



Using Text-Mining, Community Knowledge, and Generative Modeling to Quantify and Engineer Stability in MOFs

Aditya Nandy^{1,2,3,4}

¹Department of Chemistry, MIT

²Department of Chemical Engineering, MIT

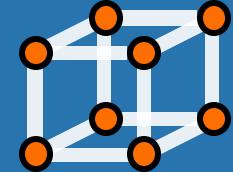
³James Franck Institute, University of Chicago (Schmidt AI Fellow)

⁴University of California, Los Angeles (started July 1, 2025)

AIMS Workshop @ NIST (July 10, 2025)

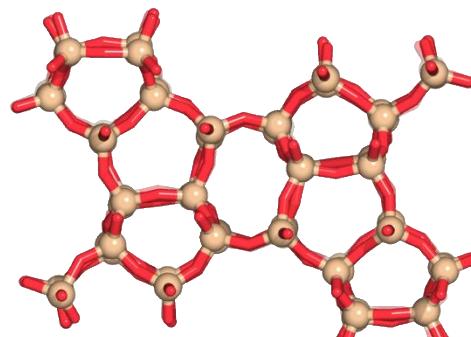
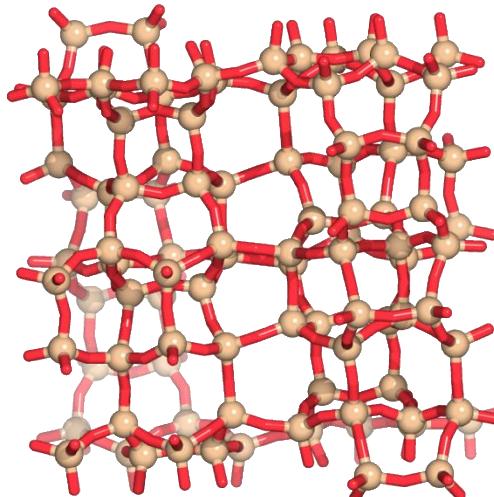
UCLA Samueli

Chemical & Biomolecular Engineering



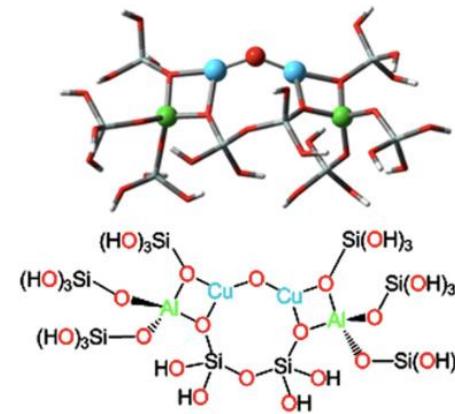
Heterogeneous Catalysis is Essential in Industry

- Heterogeneous catalysis → studied by ChemEs for > 1 century → fewer separations

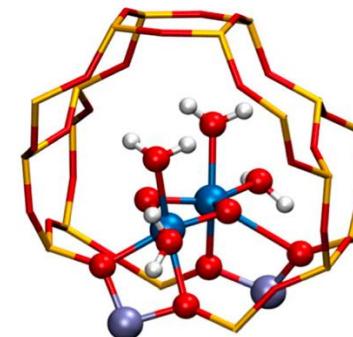


ZSM-5

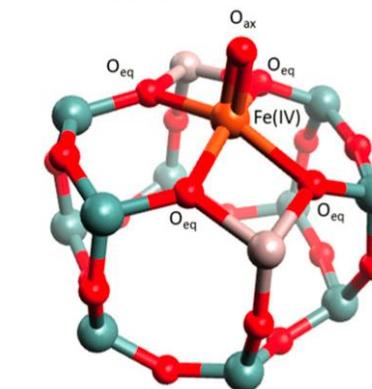
- Porous materials → zeolites → high active site density



Cu-ZSM-5¹



Fe-ZSM-5²



Fe-CHA³

- Isolated active sites → activity for various reactions

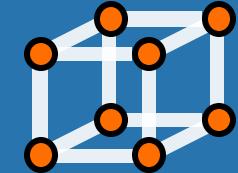
A material with a high density of isolated active sites is desirable for reactivity

¹Woertink, J. S. et al. *Proc. Natl. Acad. Sci. U.S.A.*, **2009**, *106*, 18908-18913.

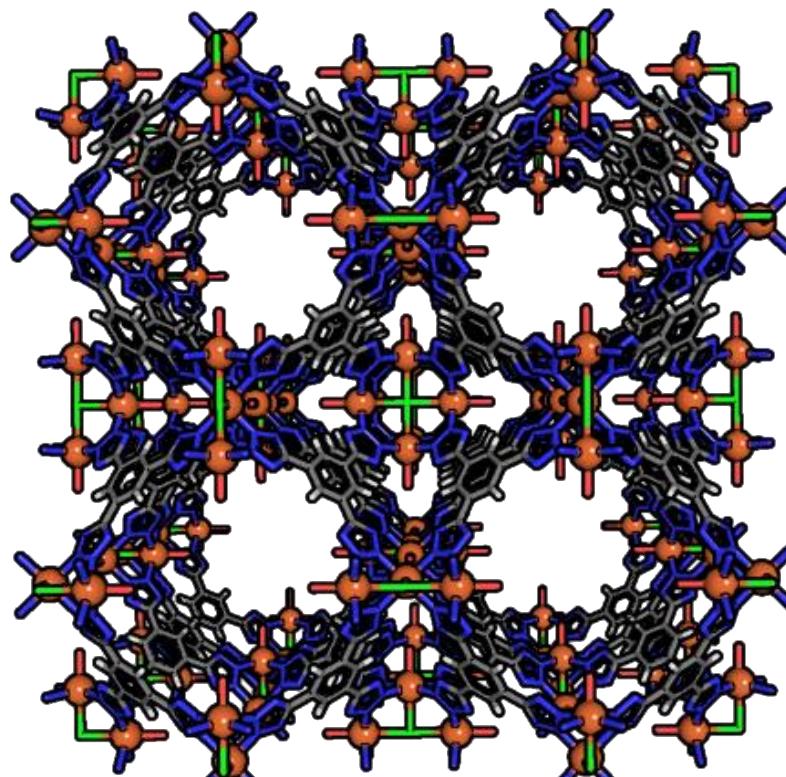
²Li, G. et al. *J. Phys. Chem. C*, **2013**, *117*, 413-426.

³Bols, M. L. et al. *J. Am. Chem. Soc.*, **2018**, *140*, 12021-12032.

Metal-Organic Frameworks as Candidate Catalysts

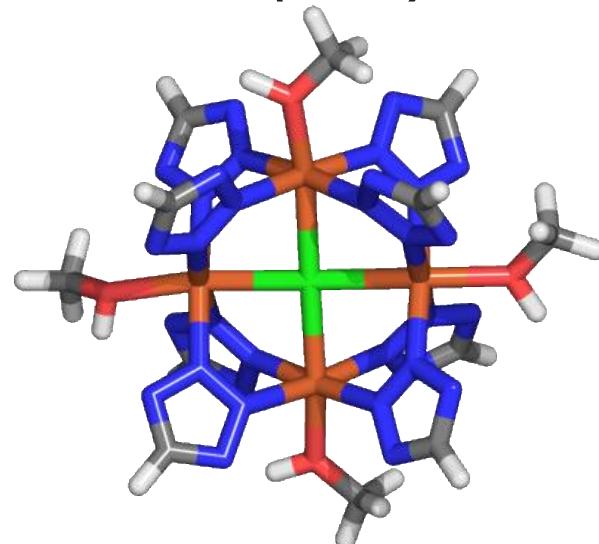


- MOFs → promise for heterogeneous catalysis but are not stable

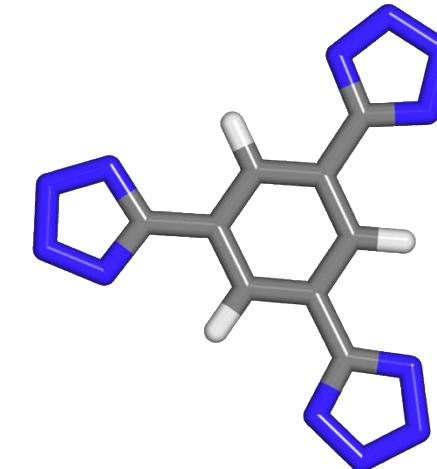


FeFe-BTT MOF

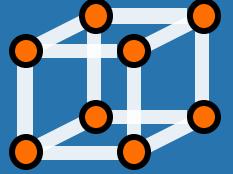
*secondary building
unit (SBU)*



linker

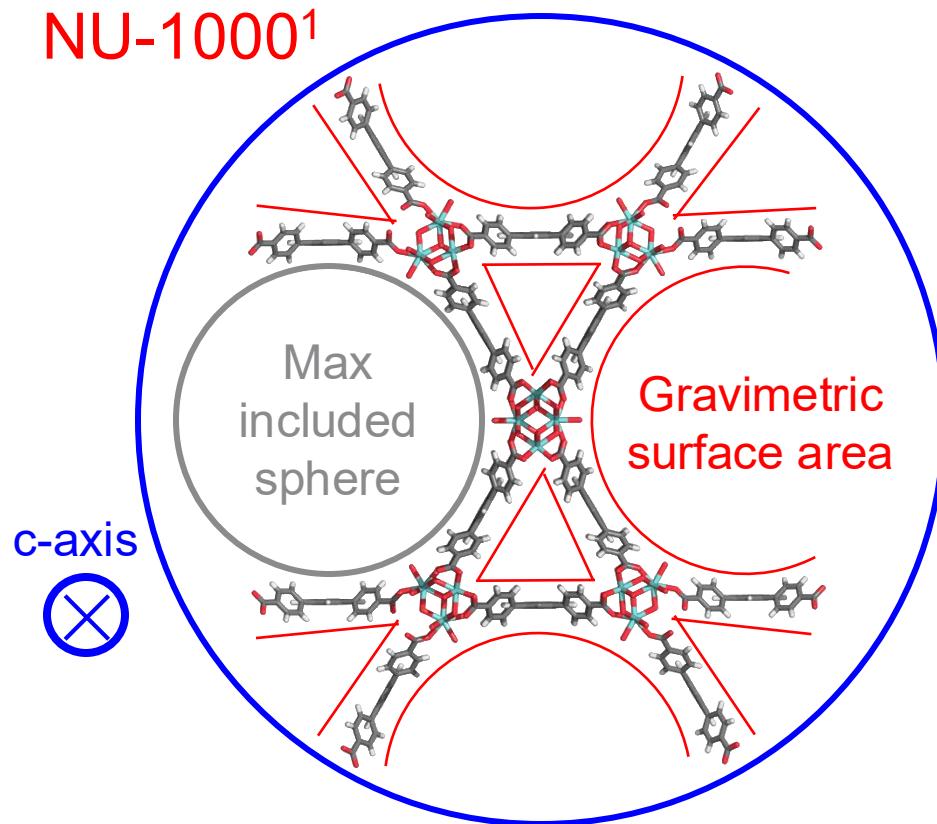


Design rules for stable MOFs are unknown

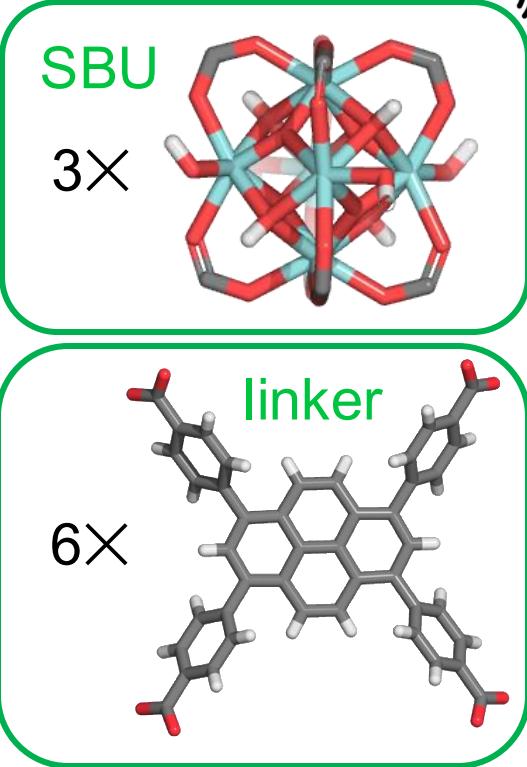
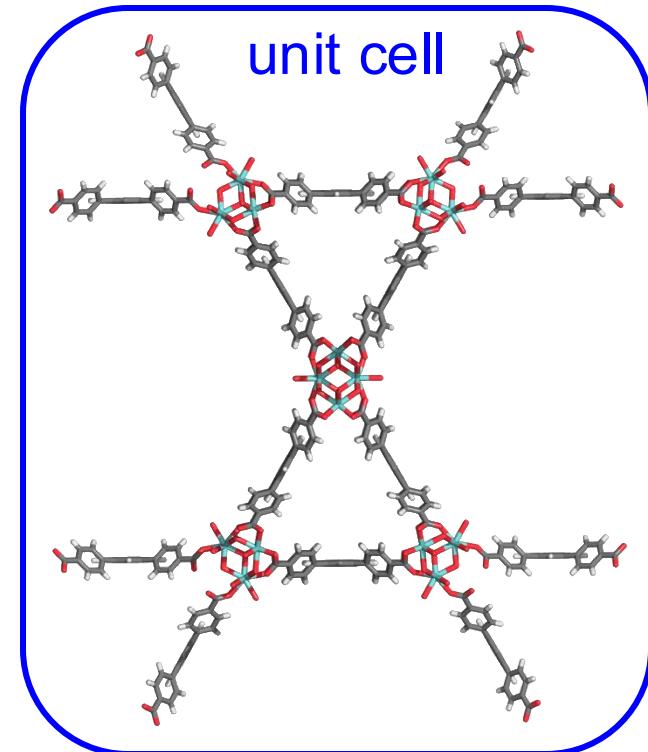


Prior MOF Fingerprints Did Not Encode Chemistry

Before: *Where is the chemistry?*



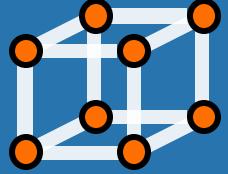
After: *graph fingerprint on MOF components*



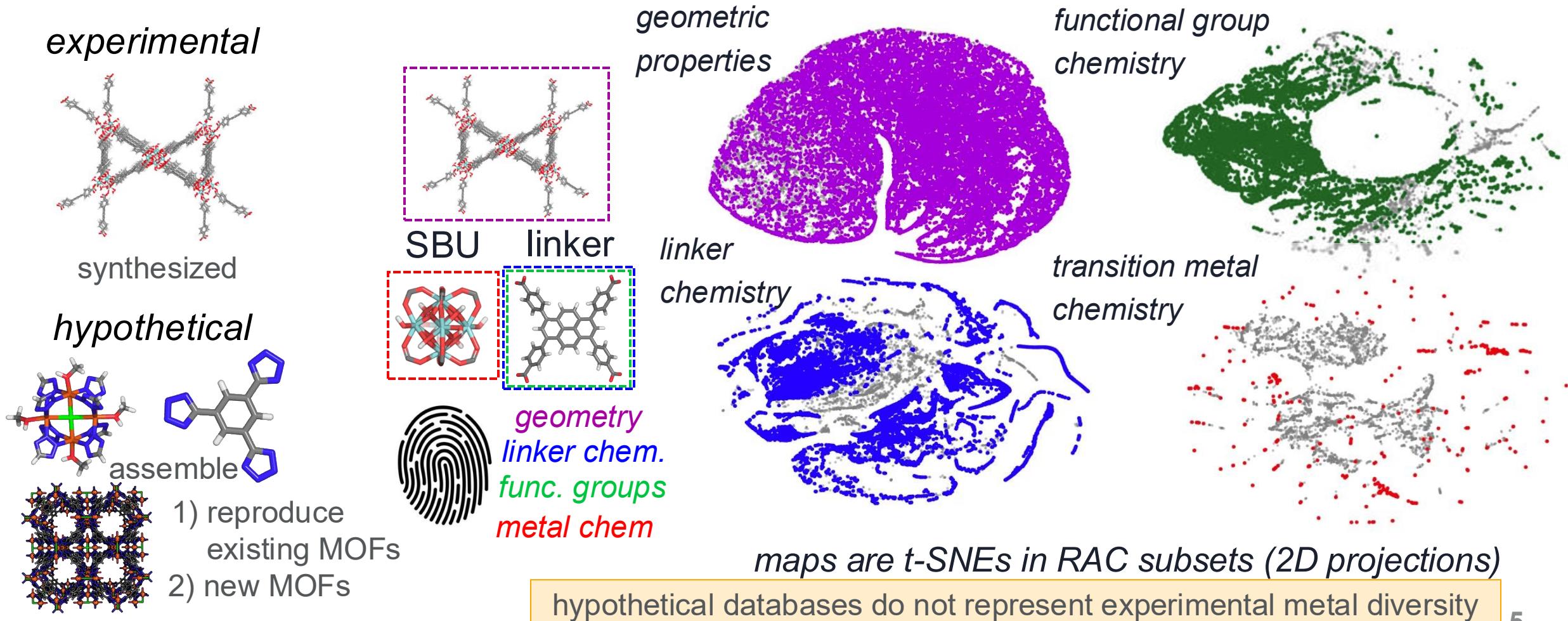
our graph fingerprint (RACs) adds chemistry to MOF descriptions

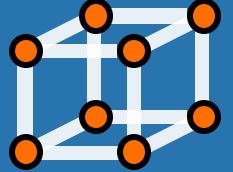
¹Islamoglu, T.; Otake, K.; Li, P.; Buru, C.T.; Peters, A.W.; Akpinar, I.; Garibay, S.; Farha, O.K. *CrystEngComm*, **2018**, *20*, 5913-5918.

²Moosavi, S. M.; Nandy, A.; Jablonka, K.; Ongari, D.; Janet, J.P.; Boyd, P.G.; Lee, Y.; Smit, B.; Kulik, H. J. *Nat. Commun.*, **2020**, 4068.

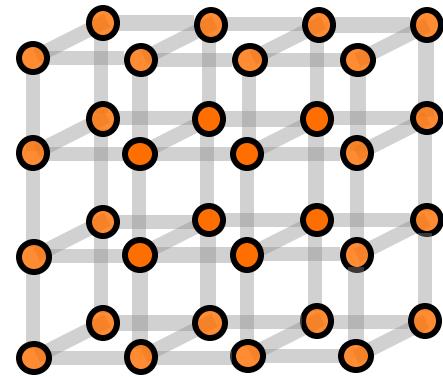


Hypothetical Databases Undersample Metal Diversity

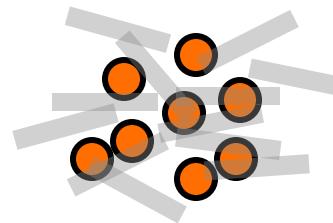




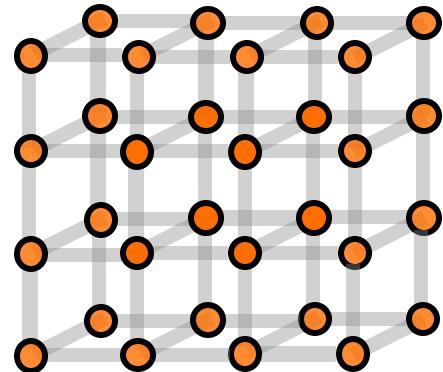
The Instability of MOFs Limits Their Use as Catalysts



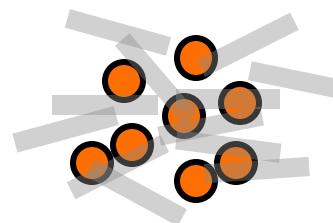
vacuum
heat



“activation stability”
classification problem

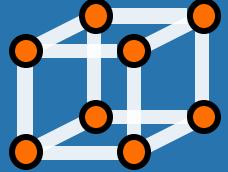


heat
until
collapse



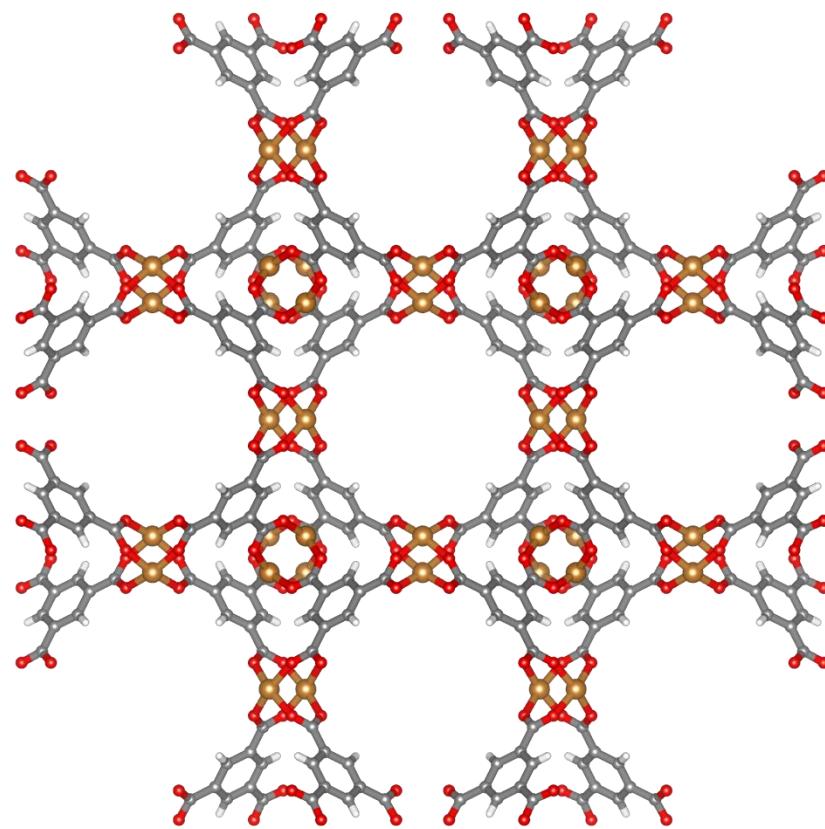
“thermal stability”
regression problem
 T_d (°C)

MOF stabilities cannot be computed but are frequently experimentally reported



Challenges to Gathering Data

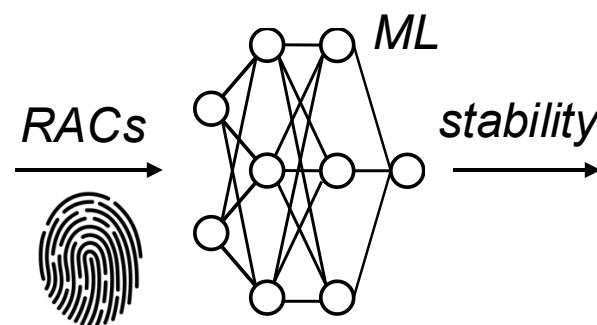
- MOFs → named inconsistently → challenging literature extraction with no structures



Popular Cu MOF → HKUST-1

In the literature:

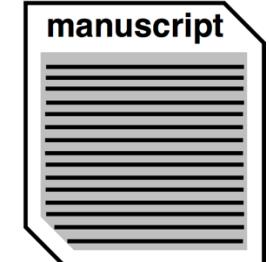
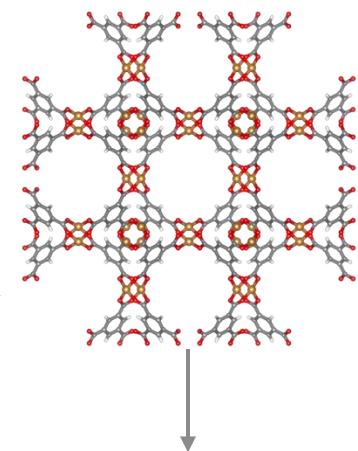
1. HKUST-1
2. Cu-BTC
3. Cu₃BTC₂
4. MOF-199
5. Basolite™ C300

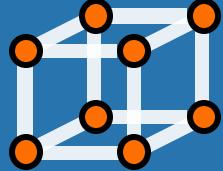


*structural
database*



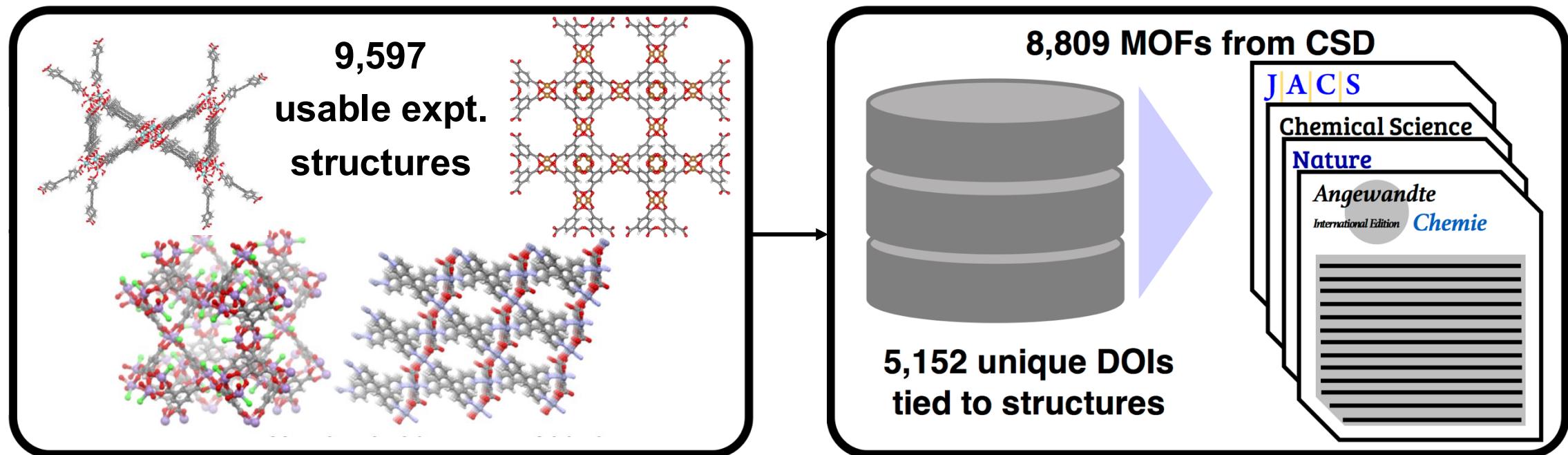
*experimental
structures*



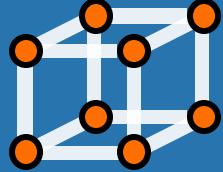


Structural Databases Contain DOIs for Text Extraction

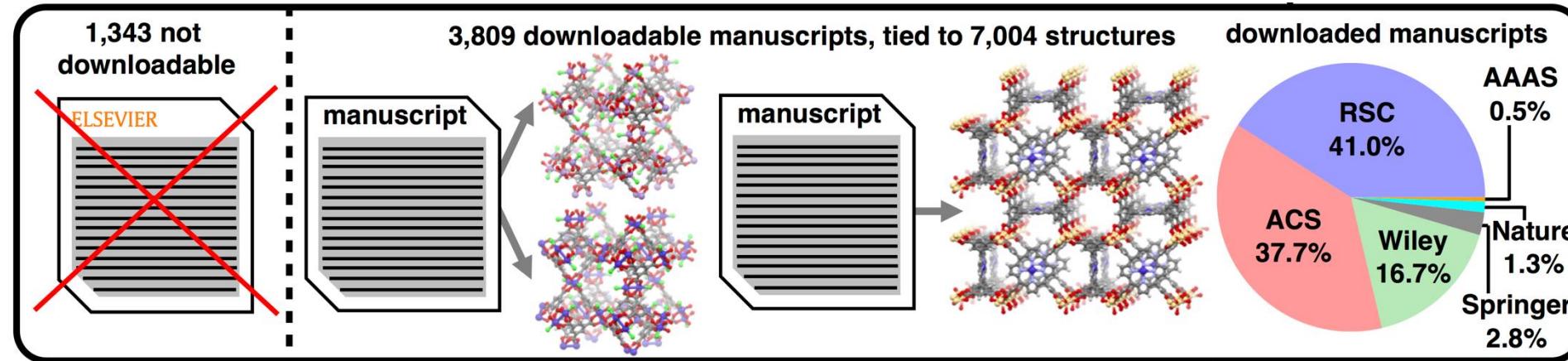
- CoRE MOF 2019¹ database → manuscripts in Cambridge Structural Database (CSD)



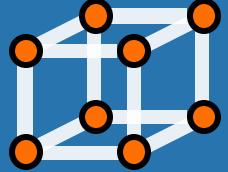
Tying MOF structures to manuscripts → enables quantitative structure-property maps



Extraction of MOF Stability Data from the Literature

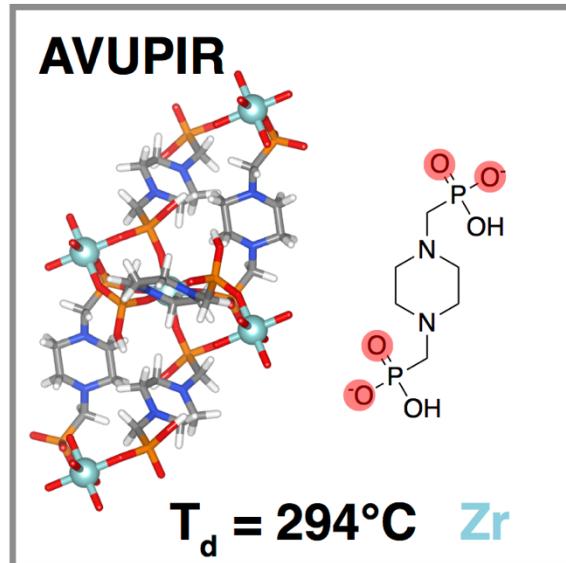


Creating structure property maps from literature data provides stability insights



Labeled Data Uncovers Oversimplified Conclusions

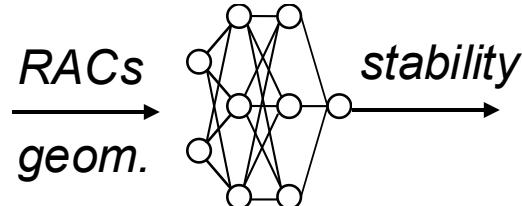
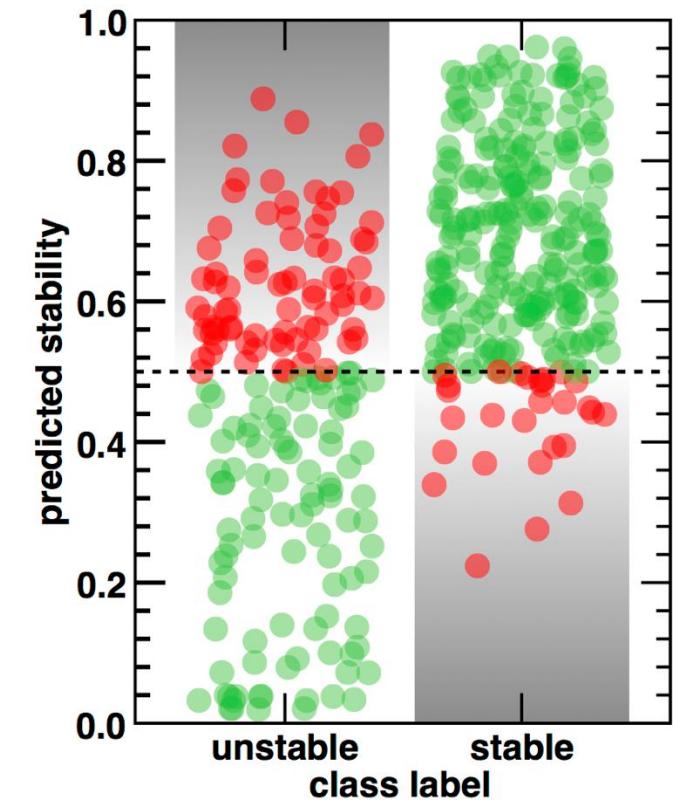
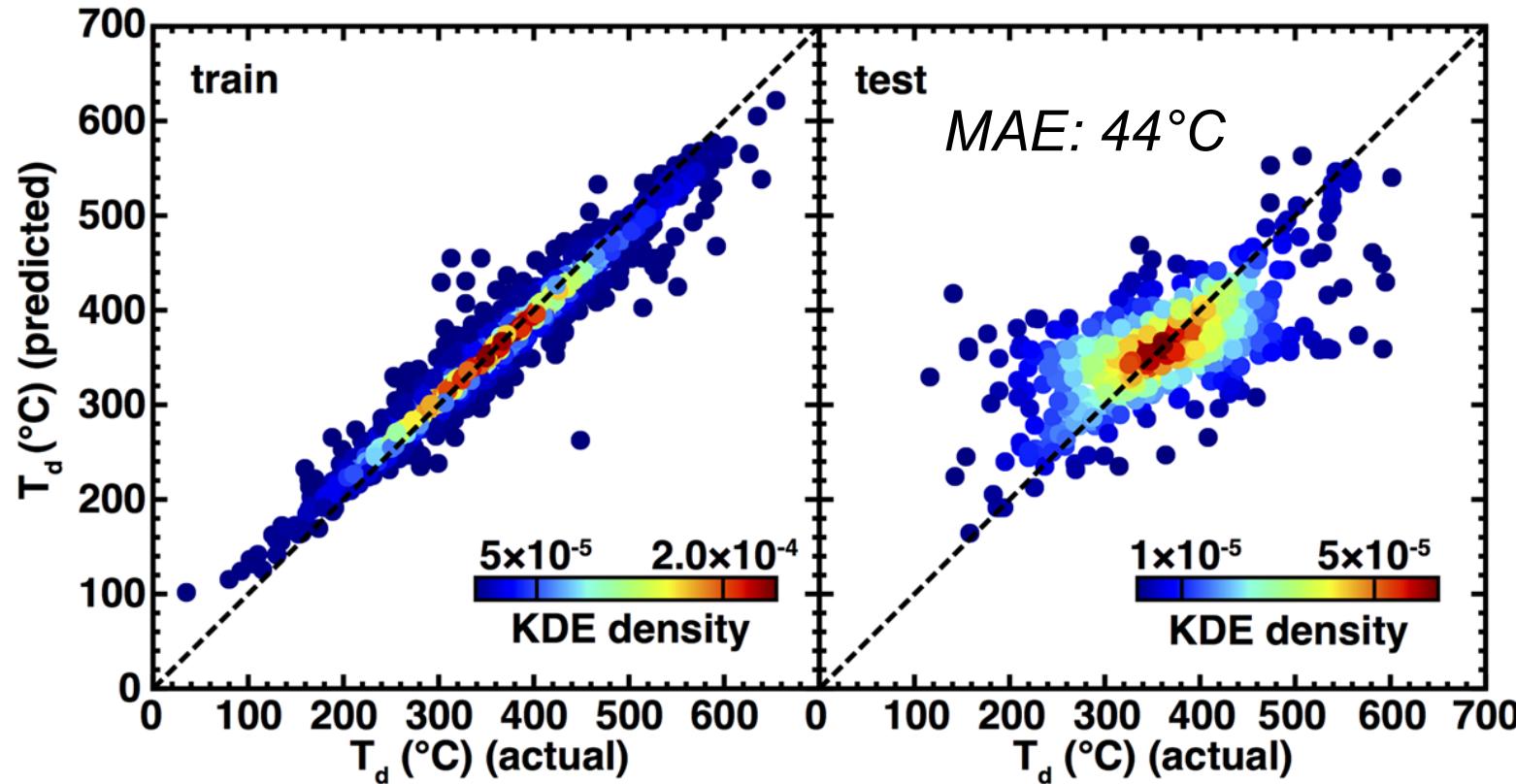
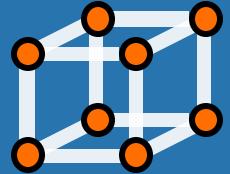
- Zr MOFs assumed to be stable in the literature



- combination of Zr nuclearity and linker identity matter

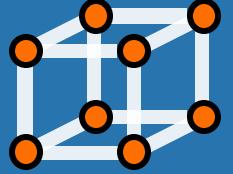
Analysis over large data set unveils limits of generality by conclusions drawn during small studies

Models Trained on Experimental Data Predict Stability

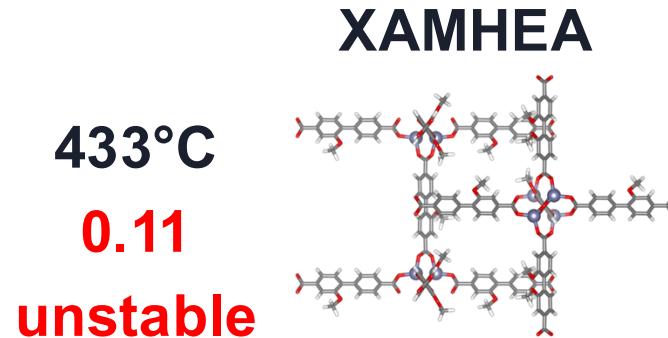


zero cost prediction
on MOF stability

Models from labeled data enable
predictions on new MOFs

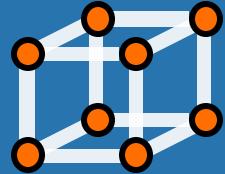


Engineering Stability into a MOF with ML Models



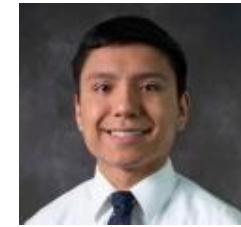
Hypothetical changes to
MOFs shows how models
think stability will improve
→ tangible suggestions

Using the Community to Improve ML Models



MOFSimplify Source Code MOF Code Dark mode

The MOFSimplify homepage features a prominent 3D molecular model of a MOF structure composed of black, red, and blue spheres. Below the model, the text "MOF Simplify" is displayed, along with "The Kulik Group at MIT" and "Powered by molSimplify". Navigation links include "Main", "Visualization", "Component Analysis", and "Data Upload". A call-to-action section asks "1) Select a MOF (metal-organic framework) for analysis" with options: "Example MOF", "Custom (upload cif file)", and "Building block assembly". A note below states "May MOFSimplify store information on your MOFs? Yes. No." Site credits at the bottom mention "Site developed and maintained by the Kulik Group at MIT" and "Contact: mofsimplify@mit.edu".

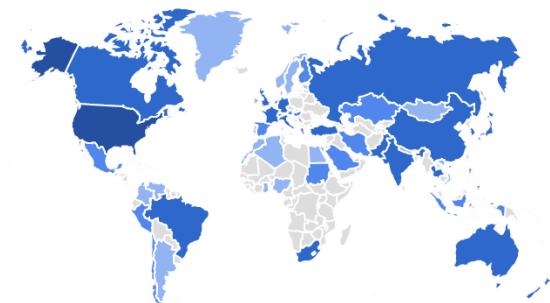


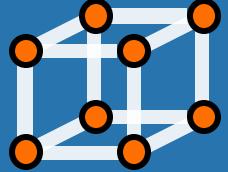
Gianmarco
Terrones

mofsimplify.mit.edu

Users New users
1.4K 1.4K

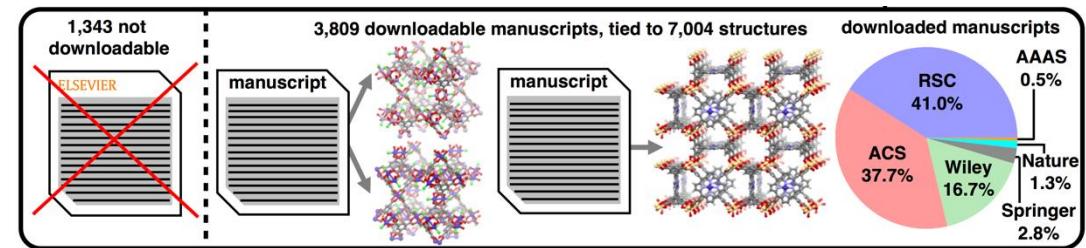
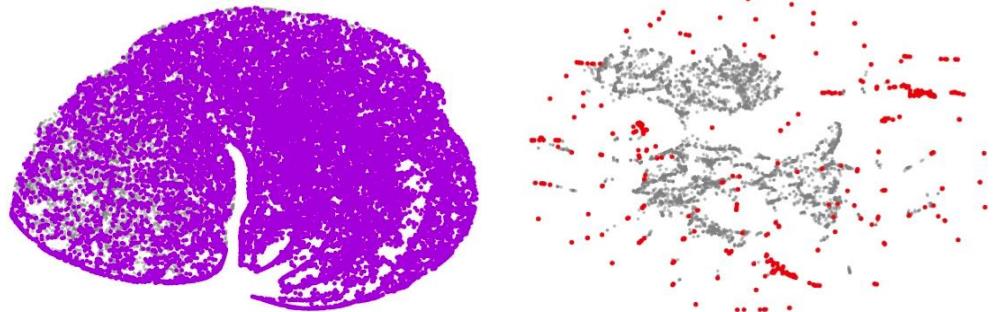
Event count ▾ Average engagement time
15K 1m 45s



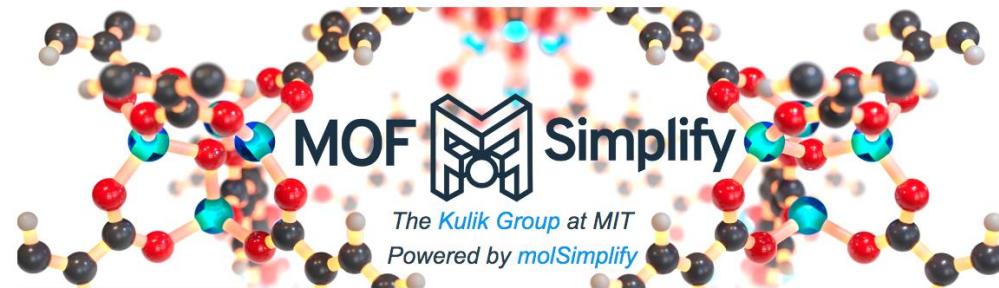
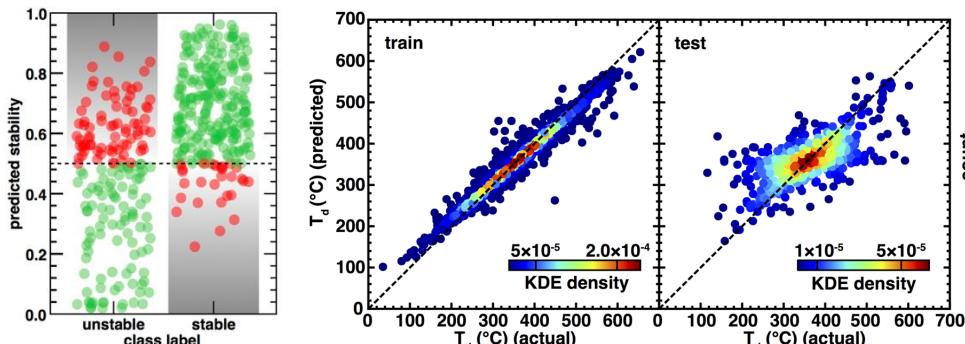


Conclusions

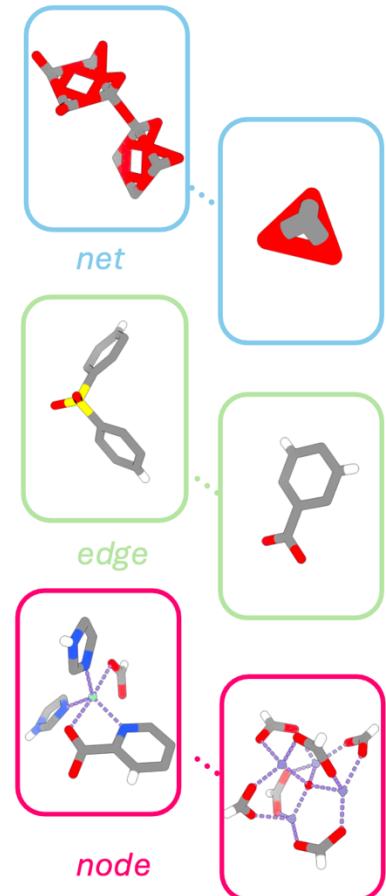
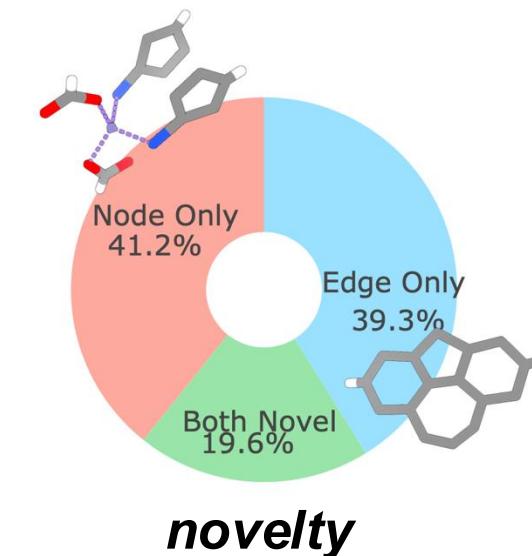
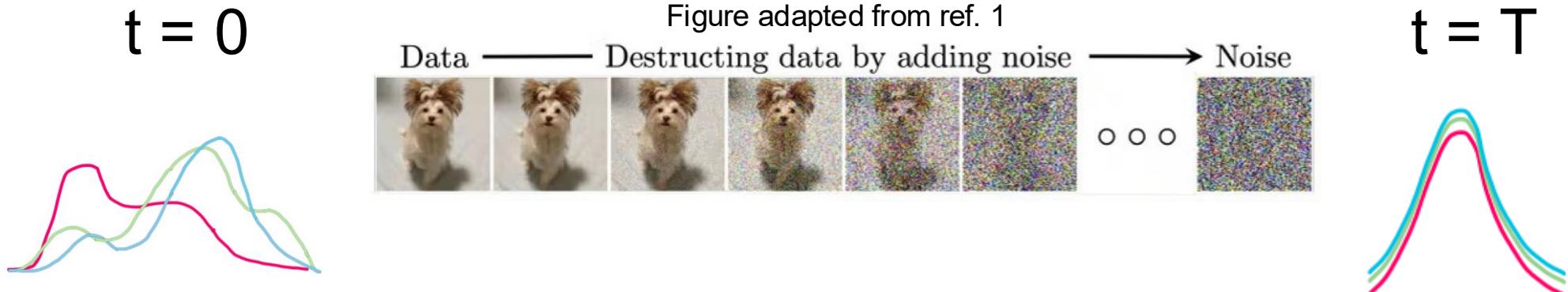
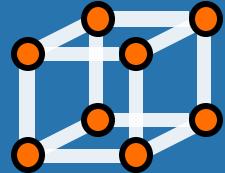
- Developing interpretable representations for MOF chemistry leads to insights on materials diversity and new methods for extracting trends from experimental data



- We can make tangible suggestions to improve stability, construct new hypothetical MOFs that are more likely to be stable, and solicit community feedback for improvements

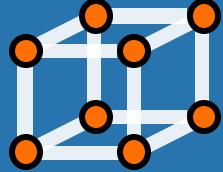


Generative Diffusion Enables Direct Sampling of Synthesized MOF Distribution



¹Sohl-Dickstein, J.; Ganguli, S. et al. *ICML*, June 2015

²Duan, C.; Nandy, A.; Dou, J.; Du, Y. et al. arXiv:2505.08531v1



Acknowledgments

Funding



ONR



GRFP

CBET



YFA



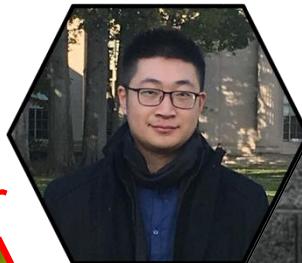
DOE EFRC



Ph.D.
Advisor



Co-workers on MOF projects



Chenru
Duan



Prof.
Heather J.
Kulik



Naveen
Arunachalam
Gianmarco
Terrones



Prof. Shuwen
Yue



Husain
Adamji



Postdoc
Advisor



Prof. Suri
Vaikuntanathan

SCHMIDT FUTURES



Collaborators on MOF projects

Prof. Y.G.
Chung



Prof. S. M.
Moosavi



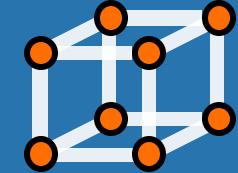
Prof. B.
Smit



Thank you for listening!

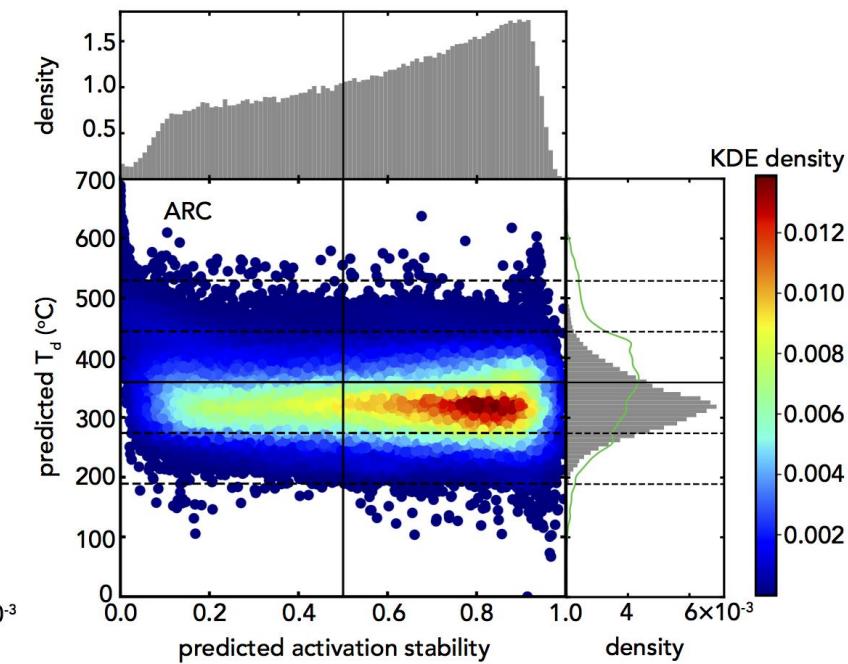
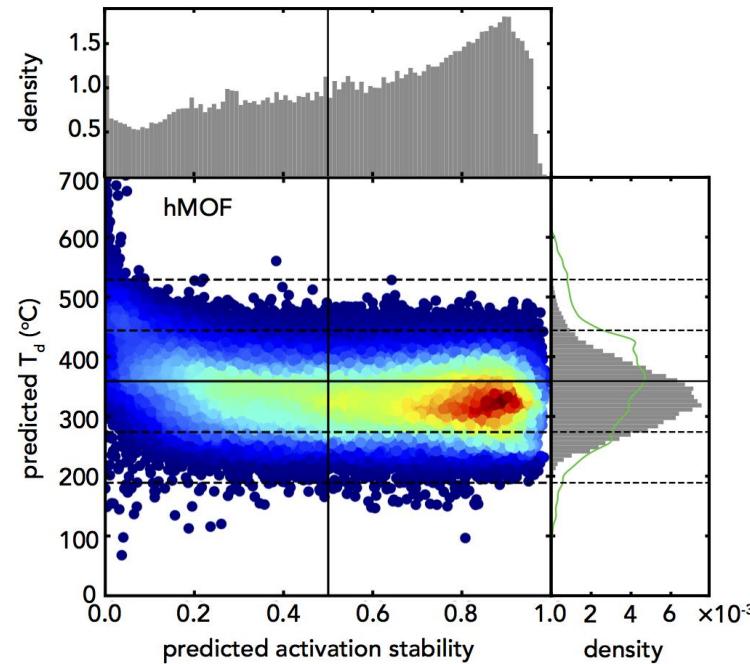
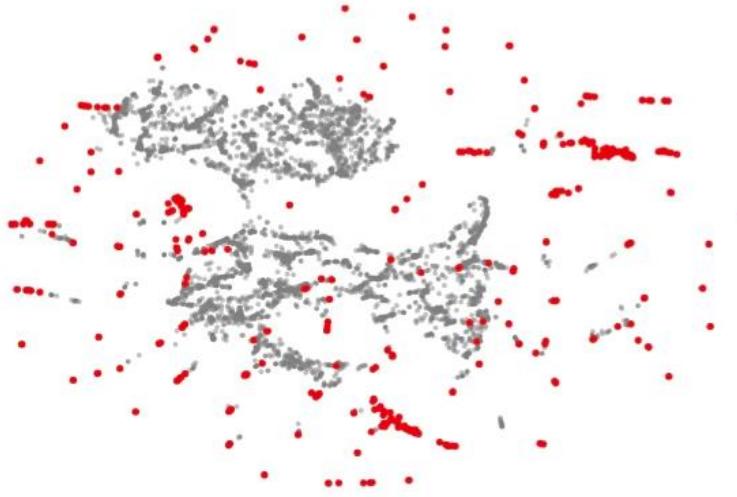


Questions?



Less Metal Diversity Leads to Reduced Stability

hypothetical databases lack metal diversity



- Average MOF in hypothetical databases are likely to be less stable than synthesized MOFs

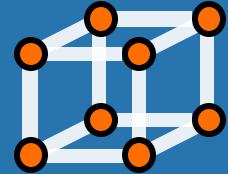
Need to harness stable building blocks to enrich MOF stability for use in thermal applications

¹Wilmer, C.E.; Leaf, M.; Lee, C.Y.; Farha, O.K.; Hauser, B.G.; Hupp, J.T.; Snurr, R.Q. *Nat. Chem.* **2012**, *4*, 83-89.

²Burner, J.; Luo, J.; White, A.; Mirmiran, A.; Kwon, O.; Boyd, P.G.; Woo, T.K. *et al. Chem. Mater.* **2023**, *35*, 900-916.

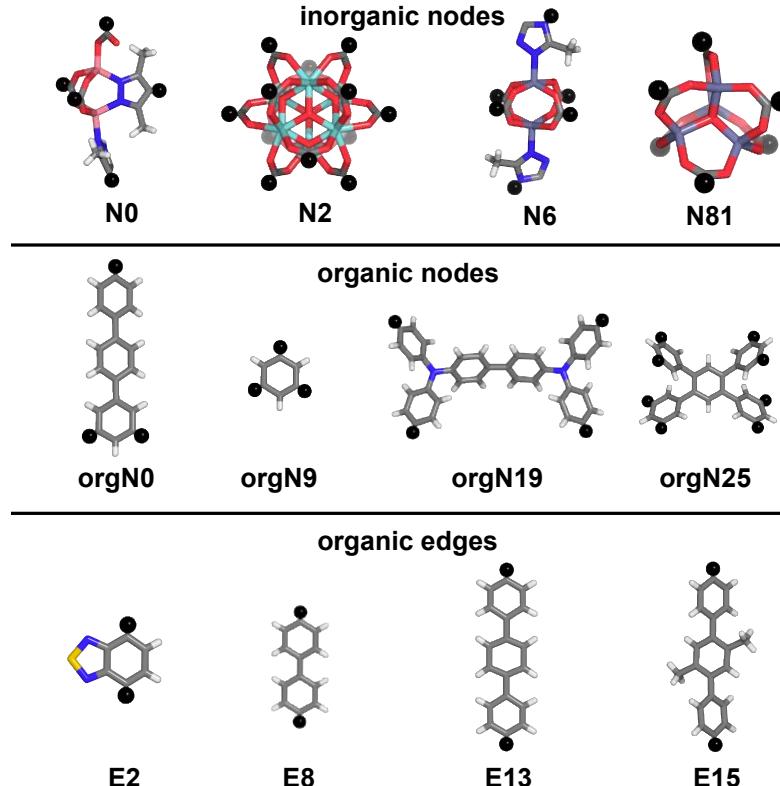
³Nandy, A.; Yue, S.; Oh, C.; Duan, C.; Terrones, G.; Chung, Y.G.; Kulik, H.J. *Matter* **2023**, *6*, 1585-1603.

Recombining Building Blocks from Stable MOFs

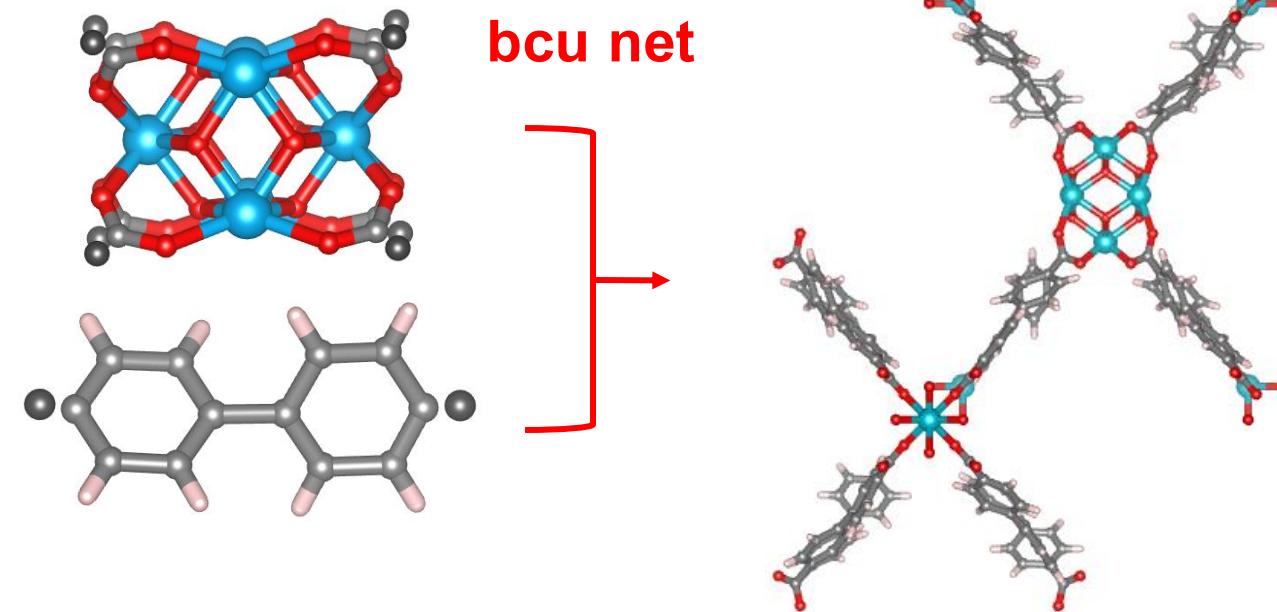


- Recombining pieces of stable MOFs makes new MOFs that have yet to be synthesized

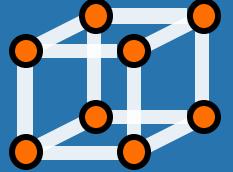
representative building blocks



combine building blocks in new arrangement

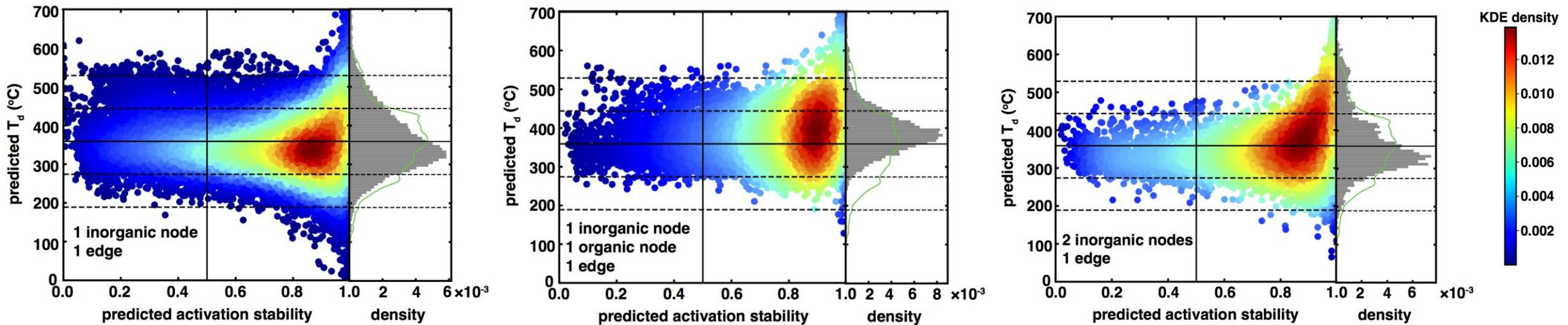


We can recombine existing building blocks to make new MOFs



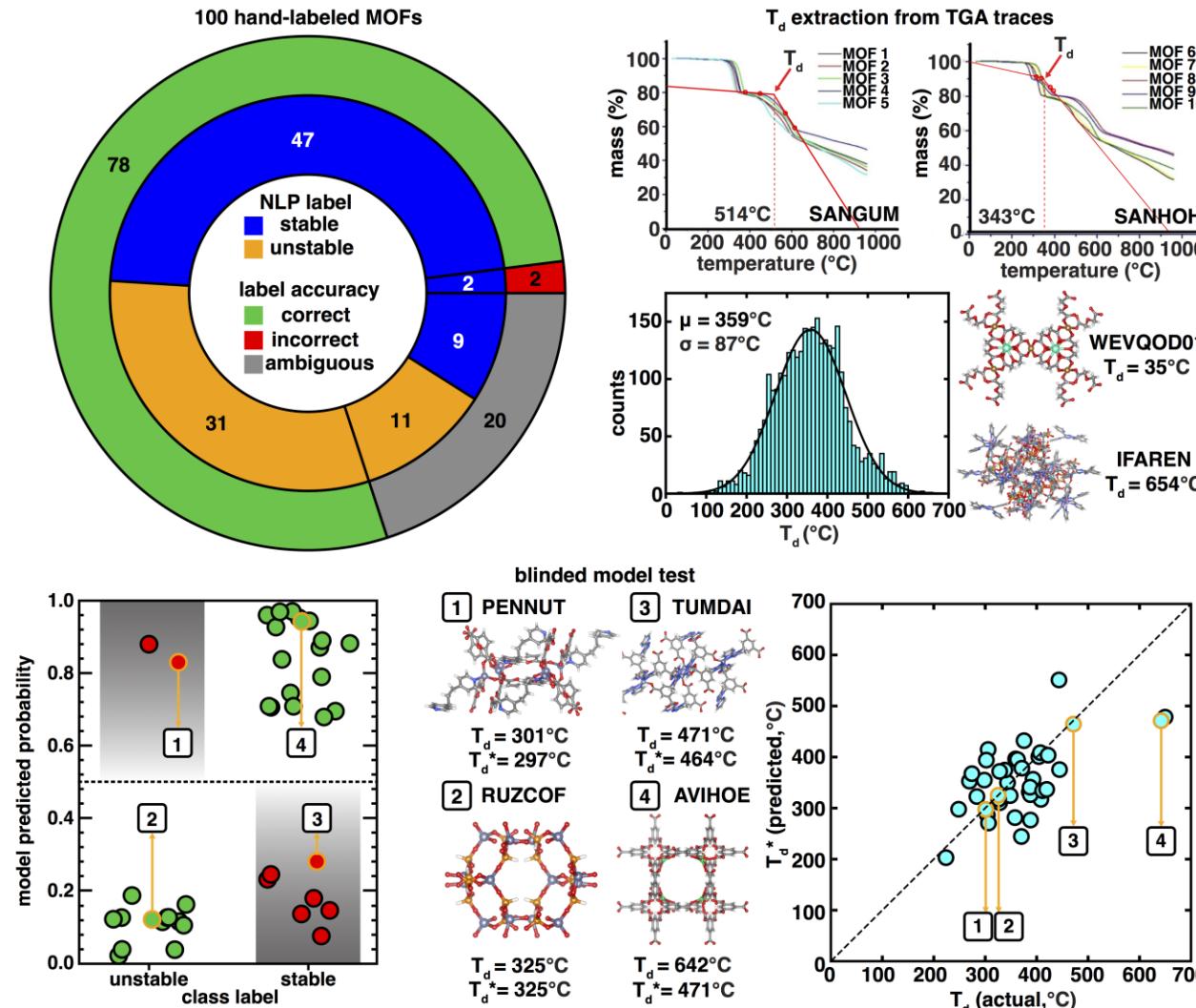
Improving Diversity Enriches Stable MOFs

- We can design more stable MOFs by improving metal and net diversity



MOF databases that incorporate stable building blocks show orders of magnitude improvement in predicted experimental stability measures (<1% to ~20%) relative to other hypothetical databases.

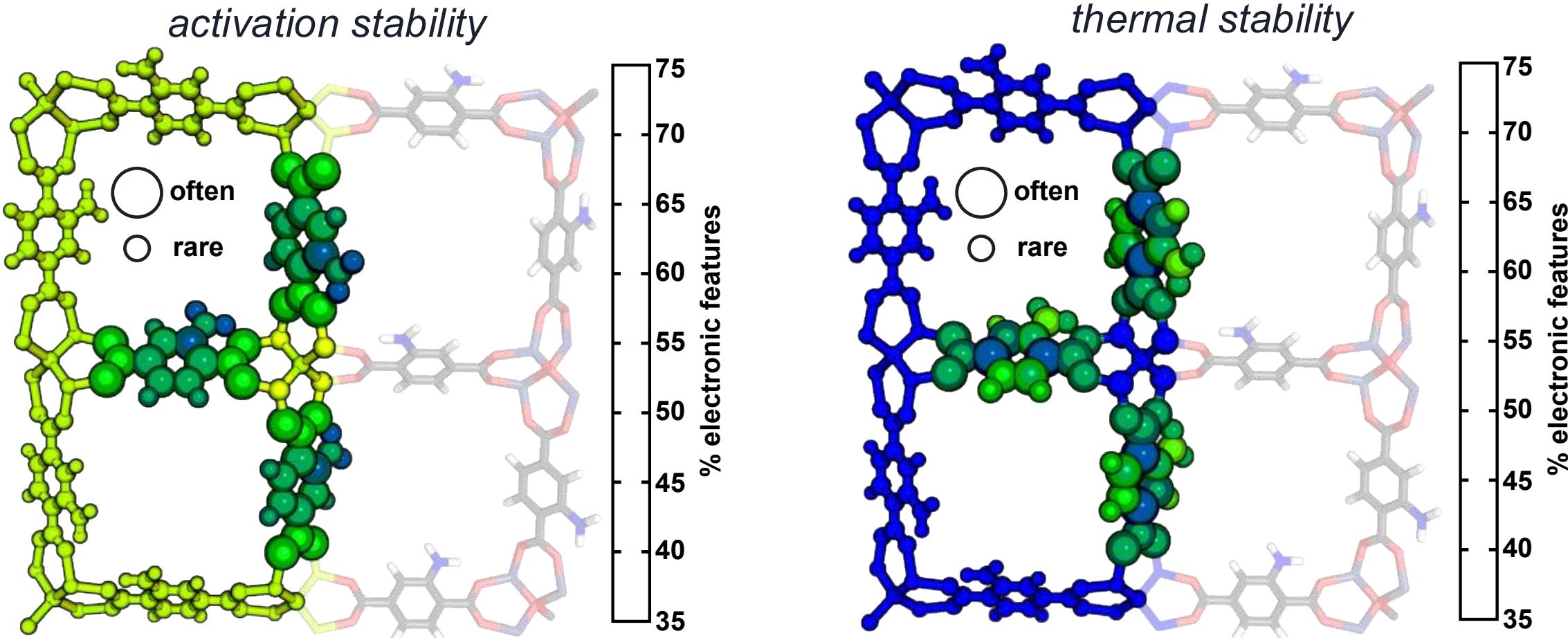
Accuracy of MOF Labels



- NLP can get the MOF label correct on ~80% of the data on a random set of 100 MOFs. Some labels that are assigned are inconclusive (referring to the wrong type of stability, or no concrete statement on stability)

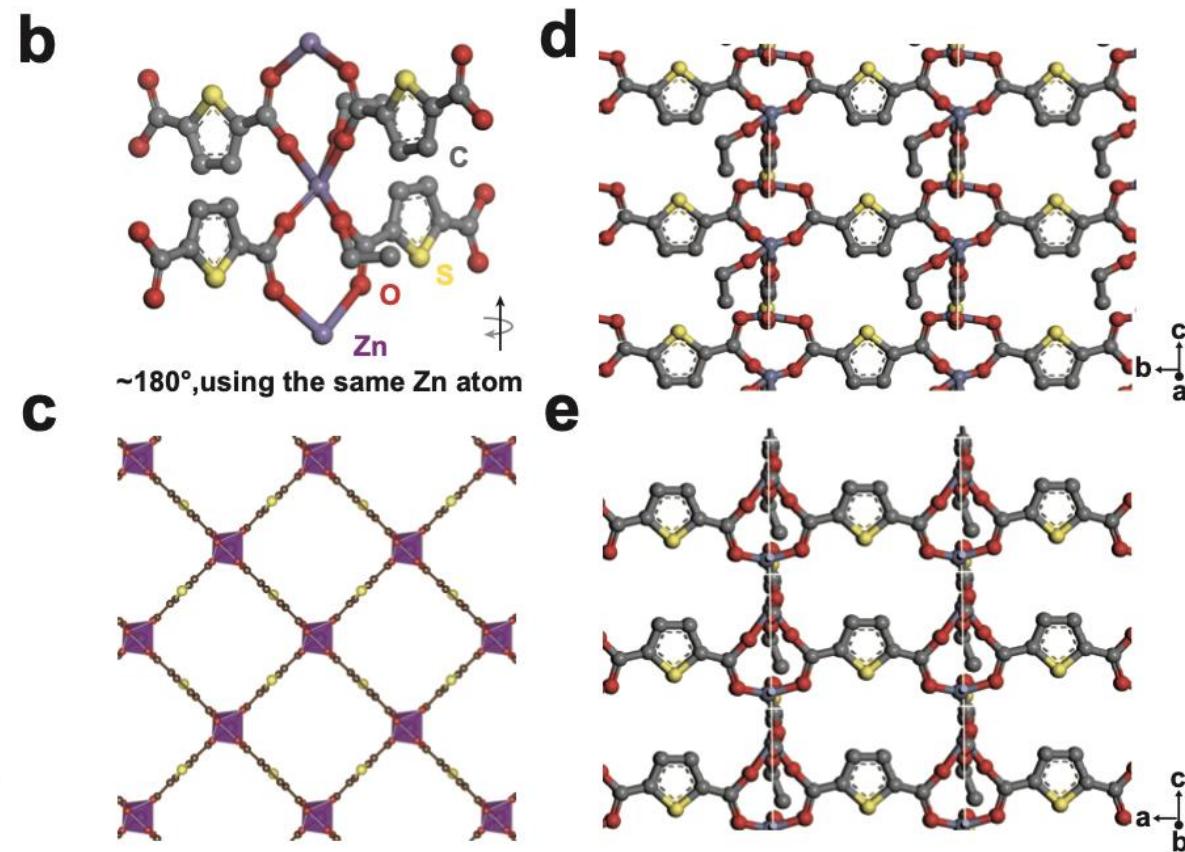
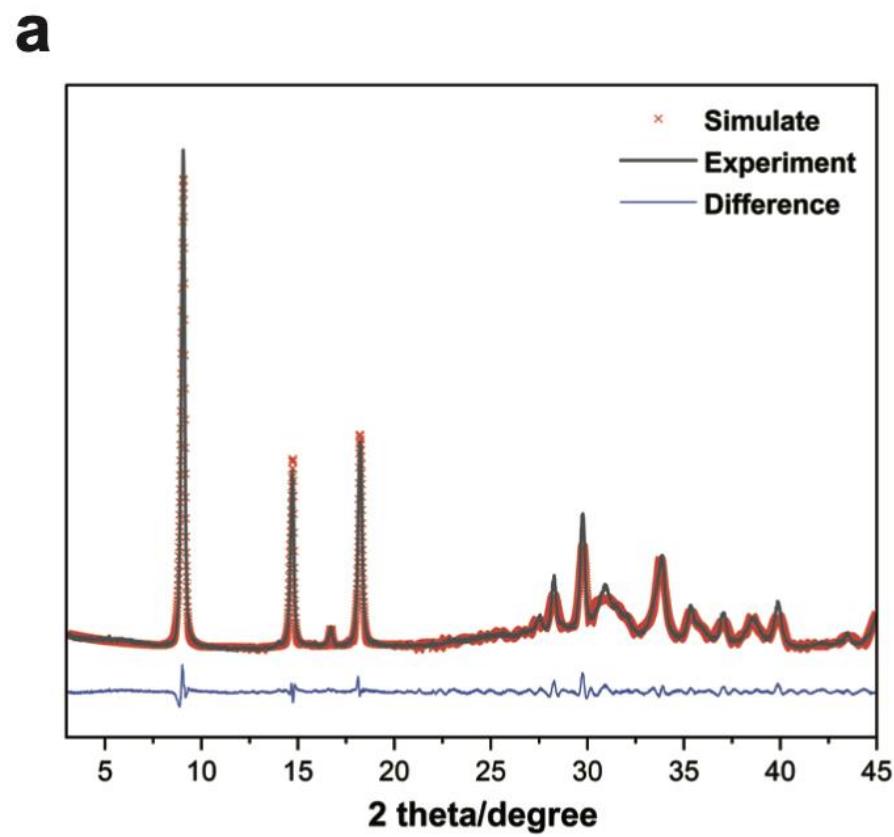
¹Nandy, A.; Terrones, G.; Duan, C.; Arunachalam, N.; Kulik, H. J. *Sci. Data.* **2022**, 9, 74.

Feature Selection Insights



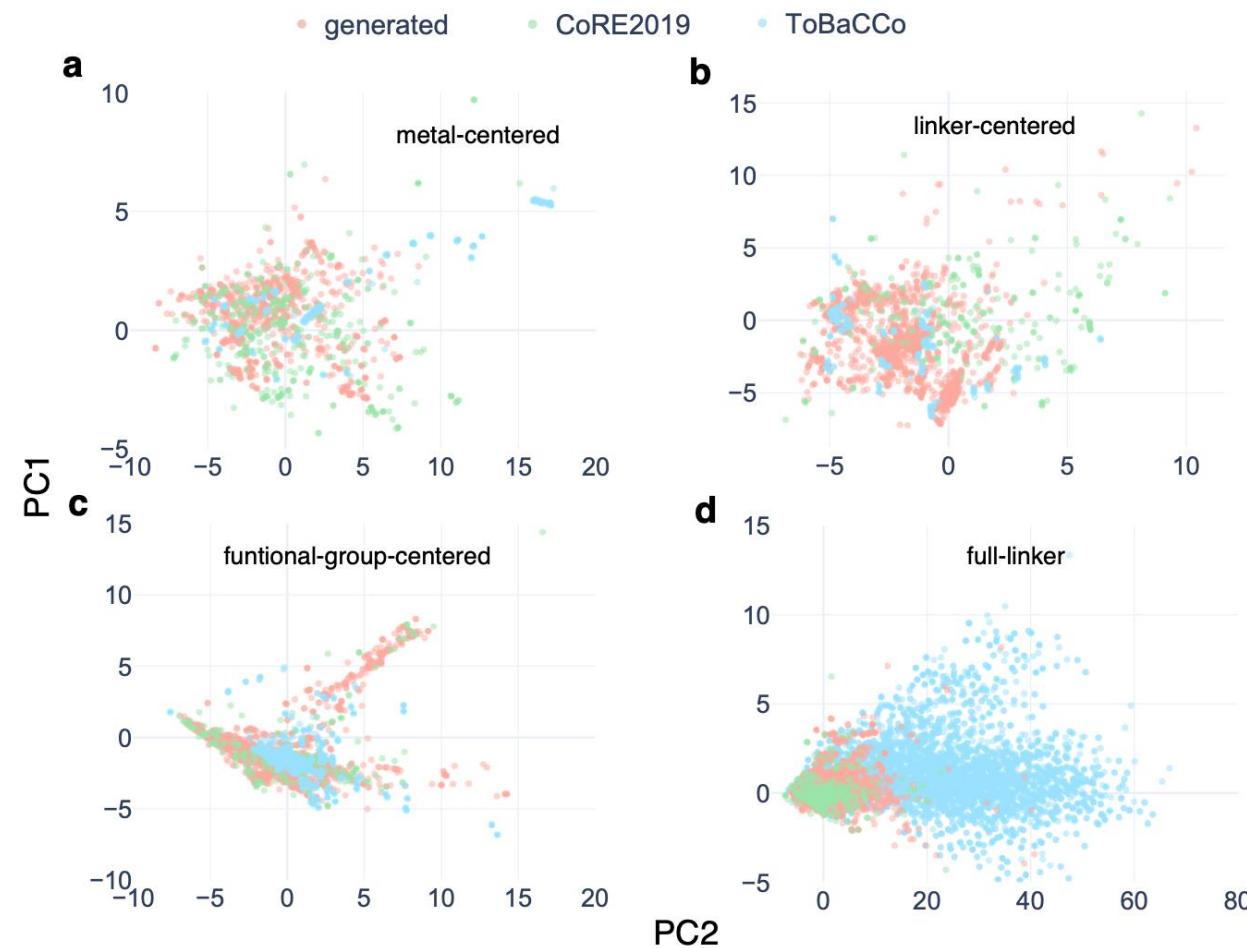
- Activation and thermal stabilities are primarily governed by linker differences

Experimental Synthesis of a Generated MOF

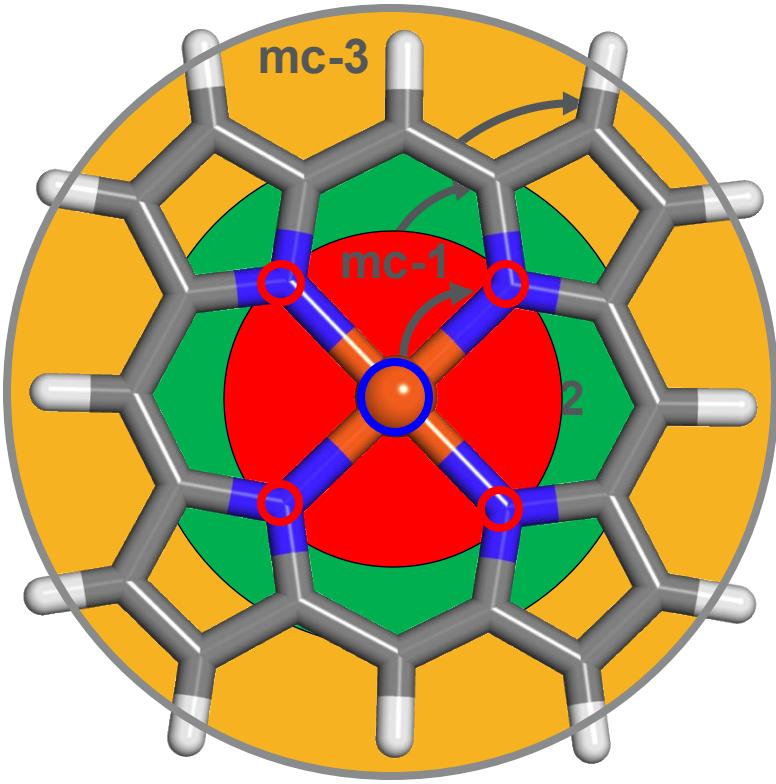


MOFs synthesized by Sizhan Liu and Jinhui Dou

Generated MOFs Match the CoRE MOF Distribution

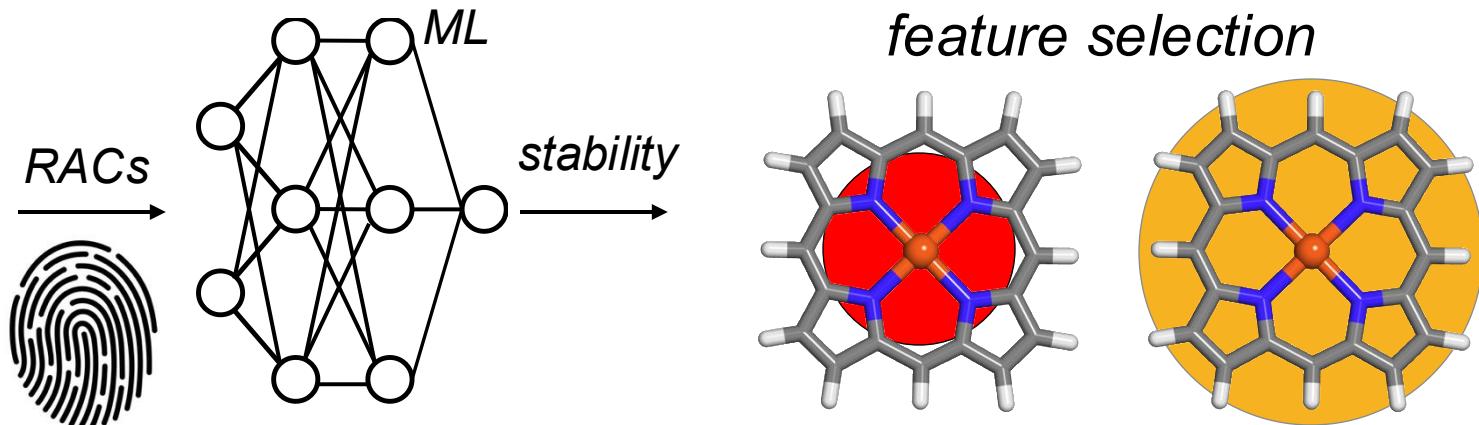


Revised Autocorrelations (RACs) are a Graph Fingerprint



Revised autocorrelations (RACs)¹

- Products² and differences on molecular graph
- Z , $S(r_{\text{cov}})$, topology (T), identity (I), x
- Metal-centered, ligand-centered, or global
- Various depths
- 155 total connectivity-only features



RACs are a zero-cost fingerprint → no calculations necessary for new MOFs

¹Janet, J.P. and Kulik, H.J. *J. Phys. Chem. A*. **2017**, *121*, 8939-8954.

²Broto, P.; Moreau, G.; Vandycke, C. *Eur. J. Med. Chem.* **1984**, *19*, 71-78.