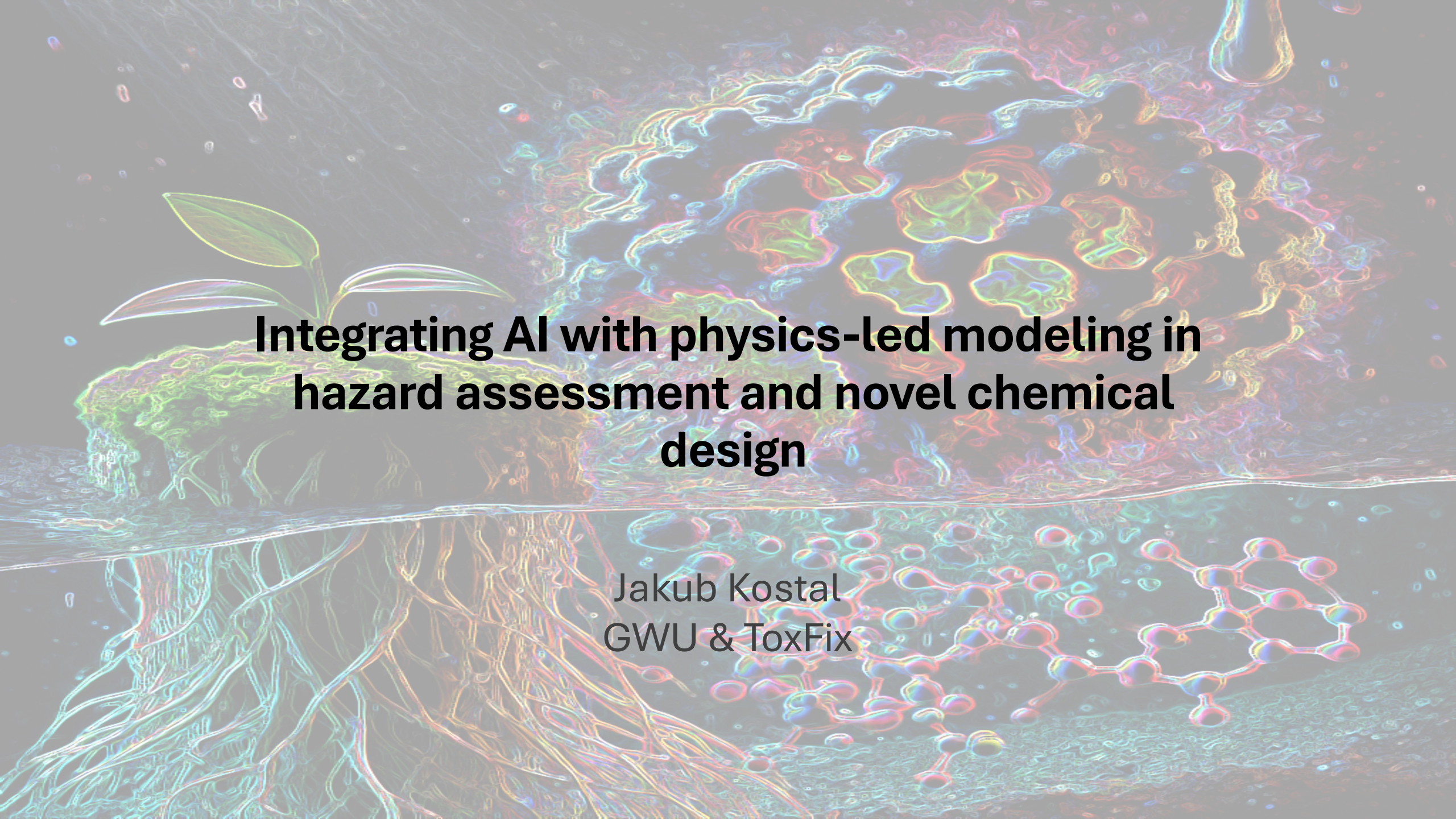
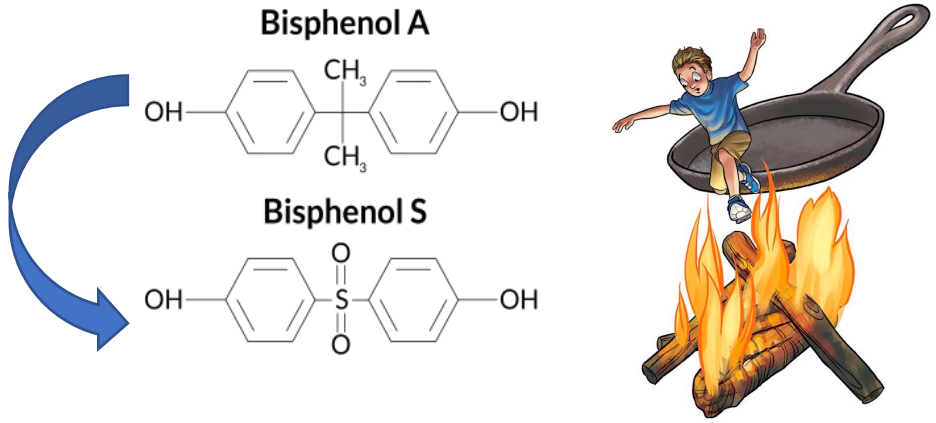


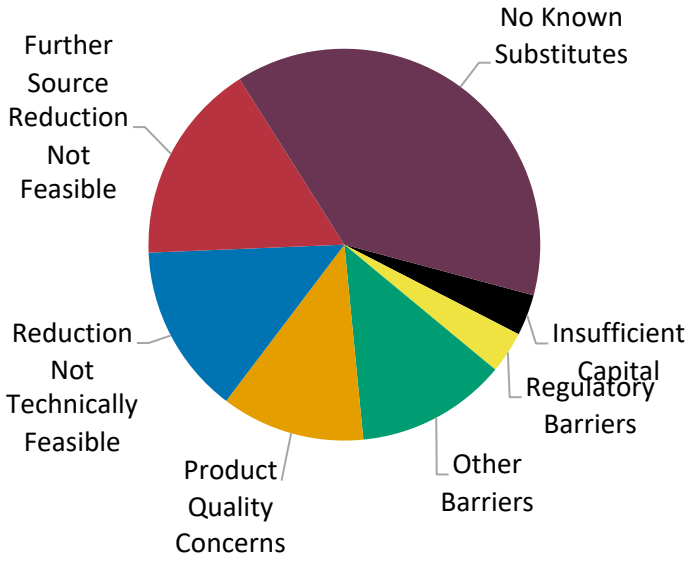
Integrating AI with physics-led modeling in hazard assessment and novel chemical design

Jakub Kostal
GWU & ToxFix

The need for *in silico* design tools in chemical discovery

- Since 1950: 140,000+ new industrial chemicals
  → ca. 5,000 are high-volume substances
- 300,000+ chemicals and chemical mixtures registered globally for commercial use
- ca. 85% lack human health safety data

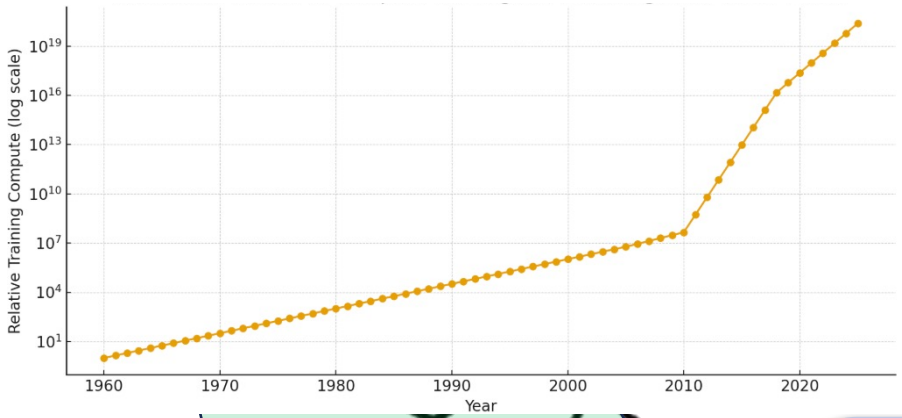


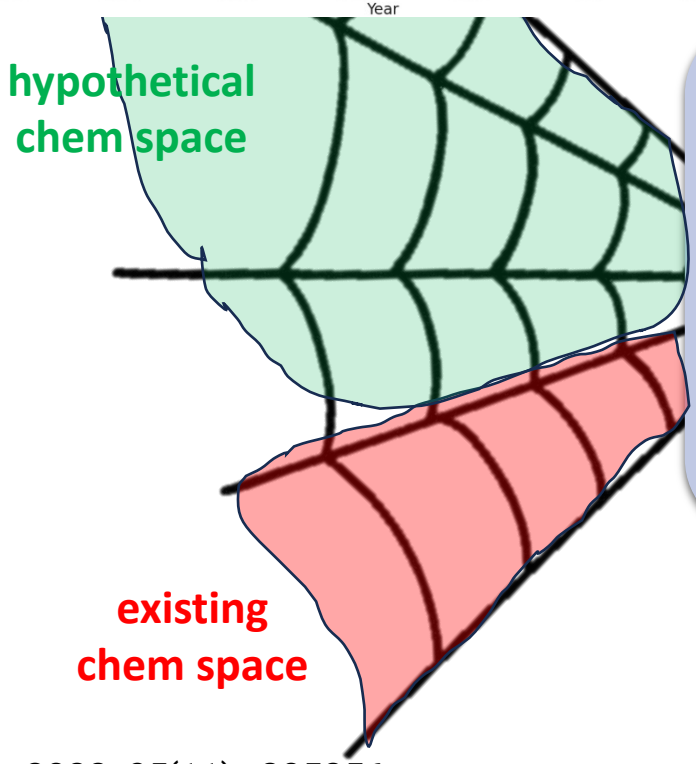


Need multicriteria *in silico* design in the upstream to mitigate high cost of tox testing in the downstream!

iScience 2022, *25*(11), 205256
Chem Res Toxicol 2020, *33*(40), 880-888

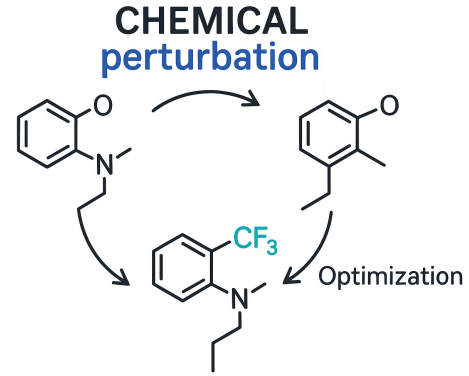
What is *in silico* safer chemical design?

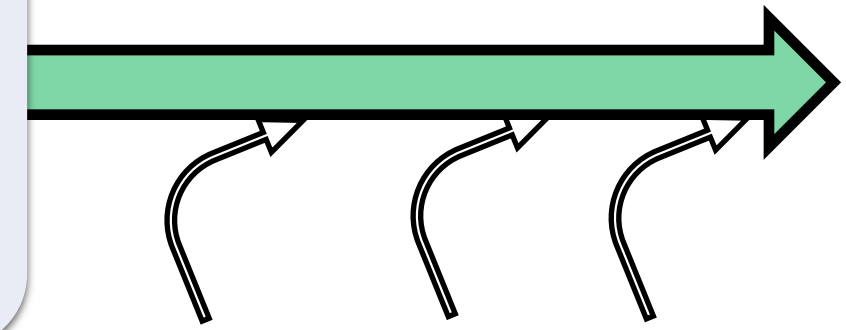




Irrational brute-force assessment of all plausible chemical solutions

Rational development of novel chemicals to achieve a desired outcome





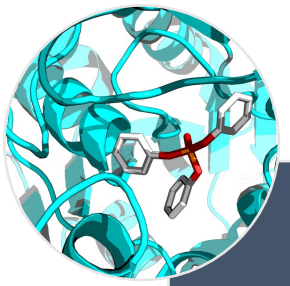
Performance + safety + depletion

iScience 2022, *25*(11), 205256
Chem Res Toxicol 2020, *33*(40), 880-888

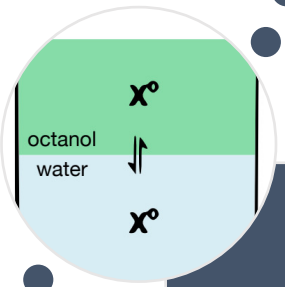
CADD

ML → generative AI

2000s



Simulations of small to complex systems in condensed phases



1960s

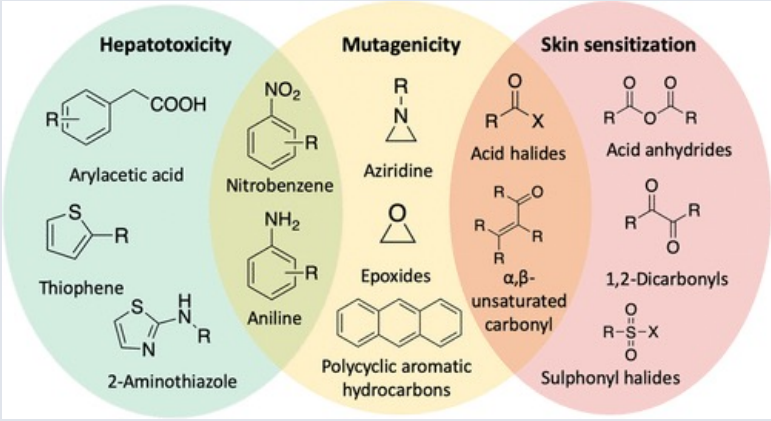
QSARs based on physchem properties

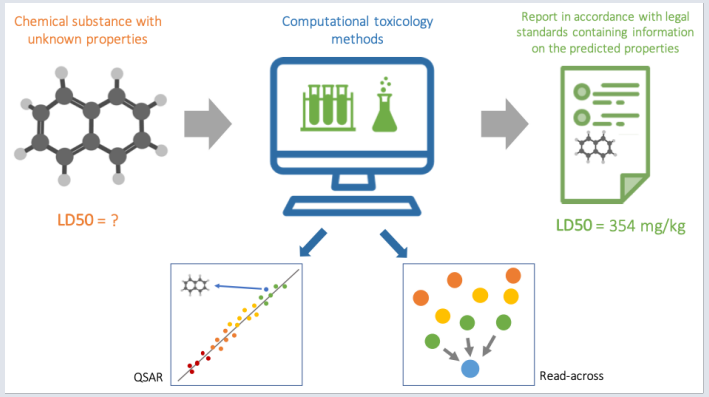
Expert rules using substructural alerts etc.

Exponential increase in compute

Pred Tox

Alerts, QSARs and read-across based on atomistic structural representation





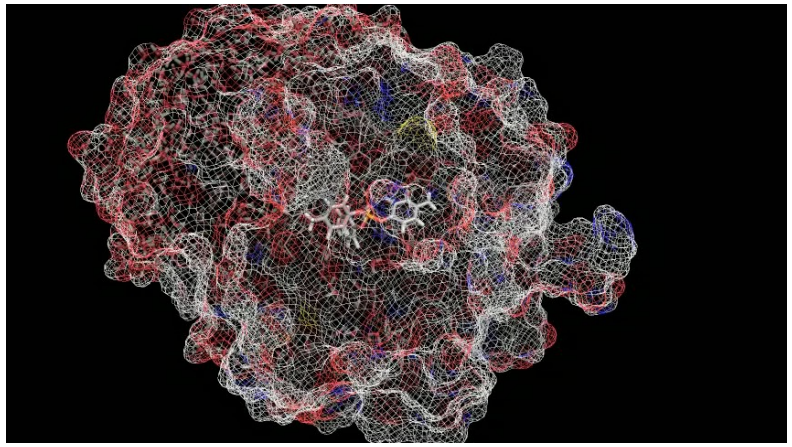
Chem. Res. Toxicol. **2020**, 4, 880–888

Roadblock #1: ‘Proving’ a negative in safer chemical design

https://www.roche.com/about/sustainability/philanthropy/science-education/biochemical-pathways/

But what to do mechanistic complexity, uncertainty and data variability??

Balancing highly specific molecular transformations with general behavior across toxic endpoints



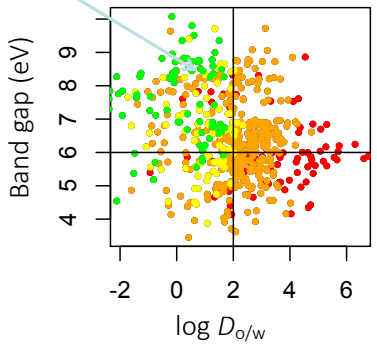
Modeling of key events in toxicity pathways

Trade-off: loss of resolution in model predictivity (but NOT robustness)

General reactivity & bioavailability metrics

Perfectly OK when data variability limits greater precision!

"safer chemical space": log$D_{o/w}$<1.7, ΔE>6 eV



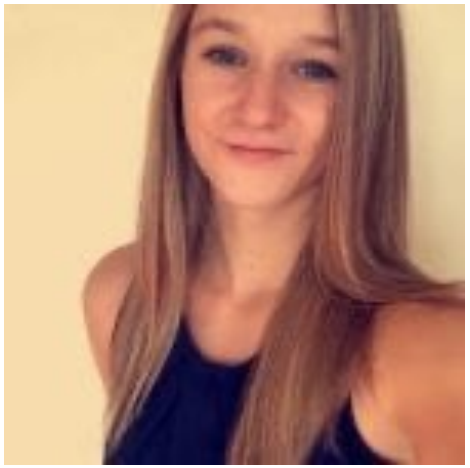
Chem Res Toxicol 2023, *36*(9), 1444-1450.

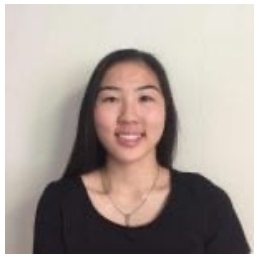
Roadblock #2: improved (non–specific) safety may decrease performance/degradation

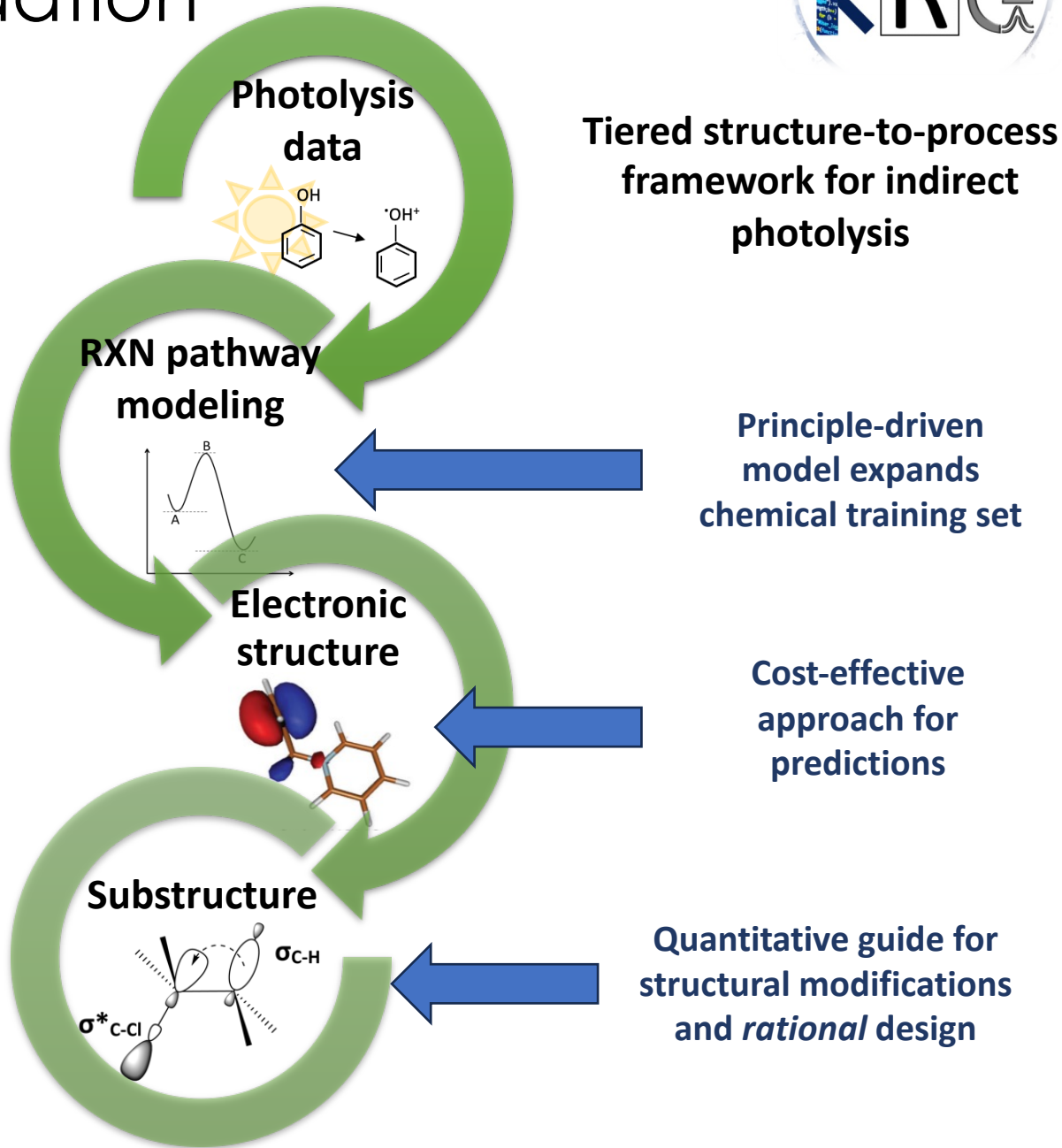
- Need systems thinking (multicriteria analysis)
- Need a way to quantify and optimize trade-offs

Case study in design: controlling safety and depletion of pesticides

- Indirect photolysis and ecotoxicity (selective vs. unintended)







Tiered structure-to-process framework for indirect photolysis

Principle-driven model expands chemical training set

Cost-effective approach for predictions

Quantitative guide for structural modifications and *rational* design

Lewer, J et al. *Environ Sci Technol* **2021,** 55(17), 11713-11722
Lewer, J et al. *Science Advances* **2022**, *8*(13), eabn2058

Computational framework to *irrationally* screen or *rationally* redesign pesticides

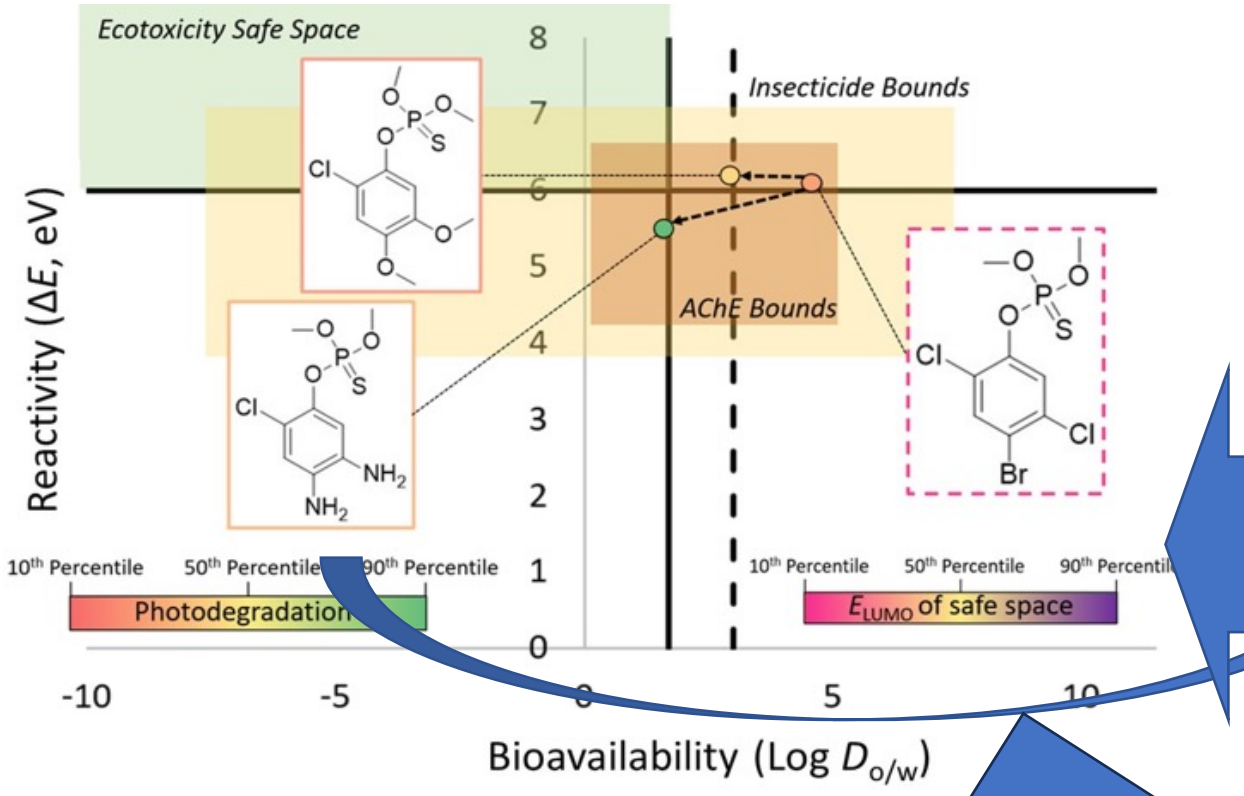
Structure-to-process *in silico* framework

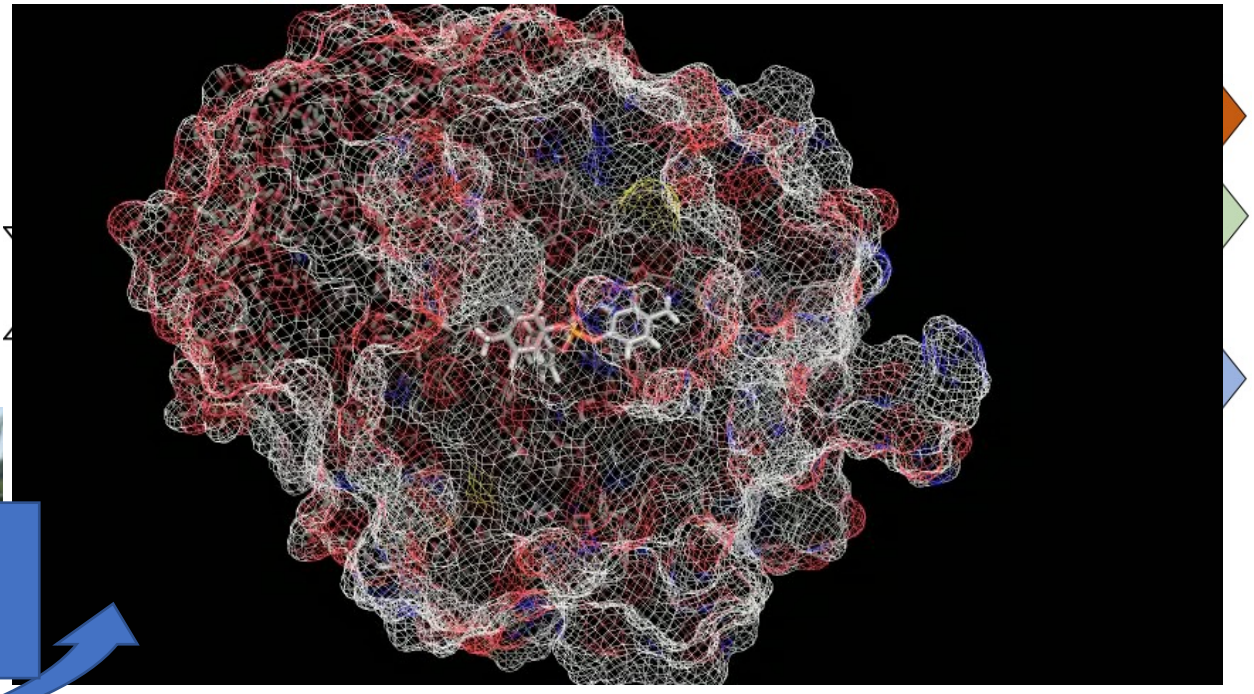
Accessible free of charge at: https://kostal.columbian.gwu.edu/software/

AA database

Design-vectoring blueprint

Redesign of existing chemicals





Orthogonalization of selective and unintended toxicity

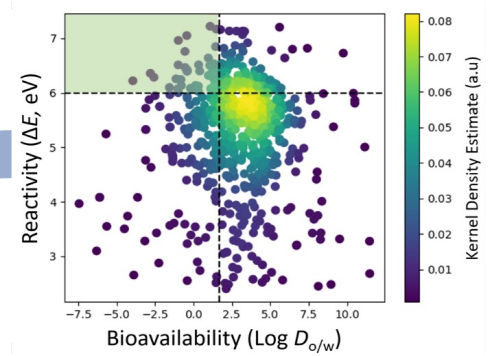
Clymer T et al. *Green Chem.* **2019**, *21*(8), 1935-1946

Kostal J et al. *PNAS* **2015**, *112*, 6289-6294

Melnikov F et al *Green Chem.* **2016**, *18*, 4432-4445

Roadblock #3: if we cannot reconcile trade-offs → broaden scope to orthogonalize design vectors

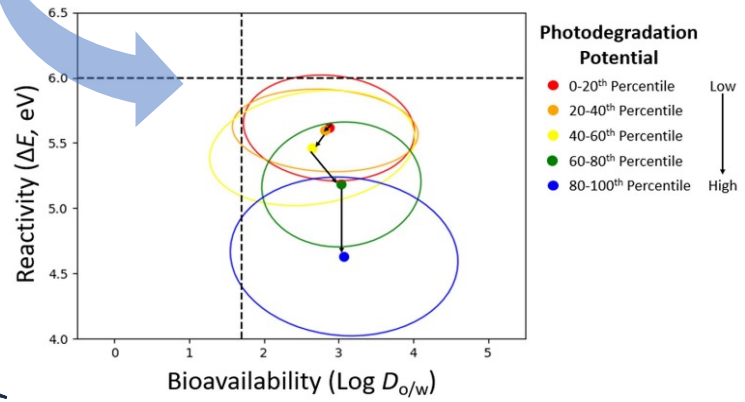
- Indirect photodegradation and toxicity partially overlap in electronic requirements
- Can we further orthogonalize the depletion and hazard/function design vectors?

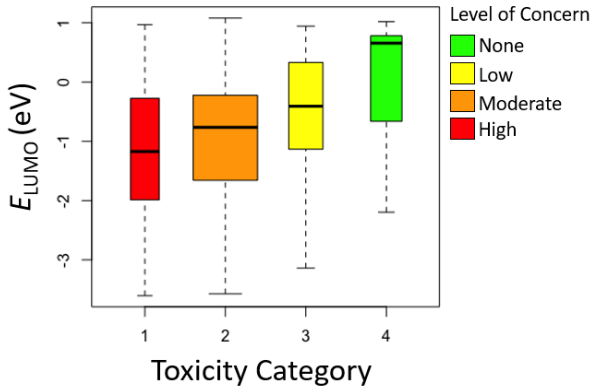


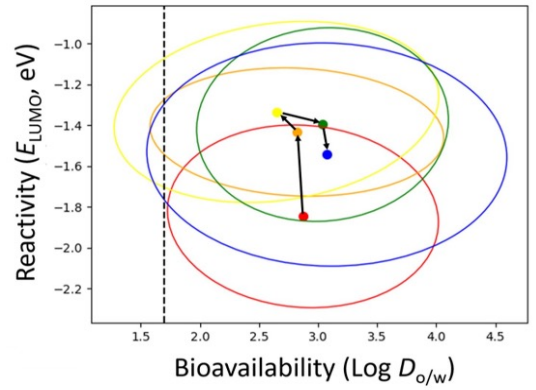
Integrate biodegradation:

- Driven by reactivity but also target-binding requirements, which are orthogonal to reactivity!

Using different orbitals for toxicity and for photodegradation 'uncouples' these design vectors

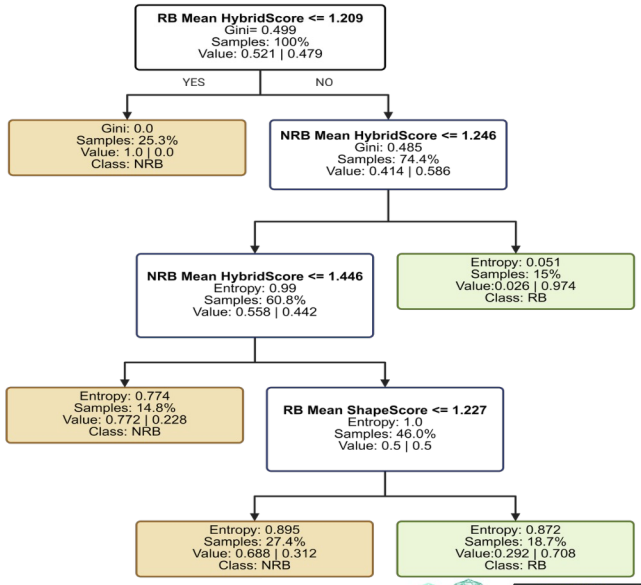


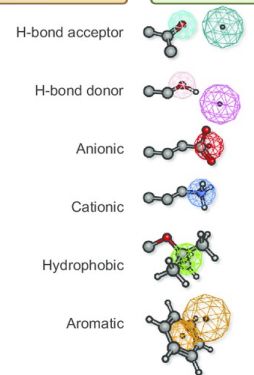


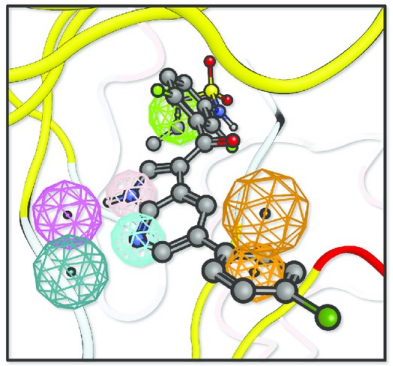


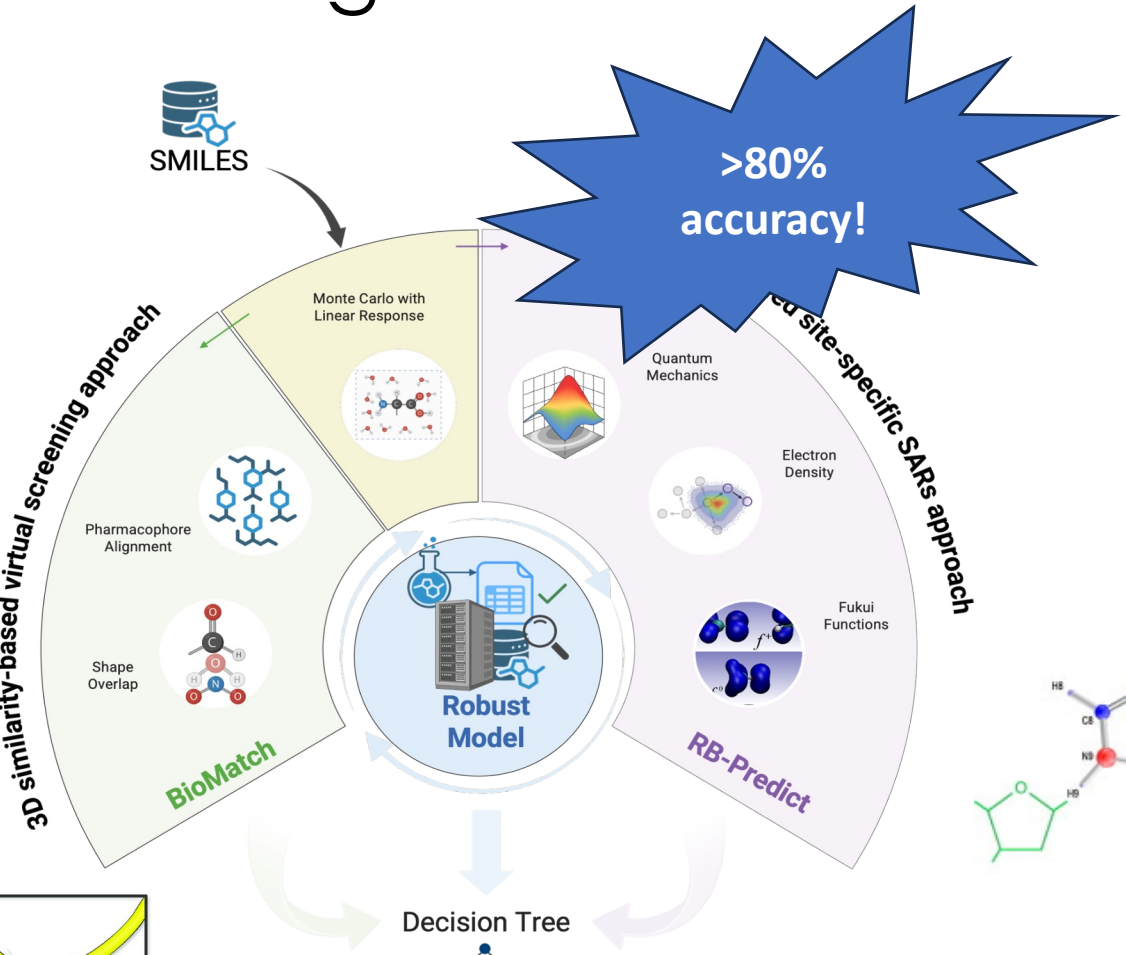
Neural-networks and gradient-boosting methods to develop physics-led biodegradation models

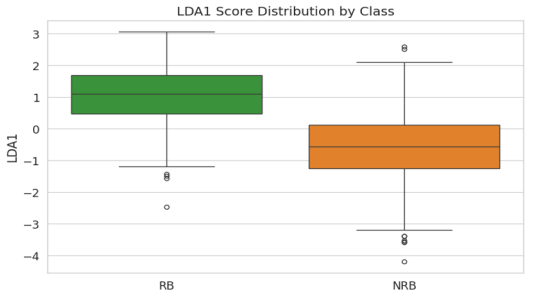
Adding biodegradation by incorporating an in-house developed model:

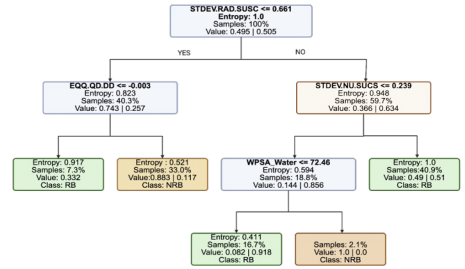


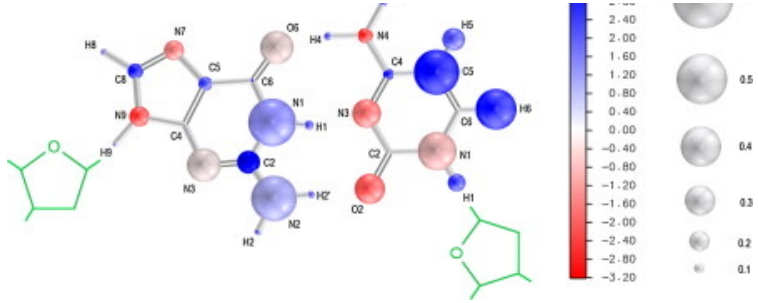












UNIVERSITY OF CAMBRIDGE

ACS GCI Pharmaceutical Roundtable

CDI Chemical Data Intelligence

If we fail to adequately reconcile trade-offs
→ need to discover/design ***brand*** new molecules

Orthogonalization fails due to modeling constraints or because of overlapping requirements

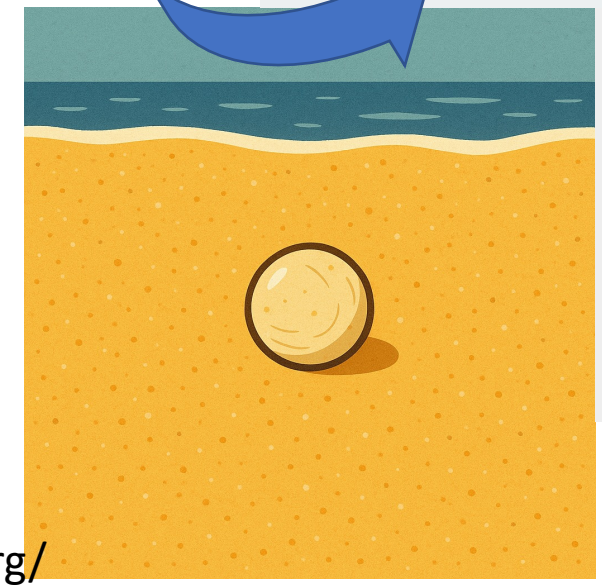
EPA Comptox: **~ 10^3 AIs**

EPAPCS	PESTICIDES|EPA: Pesticide Chemical S...	4038	2017-11-07

The entries in this list have been classified in the U.S. as pesticidal "active ingredients" (conventional, antimicrobial, or biopesticidal agents), and were sourced from the Pesticide Chemical Search database (https://iaspub.epa.gov/apex/pesticides/f?p=chemicalsearch:1) created by EPA's Office of Pesticide Programs.

Chemical Search provides a single point of reference for easy access to information previously published in a variety of locations, including various EPA web pages and Regulations.gov. Chemical search contains the following:
1) More than 20,000 regulatory documents;
2) Links to over 800 dockets in Regulations.gov
3) Links to pesticide tolerance (or maximum residue levels) information;
4) A variety of web services providing easy access to other scientific and regulatory information on particular chemicals from other EPA programs and federal government sources.

It should be noted that the Pesticide Chemical Search site is not actively maintained and the various chemicals can be out of date in terms of status.



ZINC20

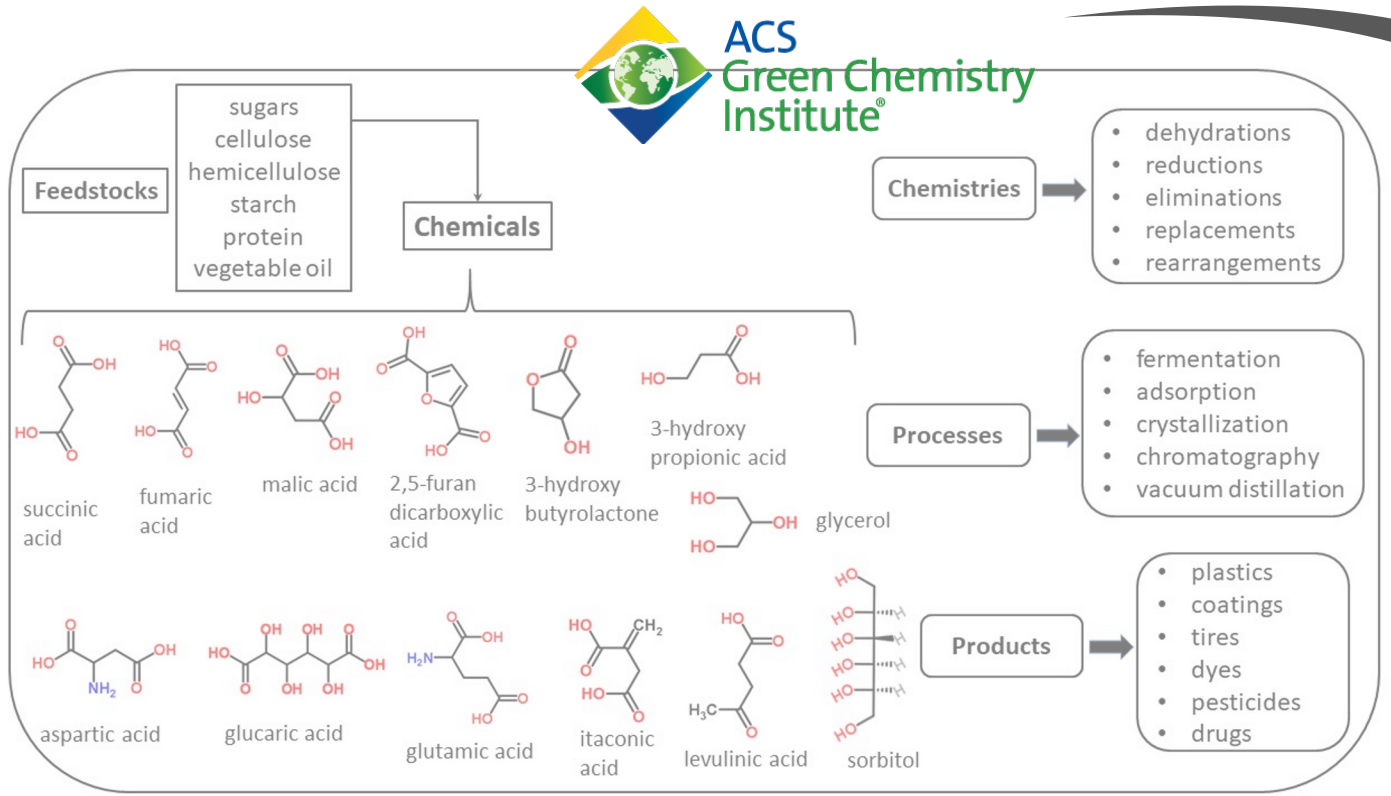
~ 10^8 AIs

Welcome to ZINC, a free database of commercially-available compounds for virtual screening. ZINC contains over 230 million purchasable compounds in ready-to-dock, 3D formats. ZINC also contains over 750 million purchasable compounds you can search for analogs in under a minute.

Available for innovation: ~ 10^{60}

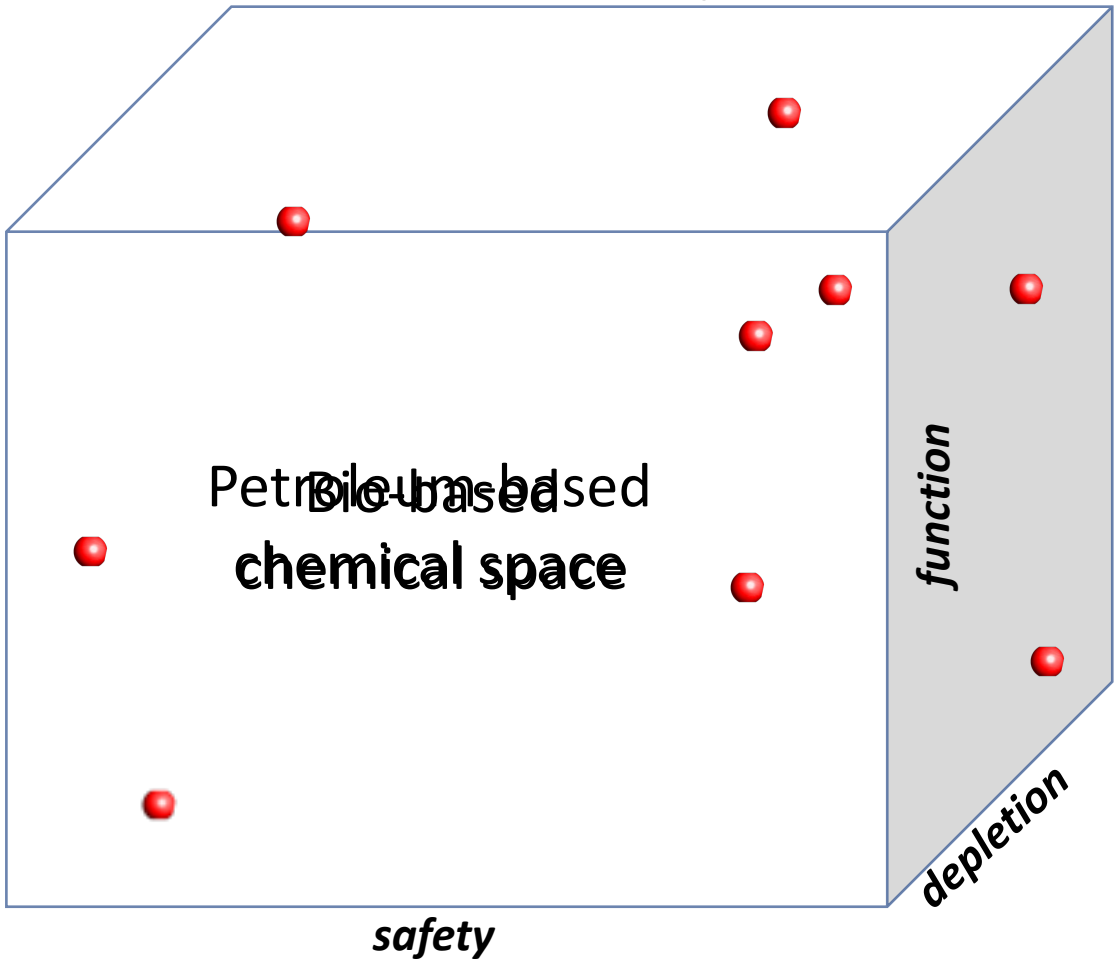
https://zinc.docking.org/

AI-enabled *irrational* design from renewables



Constable D *iScience,* 2021, *24*(12), 103489
Devineni, G. *ACS Books: Sustainable Agricultural Practices.* 2023, In press.

Problem dimensionality reduction: safety + depletion + function → ***function***



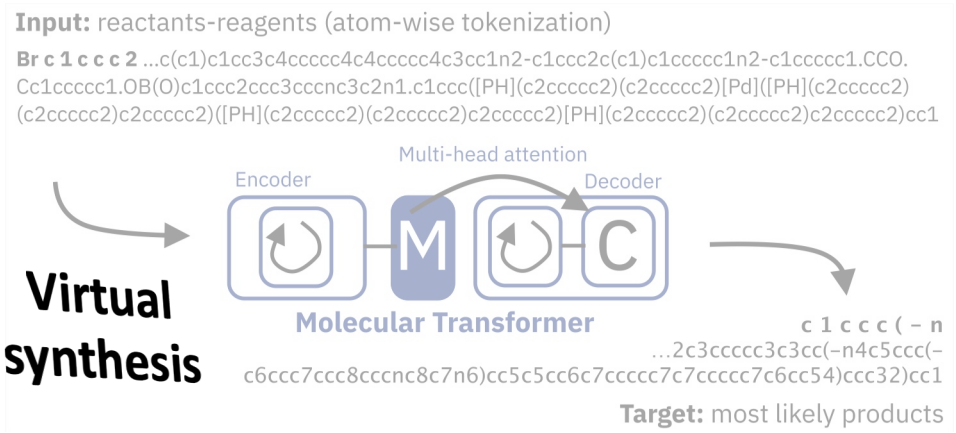
Throwing out the rulebook: de novo pesticide design

Kostal et al. ACS Zero Hunger Summit, 2023
Devineni, G. ACS Books: Sustainable Agricultural Practices. Vol 1449, Ch 2, 2023, 11-30.

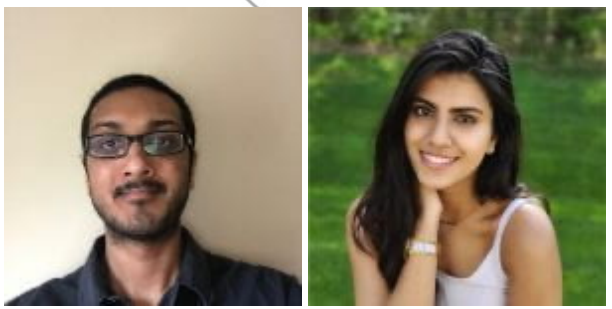
succinic acid
fumaric acid
malic acid
2,5-furan dicarboxylic acid
3-hydroxy
aspartic acid
glucaric acid
glutamic acid
levulinic acid

Renewable feedstocks

Gen AI

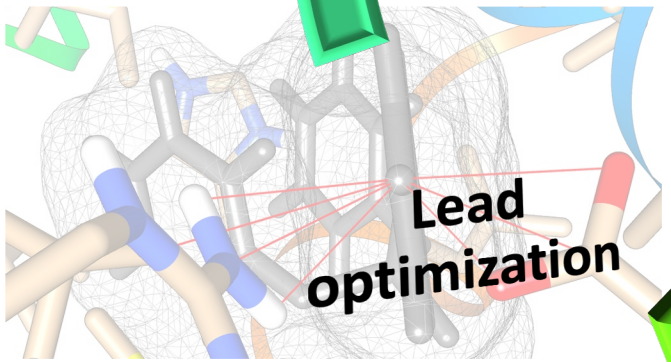


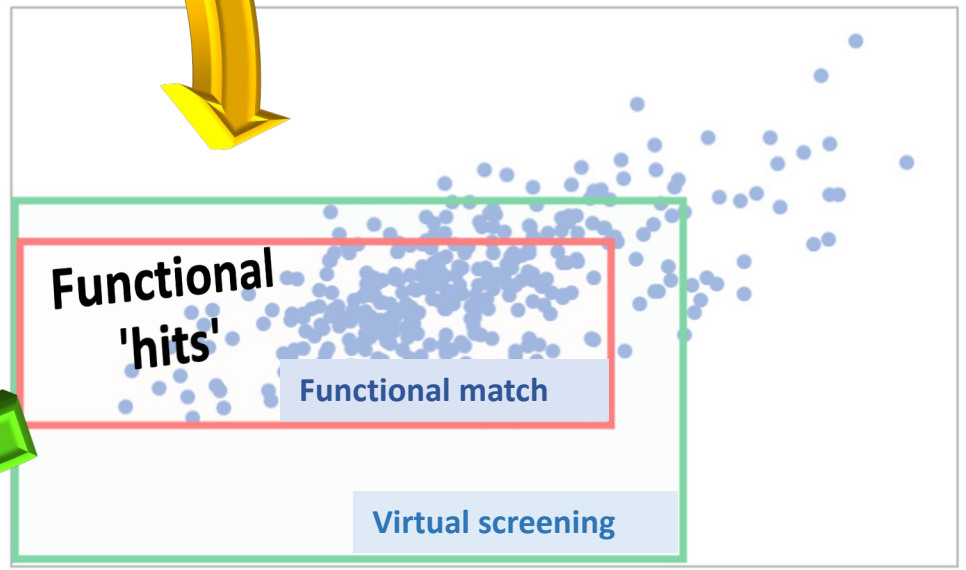
ACS Cent. Sci. 2019, 5, 1572−1583



Target activity:

new molecules in target active sites:
performance vs. undesired toxicity





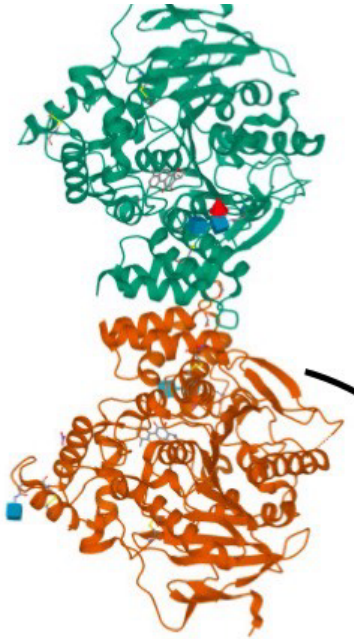
Mapping functional properties onto renewables:
Size, rigidity/flexibility, aqueous solubility + binding-site interactions

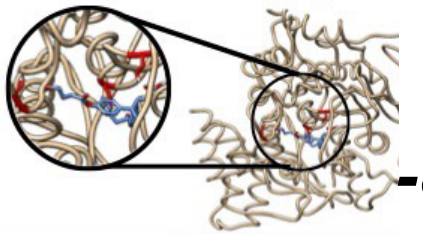
Case study: replacing OP pesticides

MW, Vol, globularity, log $P_{o/w}$

Property screen

Human





Binding affinity

-12

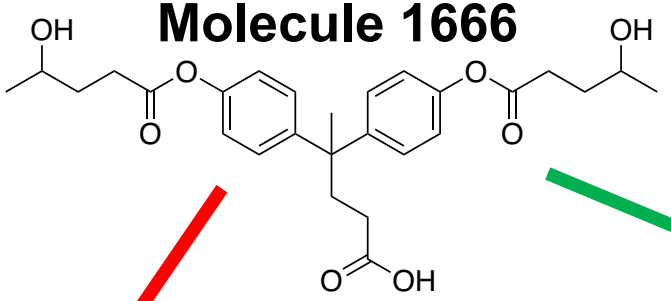
-11

-10

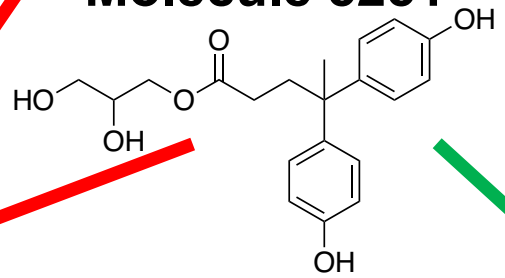
-9

-8

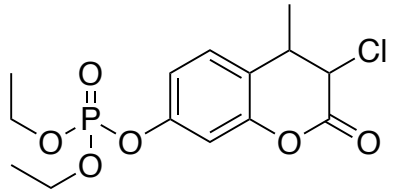
Molecule 1666



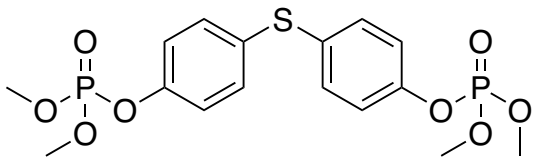
Molecule 5251



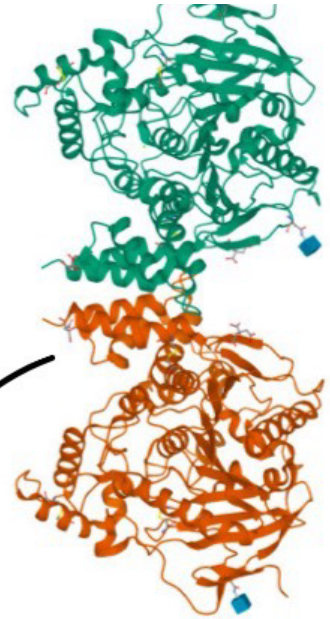
Coumaphos*

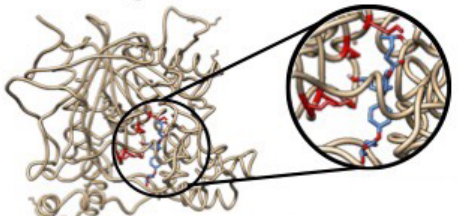


Temephos



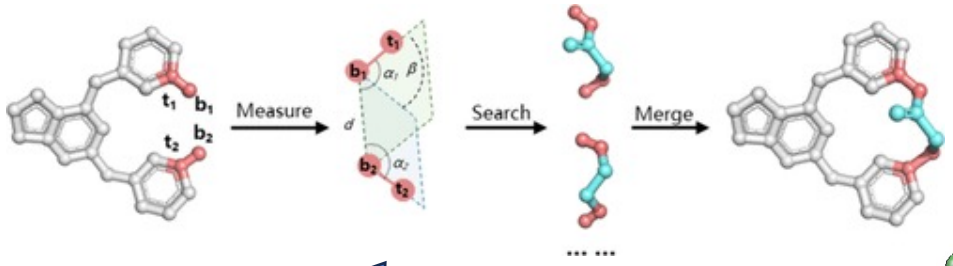
Mosquito



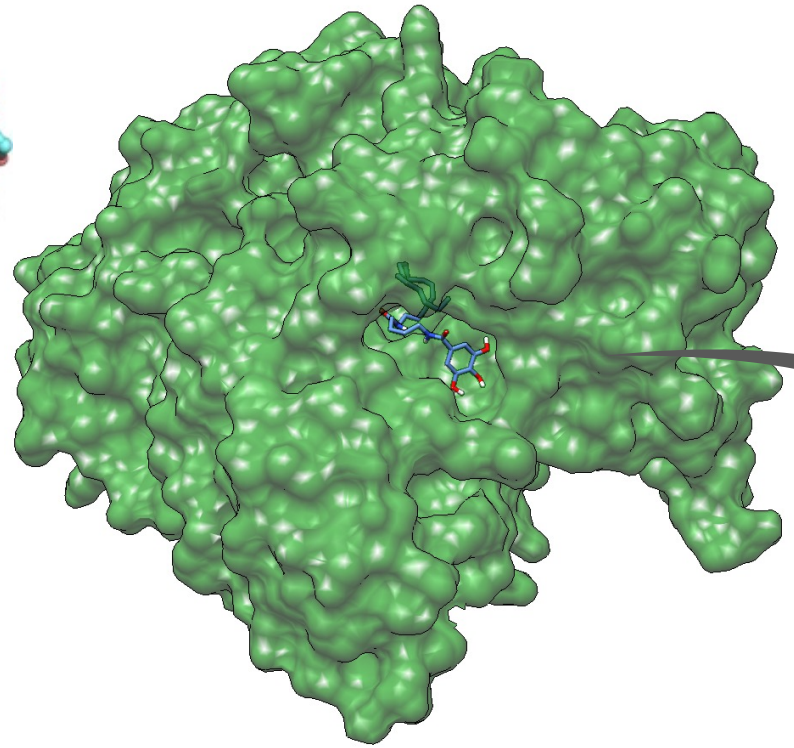


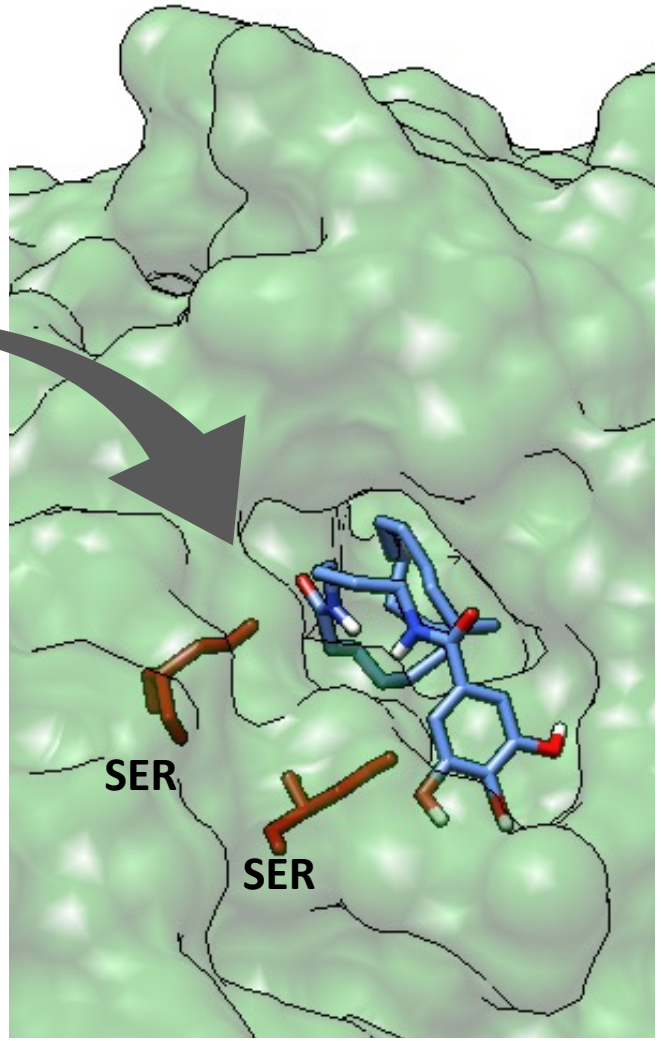
Going beyond molecular mimicry

- How do we compensate for a potential loss of covalent-binding strength?



Cyclizing biobased products to prepay ‘entropy’ and create an electrostatically stronger inhibitor





J. Chem. Inf. Model. 2024, 64, 24, 9048–9055

The Group

Current GS:

- Geetesh Devineni
- Diana Garnica Acevedo
- Jillian Brejnik
- Preston Griffin
- Ben Maybaum
- Prakruthi Koteshwar

Collaborators:

MoDRN, Yale, U Cambridge
The Brooks Group @ Baylor
US EPA, HESI, FDA, CDI

Support:

PhRMA FOUNDATION

NSF

NSF CAREER Award

HESI

GW

Find us in SEH 4400!

http://kostal.columbian.gwu.edu/

Educational:

WHEATON HIGH SCHOOL

beyondbenign
green chemistry education