EPA United States Environmental Protection Agency

New Tools for High-throughput Screening of Chemicals for Functional Substitutes

Katherine Phillips, John Wambaugh, Chris Grulke, Kathie Dionisio, Kristin Isaacs

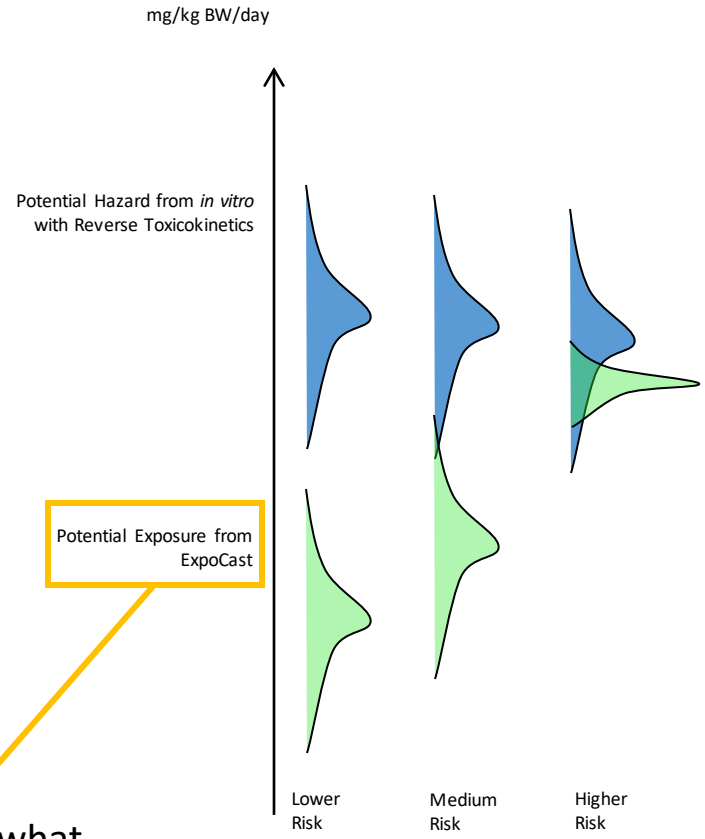
The views expressed in this presentation are those of the author(s) and do not necessarily represent the views or policies of the U.S. Environmental Protection Agency.

DOI: 10.1039/C6GC02744J

Outline

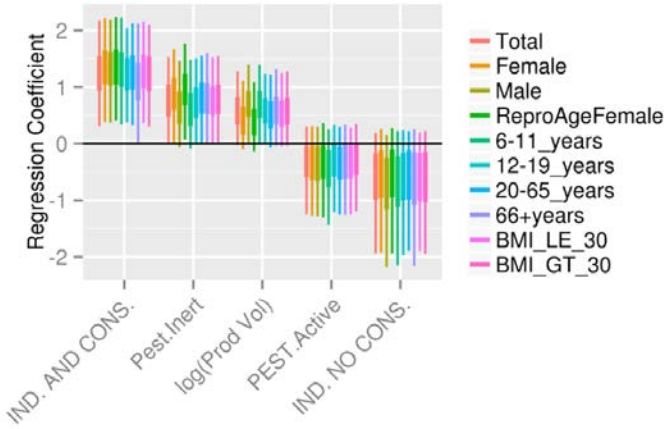
I. Introduction
II. Data Curation
    A. Data Collection
    B. Functional Use Harmonization
III. QSAR Workflow and Validation
IV. Application of Valid Models
    A. Predict functional uses of Tox21 for alternatives screening
    B. Suspect screening of consumer products

Rapid Exposure and Dosimetry

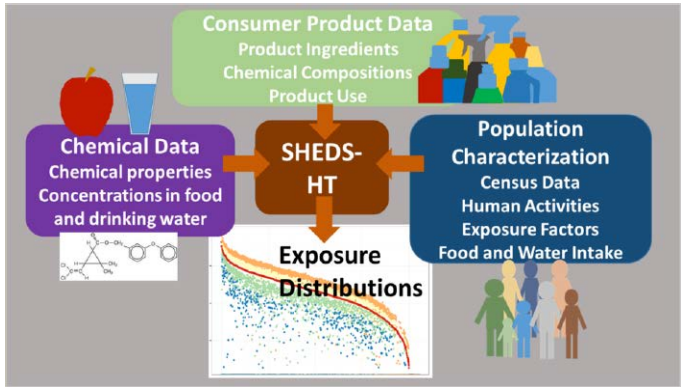


e.g. Judson *et al.*, (2011) Chemical Research in Toxicology

Need to know what people are exposed to and how much



Wambaugh 2014 used Use-based heuristics for exposure predictions



Isaacs 2014 used information on consumer products in SHEDS-HT for exposure predictions

Chemical Uses

- Volumetric Use:

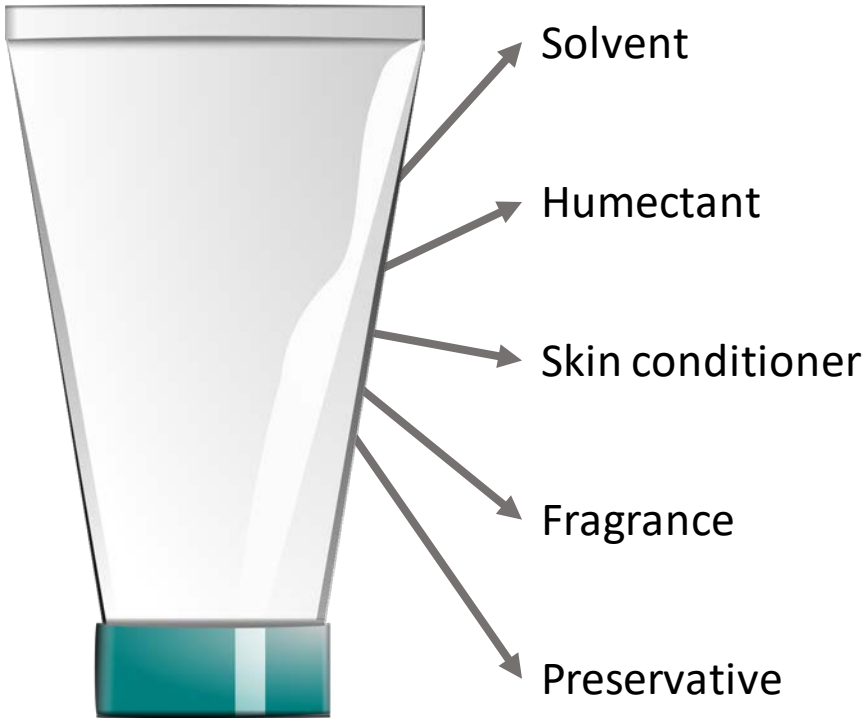
  amount used, produced, or imported

- Applicative Use:

  what products or processes a chemical is used in for a product or chemical product

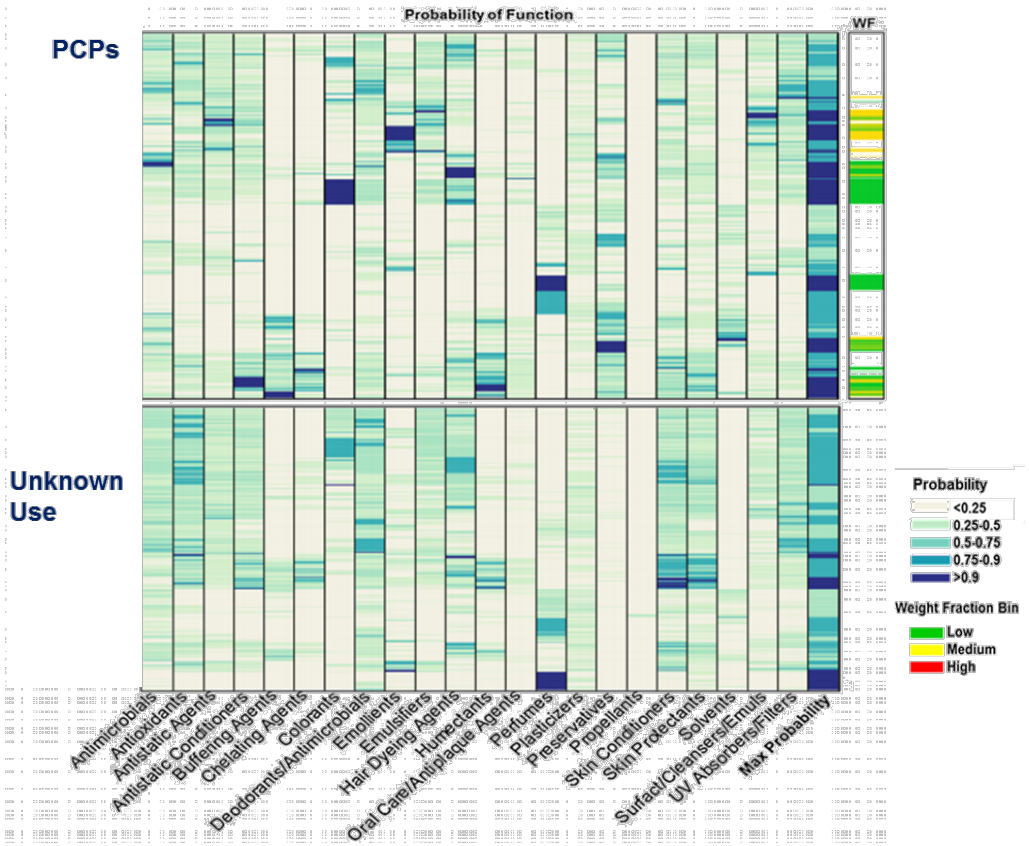
- Functional Use:

  why and how a chemical is used (i.e. its purpose) –how is it used in its application



Consumer Product Weight Fractions

- Used general use categories (applicative uses) and chemical properties
- Built models to predict function
- Used function and properties to predict weight fraction bin
- CosIng and CPCPdb databases as training set
- Only 26 functions



K. K. Isaacs, et al., *Tox. Reports*, 2016

Data Curation

Collect CASRN and functional uses from online data sources

CASRN | Functional Use

Unique CAS – Functional Use pairs

~32,263 records
224 functions

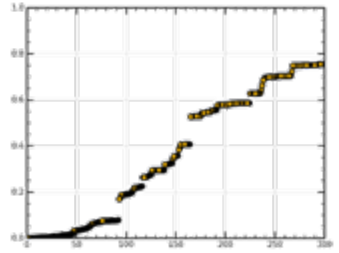
Create one-to-one mapping of chemicals to functional use

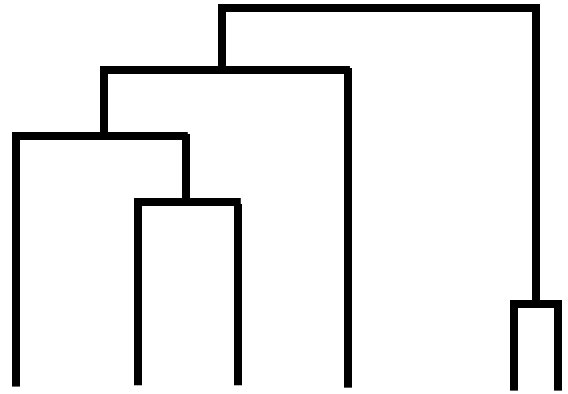
CASRN | Harmonized Function

Unique CASRNs

~14,000 unique CASRNs
137 harmonized functions

Assign each CASRNs to its new functional use identifier (cluster)





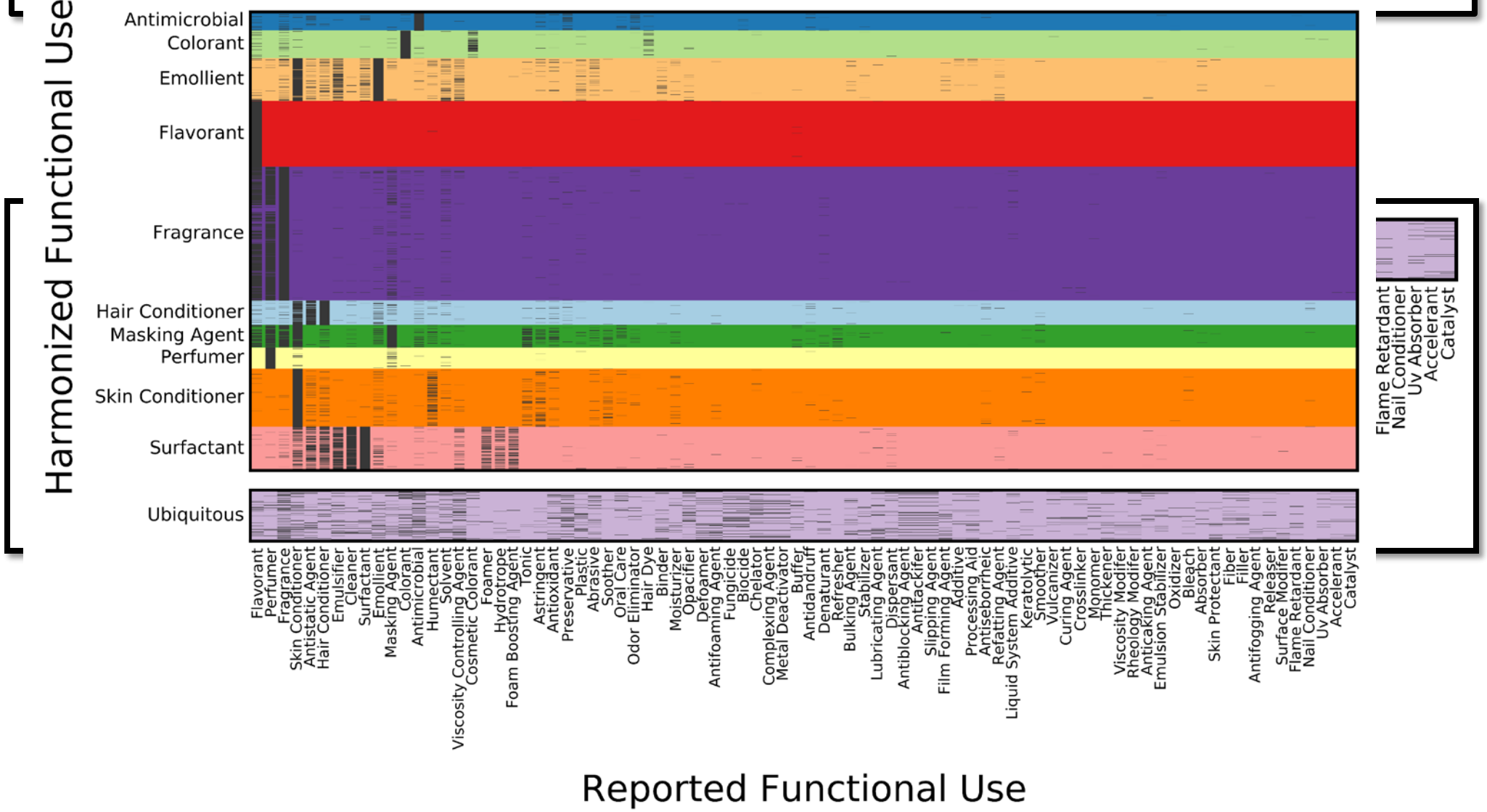
Hierarchical Clustering of Functional Use

Polyethylene – common plastic

Stearic Acid – foods, personal care, battery production, lubricants

Water

Ethanol



Chemical Descriptors

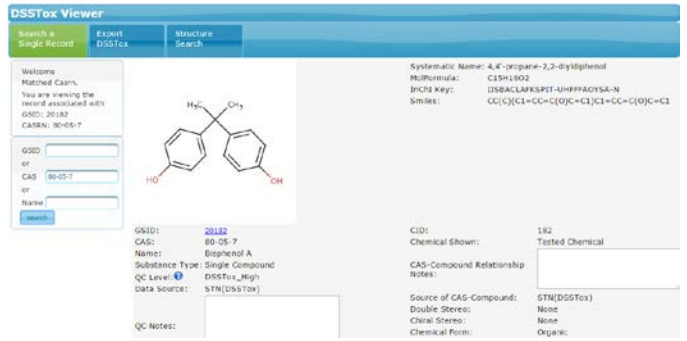
CASRNs with Harmonized Function

14,272 chemicals
137 functions

DSSTox

Curated SMILES strings

5,806 chemicals
98 functions



EPI Suite

4,791 chemicals
84 functions

11 Physicochemical Properties

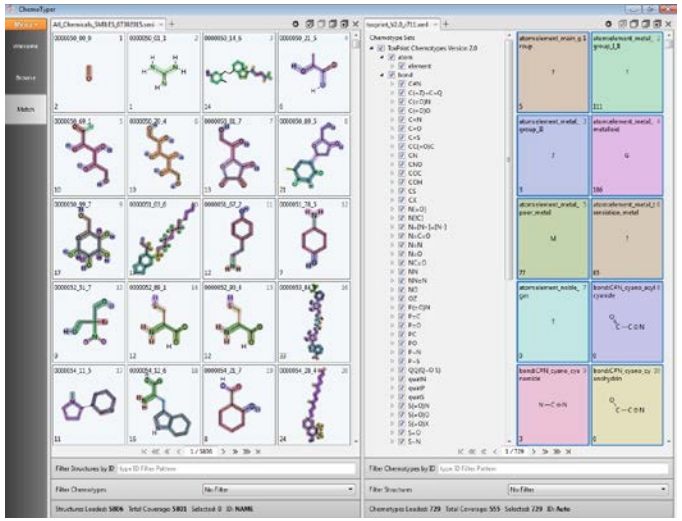
4,667 chemicals
43 functions

Chemotyper

729 ToxPrint Descriptors

5,806 chemicals
98 functions

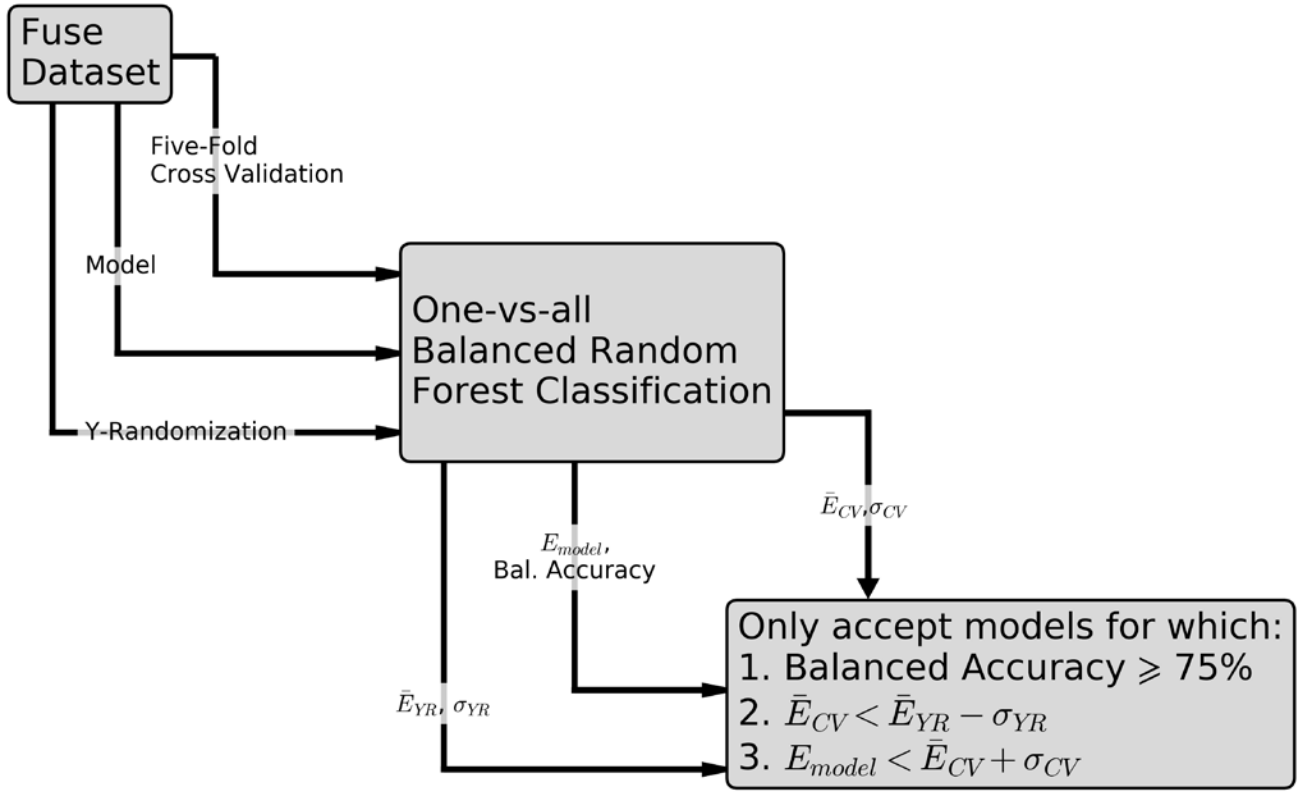
5,666 chemicals
49 functions



Require 10 Chemicals per function

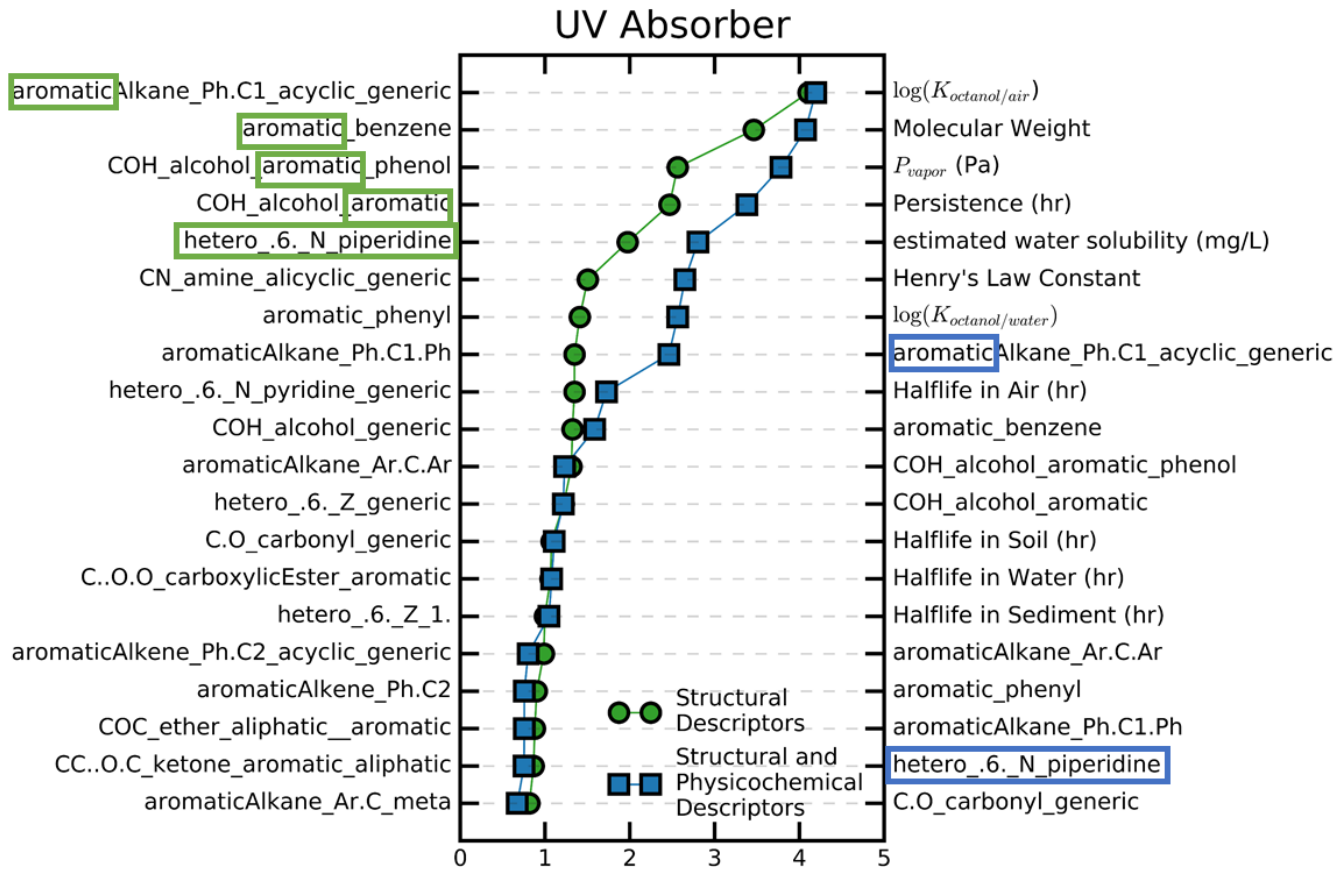
Quantitative Structure-Use Relationships

- Two model sets were built
  - Only structural descriptors
  - Structural descriptors + physicochemical properties
- One-vs-all: one model was built for every functional use
- Balanced: accounts for minority dataset



Variable Importance

- Aromatic and amine groups make excellent UV absorbers due to resonance
- Hindered amine light stabilizers are typically derived from the compound 2,2,6,6-tetramethylpiperidine



Valid Models

- 41 valid QSURS
- Model set with highest balanced accuracy was “best”
- Very vague categories yielded poor models

Bioactivity Index

- 8,600 Tox21 chemicals
- Used 16 assays that covered 5 pathways
- Pathways were known to be altered when exposed to environmental contaminants
- Bioactivity index is average of all “hit calls” for a chemical over all reported assays and duplicates
- 6,672 Tox21 chemicals had results for these assays
- 6,365 Tox21 had QSUR descriptors

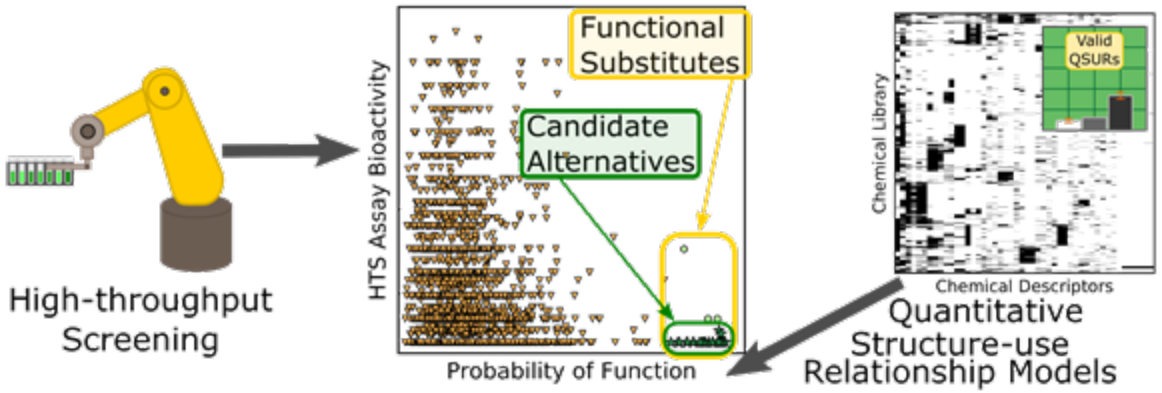
Alternatives Screening

Functional Substitutes

- Within domain of applicability of QSUR
- QSUR prediction of 80% or higher

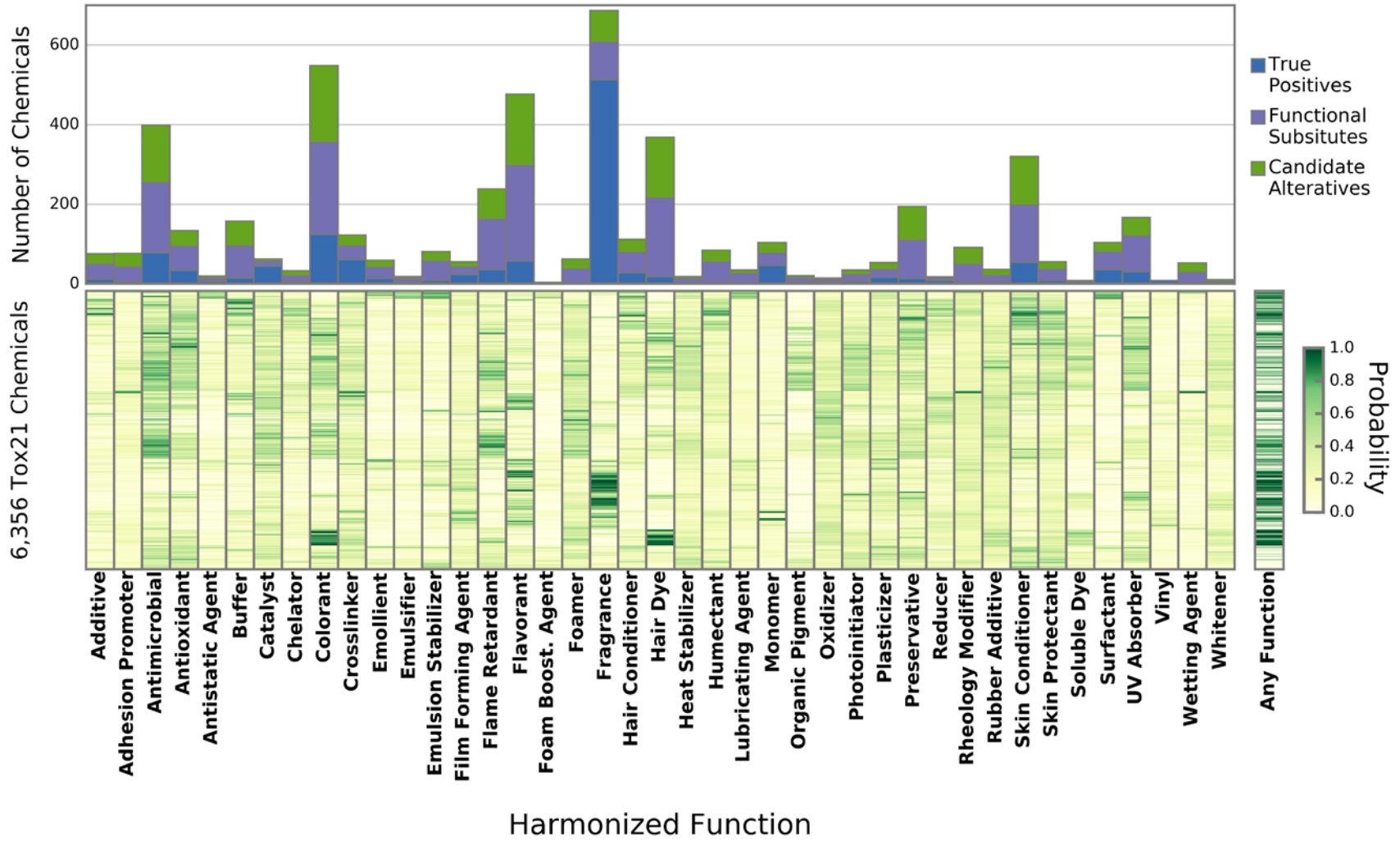
Candidate Alternatives

- Must meet requirements of functional substitute
- Bioactivity index must be lower than the 75th percentile of chemicals known to have function



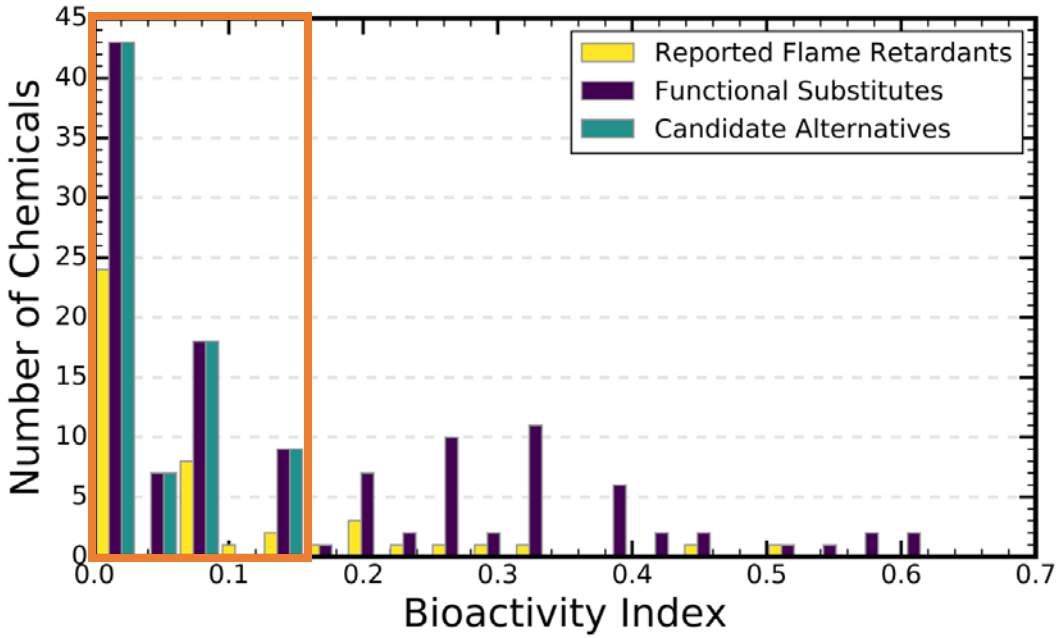
Functional Substitutes

- Models were selective
- Predictions show relationships between functions
- Models for fragrances, colorants, and flavorants yielded most positive predictions



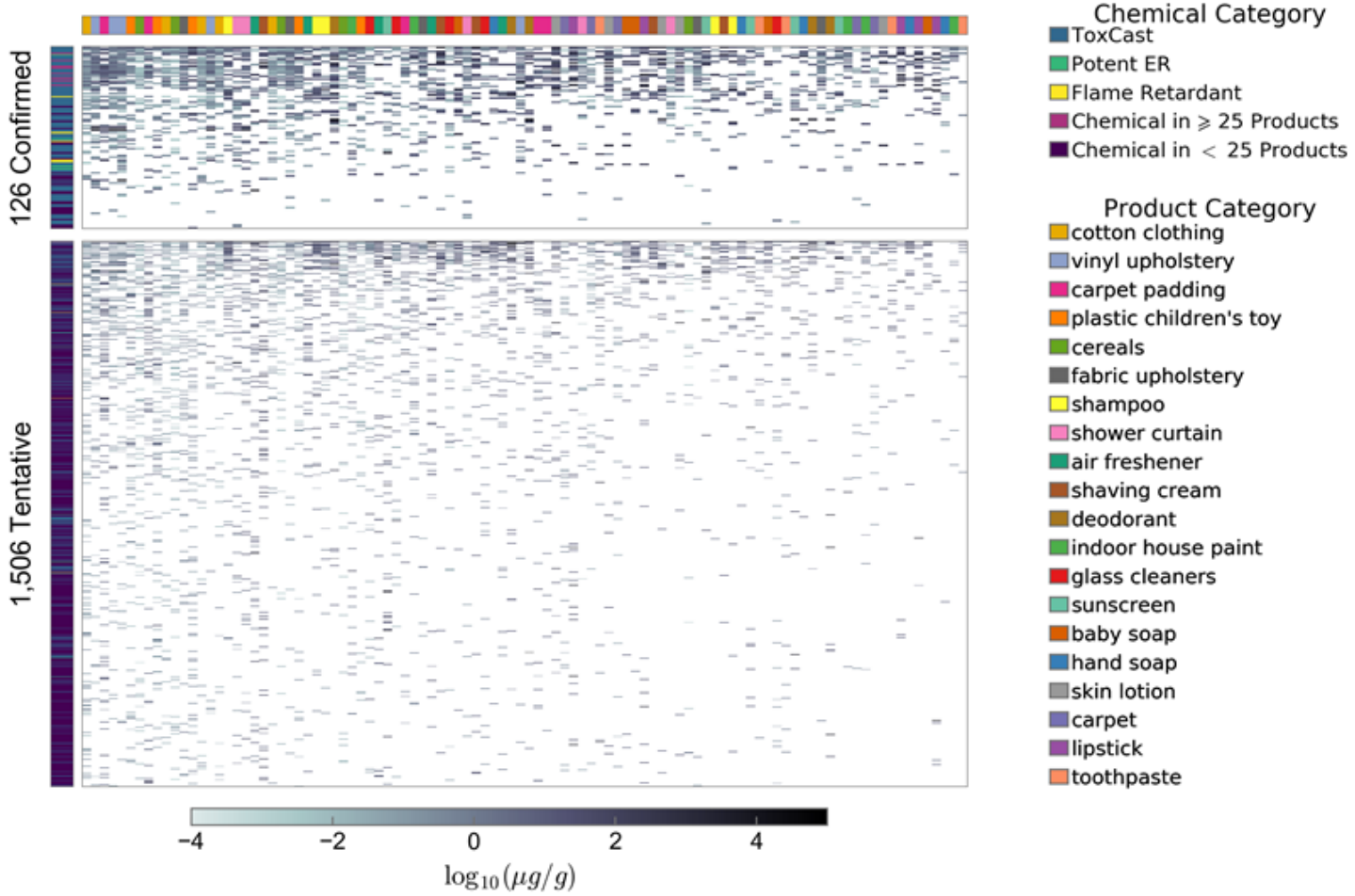
Case Study: Flame Retardants

- 45 reported flame retardants in Tox21
- 126 functional substitutes
- What is the percentile
- 77 candidate alternatives



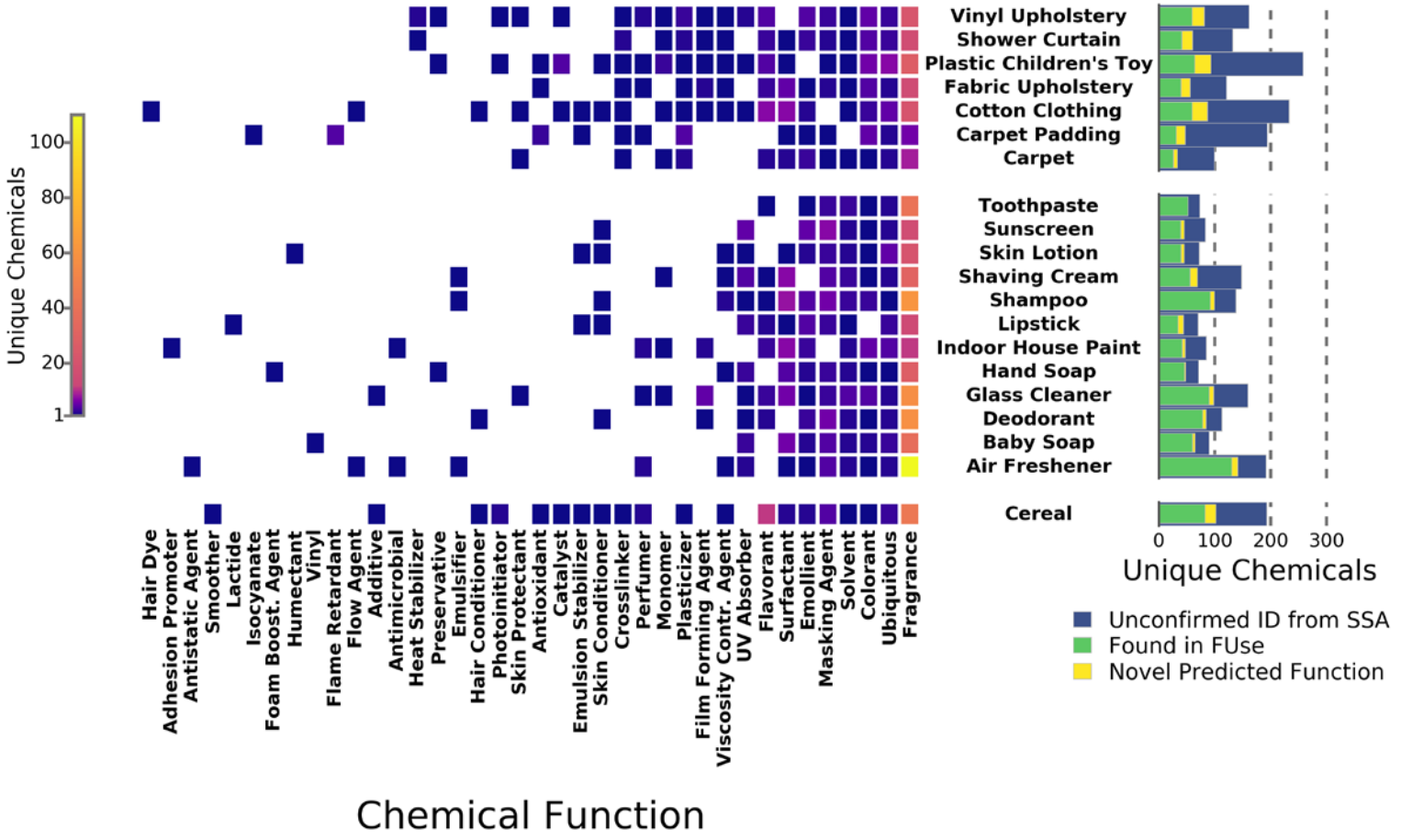
Application to Non-targeted Analysis

- 20 product categories, 5 products per category
  - 7 categories of articles
  - 12 categories of formulations
  - 1 category of food
- 126 confirmed chemicals
- 1,506 unconfirmed chemicals



Application to Non-targeted Analysis

- ~1,400 chemicals identified in analysis were not in CPCPdb – EPA database listing chemicals in consumer products
- Using functional use majority of chemicals in formulation can be rationalized
- Still missing information on articles



Conclusions

- Chemical function is a useful descriptor for chemical prioritization
- Valid models can predict 41 functional uses
- Models can rapidly screen large libraries of chemicals