DMSE DEPARTMENT OF MATERIALS SCIENCE & ENGINEERING

Massachusetts Institute of Technology

Inverse Design of Materials

through artificial intelligence and physics-based simulations

MIT ILP Japan Conference.
Tokyo January 25 2019

Rafael Gomez-Bombarelli
rafagb@mit.edu
http://gomezbombarelli.mit.edu/

Virtual discovery

Software is already matching (or beating) humans in performance and speed

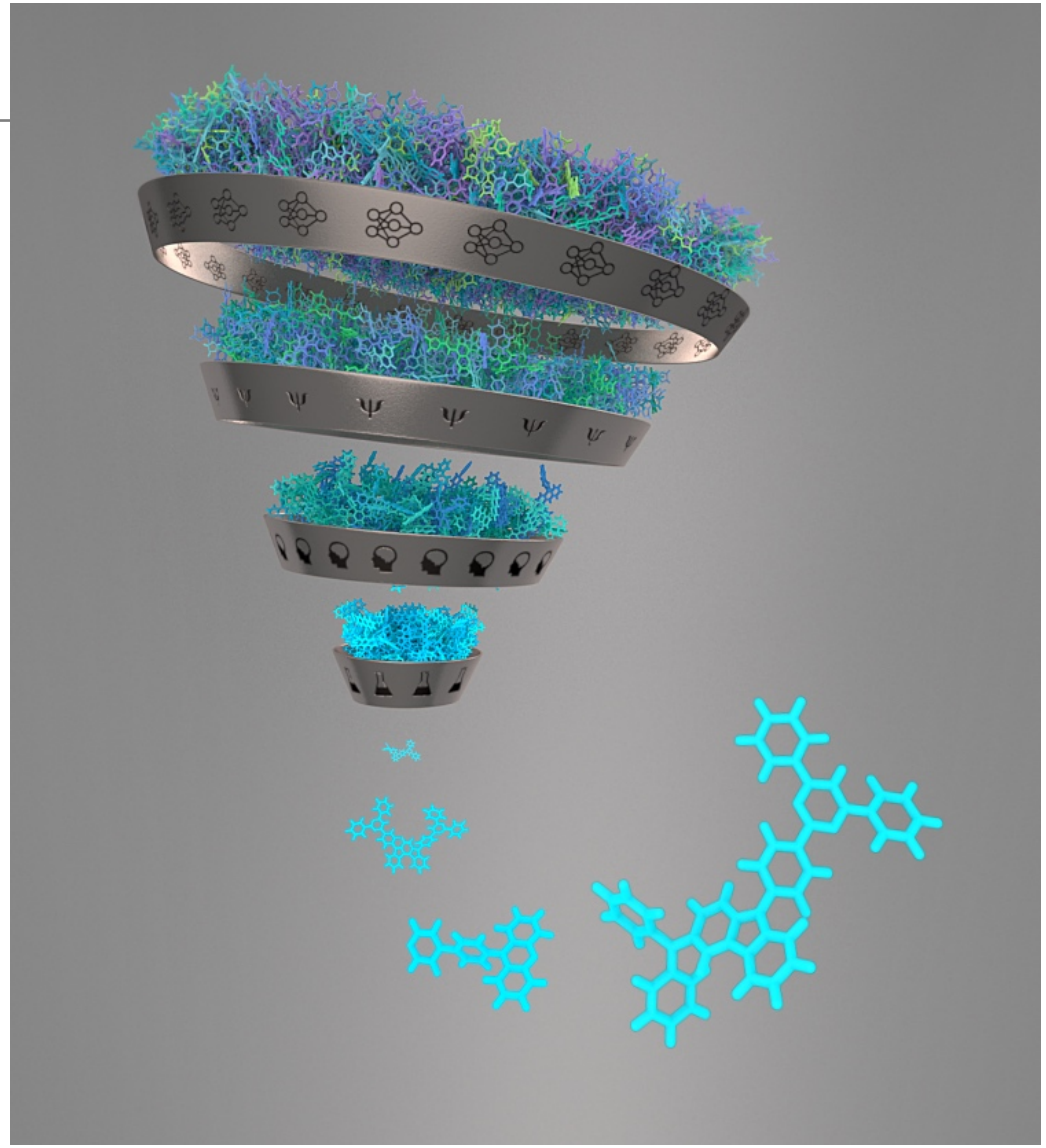
Driverless cars, AlphaGo, virtual assistants, speech & image recognition and generation, …

Increasing computing power

Faster algorithms

More data

Is it time for materials design?



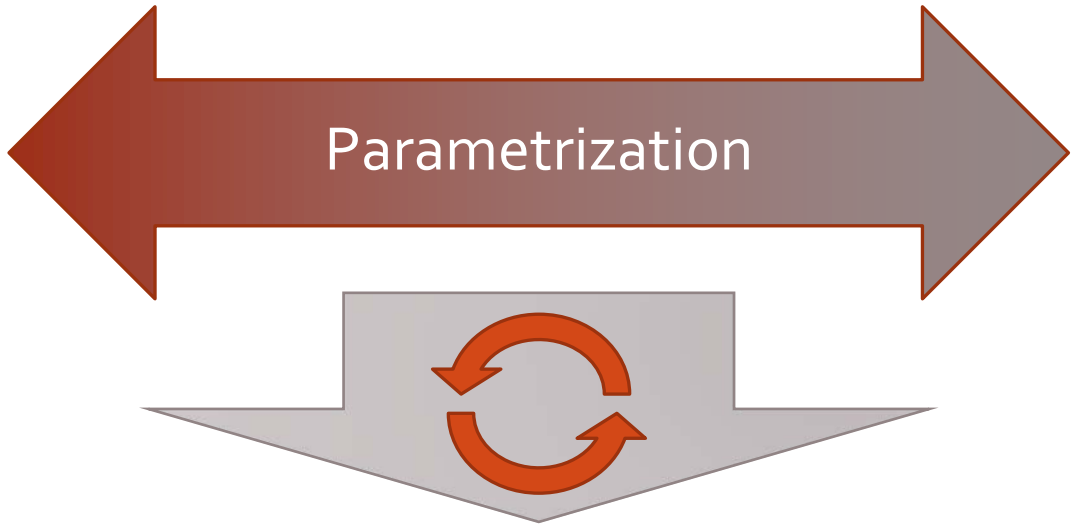
Computational spectrum - virtuous cycle

There is essentially a continuum of higher parametrization and statistical learning connecting first principles (theory-based simulations) to black-box statistical learning over experiments.

First Principles

Out-of-the-box
No/Little fitting
High-throughput

- *but* -
As good as model
Sometimes costly



Discovery

Machine learning

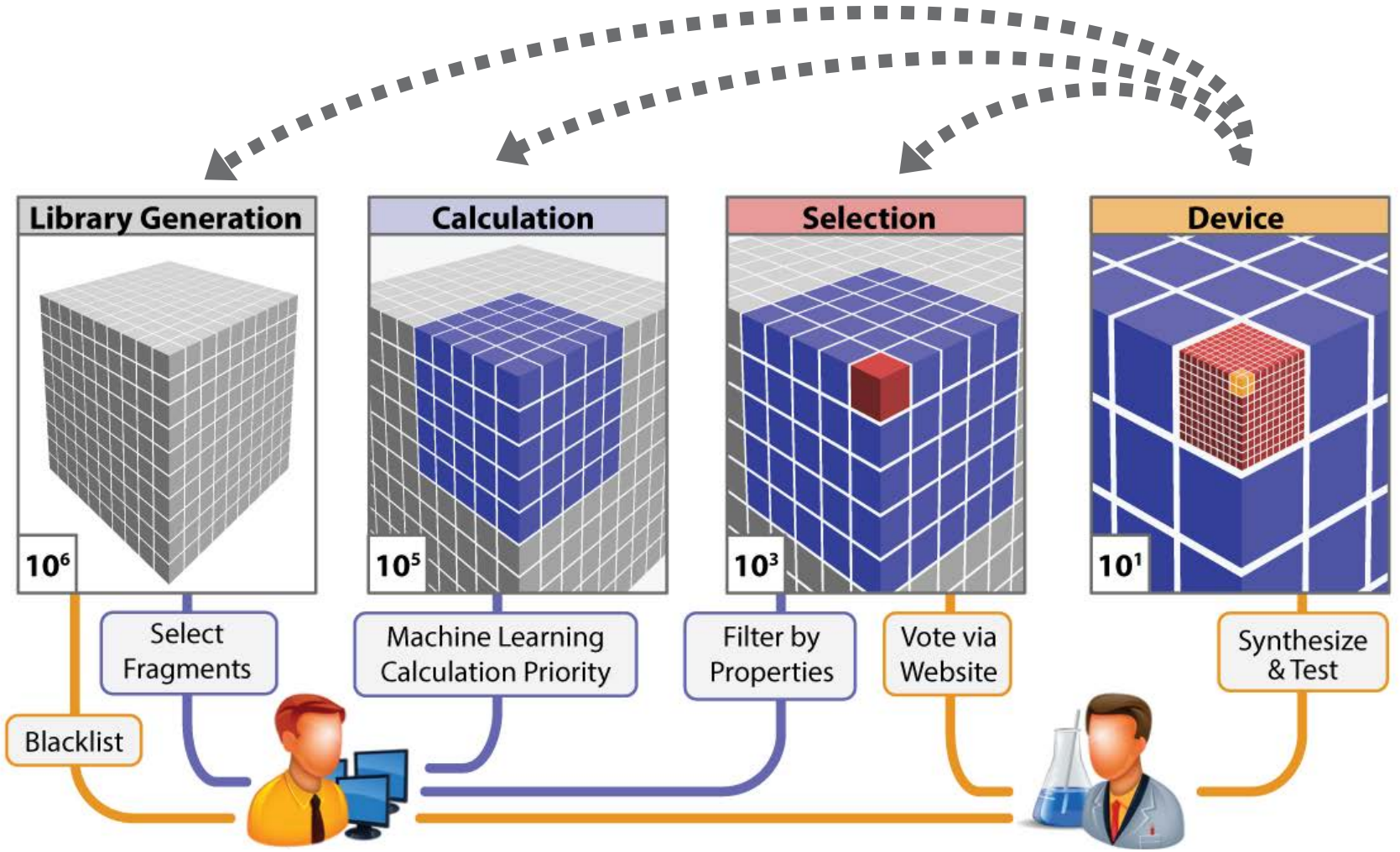
Fast
Uncanny performance
Leverage large data

- *but* -
As good as training data
Data ownership

Research background

1. High-throughput screening
2. Deep inverse design
3. Novel zeolite catalysts
4. Ongoing Research

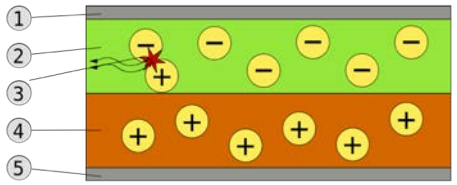
Virtual discovery approach

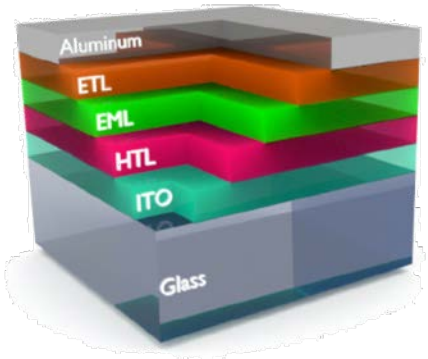


Organic light emitting diodes

Semiconductors made of molecules that can convert electricity into light.

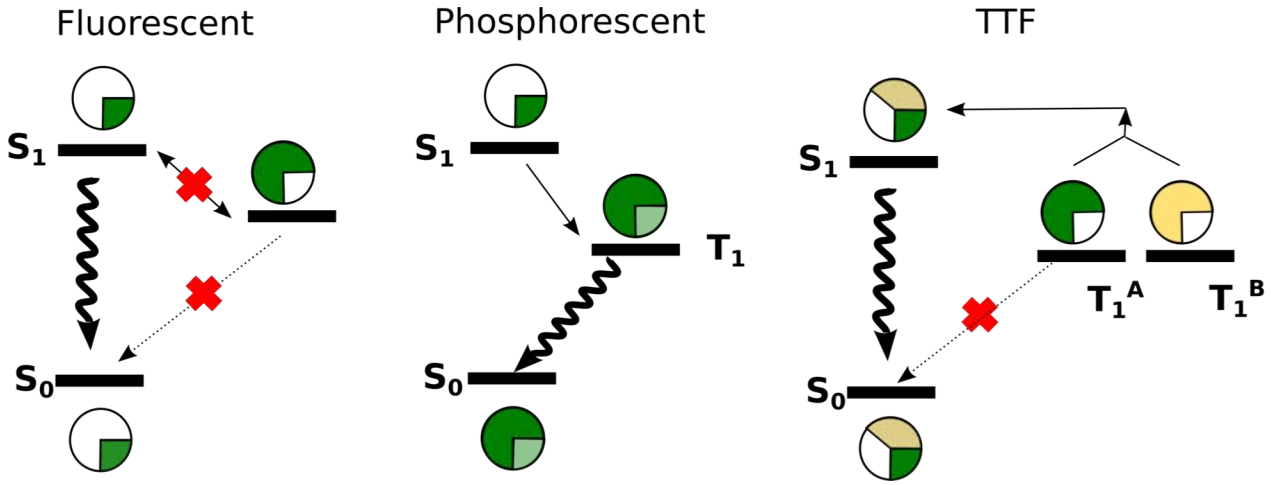
- High end displays, potentially lighting.
- Lightweight, flexible, transparent, high contrast, low power





Other OLED modes

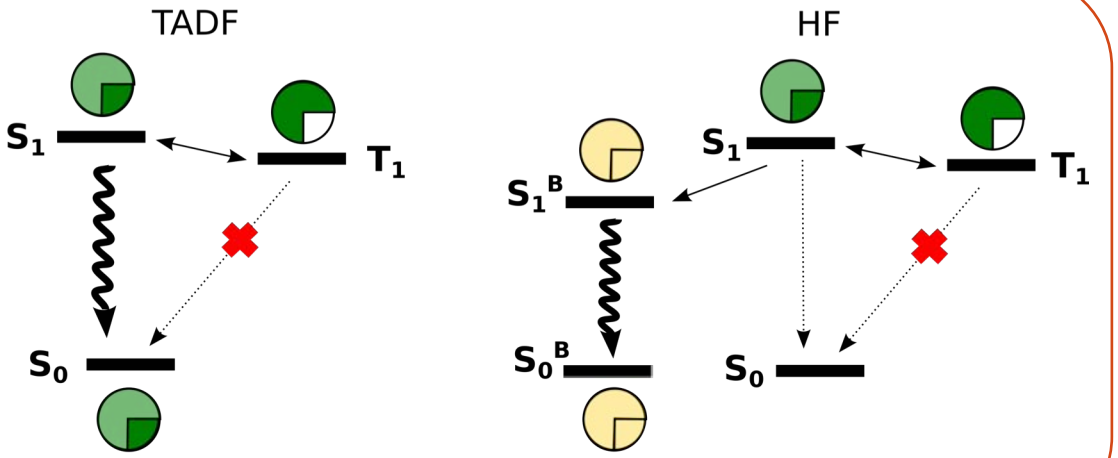
Well-known historical avenues



New TADF molecules can be efficient, cheap and potentially make blue.

We can calculate what makes them good with simulations

Uncharted territory



New physics from 2012-2014

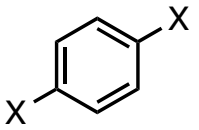
Adachi et al . *Nature* **2012**

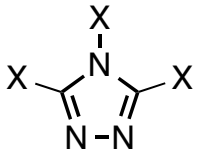
Making combinatorial libraries

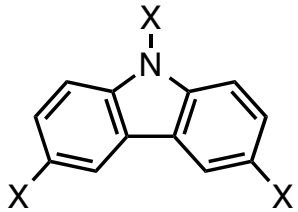
Inspired by lab chemistry

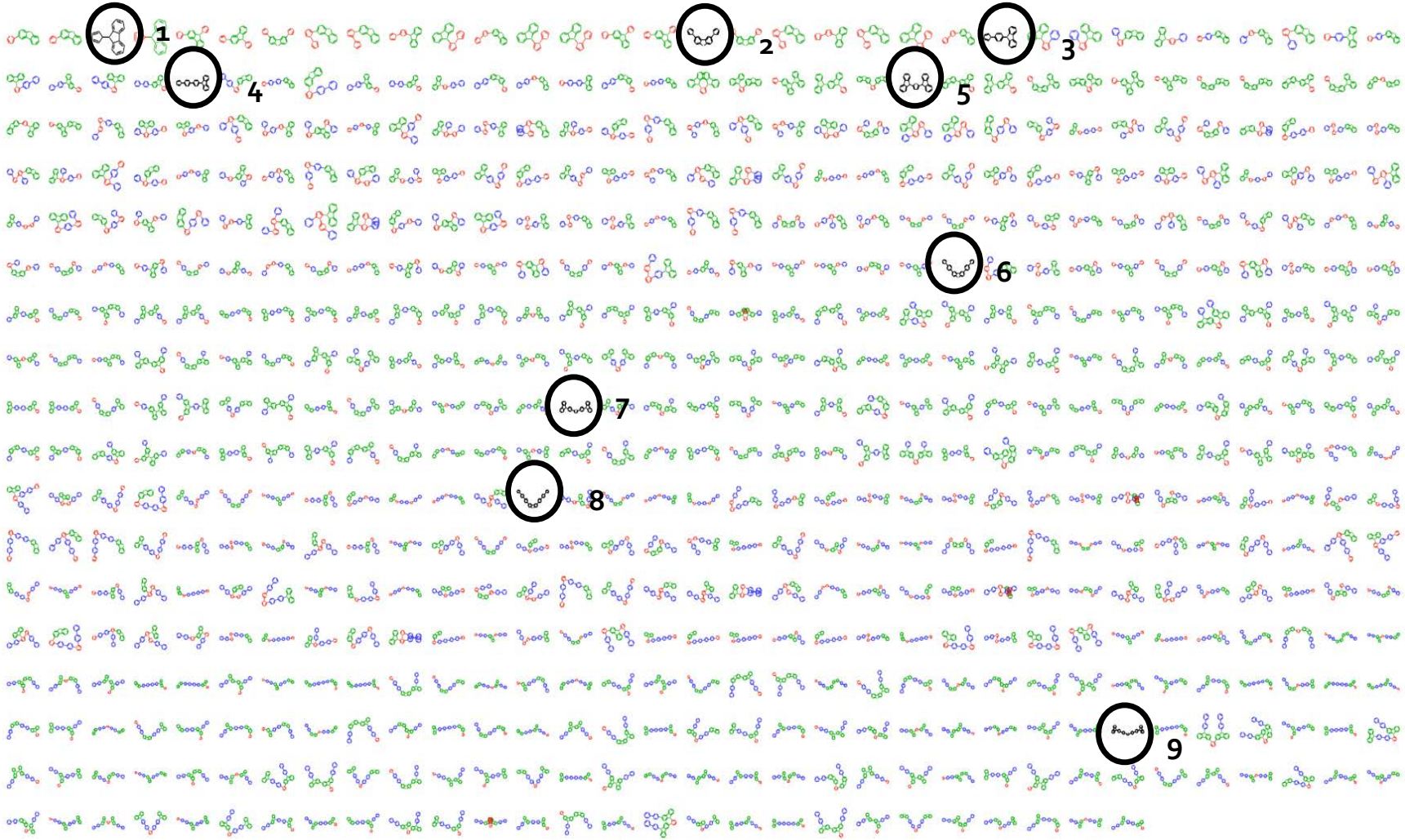
Reactants pulled from patents and literature

Connection modes that limit combinatorial explosion (cross-coupling chemistry)







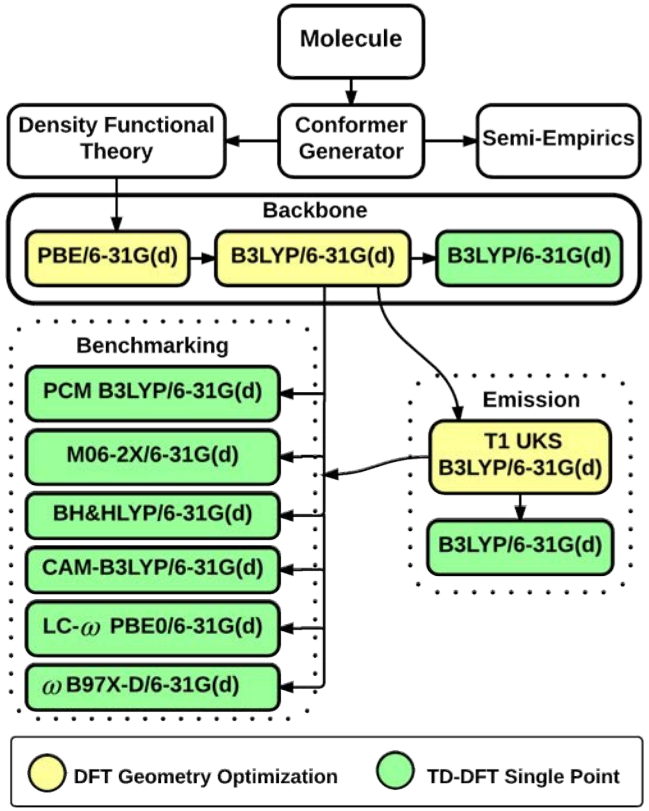


Automation and calibration

Automated calculation engine:

No operator intervention once set up

Automatic promotion to better methods



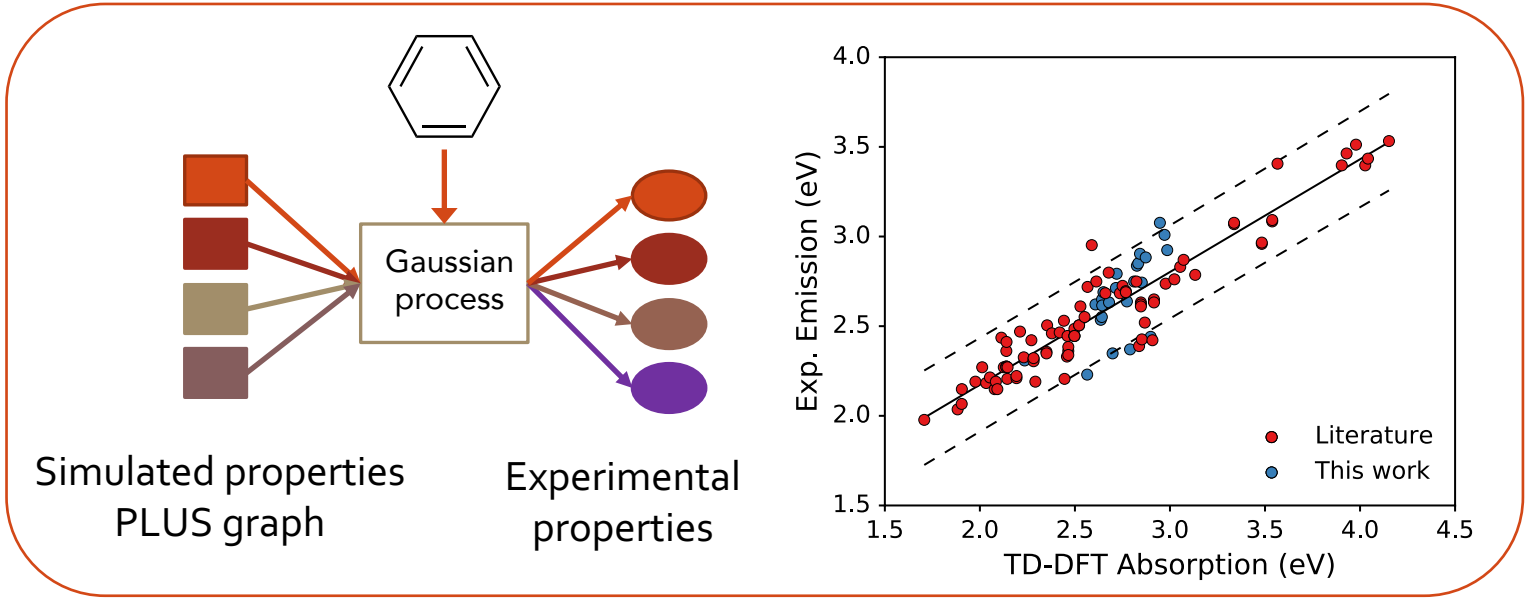
Automated calibration engine:

Machine-learning to improve theory

In house and literature data

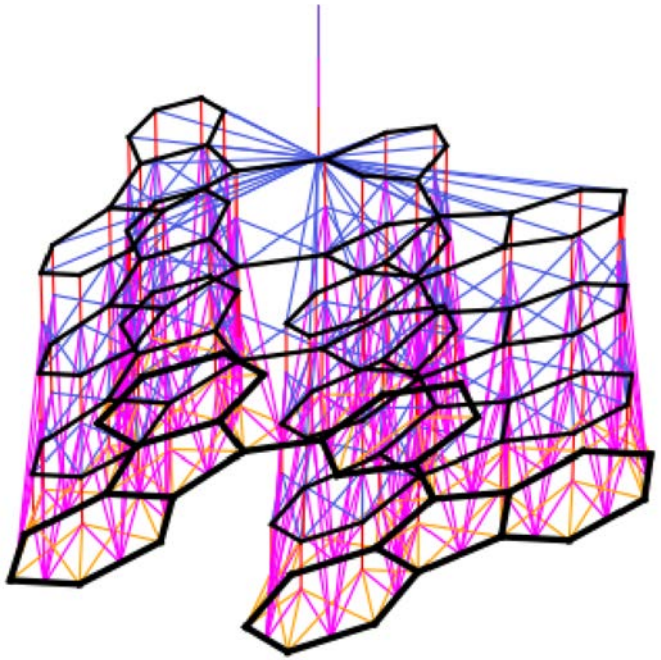
Automatically updated

Improved predictions

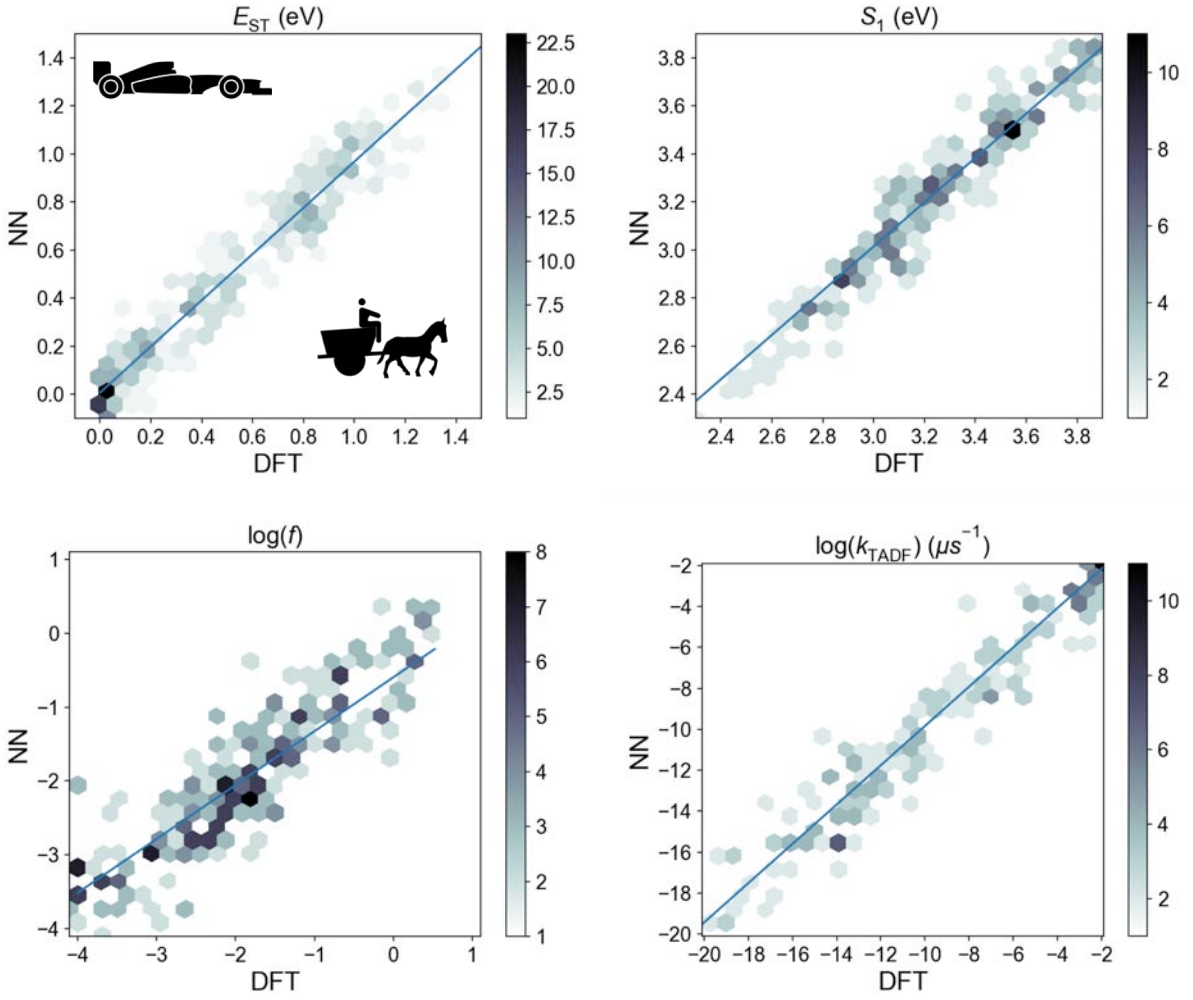


NN regression – bypassing DFT

DFT-level properties

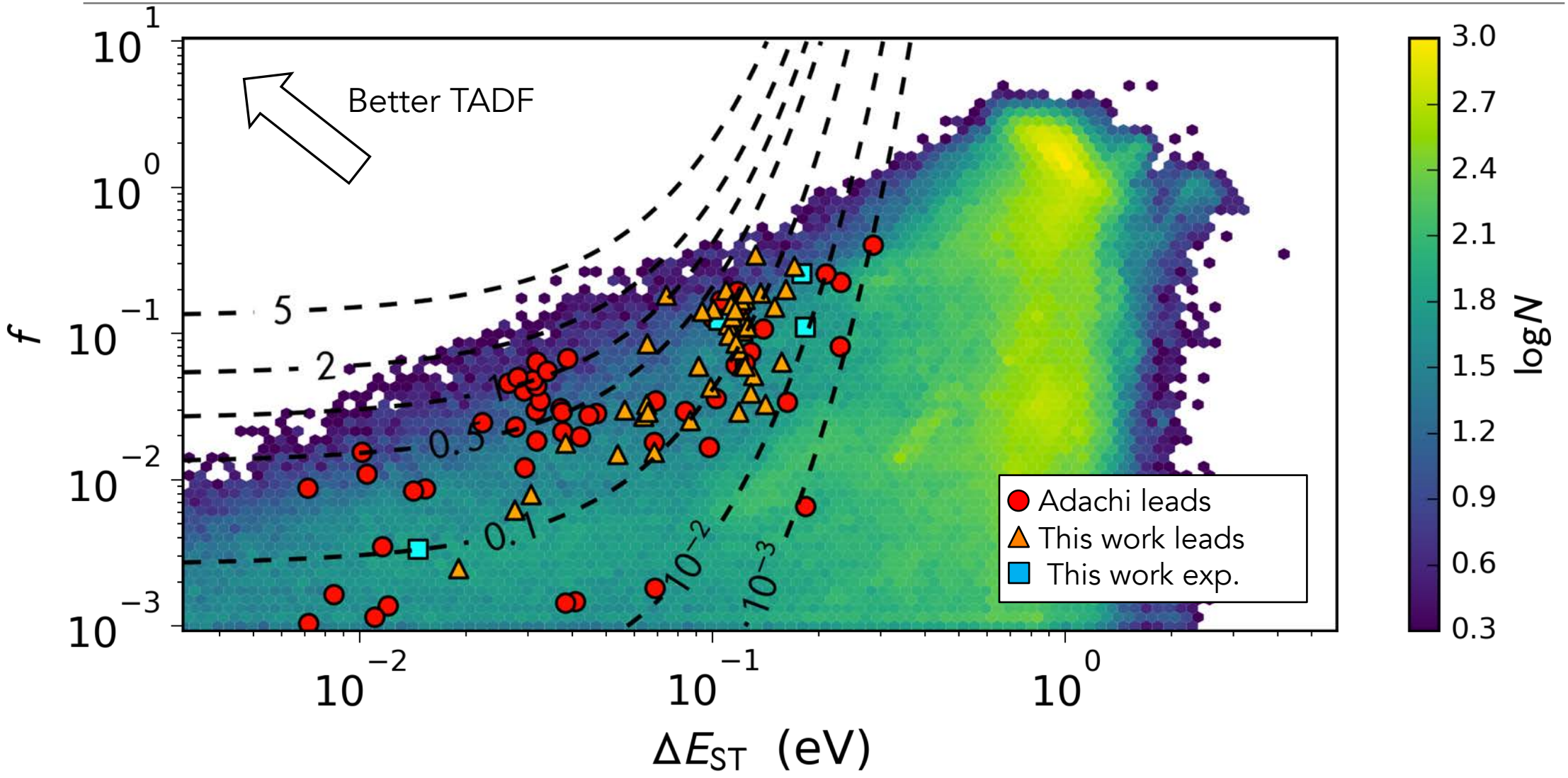


Molecule



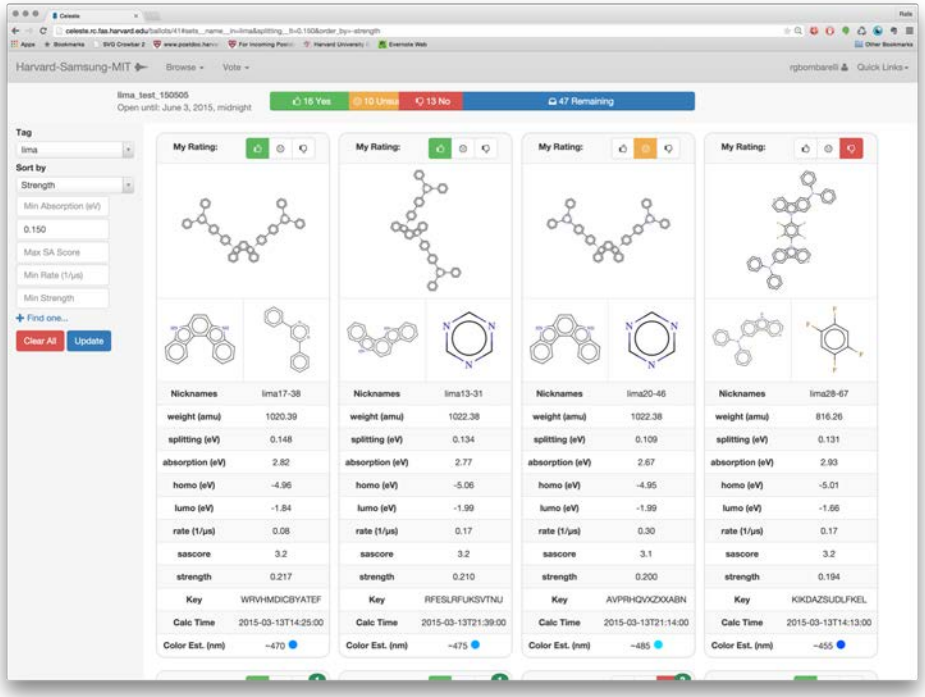
Duvenaud *et al.* arXiv:1509.09292, https://github.com/HIPS/neural-fingerprint

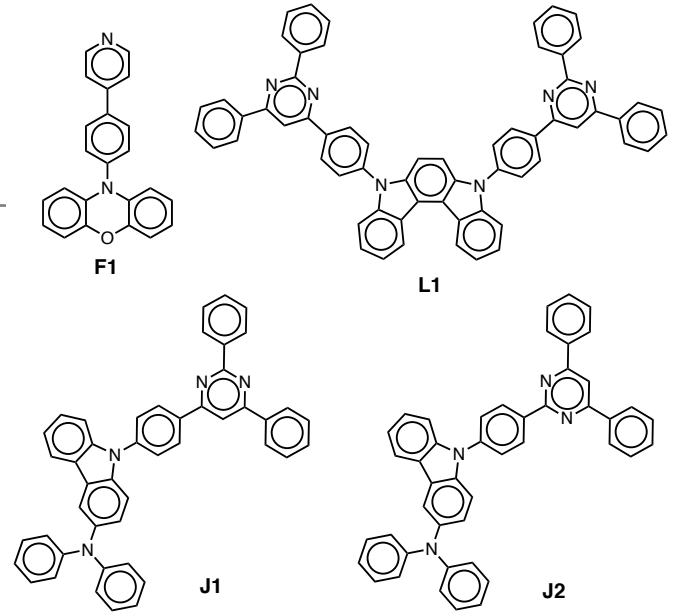
Charting chemical space

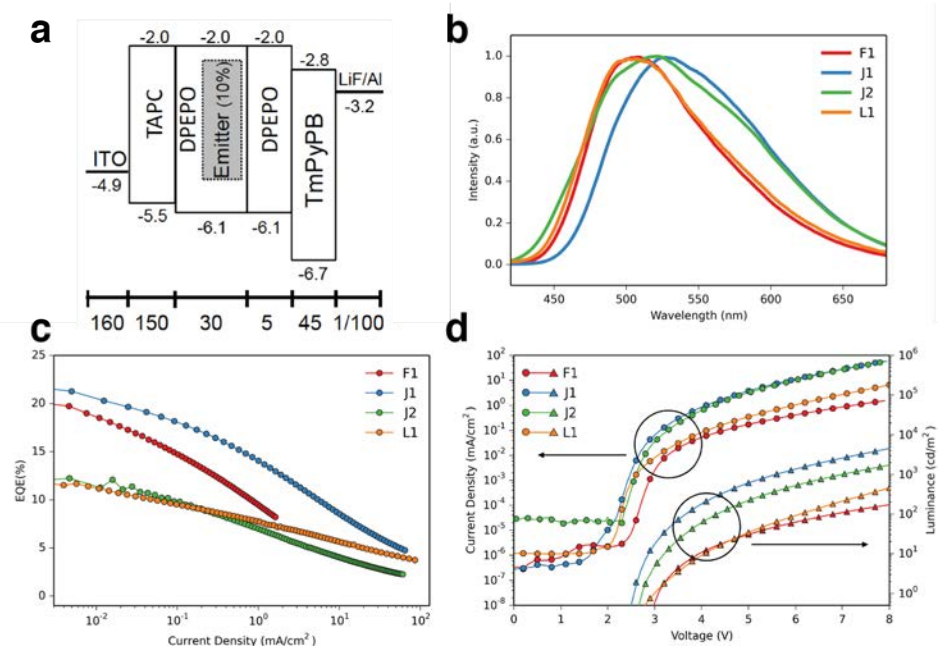


Experimental validation

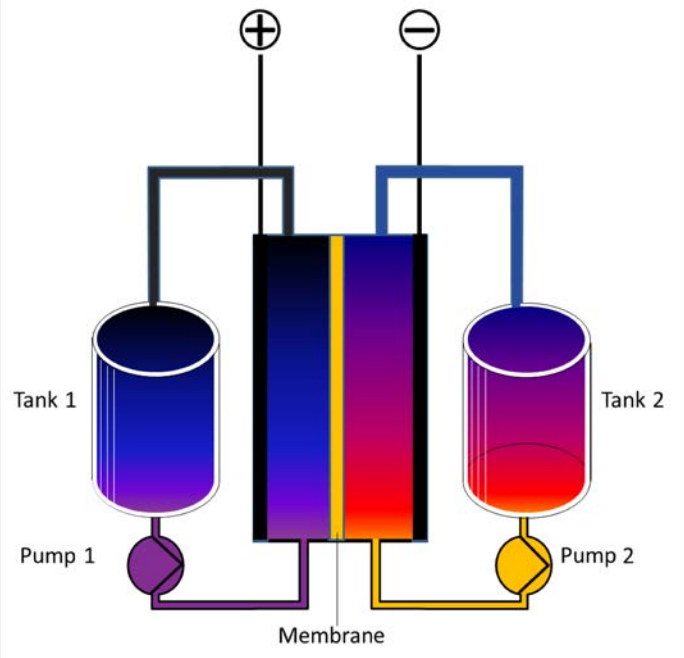
- Web-based collaborative tool
- Lead candidates synthesized
- Predictions confirmed
- Matched state of the art

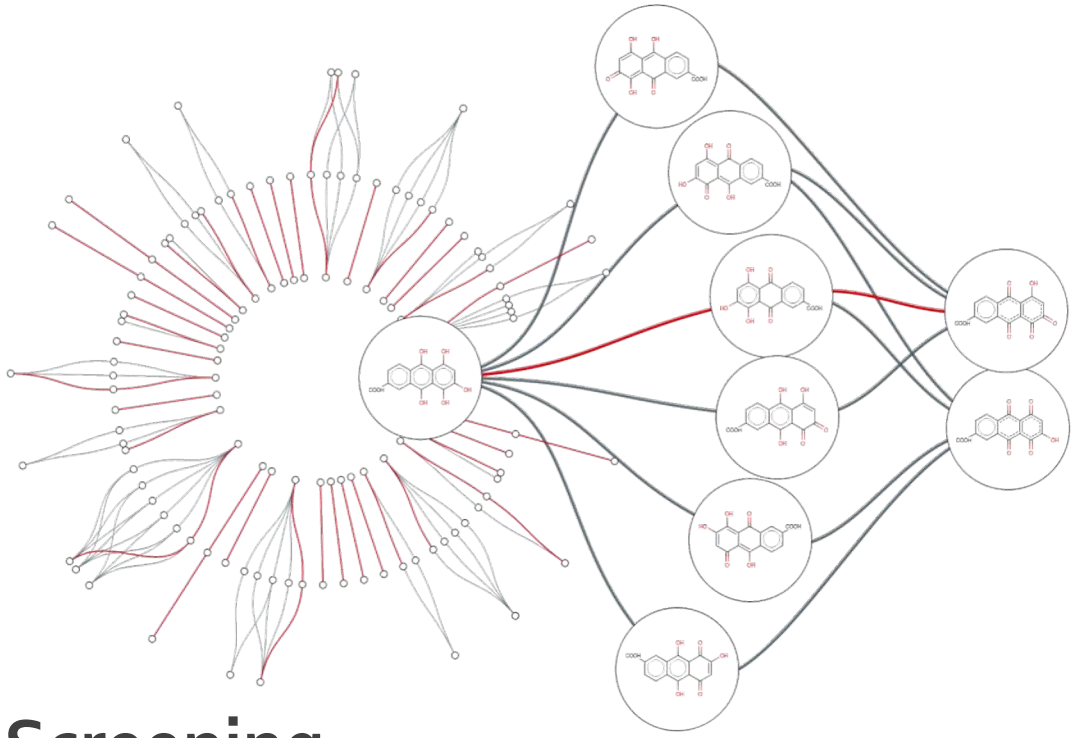






HTVS in other areas: redox flow battery





Redox flow batteries:

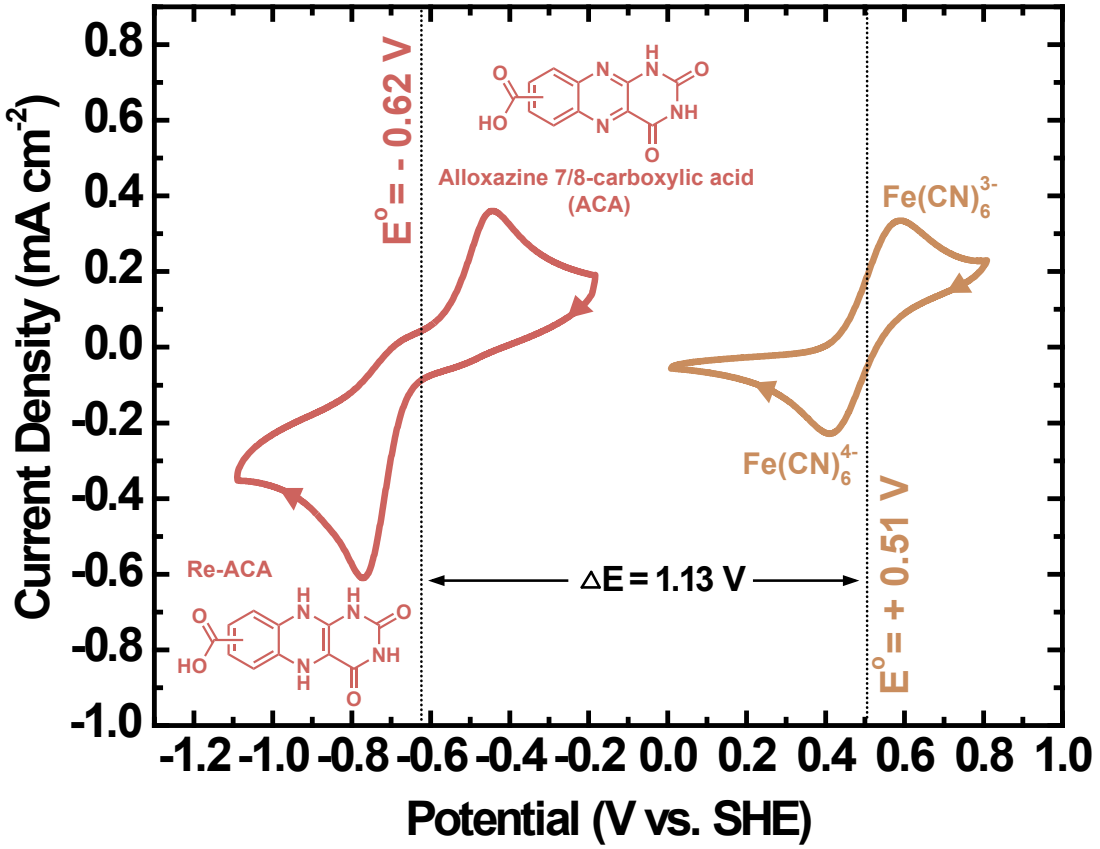
- Energy and power are independent
- Potentially low toxicity, lower safety risks
- Large scale, Efficient, Low cost
- What materials?

Screening

- Redox active, right potential
- Soluble
- Stable

HTVS in other areas: redox flow battery

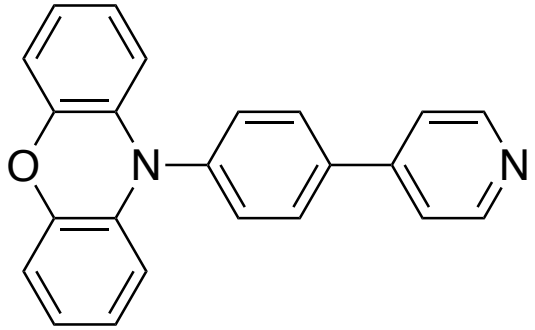
	Quinone	Alloxazine
V (V)	1.20	1.13
E Eff (%)	84	74
C Eff (%)	99.0	99.7
Cap Ret (%)	99.90	99.98
P Dens (W/cm^2)	0.45 ($0.6A/cm^2$)	0.35 ($0.6A/cm^2$)
C Dens (Ah/L)	27	54

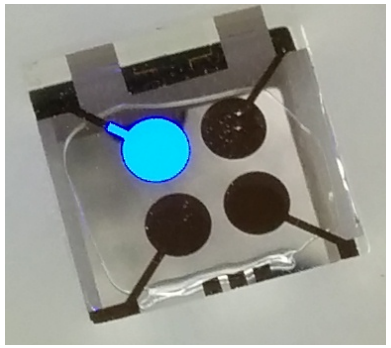


DEEP INVERSE DESIGN

Inverse design

Progress in predicting performance given candidate





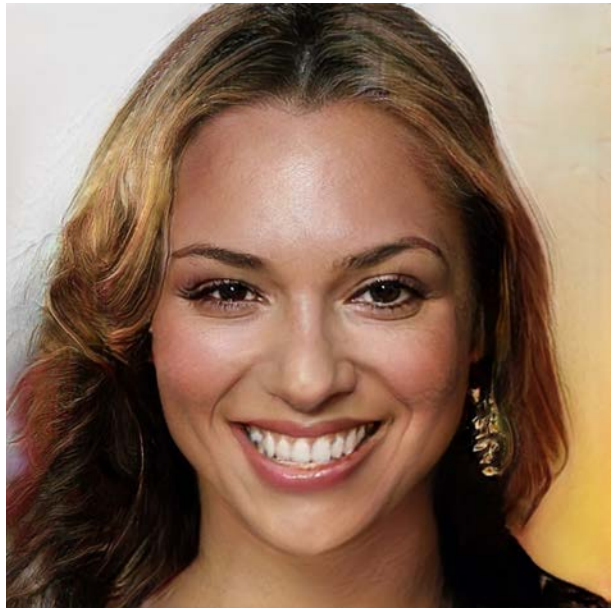
Can we generate candidate based on design targets?

?

Deep generative models

Unsupervised learning: Learning from data that has not been labeled, classified or categorized. Find a common denominator in the data.

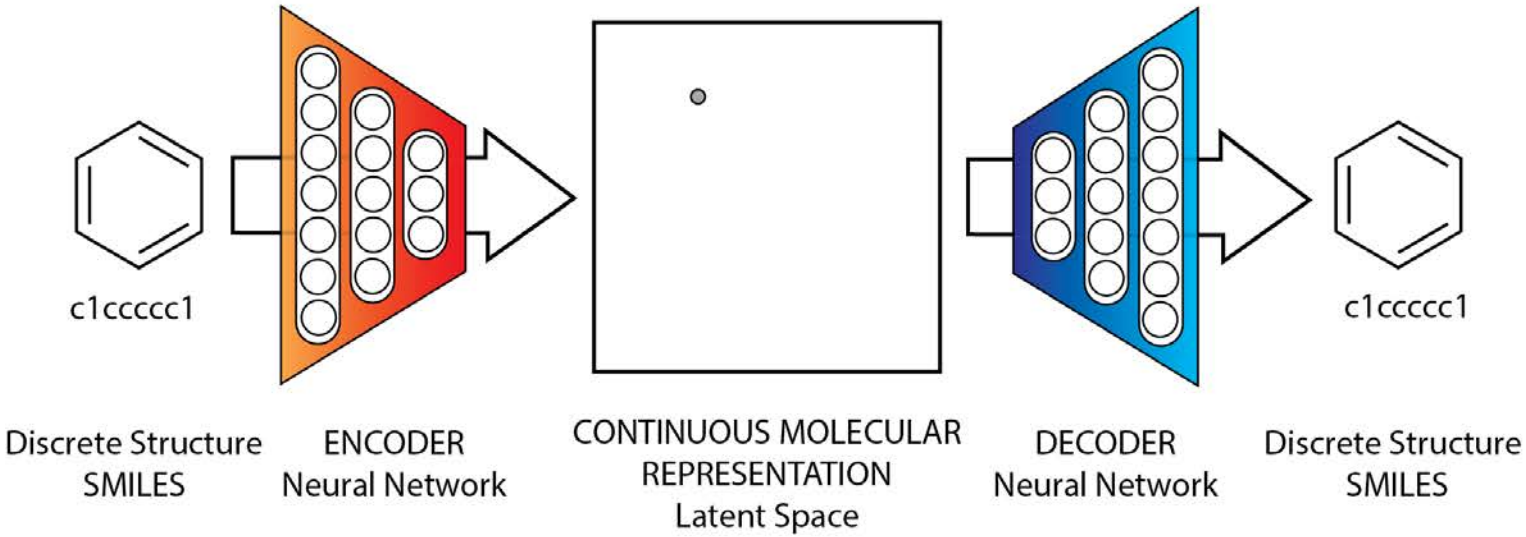
In a generative model, we then use that commonality to generate novel realistic synthetic samples.

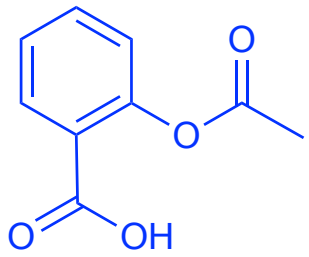


Deep molecular autoencoder

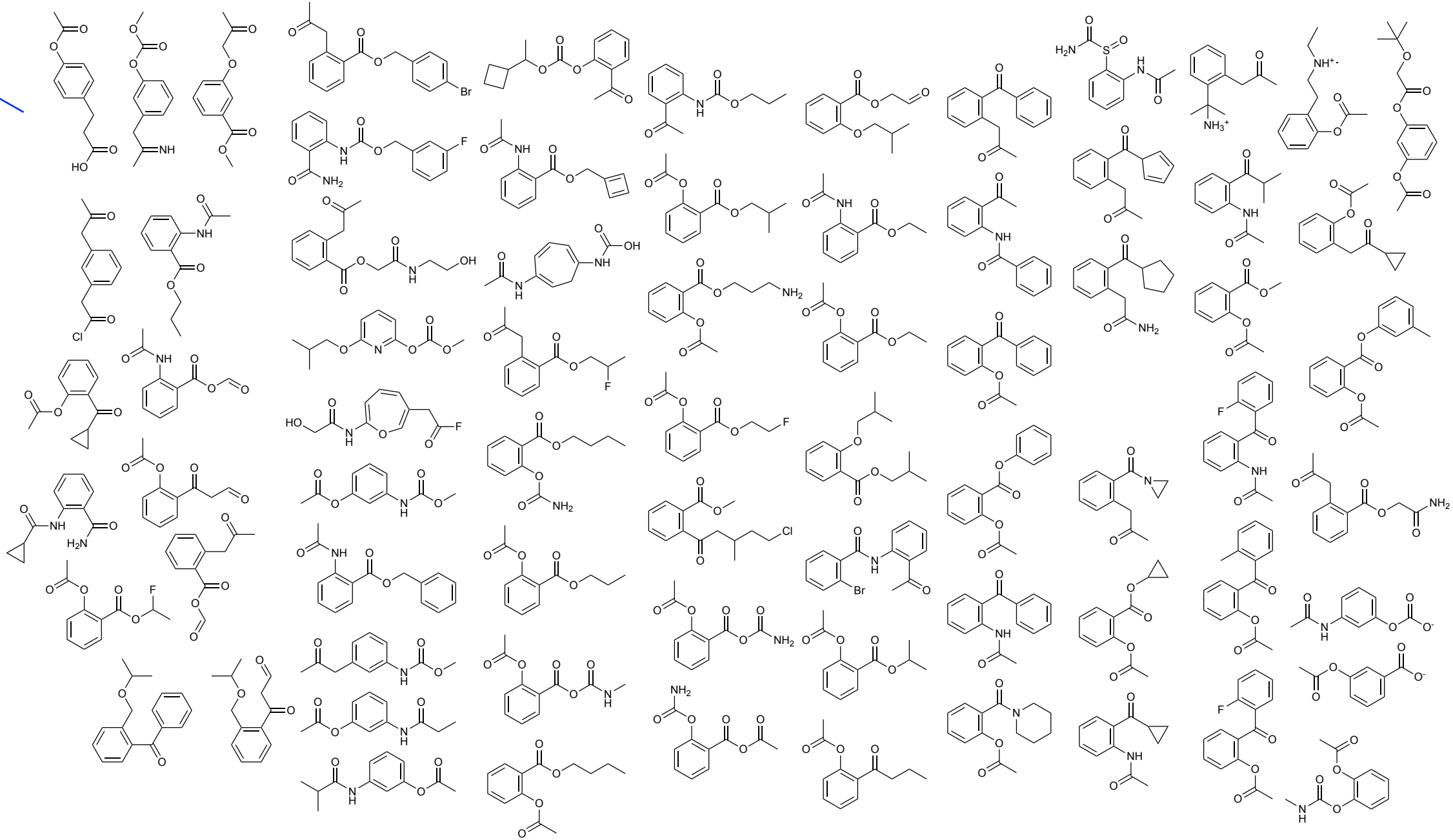
Projects discrete graph molecule onto a continuous differentiable space

Decodes any point in continuous space back out as a discrete molecular graph



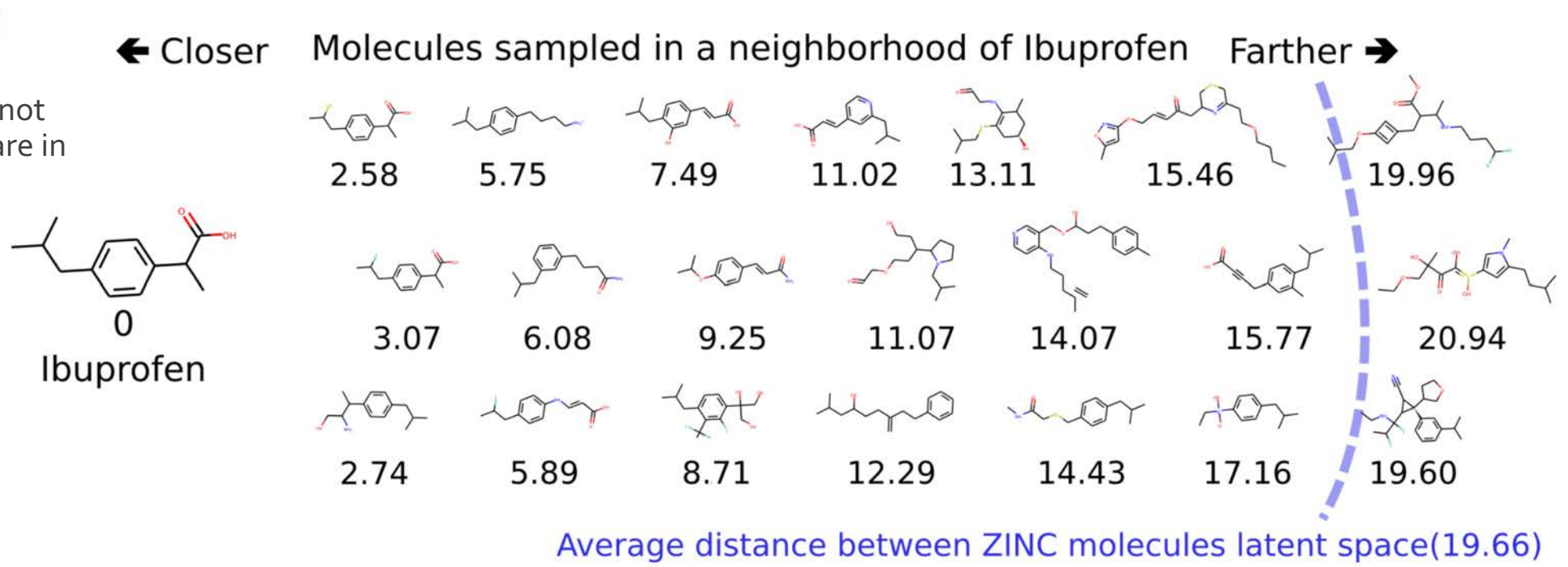


Aspirin

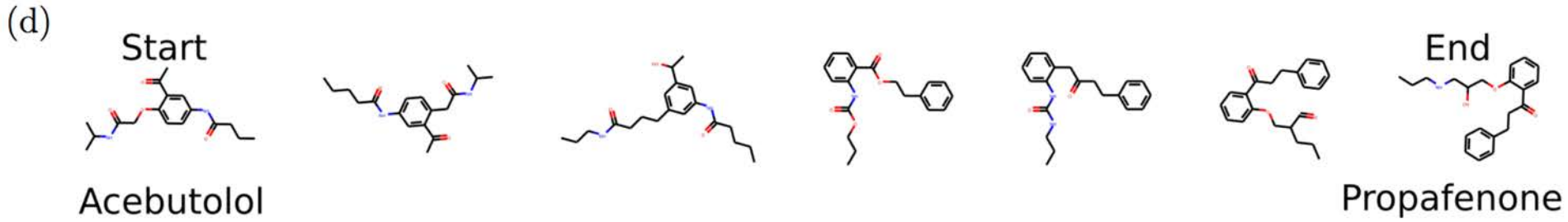


Non-linear avigation

The space is very much not linear. Most molecules are in an annulus far from the mean.

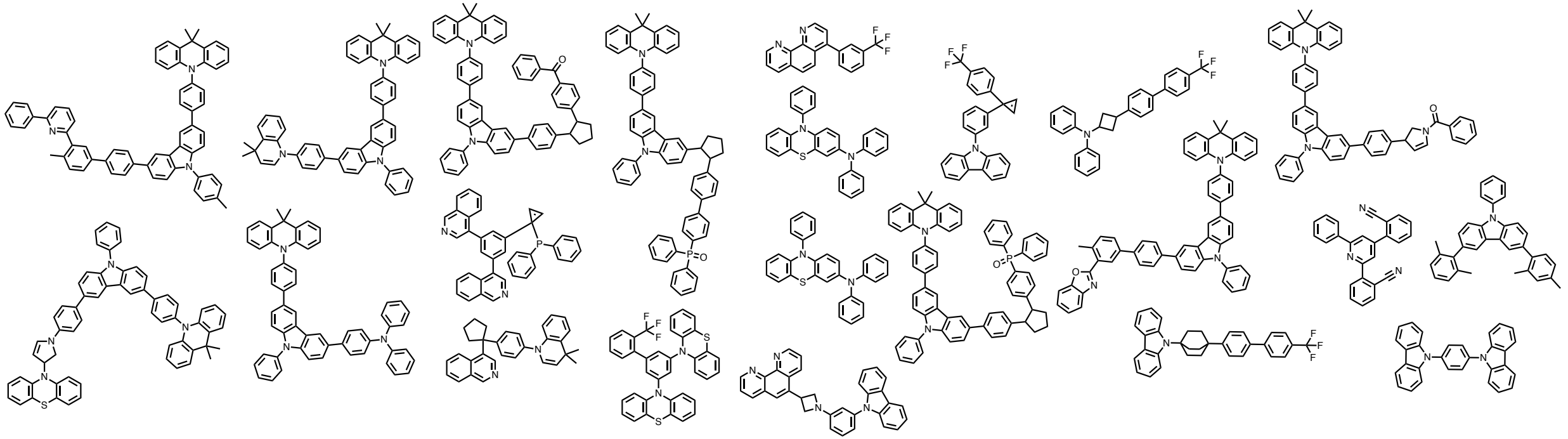


SLERP (Spherical interpolation) allows taking much more sensible steps



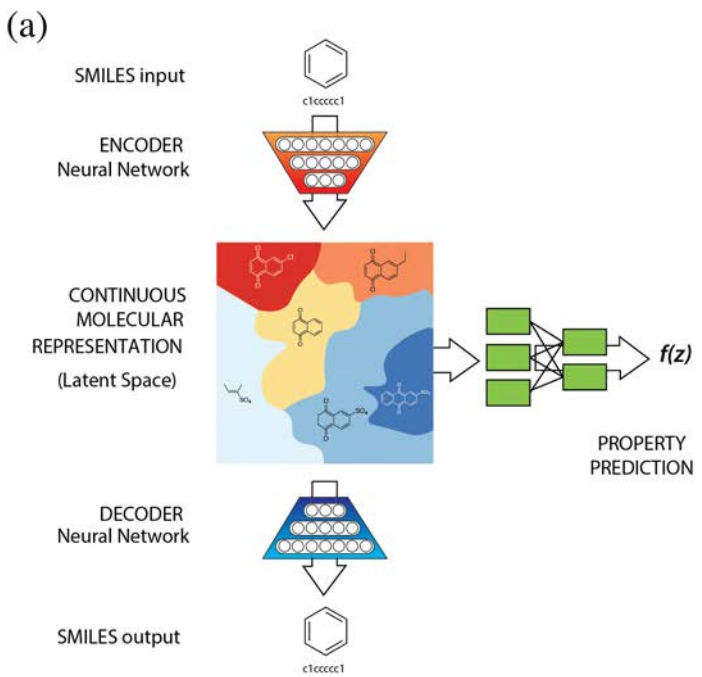
Dreaming OLEDs

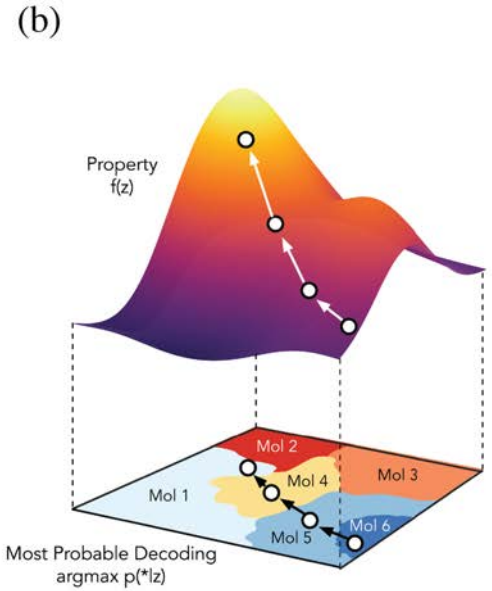
Trained on HTVS library and patented OLED
No bias, just generation

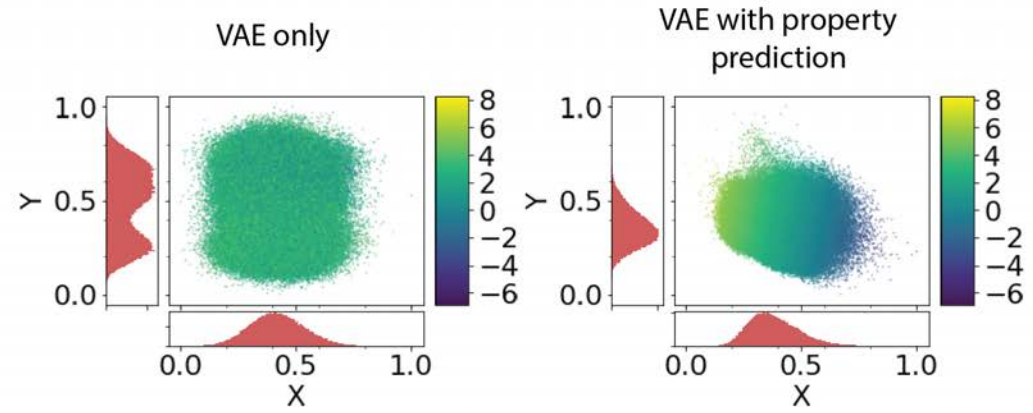


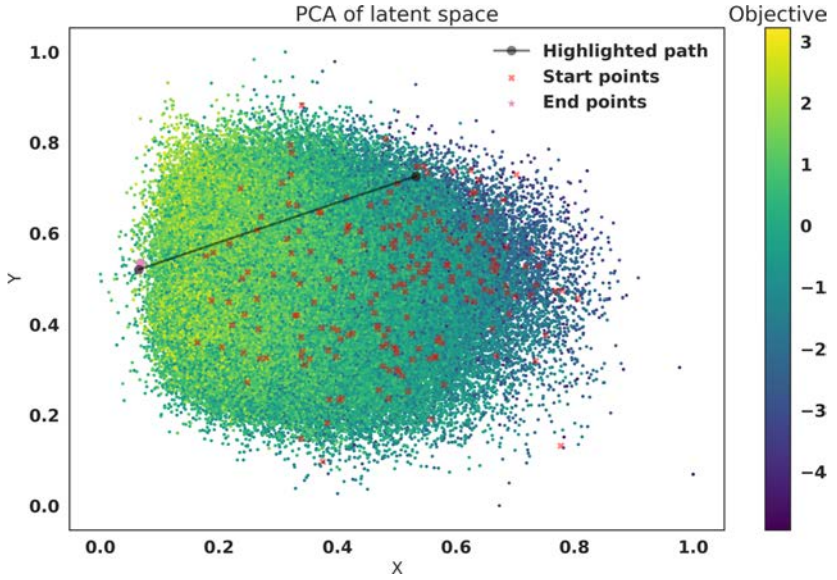
Semisupervised Molecular VAE

The latent representation now encodes mapping to properties.
Structure-property relationships over continuous space







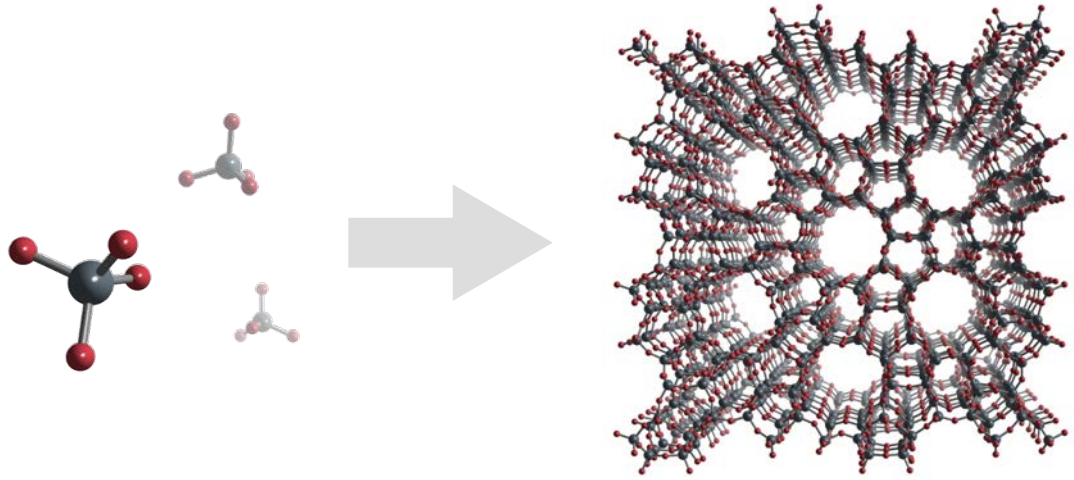


Gomez-Bombarelli et al. ACS Central Science **2018**, *4* (2), 268–276

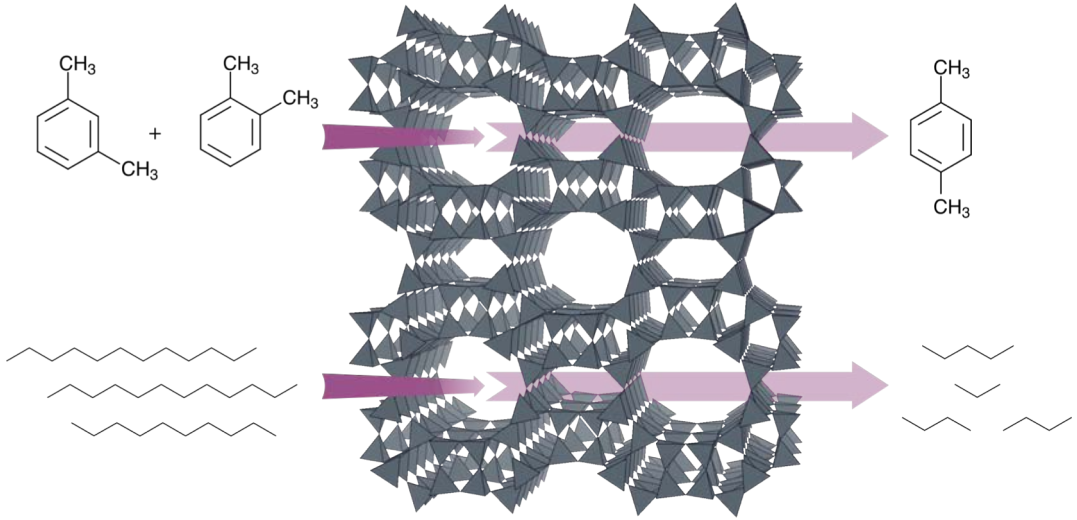
Zeolite design

Zeolites

Nanoporous materials composed of (alumino)silicates
3D Network of SiO_4 tetrahedra.



Industrial use as catalysts, molecular sieves
Highly regular, selective pores
Very robust



Zeolite conundrum

In **theory** there are millions of ways of connecting tetrahedra into a regular crystal
Computer enumeration + thermodynamics calculations estimates 100,000's

But only **235** zeolite unique frameworks have been realized since the 1750's

The International Zeolite Association keeps a database of known frameworks

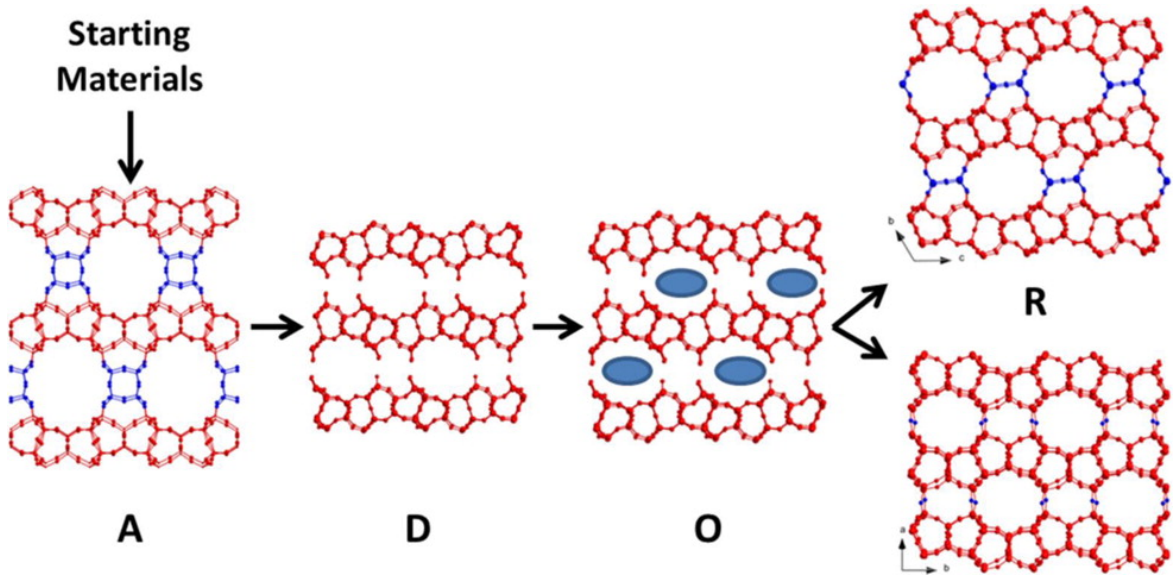
Why are some observe and some not?
Can we identify how to make new ones?

Baerlocher, C. et al. *Atlas of Zeolite Framework Types*. Elsevier, 2007.
Pophale, R. et al. *Phys. Chem. Chem. Phys.* **13** (27), 12407-12412 (2011).

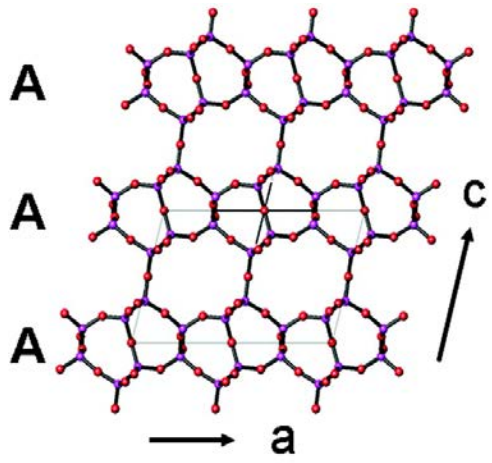
Zeolite synthesis - interconversions

More subtle directed approaches can make one-to-one conversions without organics

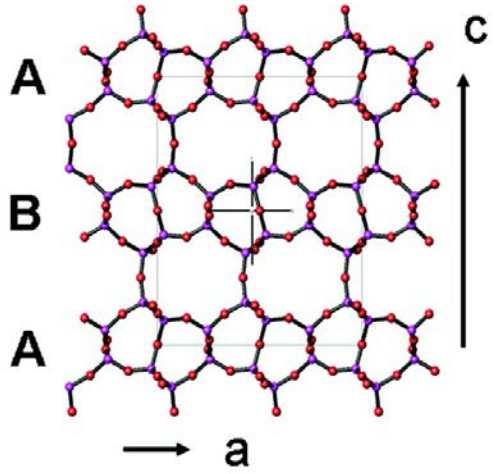
Assembly – Disassembly – Organization – Reassembly



Topotactic transformations



Zeolite framework type NSI

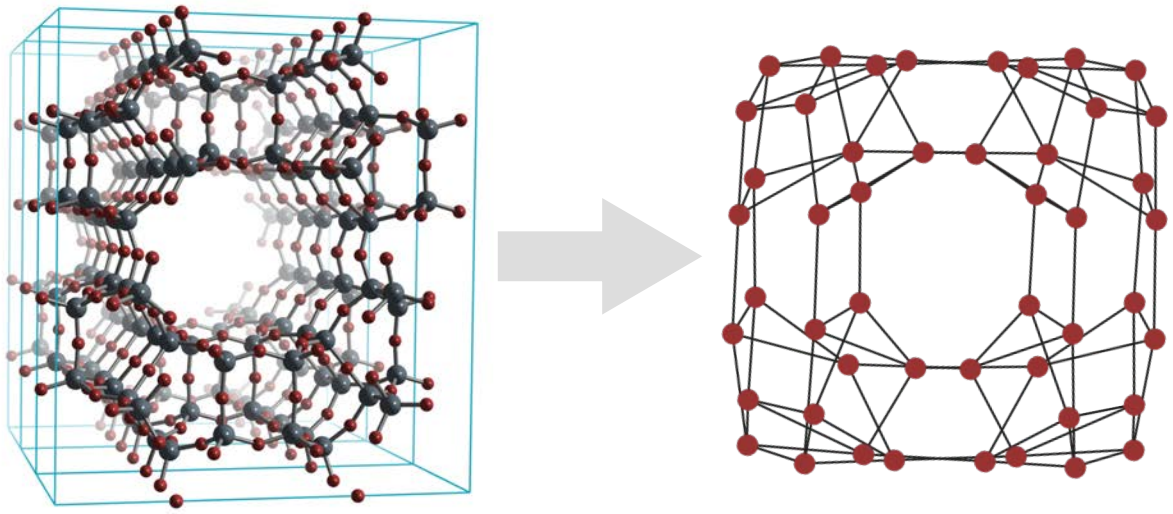


Zeolite framework type CAS

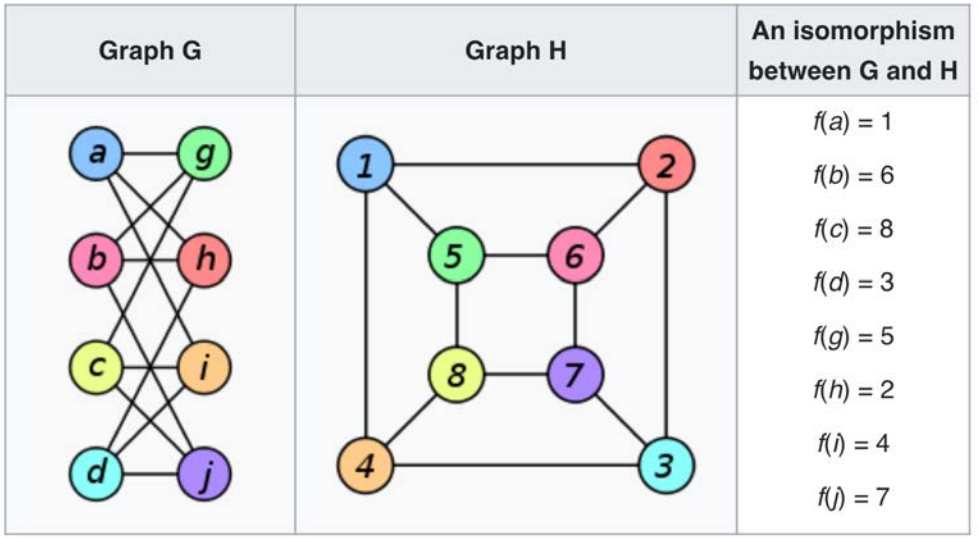
Roth, W. *et al*. *Chem. Rev.* **114**, 4807-4837 (2014)
Marler, B. and Gies, H. *Eur. J. Mineral.* **24**, 405-428 (2012)

Connectivity similarity

Crystal Graph: keep track of connectivity only. Atoms are connected also through PBC



Isomorphism: Two graphs are isomorphic if they differ just by a relabeling of the nodes



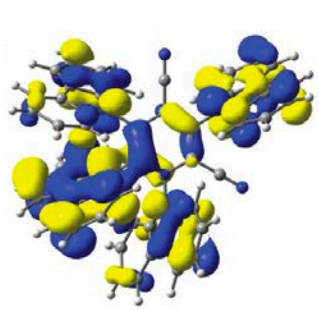
Similar to molecular formulas, ignore distance, track only connectivity

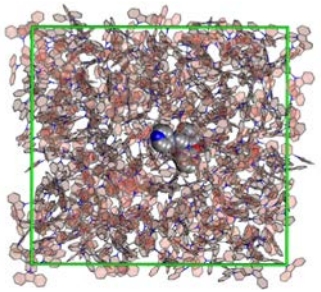
Xie, T. and Grossman, J. C. *Phys. Rev. Lett.* **120,** 145301 (2018).

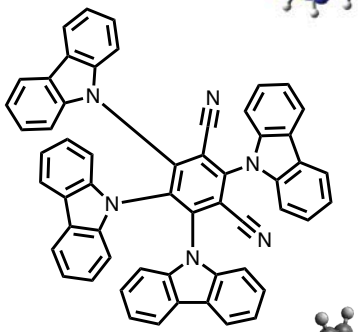
Similarity of the graphs is connected to synthetic transformations

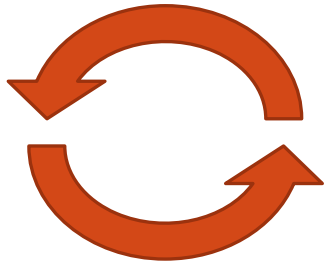
Conclusions and outlook

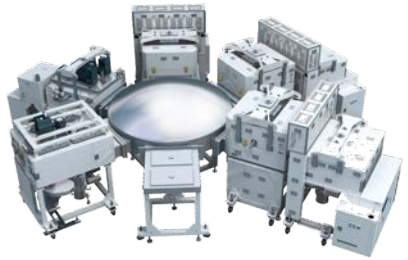
Blurring lines between ML and simulation

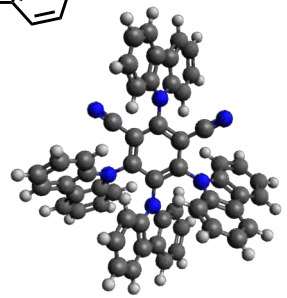


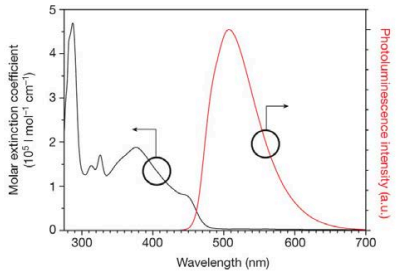












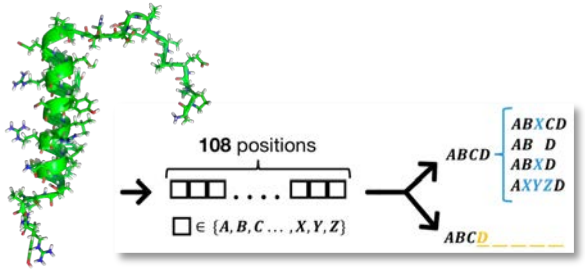
Parametrization of a physics model is a learning problem. Even building a physics model may be a learning problem

Machine learning over matter is an issue of representation: how to input a system in a way that captures known physics and chemistry

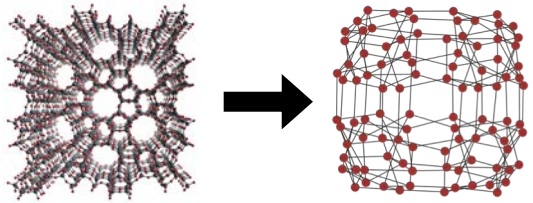
Simulations and machine learning are two sides of the same coin.

ML and simulation – blurred lines

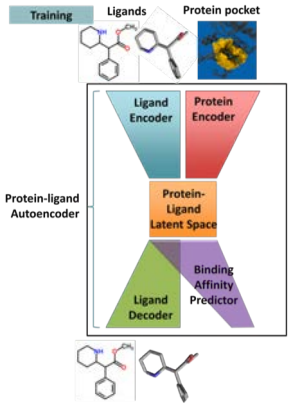
Data-driven discovery and optimization.



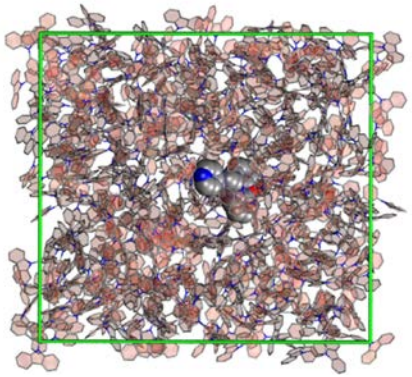
Artificial peptide design



Data driven discovery of zeolite catalysts

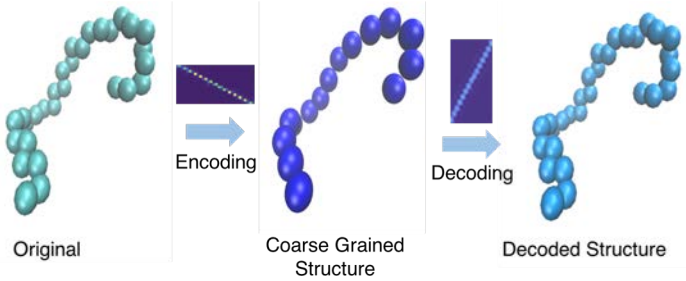


Small molecule binder

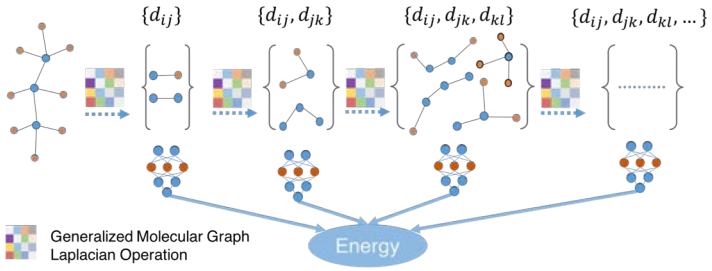


Organic electronics

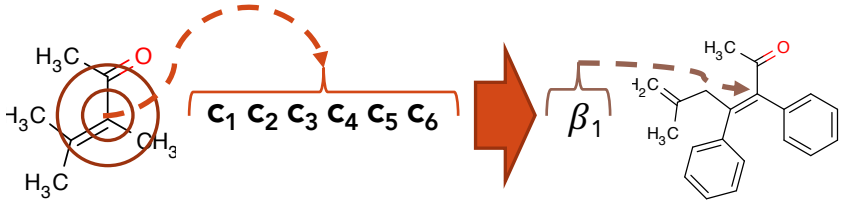
Method development and representation learning



Automatic coarse graining of atomistic simulation

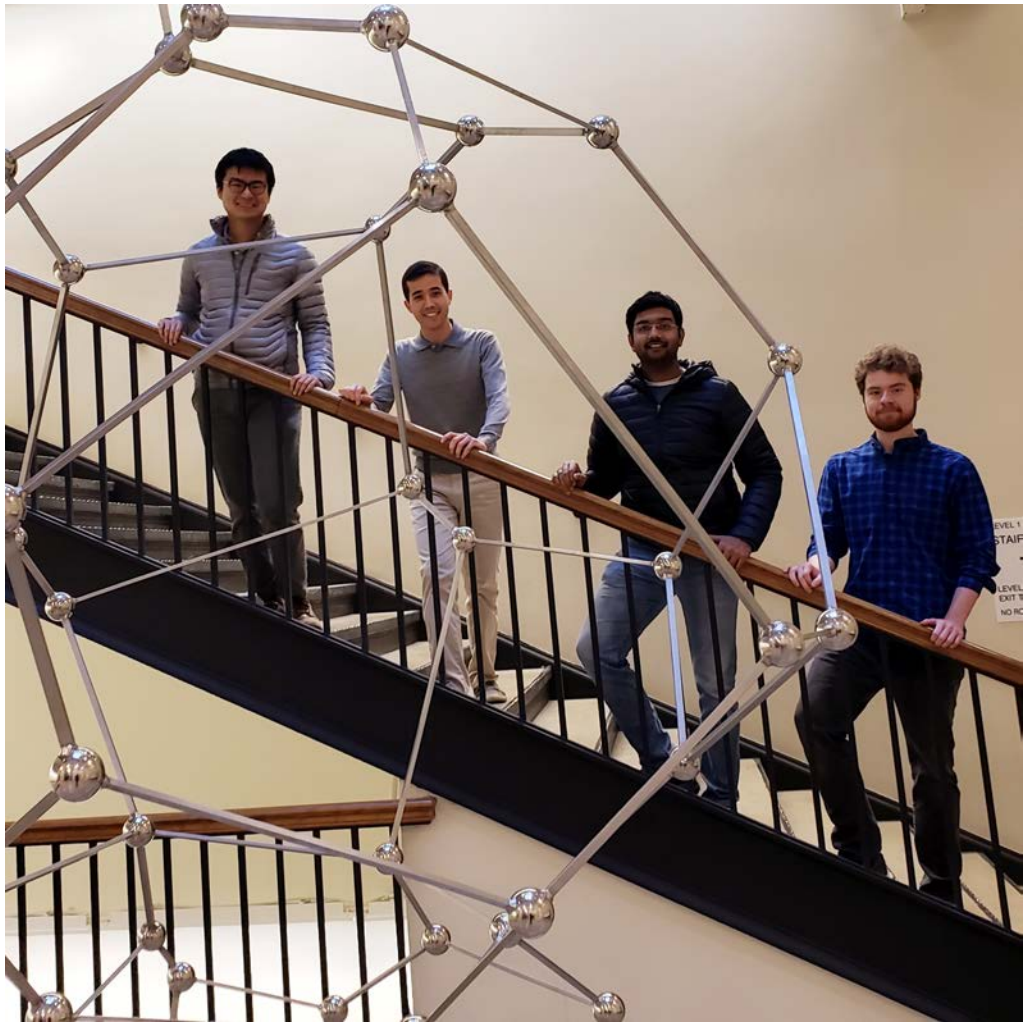


Continuous typing force fields



Basis sets

Thanks!



Former co-workers and collaborators
Jorge Aguilera, Tim Hirzel
Aspuru-Guzik, Adams, Aziz, Gordon, Baldo, Duvenaud, Hernandez-Lobato groups, SAIT, Kyulux NA

Wujie Wang
Wil Harris
Daniel Schwalbe Koda
Somesh Mohapatra

DMSE DEPARTMENT OF MATERIALS SCIENCE & ENGINEERING

Massachusetts Institute of Technology

Inverse Design of Materials

through artificial intelligence and physics-based simulations

MIT ILP Japan Conference.
Tokyo January 25 2019

Rafael Gomez-Bombarelli
rafagb@mit.edu
http://gomezbombarelli.mit.edu/