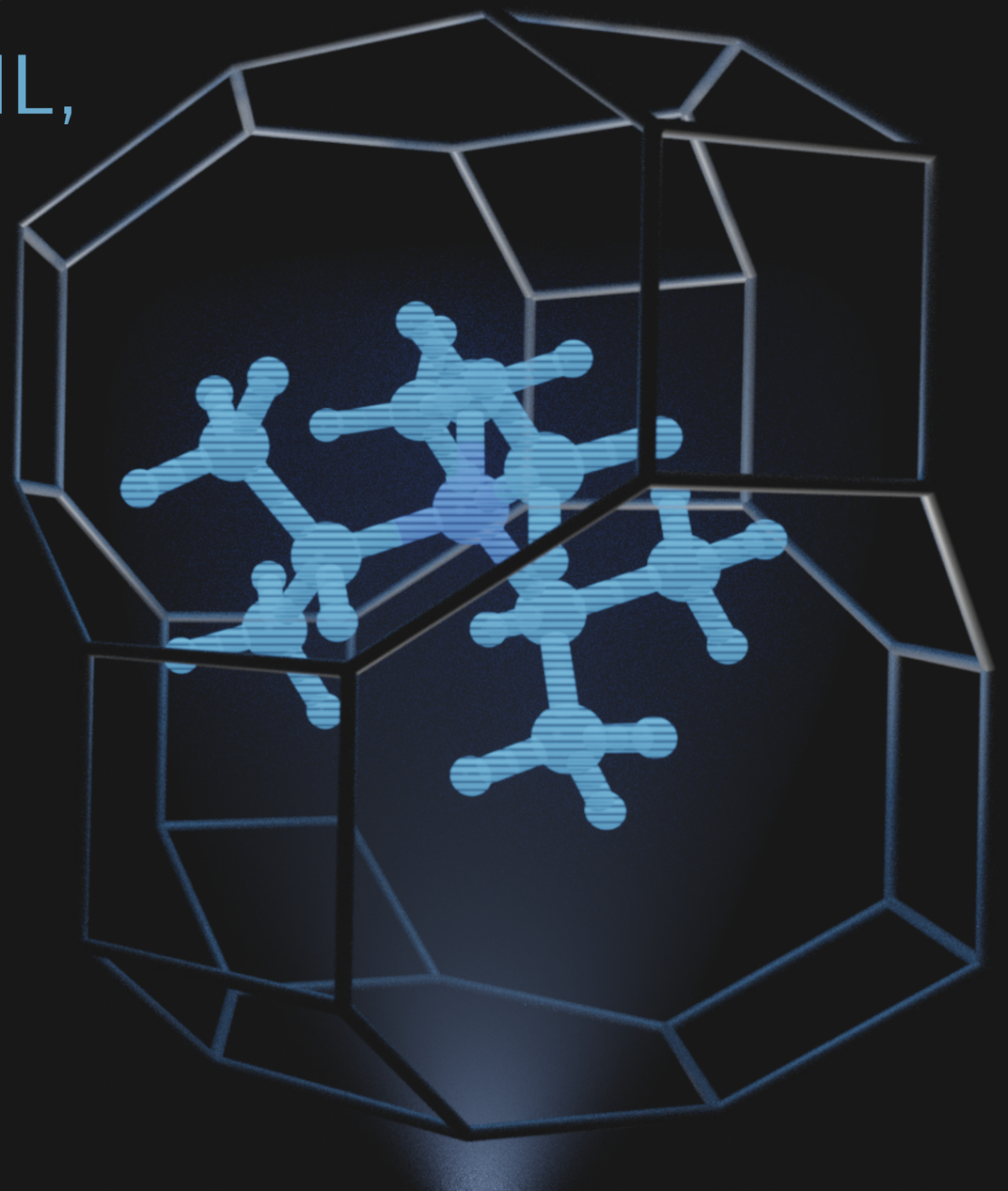
Exploring materials synthesis landscapes with high-throughput simulations, ML, and adversarial attacks



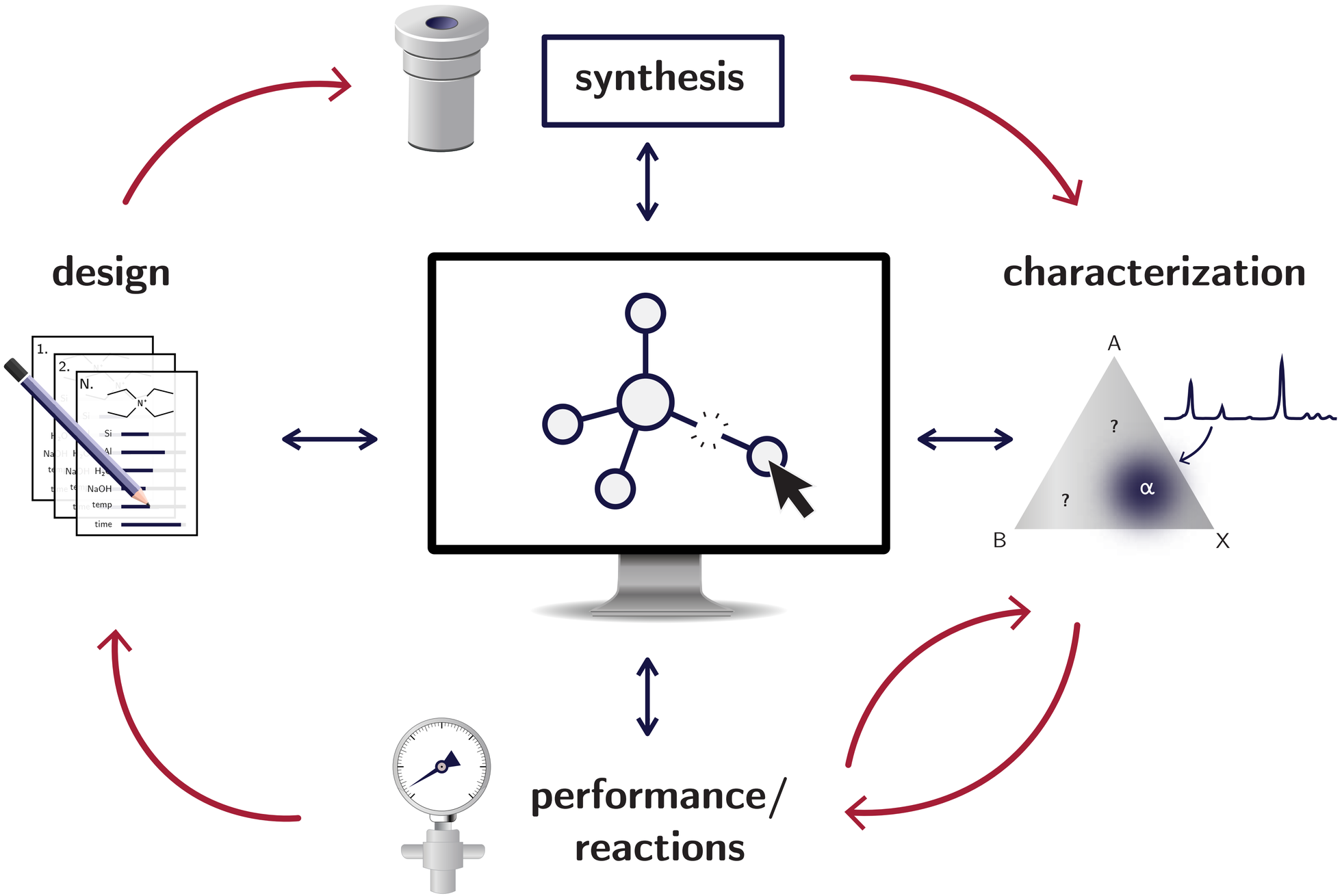
Daniel Schwalbe-Koda

Lawrence Livermore National Laboratory
https://dskoda.com, <dskoda@llnl.gov>

Brookhaven AI/ML Seminar
Dec. 13, 2022

LLNL-PRES-840662

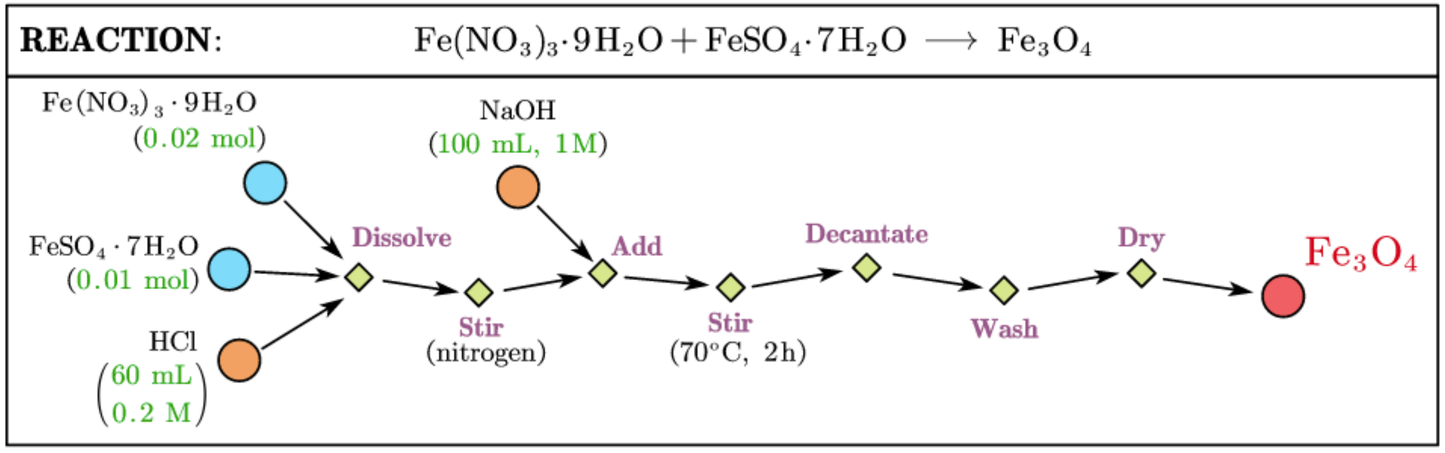
Nowadays, computation plays a role in all areas of materials discovery



Synthesis is complex...

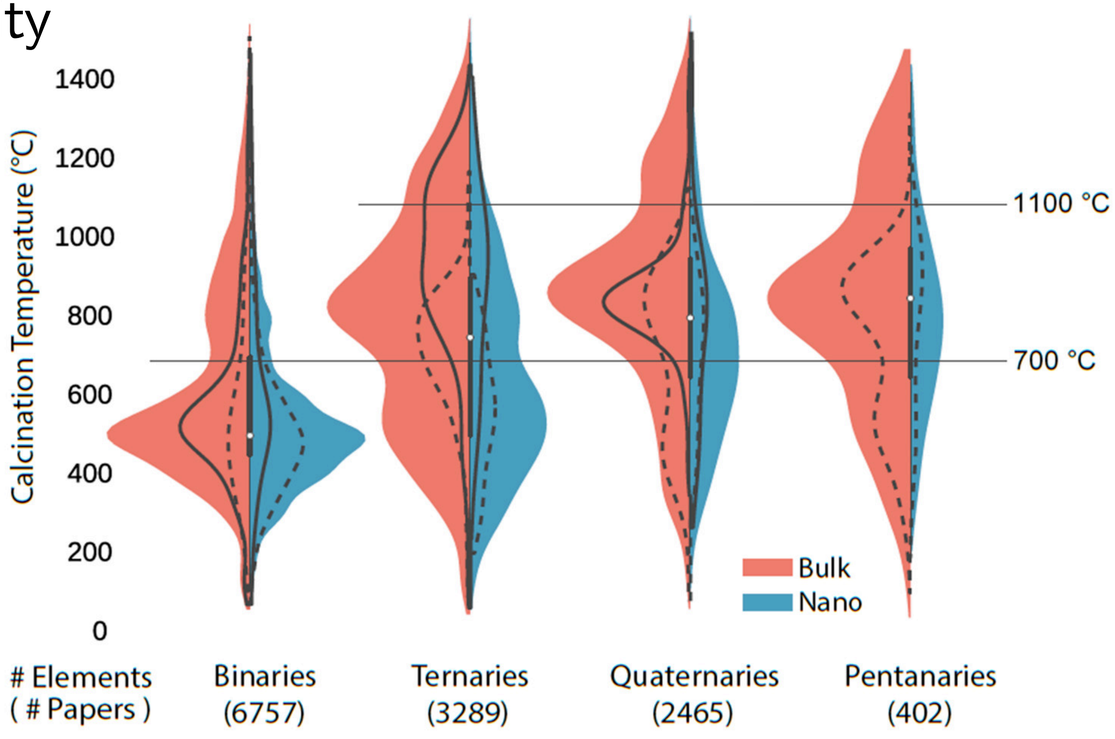
SYNTHESIS PARAGRAPH:

Fe_3O_4 were prepared by co-precipitation method. $Fe(NO_3)_3 \cdot 9H_2O$ (0.02 mol) and $FeSO_4 \cdot 7H_2O$ (0.01 mol) were **dissolved** in HCl (60 mL, 0.2 M) with vigorous **stirring** under nitrogen. NaOH solution (100 mL, 1.0 M) was then **added** dropwise to the suspension. **Stirring** was continued at 70 °C for another 2 h.The resulting precipitate (Fe_3O_4) was collected by magnetic **decantation**, **washed** three times with DI water, and then **dried** in a desiccator.



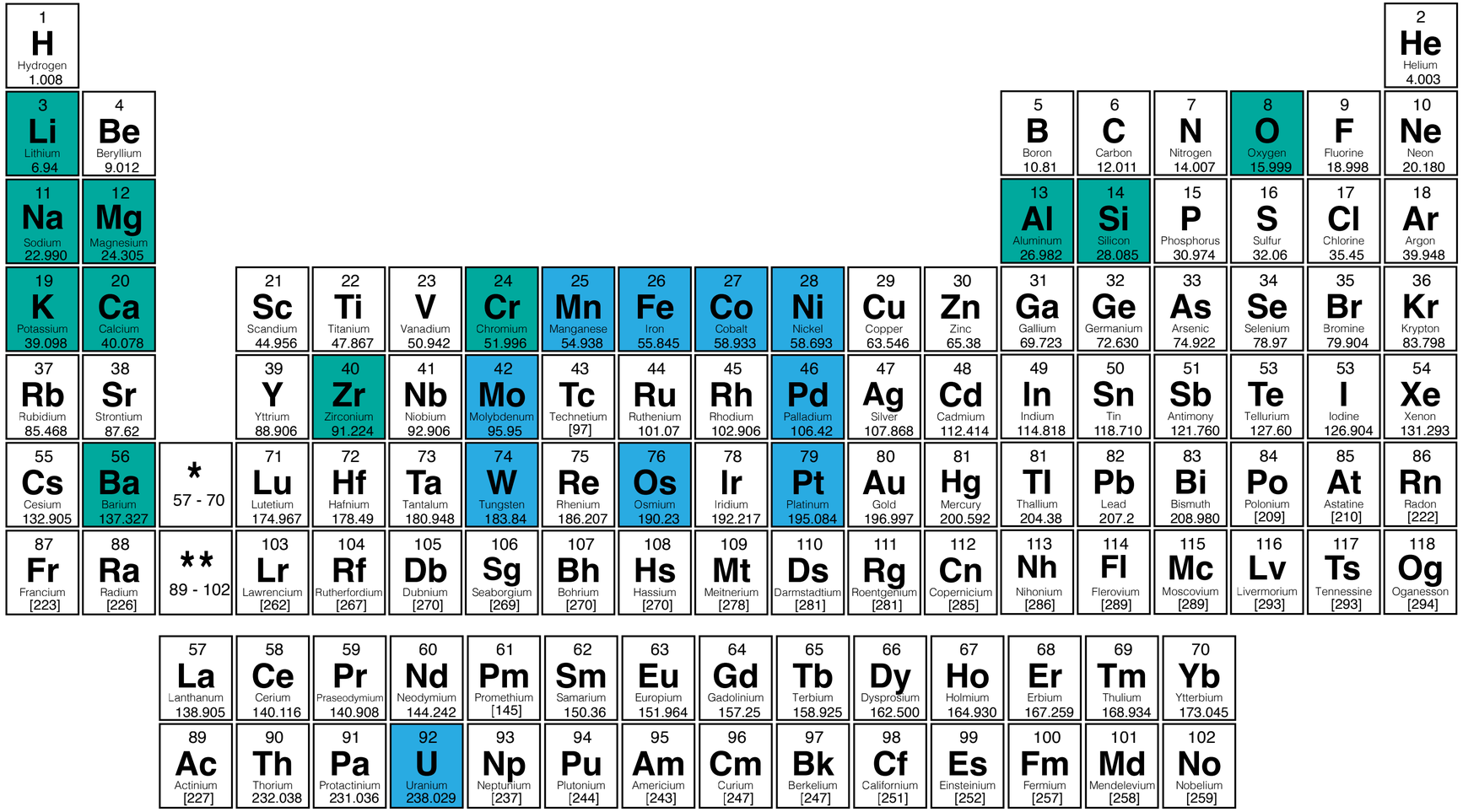
Z. Wang, ..., G. Ceder. *Sci. Data* (2022)

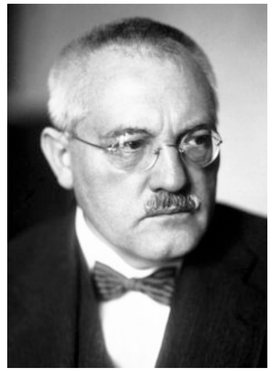
...and has a lot of variability



E. Kim, ..., E. Olivetti. *Chem. Mater.* (2017)

...however, synthesis screening is not a new problem (see 1910)





C. Bosch

A. Mittasch

Between 1910-1912:

6,500 tests with 2,500 substances

Between 1910-1922:

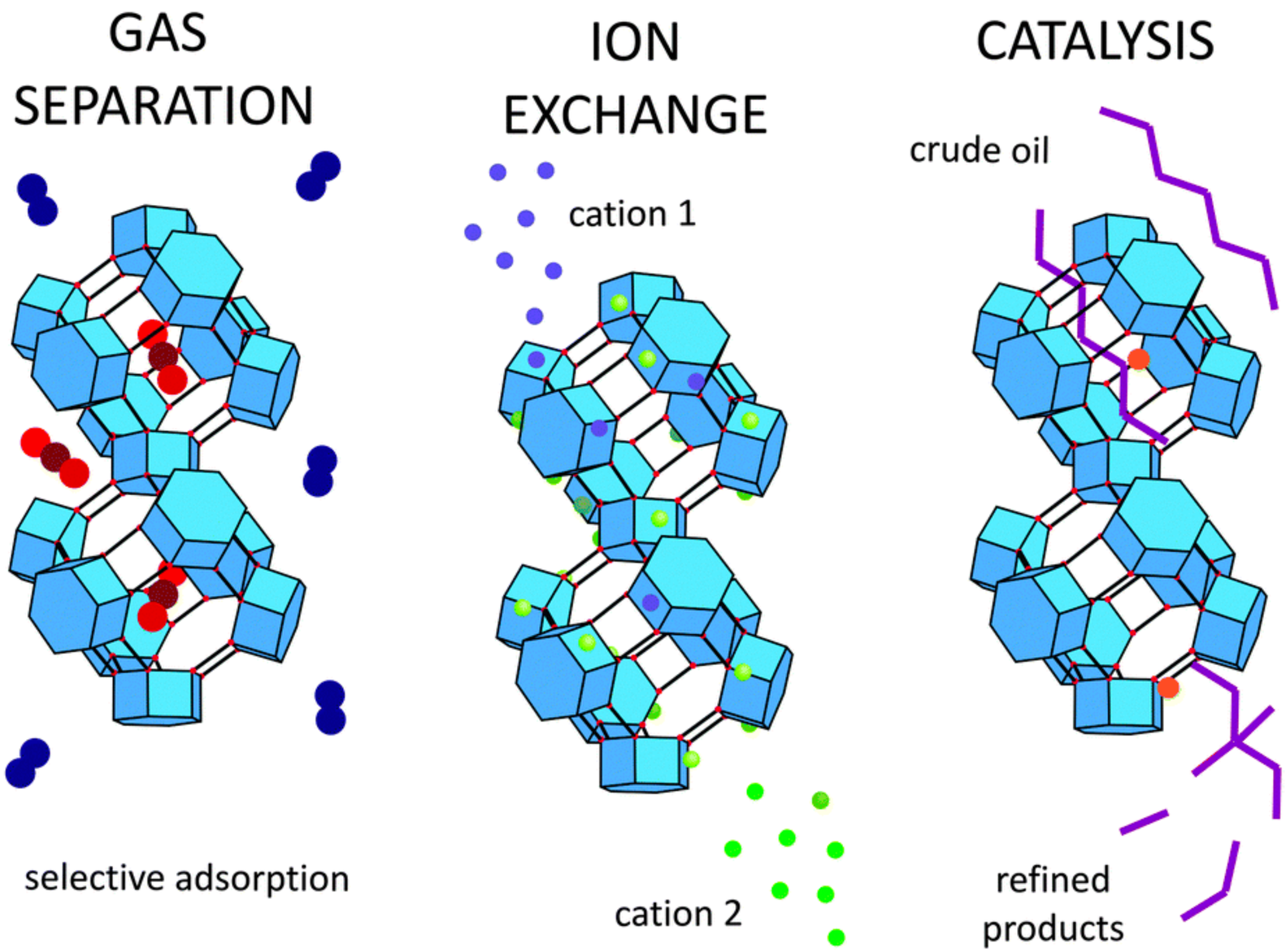
20,000 tests with 4,000 substances

Metallic catalysts explored by Bosch, Mittasch et al.

Alloys/oxides explored by Bosch, Mittasch et al.

V. Smil. *Enriching the Earth* (2001)

Case study of this talk: zeolites



Market for zeolites

18% (in value) of all catalysts

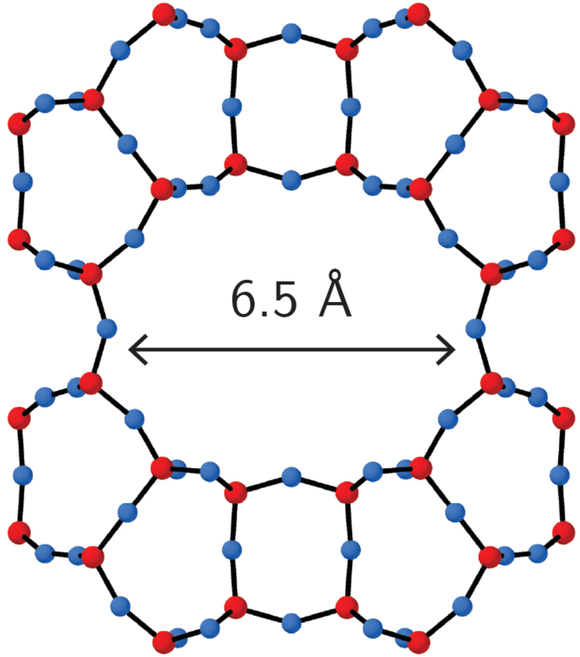
zeolite products are valued in trillions USD

Applications in emission control, CO_2 capture, biomass conversion etc.

V. Van Speybroek *et al. Chem. Soc. Rev.* (2015)

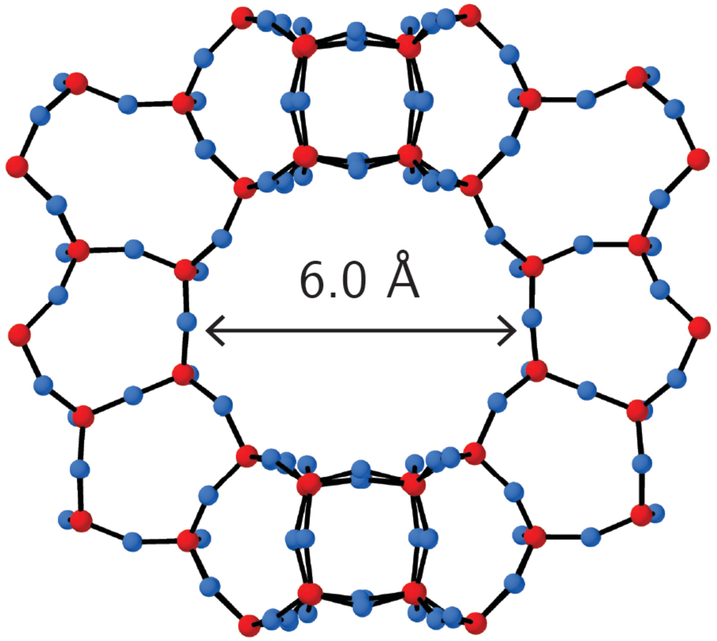
Grand View Research. *Catalyst Market Size & Share* (2020)
S. Abdo and T. Wilson. *Zeolites in Catalysis: Properties and Applications* (2017)

Synthesis controls zeolite polymorphism...



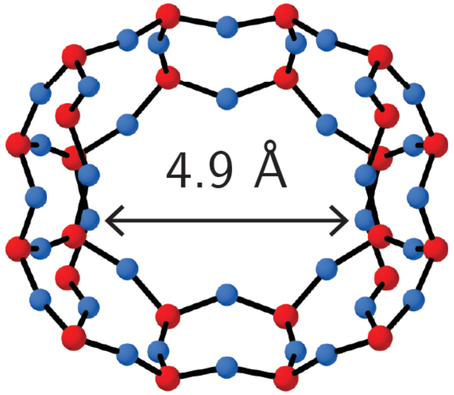
MOR

1D



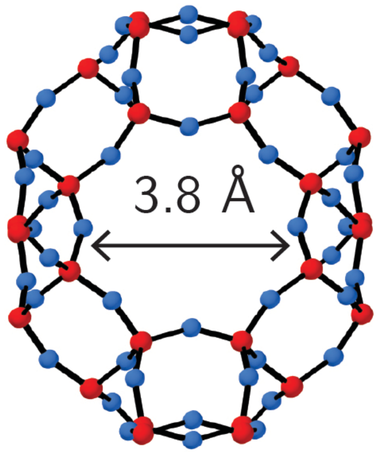
BEA

3D



MWW

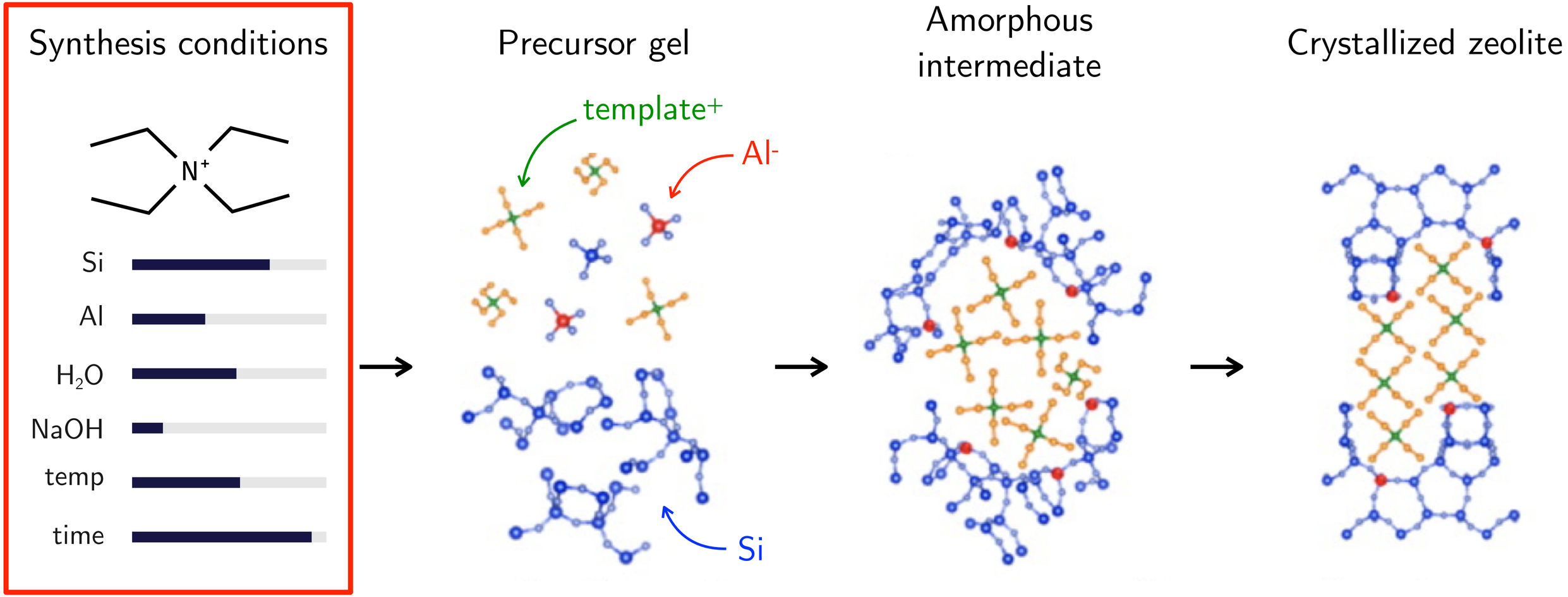
2D



CHA

3D

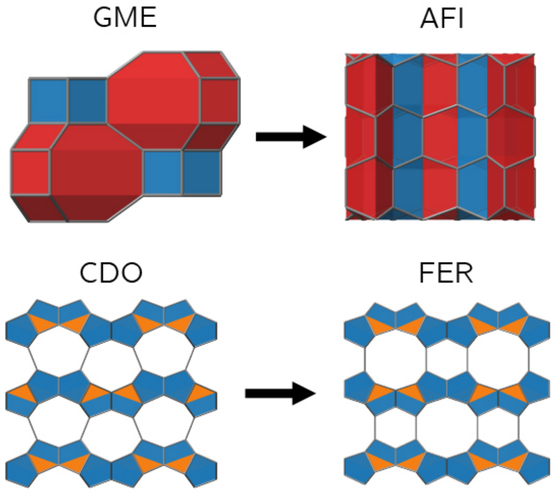
...but how are these polymorphs synthesized?



Crystallization image: T. Okubo *et al. J. Am. Chem. Soc.* (2015)

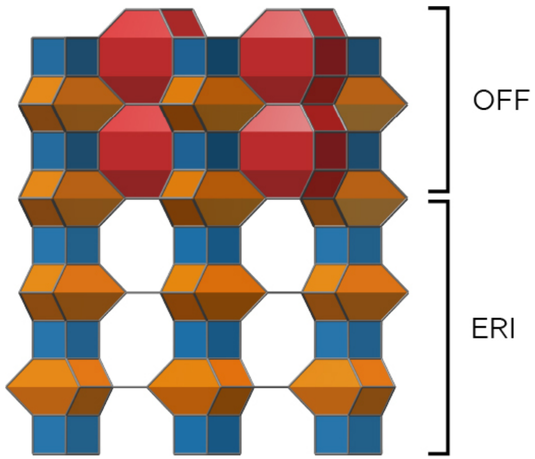
How to model synthesis in zeolites?

Phase transformations



Nat. Mater. **18**, 1177 (2019)

Intergrowth

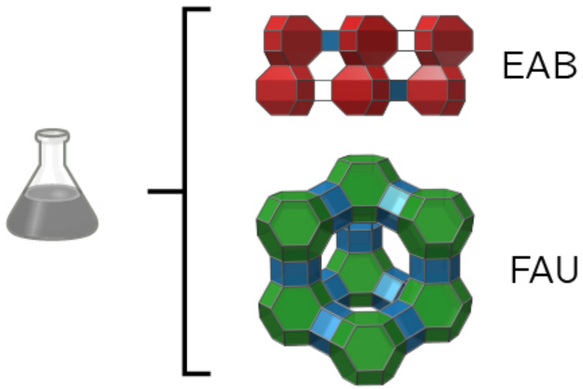


JPCL **12** (43), 10689 (2021)

Angew. Chem. **61** (28), e202201837 (2022)

Patent ES2829385 (2021)

Phase competition



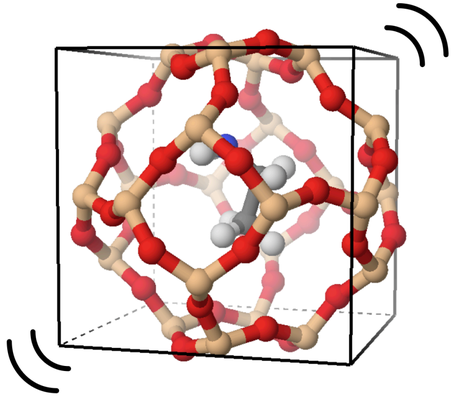
Science **374** (6565), 308 (2021)

JPCC **125** (5), 3009 (2021)

JCP **154**, 174109 (2021)

Chem. Mater. **34** (12), 5366 (2022)

Synthesis dynamics



Nat. Commun. **12**, 5104 (2021)

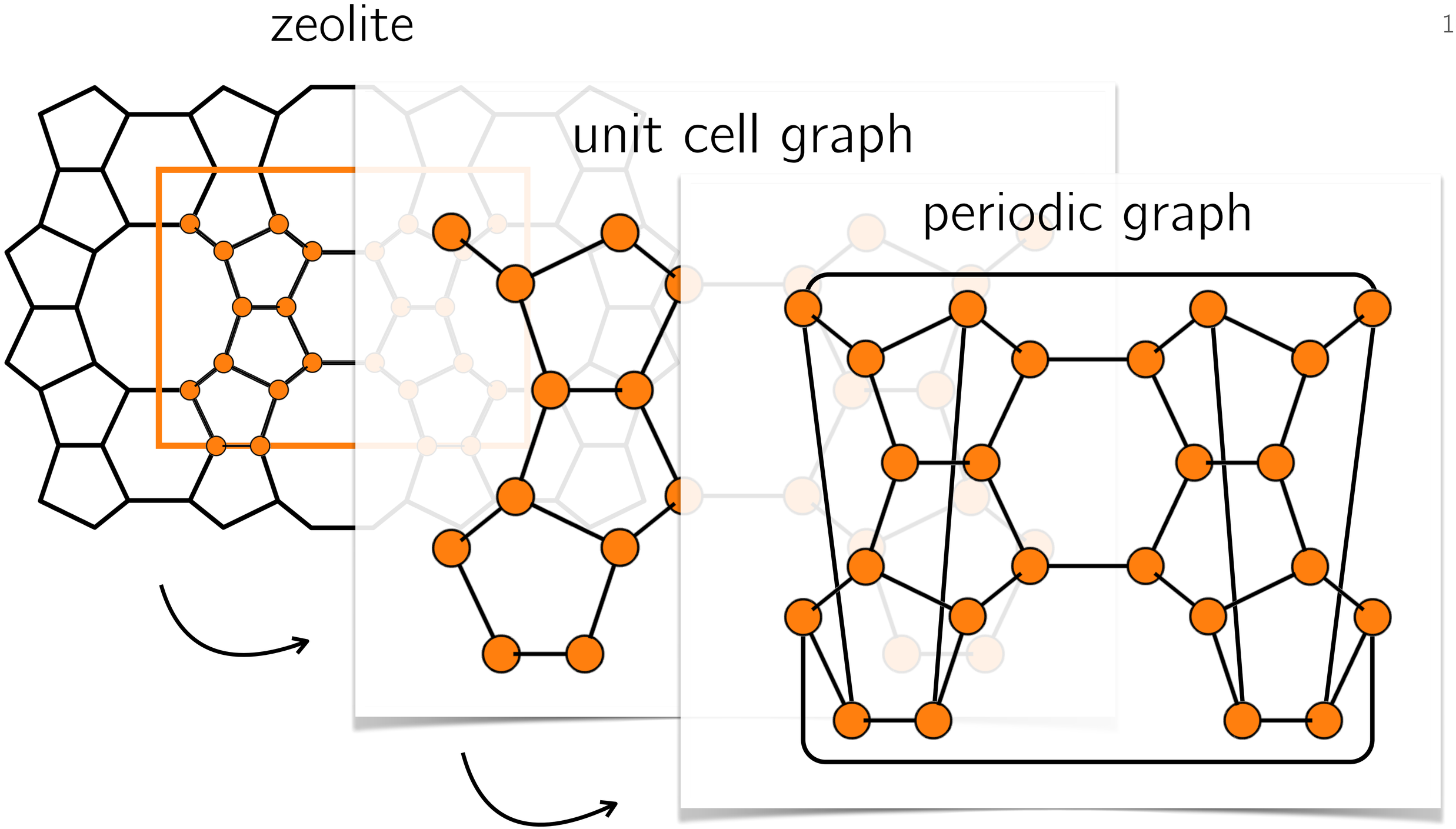
data science

ML

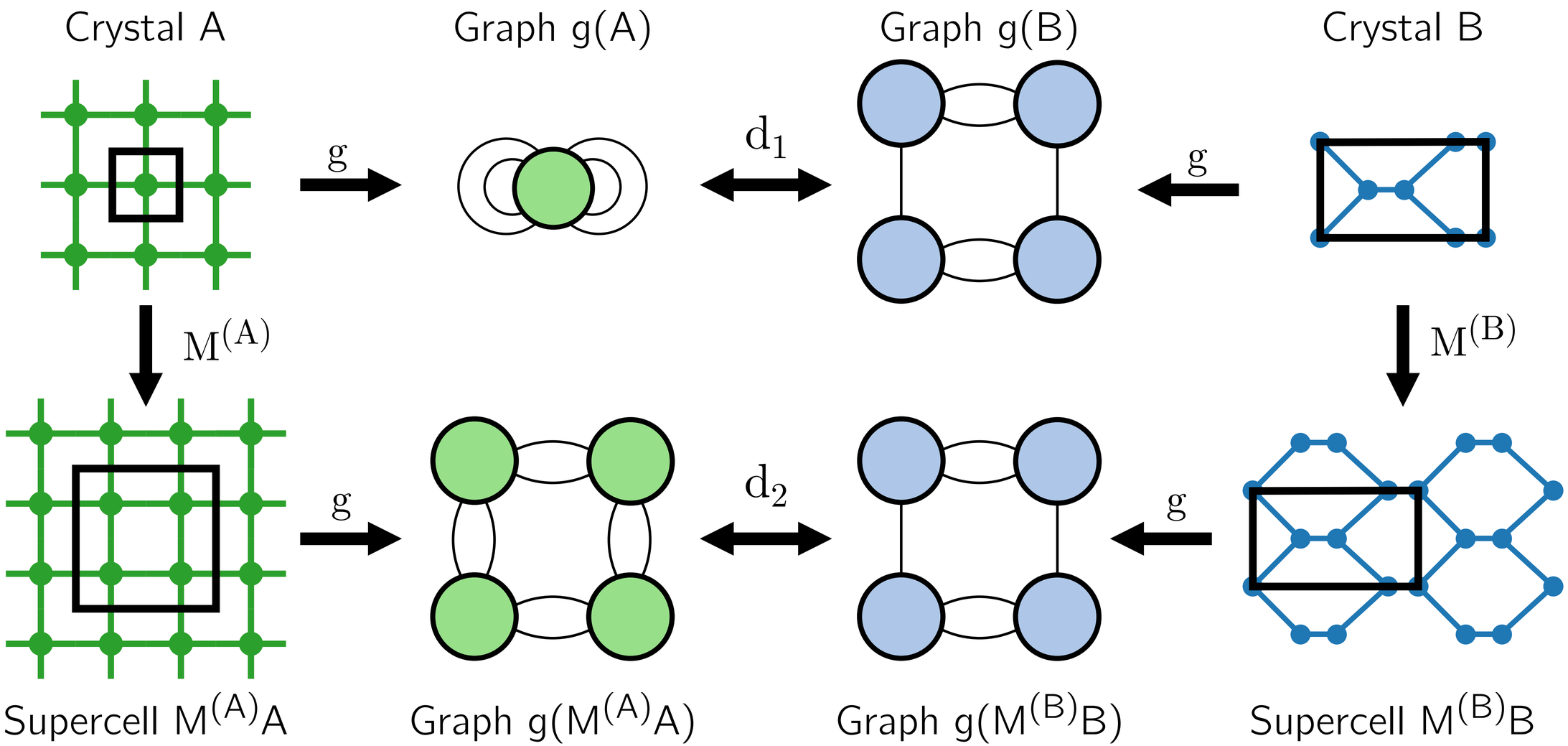
1. Representation Learning

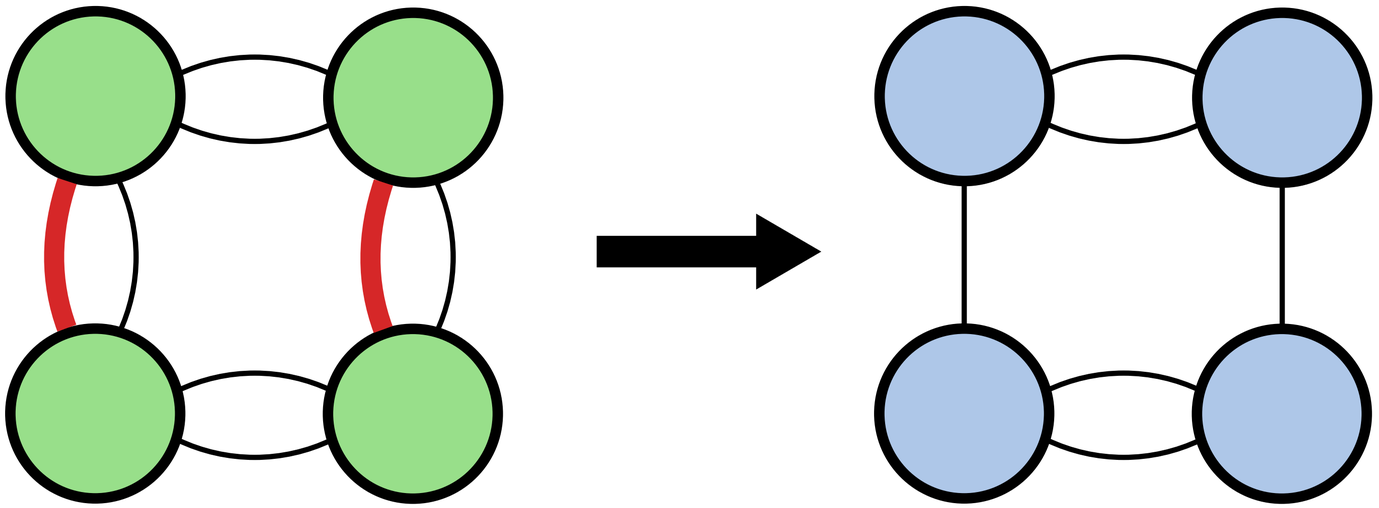
Can we correlate structure to synthesis?

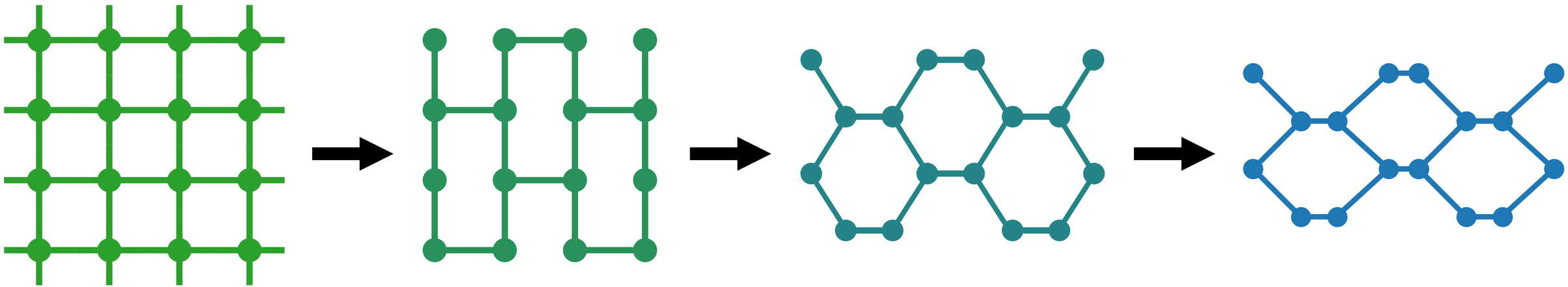
D. Schwalbe-Koda et al. *Nature Materials* (2019)


zeolite
unit cell graph
periodic graph

How to compare crystal graphs?

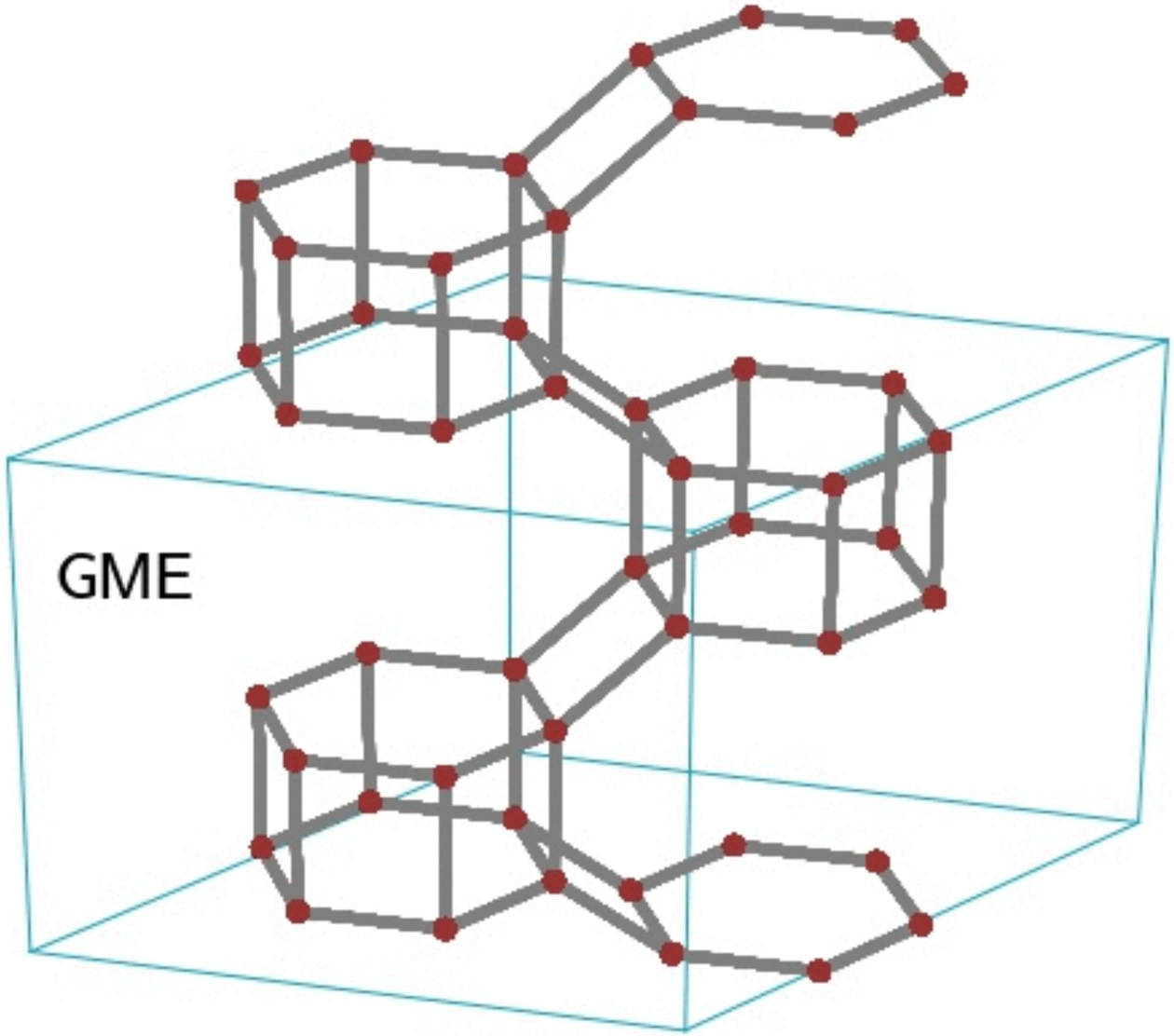




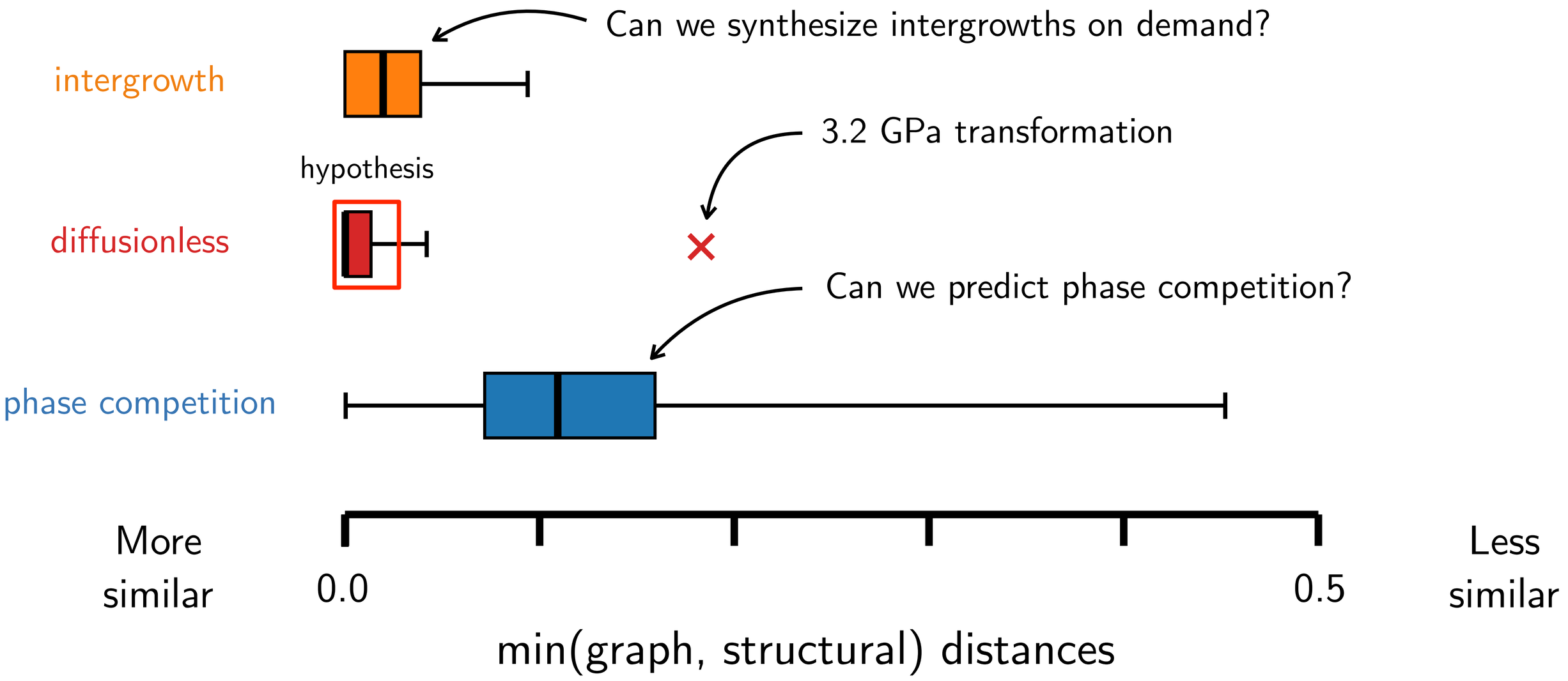


Crystal A bond breaking lattice distortion Crystal B

Transformations happen between graph-similar structures



Graph similarity explains diffusionless transformations from the literature

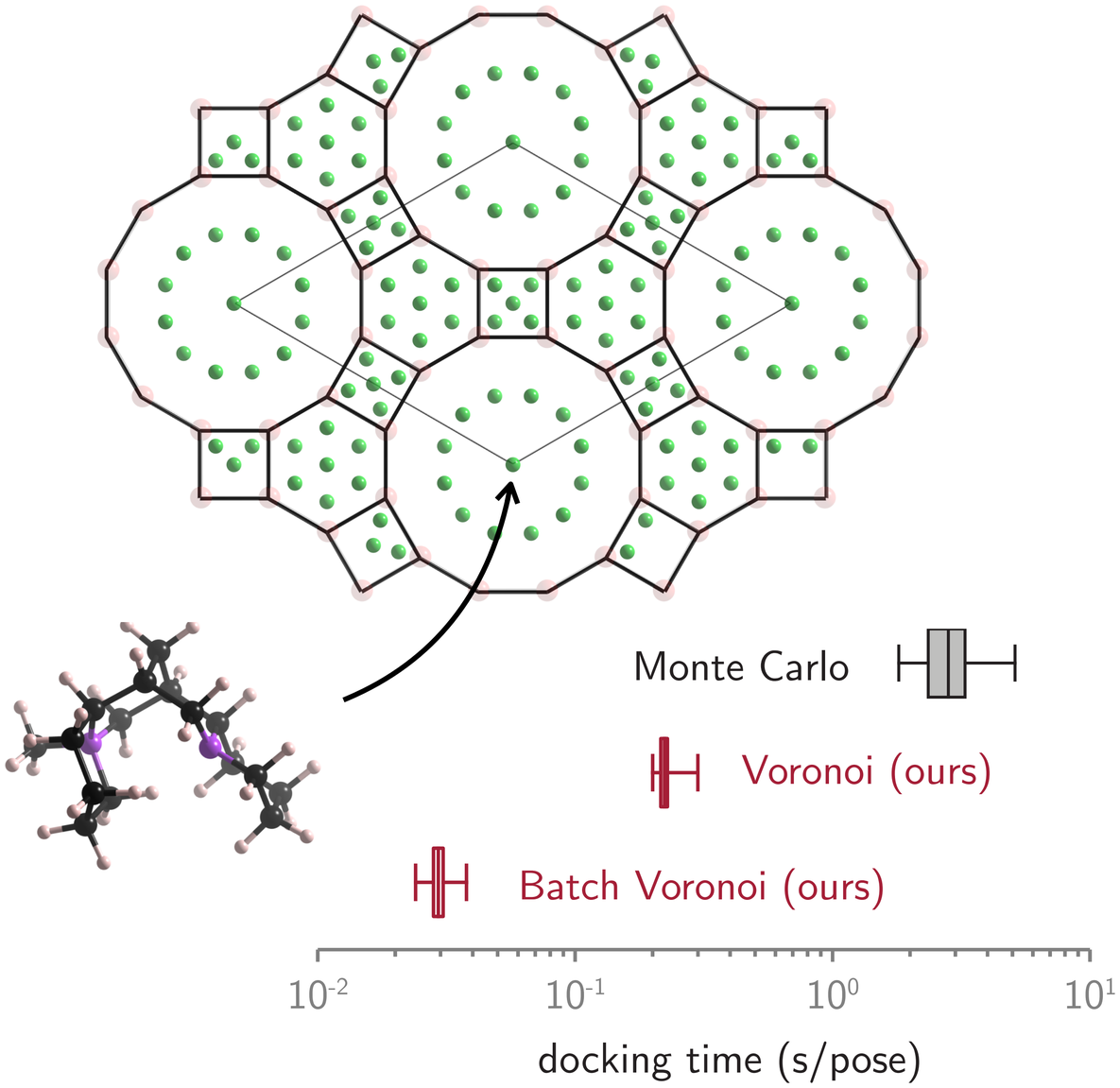


2. Dataset generation for *a priori* synthesis

How to find lock-key pairs in the molecular keychain?

D. Schwalbe-Koda et al. *JPCC* (2021)
D. Schwalbe-Koda et al. JCP (2021)
D. Schwalbe-Koda et al. *Science* (2021)

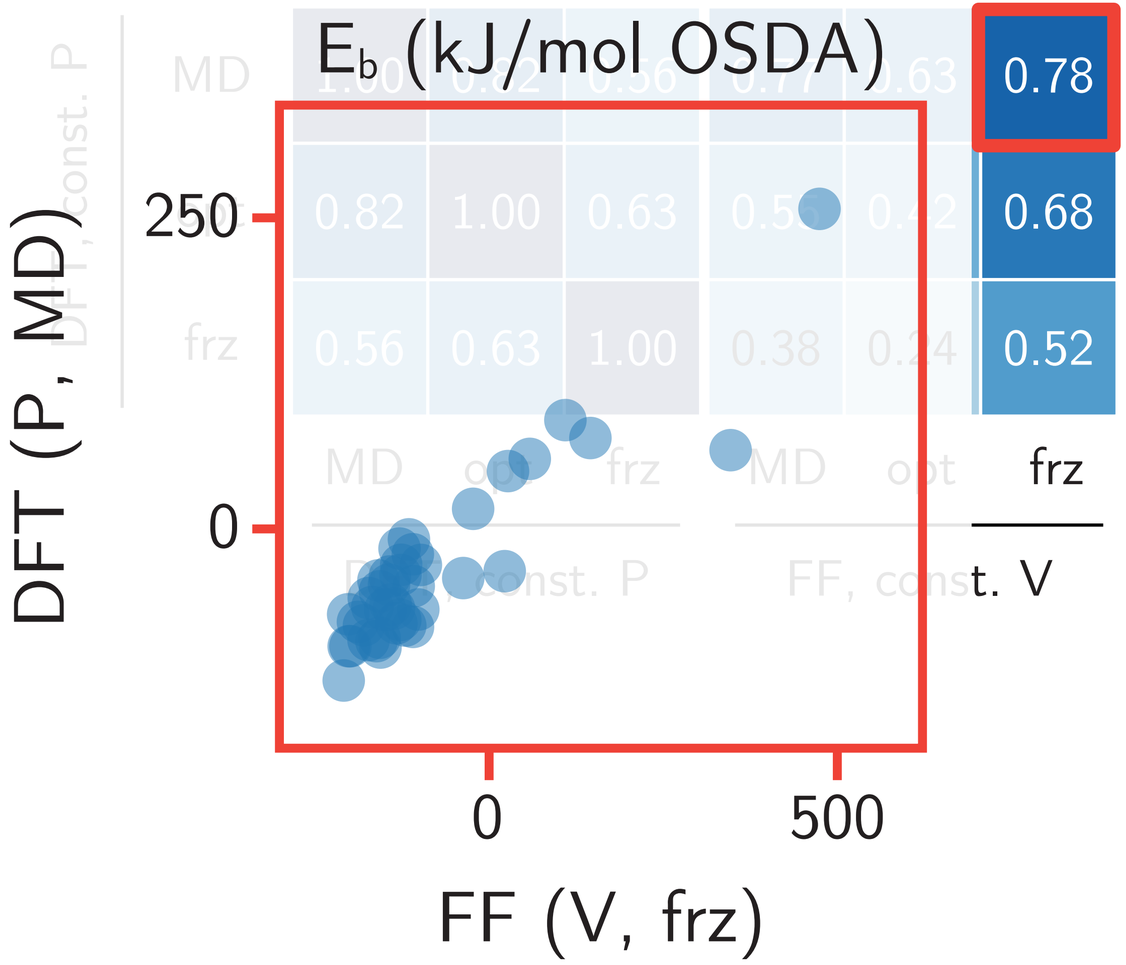
Docking algorithm



D. Schwalbe-Koda and R. Gómez-Bombarelli. *J. Phys. Chem. C.* (2021)

Benchmarks

Spearman's correlation coefficient
(higher is better)



D. Schwalbe-Koda and R. Gómez-Bombarelli. *J. Chem. Phys.* (2021)

Calculations: most known zeolites vs all known OSDAs

209

Zeolites

549

OSDAs

112,426

OSDA-zeolite pairs

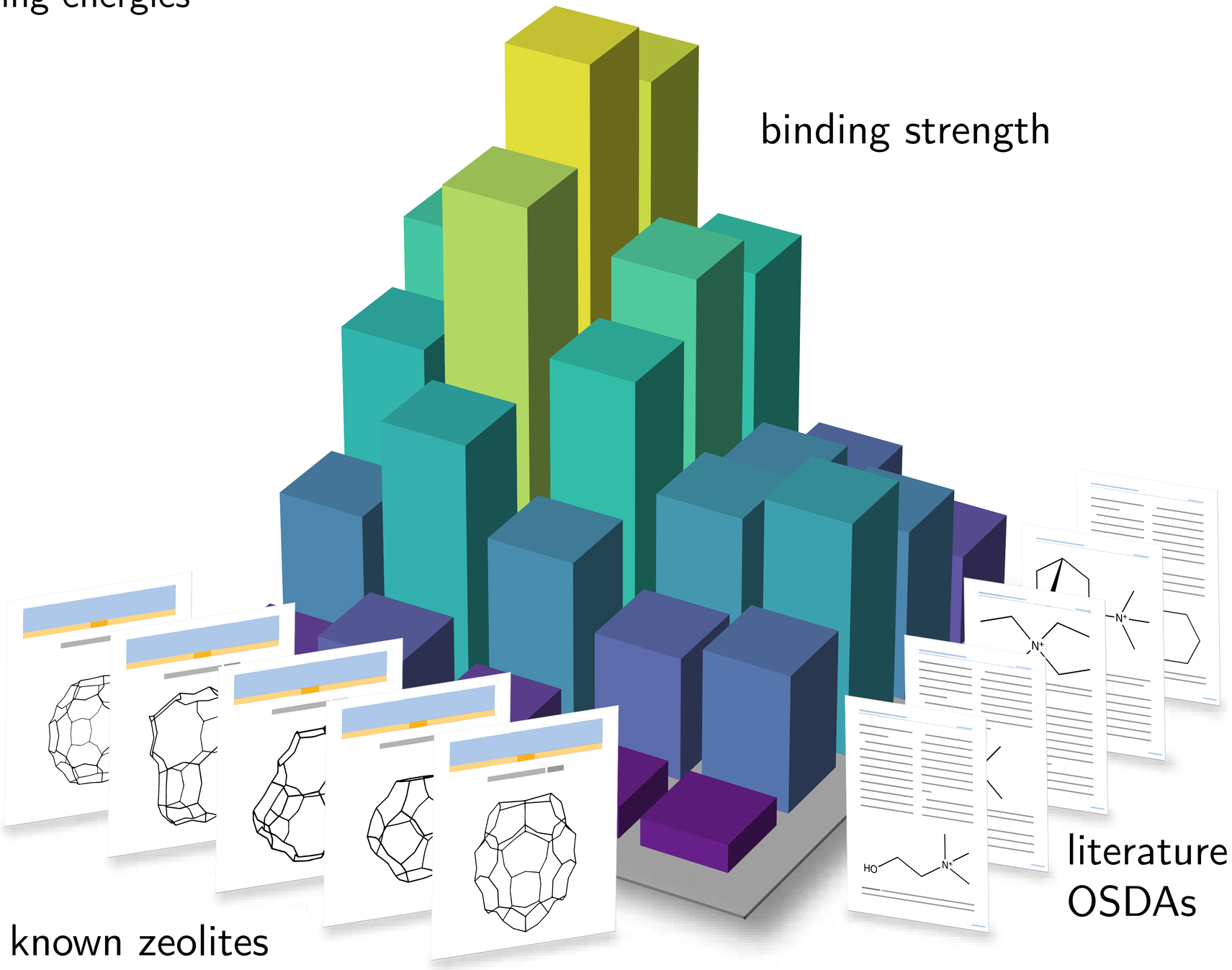
586,303

Poses calculated

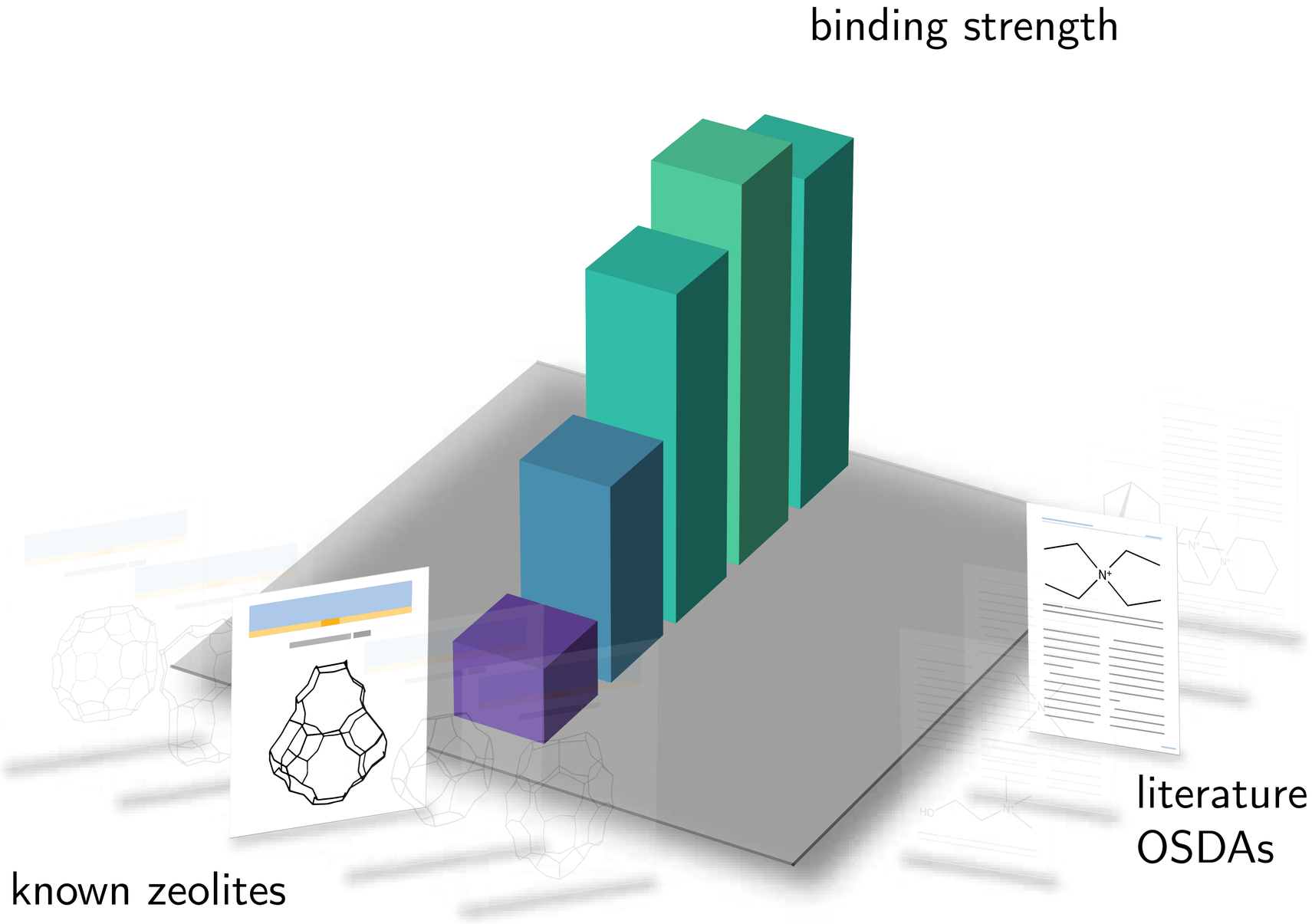
If traditional docking + FF approaches had been used: > **50,000 CPU-years** to complete these calculations

> **440 million CPU-hours**

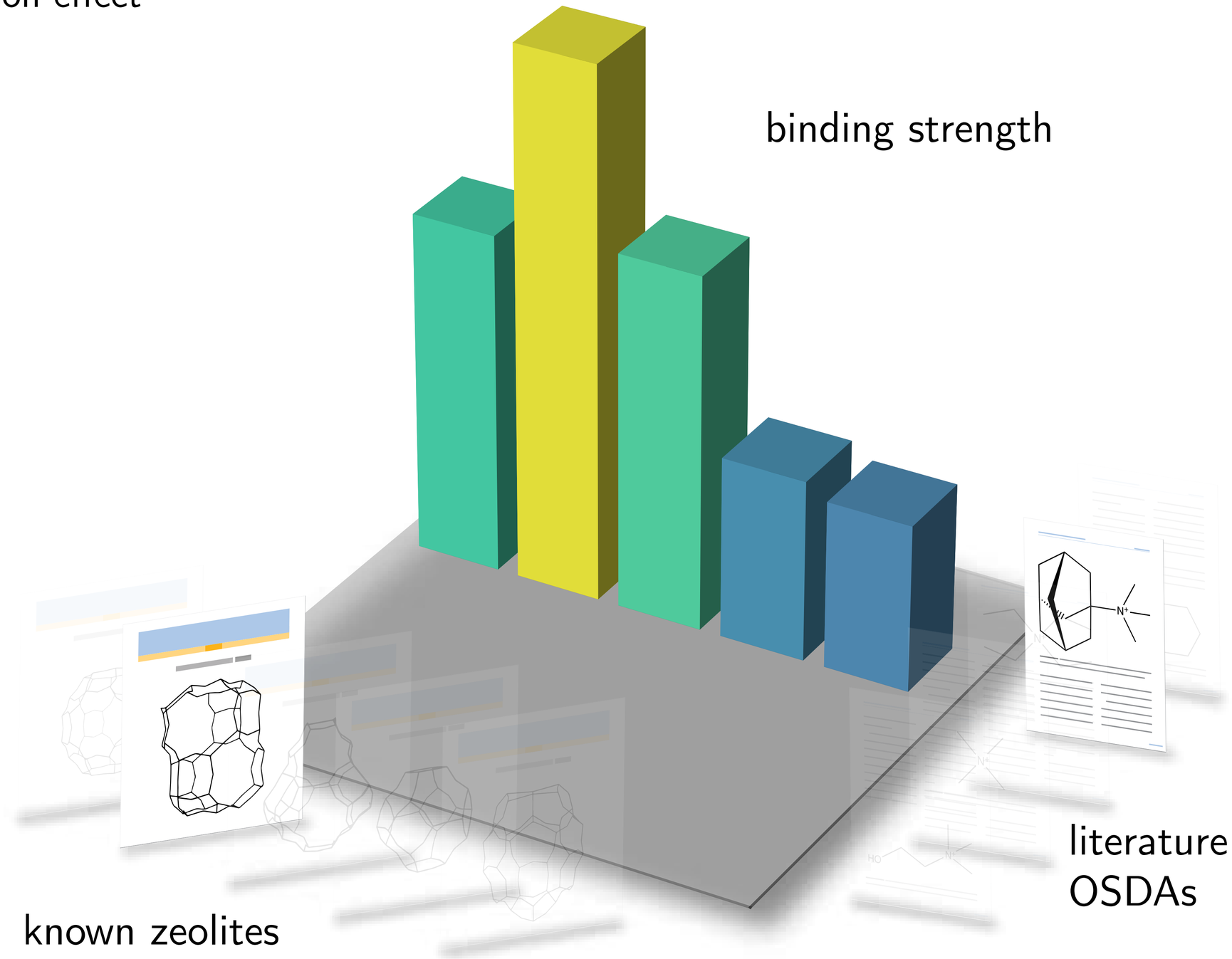
Heatmap of binding energies



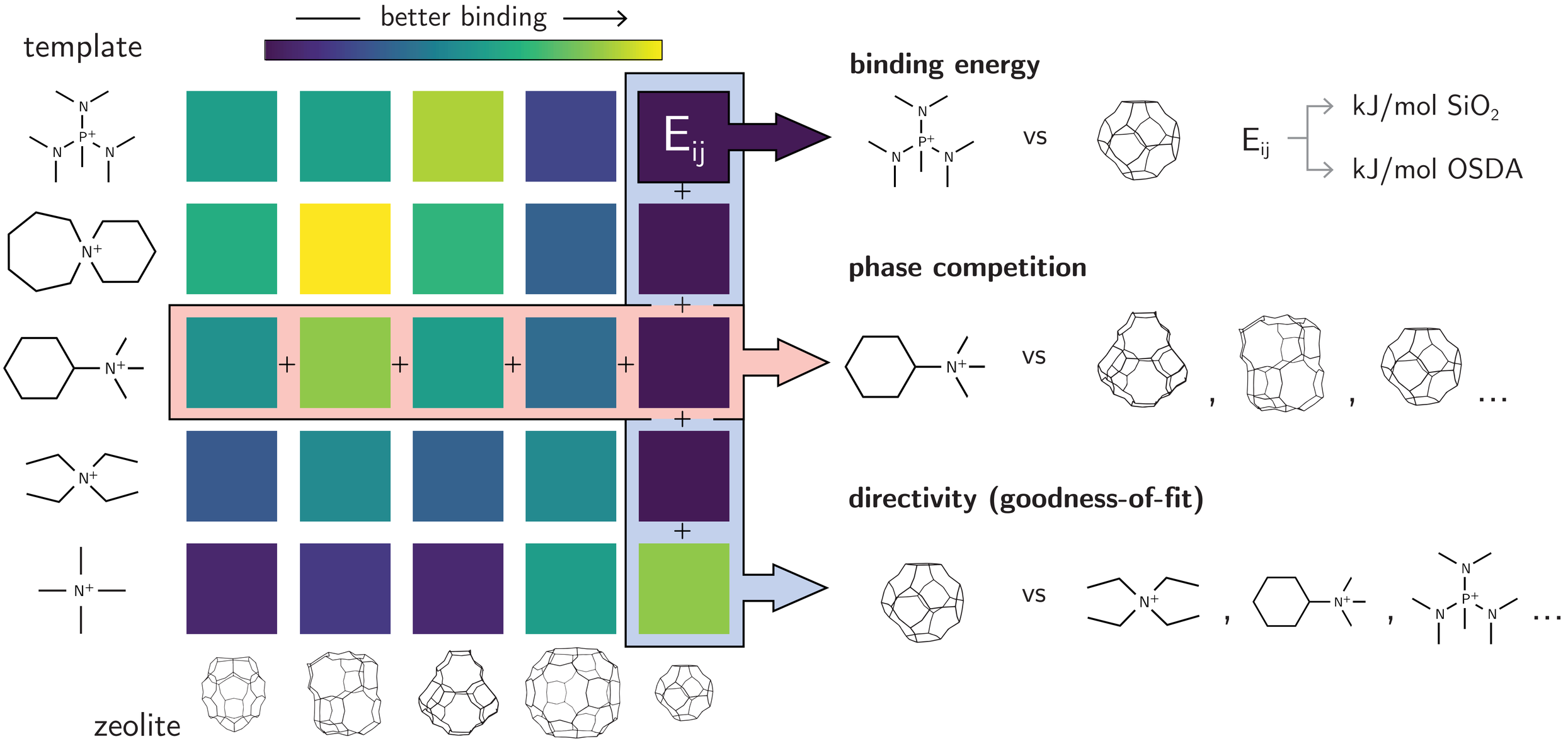
Goodness-of-fit effect



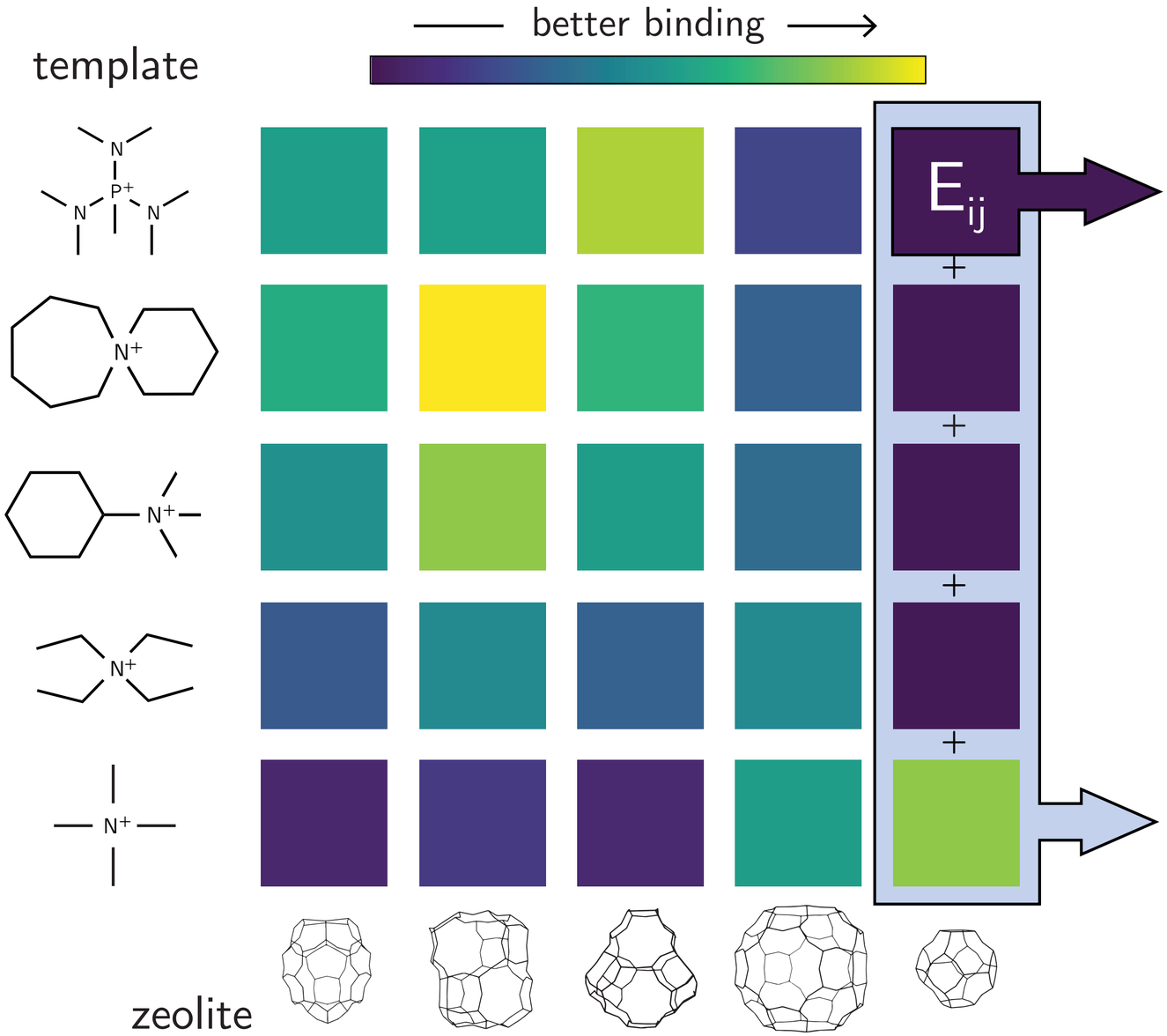
Phase competition effect



Analyzing the binding matrix



Quantifying rankings in a binding matrix



$$\frac{\exp\left(-\frac{E_{ij}^{(\mathrm{Si})}}{kT}\right)}{\sum_j \exp\left(-\frac{E_{ij}^{(\mathrm{Si})}}{kT}\right)}$$

Designing a descriptor for all binding energies: inspiration from entropy

Directivity (SiO_2)

$$D(\mathrm{SiO_2}) = \frac{\exp\left(-\frac{E_{ij}^{(\mathrm{Si})}}{kT}\right)}{\sum_j \exp\left(-\frac{E_{ij}^{(\mathrm{Si})}}{kT}\right)}$$

Competition (SiO_2)

$$C(\mathrm{SiO_2}) = \frac{\exp\left(-\frac{E_{ij}^{(\mathrm{Si})}}{kT}\right)}{\sum_i \exp\left(-\frac{E_{ij}^{(\mathrm{Si})}}{kT}\right)}$$

Normalized ranking

$$p = \left(D(\mathrm{SiO_2}) \cdot C(\mathrm{SiO_2}) \cdot D(\mathrm{OSDA}) \cdot C(\mathrm{OSDA})\right)^{1/4}$$

Directivity (OSDA)

$$D(\mathrm{OSDA}) = \frac{\exp\left(-\frac{E_{ij}^{(\mathrm{OSDA})}}{kT}\right)}{\sum_j \exp\left(-\frac{E_{ij}^{(\mathrm{OSDA})}}{kT}\right)}$$

Competition (OSDA)

$$C(\mathrm{OSDA}) = \frac{\exp\left(-\frac{E_{ij}^{(\mathrm{OSDA})}}{kT}\right)}{\sum_i \exp\left(-\frac{E_{ij}^{(\mathrm{OSDA})}}{kT}\right)}$$

Templating Energy

$$E_t = -kT \log p$$

Question: how to validate this synthesis descriptor?

Extracting ground truth data from the literature using NLP

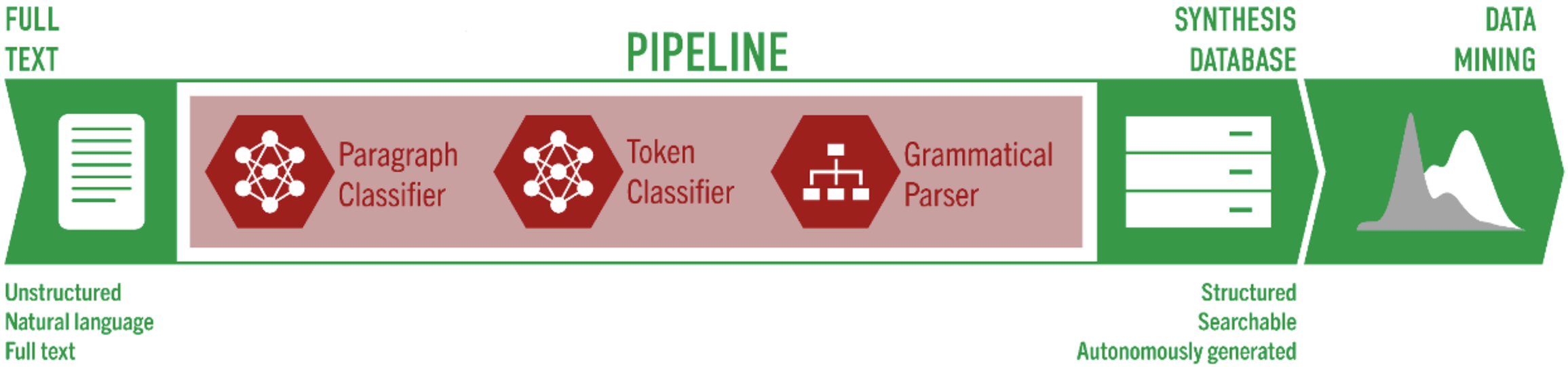
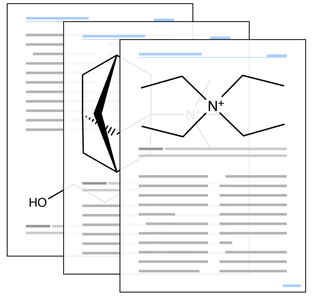


Figure courtesy of E. Olivetti

>5M papers/patents



>90,000 articles on zeolites

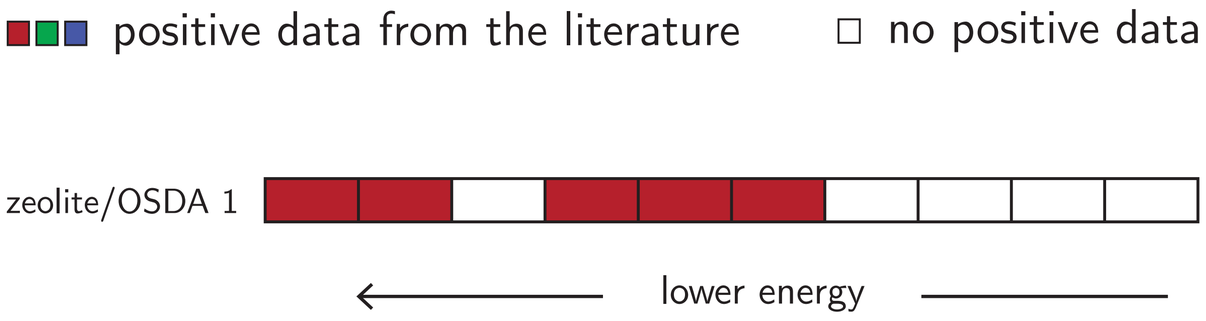
thousands of zeolite synthesis routes
hundreds of distinct OSDAs

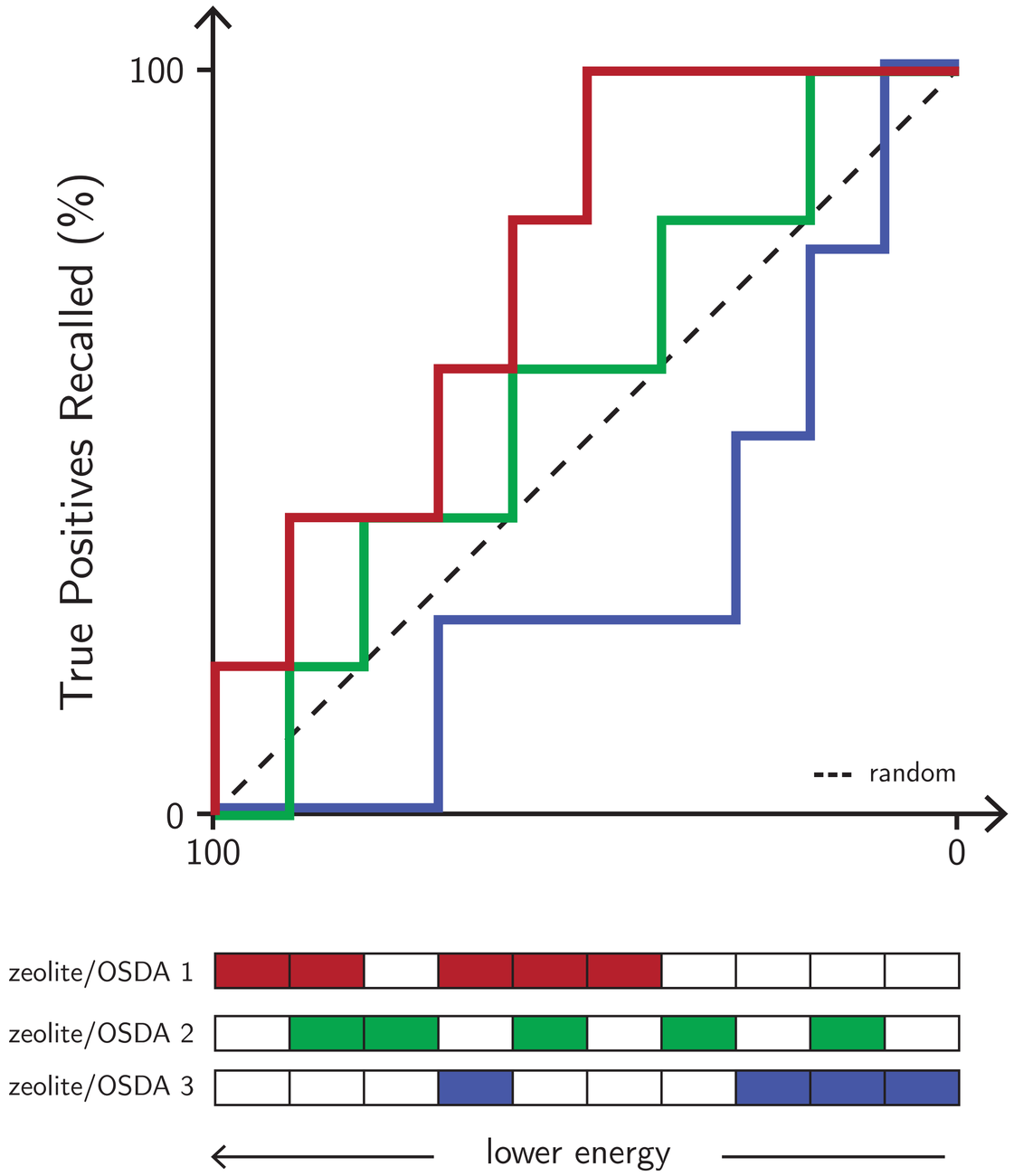
Question: how to compare the literature with theory/simulations?

with Z. Jensen and E. Olivetti (MIT)

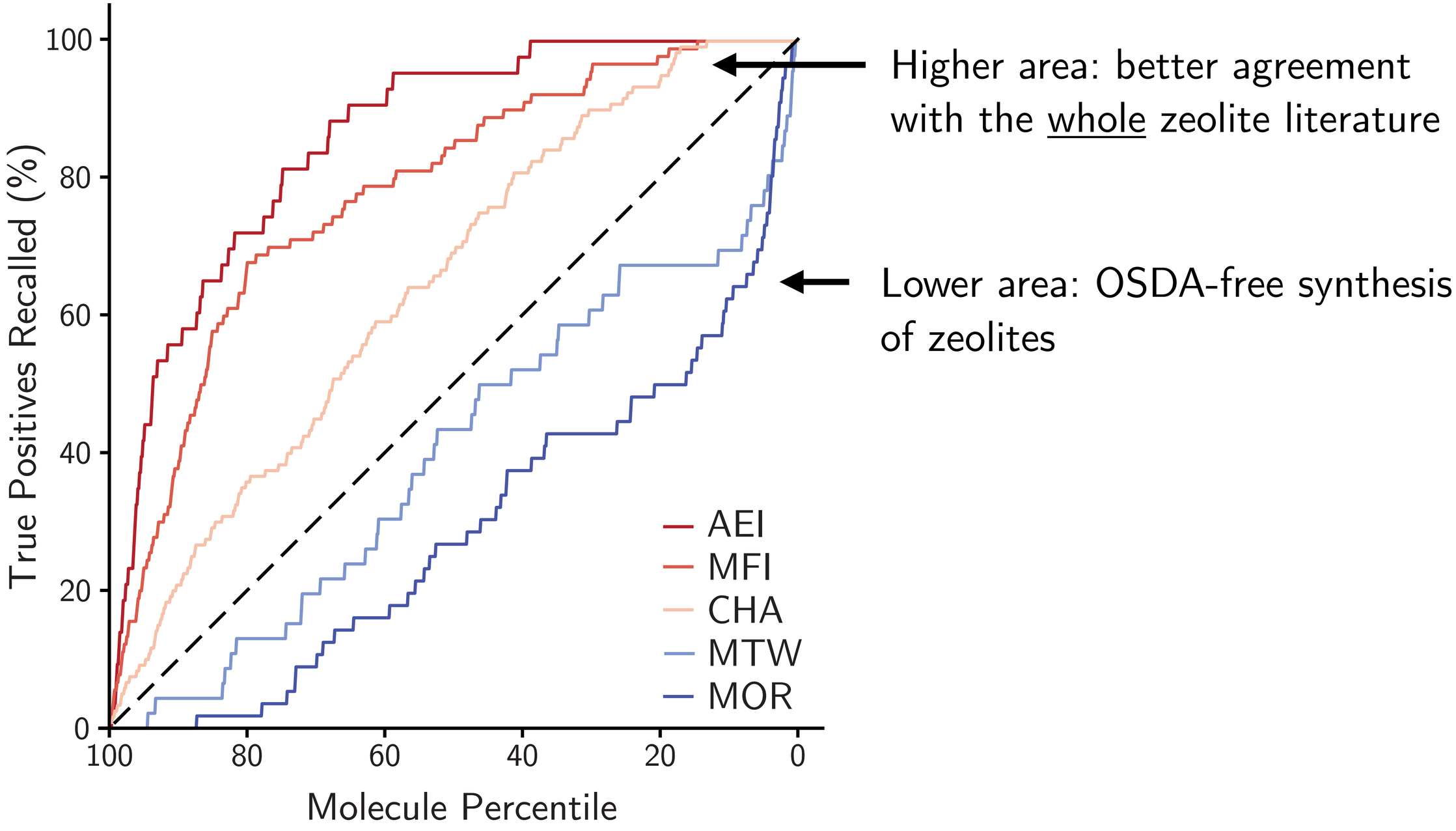
D. Schwalbe-Koda et al. *Nature Materials* (2019)
Z. Jensen et al. *ACS Cent. Sci.* (2019, 2021)

Defining a metric of agreement with the literature

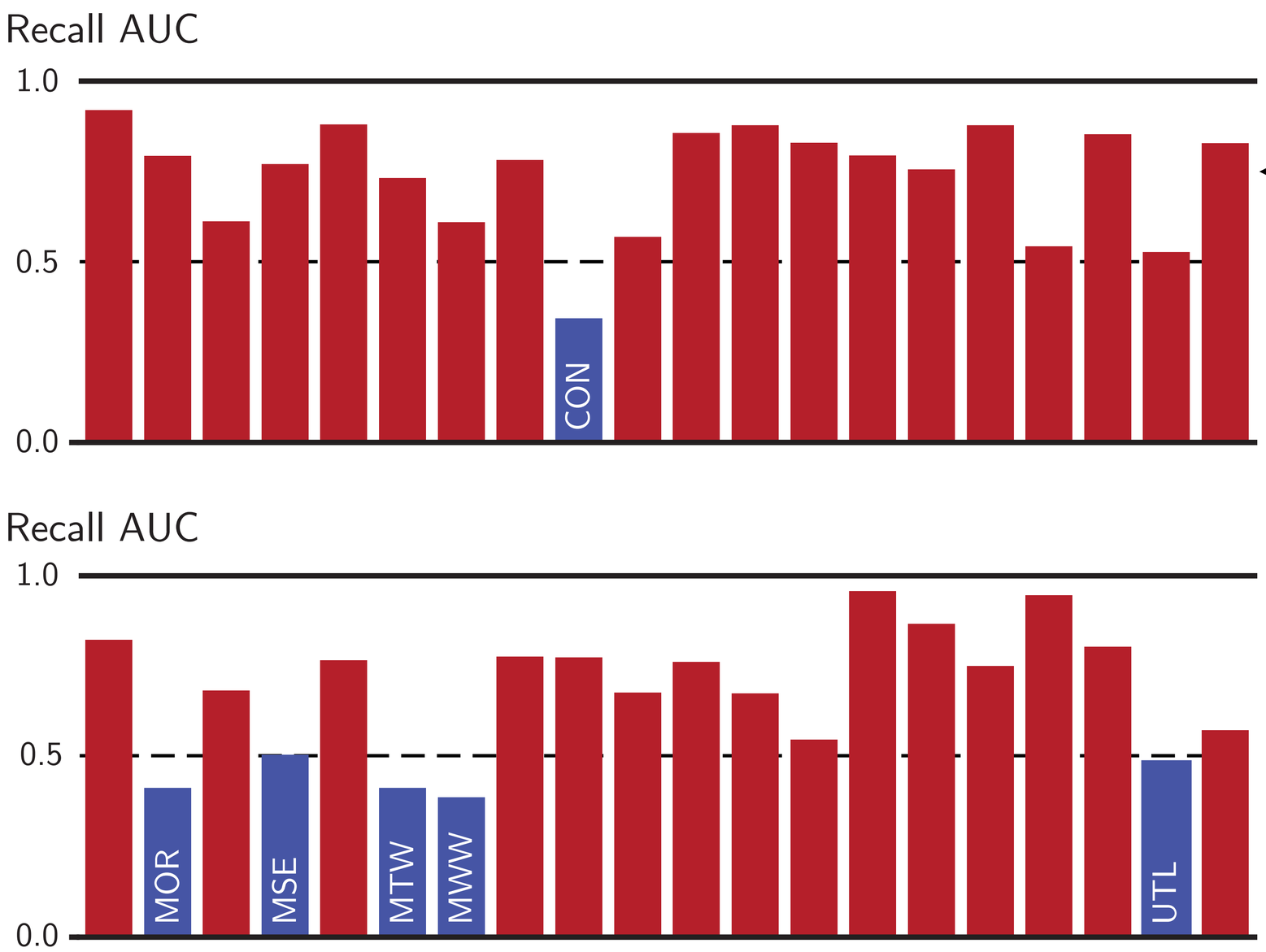




Comparing the descriptor against the entire literature (60+ years of zeolite synthesis)



Recalling literature outcomes for the 40 most common zeolites



literature for 34/40 zeolites correctly recalled

Take-home message: the simulations explain the whole literature of zeolite synthesis

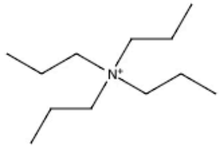
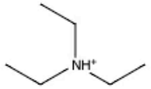
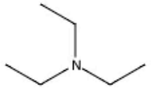
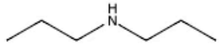
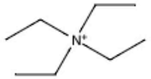
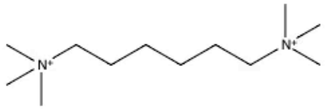
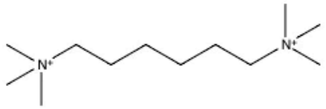
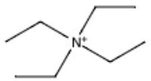
3. Data-driven zeolite synthesis

How to design synthesis routes for zeolites?

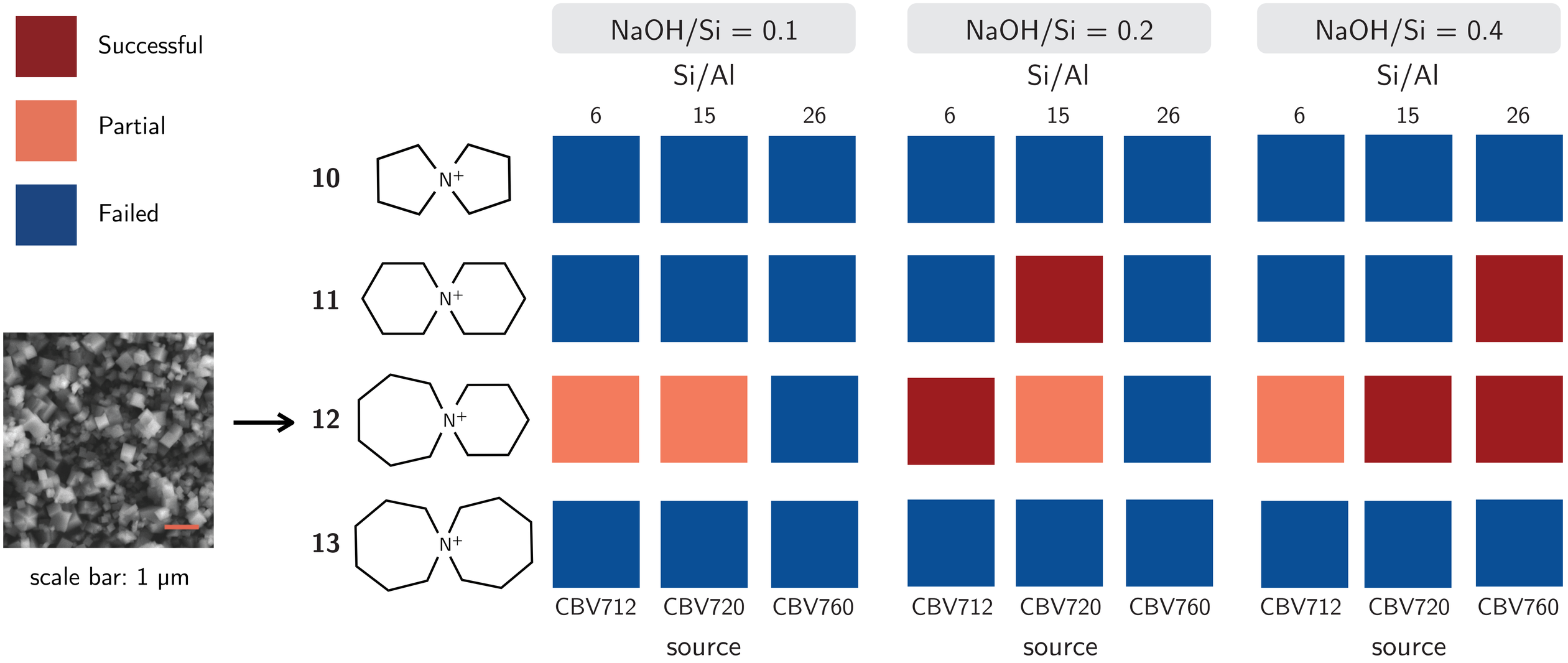
D. Schwalbe-Koda et al. *Science* (2021)
D. Schwalbe-Koda et al. *Chem. Mater.* (2022)

OSDB Affinities Frameworks OSDAs https://zeodb.mit.edu

Framework | SMILES | Binding (SiO2) < | Directivity (SiO2) < | Competition (SiO2) < | Binding (OSDA) < | Competition (OSDA) < | Templating < | Search

Host	Guest	Binding Energy (kJ/mol Si)	Directivity Energy (kJ/mol Si)	Competition Energy (kJ/mol Si)	Binding Energy (kJ/mol OSDA)	Competition Energy (kJ/mol OSDA)	Templating Energy (kJ/mol)	Rotatable Bonds	# Publications	Details
MFI		-10.23	2.64	3.72	-245.52	-18.37	16.43	8	47	Details
AFI		-4.63	6.75	13.65	-111.11	28.18	21.50	3	41	Details
AFI		-4.93	6.46	13.26	-118.29	25.30	20.97	3	41	Details
AEL		-7.19	1.73	10.08	-143.78	9.60	17.98	4	37	Details
BEA		-10.38	8.36	4.58	-110.77	62.01	19.73	4	35	Details
ITH		-2.68	8.34	10.64	-149.95	94.07	20.84	7	26	Details
EUO		-6.09	5.68	7.23	-227.41	16.61	18.83	7	25	Details
CHA		-14.29	4.03	0.67	-171.48	1.30	16.77	4	21	Details
SOD		-13.24	5.79	1.57	-79.46	35.59	14.19	0	18	Details

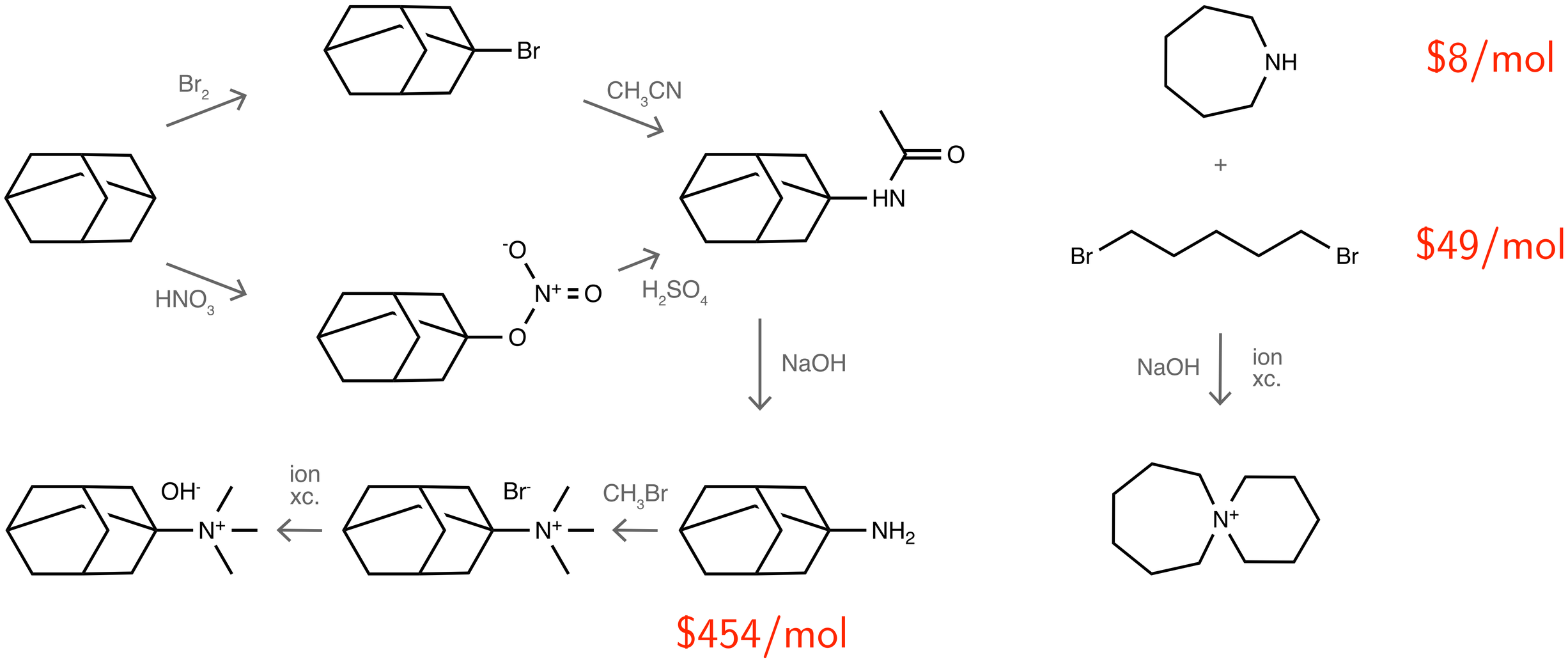
Synthesis accessibility of **CHA** zeolite



with S. Kwon, Y. Roman (MIT)
and M. Moliner, A. Corma and their team (ITQ, Spain)

Question: what is so interesting about this template?

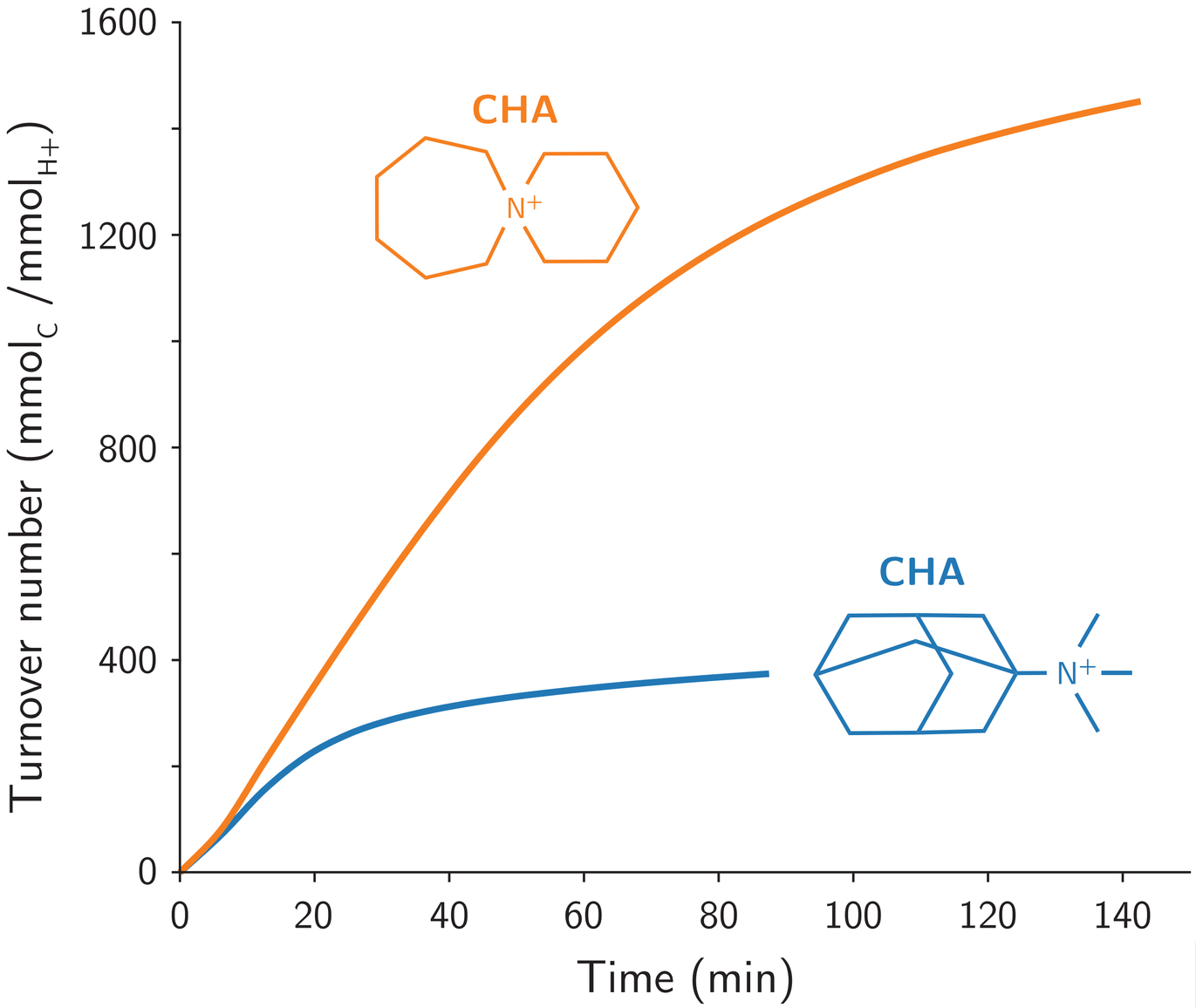
Advantage 1: possibility of lowering synthesis costs



Prices estimated from Sigma-Aldrich (2020)

Advantage 2: improvement in catalyst lifetime for MTO reaction

Higher catalyst lifetime



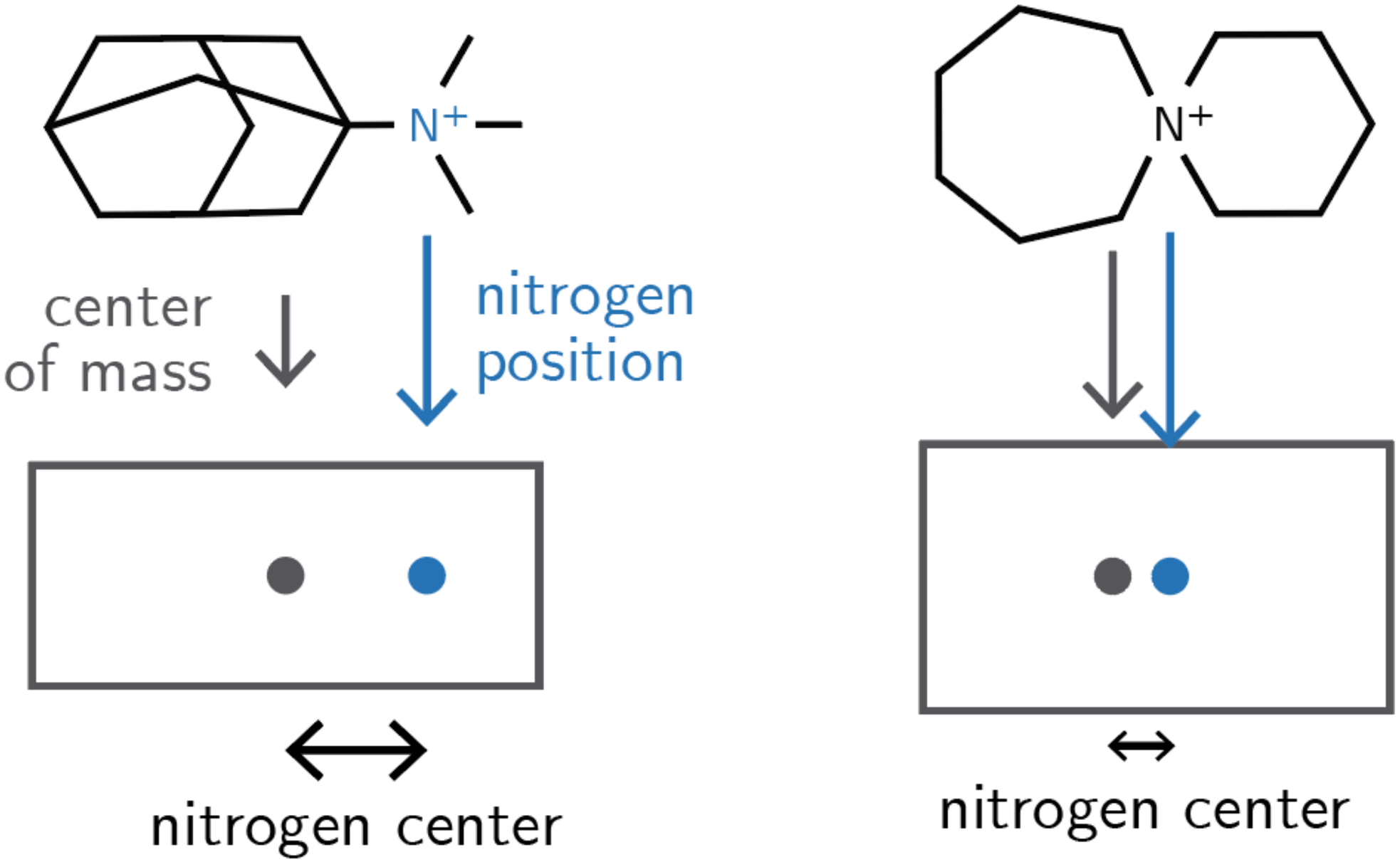
3-4 times larger turnover number

Question: how can template design influence catalytic behavior?

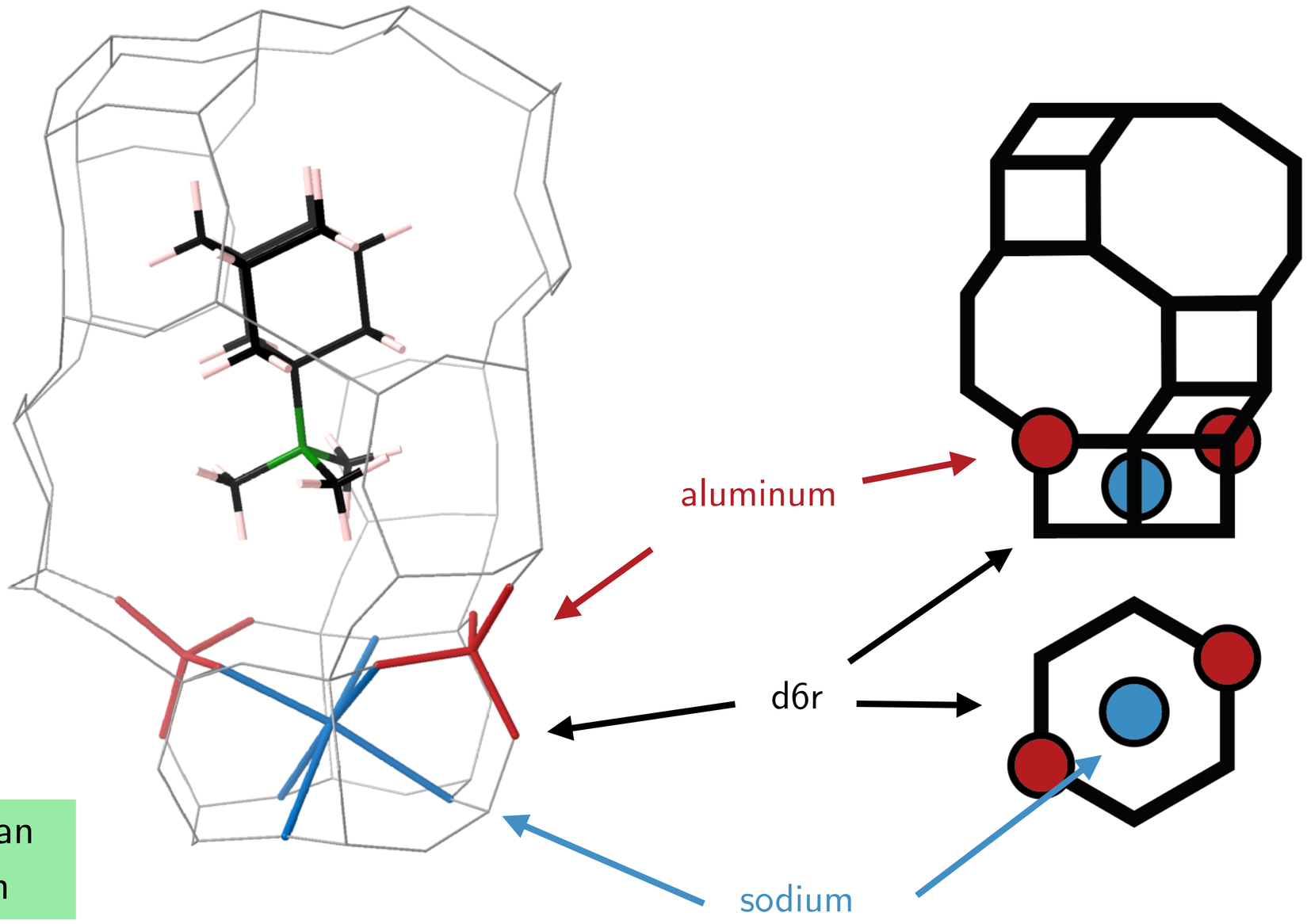
with S. Kwon, Y. Roman (MIT)
and M. Moliner, A. Corma and their team (ITQ, Spain)

Patent Appl. US 63/174,825 (2021)

OSDA design can modulate electrostatics



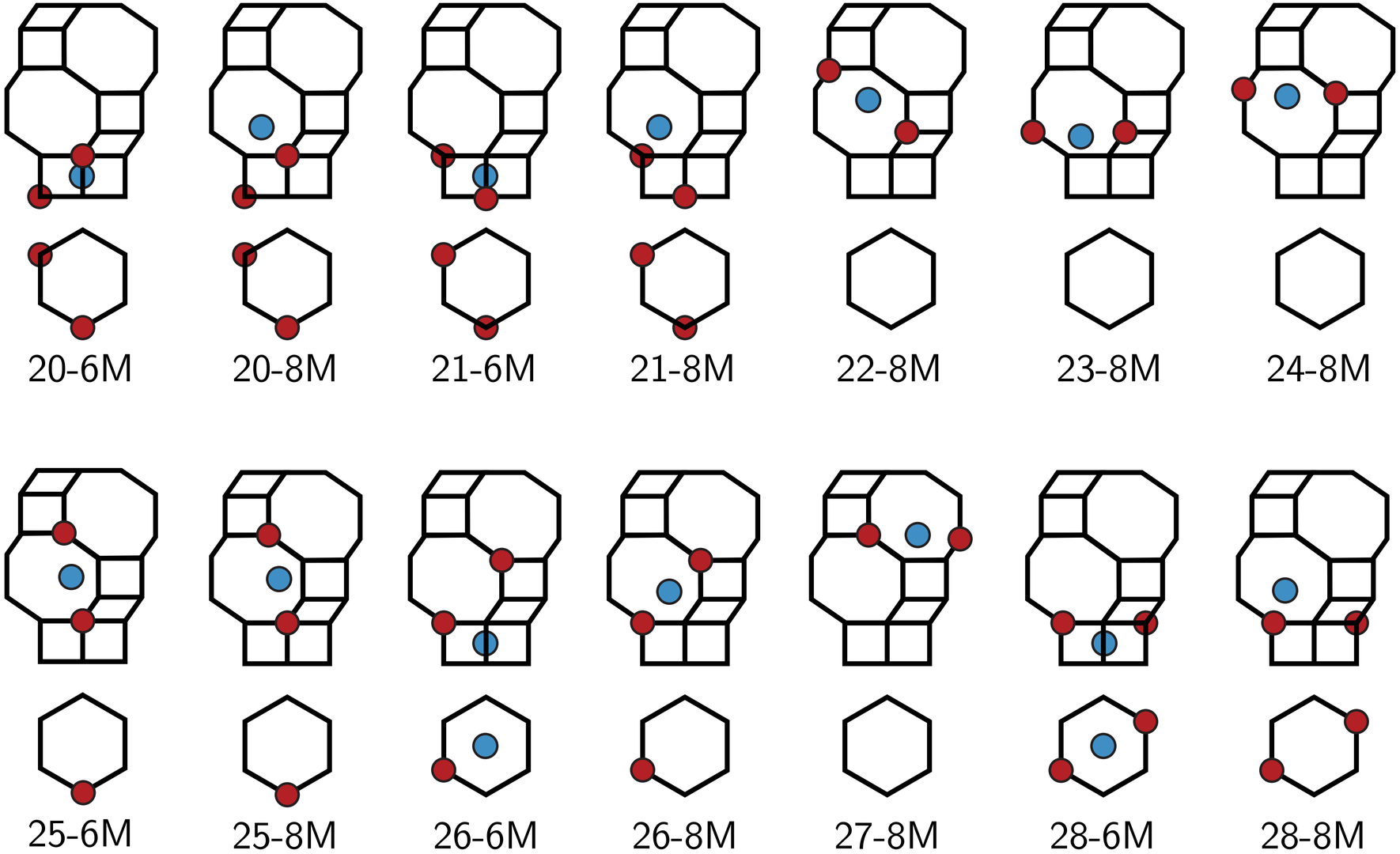
Active site distributions in zeolites are driven by electrostatics



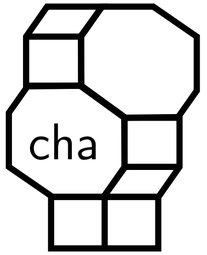
Hypothesis: OSDA design can control aluminum distribution

Calculating aluminum pairs in **CHA** using DFT

3NN



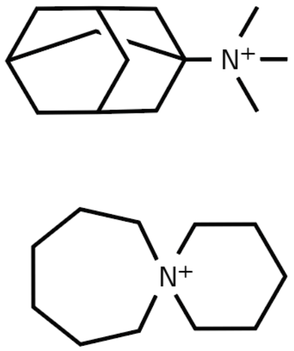
legend



sodium
aluminum

Notation from Di Iorio et al. *JACS* (2020)

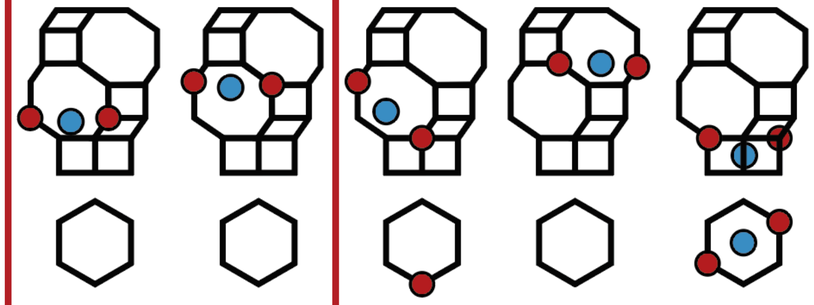
Variations in OSDA charge distribution lead to different active site placement

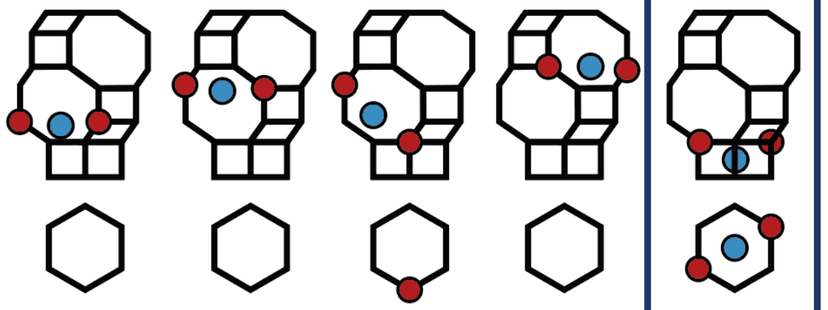


$Al^- + Na^+$ distribution

- aluminum
- sodium

	Energy (kJ/mol)					Fraction (%) @ 433 K				
	13.3	12.3	7.8	3.6	0.0	1	2	7	21	58
	2.0	0.1	6.2	0.0	1.2	15	26	5	27	19
ID	23	24	26	27	28	23	24	26	27	28

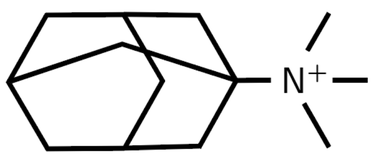




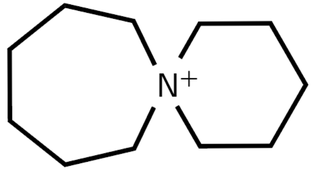
3x times less

favorable only with azaspiro template

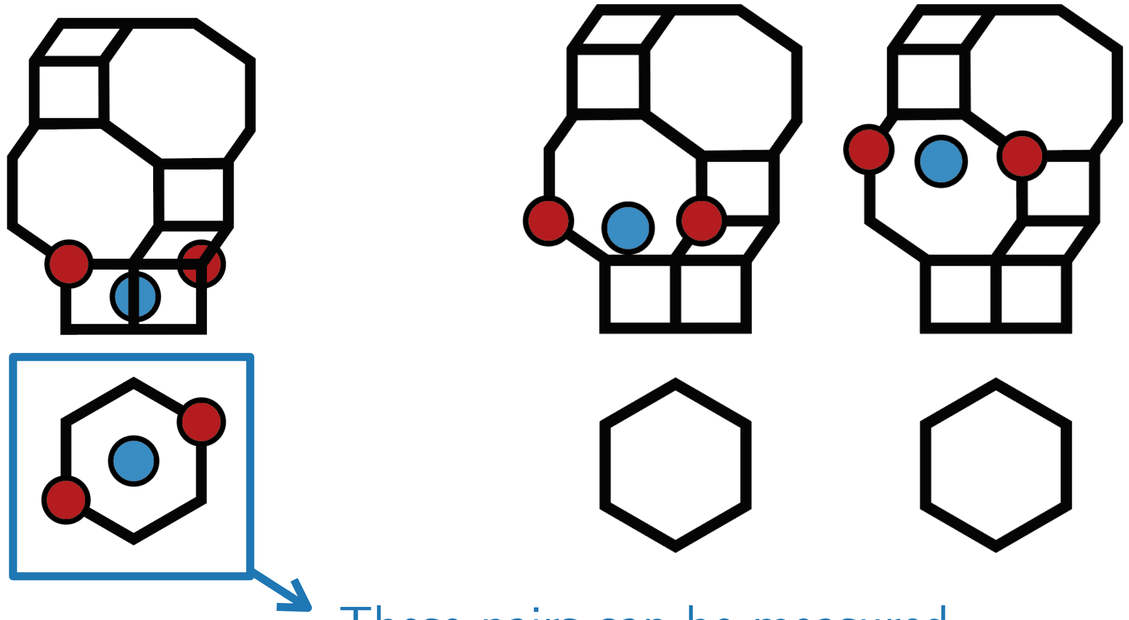
Experiments confirm theoretical predictions on control of aluminum distributions



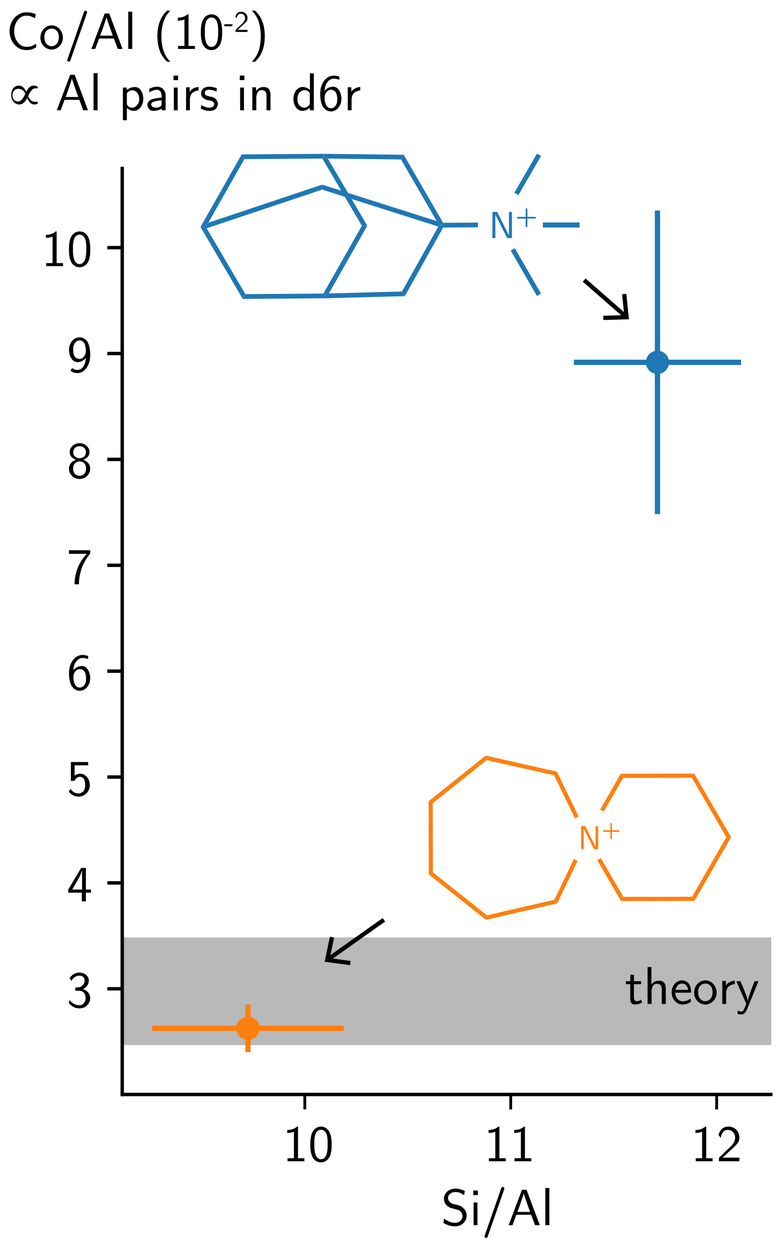
prefers



prefers

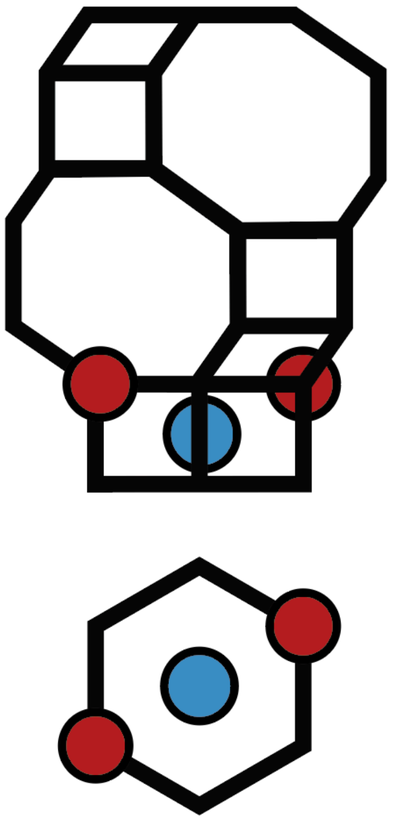


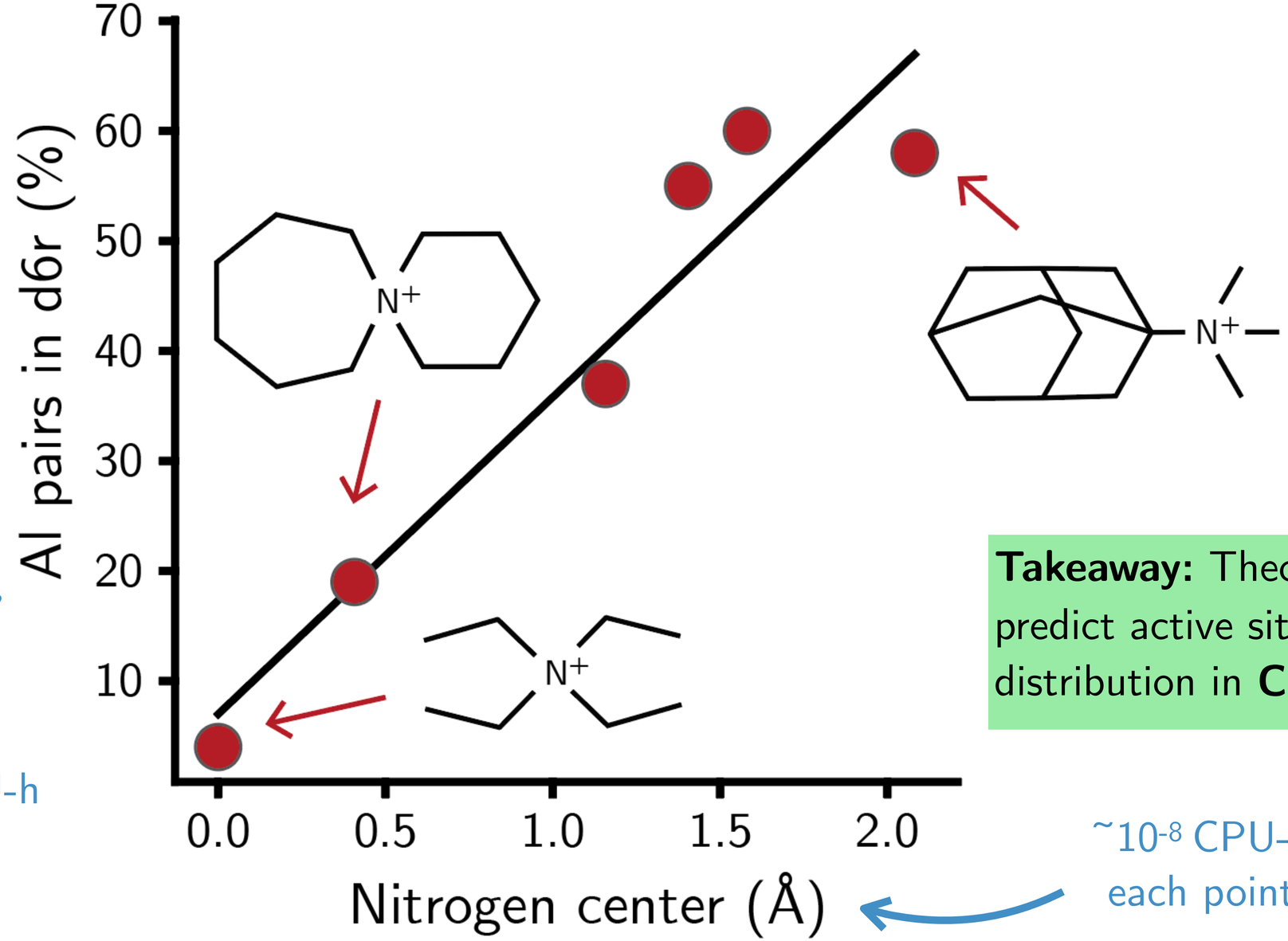
These pairs can be measured



with S. Kwon, Y. Roman (MIT)

Geometric descriptor predicts aluminum distributions in **CHA** zeolite





10^4 - 10^5 CPU-h each point

~10^{-8} CPU-h each point

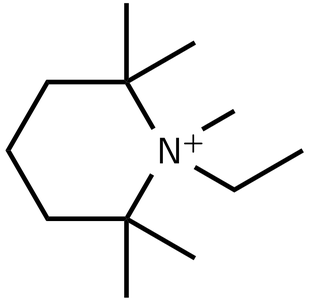
Takeaway: Theory can predict active sites distribution in **CHA**

4. Learning representations for zeolite synthesis

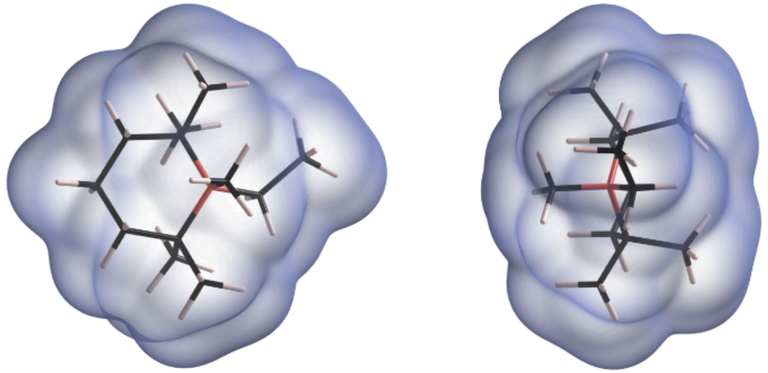
Finding representations for synthesis selectivity

D. Schwalbe-Koda et al. *Science* (2021)
D. Schwalbe-Koda et al. *J. Phys. Chem. Lett.* (2021)
D. Schwalbe-Koda et al. *Chem. Mater.* (2022)
E. Bello, D. Schwalbe-Koda et al. *Angew. Chem. Intl. Ed.* (2022)

Rationalizing zeolite intergrowths



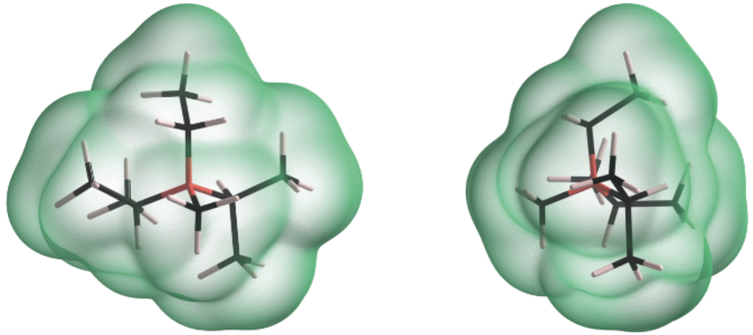
good for **AEI** zeolite



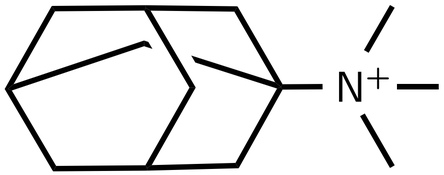
shape "interpolation"

???

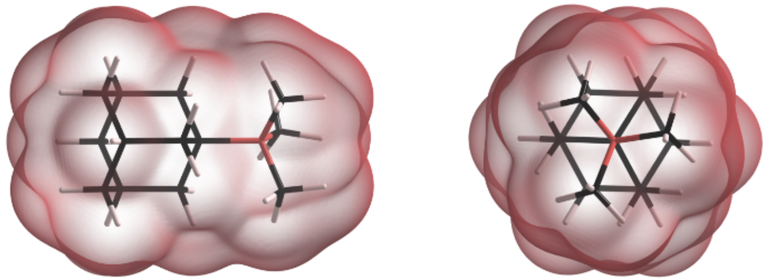
good for **CHA/AEI**?

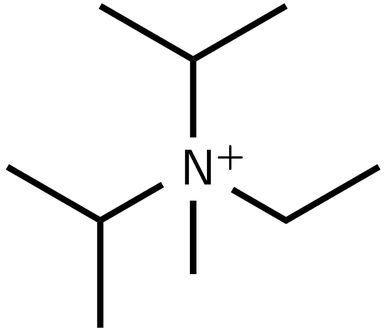


shape "interpolation"



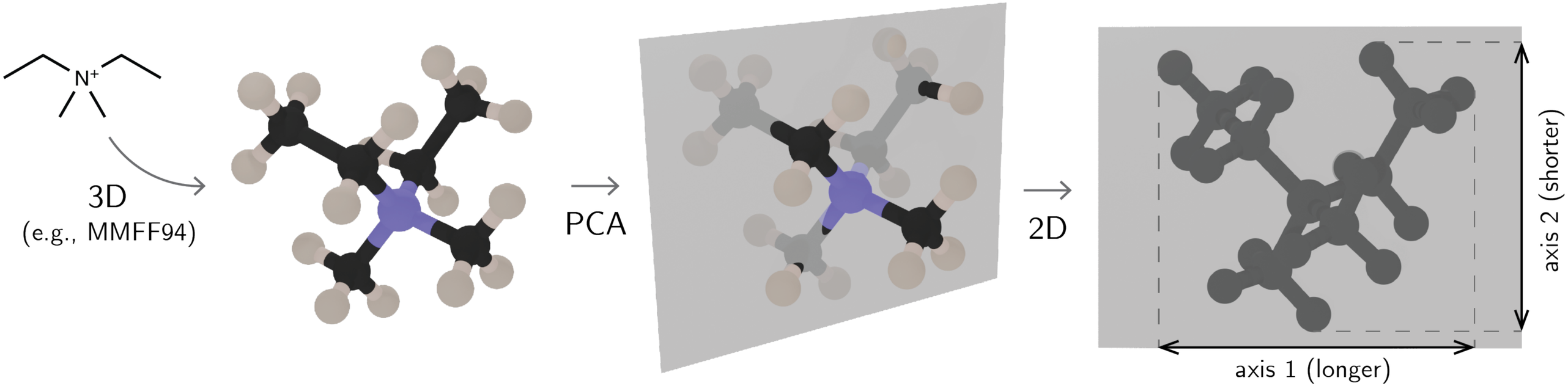
good for **CHA** zeolite



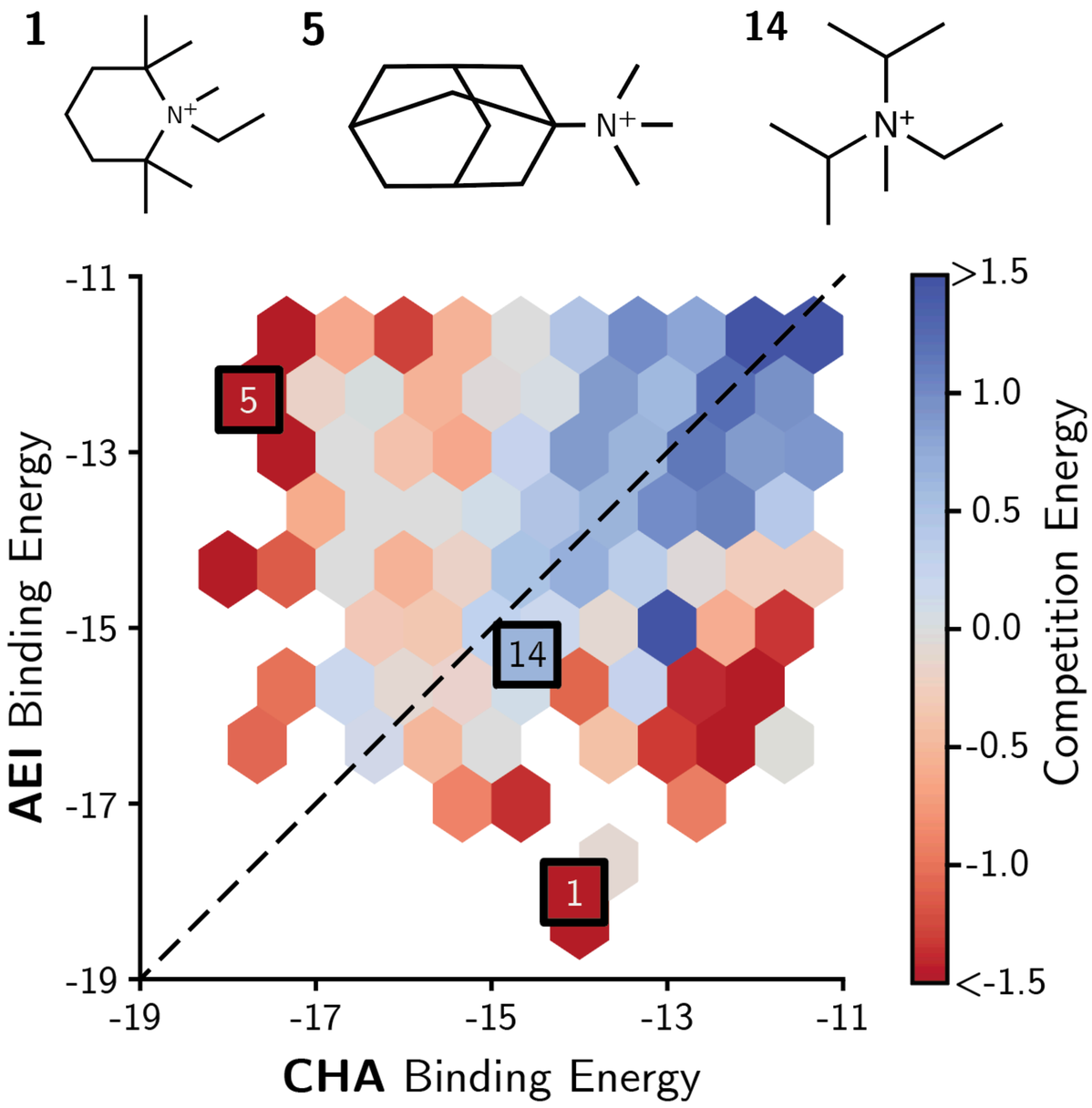


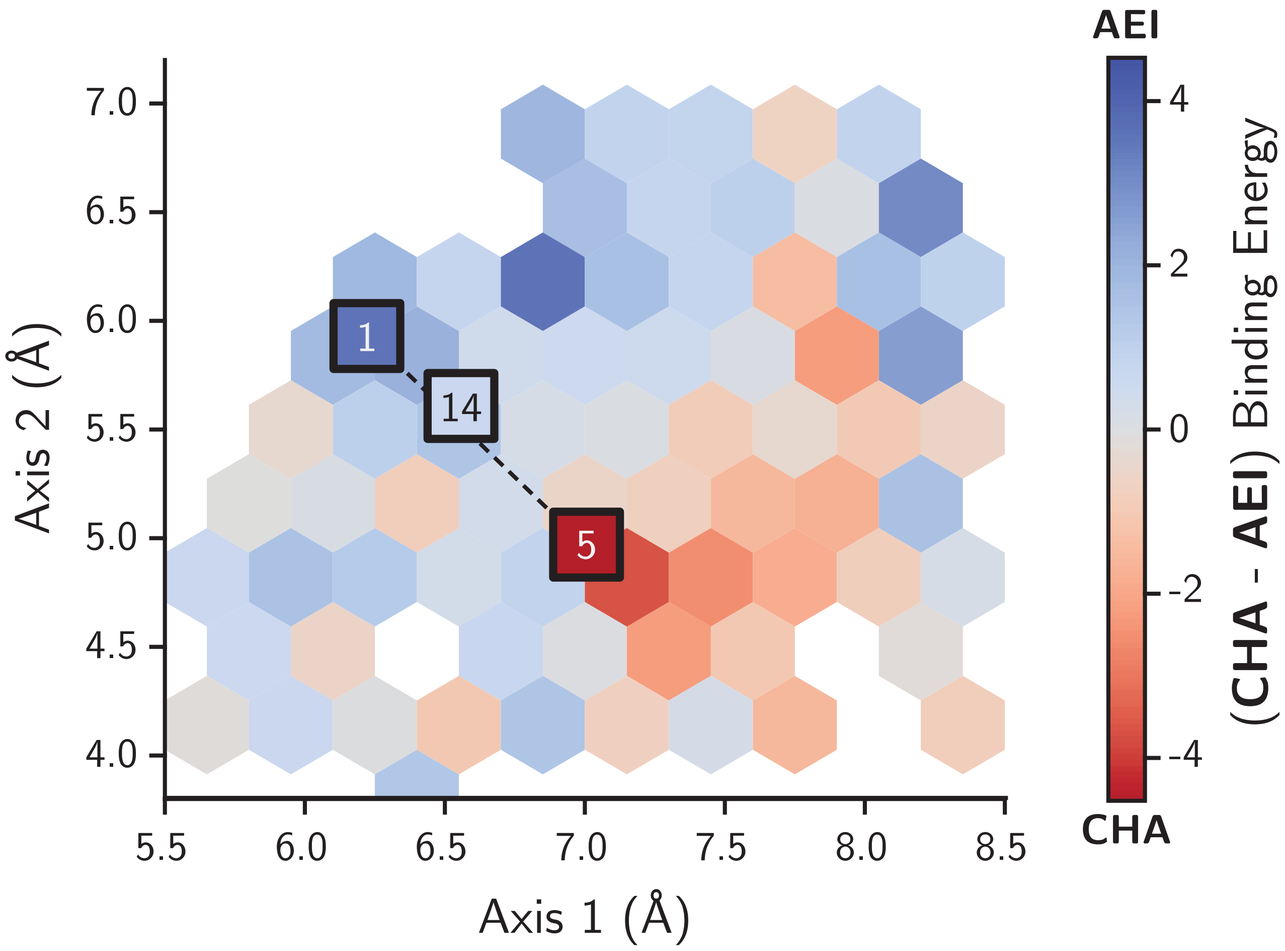
proposed OSDA for **CHA/AEI** zeolite

Simple representation of shape

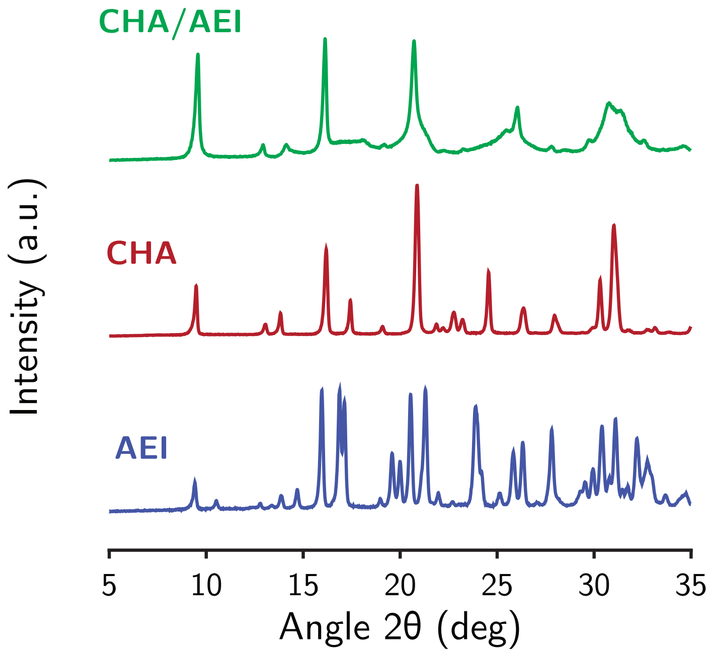


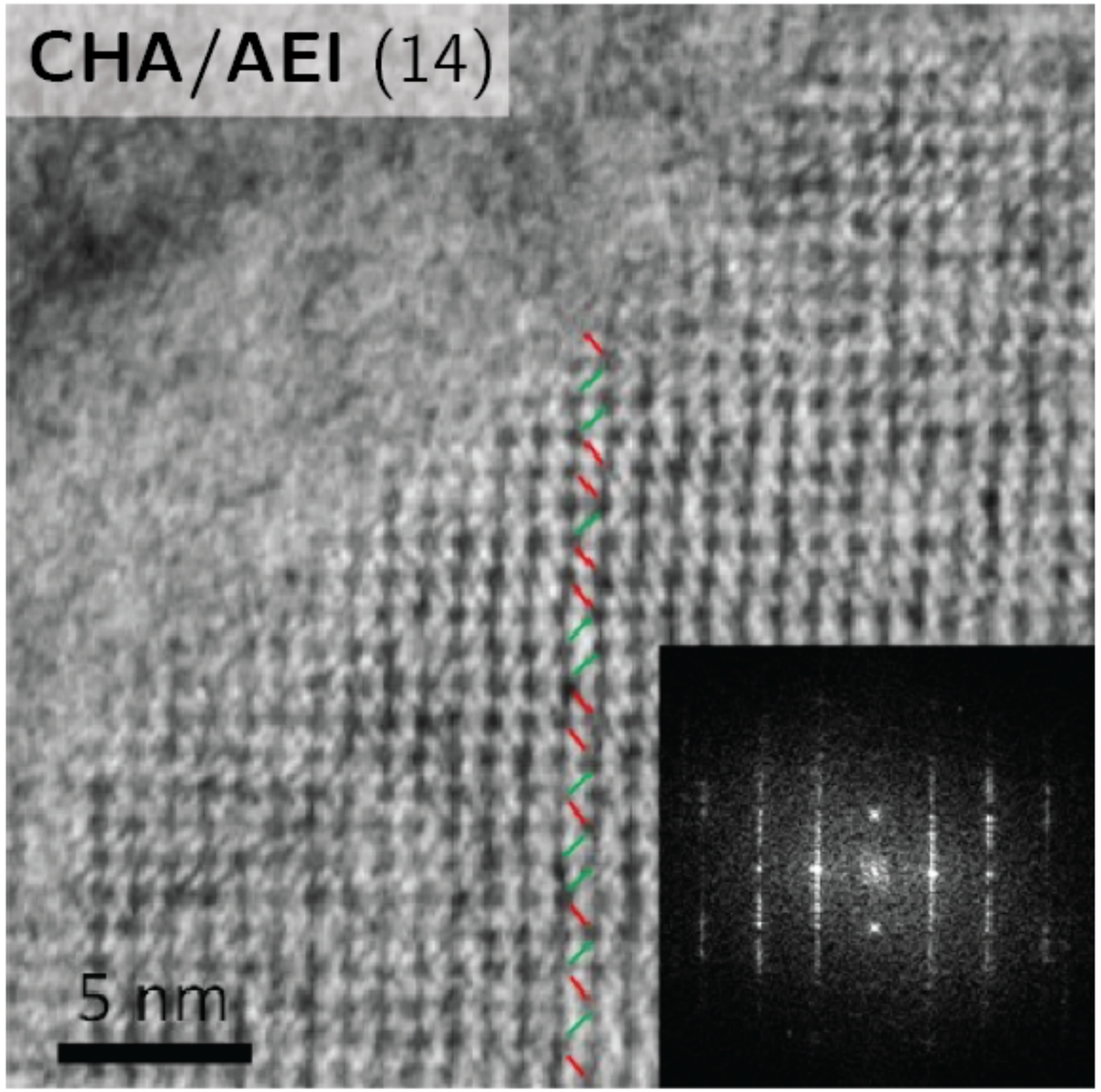
Design of bi-selective OSDA for **CHA/AEI**

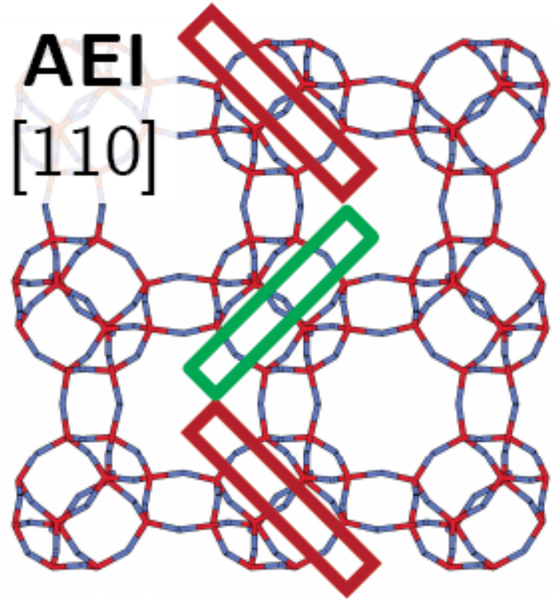


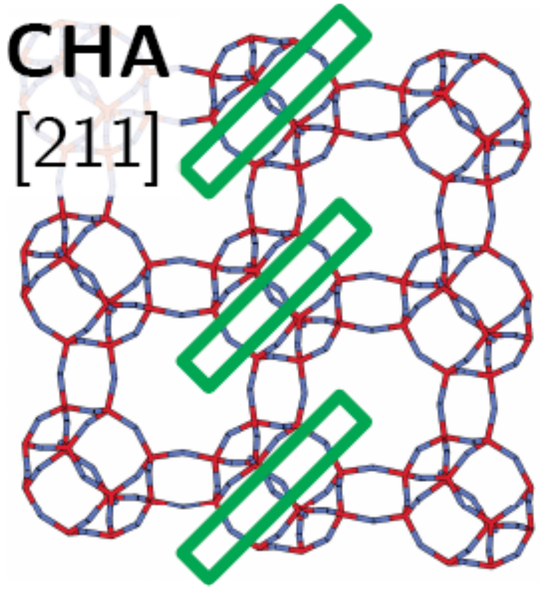


Characterization of **CHA/AEI** aluminosilicate intergrowth









with M. Moliner, A. Corma and their team (ITQ, Spain)
and T. Willhammar (Stockholm Univ., Sweden)

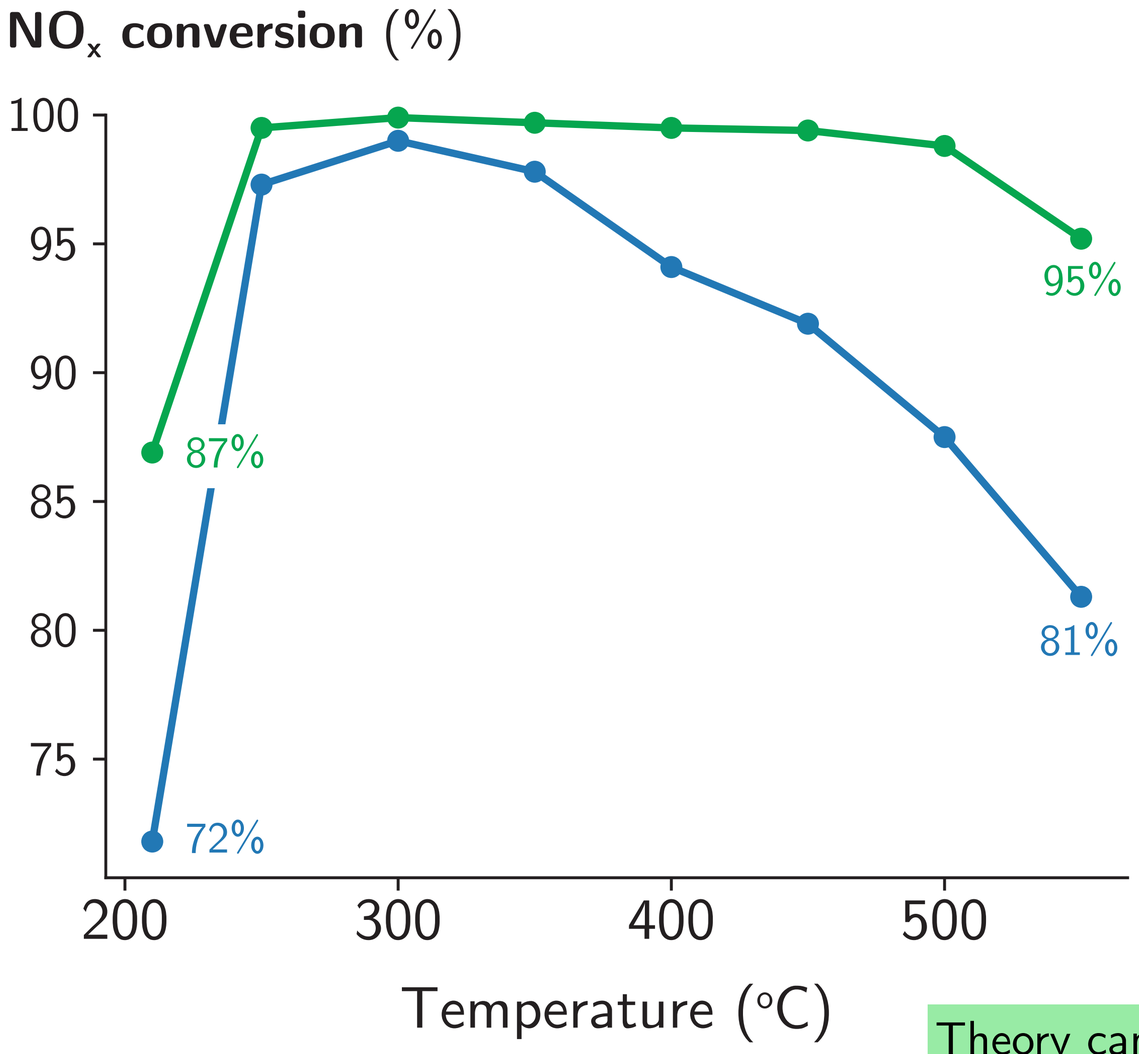
Application of **CHA/AEI** as catalysts

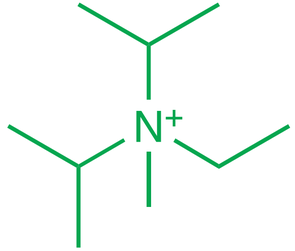
MTO

selectivity at X_{95} (%)

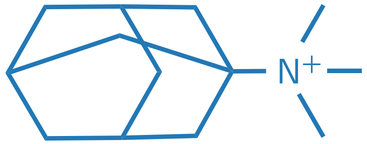
	CHA	CHA/AEI	AEI
$C_2^=$	48	40	23
$C_3^=$	36	43	48
$C_4^=$	12	16	22

NOx-SCR





Cu-CHA/AEI



Cu-CHA

Theory can control synthesis and improve catalytic behavior

Patent ES2829385A1 (2021)
with M. Moliner, A. Corma and their team (ITQ, Spain)

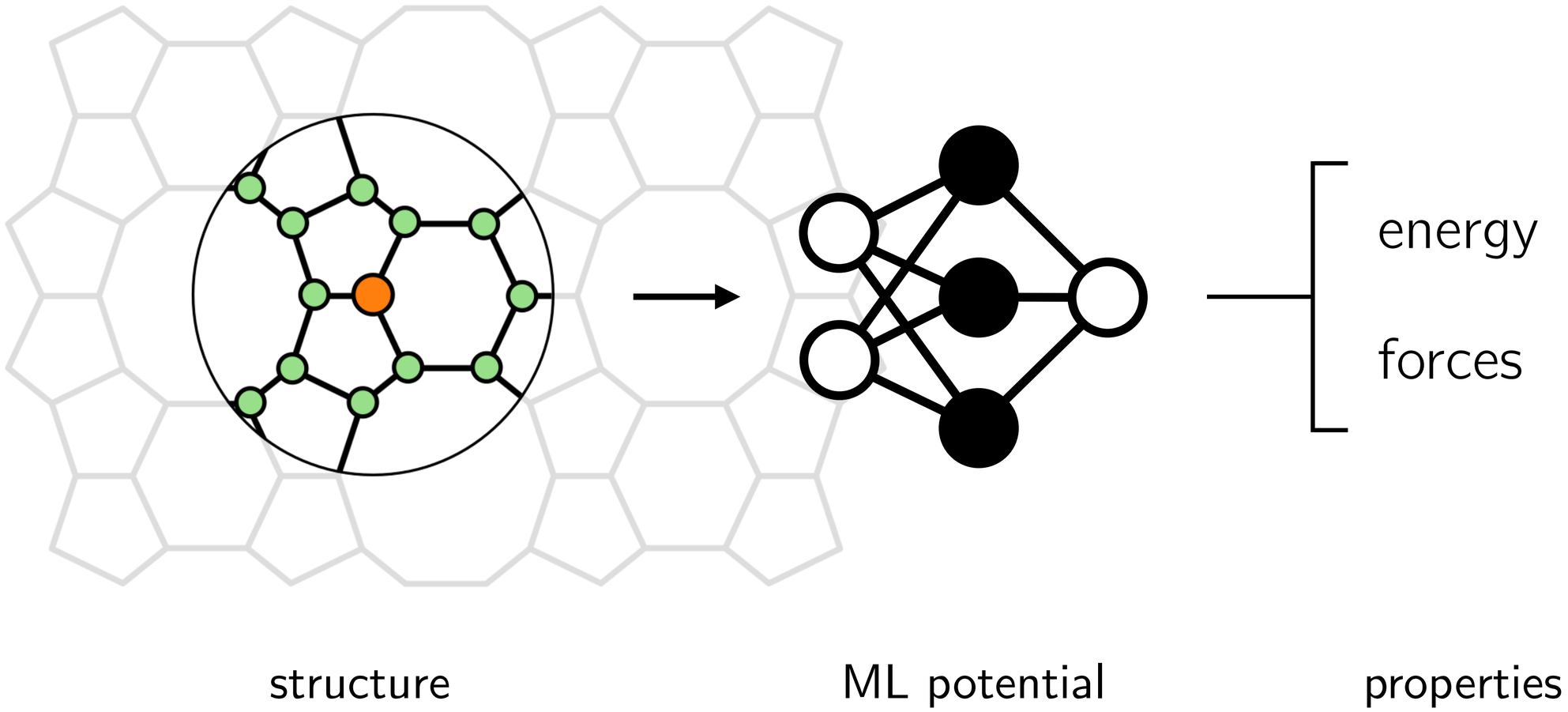
6. Towards simulating synthesis/reactions with ML

Increasing ML robustness with chemical adversarial attacks

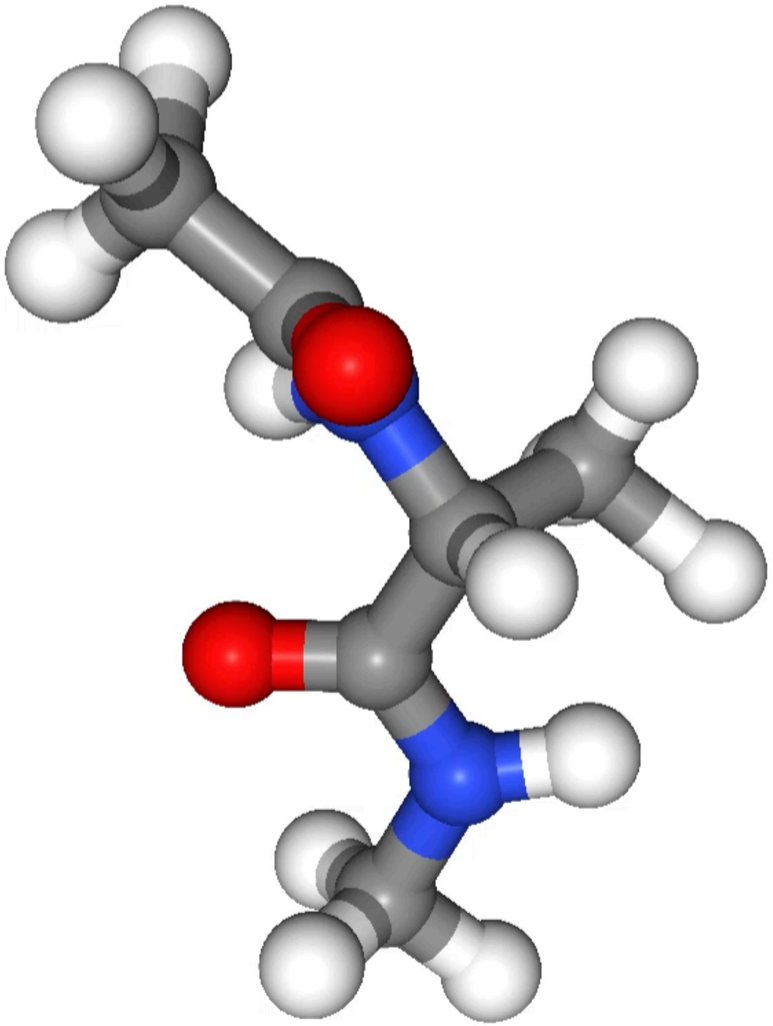
with A. R. Tan (MIT PhD'24)

D. Schwalbe-Koda*, A. R. Tan* et al. *Nat. Commun.* (2021)

We can accelerate simulations of zeolite structures using ML potentials

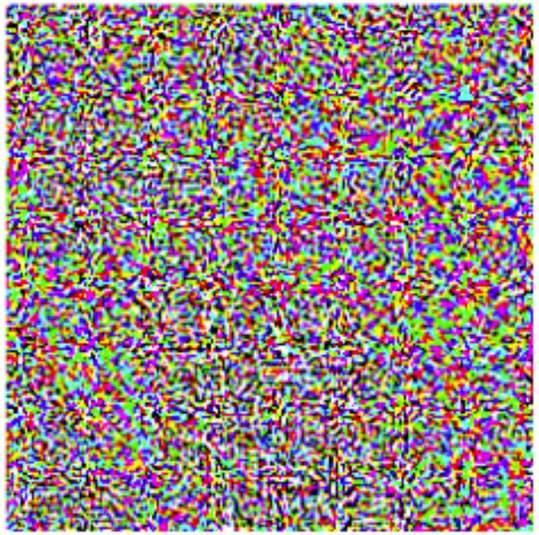


However, ill-trained ML potentials are not robust...



Neural networks are susceptible to adversarial attacks

$+ .007 \times$



$=$

$\boldsymbol{x}$

"panda"
57.7% confidence

$\text{sign}(\nabla_{\boldsymbol{x}} J(\boldsymbol{\theta}, \boldsymbol{x}, y))$

"nematode"
8.2% confidence

$\boldsymbol{x} +$
$\epsilon \text{sign}(\nabla_{\boldsymbol{x}} J(\boldsymbol{\theta}, \boldsymbol{x}, y))$
"gibbon"
99.3 % confidence

Szegedy et al. (2013)
Goodfellow et al. (2014)

Increasing NN robustness to adversarial attacks

$$\min_{\theta} \mathbb{E}_{(x,y)\sim\mathscr{D}} \left[\max_{\delta\in\Delta} \mathscr{L}(h_\theta(x+\delta), y) \right]$$

find the NN weights that minimize

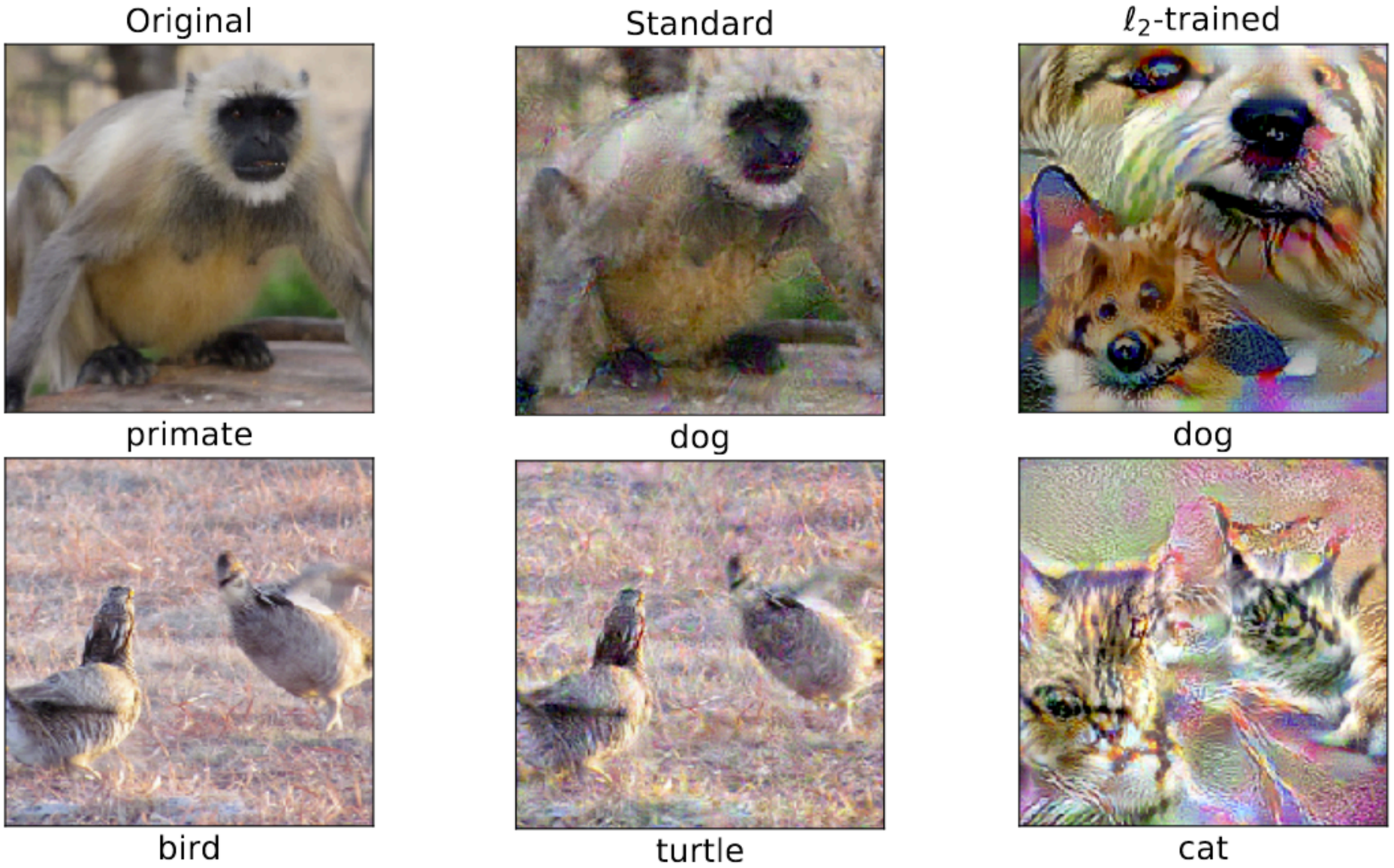
across the whole dataset under study

and for perturbations δ in the set of allowed perturbations Δ

the perturbed loss function

Tsipras et al. ICLR'19

Qualitative results of robust NNs



Question: how to do this for NN potentials?

Tsipras et al. ICLR'19

Objective of a robust neural network regressor

E_δ, F_δ come from reference calculations in an AL loop

$$\min_{\theta} \mathbb{E}_{(X,E,\mathbf{F})\sim\mathscr{D}} \left[\max_{\delta\in\Delta} \mathscr{L}(X_\delta, E_\delta, \mathbf{F}_\delta; \theta) \right]$$

find the NN weights that minimize

across the whole dataset under study

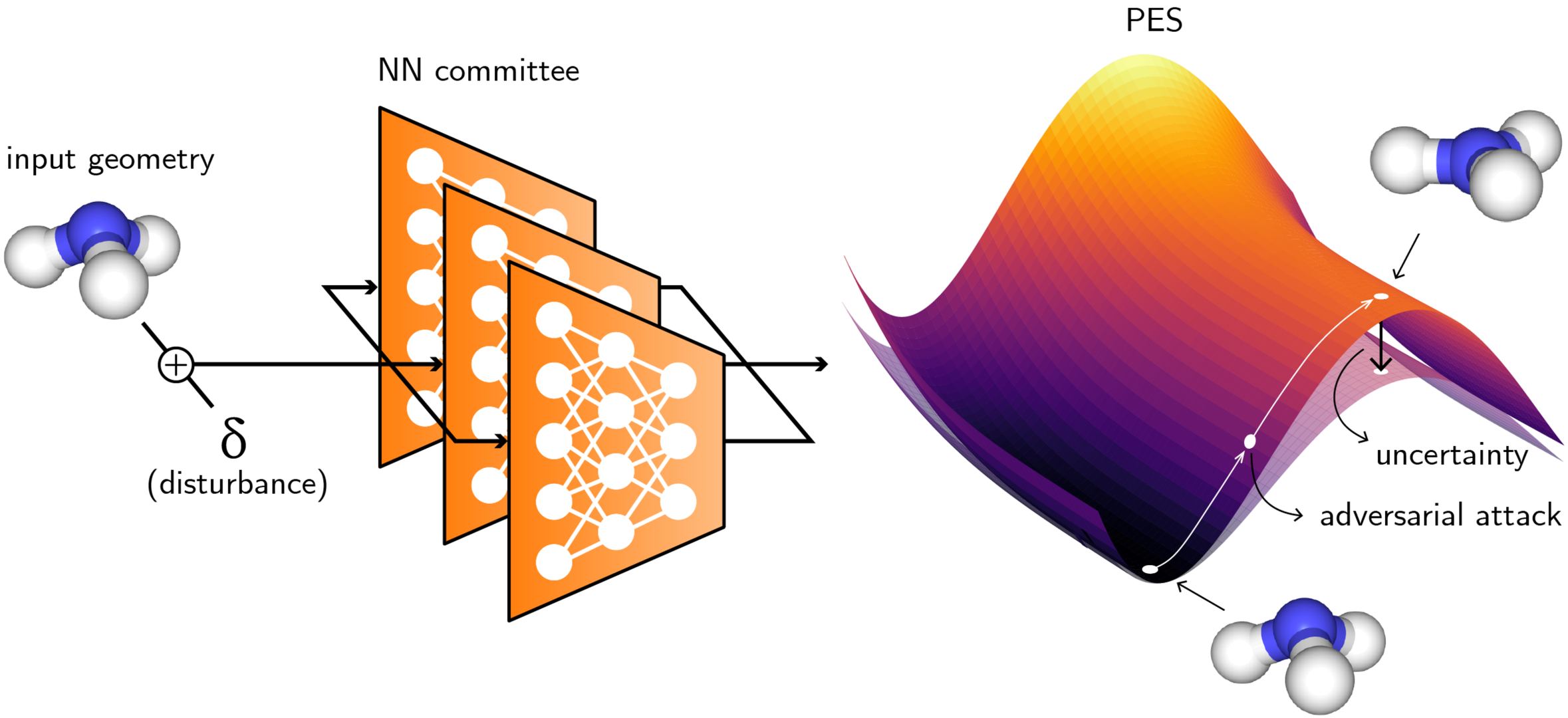
and for perturbations δ in the set of allowed perturbations Δ

the perturbed loss function of the <u>regressor</u>

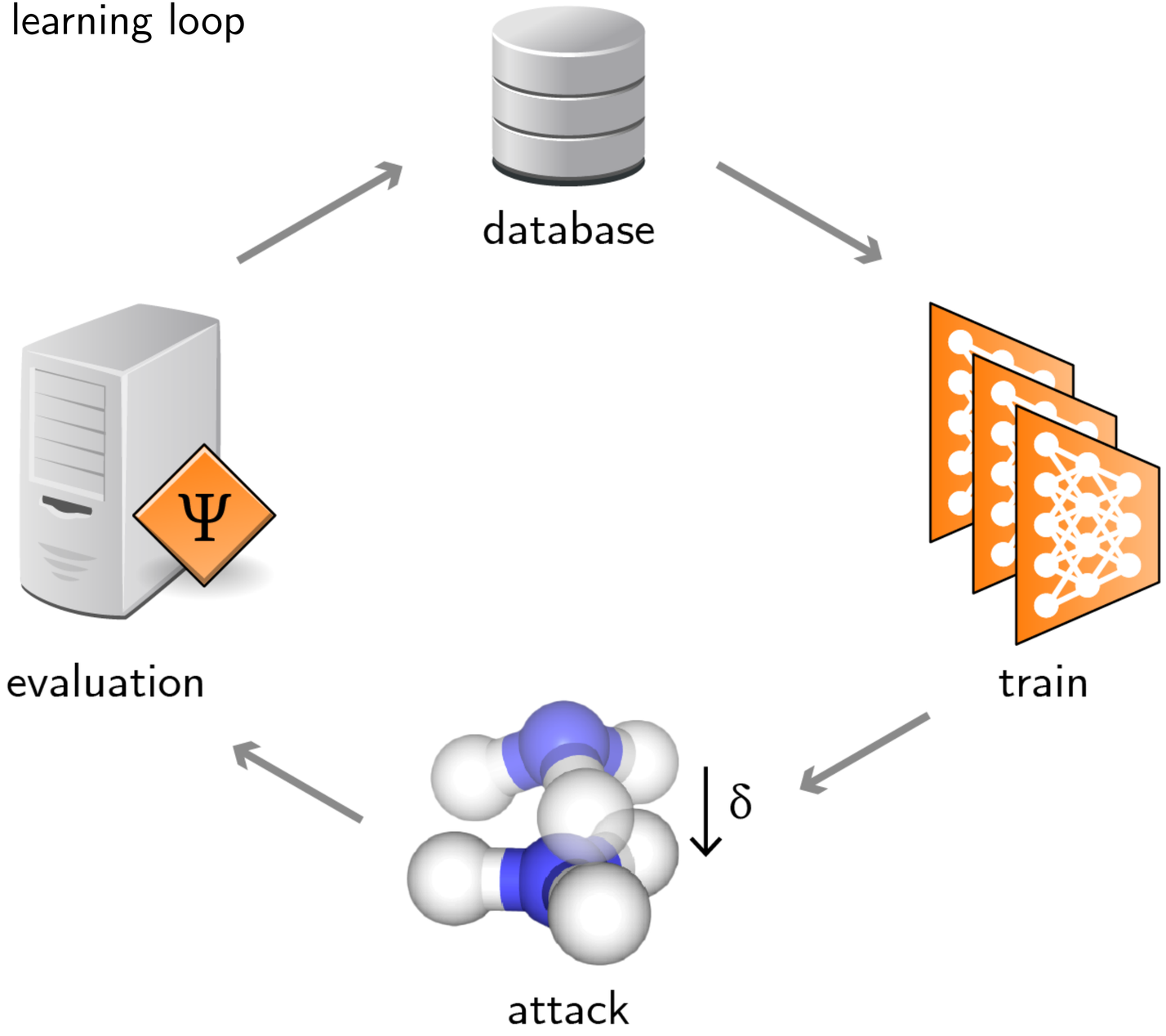
Question: how to generate the perturbed samples and their ground truth values?

Idea: find geometries that maximize the epistemic uncertainty of the NN potential!

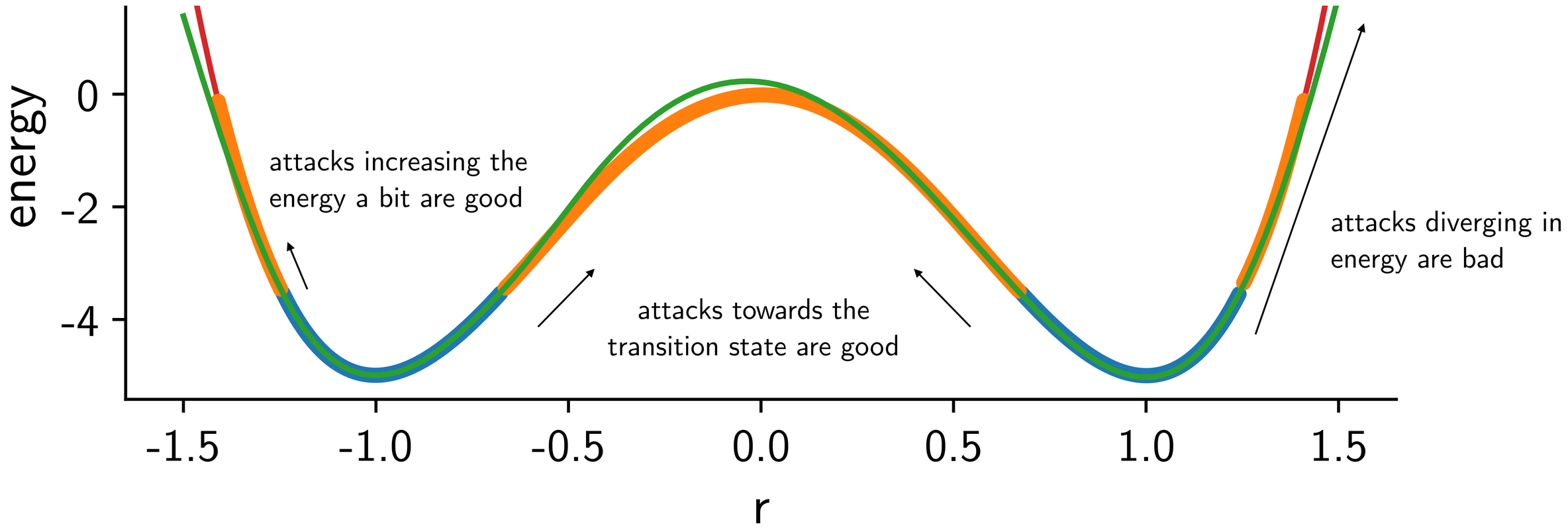
Adversarial loss depends on the uncertainty



Robust training is an active learning loop



Sample new points through adversarial attacks



Goals of a good adversarial attack:

- Find points of maximum uncertainty
- Penalize going towards crazy high energies

$$\min_{\theta} \mathbb{E}_{(X,E,\mathbf{F})\sim\mathscr{D}} \left[\max_{\boxed{\delta\in\Delta}} \mathscr{L}(X_\delta, E_\delta, \mathbf{F}_\delta; \theta) \right]$$

Loss function for adversarial attack

Construct partition function from training set:

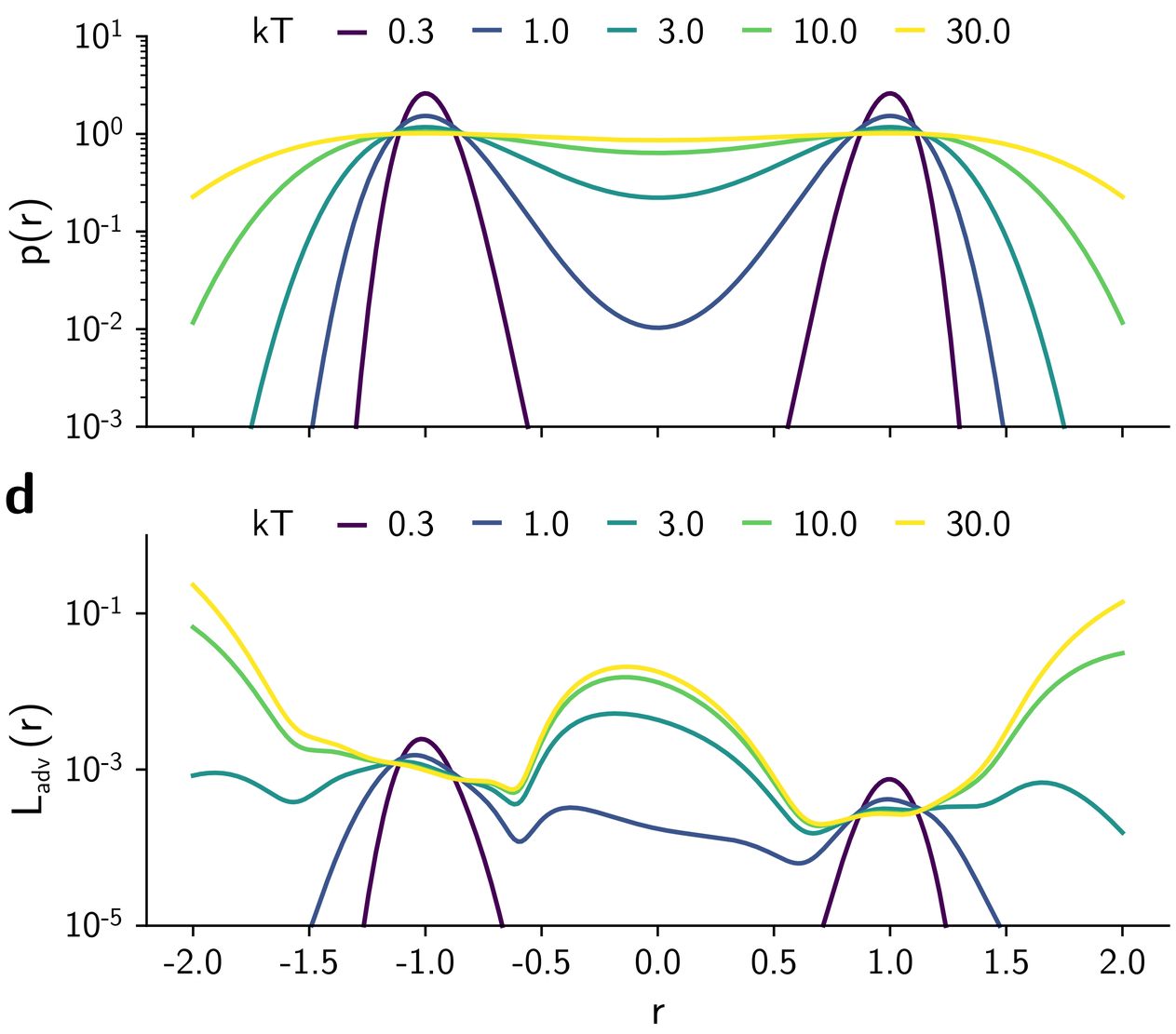
$$Q = \sum_{(X,E,\mathbf{F})\in\mathscr{D}} \exp\left(\frac{-E}{kT}\right)$$

Estimate Boltzmann probability given the mean energy from NN ensemble:

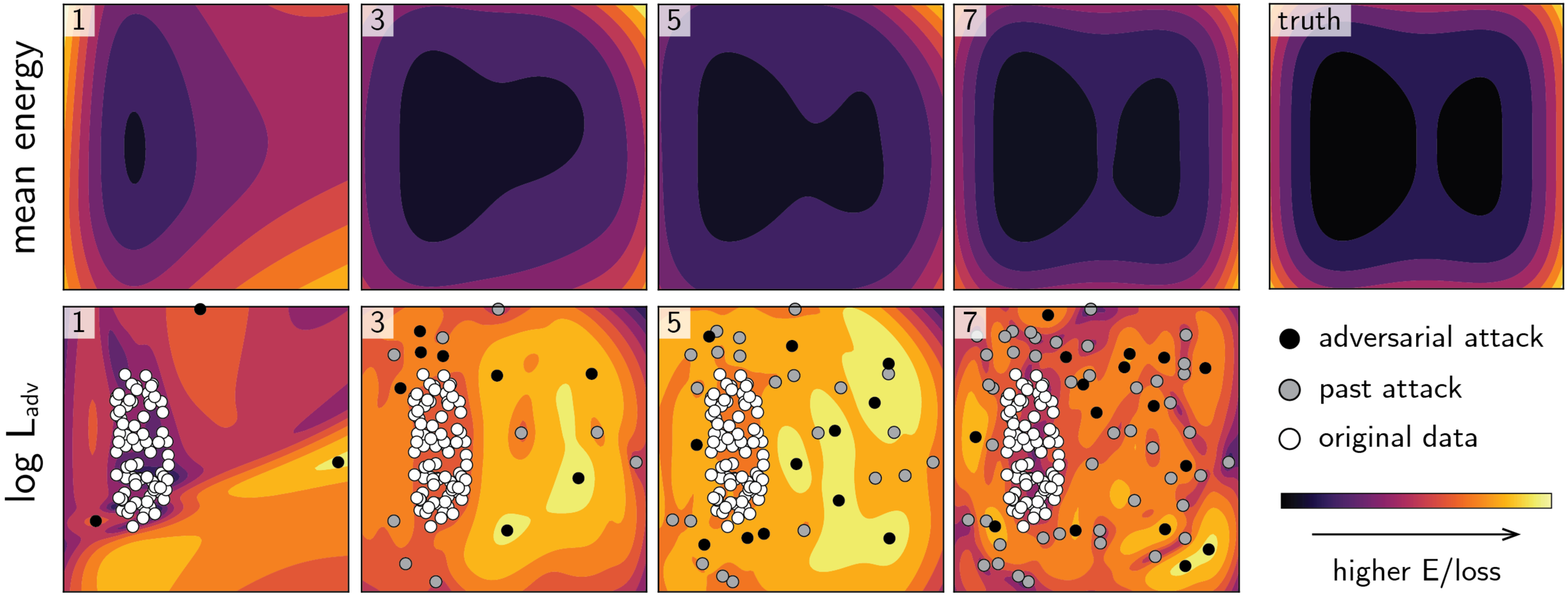
$$p(X_\delta) = \frac{1}{Q}\exp\left(\frac{-\bar{E}(X_\delta)}{kT}\right)$$

The final adversarial objective then becomes

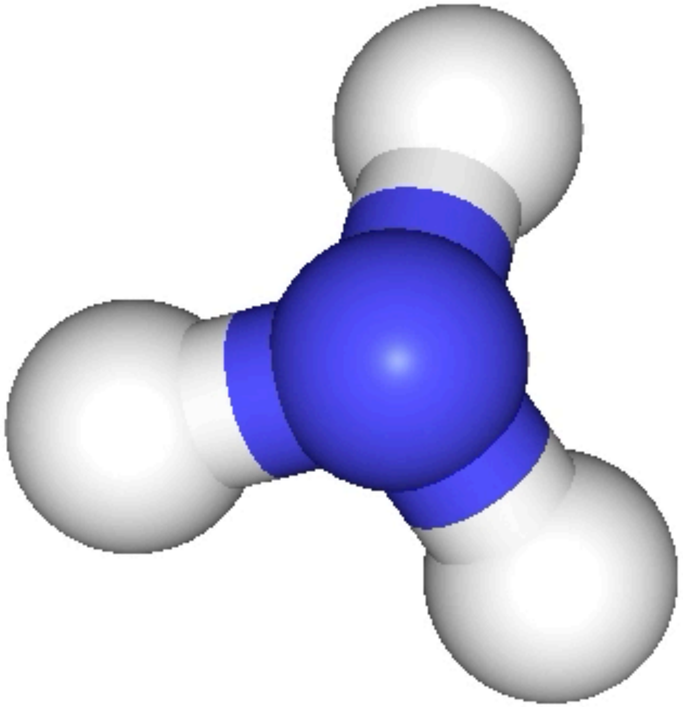
$$\max_\delta \mathscr{L}(X,\delta;\theta) = \max_\delta p(X_\delta)\cdot\sigma_F^2(X_\delta)$$

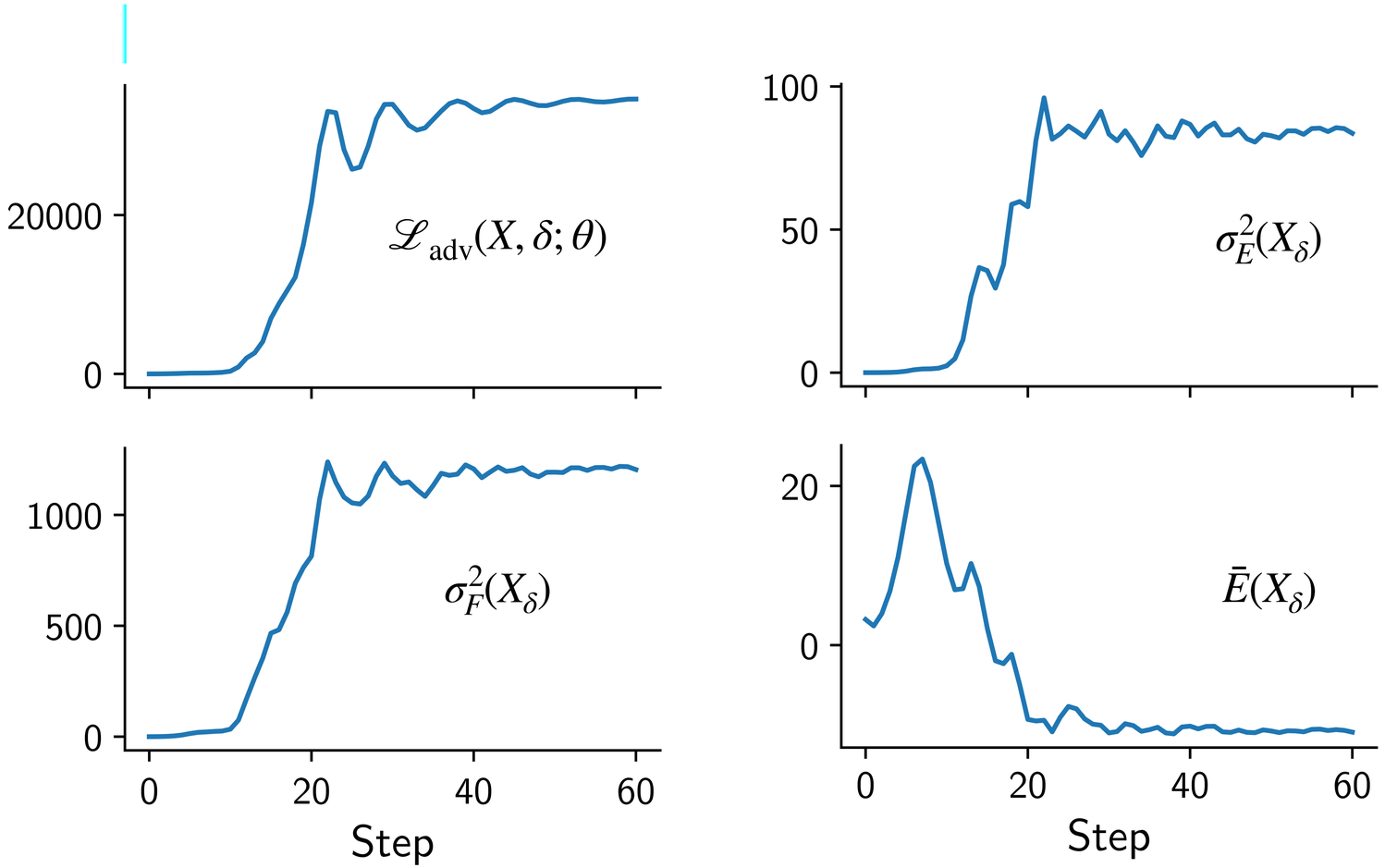


Adversarial attacks for 2D double well

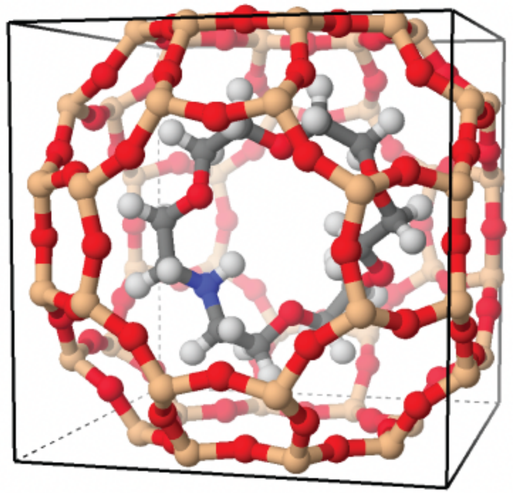


How adversarial attacks look like for molecules?



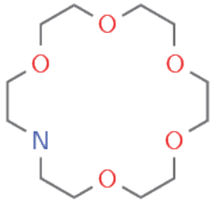


How efficient is the active learning with this technique?

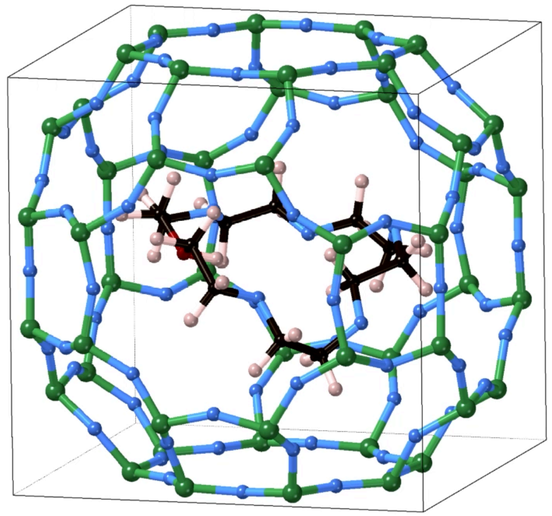


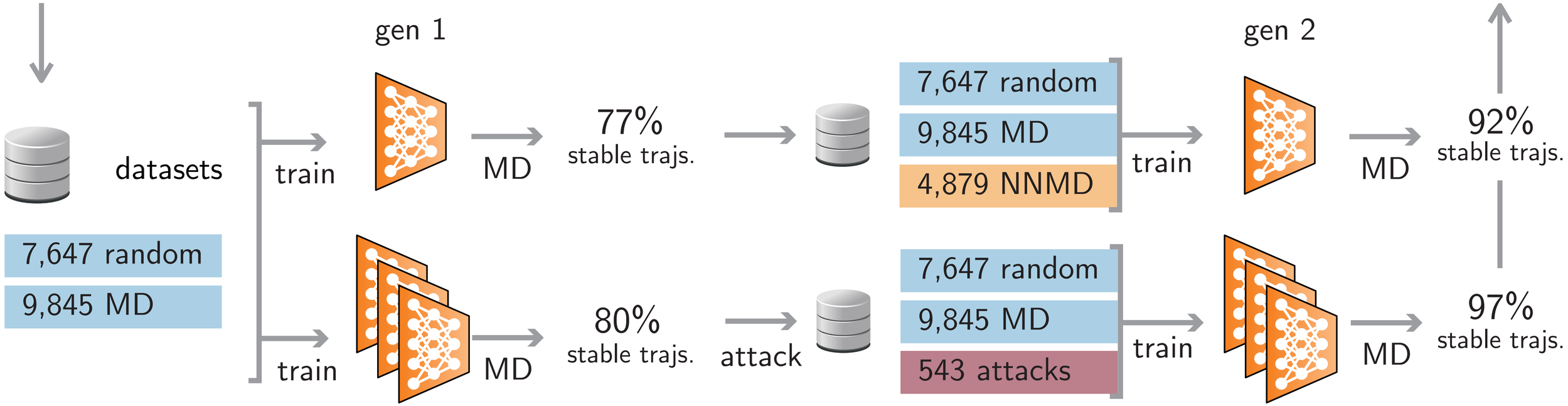
LTA

+



and 542 other zeolite-OSDA pairs

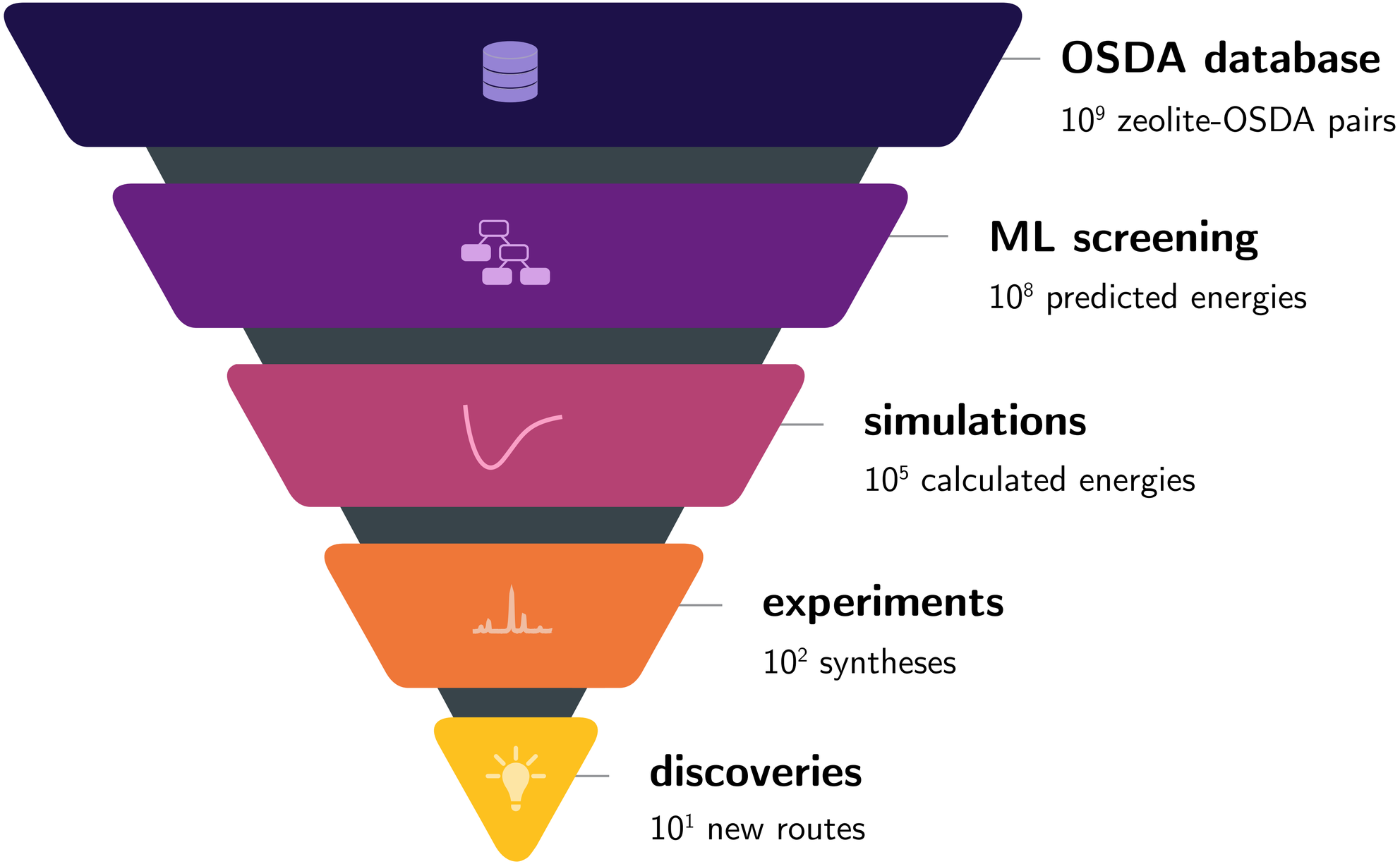


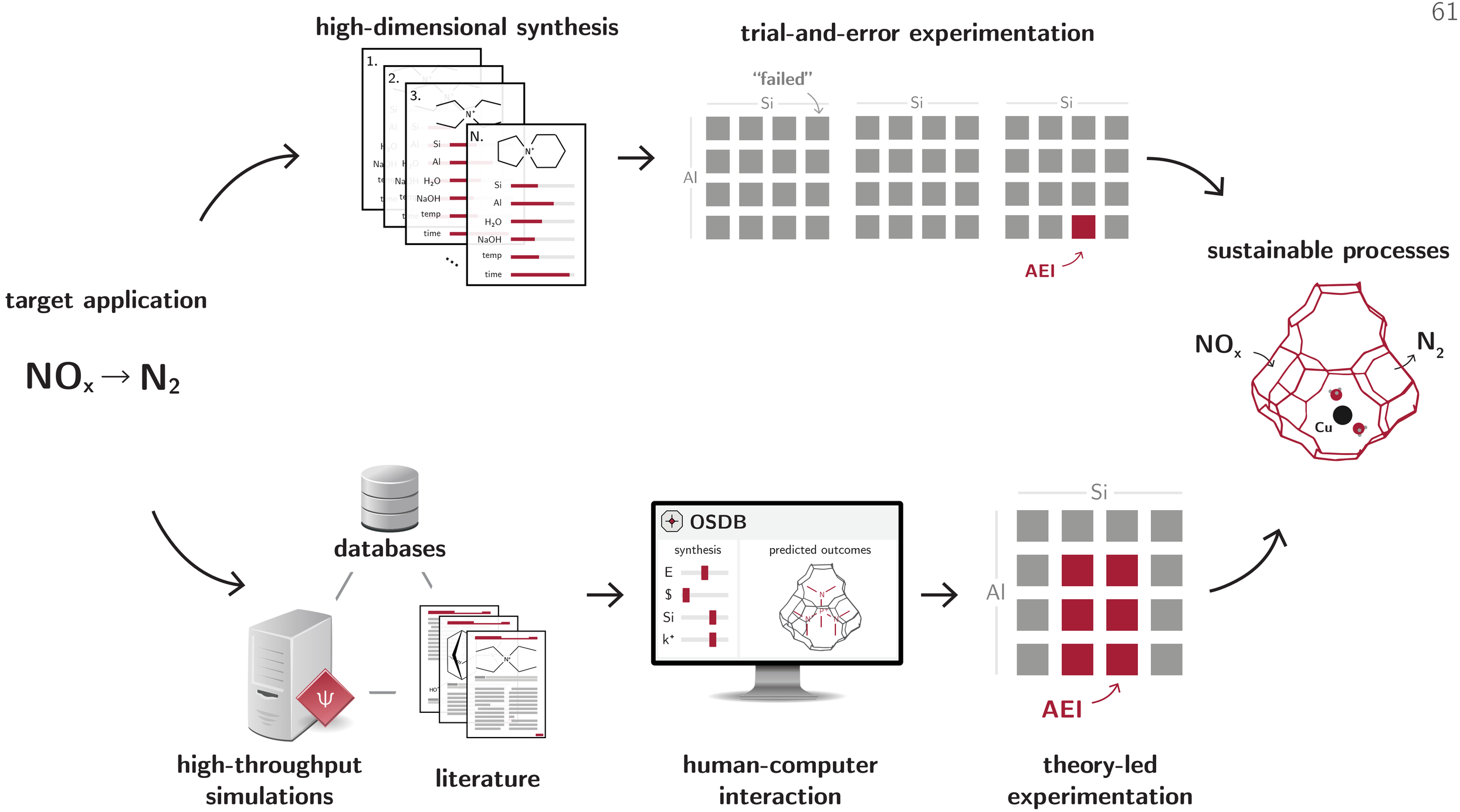


Summary

Take-home messages

Machine learning guides the high-throughput synthesis of zeolites




high-dimensional synthesis
1.
2.
3.
N.
Si
Al
H_2O
NaOH
temp
time
trial-and-error experimentation
"failed"
Si
Al
AEI
sustainable processes
NO_x
N_2
Cu
target application
$NO_x \rightarrow N_2$
databases
high-throughput simulations
literature
OSDB
synthesis
predicted outcomes
E
$
Si
k+
human-computer interaction
theory-led experimentation

Acknowledgements

MIT
Prof. Rafael Gómez-Bombarelli
Prof. Elsa Olivetti
Prof. Yuriy Román-Leshkov
Aik Rui Tan
Omar A. Santiago-Reyes
Soonhyoung Kwon
Dr. Zach Jensen

ITQ-UPV, Spain
Prof. Avelino Corma
Dr. Manuel Moliner
Dr. Cecilia Paris
Estefania Bello

Stockholm University
Dr. Tom Willhammar

LLNL
PLS/Materials Science Division

Funding

MIT Energy Initiative

XSEDE
Extreme Science and Engineering Discovery Environment

Massachusetts Institute of Technology

Lawrence Livermore National Laboratory

LLNL-PRES-840662, LDRD #22-ERD-055
Prepared by LLNL under Contract DE-AC52-07NA27344.