

Exploring the application of informatics tools to visualize the chemical space of compounds released from medical devices

Candice J. Gordon (BS student, CUNY-York College), Dr. Keaton Nahan (CDRH/OSEL), Dr. Diego Rua (CDRH/OSEL), Dr. Robert M. Elder (CDRH/OSEL)

Introduction

- Medical devices contain many different materials and chemicals, collectively termed a *chemical space*
 - polymers, metals, ceramics, bio-based, ...
 - additives: dyes, antioxidants, plasticizers, ...
 - contaminants: heavy metals, processing aids, ...
- Device safety can be assessed using *in vitro* solvent extraction to determine what compounds may be released *in vivo* (extractables & leachables)
- Regulatory submissions often report the extractables, but not in a standardized, machine-readable form
- A database of these extractables would preserve institutional memory and streamline regulatory review
- Does the extractables space present unique challenges compared to other regulatory chemical spaces?

Methods

- Developed an app to text-mine regulatory submissions for CAS numbers of extractables (~600 compounds)

Fig 1: Example of CAS numbers identified and highlighted in a PDF

ChemScan: Detect extractables in submissions	
Select a file: <input type="text" value="Browse..."/> Chem_Char_Edwards_Quality_Labs.pdf	Glycerin (CAS# 56-81-5)
First page: <input type="text" value="1"/>	Diethoxydimethyl-silane (CAS# 78-62-6); Dimethylsil; Dimethylsilanediol (1066-42-8).....
Last page: <input type="text" value="18"/>	Oxiranemethanol (CAS# 57044-25-4).....
<input type="button" value="Run ChemScan"/>	Methoxy-phenyl-oxime (CAS# N/A)
	Ethylene Glycol aka 1,2-Ethanediol (CAS# 107-21-1) ..

- Data-mined >850,000 molecular structures from several regulatory databases (foods, drugs, environmental, etc.)
- Represented structures as numerical descriptors using cheminformatics (>200 variables)
- Visualized the chemical spaces via dimensionality reduction (using t-SNE)

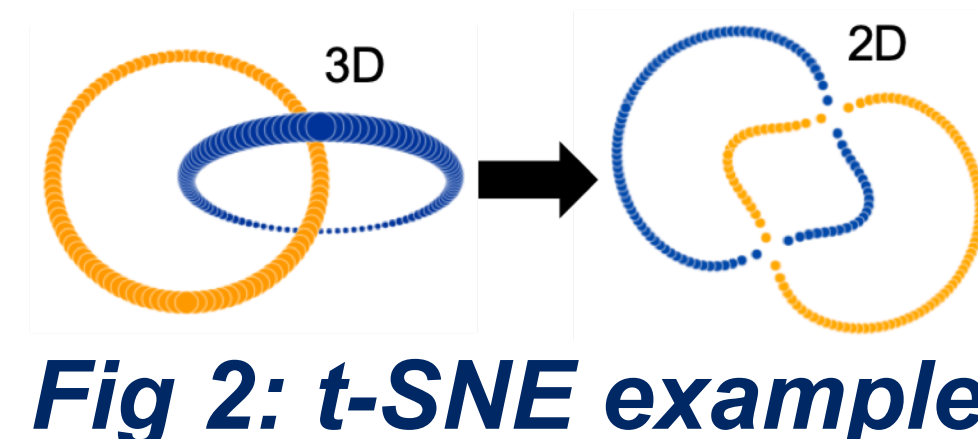


Fig 2: t-SNE example

Results

Fig 3: The probability density of compounds in reduced dimensions produces a “map” of chemical space. Stars mark the center of each distribution.

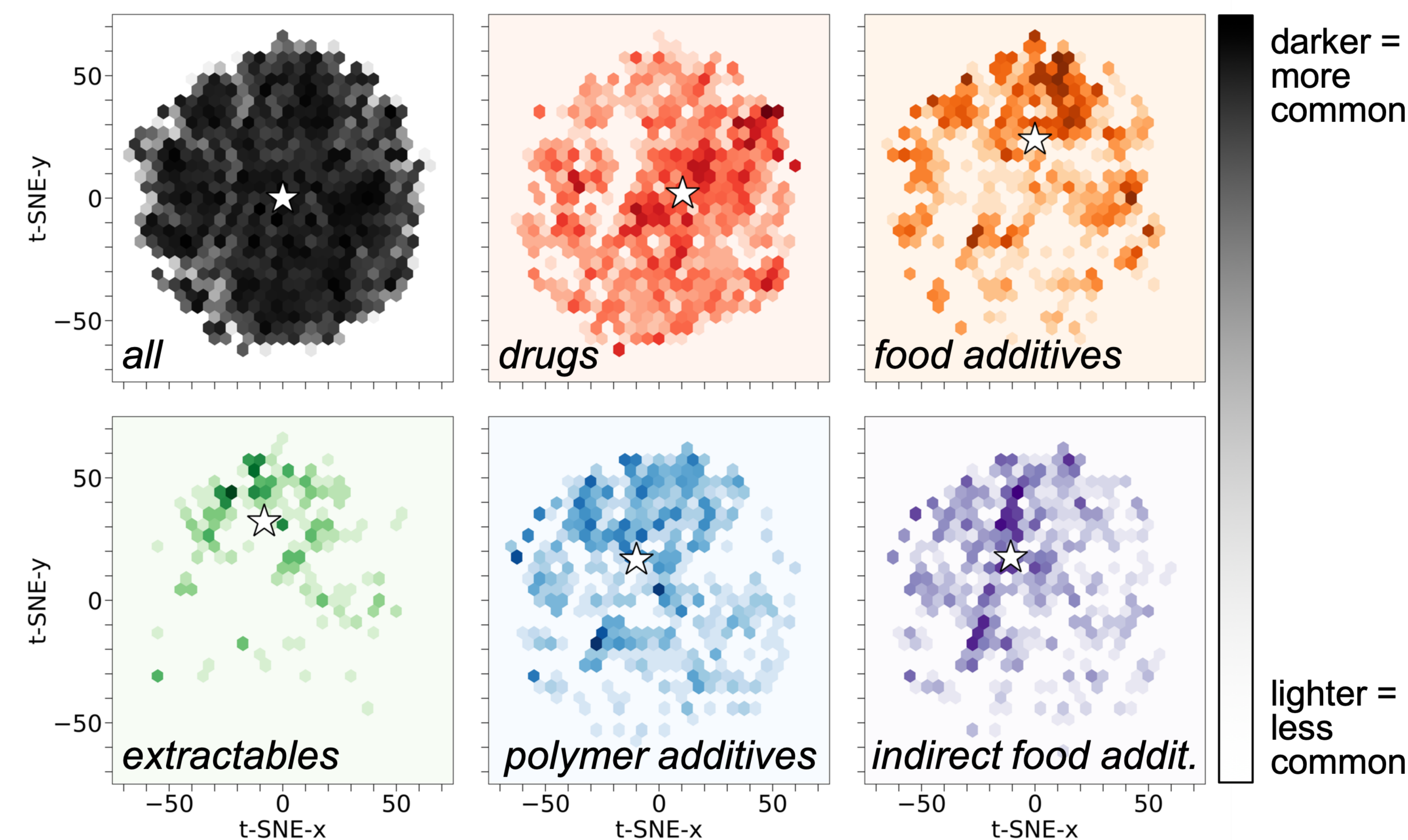
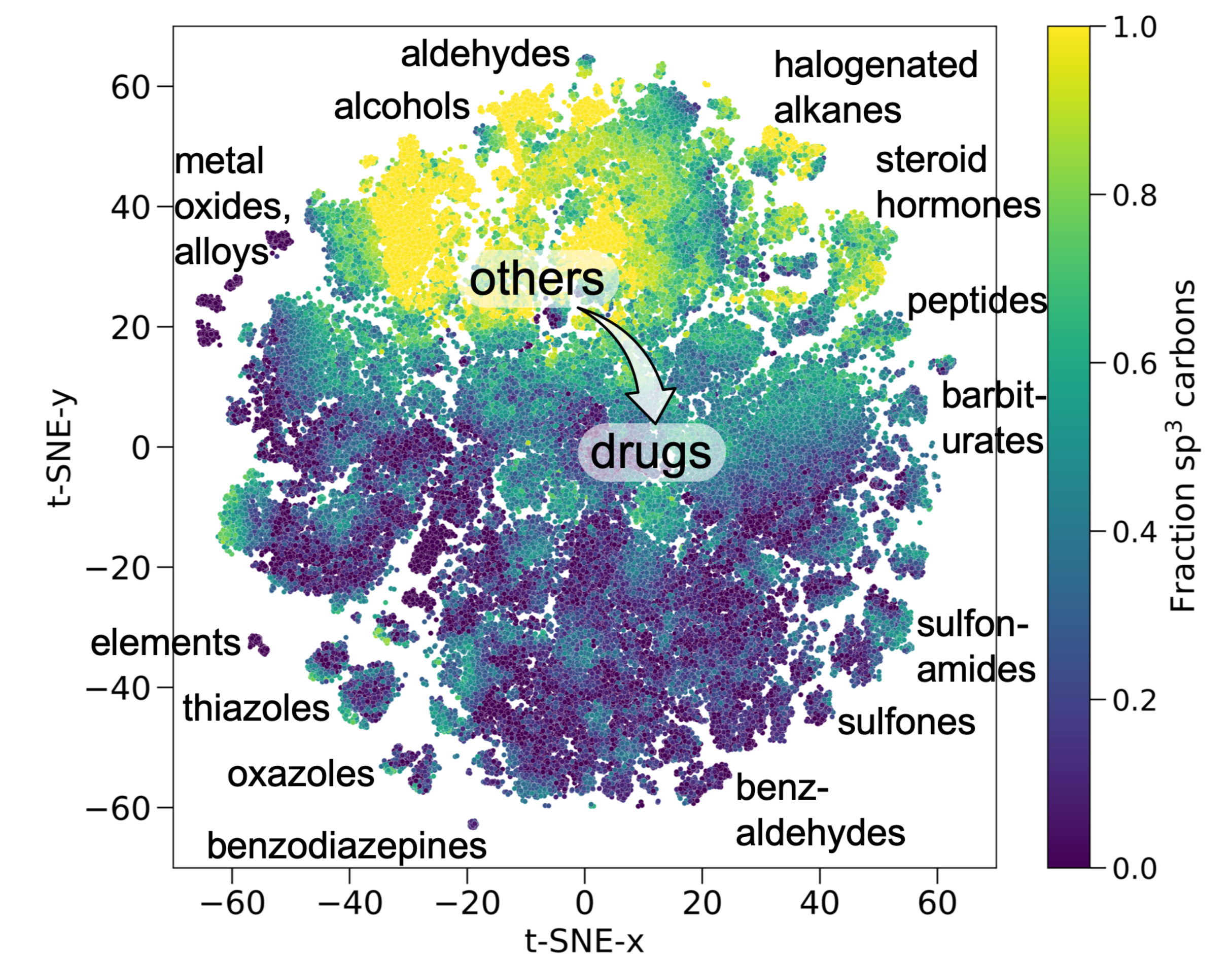


Fig 4: Scatter plot of all compounds, colored by the fraction of sp³ hybridized carbons. Several regions are labeled to indicate the types of chemicals they contain.



Conclusions & Future Work

- Extractables overlap with polymer additives, as expected
- The drug space is largely separate from other regulatory spaces, in part because drugs have fewer sp³ carbons
- Text-mining for CAS numbers appears to work well, but we plan to conduct thorough validation, increase the scale of data collection, and build a database
- Possible future work: link with computational toxicology tools to alert reviewers to compounds of potential concern

Acknowledgements: CJG was supported by funding from the U.S. Department of Energy Oak Ridge Institute for Science and Education (ORISE). We thank Dr. K. Myers, Ms. A. Clingerman, and Dr. E. Margerrison for their help making CJG's internship possible. We also thank Dr. J. Park for providing polymer additive data, and Drs. K. Vorvolakos and P. Goering for managerial support.

Disclaimer: The findings and conclusions in this presentation have not been formally disseminated by the Food and Drug Administration and should not be construed to represent any agency determination or policy.

Data mining can automate the collection of useful information from the FDA's vast trove of regulatory submissions, which will streamline the review process and preserve institutional memory.