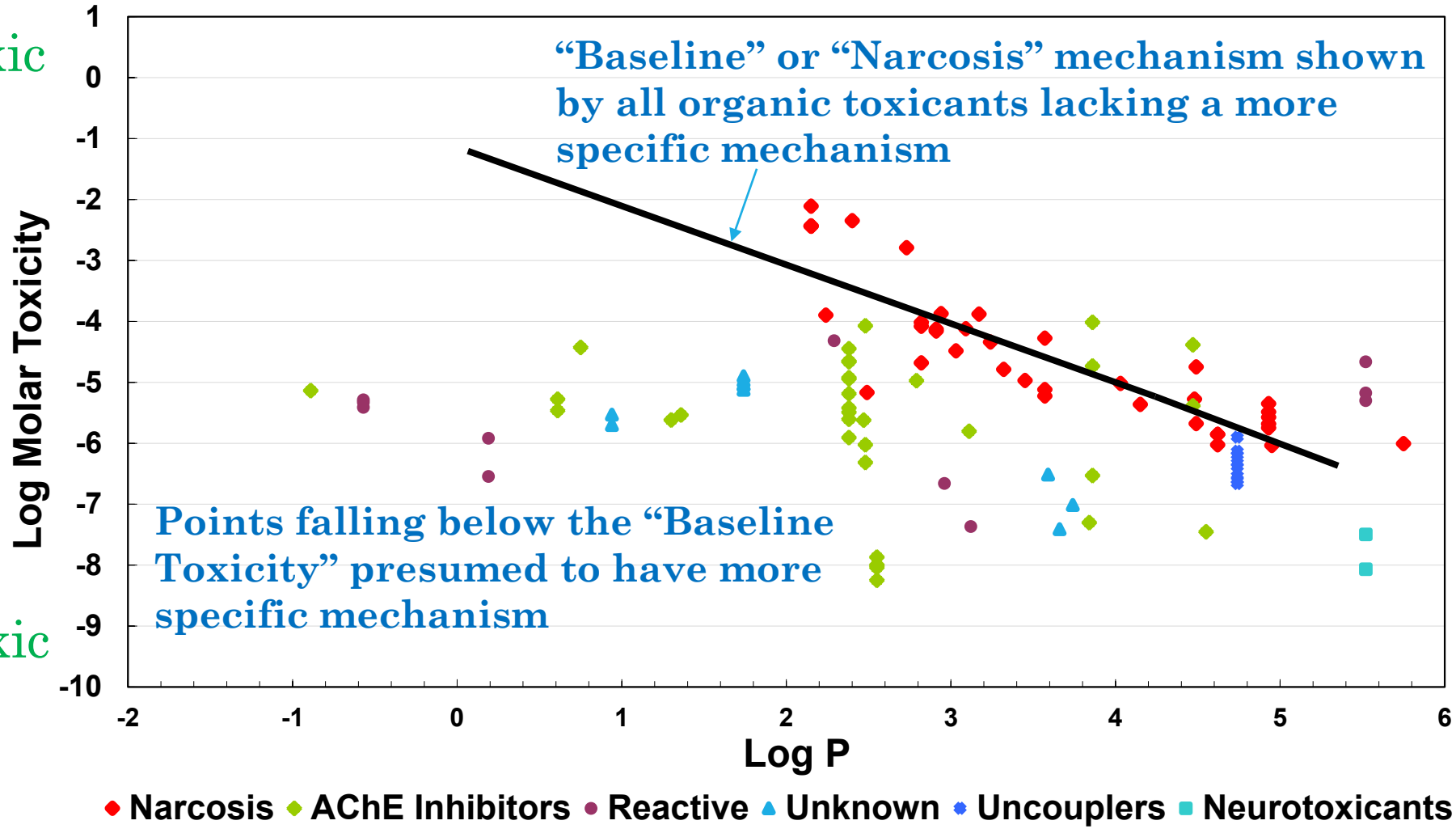
- Profiler in OECD QSAR Toolbox

Narcosis vs. specific-acting toxicity MOA

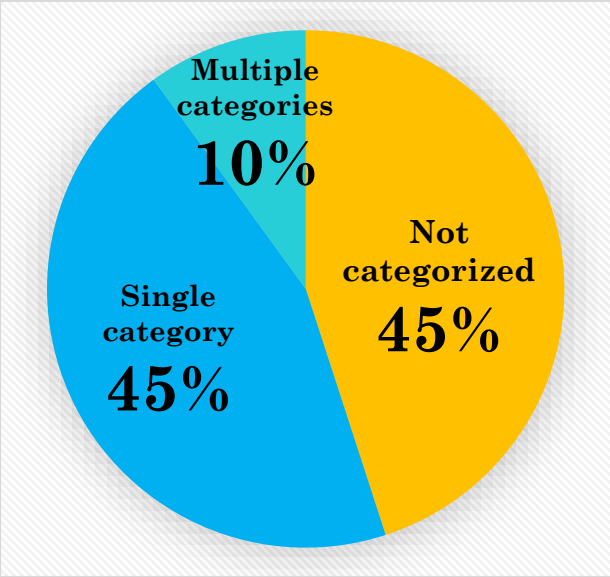
Less Toxic

Regulators (ECCC) consider MOA information to determine the size of assessment factors

More Toxic



Potential approach for updating chemical categories



- Almost half of all New Chemical inventories across regulatory jurisdictions cannot be categorized using NCC or ECOSAR
- Some fall into multiple categories

How to update?

- Incorporate New Approach Methodologies (NAMs) – *i.e.*, ToxCast and Tox21 biological activity information
- Apply cheminformatic approaches

General approach

Training set chemicals

- Well-defined MOA (narcosis vs. specific-acting)
- NAM data *in vitro* toxicity data
- *in vivo* toxicity data
- Representative of chemicals of interest for prediction

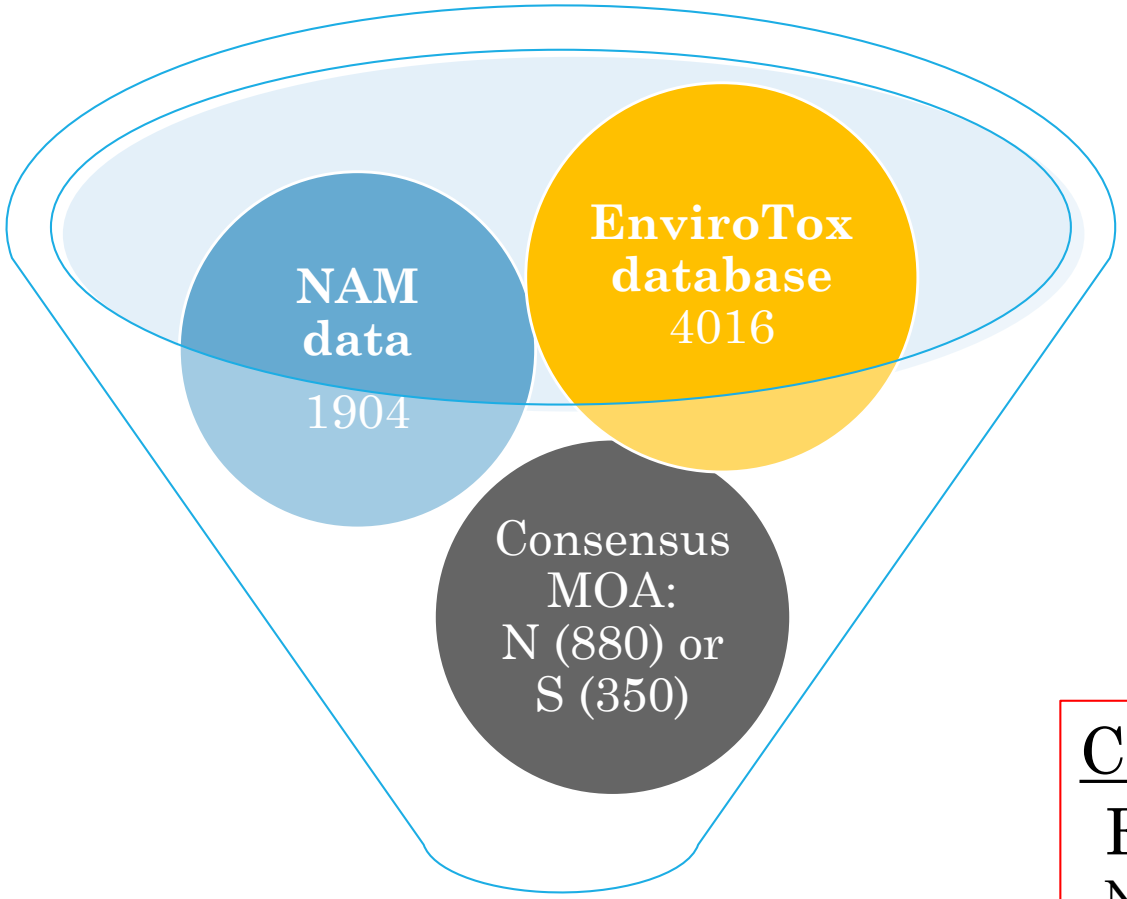
Characterize training set

- 1. ECOSAR classes
- 2. NCC
- 3. Chemotype fingerprints (ToxPrints)

Model

- NAM data, chemotypes and combination of both
- Evaluate different machine learning algorithms

Training set chemicals



Training set chemicals

1. Chemicals with *in vivo* eco-data – from the EnviroTox[1] database – 4016
2. Sub-selection for chemicals with NAM data (ToxCast and Tox21) - 1904
3. MOA predictions based on 4 publicly-available classification models
   - VERHAAR, ASTER, OASIS, TEST
   - Each predicts Narcotic, Specific-Acting or Unclassified

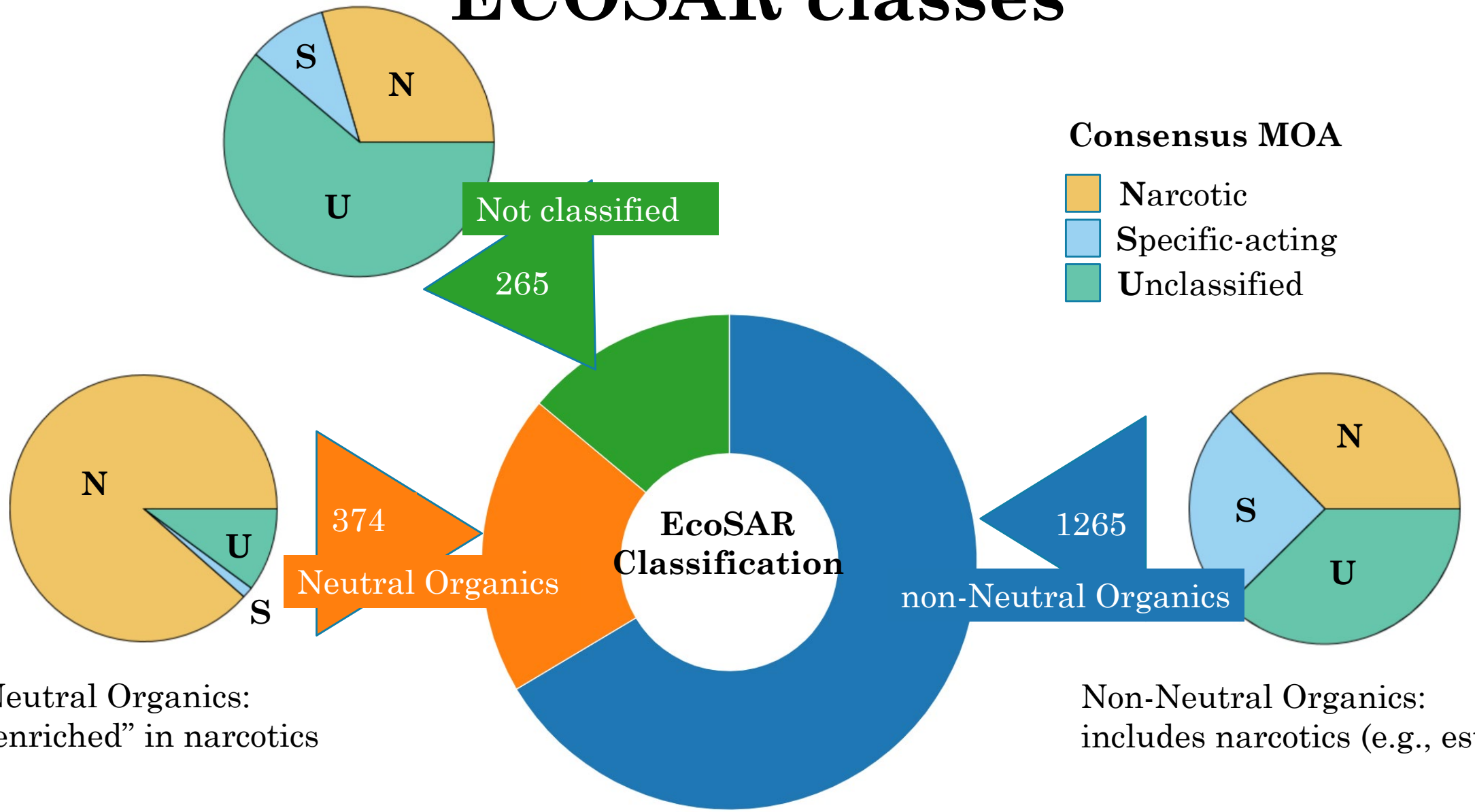
Consensus MOA with confidence scores[2]

Examples:	Results:
NNNN = N, score =3	880 Narcotic
NNSN = N, score= 2	350 Specific-acting
SUSS = S, score= 2	674 Unclassified
NUNS = U, score = 0	

[1]Health and Environmental Sciences Institute (HESI). 2019. EnviroTox Database & Tools. Version 1.1.0 Available: http://www.envirotoxdatabase.org/

[2] Kienzler et al.. Environ Toxicol and Chem. 2019, 38(10) 2294-2304

Characterize training set chemicals: ECOSAR classes



Neutral Organics:
"enriched" in narcotics

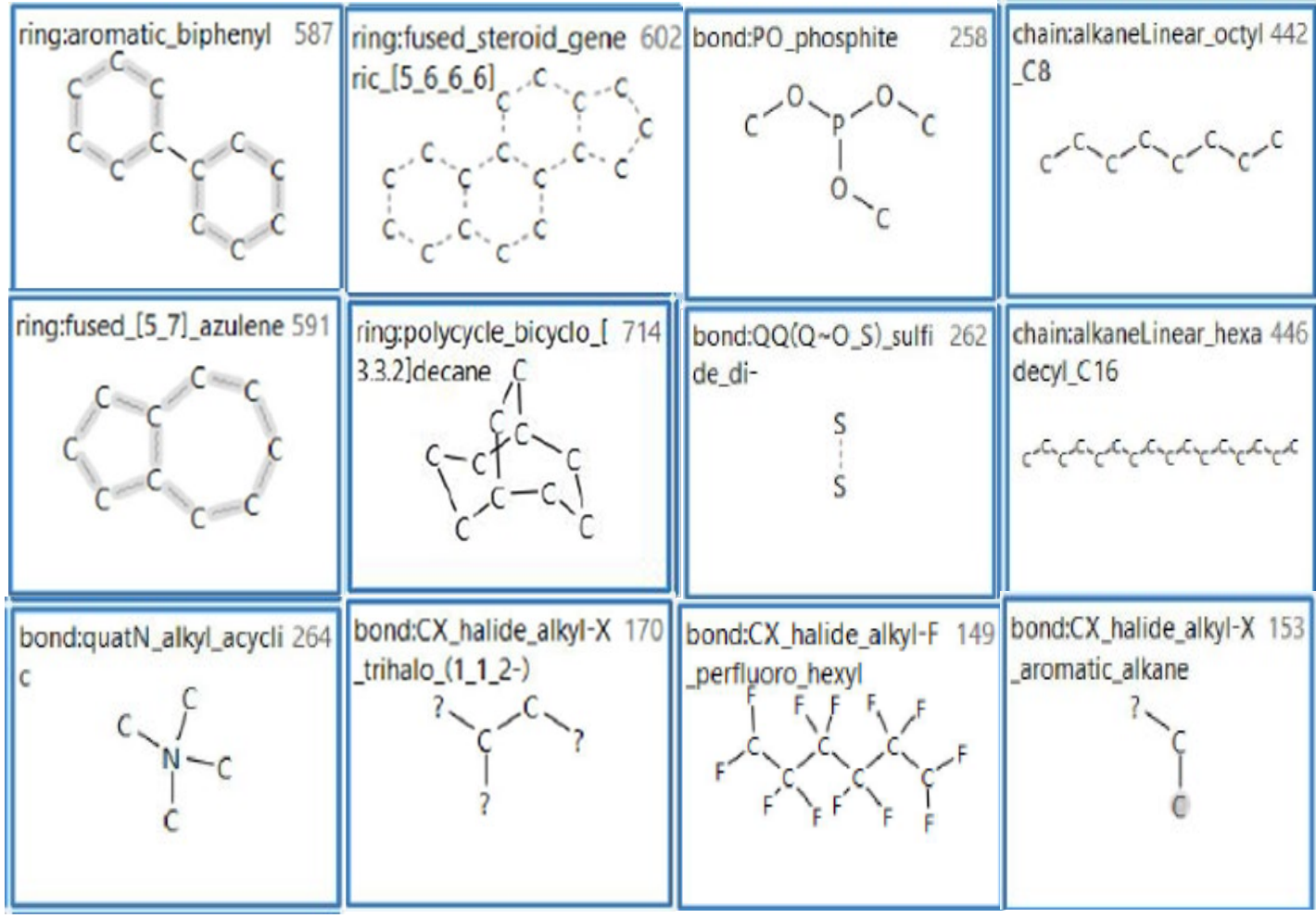
Non-Neutral Organics:
includes narcotics (e.g., esters)

Characterize Training Set Chemicals: ToxPrints

- Pull in chemotype information for our chemicals via ToxPrints
  - Publicly available tool
  - EPA Comptox Chemistry Dashboard

ToxPrints:

- ✓ *729 chemical features*
- ✓ *Chemically interpretable*
- ✓ *Coverage of diverse chemistry*
- ✓ *Includes scaffolds, functional groups, chains, rings, bonding patterns, atom-types*



Yang et al. J. Chem. Inf. Model. 2015. Richard et al., Chem. Res. Toxicol. 2016, 29(8) 1225 – 1251; Strickland et al., Arch Toxicol. 2018 92(1) 487 – 500; Wang et al., Environment International 2019, 126 377 – 386

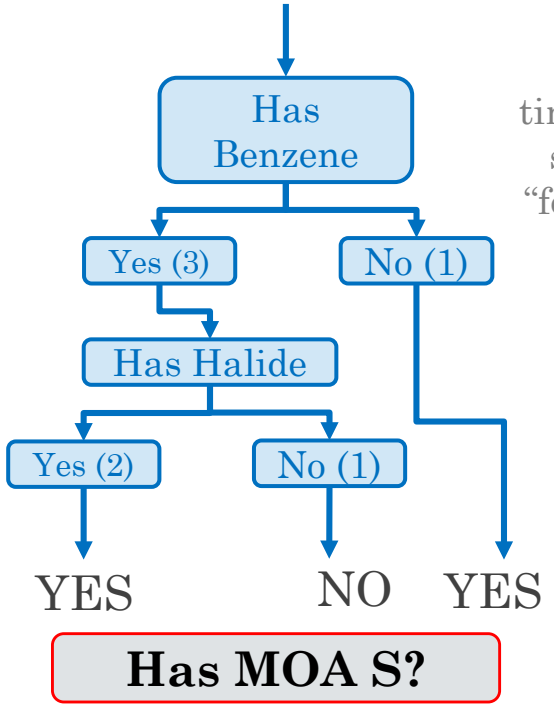
Classification model development

Train the Model

ToxPrints, NAM data or both

	cMOA	Has Metal	Has Halide	Has Benzene	Has Alkyne
Chemical 1	N	0	1	0	0
Chemical 2	N	0	0	1	1
Chemical 3	S	0	1	1	0
Chemical 4	S	1	1	1	1

cMOA Features



Repeated many times with different samples to build "forest" of classifier trees.

Figure adapted from Katherine Phillips

Valid models **must**:

- accurately predict the training set
- predict beyond the training set
- be more predictive than a model built on randomized data

Predict with the Model

Chemical 5	0	1	1	0
Chemical 6	0	1	0	0
Chemical 7	0	0	1	1

Features

In Applicability Domain of Model?

- Has Benzene → Has Halide → YES
- No Benzene → NO
- Has Benzene → No Halide → YES

Repeated with each "tree"

p

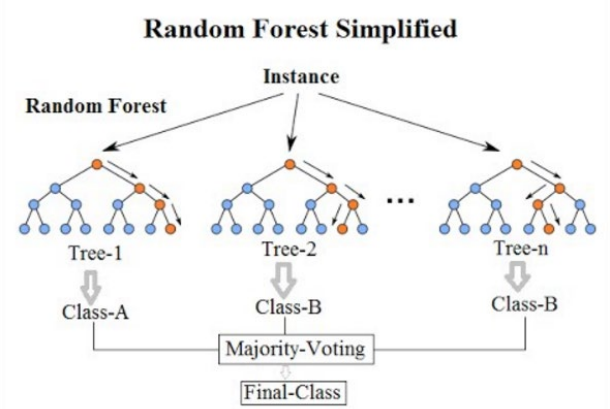
Probability for N or S

Predict target with ***valid*** models using features

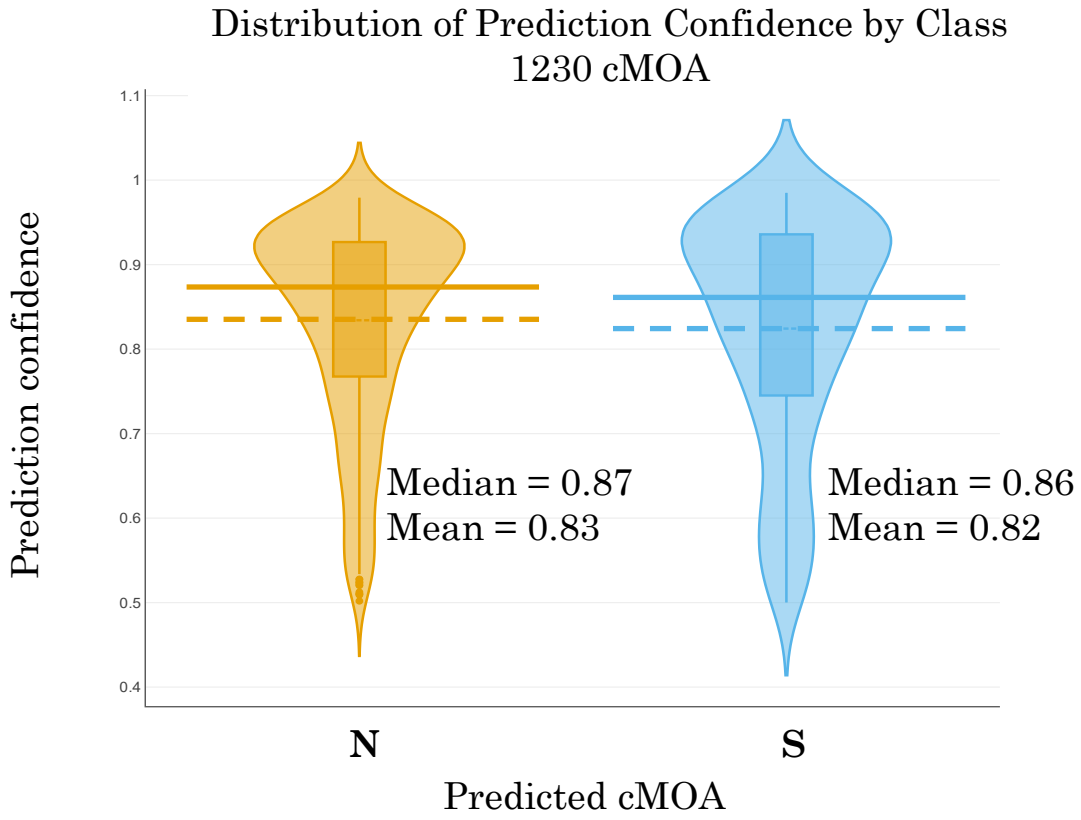
Has MOA S?

Preliminary results

- Random Forest provided the best model results:
  - Trained on a “balanced” down-sampled subset (675 cMOA N+S)
  - Training Out-of-Bag (OOB) error rate = 10.2%
  - Total Accuracy on the full N+ S data set = 94.5% (1230 cMOA N+S)
  - 68 chemicals misclassified:
    - 11 F_{pos}{predicted S}
    - 57 F_{neg}{predicted N}

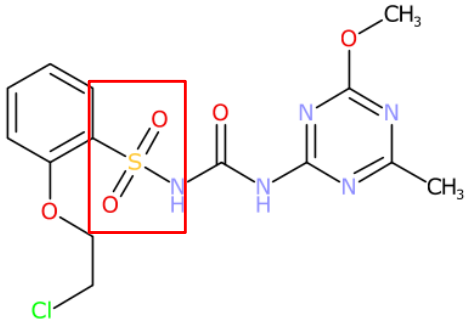


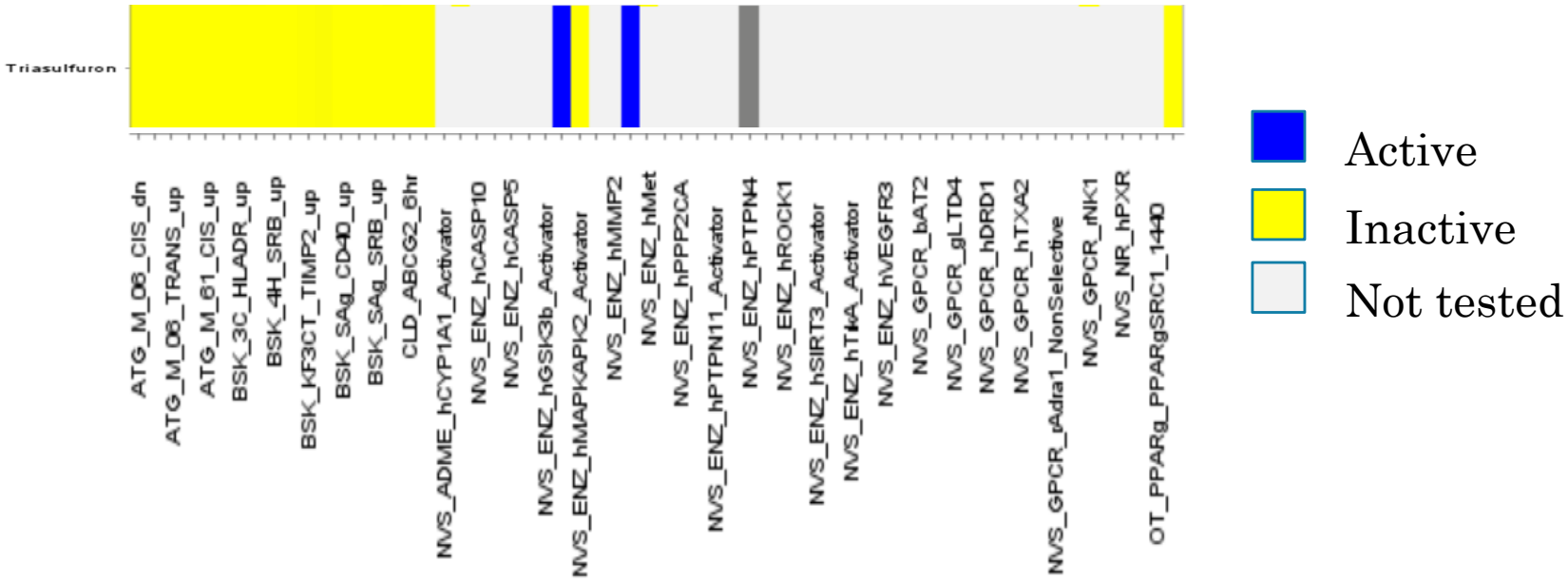
https://medium.com/@williamkoehrsen/random-forest-simple-explanation-377895a60d2d



Example: Differences in model prediction vs. cMOA: Triasulfuron

- N-sulfonylurea herbicide
- Model prediction: Specific-acting
- EnviroTox consensus MOA: Narcotic
- ECOSAR classification: Sulfonyl Urea
- S(=O)_sulfonyl ToxPrint is enriched in the specific-acting MOA space and 47 assays

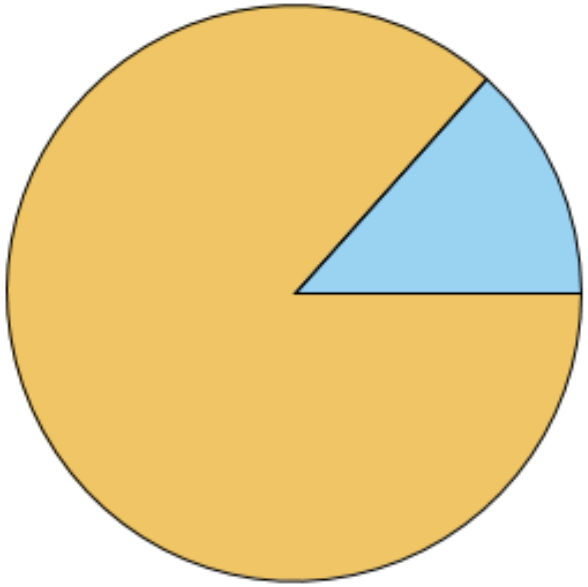




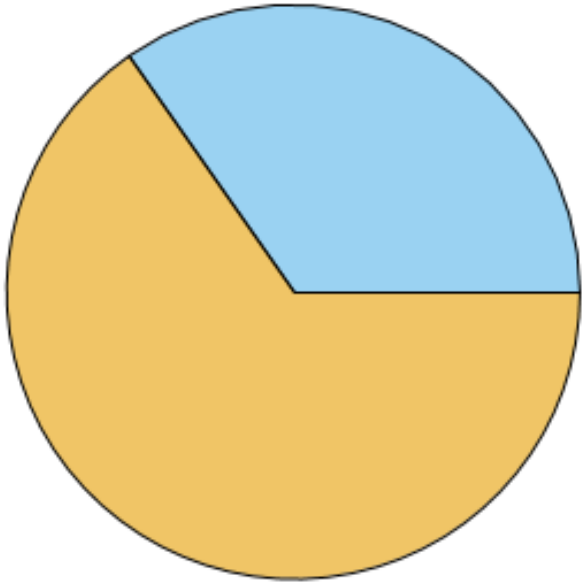
Predicted MOAs of the Unclassified set

- 674 chemicals in the EnviroTox dataset that had low confidence or ambiguous consensus
- Applied model to the Unclassified set and compared predictions to ECOSAR classification

361 predicted as Narcotic



313 predicted as Specific-acting



ECOSAR Classified

ECOSAR Not Classified

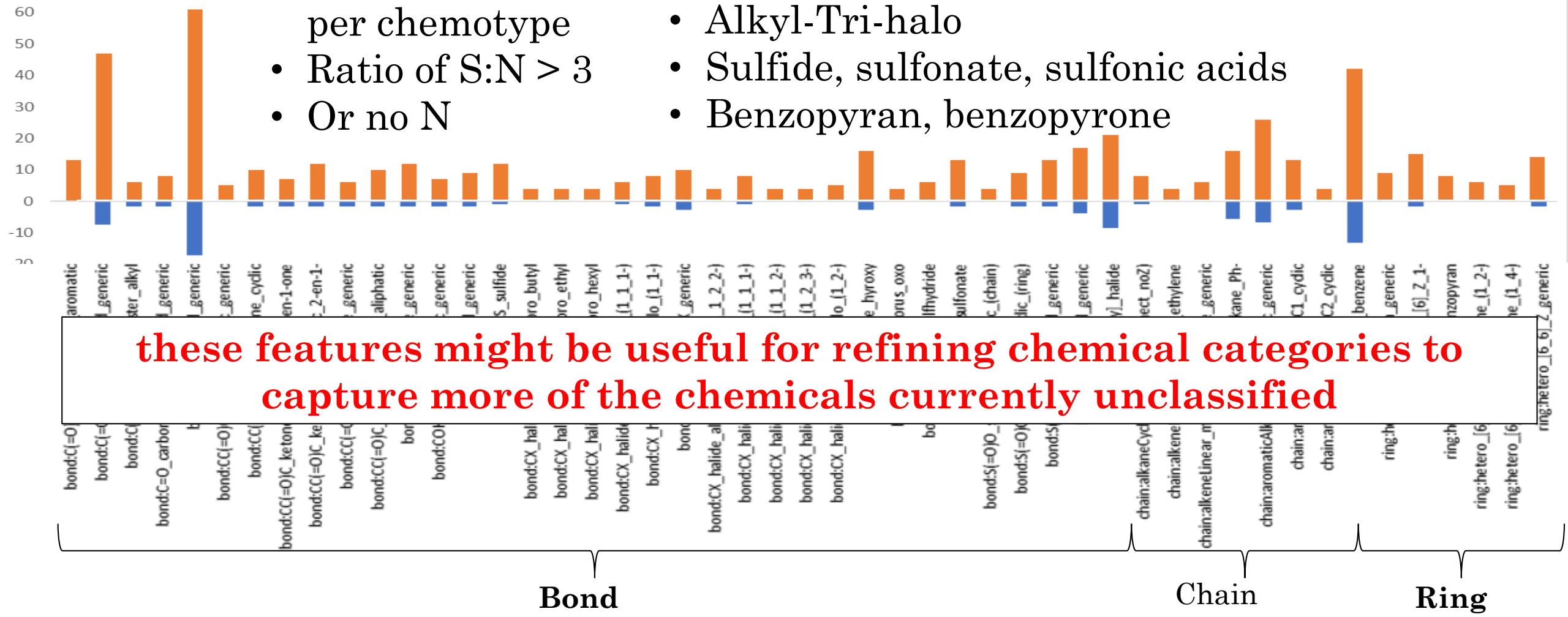
Unclassified chemicals, predicted Specific-Acting: Enriched ToxPrints

Criteria:

- ≥ 3 chemicals per chemotype
- Ratio of S:N > 3
- Or no N

Results:

- Ketones
- Alkyl-Tri-halo
- Sulfide, sulfonate, sulfonic acids
- Benzopyran, benzopyrone



Summary

- Identified relevant NAM information to develop a classification model for specific-acting MOAs
- Explored differences in predicted and consensus MOA via chemotype enrichments
- Used model to inform ECOSAR unclassified chemicals
  - Majority of unclassified chemicals were predicted to have a specific acting MOA
  - Identified primary chemotypes for specific acting MOAs

Next steps/ongoing work

- Leverage more invitroDB chemicals beyond the 1905 EnviroTox chemicals
  - Generated KNIME workflow for the consensus MOA calls
  - Greater coverage of the NAM assay space
    - >7000 chemicals with MOA calls
    - Integration of HTS and transcription assay data
- Use methods to inform classification models for TSCA (New Chemical Categories)
- Use chemotype enrichments to identify potential bioassays with bioactivity

Thank you!