From crystal structure to properties: accelerating discovery of novel materials

François-Xavier Coudert

@fxcoudert

www.coudert.name

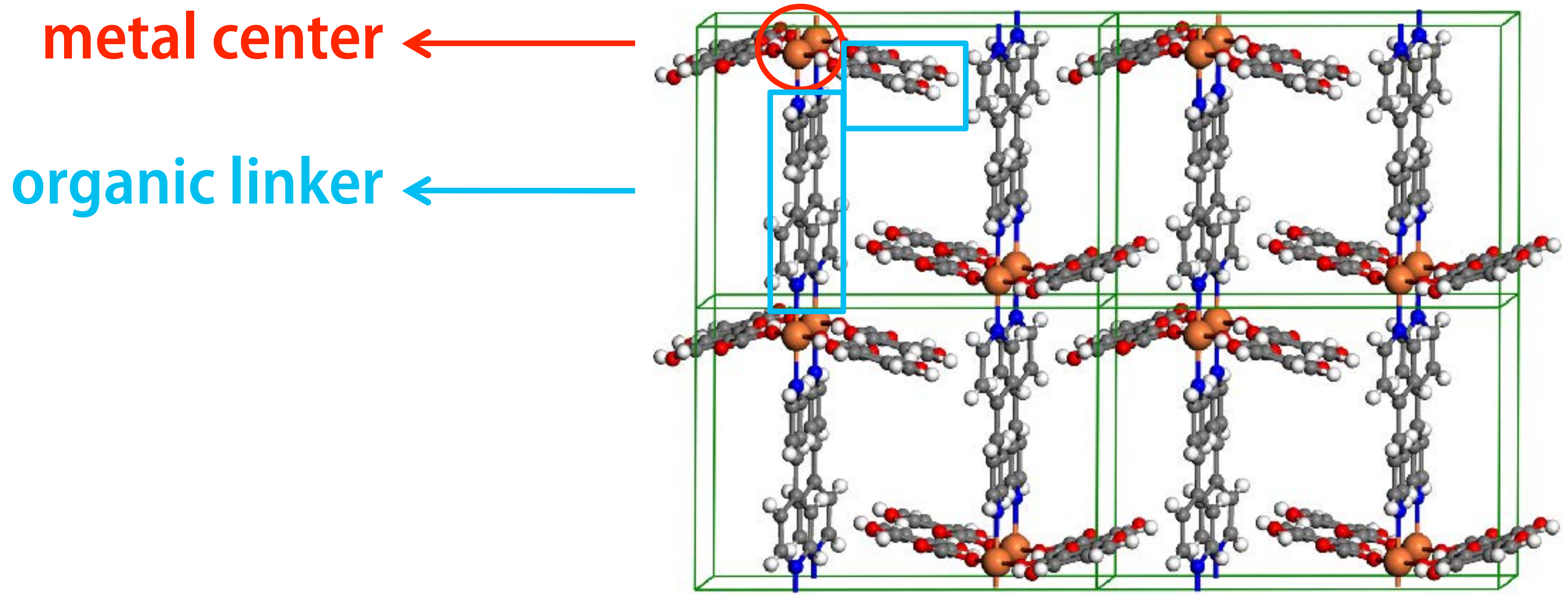
PSL UNIVERSITÉ PARIS

Chimie ParisTech

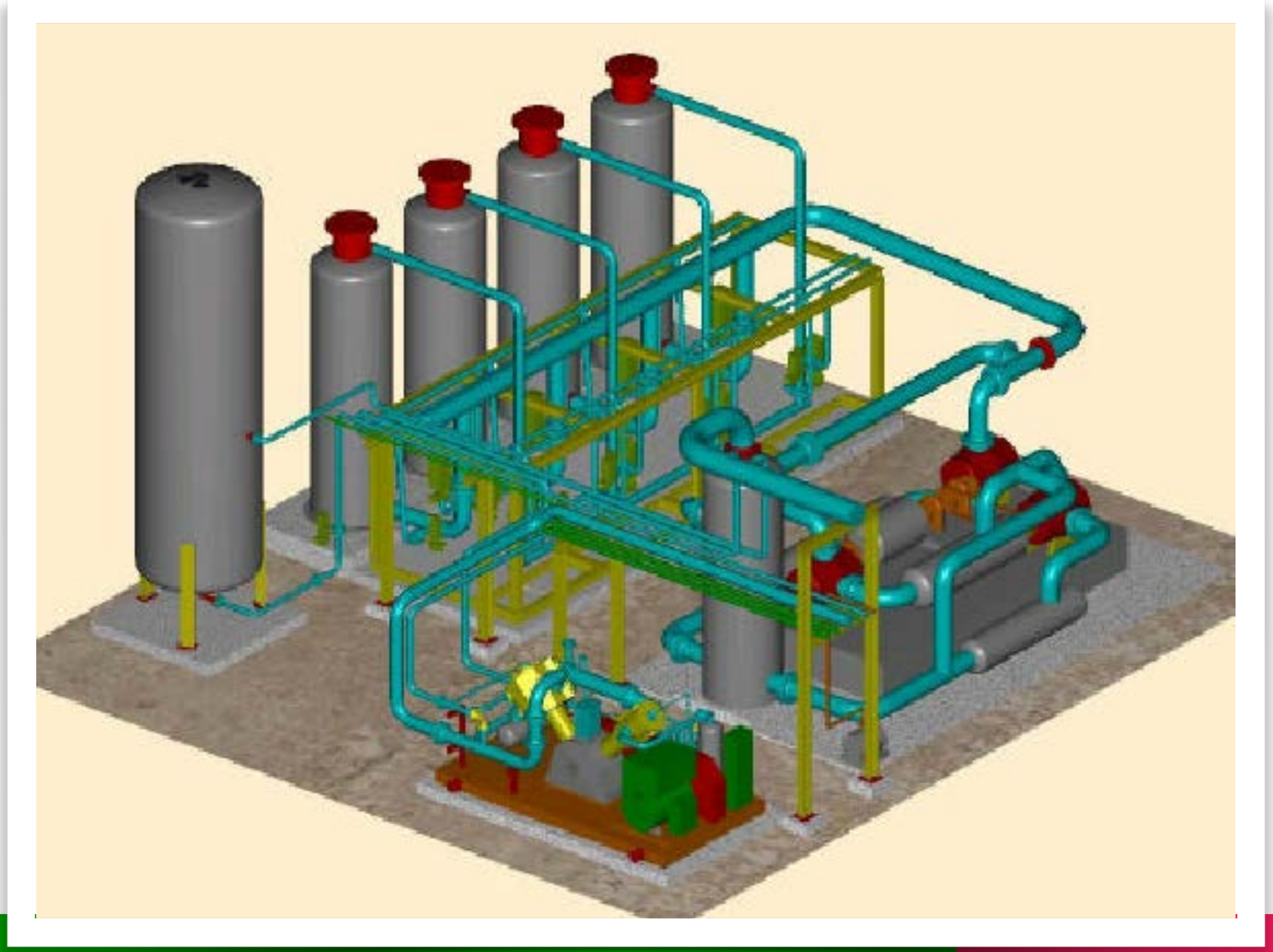
cnrs

Metal–Organic Frameworks

Cristalline, organic–inorganic hybrid nanoporous materials



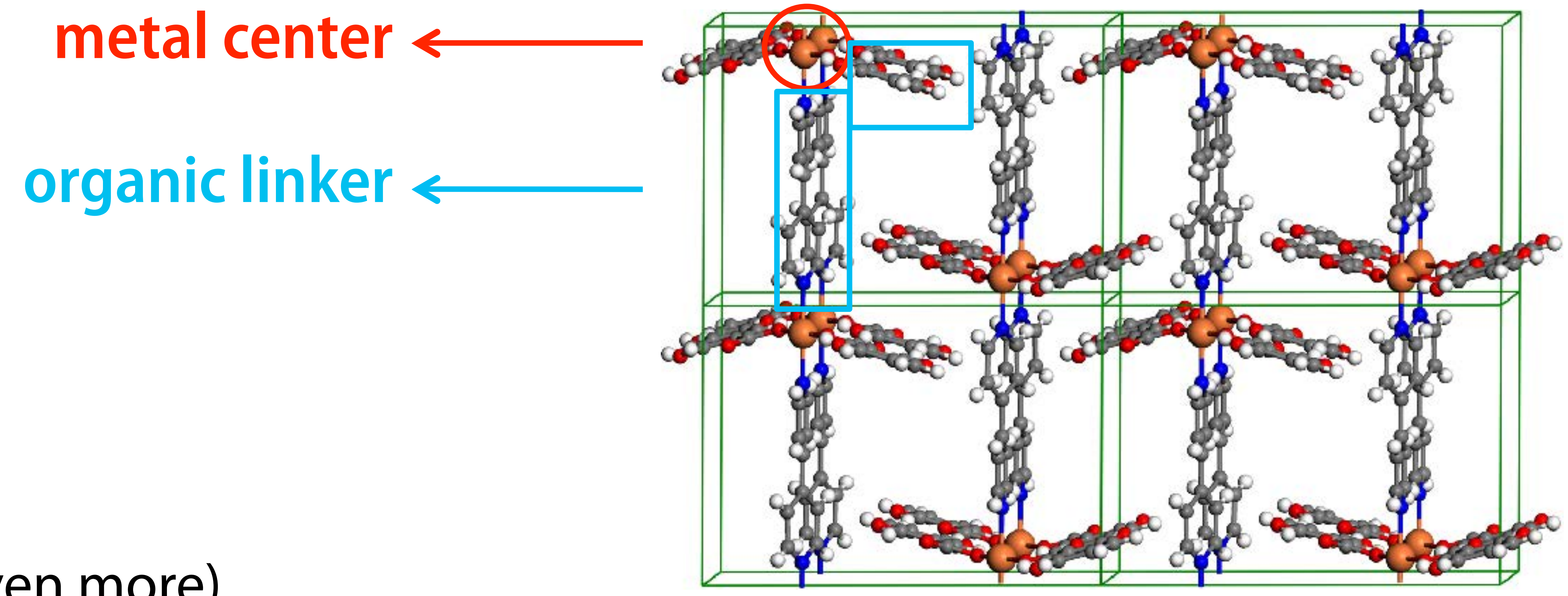
- Flexibility of coordination chemistry: pore geometry and topology
- Versatility of organic chemistry: pore size and internal surface



- Applications: gas adsorption, catalysis, sensing, delivery, …
- **High structural flexibility of their frameworks**
- **Important limitation for applications: hydrothermal & mechanical stability**

Metal–Organic Frameworks

Cristalline, organic–inorganic hybrid nanoporous materials



- Metals (many) and inorganic blocks (even more)
- Organic linkers… and functionalization
- Multivariate MOFs are possible
- Topology
- Guest molecule… or guests

- Thermodynamic space: temperature, pressure, composition

Soft Porous Crystals

nature chemistry

REVIEW ARTICLE

PUBLISHED ONLINE: 23 NOVEMBER 2009 | DOI: 10.1038/NCHEM.444

Soft porous crystals

Satoshi Horike[1,2], Satoru Shimomura[1] and Susumu Kitagawa*[1-3]

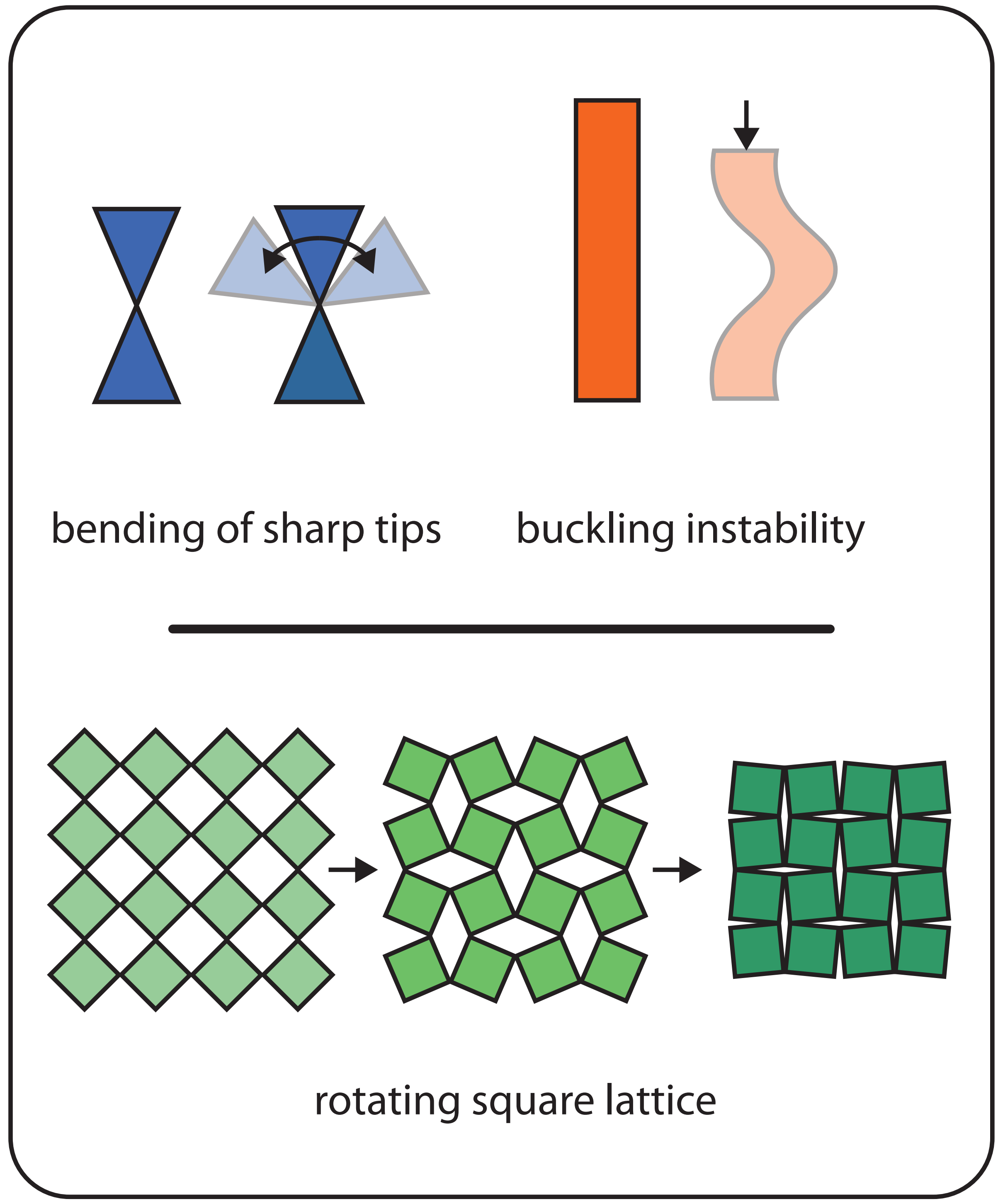
The field of host–guest complexation is intensely attractive from diverse perspectives, including materials science, chemistry and biology. The uptake and encapsulation of guest species by host frameworks are being investigated for a wide variety of purposes, including separation and storage using zeolites, and recognition and sensing by enzymes in solution. Here we focus on the concept of the cooperative integration of 'softness' and 'regularity'. Recent developments on porous coordination polymers (or metal–organic frameworks) have provided the inherent properties that combine these features. Such soft porous crystals exhibit dynamic frameworks that are able to respond to external stimuli such as light, electric fields or the presence of particular species, but they are also crystalline and can change their channels reversibly while retaining high regularity. We discuss the relationship between the structures and properties of these materials in view of their practical applications.

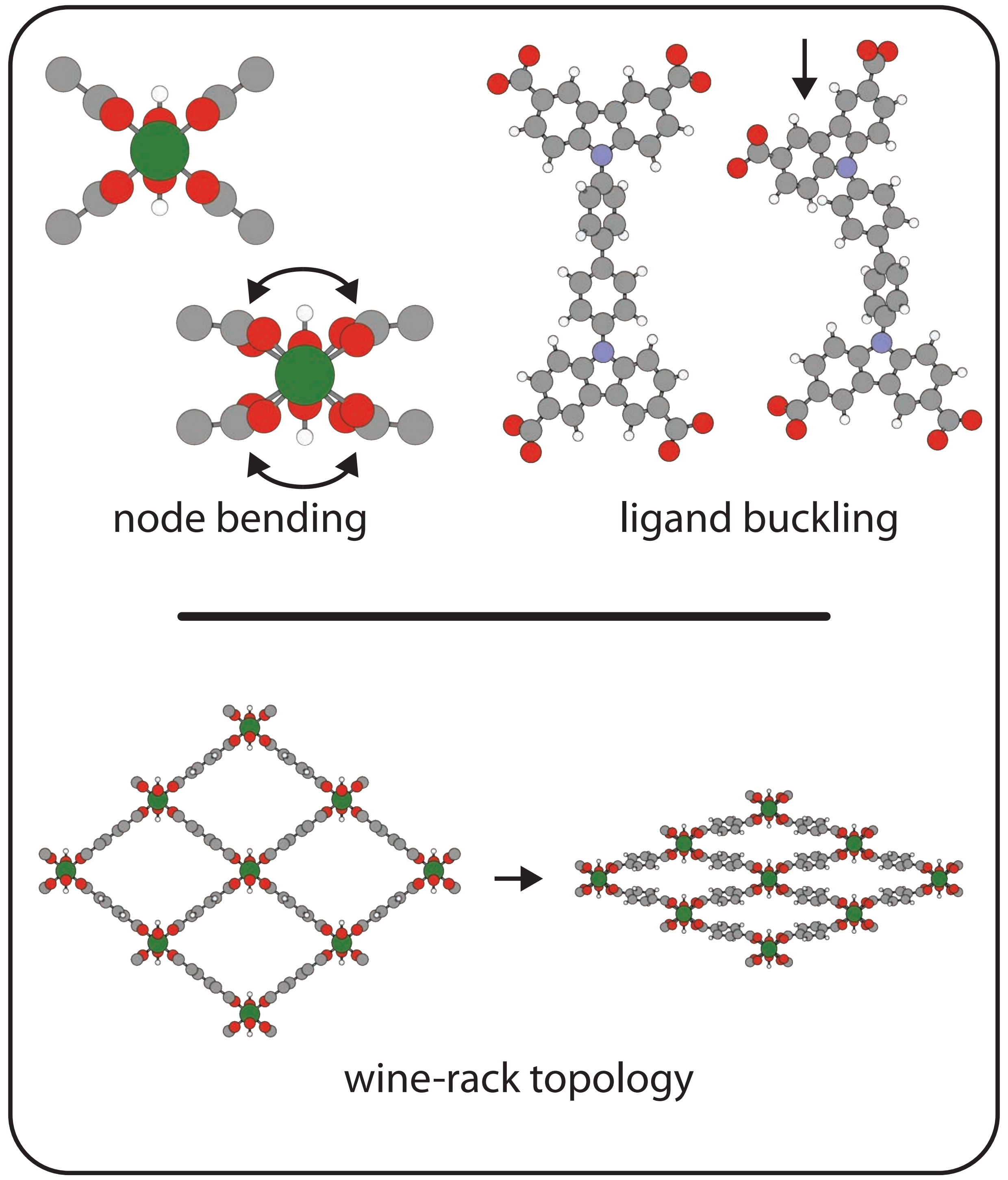
"Soft porous crystals are defined as porous solids that possess both a highly ordered network and **structural transformability**. They are bistable or multistable crystalline materials with long-range structural ordering, a **reversible transformability between states**, and permanent porosity"

Meta-MOFs

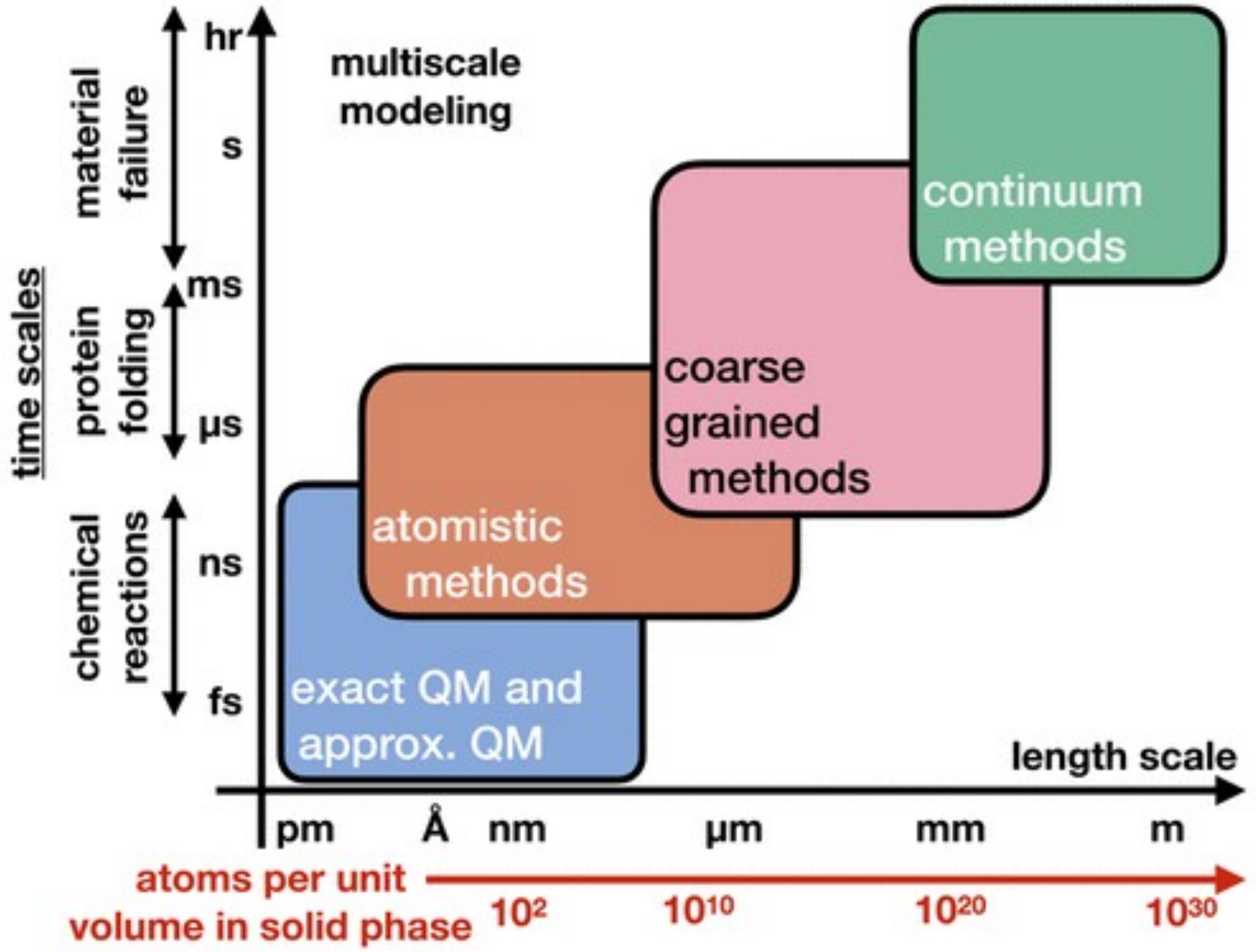
Coudert & Evans, *Coord. Chem. Rev.* **2019**

metamaterials	meta-MOFs
composite material with a structure that exhibits properties not usually found in natural materials	**negative thermal expansion**
	negative compressibility
	negative adsorption
	breathing
μετά = beyond	**chiral induction**
	…





Molecular modelling



Keith et al, *Chem. Rev.* **2021**

Molecular modelling

Thermodynamics

adsorption isotherms

phase diagram, nature of transitions

Dynamics

IR/Raman, vibrational spectra

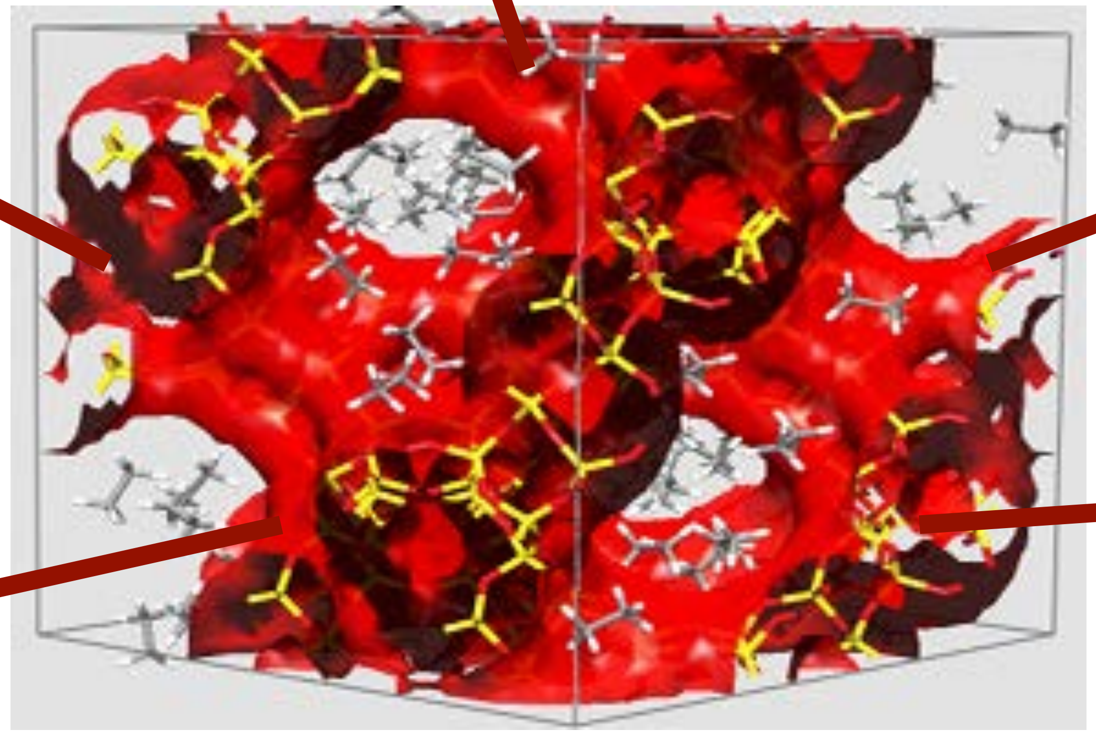
orientation/transl. dynamics

H bond networks

Structure

localization of adsorbates

adsorbent structural transitions



Electronic properties

dipole moments

hydrogen bonds

Reactivity

(hydro)thermal stability

chemical reactions, catаysis

+ machine learning

to allow faster exploration of the huge space of materials

ML methods for chemical sciences

- **Property prediction**: from structure or from composition (supervised learning, data obtained experimentally or computationally)
- **High-throughput screening**: applying predictor at large scale
- Analysis and exploration of diversity, **clustering of molecules**
- **Generative ML methods**: creating new molecules, new materials
- **Text and data mining**: a lot of information in published literature, in notebooks
- **AI for synthesis prediction**: propose a synthesis method/protocal, possibly drive robotic chemistry lab
- ML to **improve computational chemistry**: using machine learning to design new force fields, new DFT functionals, etc.
- … and many more…

Why data-based methods?

Daglar & Keskin, 2020

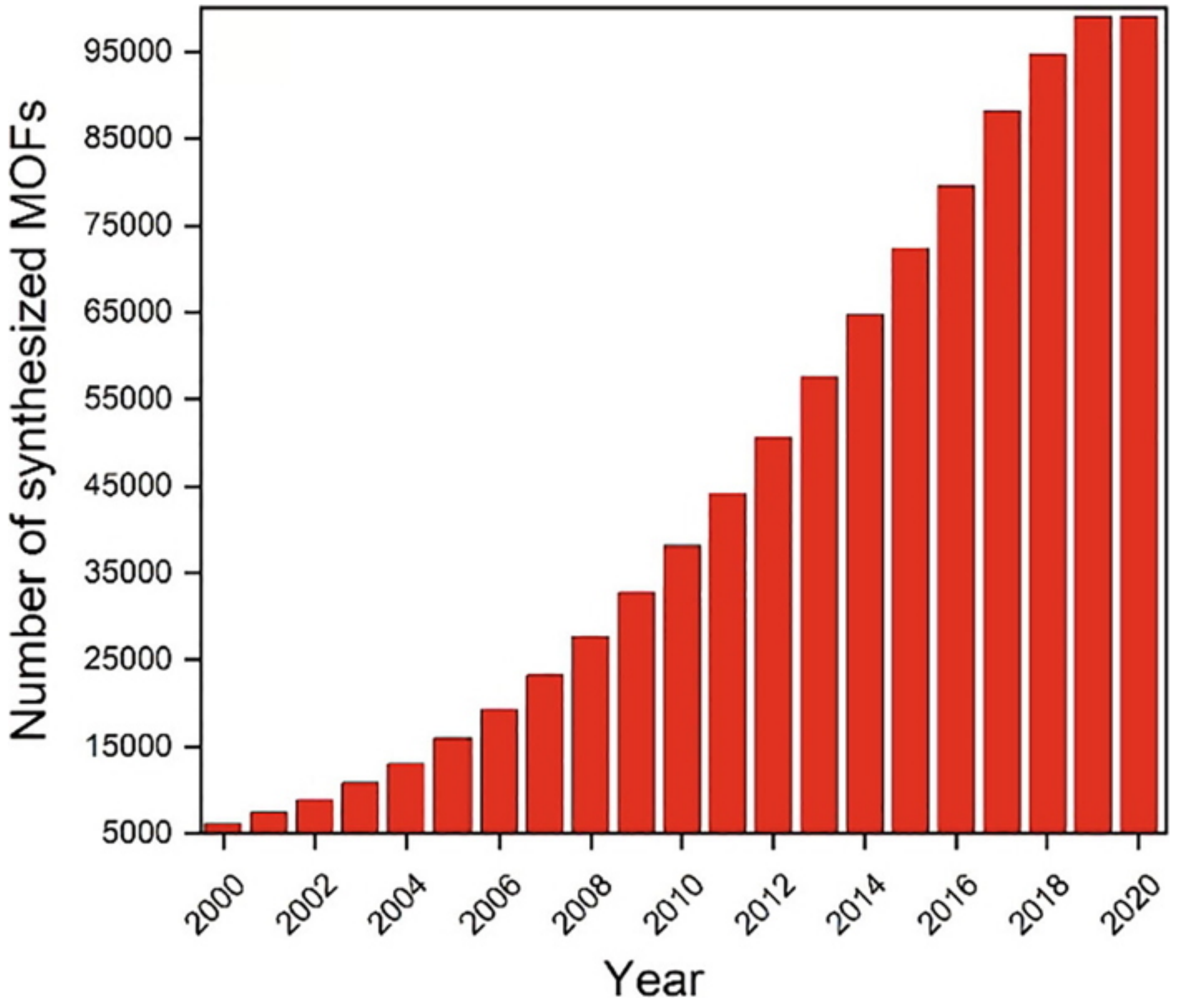


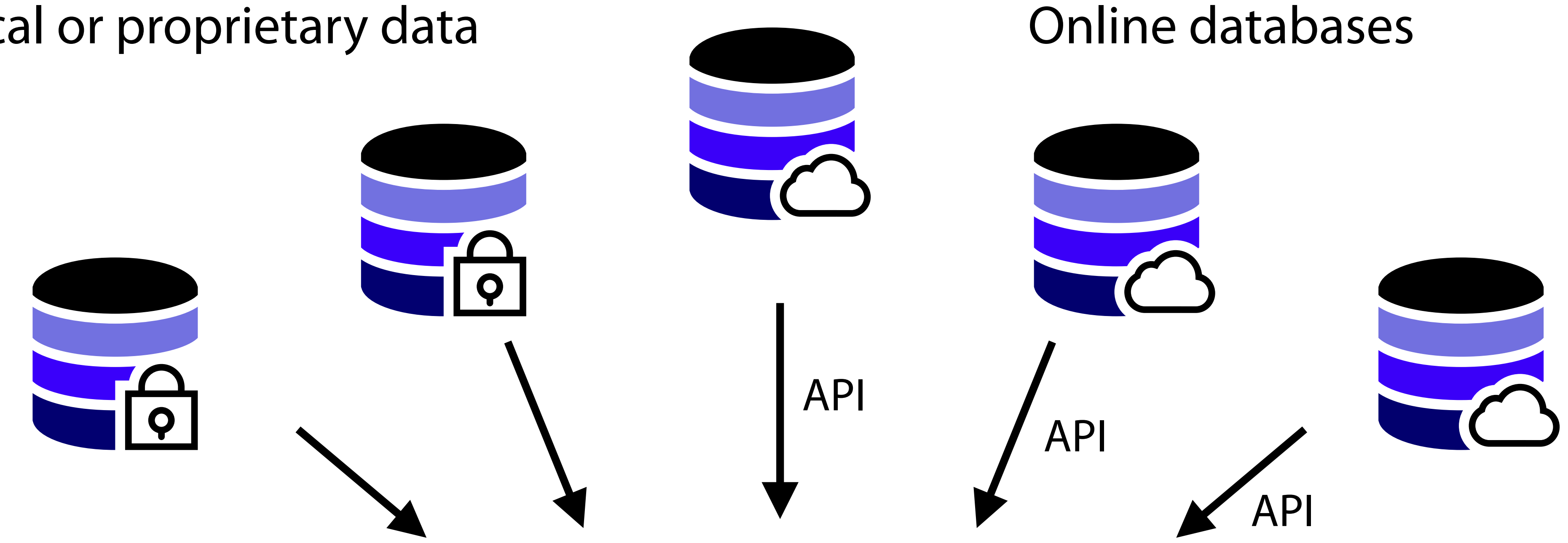
Figure 4. Most frequent emojis in followers' profiles. National flags were excluded from the analysis.

Coudert, *Chem. Mater.* 2023

- ⭐ Since Sunday,
  61 MOF papers were tweeted

Local or proprietary data

Online databases



Cleaning & preprocessing

↓

Training data

↓

Featurization

↓ Learning process

Input data → **ML-trained model** → **Predictions**

⚠ we have many databases of structures, but **information on properties is scarce**

⚠ **huge** crystalline bias

Can we predict mechanical properties of crystalline materials?

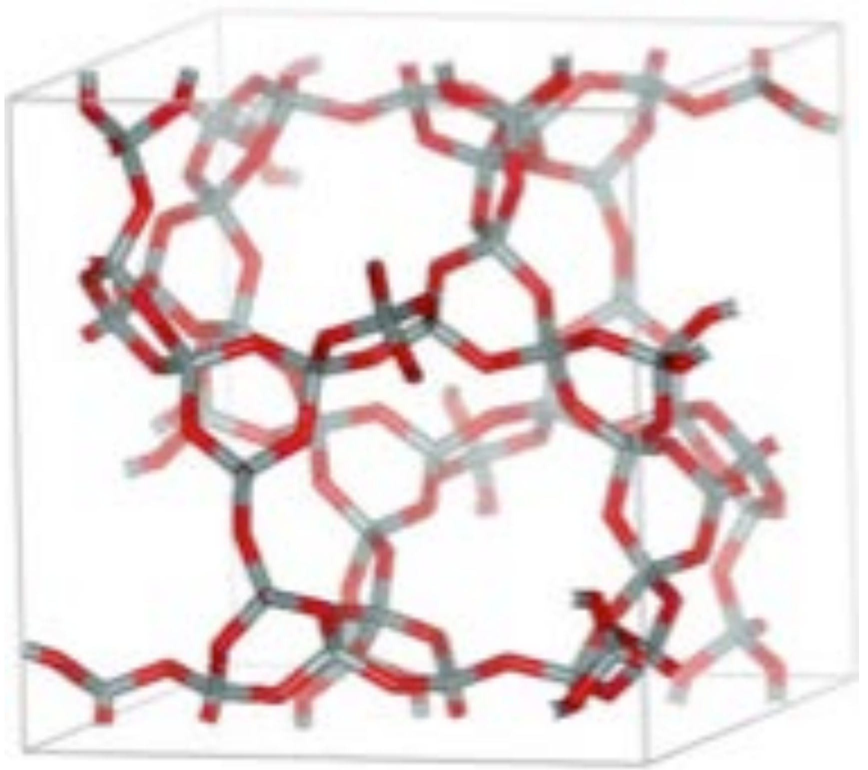
PSL UNIVERSITÉ PARIS

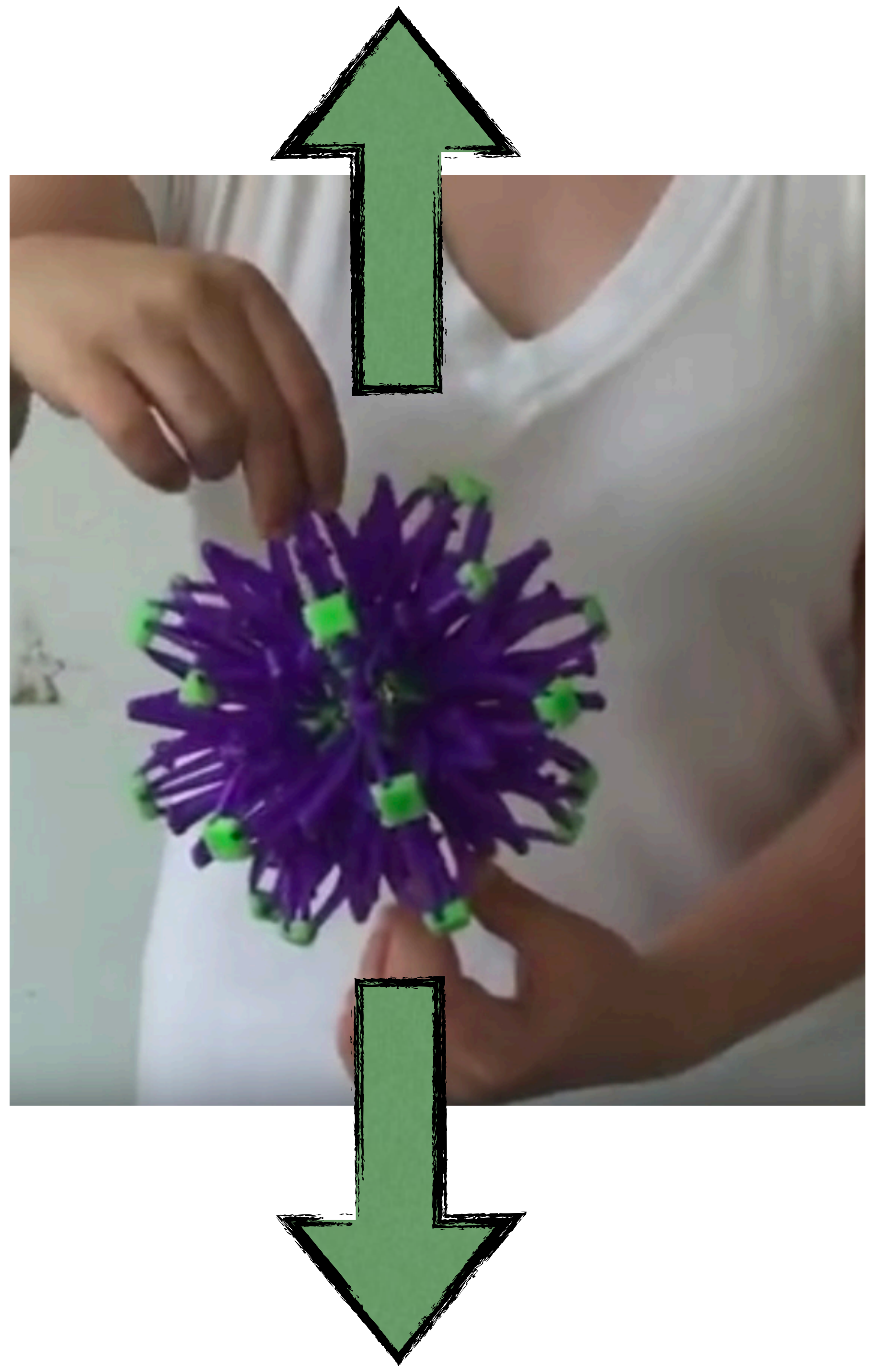
cnrs

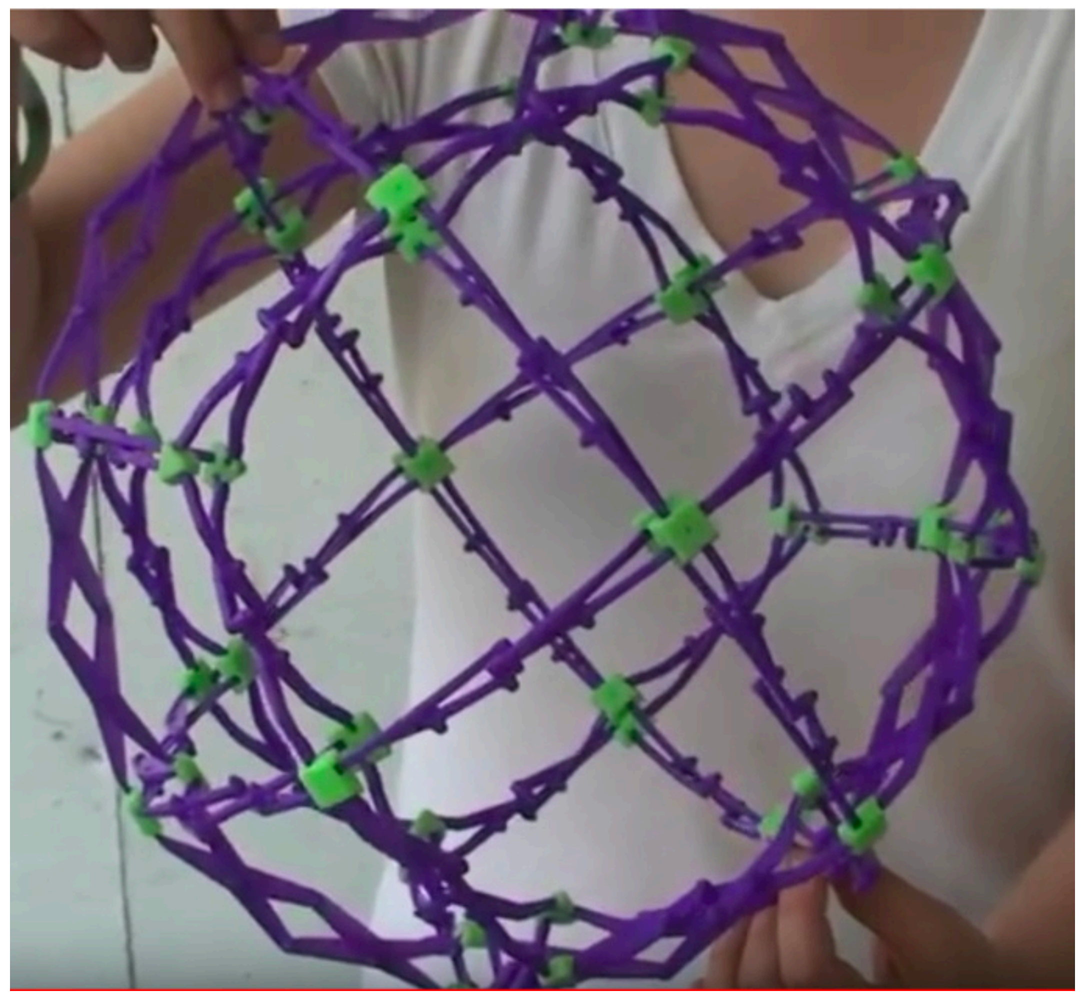
Chimie ParisTech

Mechanical properties of crystals

- In 2016, we identified by chance a zeolite with **isotropic auxeticity**







- Only **5 known crystals** with this property!
- Also considered “rare”: **negative linear compressibility**

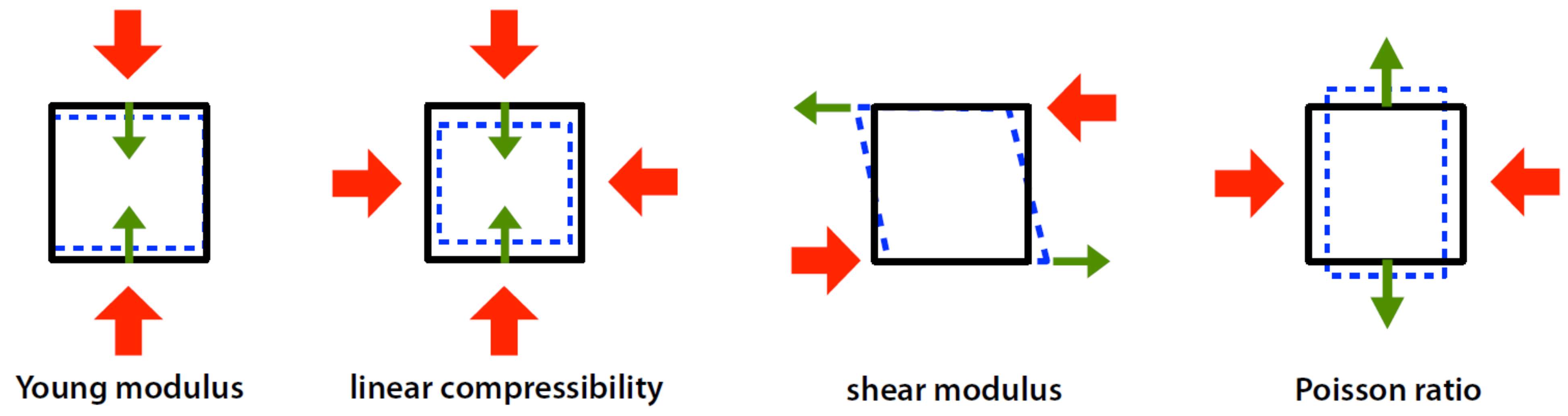
- How rare are other so-called “rare” mechanical properties?

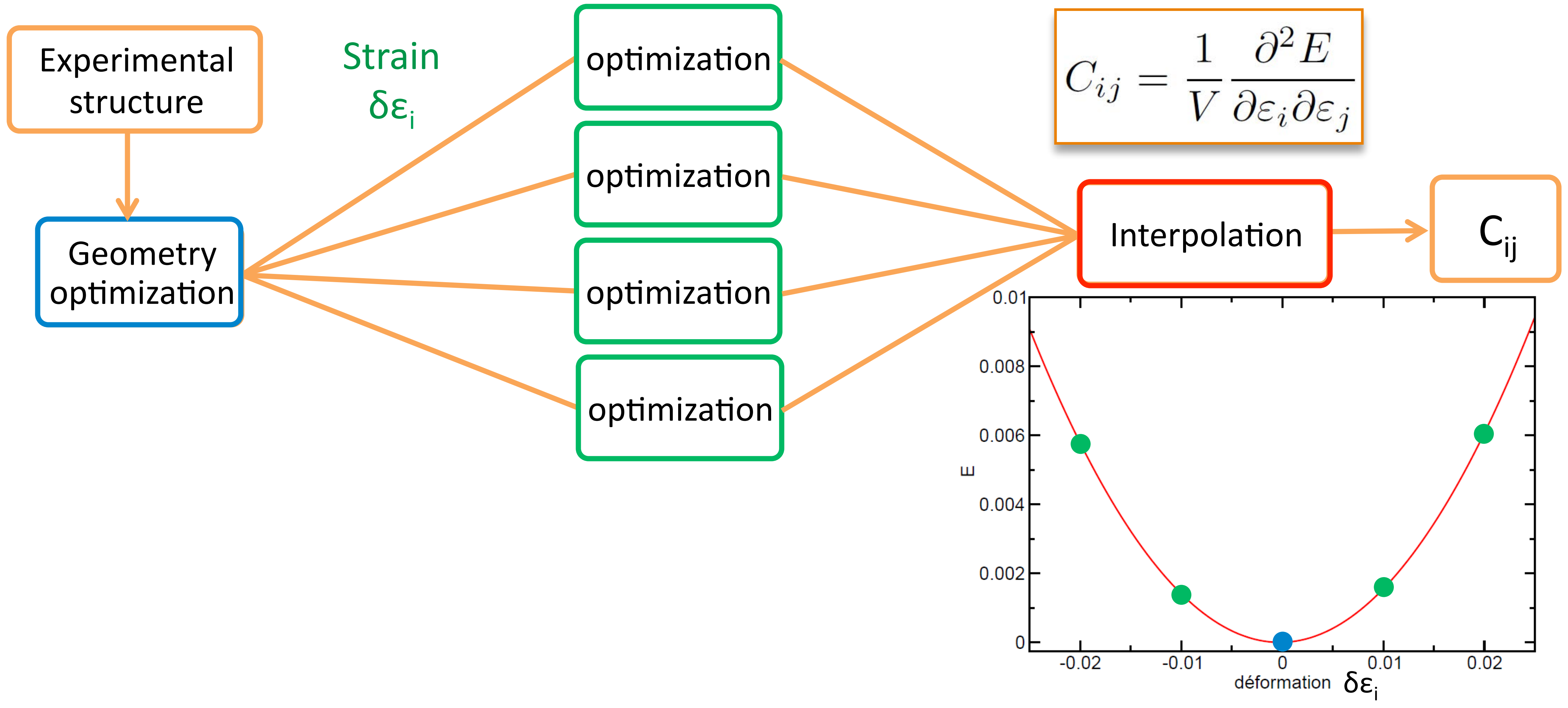
Mechanical properties of crystals

$$\boxed{\sigma_{ij} = C_{ijkl}\,\epsilon_{kl}}$$

$$C_{ijkl} = \begin{pmatrix} C_{11} & C_{12} & C_{13} & & & \\ \cdot & C_{22} & C_{23} & & & \\ \cdot & \cdot & C_{33} & & & \\ & & & C_{44} & & \\ & & & & C_{55} & \\ & & & & & C_{66} \end{pmatrix}$$

Elastic tensor

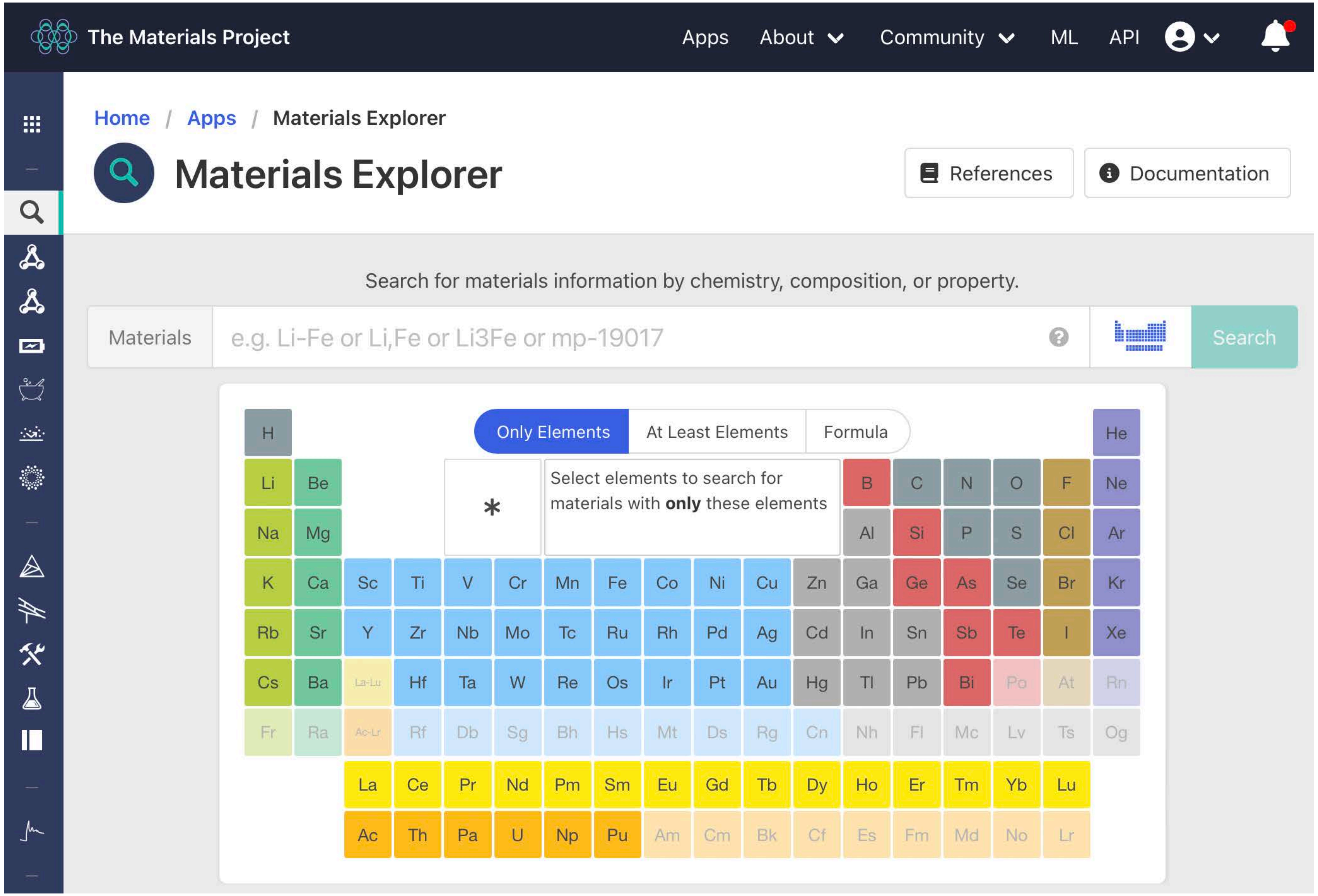


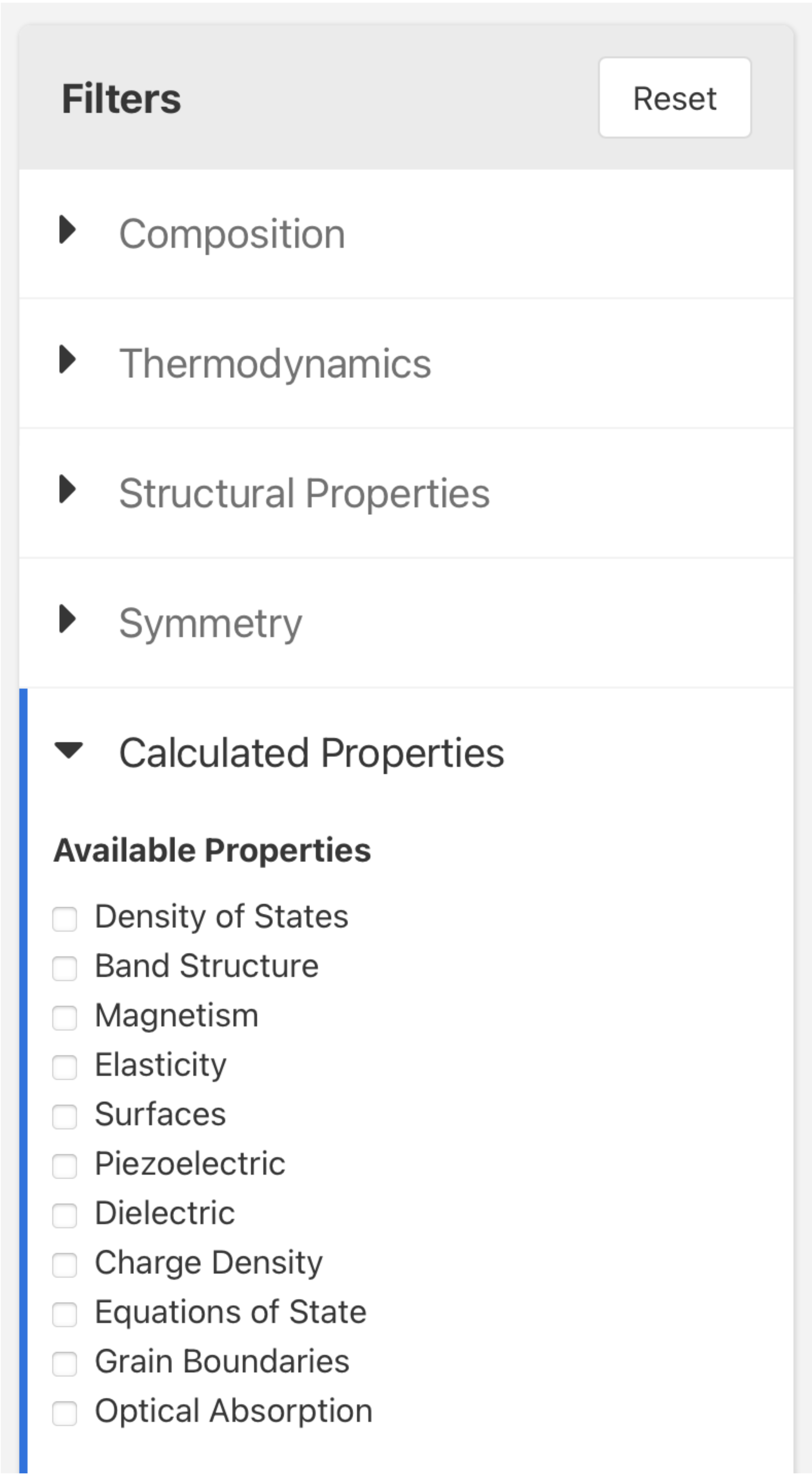


$$\boxed{C_{ij} = \frac{1}{V}\frac{\partial^2 E}{\partial \varepsilon_i \partial \varepsilon_j}}$$

- Elasticity is an anisotropic property
- Experimentally difficult to determine
- “Relatively easy” to compute from DFT

- Most people only care about the bulk modulus, but there is a lot more information!

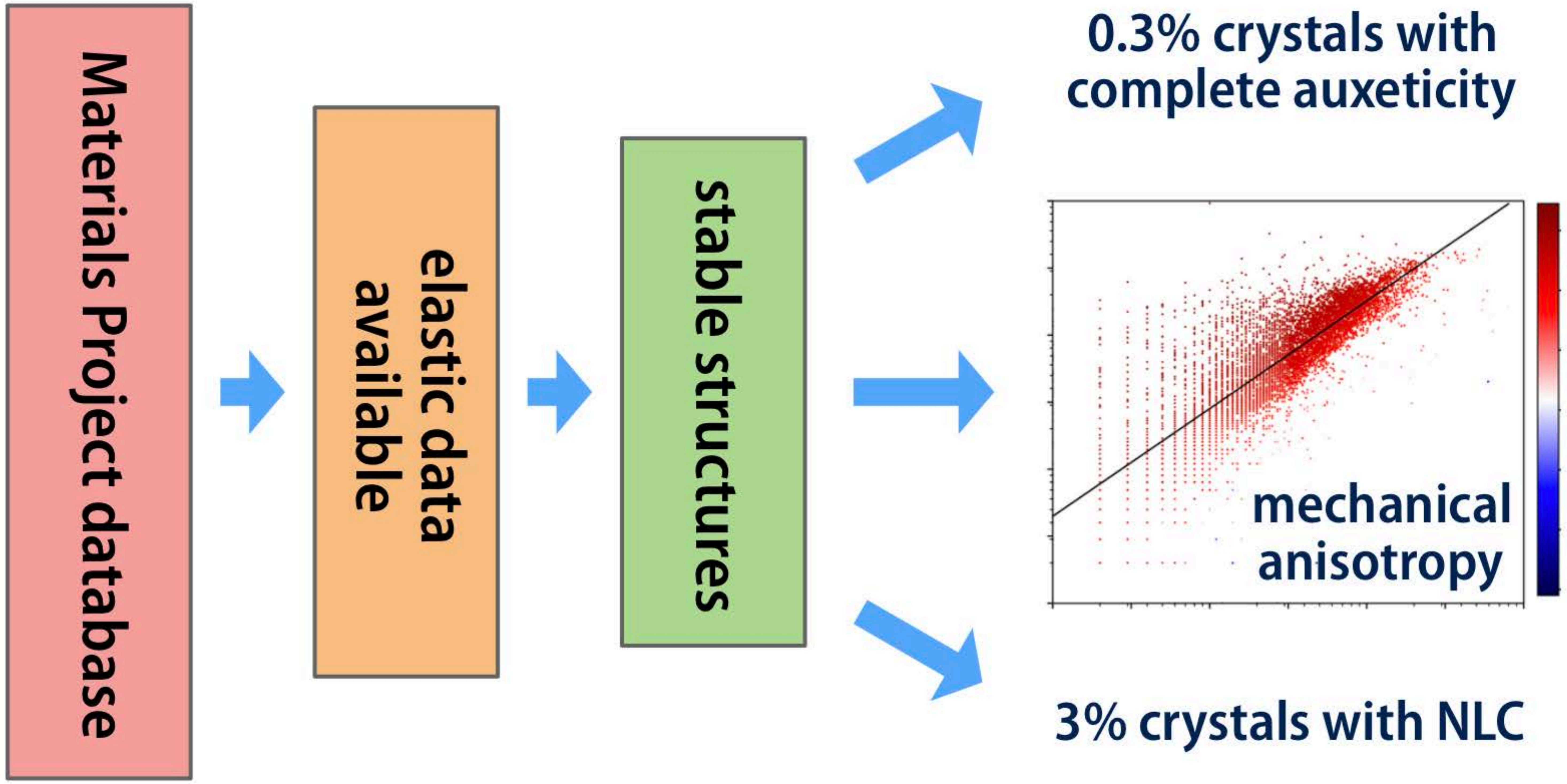

The Materials Project
Apps
About
Community
ML
API
Home / Apps / Materials Explorer
Materials Explorer
References
Documentation
Search for materials information by chemistry, composition, or property.
Materials
e.g. Li-Fe or Li,Fe or Li3Fe or mp-19017
Search
Only Elements
At Least Elements
Formula
*
Select elements to search for materials with **only** these elements
H He
Li Be B C N O F Ne
Na Mg Al Si P S Cl Ar
K Ca Sc Ti V Cr Mn Fe Co Ni Cu Zn Ga Ge As Se Br Kr
Rb Sr Y Zr Nb Mo Tc Ru Rh Pd Ag Cd In Sn Sb Te I Xe
Cs Ba La-Lu Hf Ta W Re Os Ir Pt Au Hg Tl Pb Bi Po At Rn
Fr Ra Ac-Lr Rf Db Sg Bh Hs Mt Ds Rg Cn Nh Fl Mc Lv Ts Og
La Ce Pr Nd Pm Sm Eu Gd Tb Dy Ho Er Tm Yb Lu
Ac Th Pa U Np Pu Am Cm Bk Cf Es Fm Md No Lr


Filters
Reset
Composition
Thermodynamics
Structural Properties
Symmetry
Calculated Properties
Available Properties
Density of States
Band Structure
Magnetism
Elasticity
Surfaces
Piezoelectric
Dielectric
Charge Density
Equations of State
Grain Boundaries
Optical Absorption

Quantifying anomalous behavior

- Materials Project: 133,691 inorganic compounds
- Elastic data at DFT level for 13,621 structures

- Systematic tensorial analysis to answer this simple question:
  mechanical metamaterials are rare, but how rare exactly?



Chibani & Coudert, *Chem. Sci.* **2019**

Quantifying anomalous behavior

Table 1 List of completely auxetic materials in the Materials Project database, with extremal values of directional Poisson's ratio, and isotropic average

Material ID	Structure	Synthesized	ν_{min}	ν_{max}	$\langle\nu\rangle$
mp-1021516	K_2Sn	No	−0.26	−0.20	−0.21
mp-9580	$TlGaSe_2$	Yes	−0.94	−0.24	−0.59
mp-982773	Na_3Tl	No	−0.50	−0.20	−0.4
mp-862769	$RbGe_3$	No	−1.25	−0.17	−0.18
mp-974789	Rb_3Sn	No	−0.75	−0.73	−0.62
mp-7621	$KTcO_4$	Yes	−0.41	−0.04	−0.2
mp-36508	$SnHgF_6$	No	−1.08	−0.10	−0.45
mp-15639	$HgRhF_6$	Yes	−0.53	−0.14	−0.4
mp-999274	$RbNaH_2$	Yes	−0.77	−0.47	−0.67
mp-697133	Cs_2CaH_4	Yes	−0.56	−0.32	−0.47
mp-27718	$CsHgBr_3$	Yes	−0.15	−0.06	−0.12
mp-865080	$NaCeAu_2$	No	−0.35	−0.29	−0.3
mp-13925	Cs_2NaYF_6	Yes	−0.85	−0.77	−0.82
mp-7961	Sr_3SnO	Yes	−0.08	−0.08	−0.09
mp-989580	Cs_2KNF_6	No	−0.18	−0.07	−0.14
mp-989523	Rb_2NaAsF_6	No	−0.31	−0.20	−0.26
mp-4051	$AlPO_4$	Yes	−0.58	−0.05	−0.28
mp-631316	Li_2GaSb	No	−0.05	−0.05	−0.05
mp-866229	Ca_2SnHg	No	−0.74	−0.65	−0.7
mp-2739	TeO_2	Yes	−0.77	−0.37	−0.54
mp-989536	Cs_2LiNF_6	No	−0.78	−0.75	−0.75
mp-867920	$K_2Rh_2O_5$	No	−0.57	−0.00	−0.27
mp-21200	$PuGa_2$	Yes	−0.45	−0.07	−0.28
mp-989590	Ca_6Sn_2NF	No	−0.58	−0.53	−0.55
mp-20457	InP	Yes	−0.86	−0.77	−0.81
mp-1025524	Zr_2TlC	Yes	−0.20	−0.02	−0.07
mp-1017566	$GePbO_3$	Yes	−0.50	−0.26	−0.38
mp-1008282	Cr_3Fe	Yes	−0.25	−0.04	−0.13

- No clear systematic…
- What do all these materials have in common?
- Can such complex relationships be captured by chemical descriptors? topological descriptors?
- A good case study for deep learning?

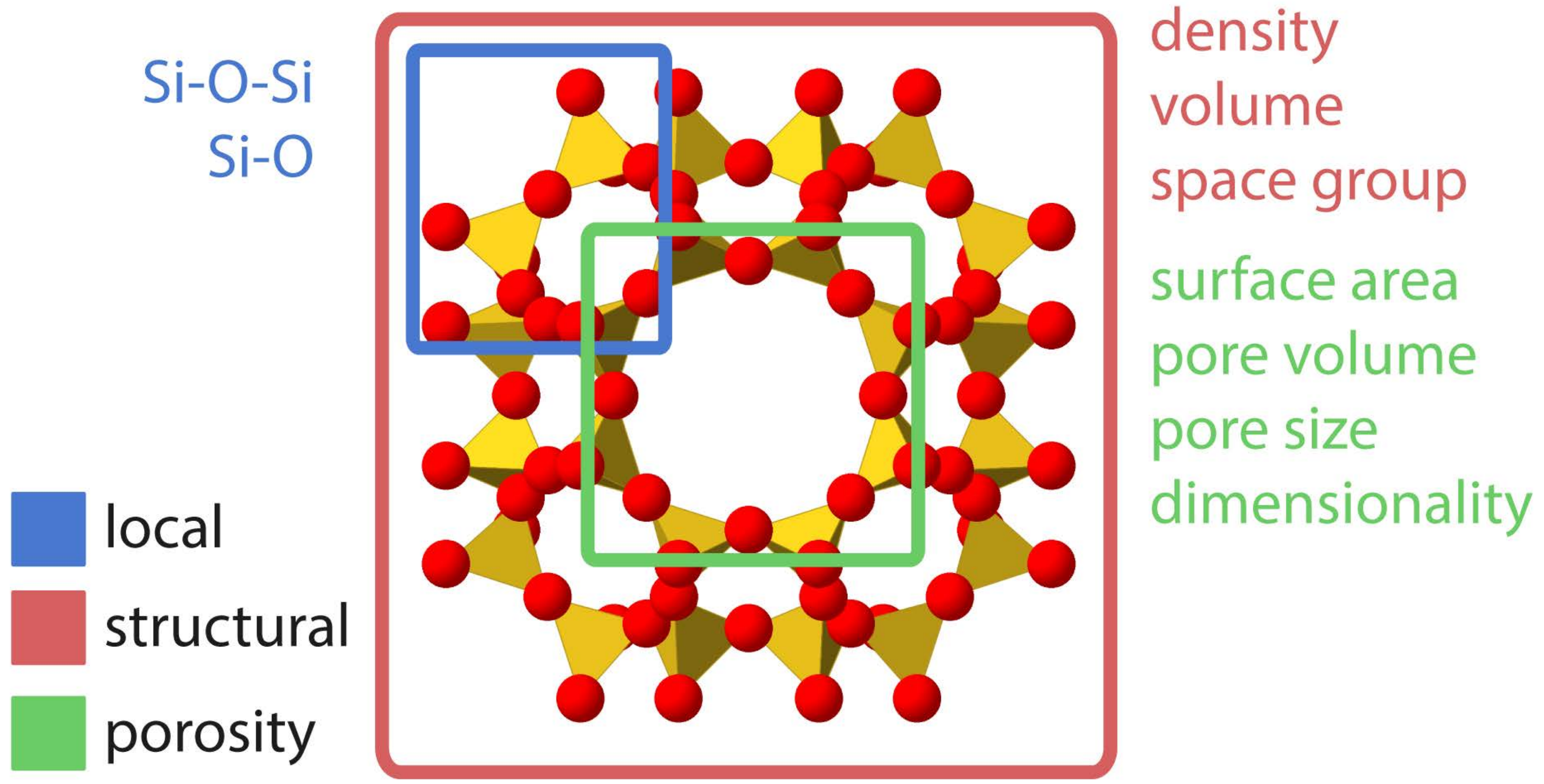
Chibani & Coudert, *Chem. Sci.* **2019**

Predicting mechanical properties

We have a smaller data set (SiO_2 zeolites) that is chemically homogeneous

Different kinds of descriptors are available, with different information:

- Hand-picked geometrical descriptors, relying on our know
- Unbiased/agnostic local geometrical descriptors (e.g. Smooth Overlap of Atomic Positions + PCA)
- Porous network characteristics (Zeo++)
- Topological information?



- Geometrical descriptors are best
- SOAP + PCA performs generally as well as "smart" descriptors

Evans & Coudert, *Chem. Mater.* **2017**

Hunting for anisotropic zeolites

- Anisotropic mechanical properties are much harder to predict
- Force fields generally perform badly
- What we are looking for is a very rare property

- Let's try a multi-step approach

Gaillac, Chibani & Coudert, *Chem. Mater.* **2020**

Hunting for anisotropic zeolites

590,811 hypothetical structures from Pophale *et al.*

BKS force field

462,248 mechanically stable

predicted auxetic: 578

random subset: 742

DFT

DFT

392 stable structures

599 stable structures

machine learning

trained predictor

Gaillac, Chibani & Coudert, *Chem. Mater.* **2020**

Hunting for anisotropic zeolites

- Force field predicts structures adequately, average mechanical properties "okay", but anisotropic properties are terrible
- GBR model based on geometric descriptors only, trained on DFT data, achieves much better accuracy

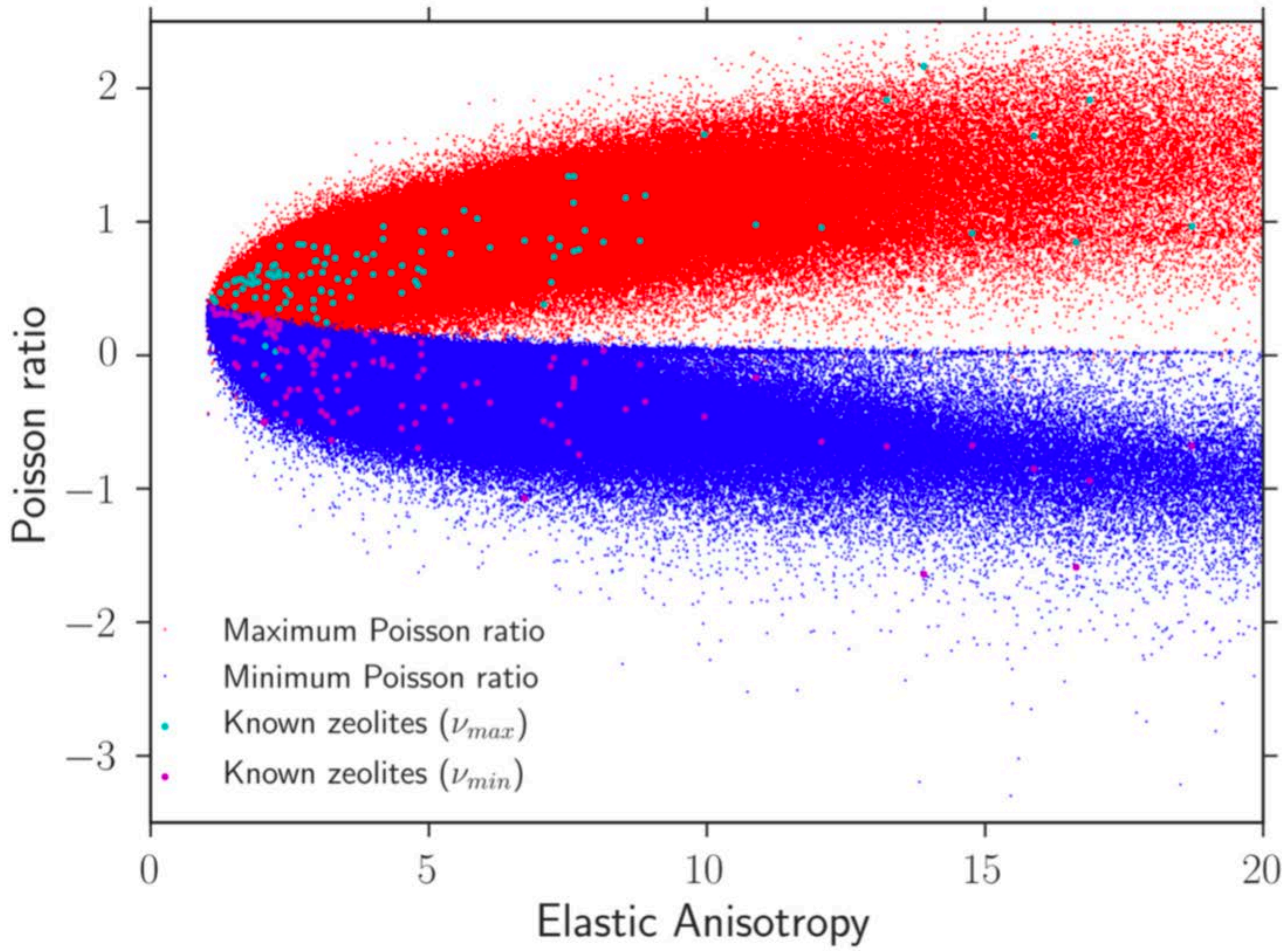


Table 4. Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) for the Three Subsets and Their Assembly for the Prediction of the Poisson's Ratio

subset (method)	RMSE (ν_{min})	MAE (ν_{min})	RMSE (ν_{max})	MAE (ν_{max})
all (GBR)	0.39	0.26	0.46	0.32
all (BKS)	1.4	0.51	9.8	2.1

- Future work: extend to zeolitic frameworks with different chemical composition ($AlPO_4$, gallogermanates, etc.) and extra-framework cations

Predicting the full tensor?

Digital Discovery

ROYAL SOCIETY OF CHEMISTRY

PAPER

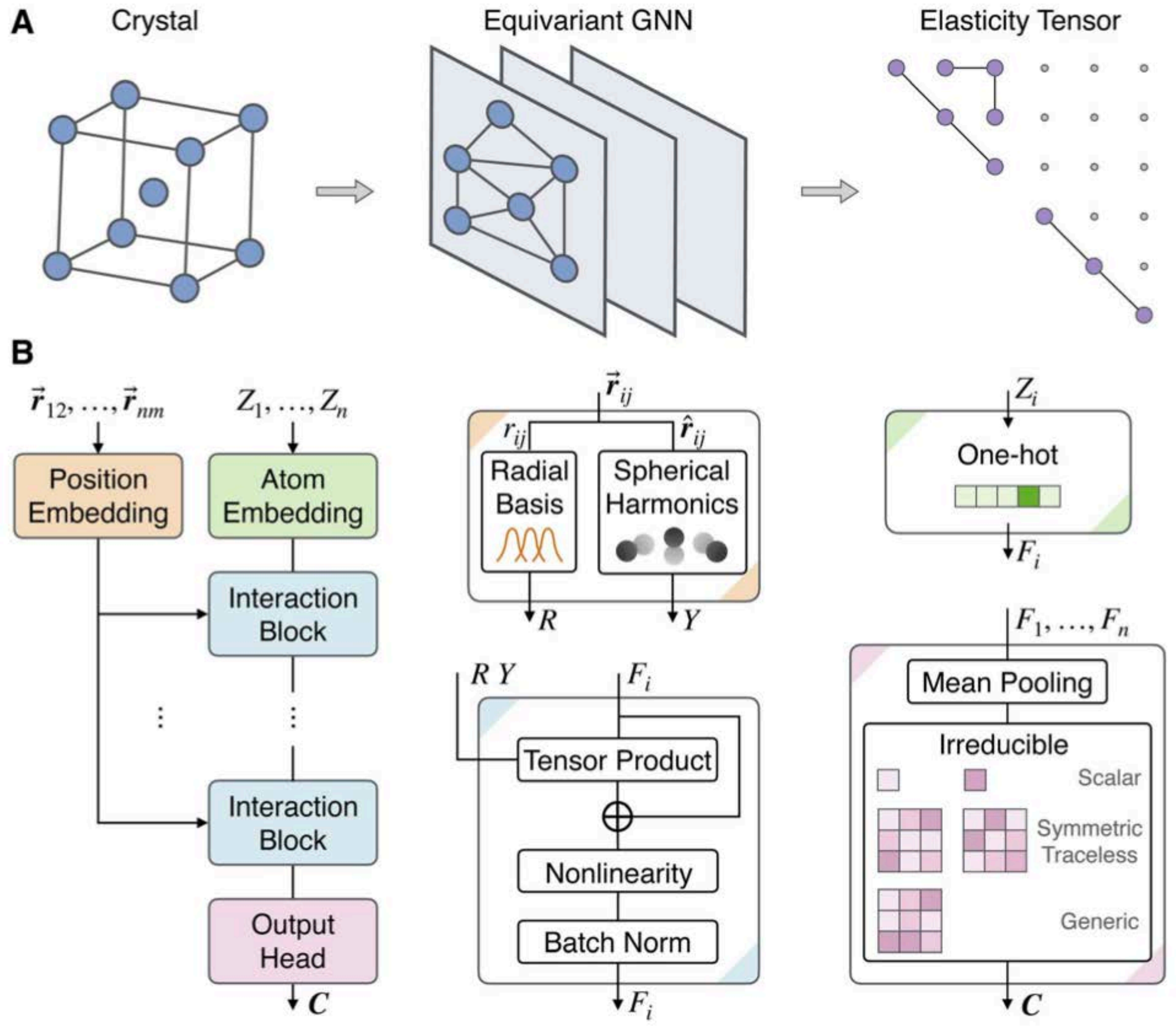
View Article Online
View Journal | View Issue

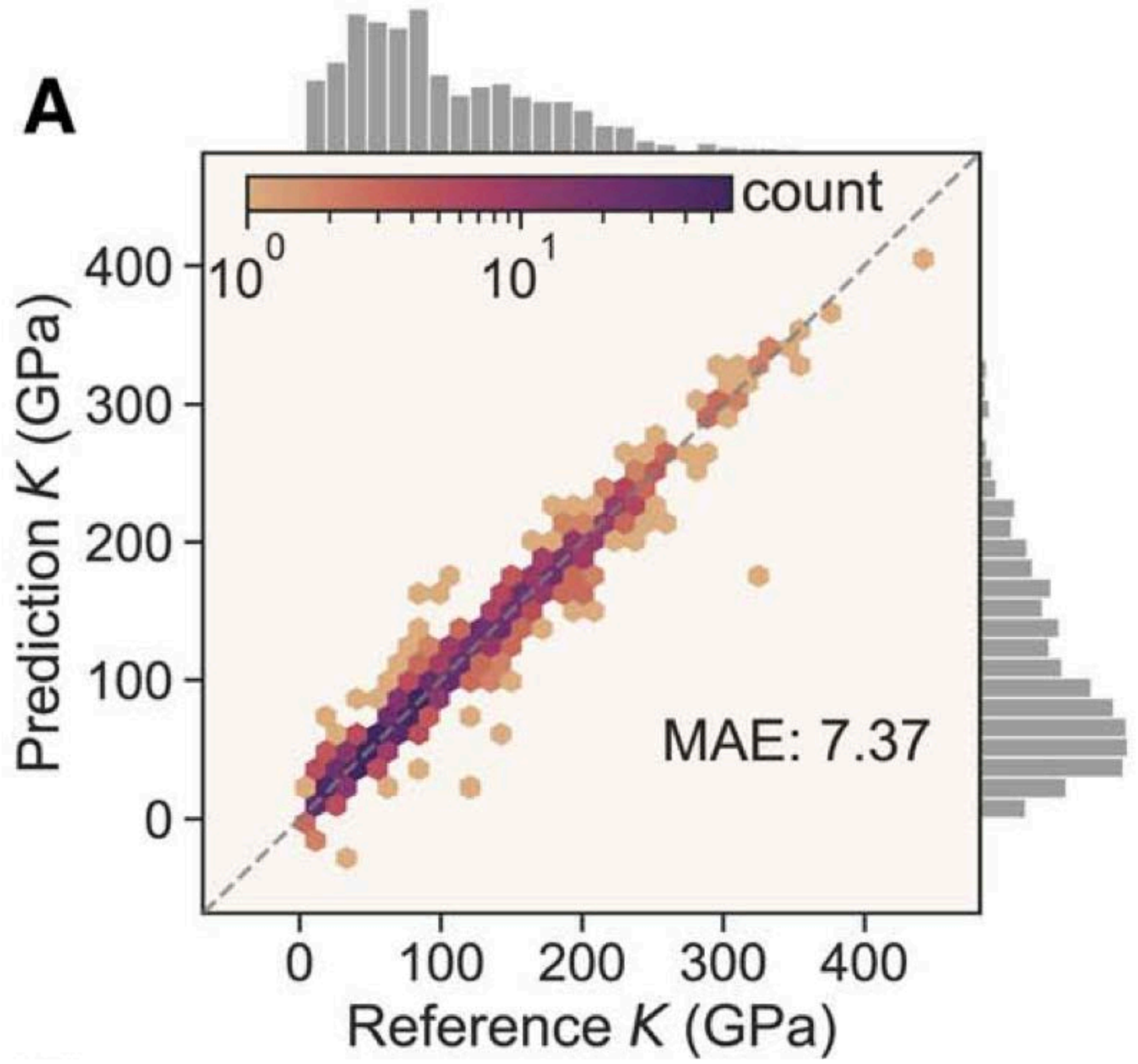
Check for updates

Cite this: *Digital Discovery*, 2024, 3, 869

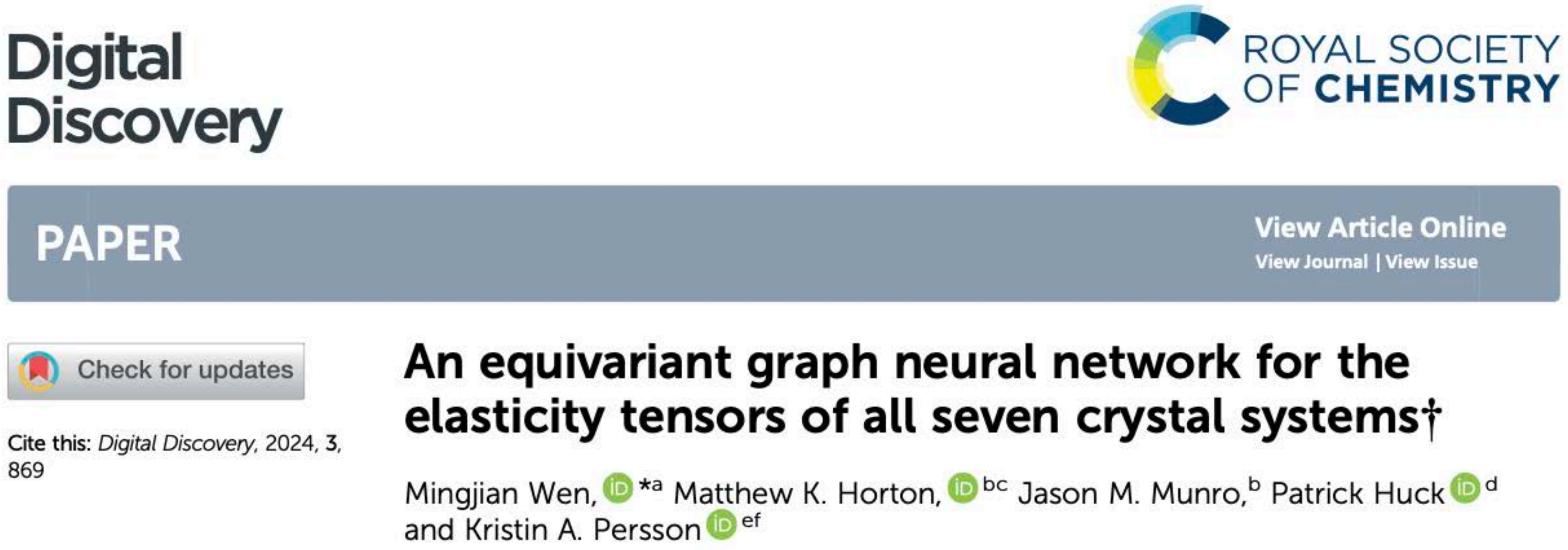
An equivariant graph neural network for the elasticity tensors of all seven crystal systems†

Mingjian Wen,[*a] Matthew K. Horton,[bc] Jason M. Munro,[b] Patrick Huck[d] and Kristin A. Persson[ef]





Predicting the full tensor?


Digital Discovery

ROYAL SOCIETY OF CHEMISTRY

PAPER

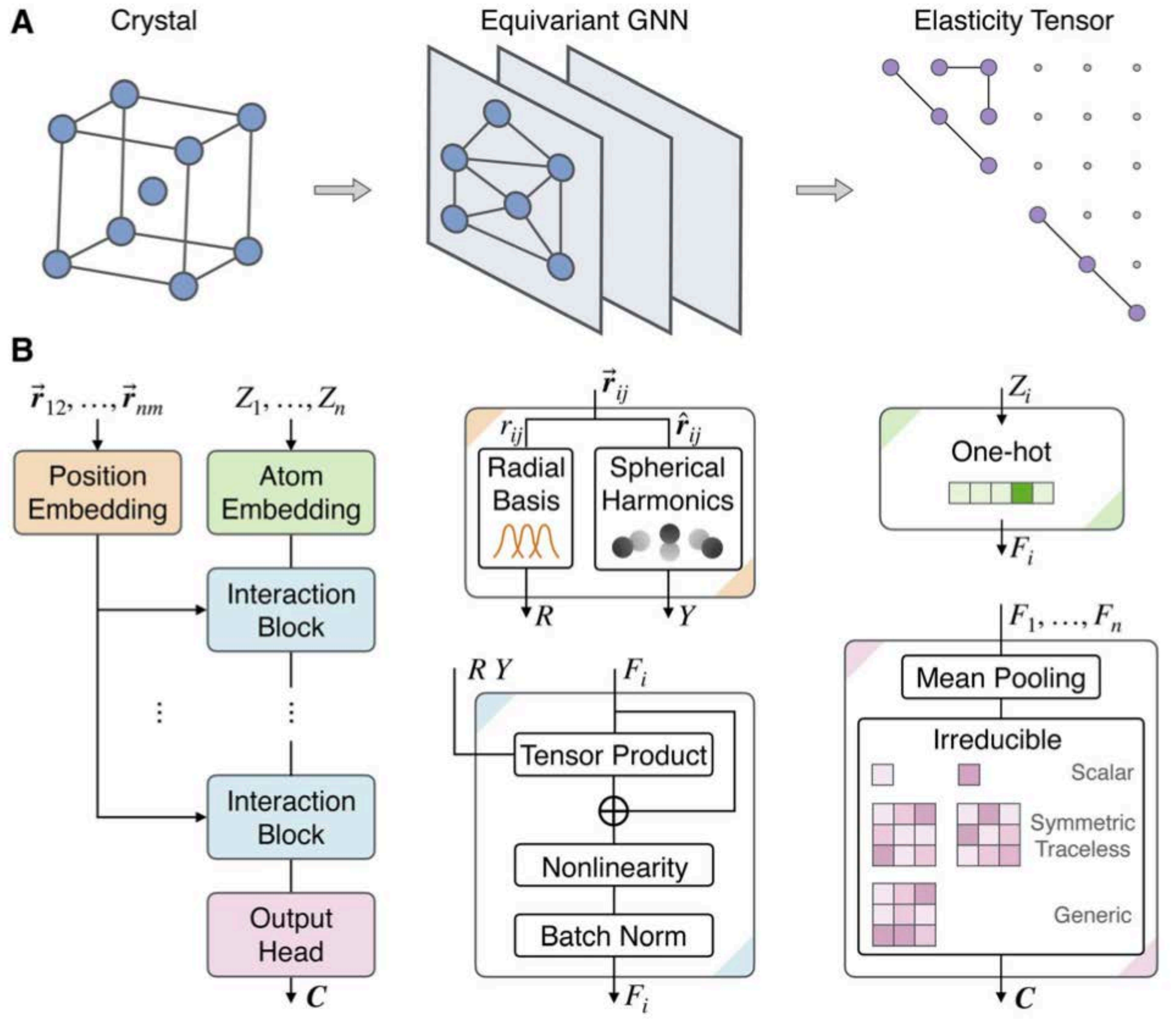
View Article Online
View Journal | View Issue

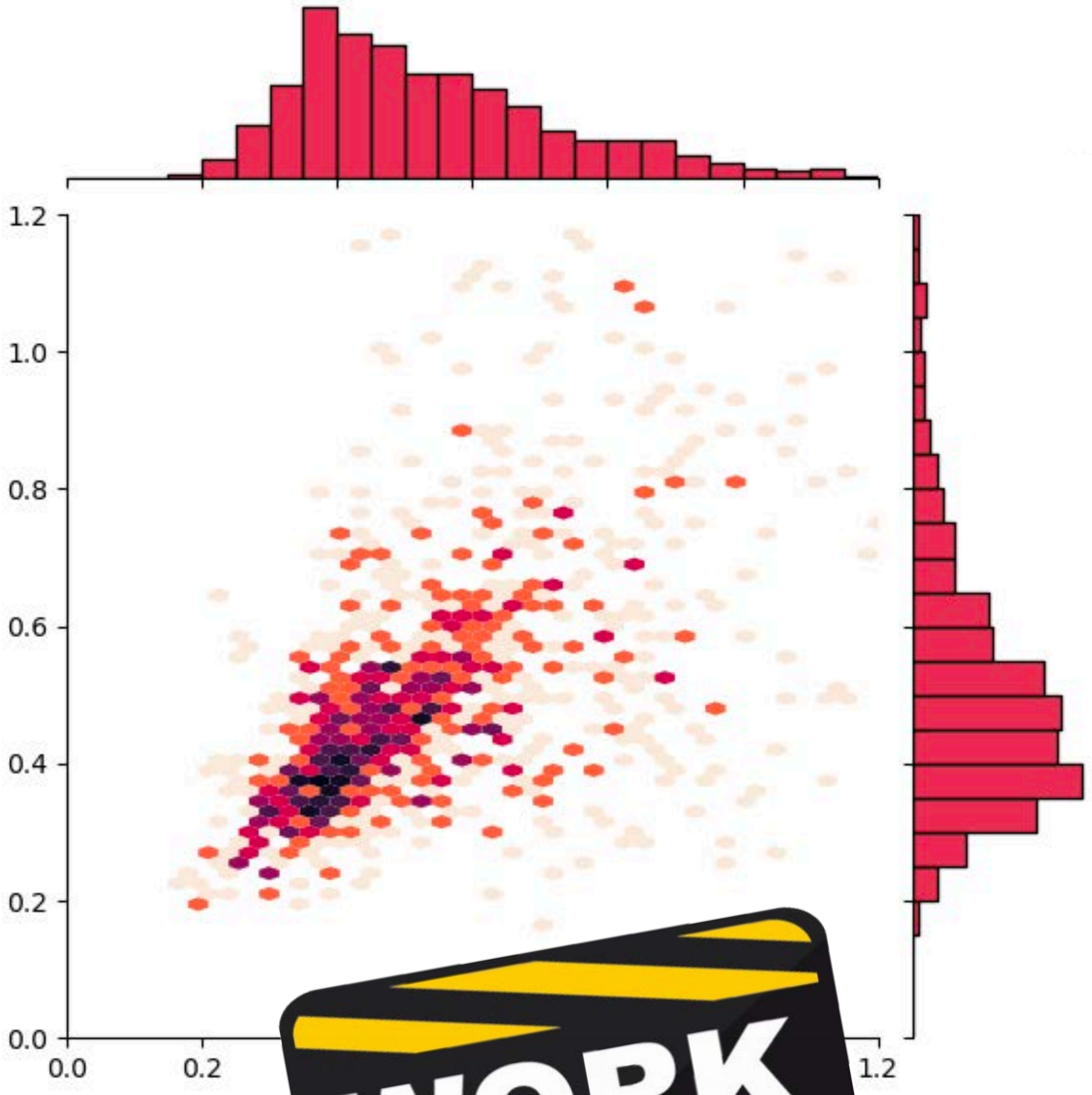
Check for updates

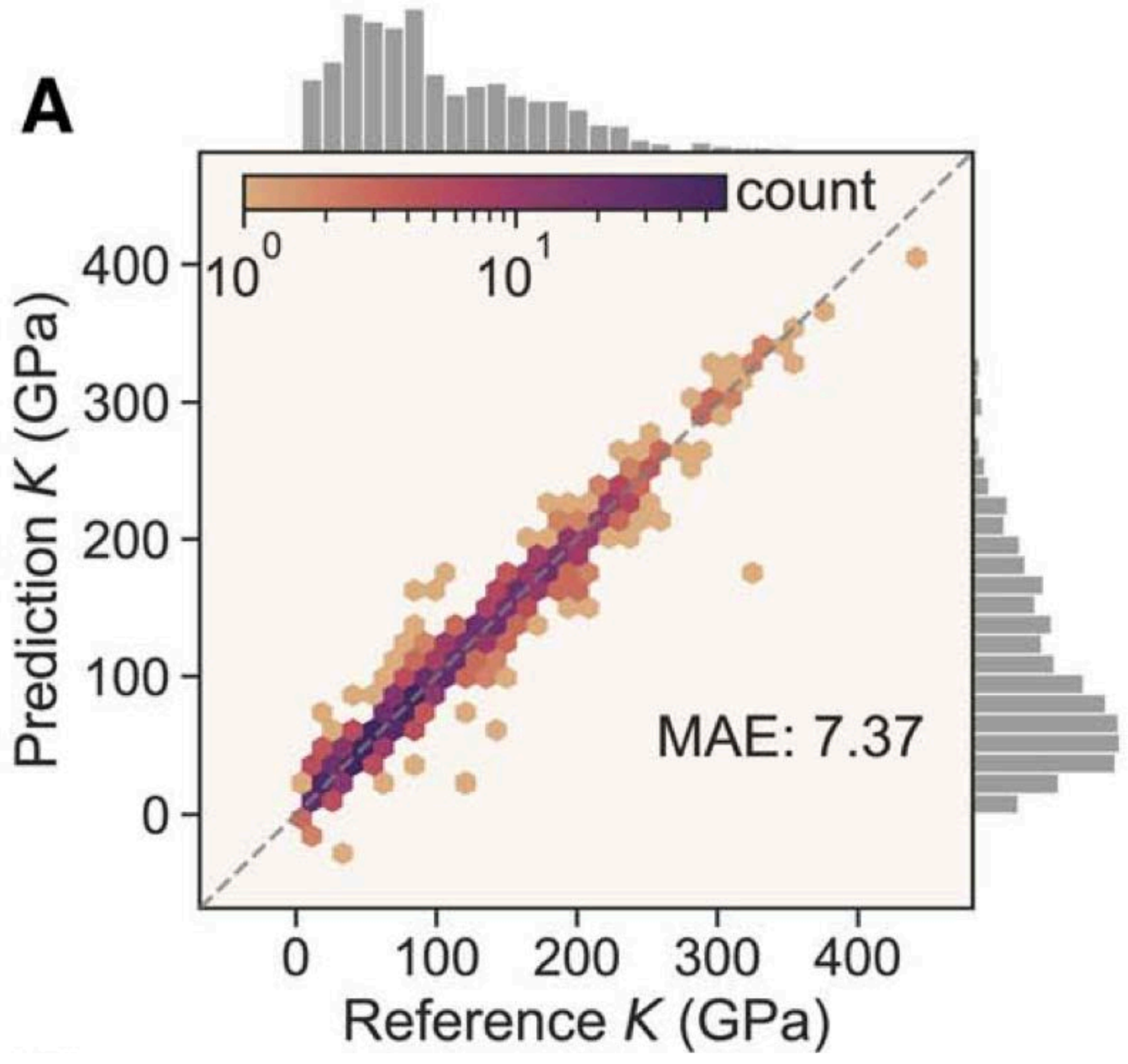
Cite this: *Digital Discovery*, 2024, 3, 869

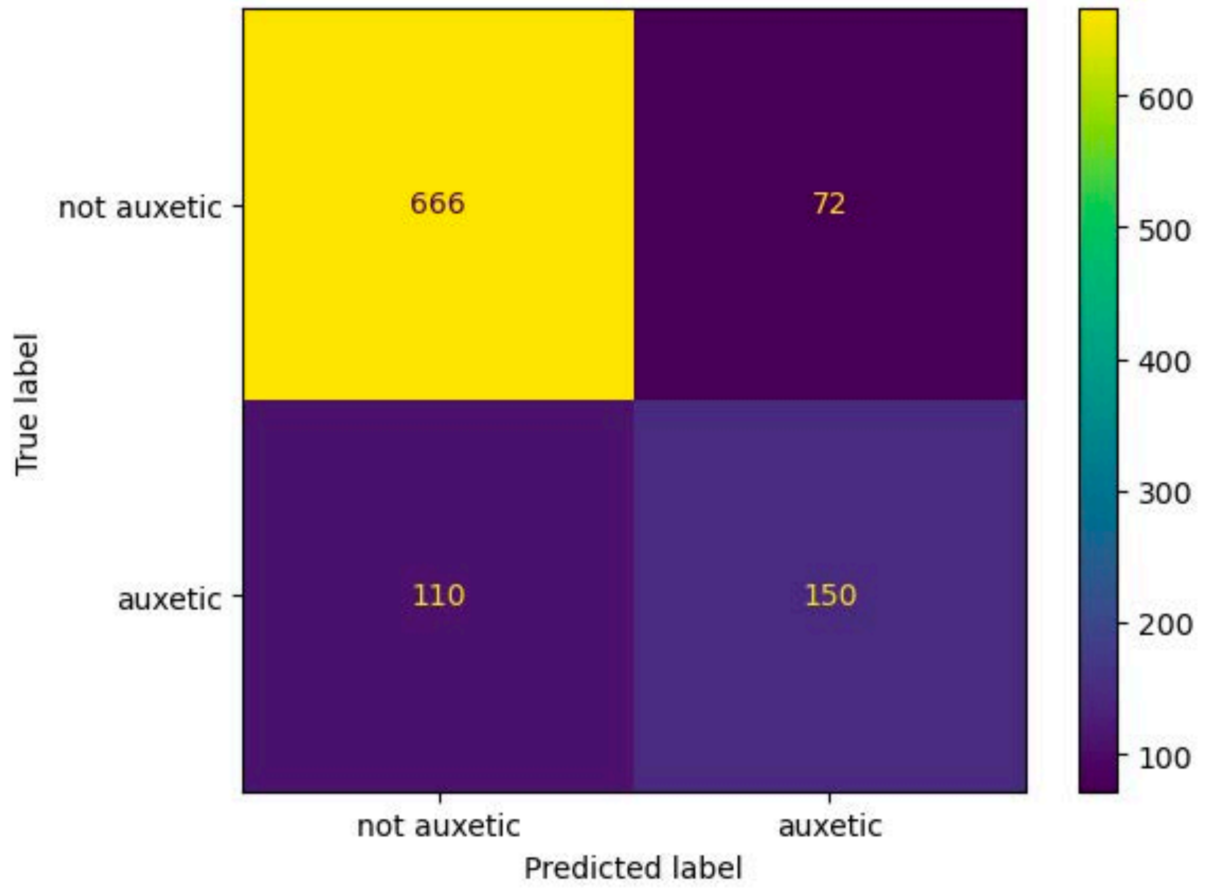
An equivariant graph neural network for the elasticity tensors of all seven crystal systems†

Mingjian Wen,[*a] Matthew K. Horton,[bc] Jason M. Munro,[b] Patrick Huck[d] and Kristin A. Persson[ef]









Can we identify top adsorbents among known nanoporous materials?

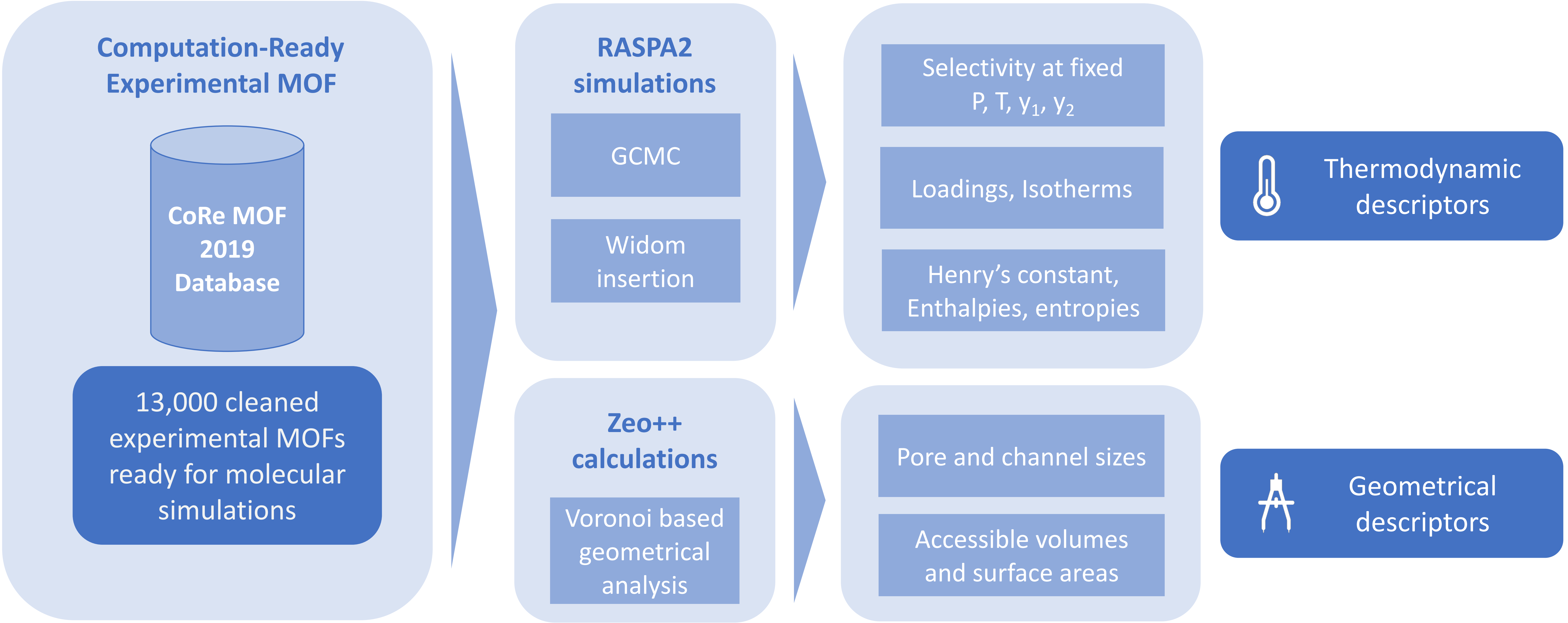
PSL UNIVERSITÉ PARIS

Chimie ParisTech

cnrs

Methodology for adsorption screening

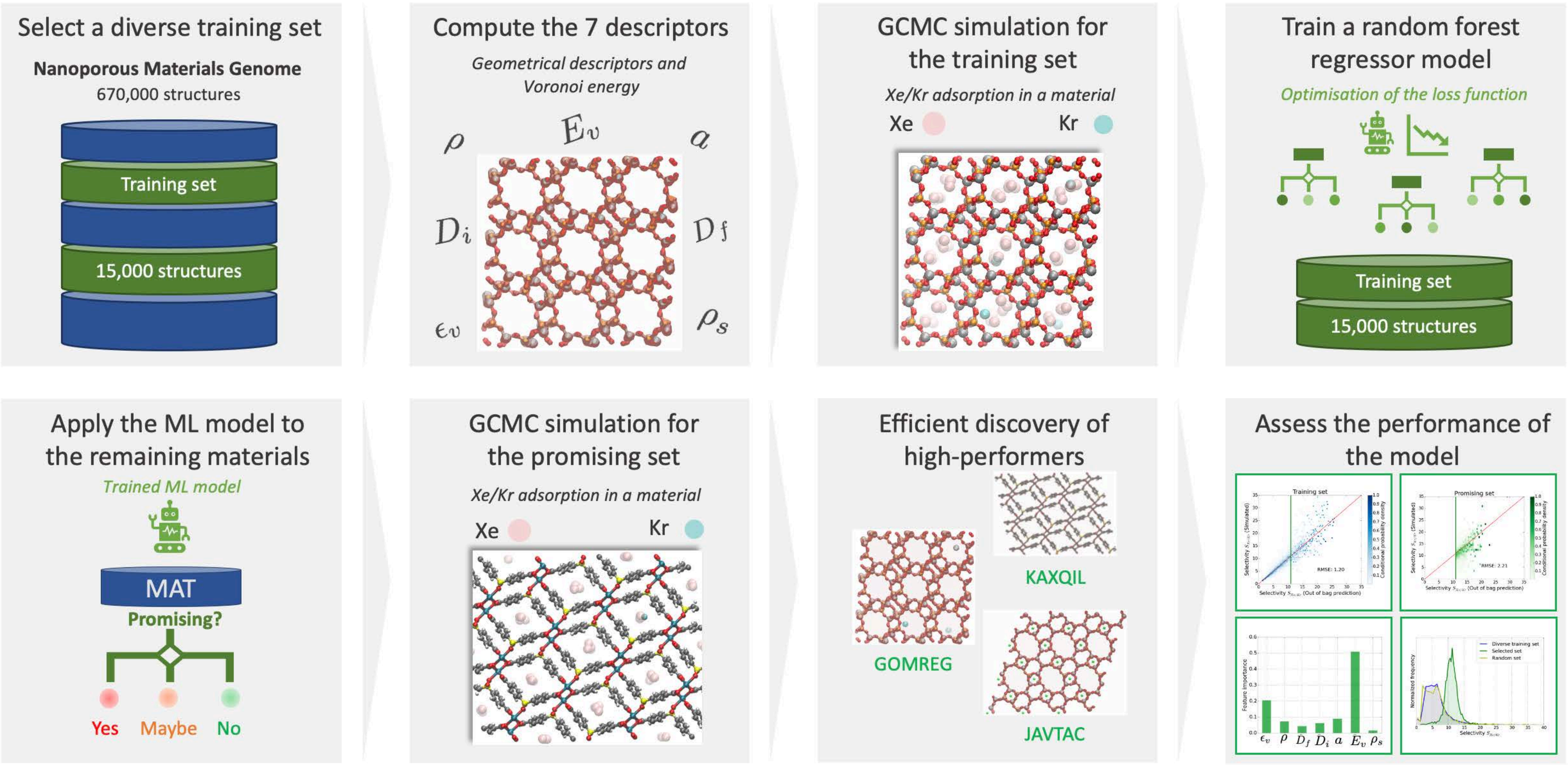
- Adsorption is cheap to evaluate, under certain restrictions *(rigid material, physisorption, polarizability not involved, …)*
- Many studies, often based on similar workflows



- What about **transport**? Diffusion is harder to compute
- Local geometric descriptors are not sufficient

Methodology for adsorption screening

- We can combine GCMC simulations and ML to accelerate discovery



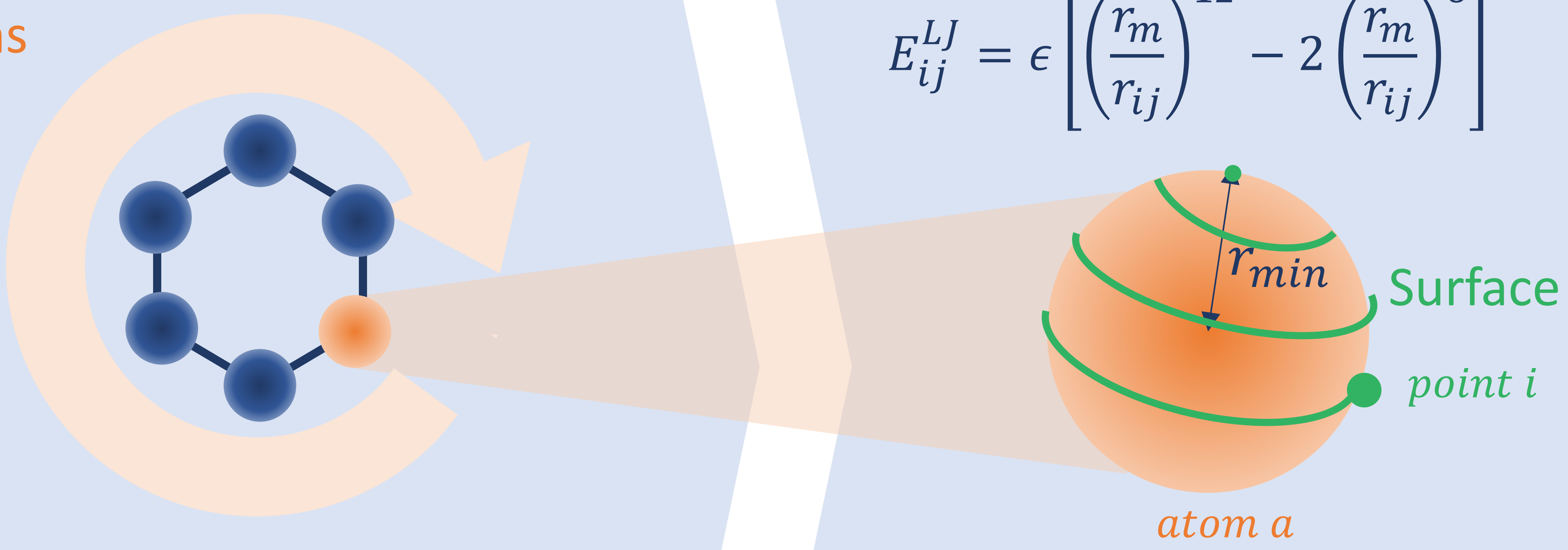
- How do we go beyond that? What are the next steps?

Methodology for adsorption screening

- ☆ Do we really need GCMC? Interactions are mostly local / short-range
- ☆ GCMC requires 3D sampling, but the internal surface is 2D
- ☆ We introduce RAESS : Rapid Adsorption Enthalpy Surface Sampling

1. Loop over the atoms

Loop over unique atoms



Material framework

Atom of the material

2. Sample points on the sphere

Lennard-Jones energy:

$$E_{ij}^{LJ} = \epsilon \left[\left(\frac{r_m}{r_{ij}} \right)^{12} - 2 \left(\frac{r_m}{r_{ij}} \right)^{6} \right]$$

r_{min} Distance to the minimum of the LJ potential

$-\epsilon$ Minimum of LJ potential

3. Adsorption energy calculation

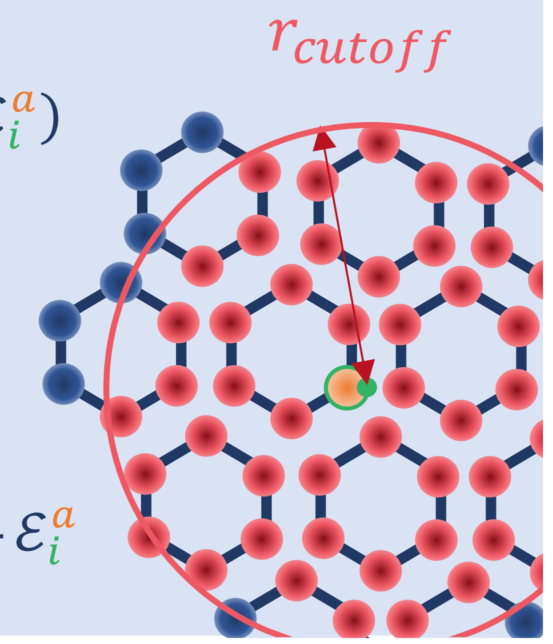
Interaction energy of an adsorbate at point 1 with the structure:

$$\mathcal{E}_i^a = E_{ia}^{LJ} + \sum_j E_{ij}^{LJ}$$

where j represents the red atoms of the structure

Boltzmann average:

$$Z = \sum_a \sum_i \exp(-\beta \mathcal{E}_i^a)$$



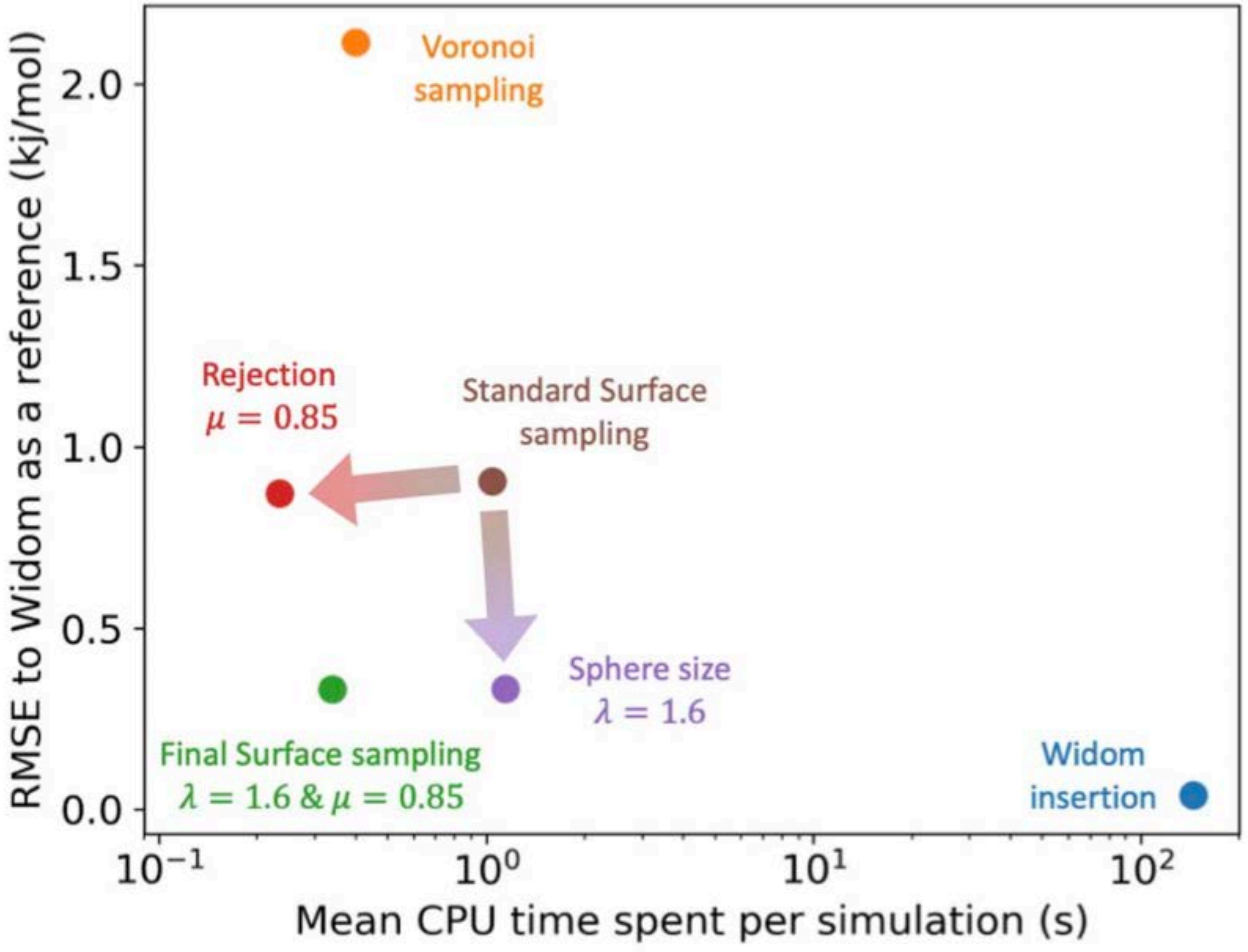
Approximation of the adsorption energy:

$$E_{ads} = \sum_a \sum_i \frac{\exp(-\beta \mathcal{E}_i^a)}{Z} \mathcal{E}_i^a$$

E. Ren and FXC, *Chem. Sci.* **2023**

Methodology for adsorption screening

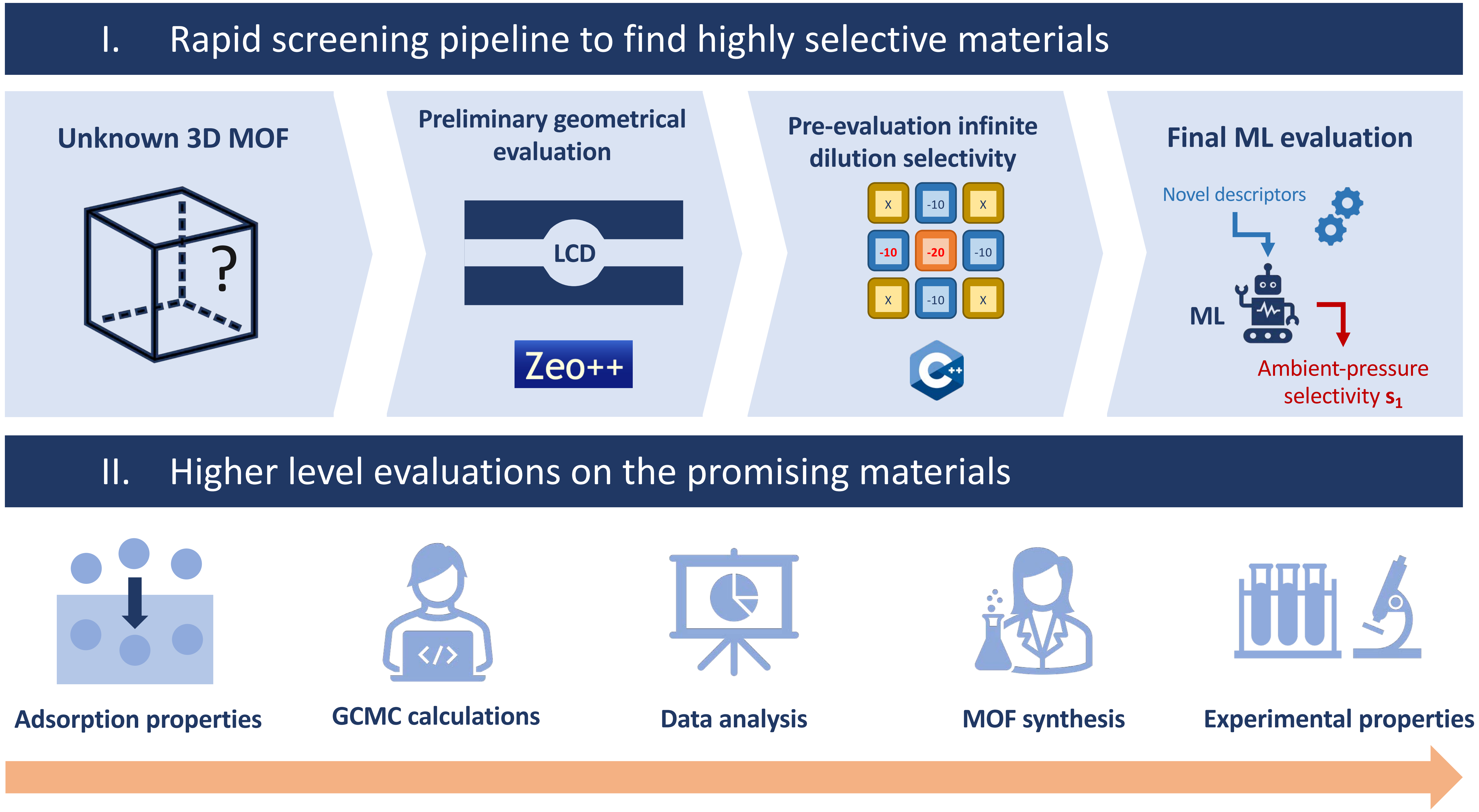
- Do we really need GCMC? Interactions are mostly local / short-range
- GCMC requires 3D sampling, but the internal surface is 2D
- We introduce RAESS : Rapid Adsorption Enthalpy Surface Sampling



E. Ren and FXC, *Chem. Sci.* **2023**

Methodology for adsorption screening

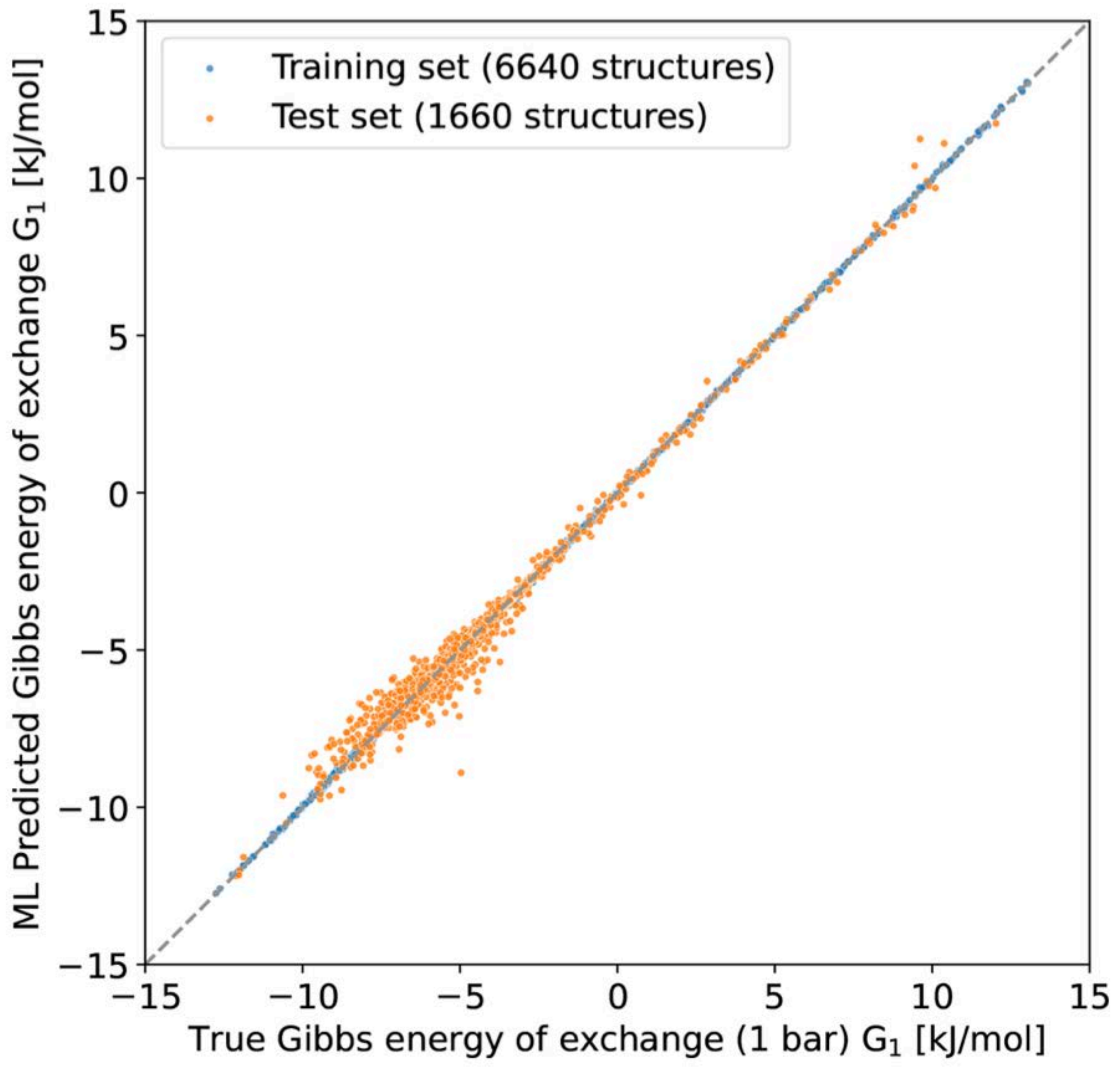
- ☆ RAESS is good, but it predicts zero-loading adsorption enthalpy
- ☆ How do we predict ambient pressure loading, or selectivity?



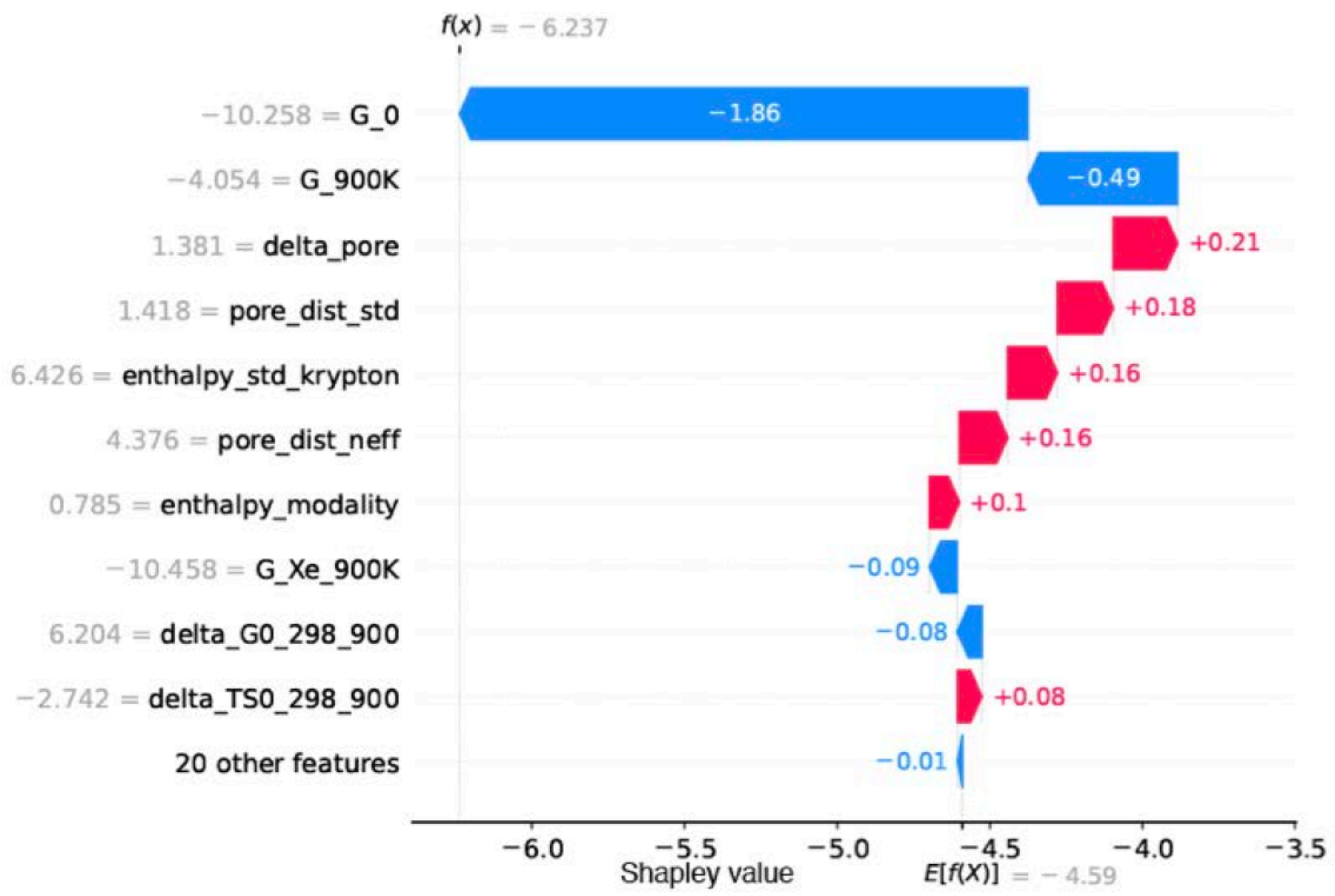
E. Ren and FXC, *Chem. Mater.* **2023**

Methodology for adsorption screening

- RAESS is good, but it predicts zero-loading adsorption enthalpy
- How do we predict ambient pressure loading, or selectivity?
- ML model takes RAESS data (and other descriptors) as input



- Very important: explanability!

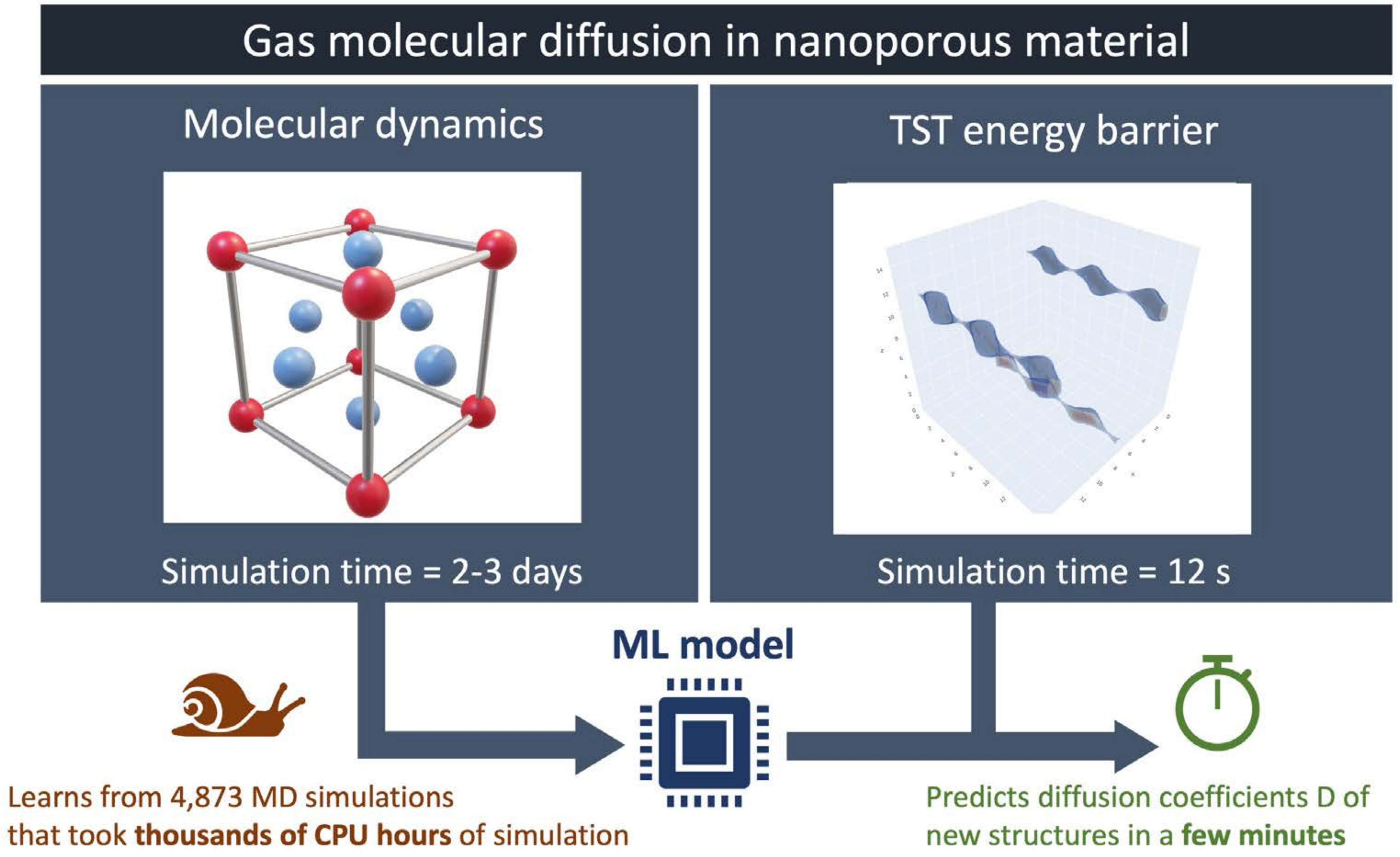


(a) VIWMIZ: true $\Delta_{\mathrm{exc}}G_1 = -6.63$

E. Ren and FXC, *Chem. Mater.* **2023**

Harder task: predicting diffusion

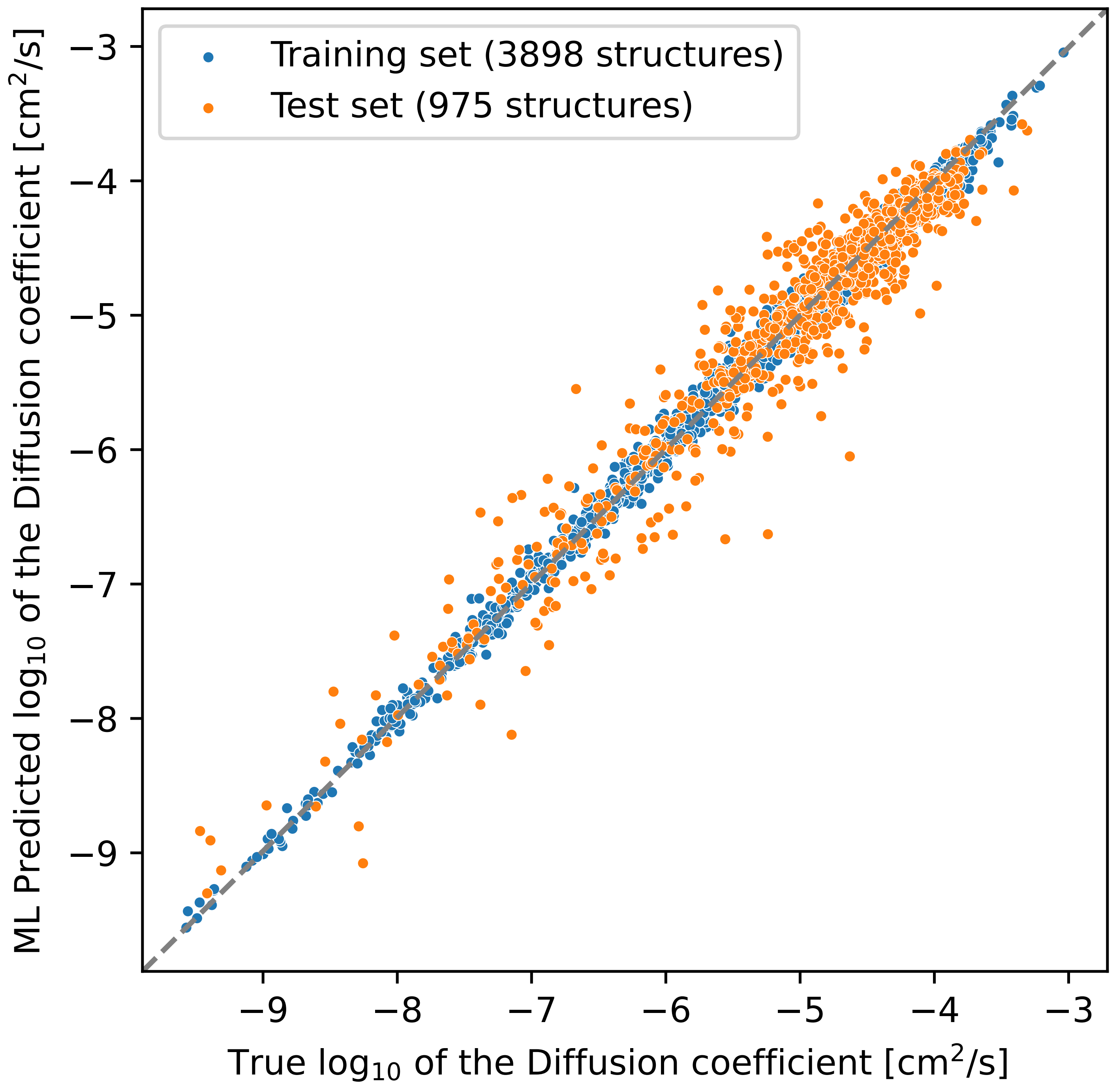
- Transport properties are also crucial for applications, but difficult to study in high-throughput setups: MD simulations are slow…
- Geometrical features are key, but not sufficient for accurate prediction of diffusion coefficients
- We propose a scheme based on:
  geometric descriptors + fast-to-compute energetic features

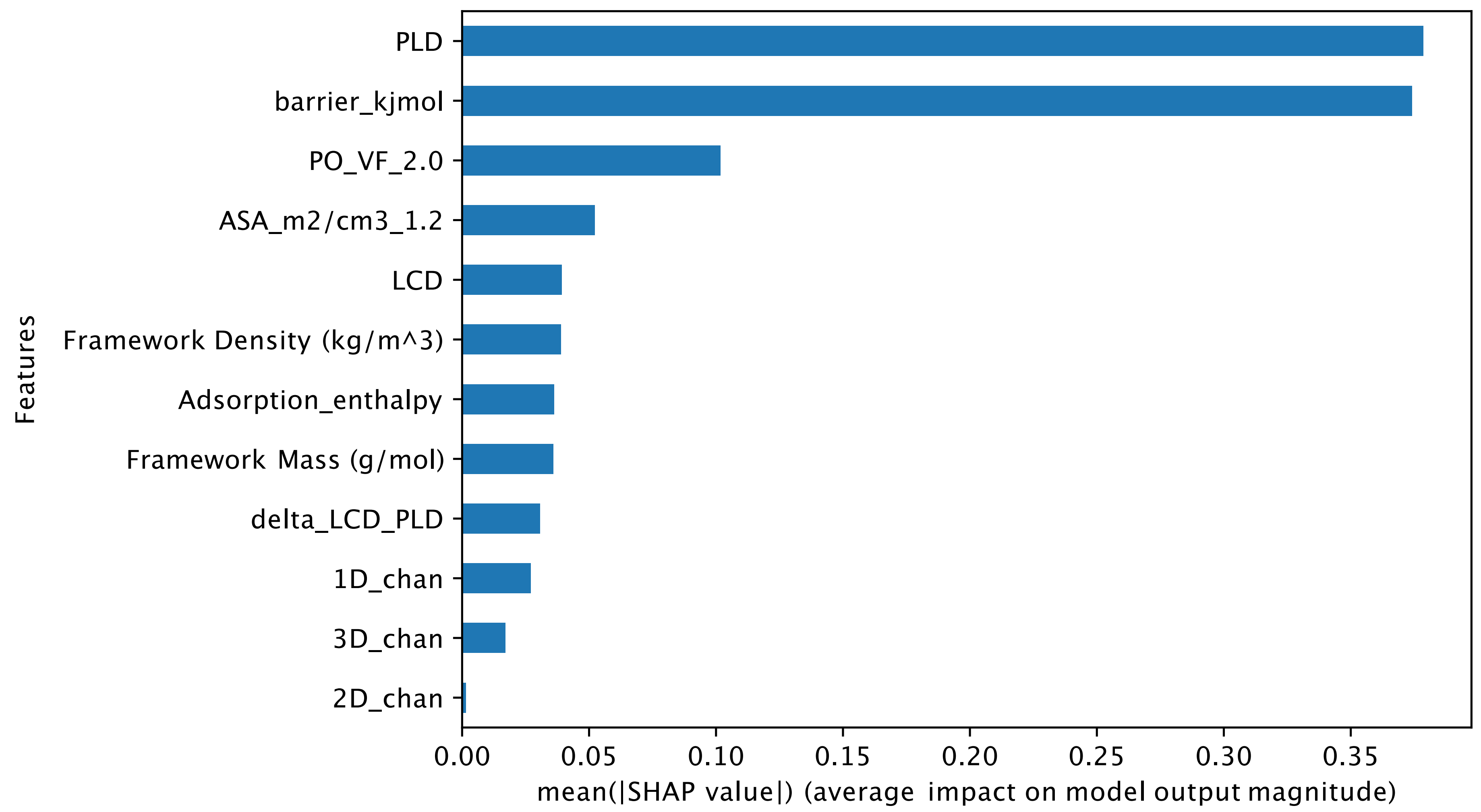


E. Ren and FXC, *J. Phys. Chem. C* **2024**

Harder task: predicting diffusion

- Transport properties are also crucial for applications, but difficult to study in high-throughput setups: MD simulations are slow…
- Geometrical features are key, but not sufficient for accurate prediction of diffusion coefficients
- We propose a scheme based on:
  geometric descriptors + fast-to-compute energetic features

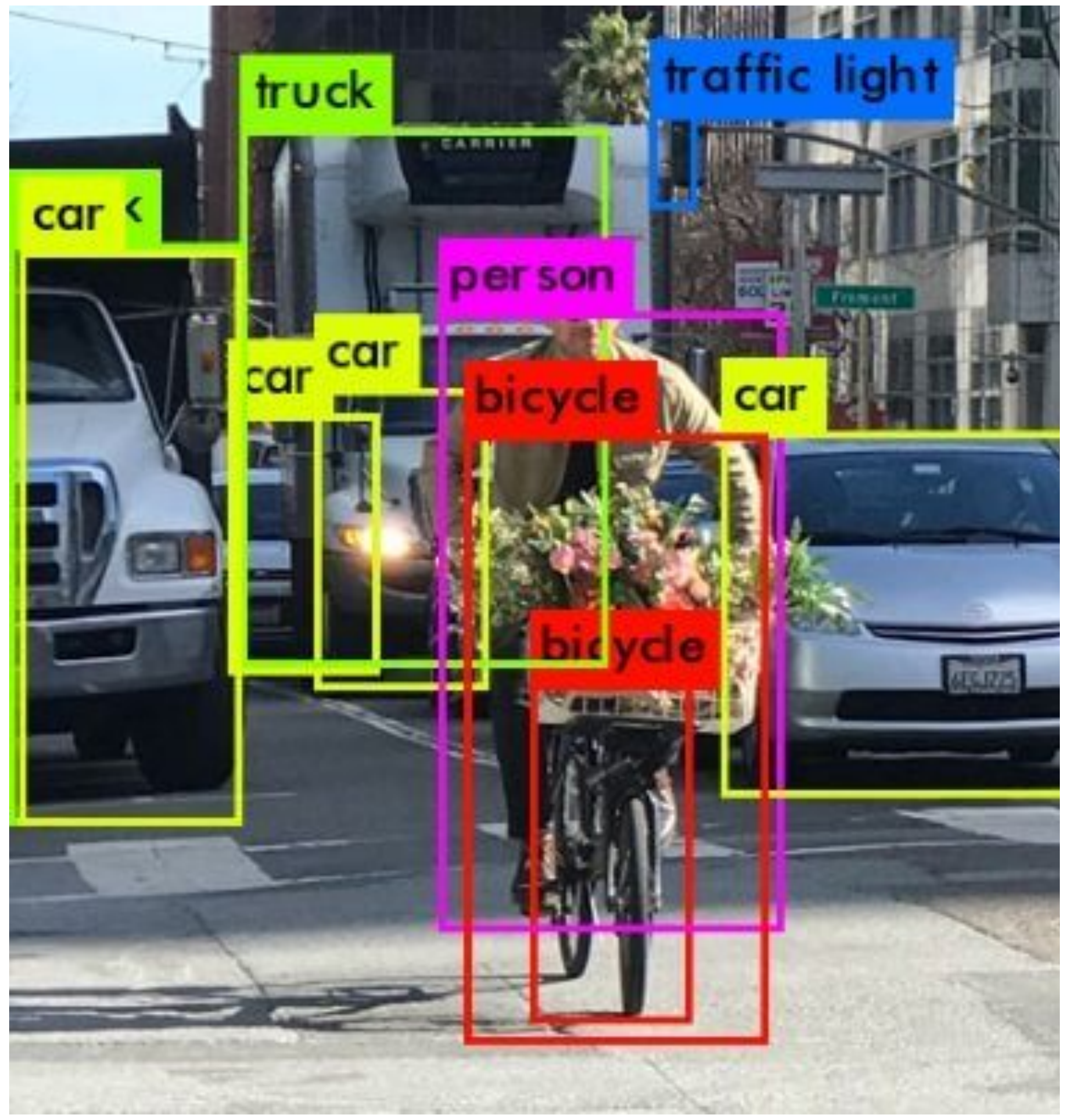




E. Ren and FXC, *J. Phys. Chem. C* **2024**

Harder task: predicting diffusion

- How do we get even better?
- Can a 3D convolutional neural network directly "read" the potential energy surface?



Article | Open access | Published: 26 January 2024

Gas adsorption meets deep learning: voxelizing the potential energy surface of metal-organic frameworks

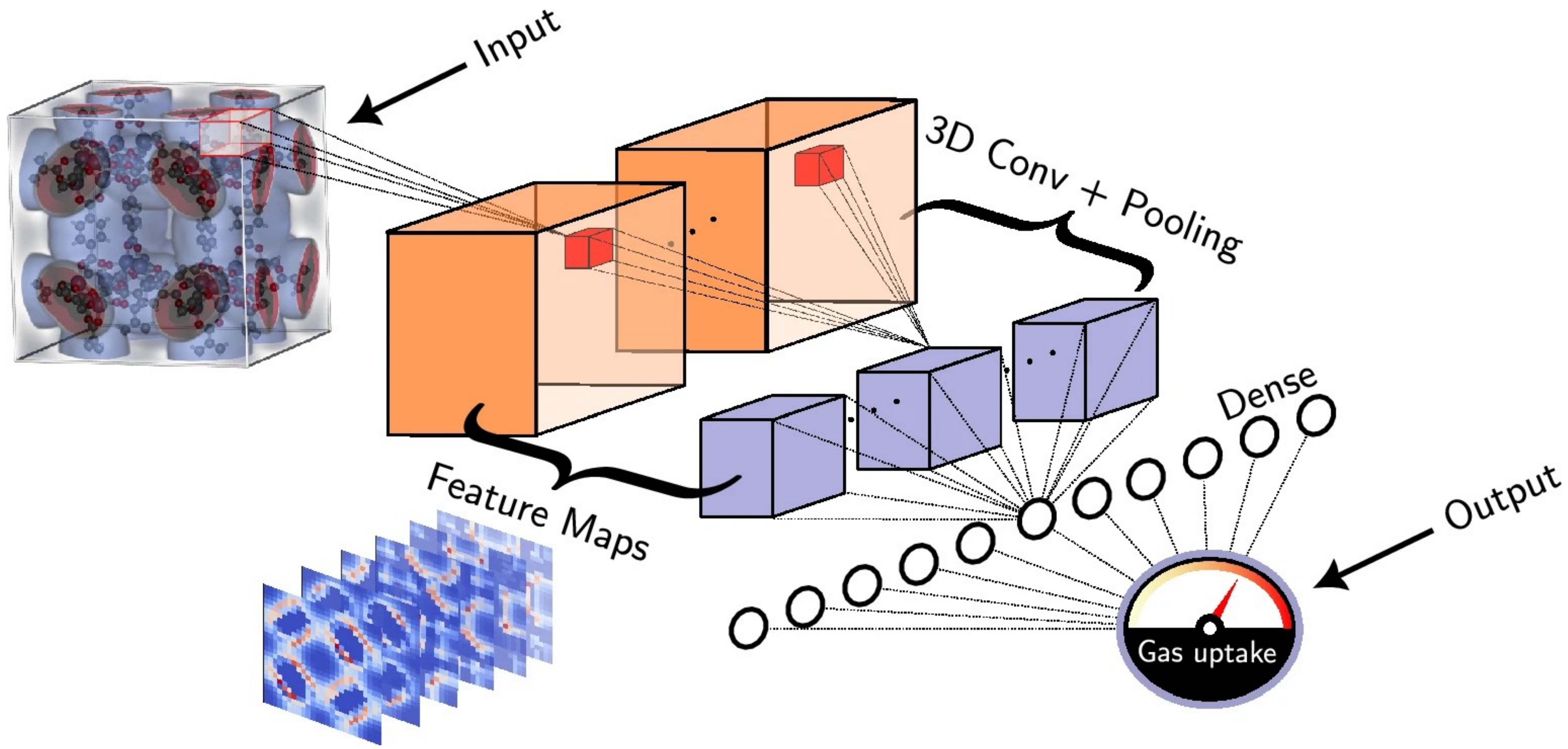
Antonios P. Sarikas, Konstantinos Gkagkas & George E. Froudakis

Scientific Reports **14**, Article number: 2242 (2024) | Cite this article

Harder task: predicting diffusion

Antonios P. Sarikas, Konstantinos Gkagkas & George E. Froudakis

Scientific Reports **14**, Article number: 2242 (2024) | Cite this article



- Calculation of potential energy surface is fast
- **Can the 3D-CNN learn Boltzmann equation? Yes**

$$\Delta H^0 = \left\langle E_i \ e^{-\beta E_i} \right\rangle_i$$

- How about guest–guest interactions?
  We can view the trained model as "mean field"
- But: can it perform better?

- How about eletrostatic effects?

Adding topological information to framework materials databases

Chimie ParisTech

Contributing to reference databases

Matter

CellPress

Article

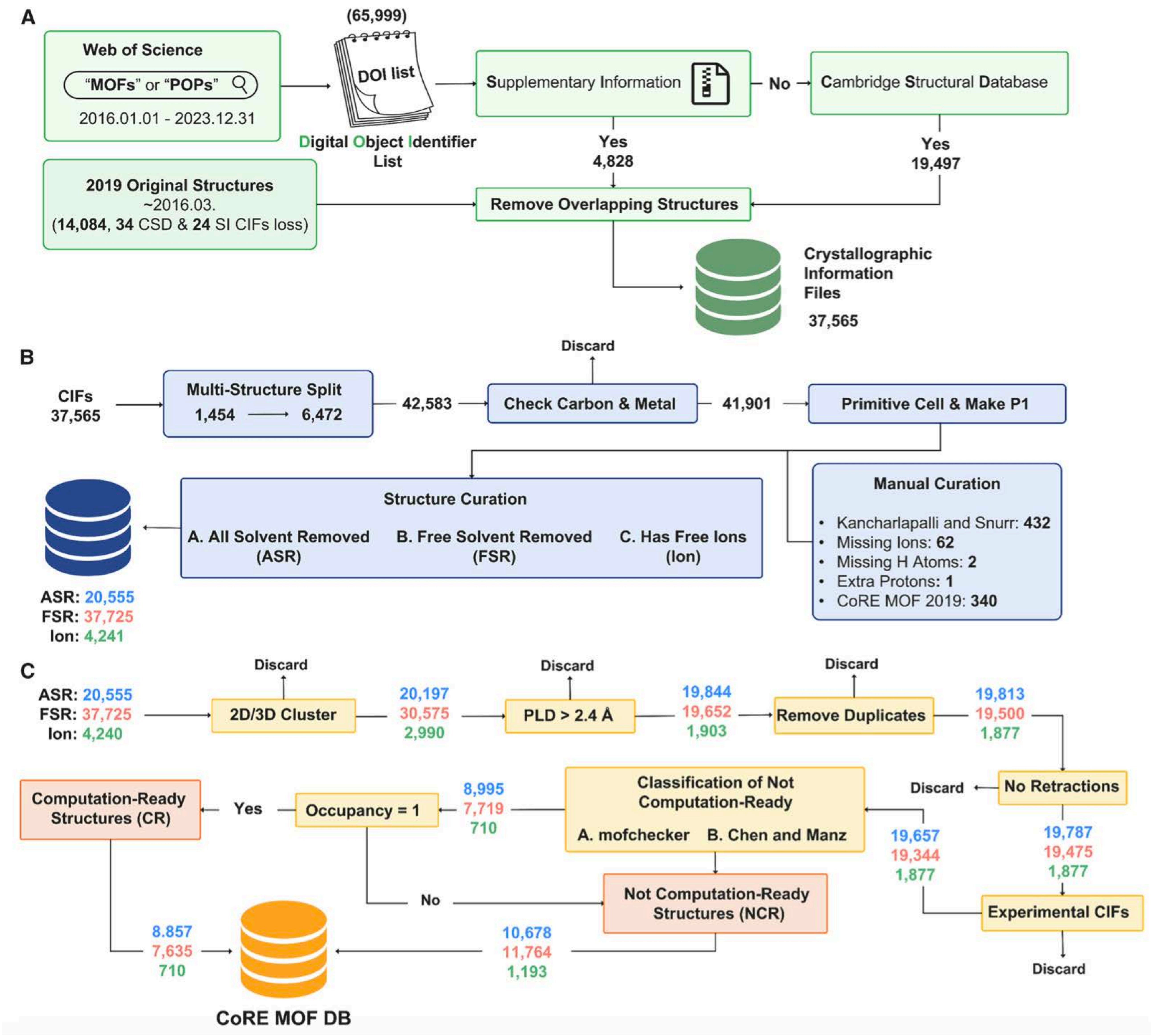
CoRE MOF DB: A curated experimental metal-organic framework database with machine-learned properties for integrated material-process screening

Guobin Zhao,[1] Logan M. Brabson,[2] Saumil Chheda,[3,4,5] Ju Huang,[6] Haewon Kim,[1] Kunhuan Liu,[7] Kenji Mochida,[7] Thang D. Pham,[7] Prerna,[3,4] Gianmarco G. Terrones,[8] Sunghyun Yoon,[1] Lionel Zoubritzky,[9,10] François-Xavier Coudert,[9,15] Maciej Haranczyk,[11,15] Heather J. Kulik,[8,12,15] Seyed Mohamad Moosavi,[6,15] David S. Sholl,[13,15] J. Ilja Siepmann,[3,4,15] Randall.Q. Snurr,[7,15] and Yongchul G. Chung[1,14,15,16,*]

- 8 research groups
- Lead by Greg Chung (Pusan National University)
- Emphasis on **quality over quantity**

CoRE MOF

Contributing to reference databases



- Database of structures, enriched with calculated properties
- **New in 2025: topology!**
- Predicted solvent-removal and water stability, through ML models
- Heat capacity, decomposition temperature
- Chemical decomposition
- Partial atomic charges

Determination of topology: CrystalNets

Home　　Edit on GitHub

CrystalNets.jl

`CrystalNets.jl` is a a Julia package for automatic detection and identification of net topologies underlying crystalline materials. Its inputs can be chemical files in any format recognized by chemfiles.

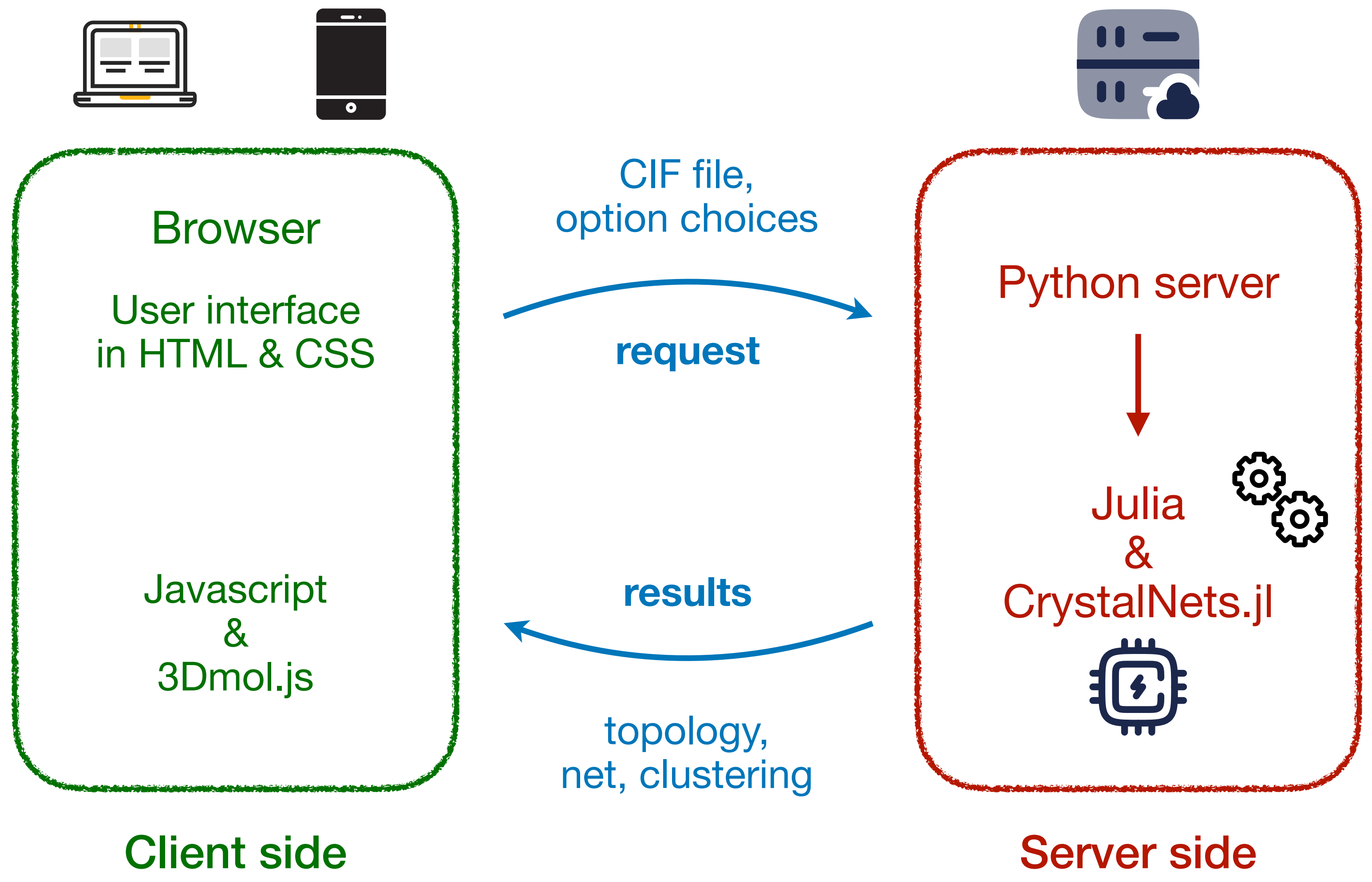
```
julia> determine_topology("/path/to/intertwinned/structures.cif")
2-element Vector{Tuple{Vector{Int64}, TopologyResult}}:
 ([2, 3, 4, 6], pcu)
 ([1, 5, 7, 8], srs)
```

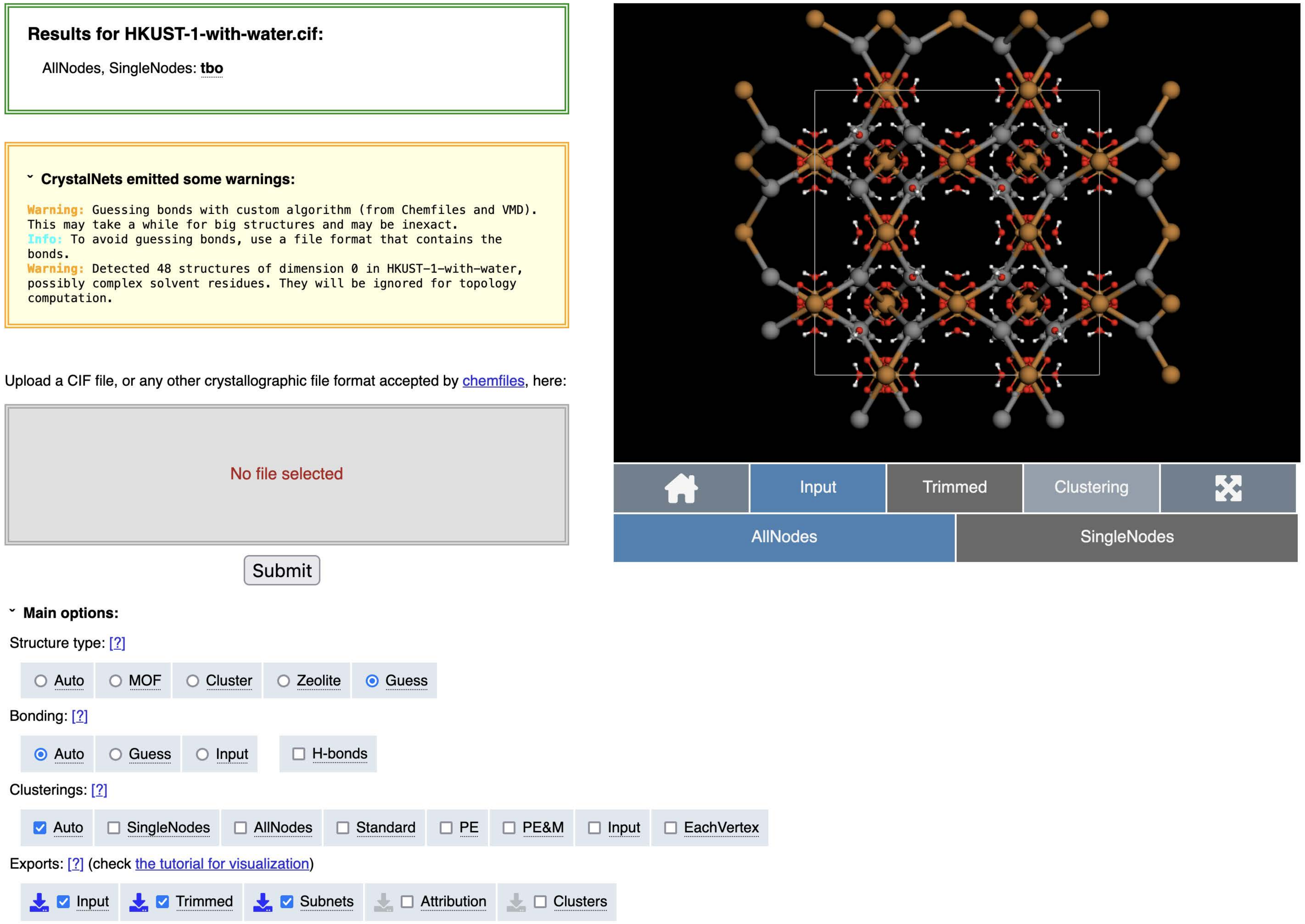
- ☆ Open source
- ☆ Algorithm more stable
- ☆ Highly optimised,
  faster than Systre or Topos

- ☆ **Since 2025: handles *unstable nets***
  (a combinatorial nightmare solved!)

Zoubritzky et al, *SciPost Chemistry*, **2022**

Determination of topology: CrystalNets

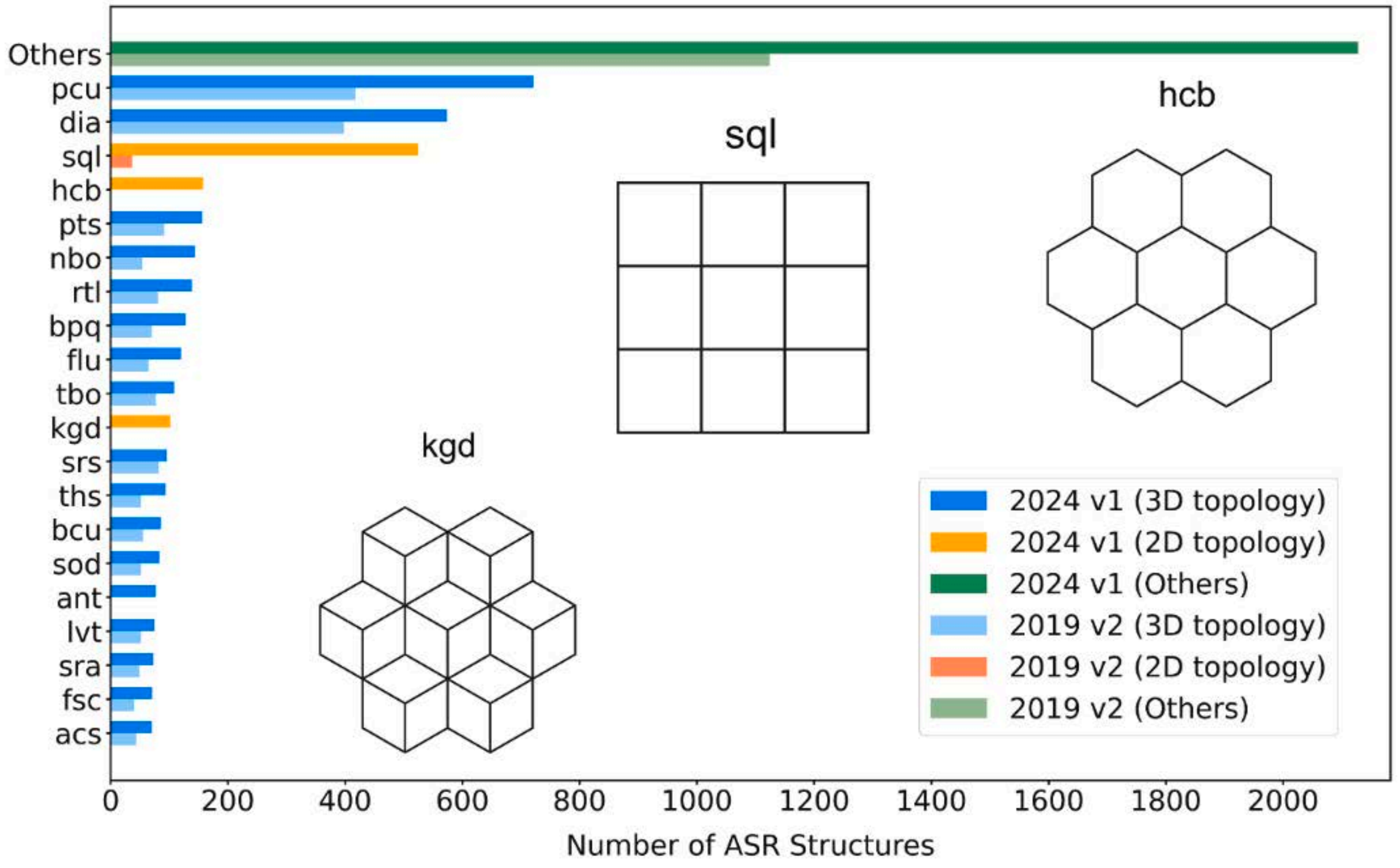


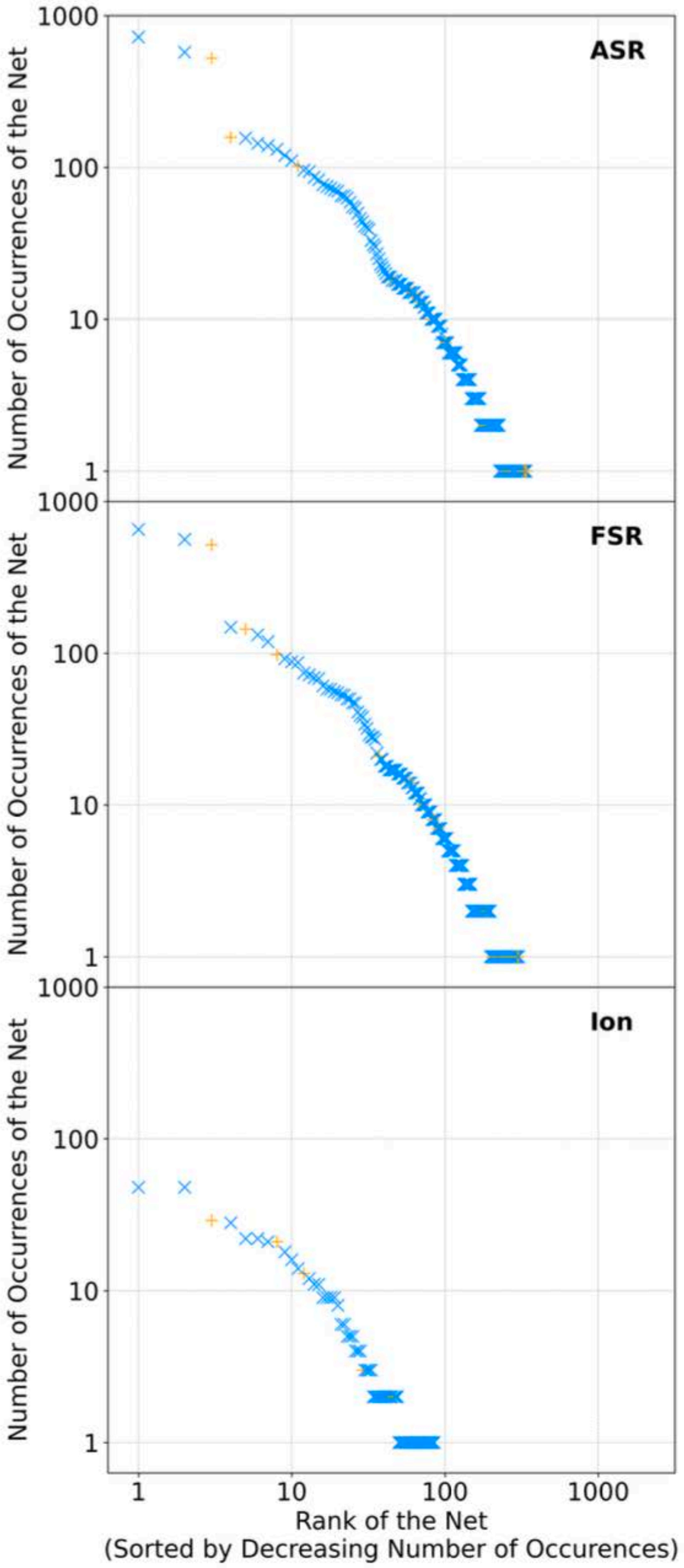


Zoubritzky et al, *ChemRxiv*, **2025**

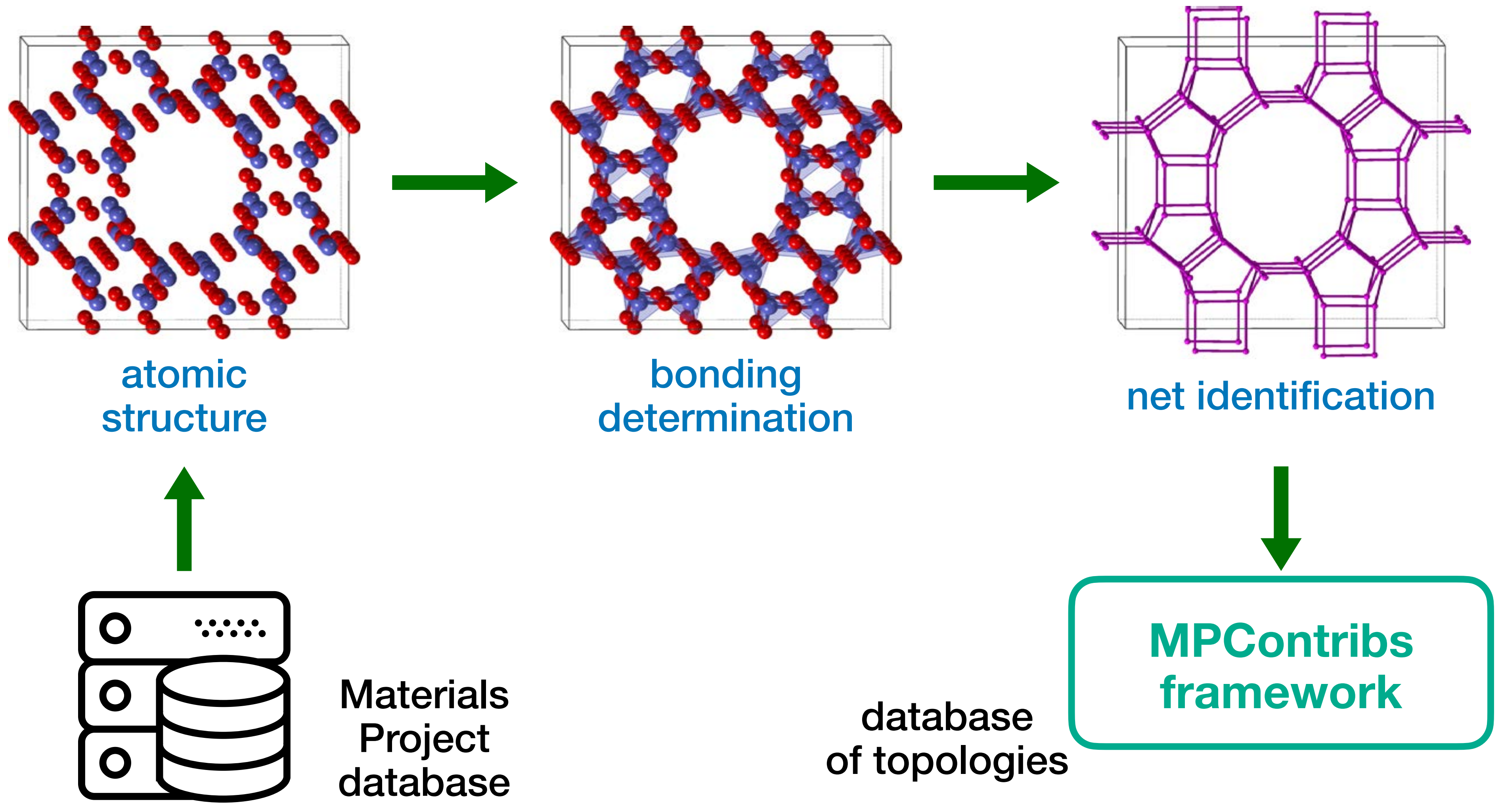
Why add topology information?

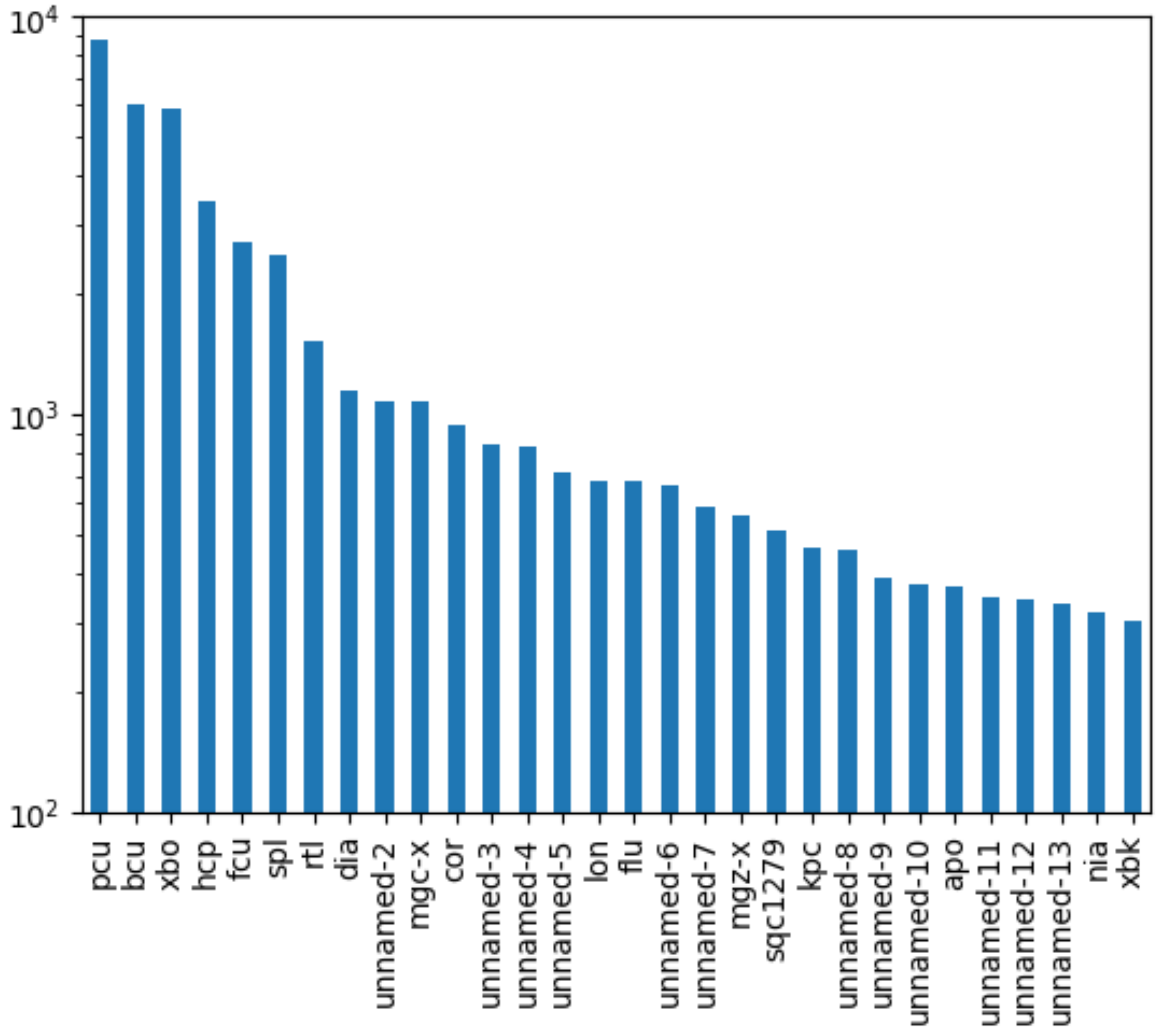
- Statistical analysis of topology
- Correlation between properties and topology, provide insight into the database
- Provide topology descriptors for training machine learning predictors of structure/property relationships



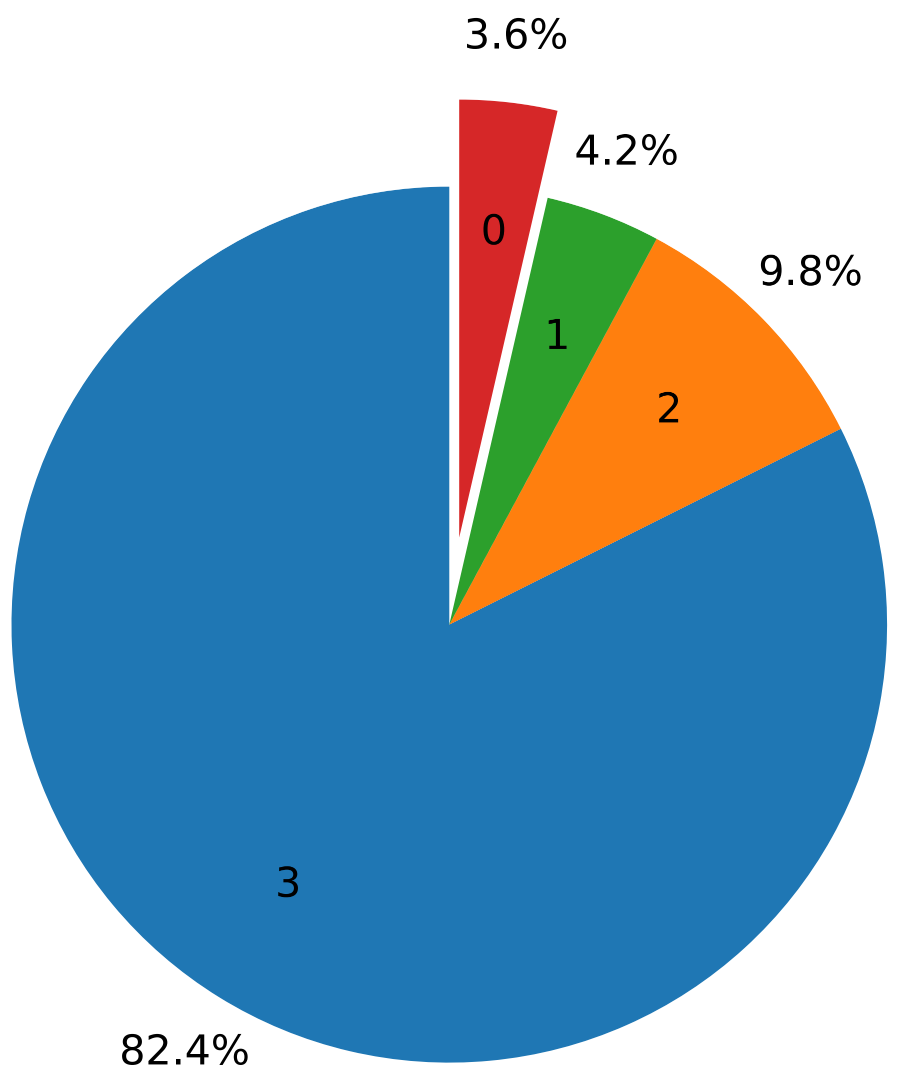


Search materials by topology: Materials Project





Dimensionality

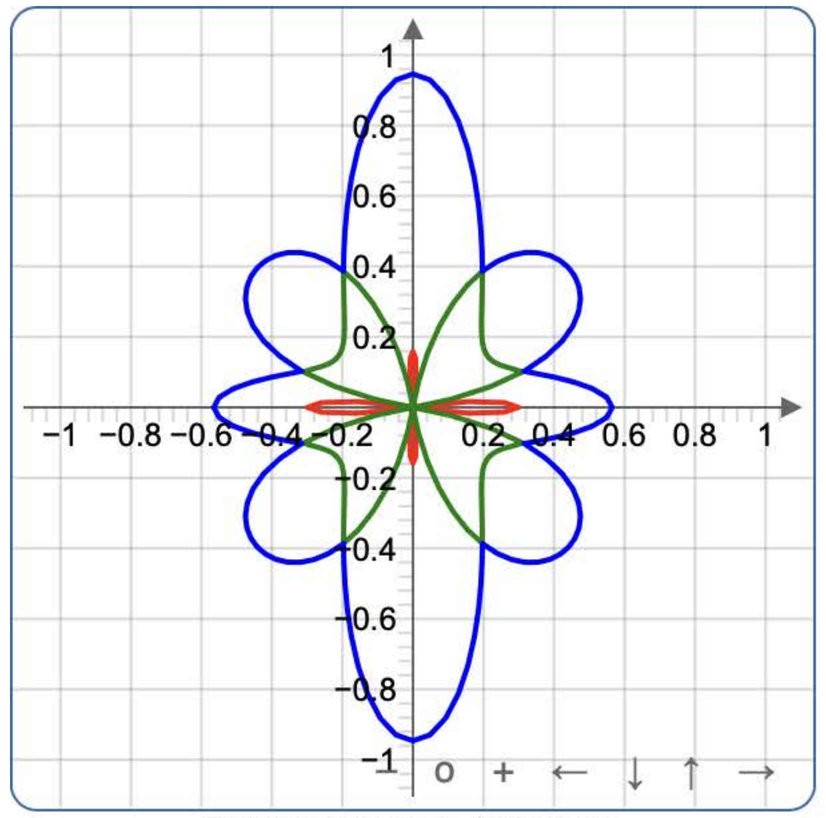


- More than 170,000 structures analyzed
- Distribution has a very long tail
- There is more diversity in topology that we typically consider (do we have a bias toward known topologies?)

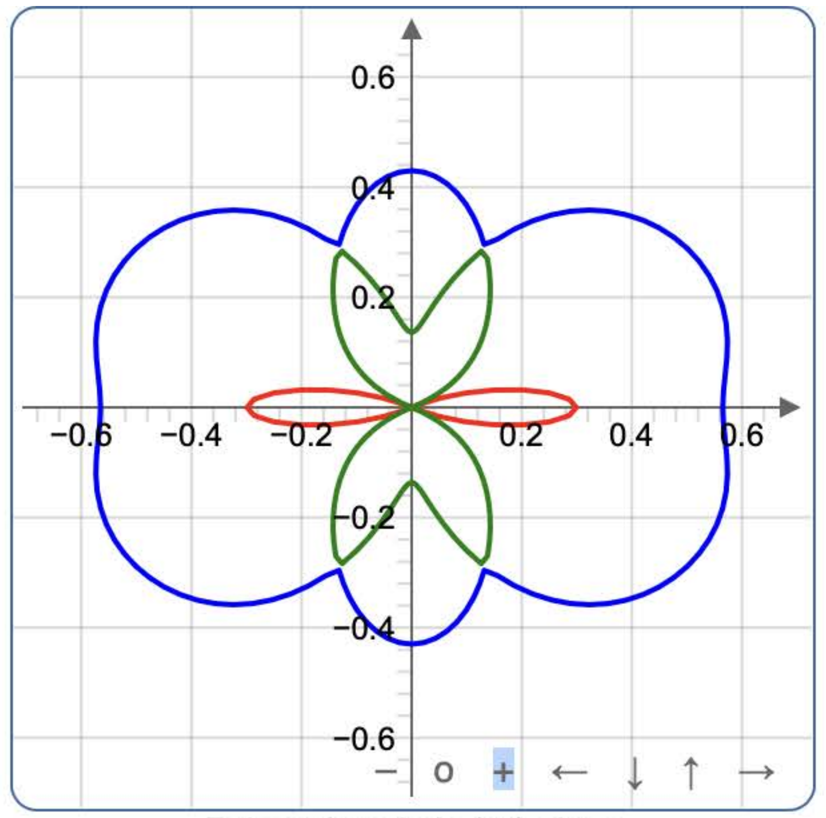
Search materials by topology: Materials Project

Example use case of "search by topology":

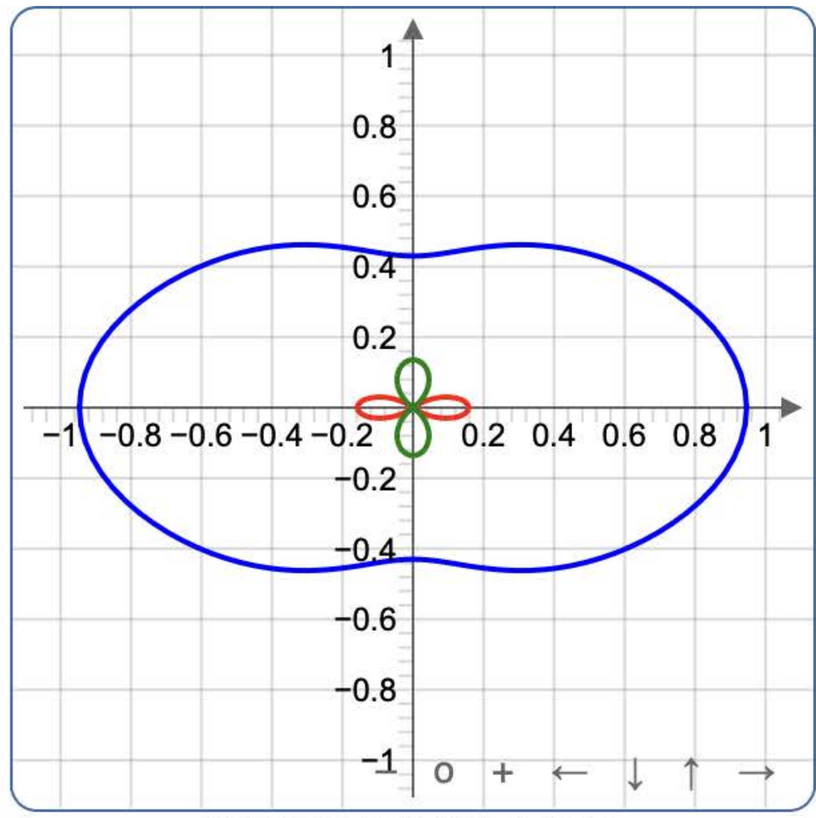
- MIL-53 MOF has "wine rack" framework, flexibility and auxeticity (negative Poisson's ratio)
- It has net topology **rna** or **bpq** (depending on choice of clustering)
- Are there inorganic analogues to MIL-53?



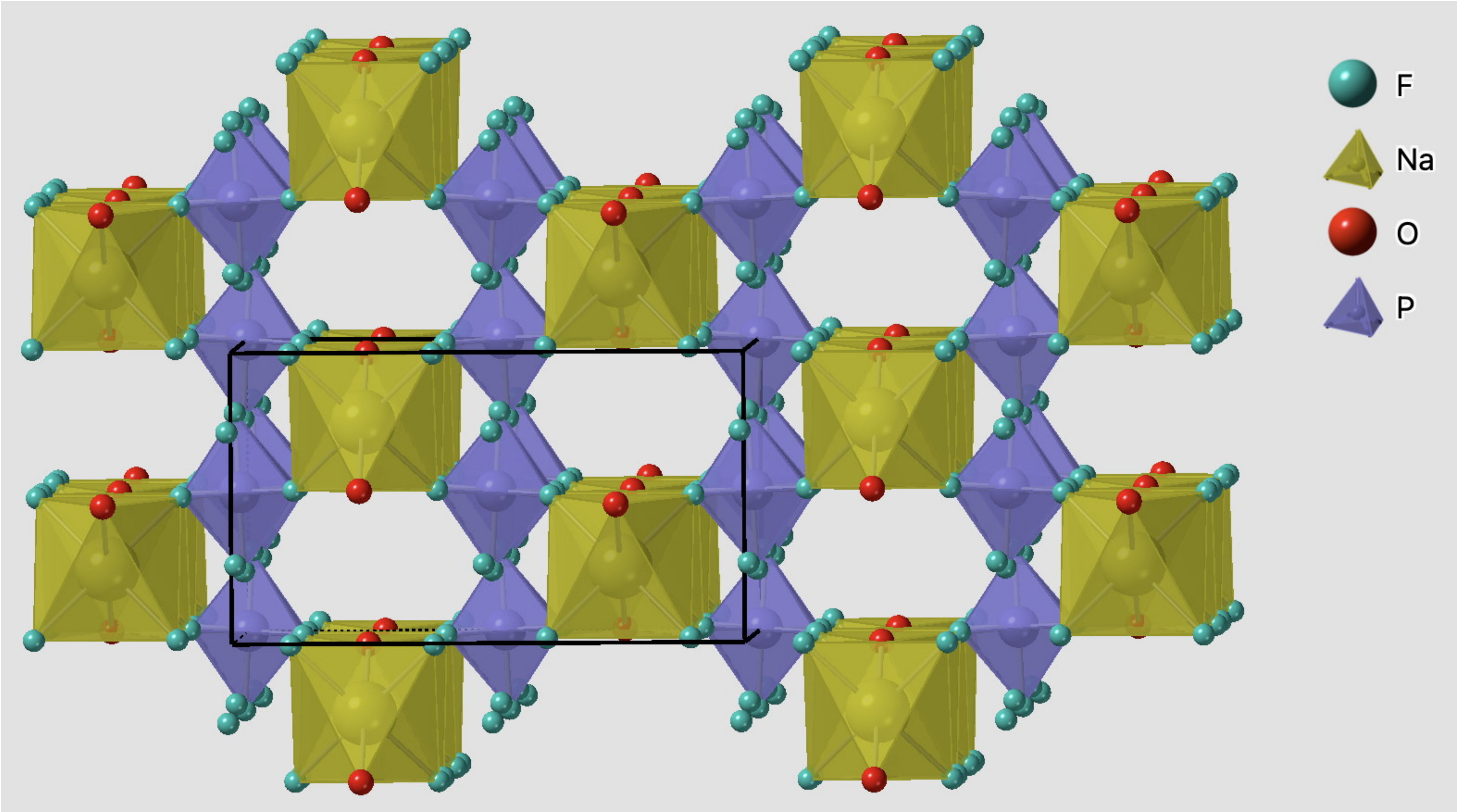
Poisson's ratio in (xy) plane



Poisson's ratio in (xz) plane



Poisson's ratio in (yz) plane



$NaPF_6H_2O$ (mp-767419)

Results:

- 25 structures in MP database, including 16 experimental materials
- One family emerges: $ABF_{6-n}(H_2O)_{1+n}$ with n = 0 or 1
- BF_6 octahedra that share corners with AO_2F_4 octahedra
- Confirm properties with DFT calculations: auxeticity confirmed

The need for databases

"This exciting paper shows AI design of materials, robotic synthesis. 10s of new compounds in 17 days."

nature

Explore content ⌄ About the journal ⌄ Publish with us ⌄

nature > articles > article

Article | Open access | Published: 29 November 2023

An autonomous laboratory for the accelerated synthesis of inorganic materials

Nathan J. Szymanski, Bernardus Rendy, Yuxing Fei, Rishi E. Kumar, Tanjin He, David Milsted, Matthew J. McDermott, Max Gallant, Ekin Dogus Cubuk, Amil Merchant, Haegyeom Kim, Anubhav Jain, Christopher J. Bartel, Kristin Persson, Yan Zeng & Gerbrand Ceder

Nature **624**, 86–91 (2023) | Cite this article

TABLE I. The distribution of errors in the 36 claimed "successful" syntheses. The "X" symbol denotes that the error is present. Error 1 is a very poor fit, such that the fitted model is meaningless. Error 2 is where a different CIF has been used for refinement compared with that in the paper and in the Materials Project. Error 3 is where the predicted structure has ordered cations but there is no evidence for order, and a known disordered version of the compound exists. Error 4 is where the compound is correctly identified but is already reported.

Claimed Phases	1	2	3	4	Claimed Phases	1	2	3	4
Ba_2ZrSnO_6	X	X	X		$Mg_3MnNi_3O_8$	X		X	
$Ba_6Na_2Ta_2V_2O_{17}$	X		X		Mg_3NiO_4		X	X	
$Ba_6Na_2V_2Sb_2O_{17}$	X				$MgCuP_2O_7$		X	X	
$CaCo(PO_3)_4$			X		$MgNi(PO_3)_4$	X		X	
$CaFe_2P_2O_9$					$MgTi_2NiO_6$			X	
$CaMn(PO_3)_4$			X		$MgTi_4(PO_4)_6$				X
$CaNi(PO_3)_4$			X		$MgV_4Cu_3O_{14}$	X	X	X	
$FeSb_3Pb_4O_{13}$			X		Mn_2VPO_7	X		X	
$Hf_2Sb_2Pb_4O_{13}$			X		$Mn_4Zn_3(NiO_6)_2$			X	
$InSb_3Pb_4O_{13}$			X		$MnAgO_2$	X			X
$K_2TiCr(PO_4)_3$			X		$Na_3Ca_{18}Fe(PO_4)_{14}$	X			
$K_4MgFe_3(PO_4)_5$	X				$Na_7Mg_7Fe_5(PO_4)_{12}$	X			
$K_4TiSn_3(PO_5)_4$	X				$NaCaMgFe(SiO_3)_4$		X	X	
$KBaPrWO_6$	X				$NaMnFe(PO_4)_2$	X			
KMn_3O_6	X	X	X		$Sn_2Sb_2Pb_4O_{13}$			X	
$KNaP_6(PbO_3)_8$	X	X	X		$Y_3In_2Ga_3O_{12}$	X			X
$KNaTi_2(PO_5)_2$			X		$Zn_2Cr_3FeO_8$			X	
$KPr_9(Si_3O_{13})_2$	X	X			$Zr_2Sb_2Pb_4O_{13}$			X	

*"We discuss all 43 synthetic products and point out four common shortfalls in the analysis. Conclusion: **no new materials have been discovered** in that work."*

Palgrave et al, *Phys. Rev. X*, **2024**

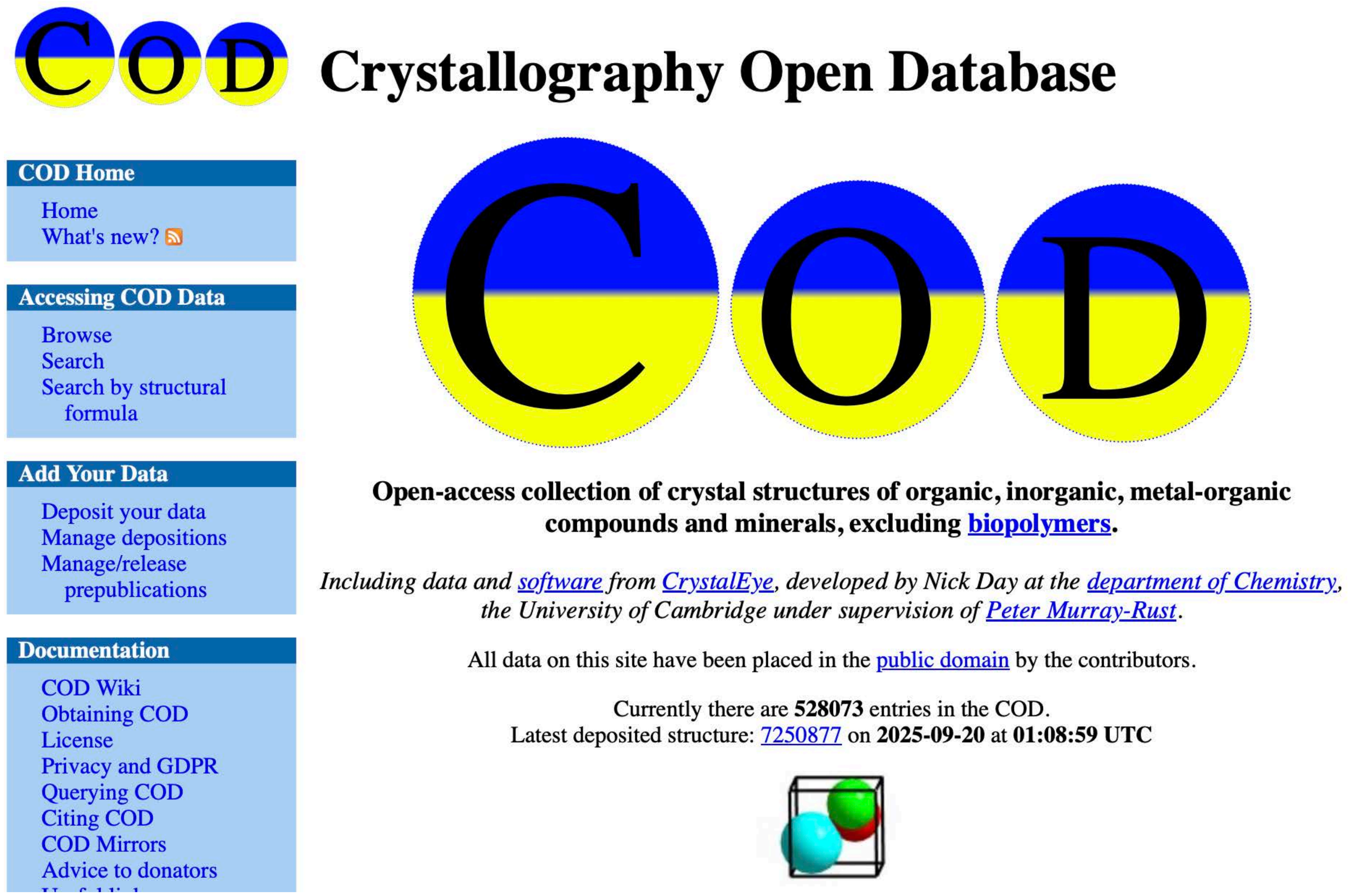
The need for open databases

- Data is obtained in large part by researchers and published in public repositories
- A large part of that effort is based on public funding
- However, **"public" does not mean "open"**
- Licences restrict programmatic access, modification and redistribution of data

Common "public" repositories with restrictive licensing:

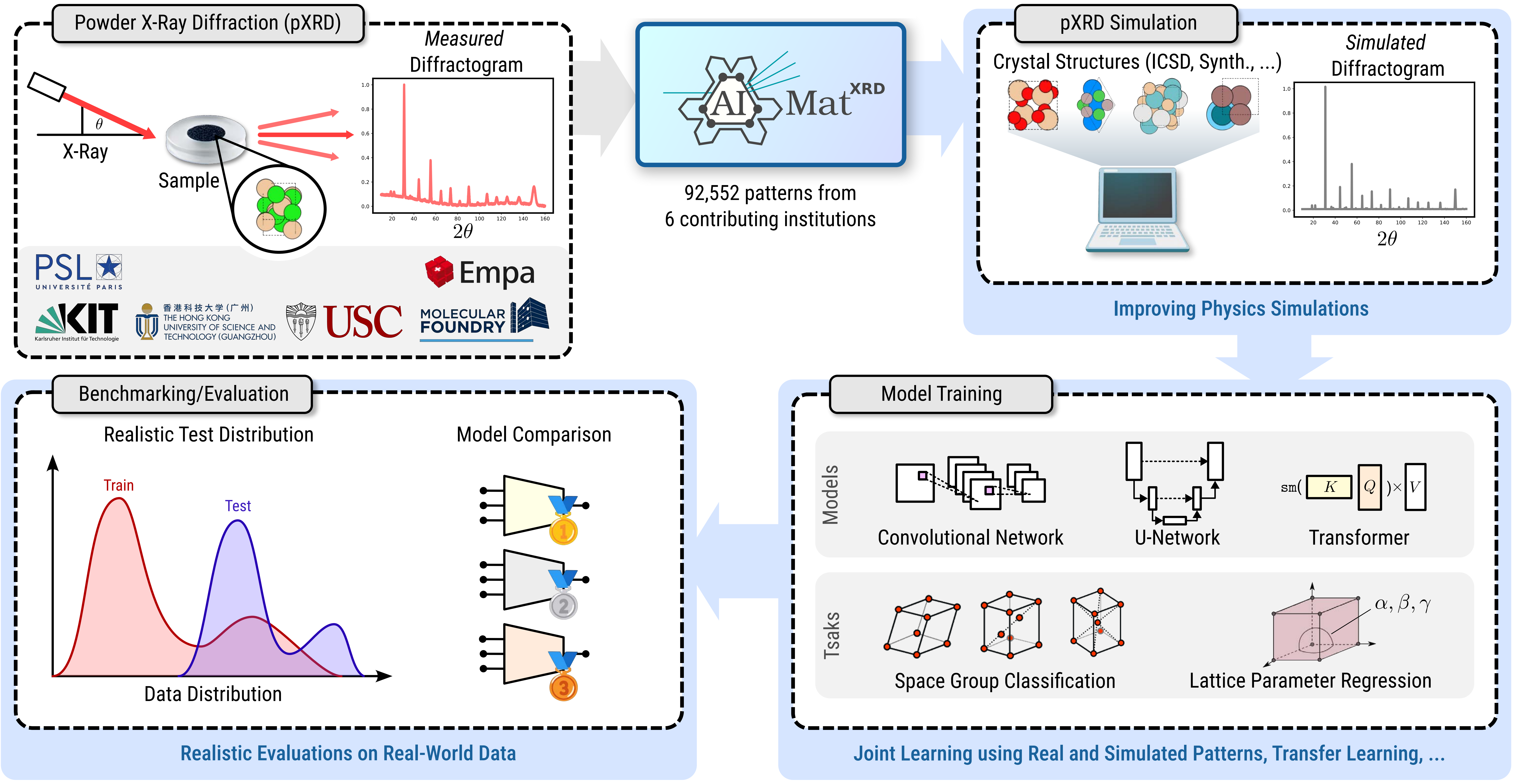
- *CSD (Cambridge Structural Database)*
- *Inorganic Crystal Structure Database (ICSD)*
- *Powder Diffraction File*

- **Please contribute your structures to the COD!**



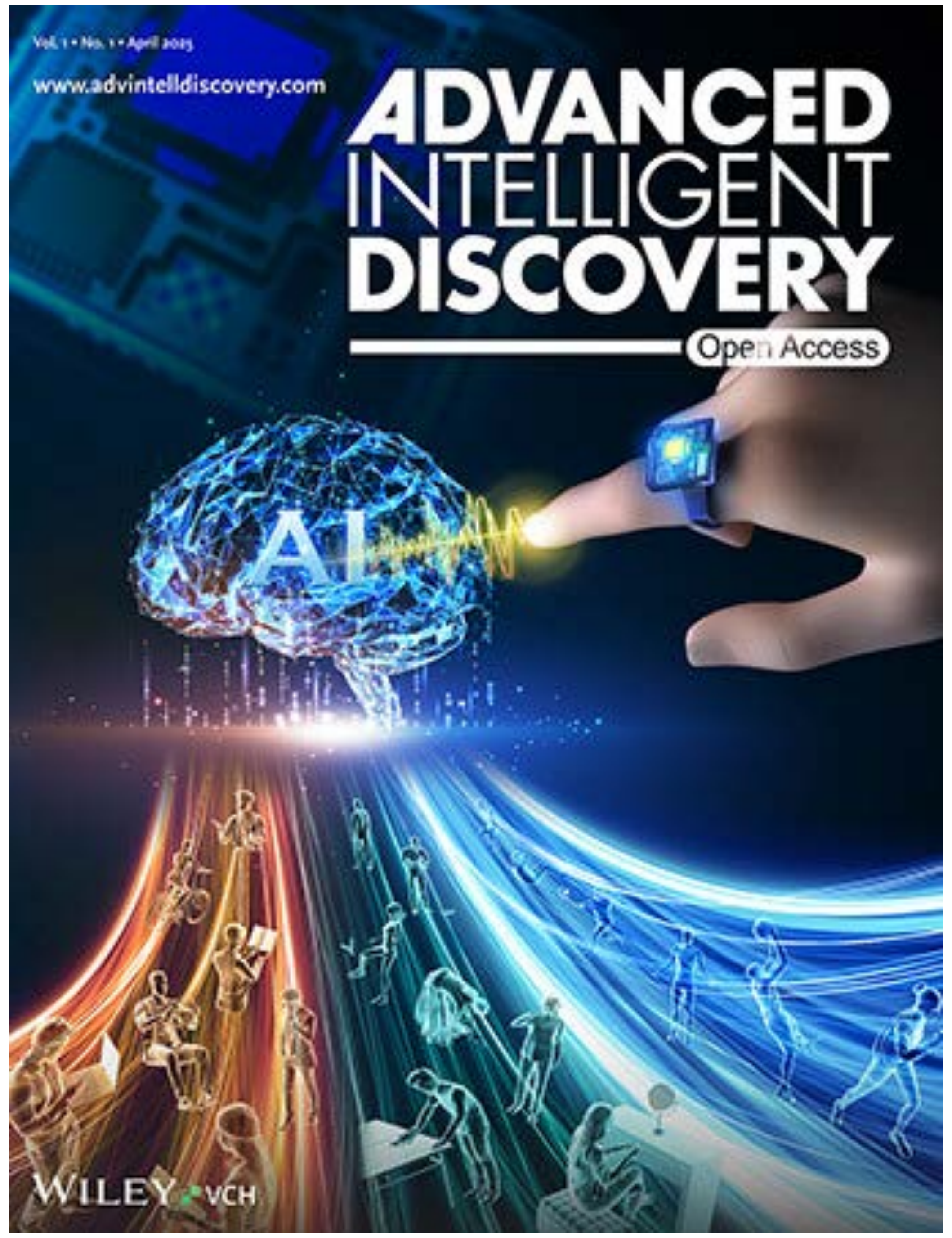
Contributing to reference databases

opXRD database

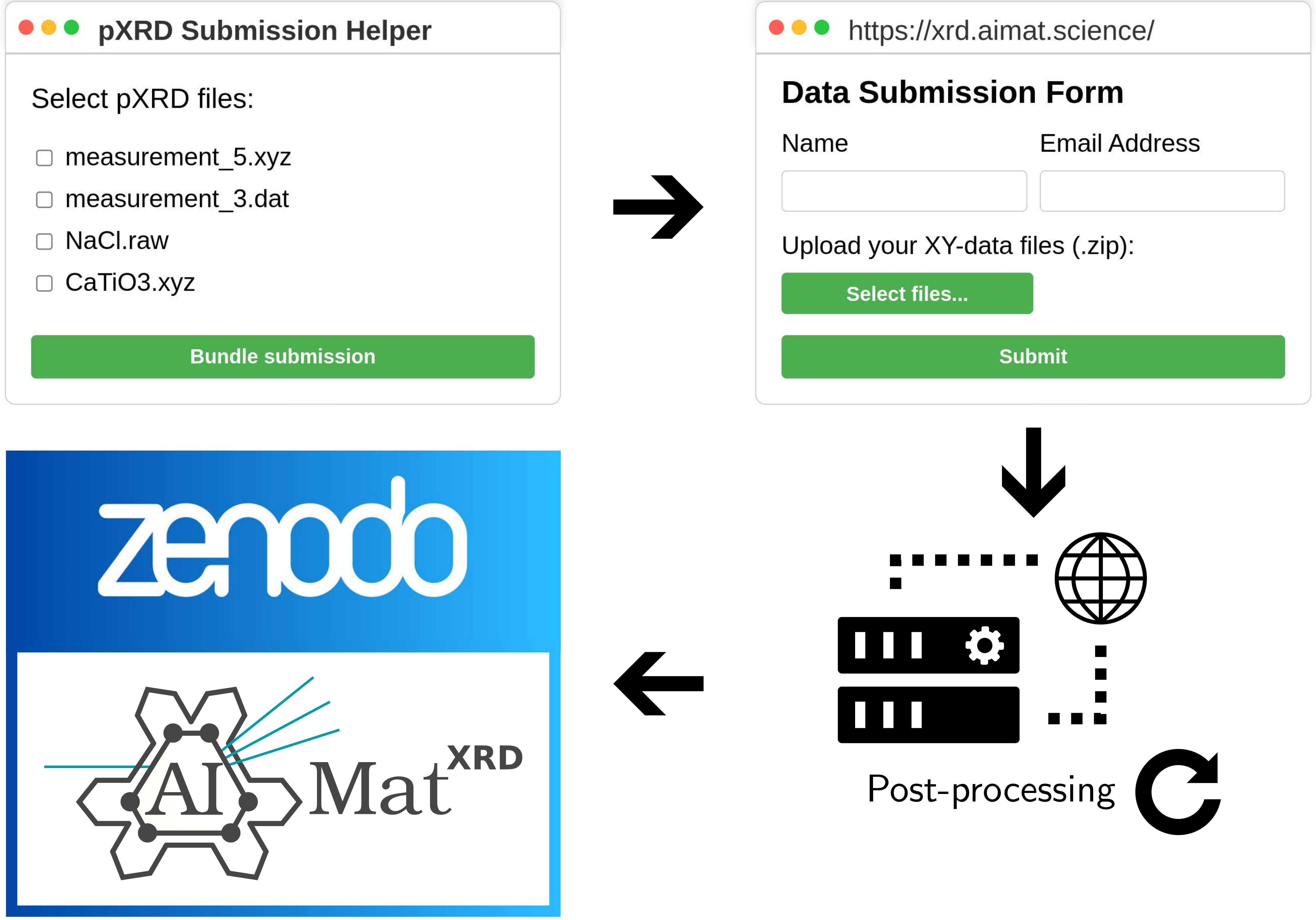


Contributing to reference databases

- Effort lead by Pascal Friederich and AiMat group @ KIT



- Contribute your group's data!



On this topic…

- Reproducibility of models

Nature Chemistry, 2021

Check for updates

comment

Best practices in machine learning for chemistry

Statistical tools based on machine learning are becoming integrated into chemistry research workflows. We discuss the elements necessary to train reliable, repeatable and reproducible models, and recommend a set of guidelines for machine learning reports.

Nongnuch Artrith, Keith T. Butler, François-Xavier Coudert, Seungwu Han, Olexandr Isayev, Anubhav Jain and Aron Walsh

- Introductory review

APL Materials RESEARCH UPDATE scitation.org/journal/apm

Machine learning approaches for the prediction of materials properties

Cite as: APL Mater. 8, 080701 (2020); doi: 10.1063/5.0018384
Submitted: 15 June 2020 • Accepted: 16 July 2020 •
Published Online: 4 August 2020

View Online Export Citation CrossMark

Siwar Chibani and François-Xavier Coudert[a)]