

# Machine learning models for accelerating materials discovery

Christopher Sutton

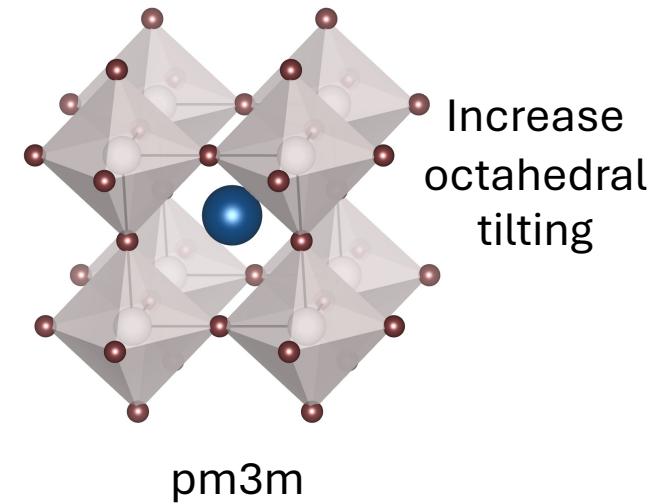
July 17, 2024



 **South Carolina**

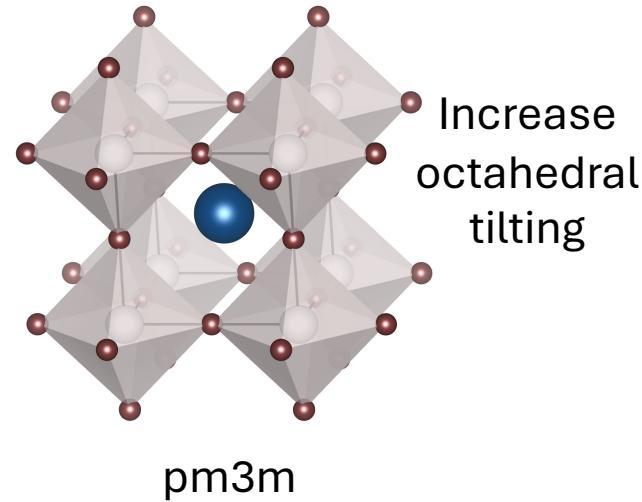
Need to determine the relevant structure  
to accurately calculate the material properties

The cubic phase is the prototypical perovskite structure  
used in calculations<sup>1,2</sup>

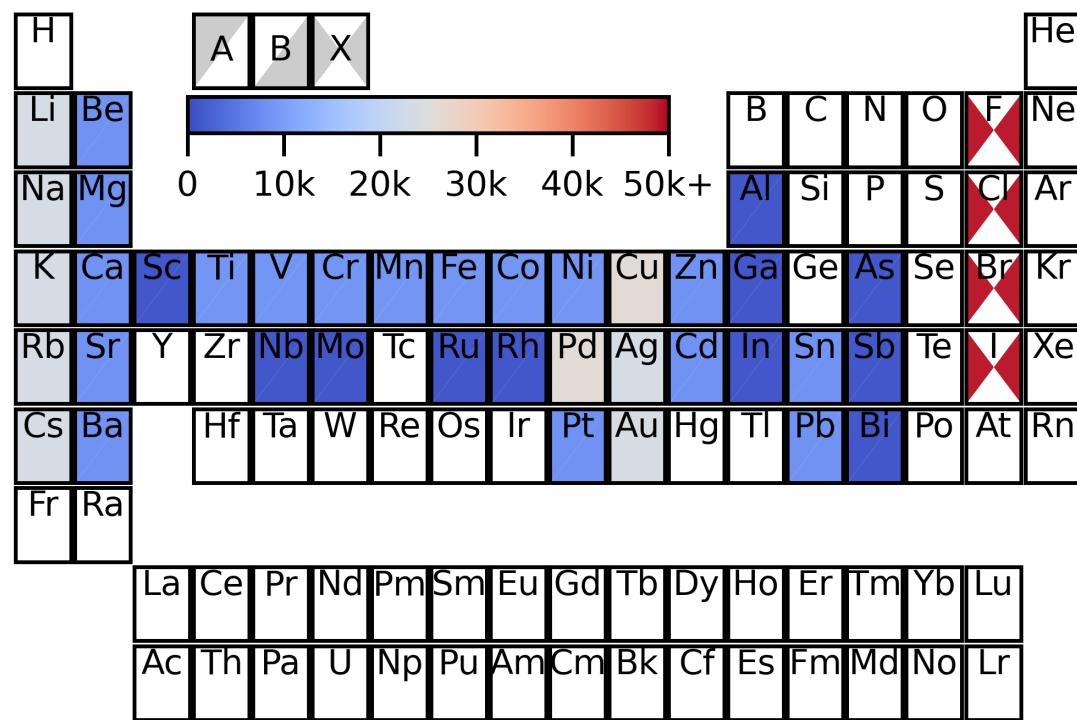


1. I.E. Castelli, T. Olsen, S. Datta, D. D. Landis, S. Dahl, K. S. Thygesen, and K. W. Jacobsen, Energy & Environmental Science (2012)
2. J. Schmidt, J. Shi, P. Borlido, L. Chen, S. Botti, and M. A. Marques, Chem. Mater. (2017).

Need to determine the relevant structure  
to accurately calculate the material properties



The cubic phase is the prototypical perovskite structure used in calculations<sup>1,2</sup>



1. I.E. Castelli, T. Olsen, S. Datta, D. D. Landis, S. Dahl, K. S. Thygesen, and K. W. Jacobsen, Energy & Environmental Science (2012)

2. J. Schmidt, J. Shi, P. Borlido, L. Chen, S. Botti, and M. A. Marques, Chem. Mater. (2017).

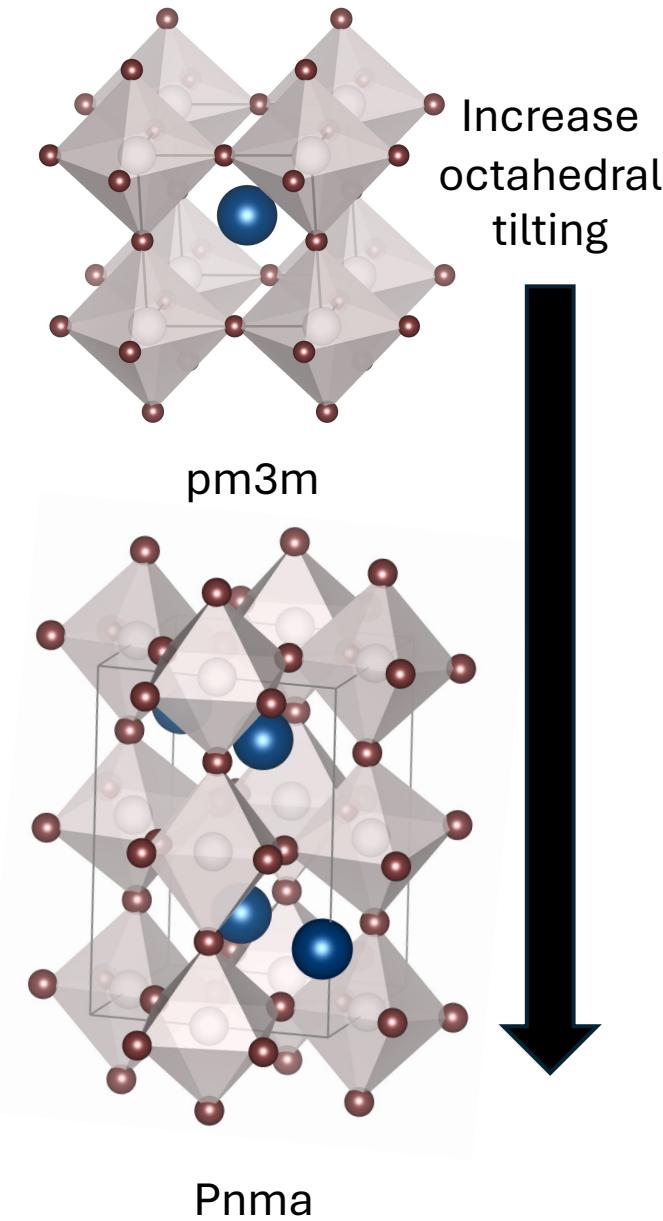
Need to determine the relevant structure  
to accurately calculate the material properties

The cubic phase is the prototypical perovskite structure  
used in calculations<sup>1,2</sup>

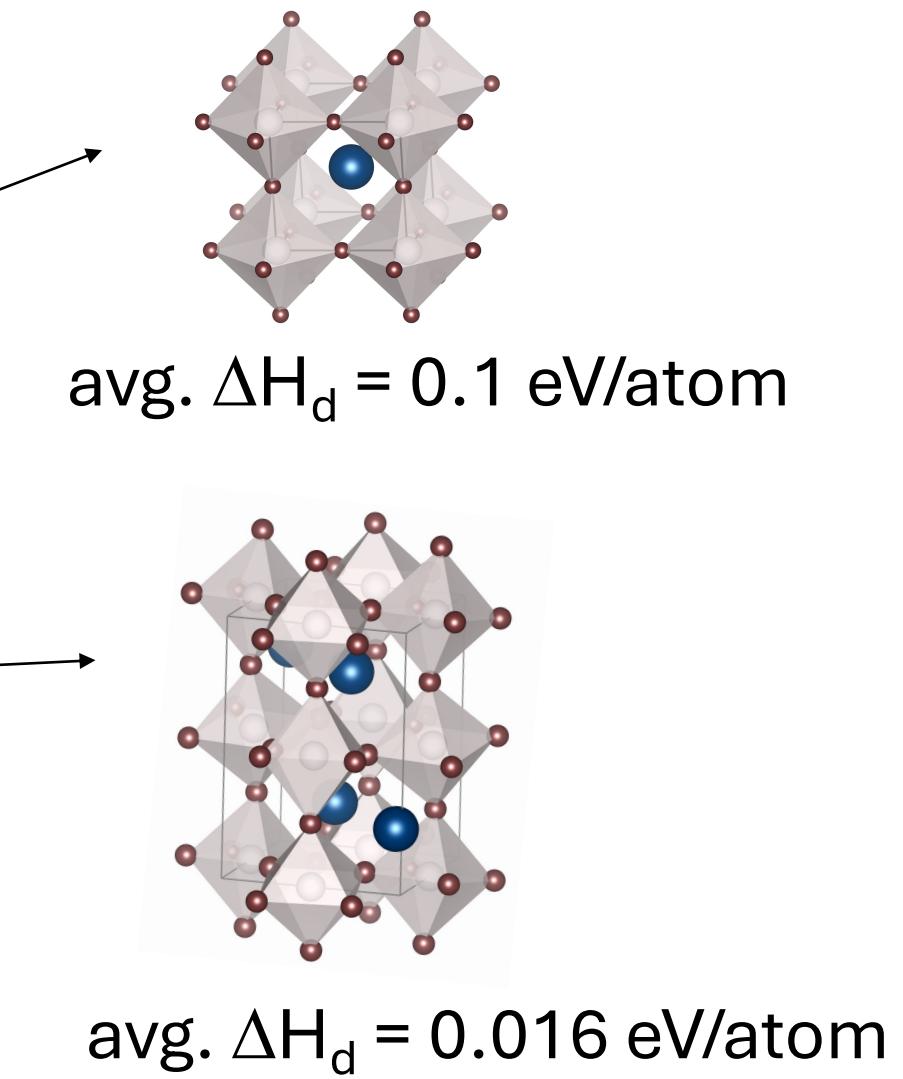
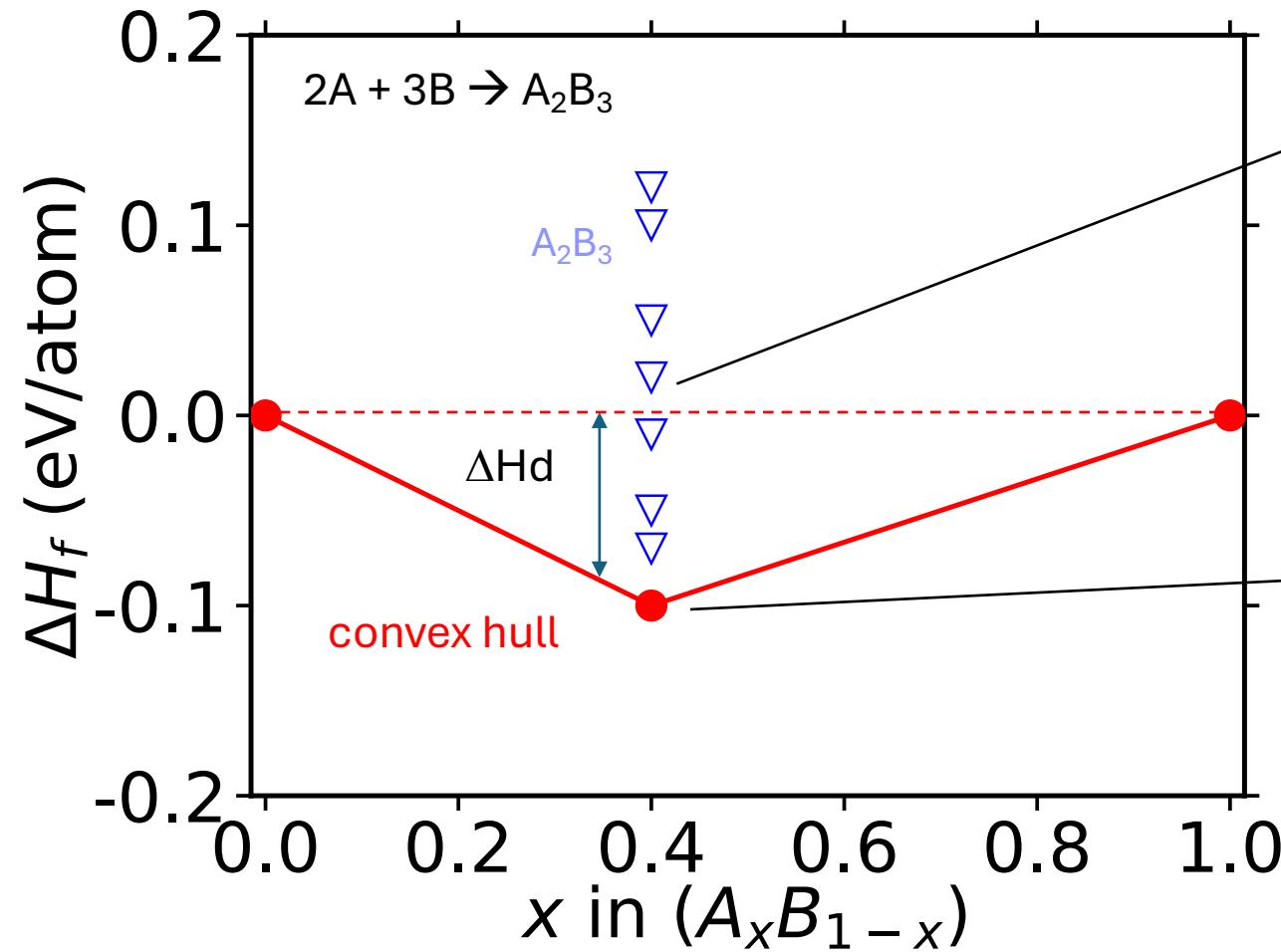
The cubic structure can distort to 23 different perovskite  
phases (referred to as Glazer tilts)

A. M. Glazer, The classification of tilted octahedra in  
perovskites. *Acta Cryst. B* 28, 3384–3392 (1972)

Various factors lead to distortions (such differences in  
ionic radii) but these other phases are often neglected  
in high-throughput screening tasks



Previous work showed that the average difference between the cubic and ground state is  $\sim 0.1$  eV/atom

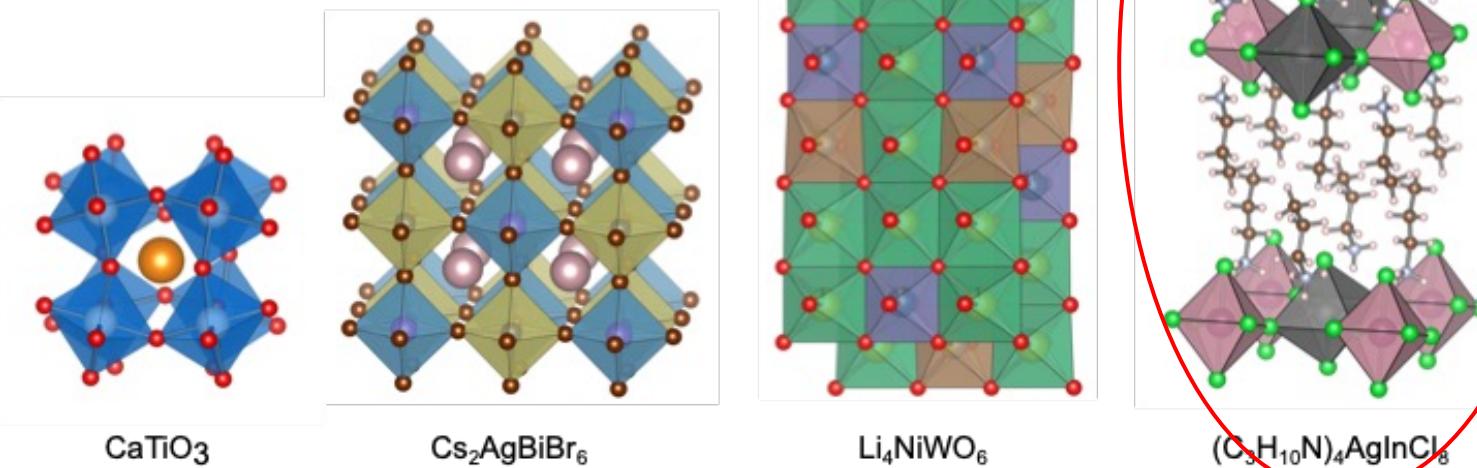


# Structural diversity of compounds complicates the use of structural-based ML models

These large systems necessitates using machine learning to bypass DFT

$$f(\text{structure}) \rightarrow \text{e.g., } \Delta H_f \text{ or } E_g$$

... requires a cost-effective input structure for the ML model



single  
perovskite

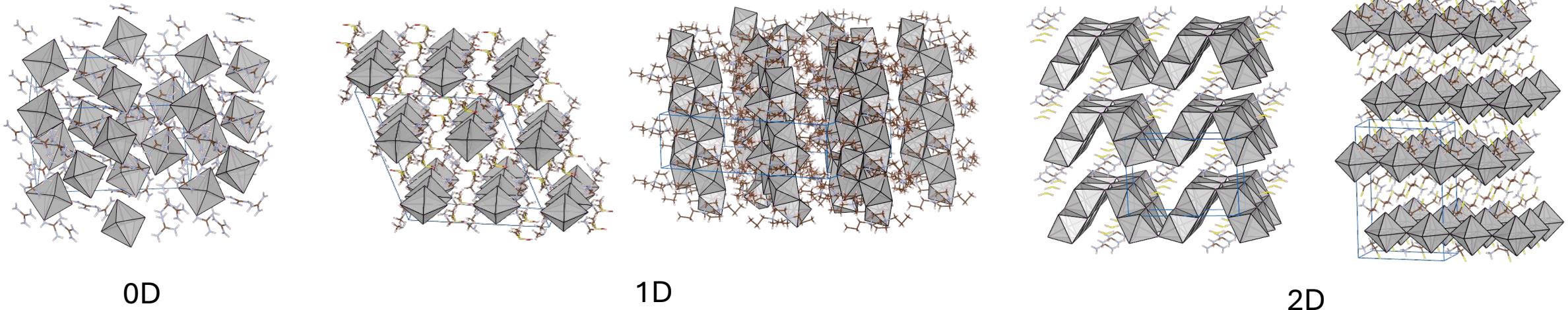
double  
perovskite

nonstoichiometric  
oxide

hybrid  
perovskite

# An immense diversity of structures have been synthesized for hybrid organic/inorganic perovskites

These systems possess a diverse octahedral networks, depending on the specific organic spacers



Modification of the octahedra network connectivity from purely corner- sharing to edge-sharing increases the bandgap by  $\sim 0.5$  eV, and a further  $\sim 0.5$  eV increase from purely edge-sharing to face-sharing.

M. E. Kamminga, et al. Inorganic Chemistry 56 (2017)

# Developed a generalizable ML force-fields and structure prediction algorithm to predict the crystal structures of new compounds

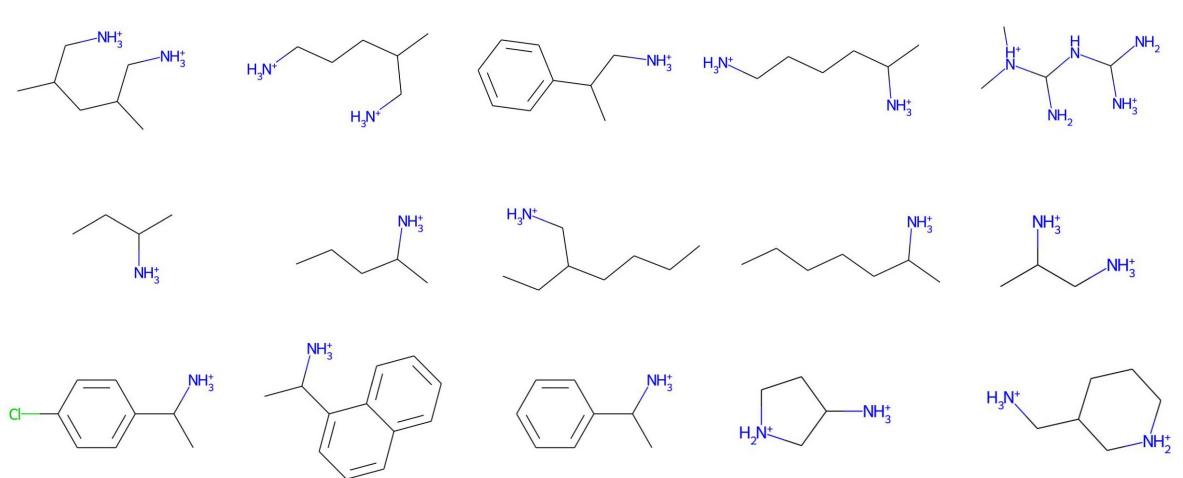


Gabor  
Csanyi

Will  
Baldwin

Nima  
Karimitari

## 1. Select organic cation

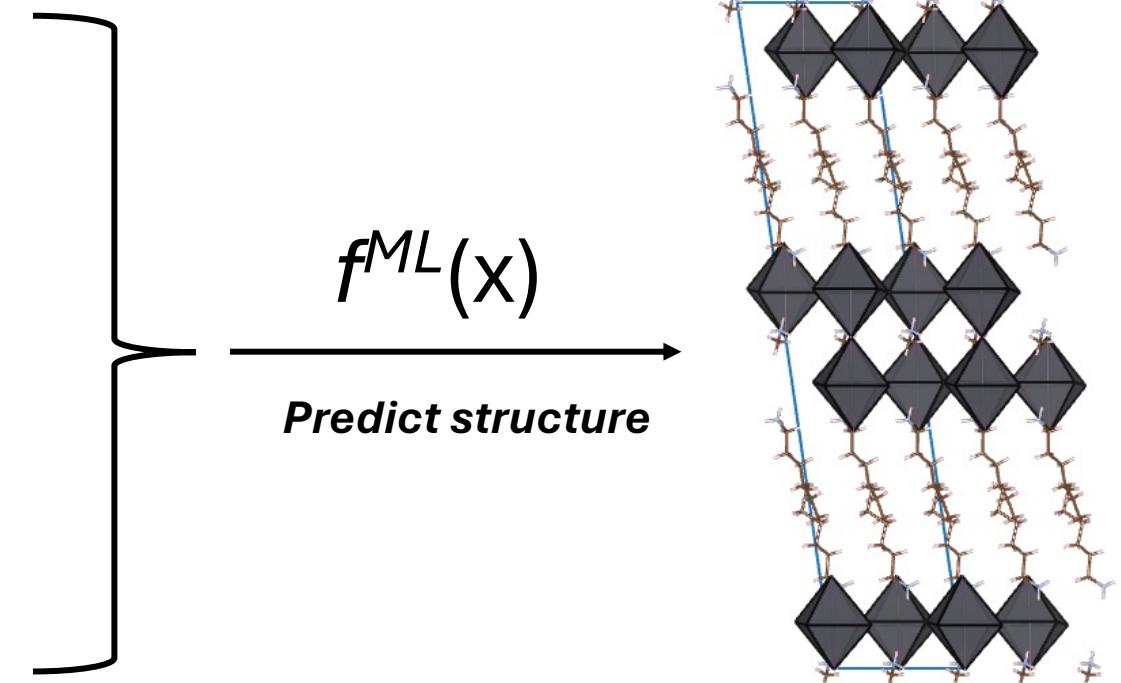


## 2. Select metal cation and halide combination



Currently extending to other cations: Mn, Fe, Co, Ni, Cu, Zn, Cd, Pb, Sn, Bi, Sb

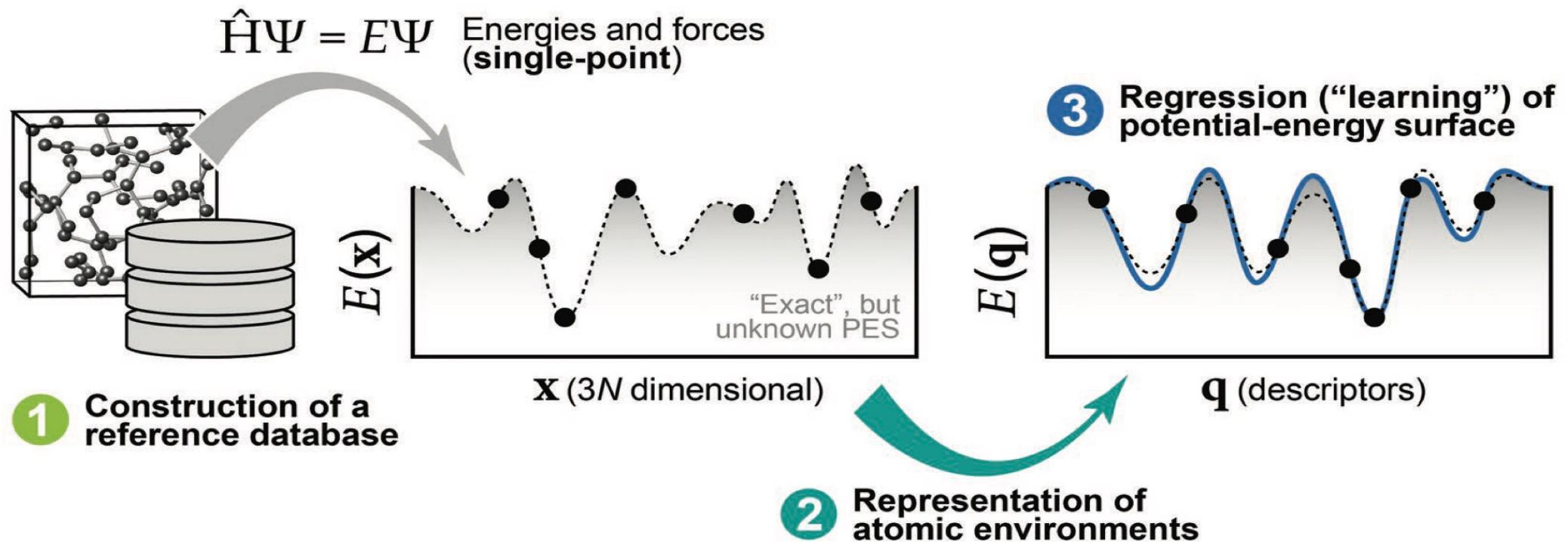
N Karimitari, WJ Baldwin, EW Muller, ZJL Bare, WJ Kennedy, G Csányi, C Sutton “Accurate Crystal Structure Prediction of New 2D Hybrid Organic Inorganic Perovskites” under review JACS, arXiv:2403.06955 (2024)



$$f^{\text{ML}}(x)$$

**Predict structure**

Goal: Train a generalizable MLIP for halide perovskites  
that reproduces the accuracy of the reference electronic structure model  
(but without the electrons)



# Message passing NNs mimic many-body interactions

3

Regression (“learning”) of potential-energy surface

Fitting to a parametric function of the local (site) energy :

$$E_i = V^{(1)}(\mathbf{r}_i) + \frac{1}{2} \sum_j V^{(2)}(\mathbf{r}_i, \mathbf{r}_j) + \frac{1}{3!} \sum_{jk} V^{(3)}(\mathbf{r}_i, \mathbf{r}_j, \mathbf{r}_k) + \dots$$

MACE<sup>1</sup> is a graph neural network, where nodes are atomic sites and edges are bonds

$$\sigma_i^{(t)} = (\mathbf{r}_i, \boldsymbol{\theta}_i, \mathbf{h}_i^{(t)})$$

$\mathbf{r}_i$  → position

$\boldsymbol{\theta}_i$  → Attributes (e.g., Z)

$\mathbf{h}_i^{(t)}$  → learnable parameters

Main steps:

Message construction

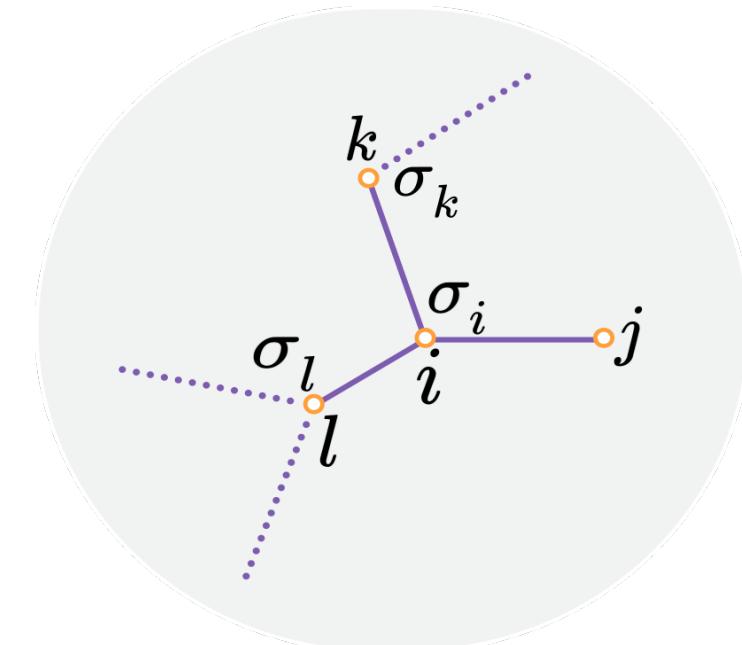
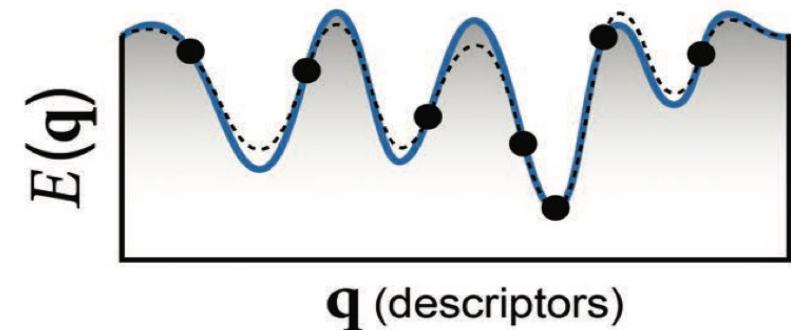
$$\mathbf{m}_{i,j}^{(t)} = f_{message}(\sigma_i^{(t)}, \sigma_j^{(t)}, \mathbf{e}_{ij})$$

Aggregation/ Update:

$$\sigma_i^{(t+1)} = (\mathbf{r}_i, \boldsymbol{\theta}_i, \mathbf{h}_i^{(t+1)}) = (\mathbf{r}_i, \boldsymbol{\theta}_i, U_t(\sigma_i^{(t)}, \mathbf{m}_i^{(t)}))$$

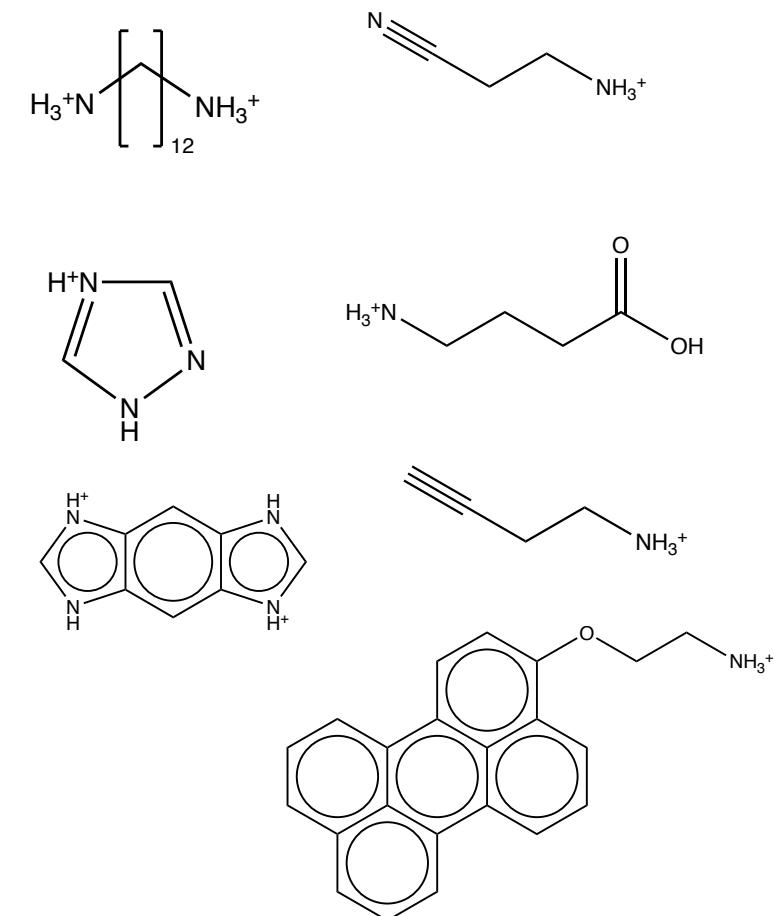
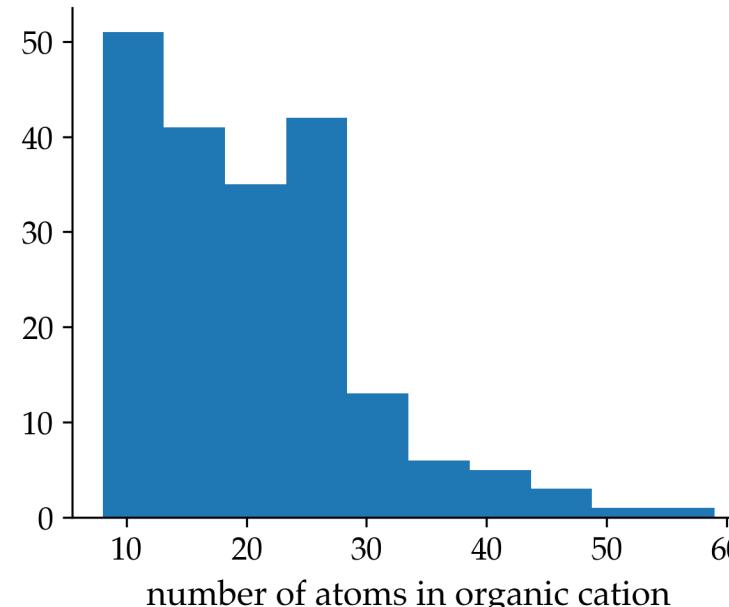
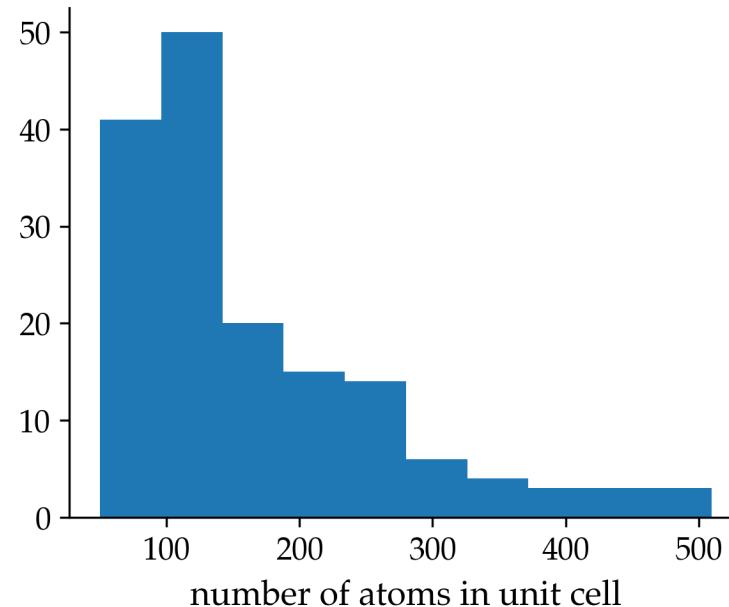
Readout:

$$E_T = \sum_i f_{readout}(\sigma_i^{(t)})$$

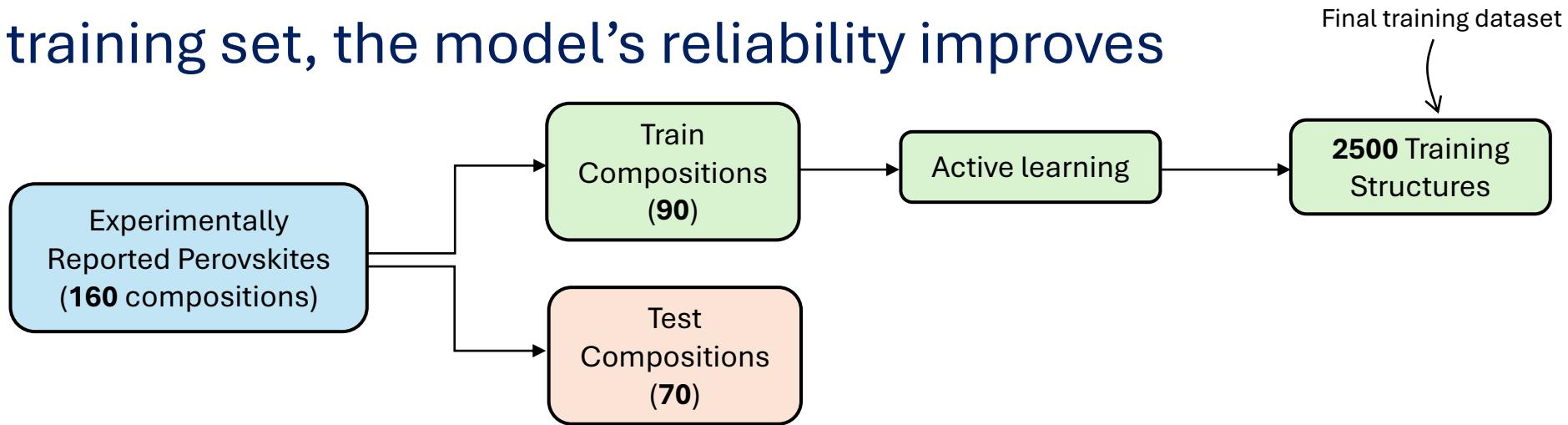


# Dataset construction: a diverse set of structures and organics molecules

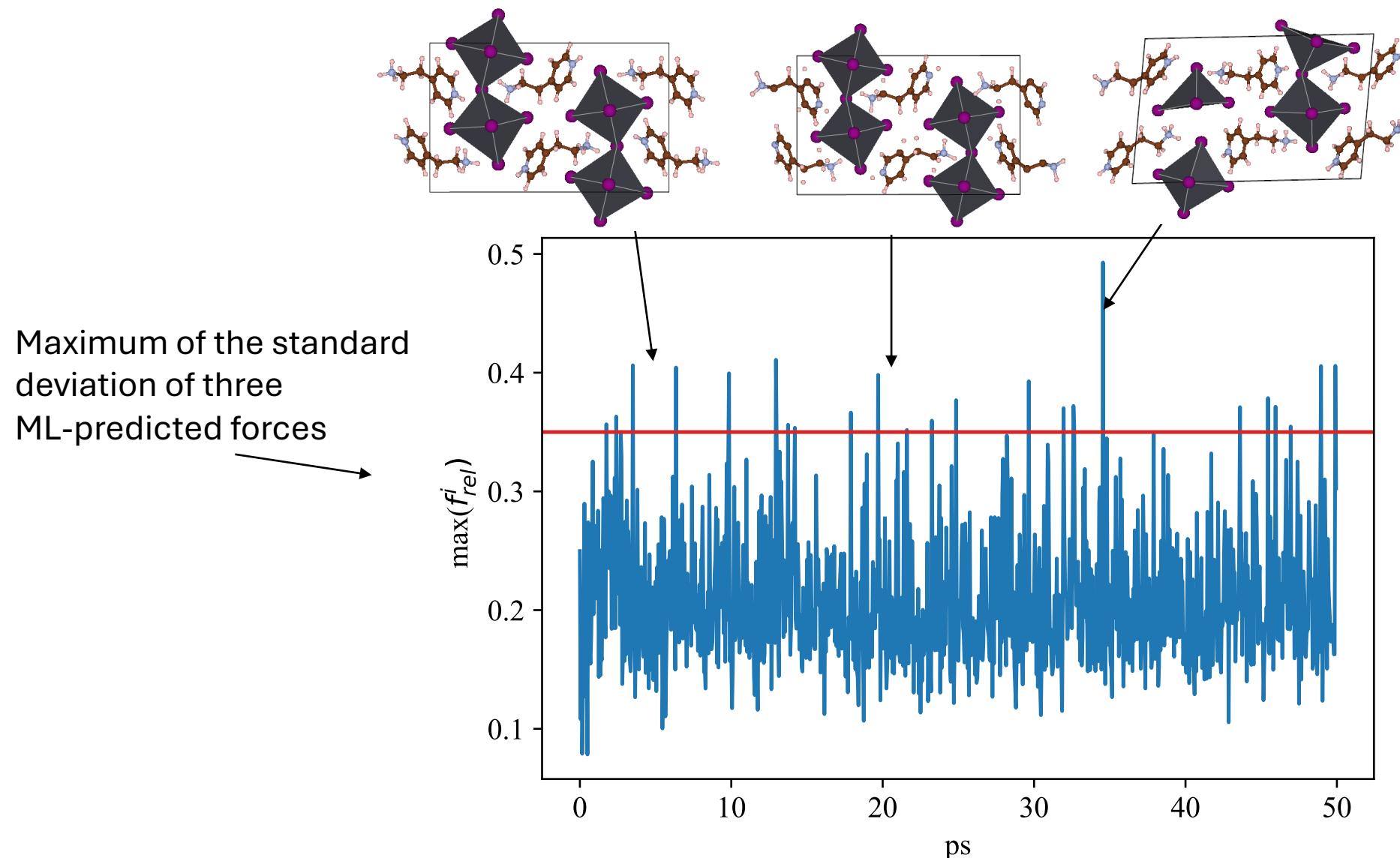
- Total dataset: 160 of 700 experimentally reported compounds
- 90 for training, 70 in test set



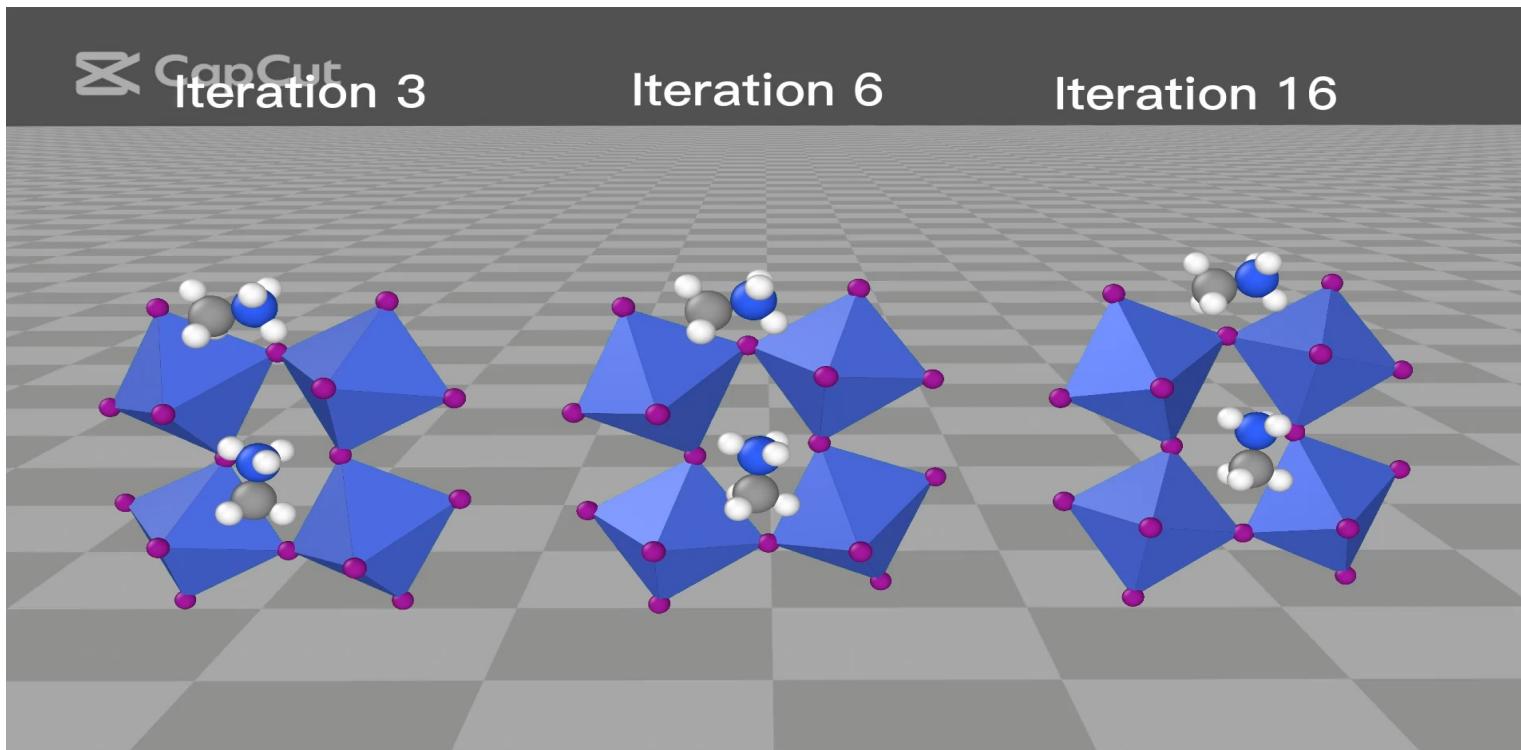
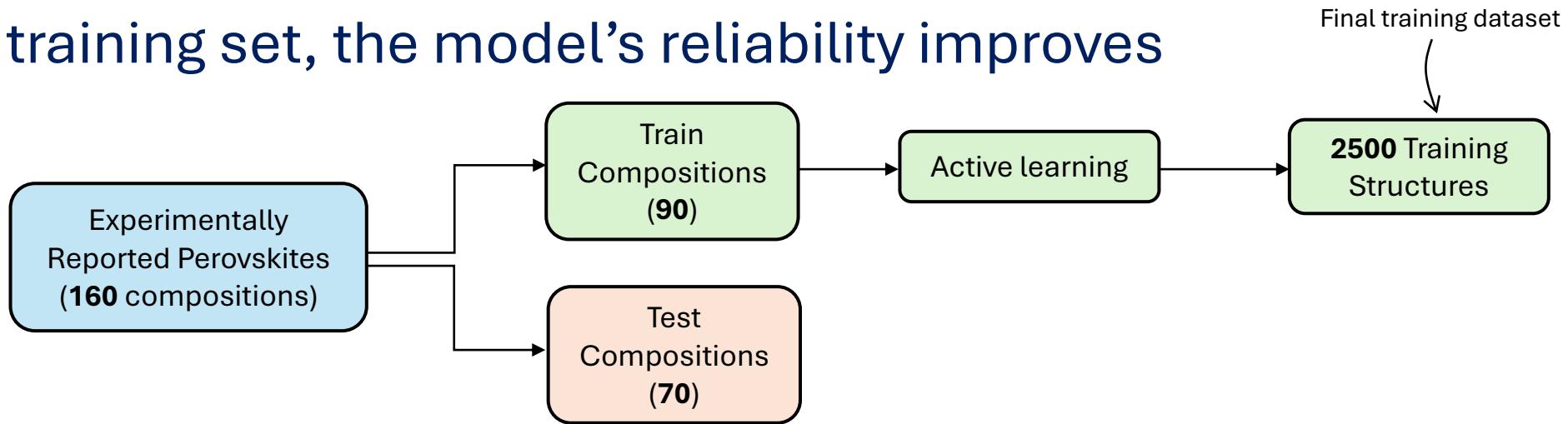
As more uncertain samples are incorporated  
into the training set, the model's reliability improves



Samples with a high uncertainty are added to the training set,  
the model was then re-trained with these new samples

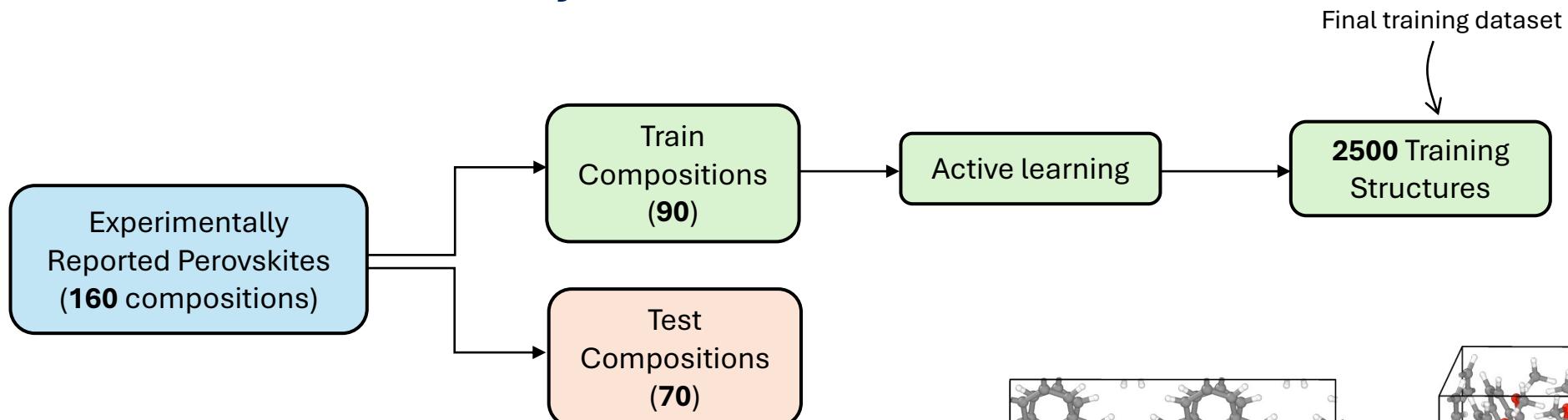


As more uncertain samples are incorporated into the training set, the model's reliability improves

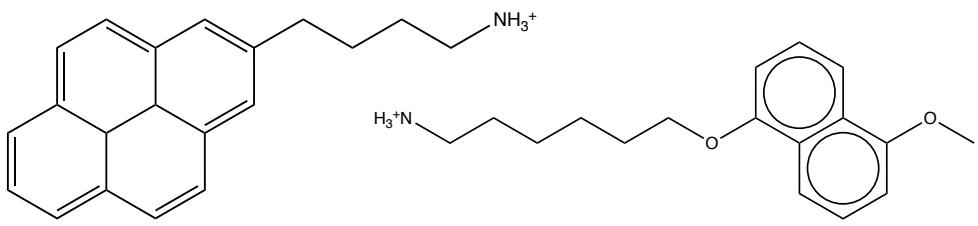
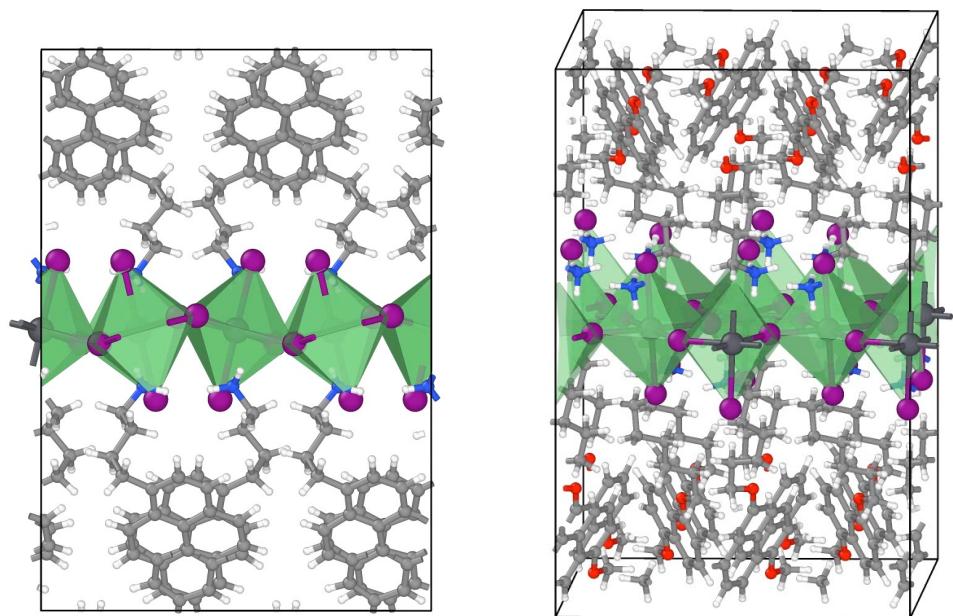
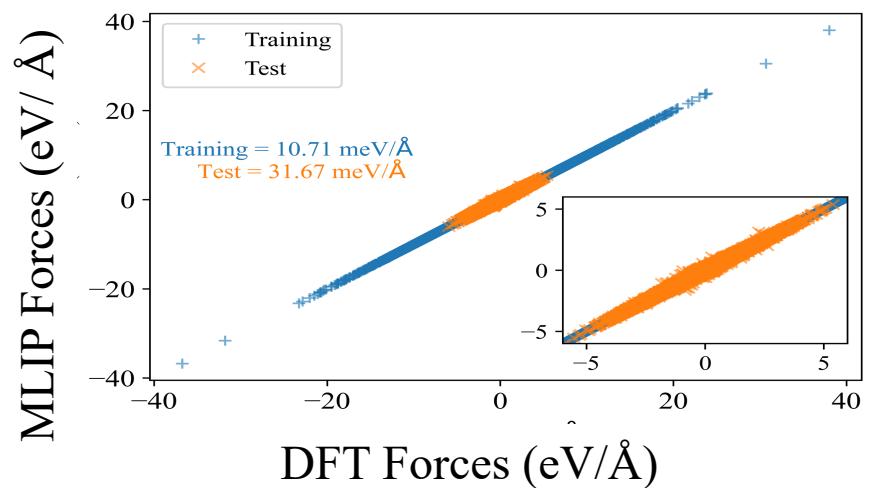
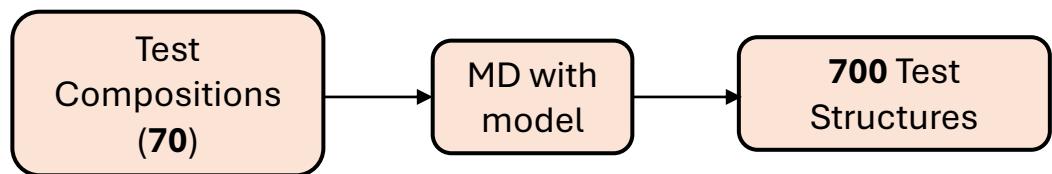


Video credit: Will Baldwin

# Confirmation of the accuracy of the MLFF

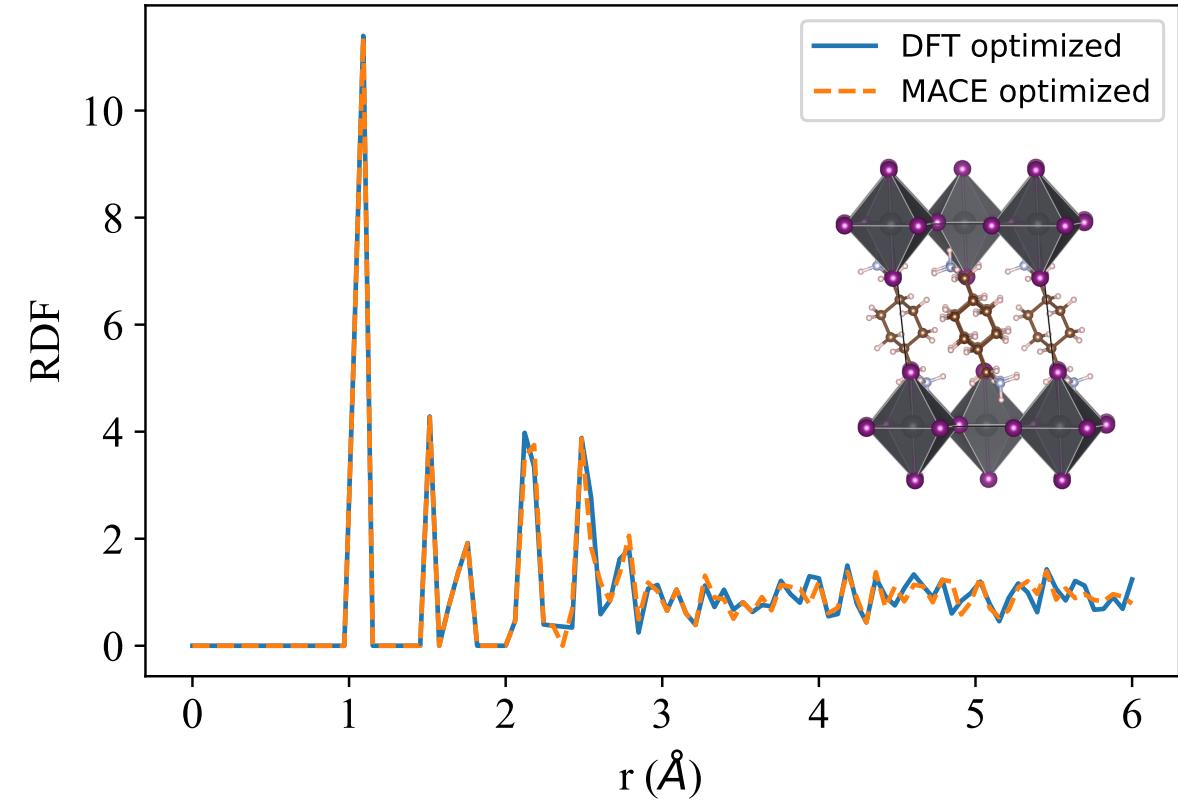
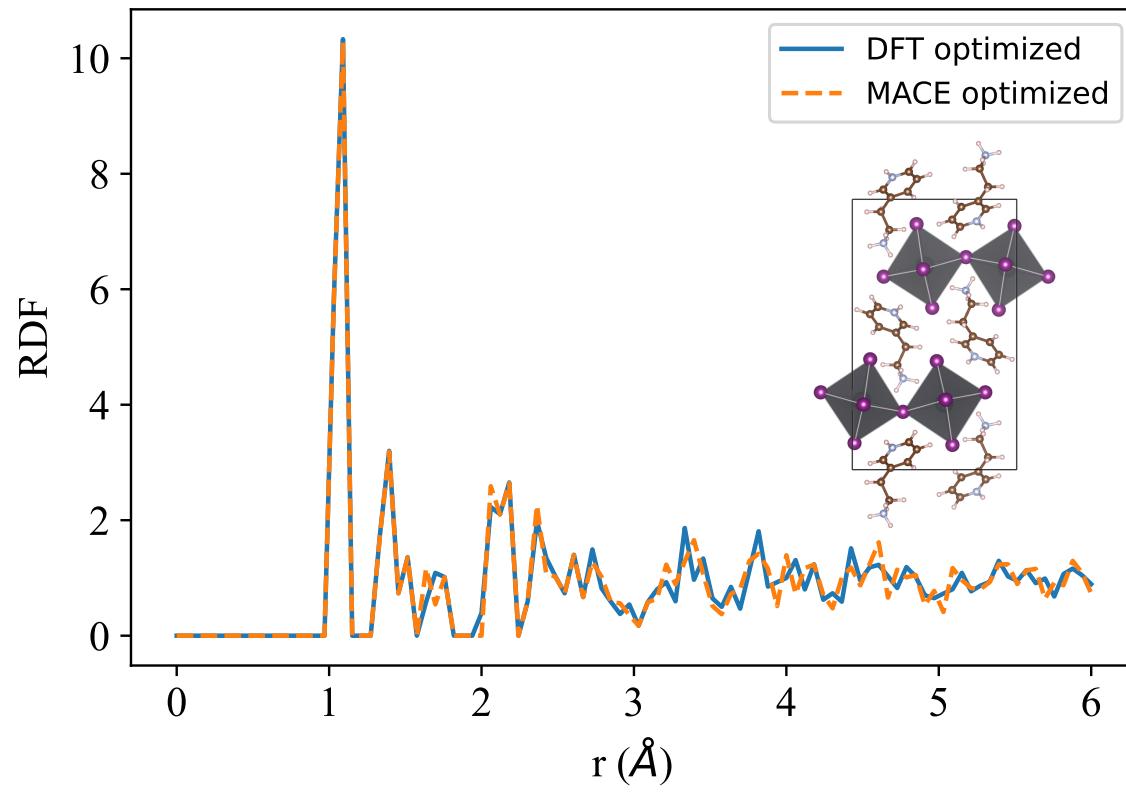


## Model Performance/Validation



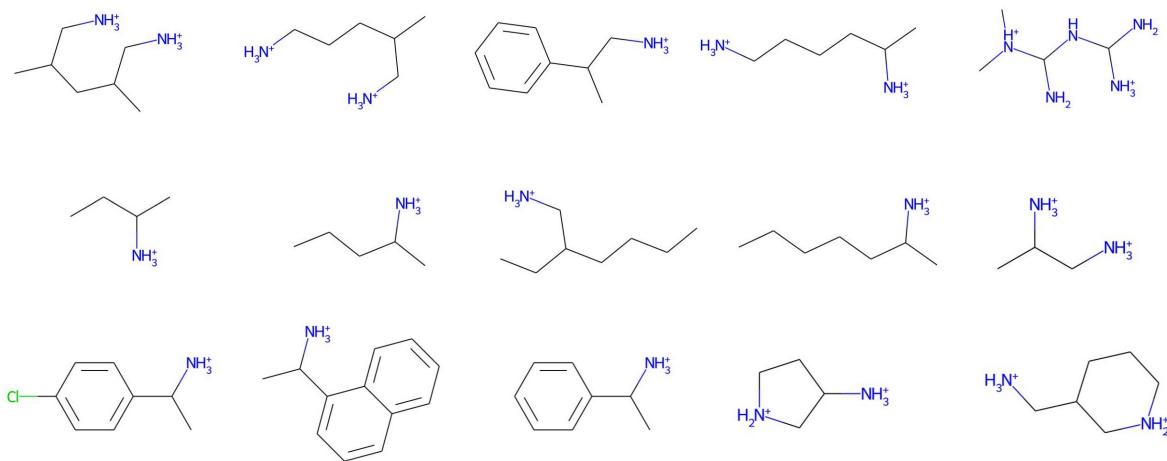
# MLIP-structure optimization recovers the DFT-optimized structure

Comparison of independently optimized using DFT and MLIP

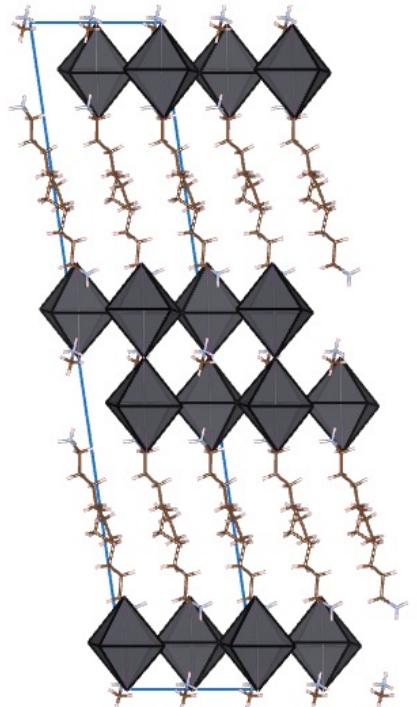
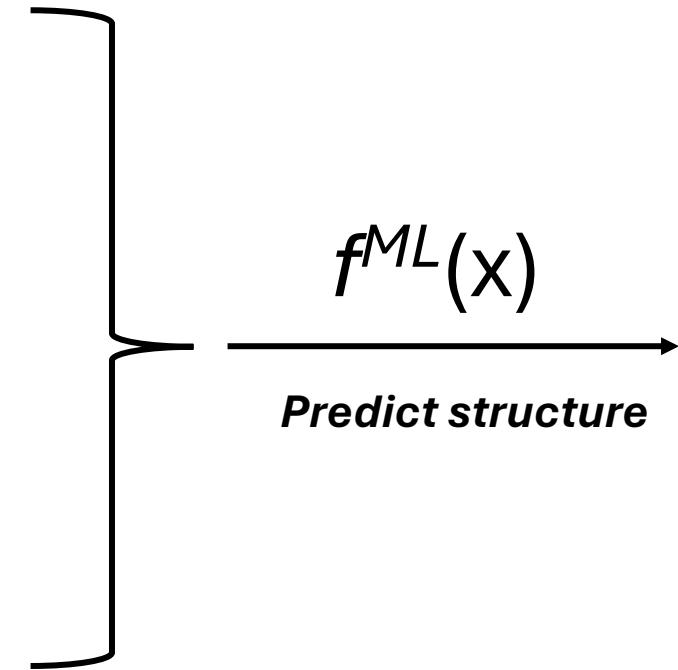


# Perform structure prediction of totally new compounds with our ML potential

## 1. Select organic cation

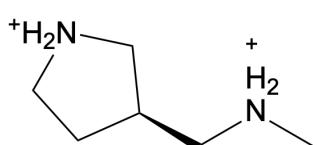


## 2. Select metal cation and halide combination



# Overview of structure prediction approach

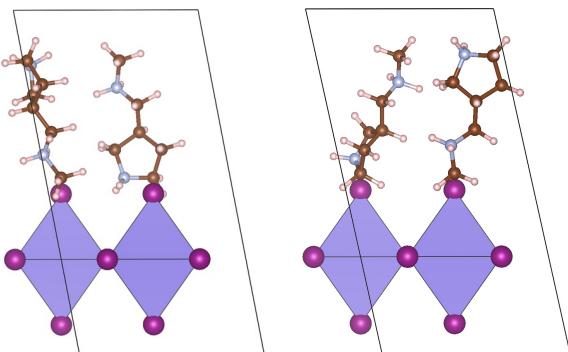
Input organic and  
inorganic formula



$\text{PbI}_4$

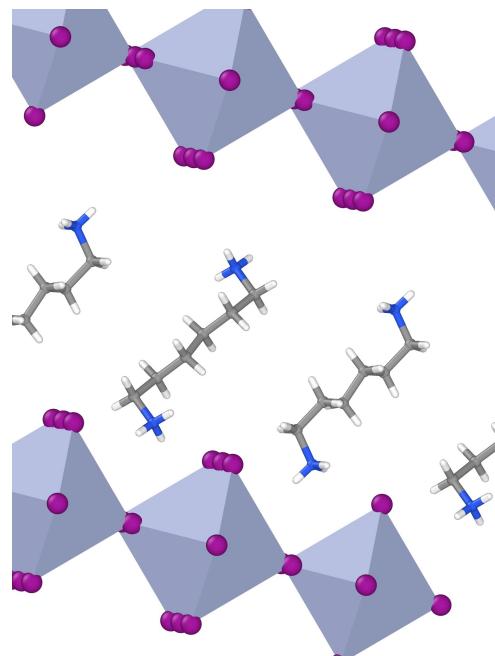
[unit cell size]

Generate random  
candidate structures



+ ...

Relax all with the MLIP

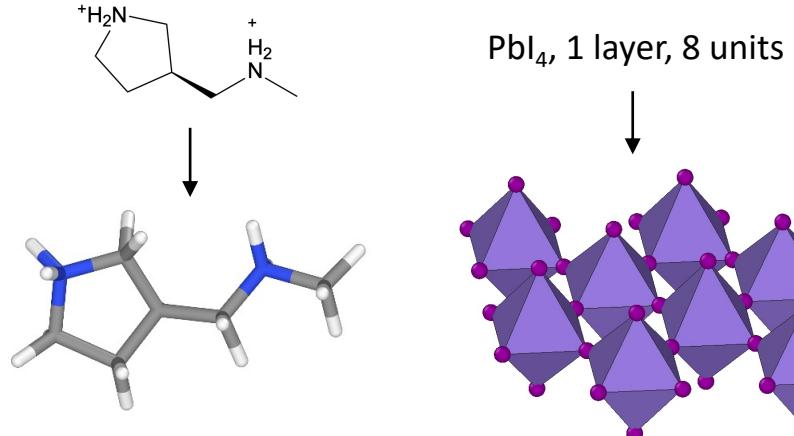


+ ...

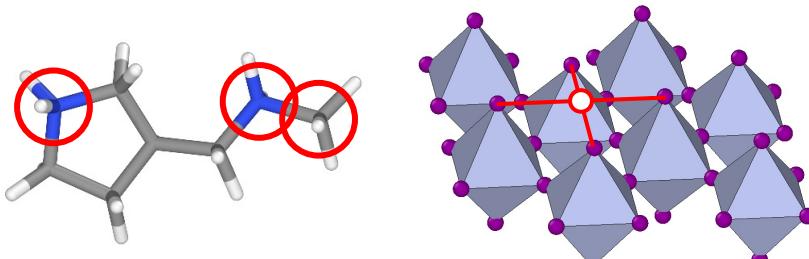
Lowest energy  
structure is best  
guess for the  
**real structure**

# Overview of procedure for generating random structures

1. Create 3D geometries for the cation and monolayer

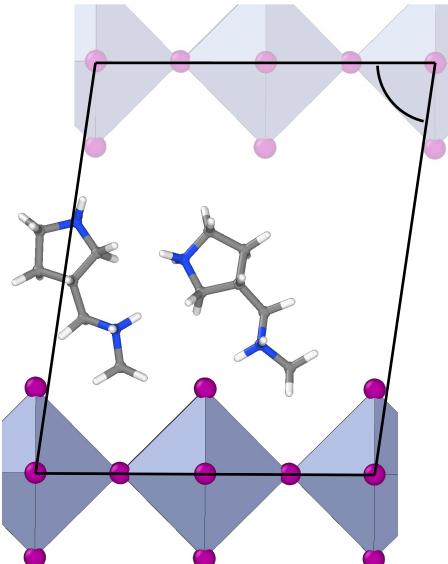
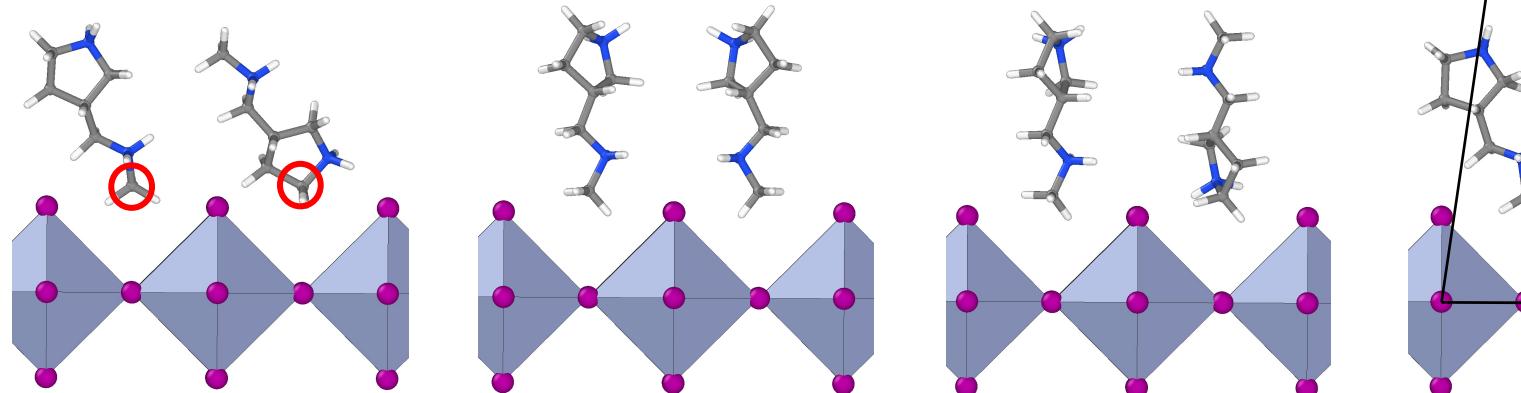


2. Identify Reference points on the Cation and monolayer

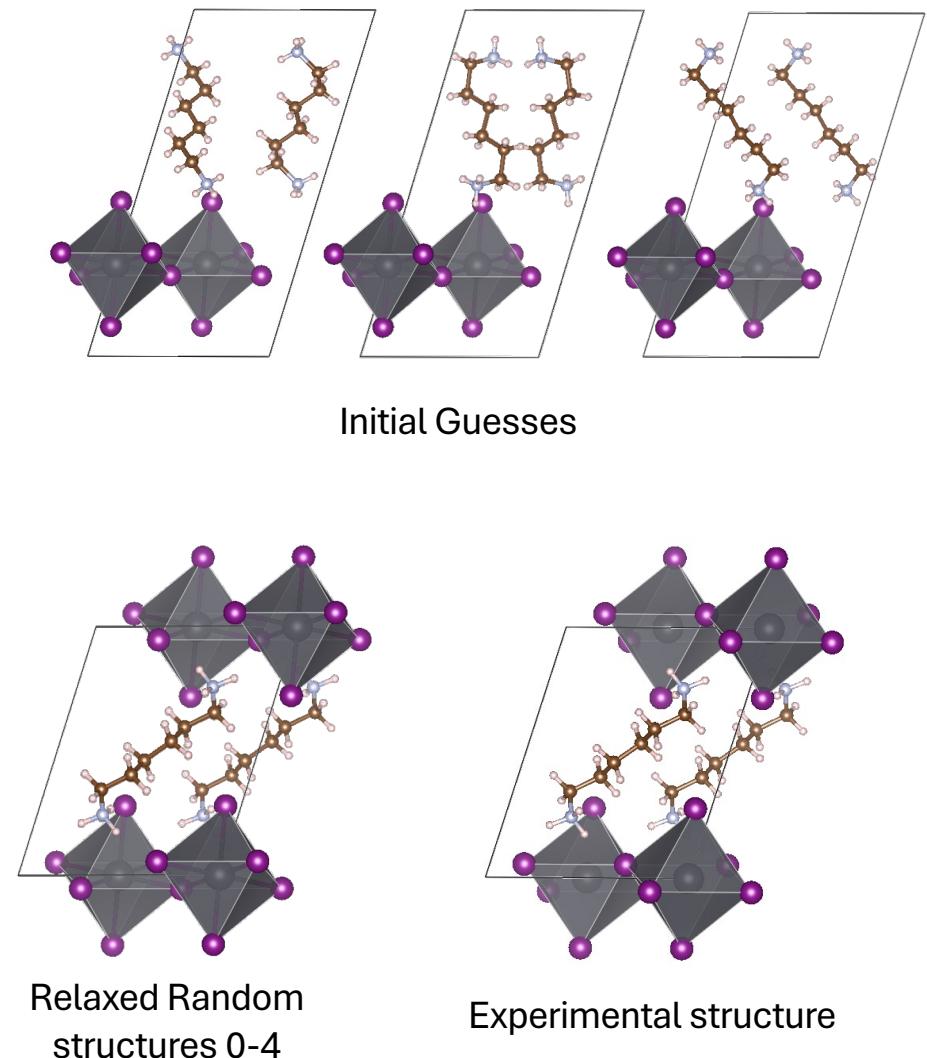
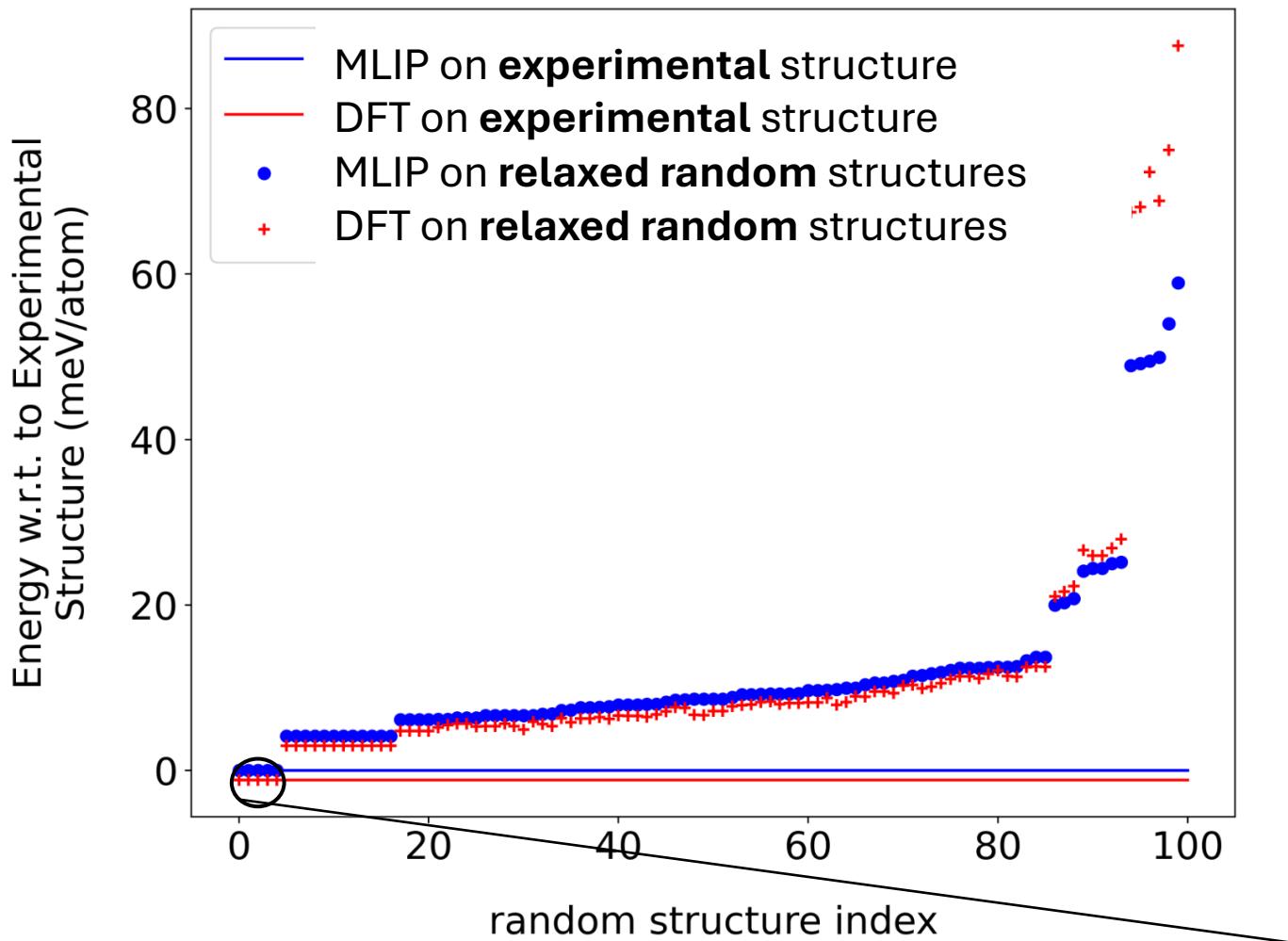


3. Generate Candidate starting structures

- Match up reference points on the two geometries
- Sample over reflections, rotations, cell shapes

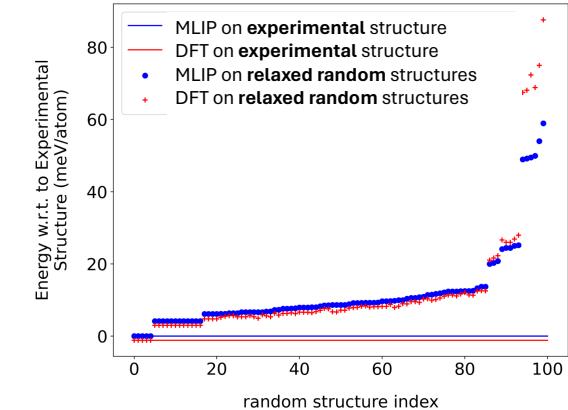
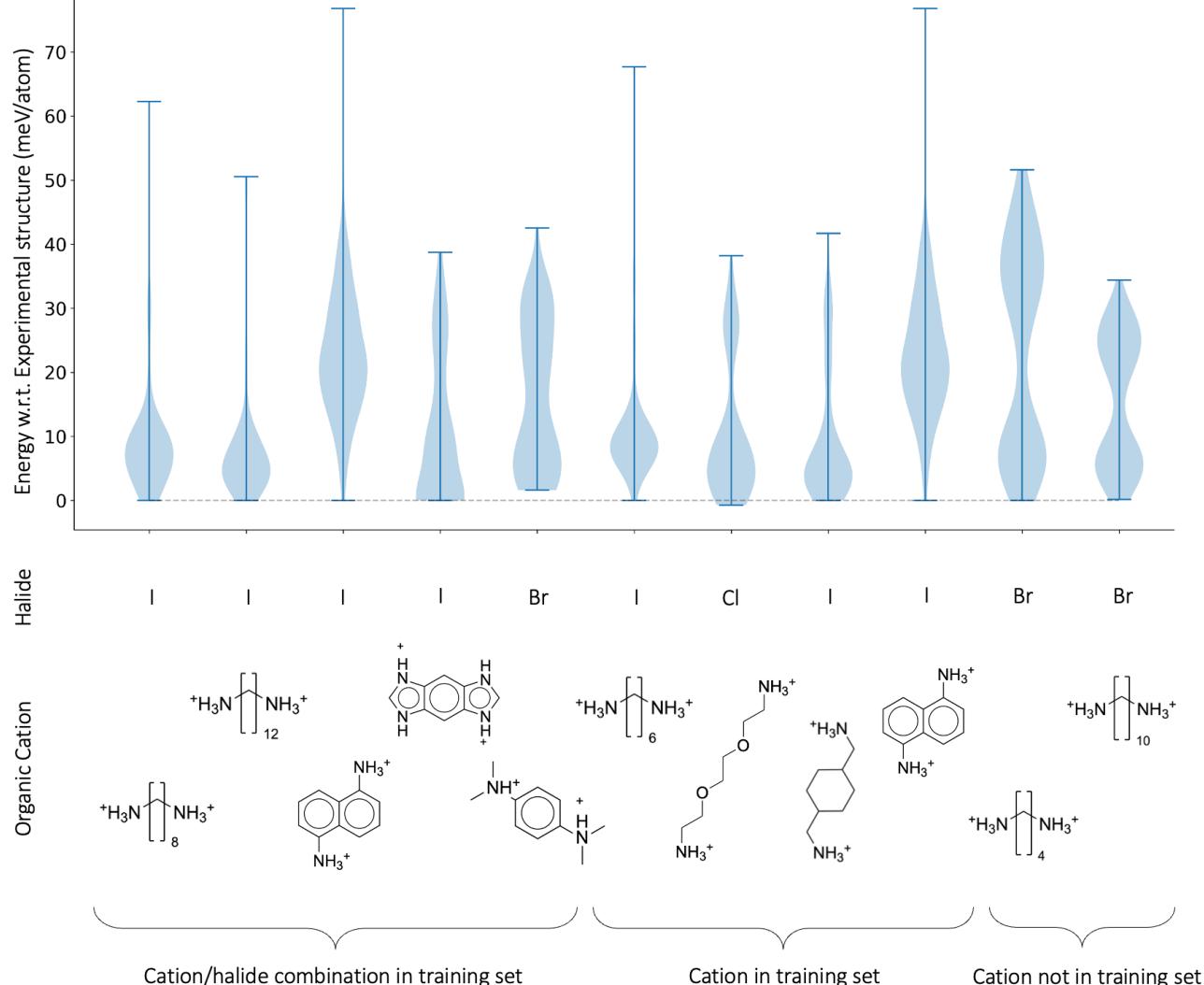


# The MLIP and structure searching approach re-discovers the correct structure of an experimentally known system



# Independent prediction of the ground-state structure using our approach is performed for 13 samples

- In all but one case we find the experimental structure
- in 2 cases we **also** find a lower energy structure, according to DFT

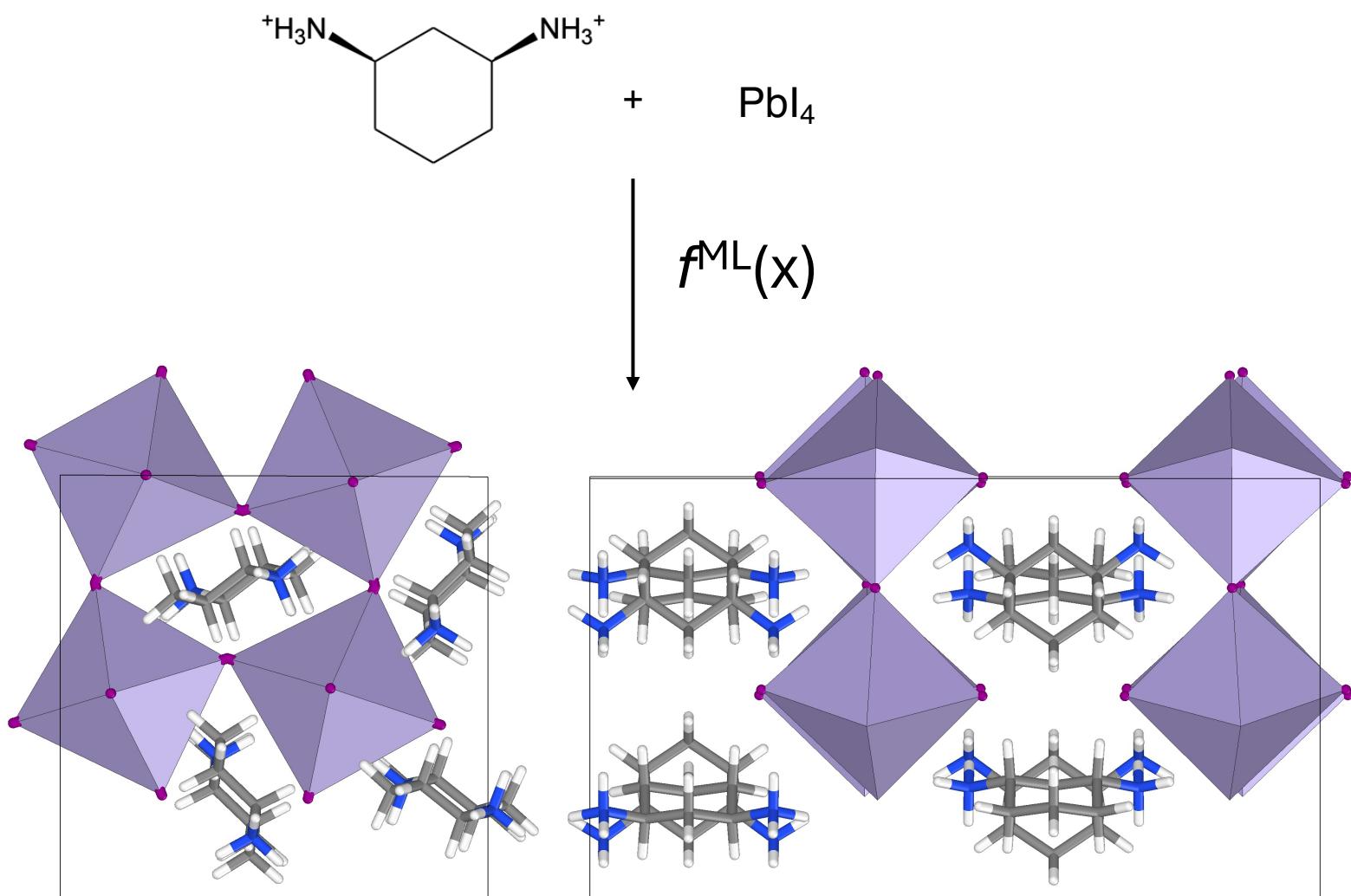


# Predicting (and synthesizing) a new 2D HOIP

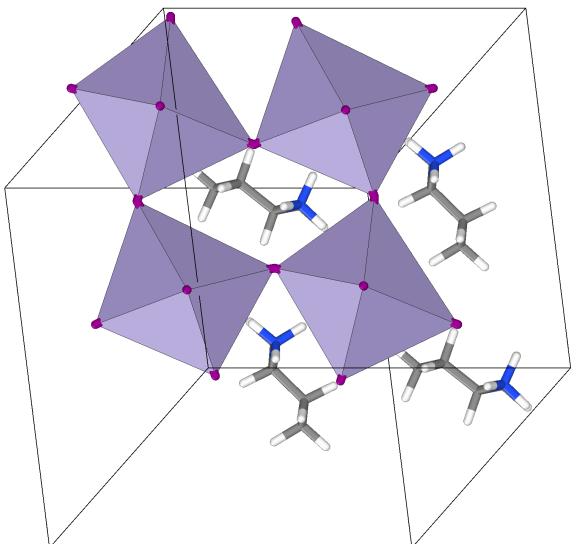
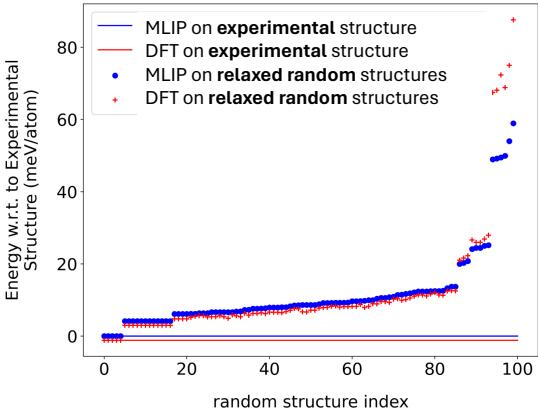
- Prediction of an organic cations with no previously reported perovskite
- 6000 samples
- 8 molecules per unit cell, 232 atoms
- **The same structure was obtained by synthesis**

Synthesis and experimental characterization by

- Evan Muller
- Joshua Kennedy

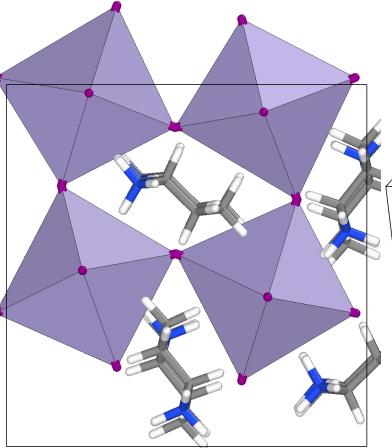


# The MLIP/structure prediction method also identifies the landscape of low energy structures



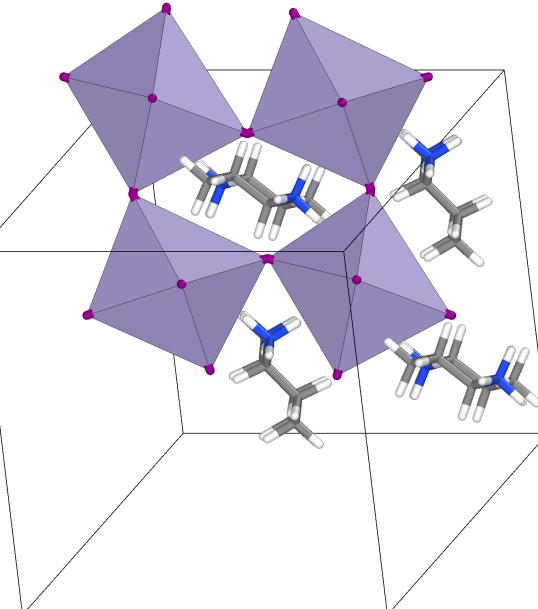
Minimum +1

$\Delta E = 0.13 \text{ meV/atom}$



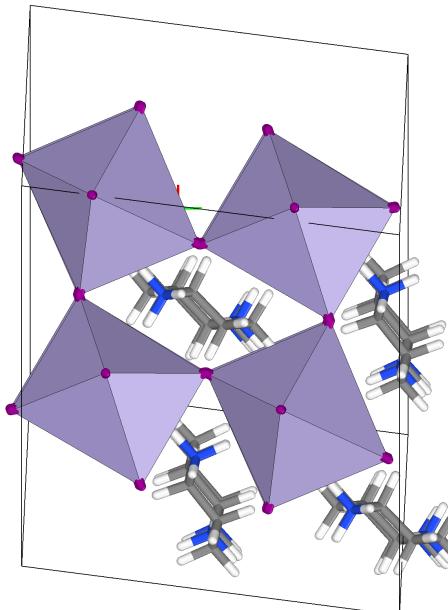
Minimum +2

$\Delta E = 0.26 \text{ meV/atom}$



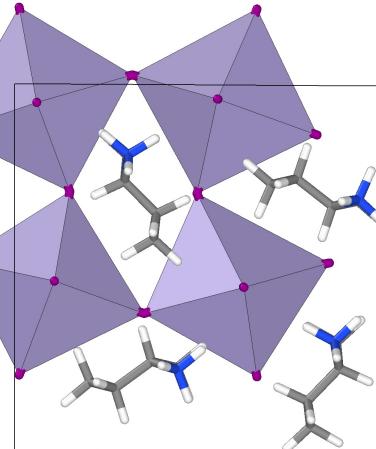
Minimum +3

$\Delta E = 0.29 \text{ meV/atom}$



Minimum +4

$\Delta E = 0.30 \text{ meV/atom}$



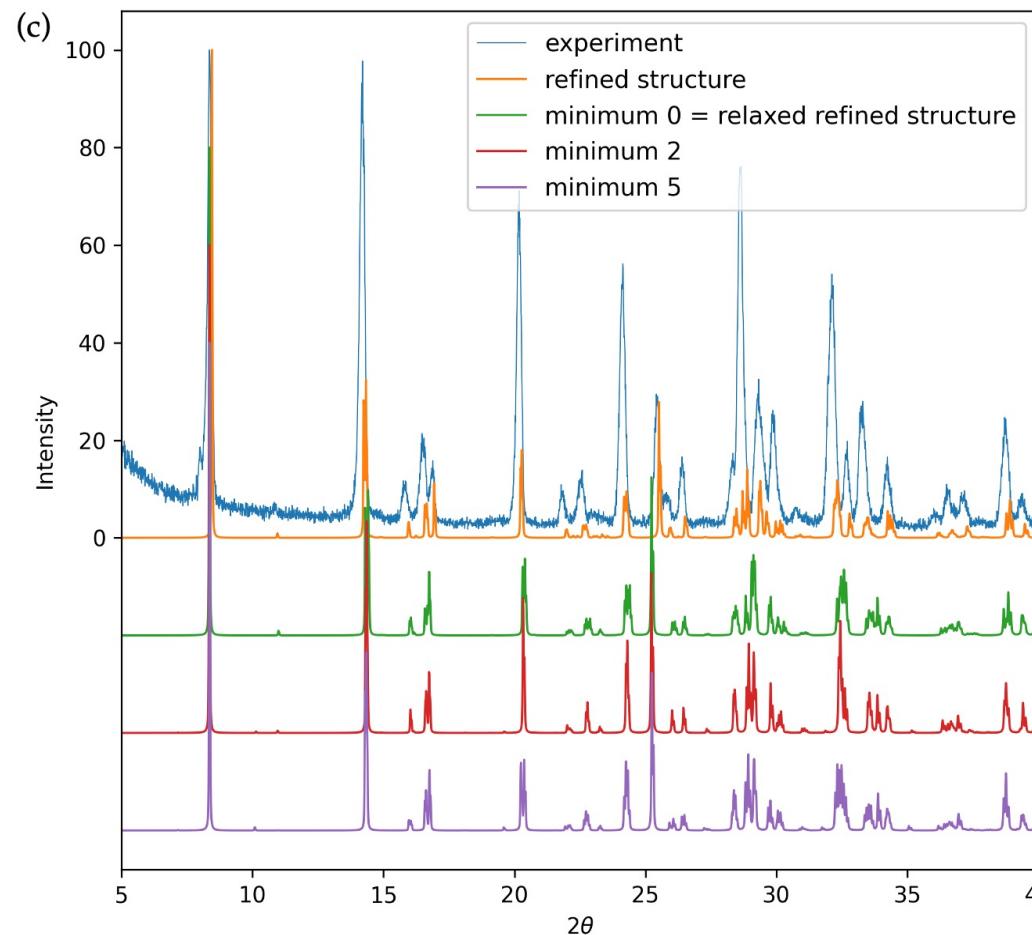
Minimum +5

$\Delta E = 0.50 \text{ meV/atom}$

# Other MLIP-predicted structures look very similar based on XRD

Could these other phases be stabilized through modifications of the synthesis?

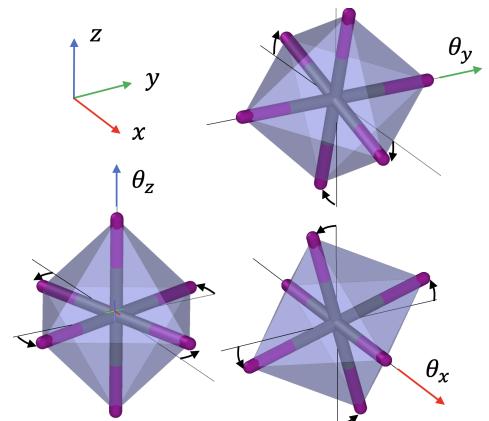
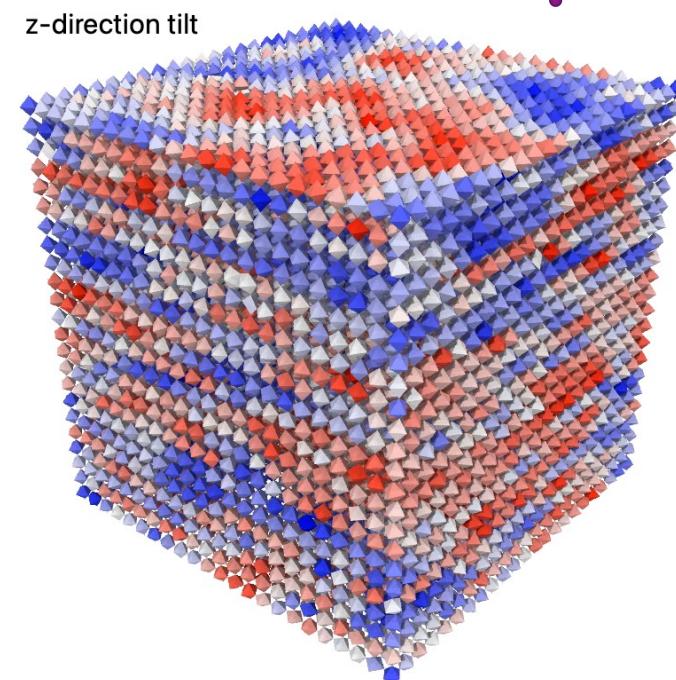
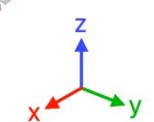
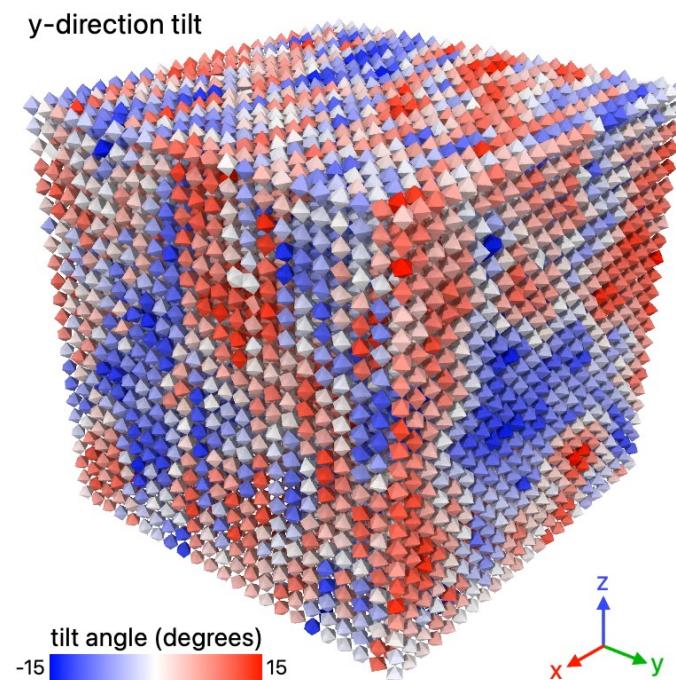
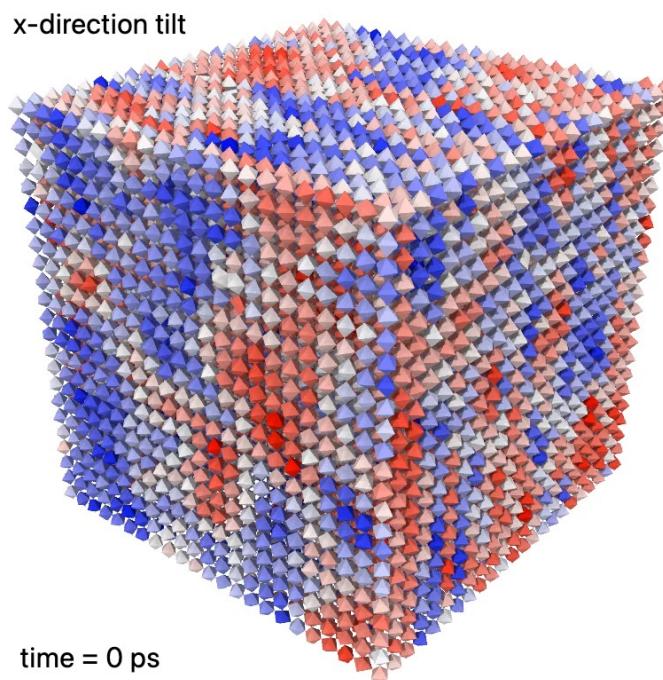
SFFP: My student Jack Barber is working with Josh Kennedy et al. in modifying the synthesis to try to find different phases



# Modeling local distortions in halide perovskites

$\text{CsPbI}_3$  T = 550K (cubic phase)

MLIP/MD simulation of on a 69,120 atom simulation box, equilibrated for 1 ns



# Local symmetry breaking drives picosecond spin domain formation in polycrystalline halide perovskite films

Received: 4 August 2022

Accepted: 6 April 2023

Published online: 12 June 2023

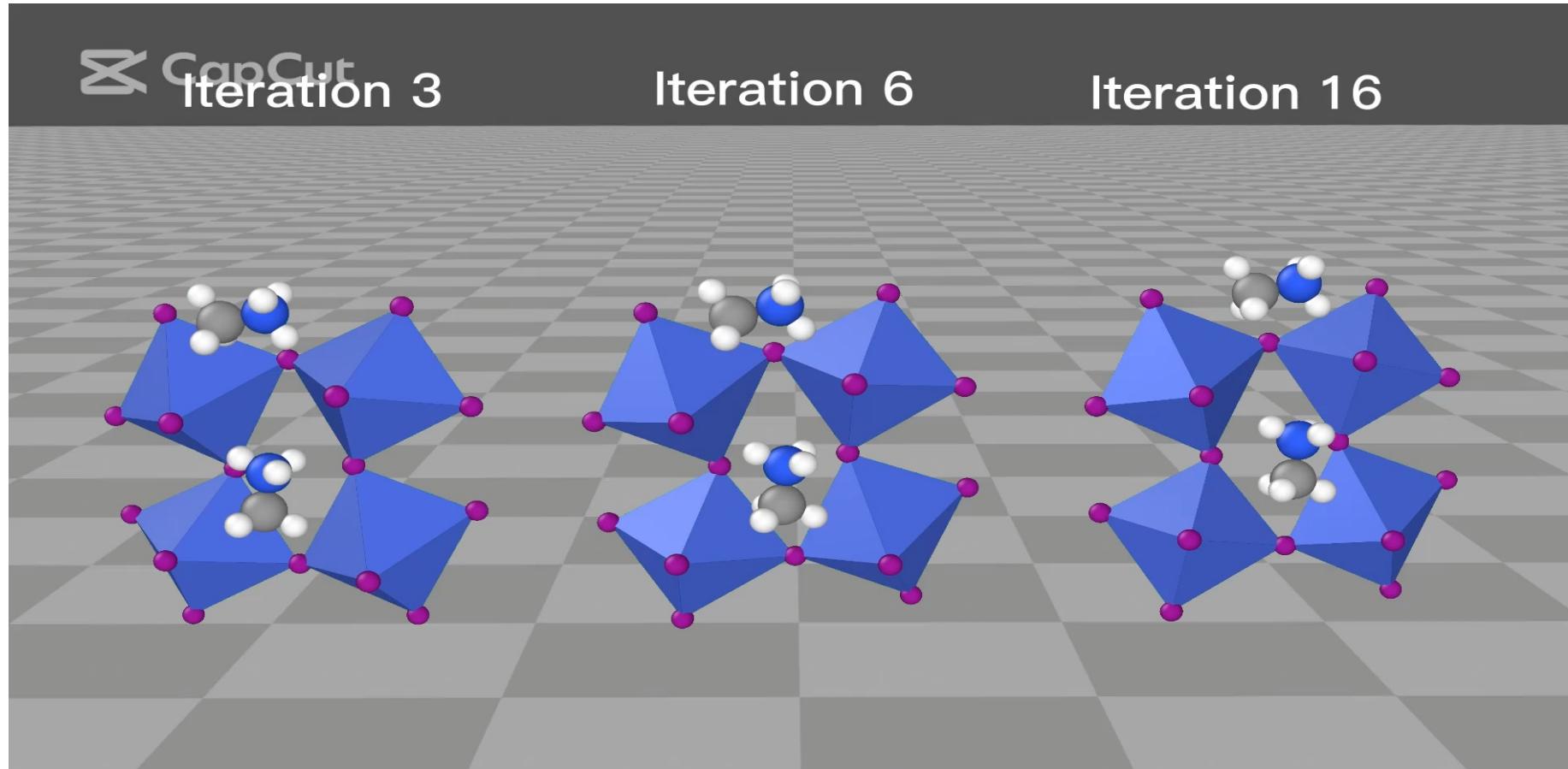


Arjun Ashoka<sup>1</sup>, Satyawan Nagane<sup>②</sup>, Nives Strkalj<sup>③</sup>, Ashish Sharma<sup>1</sup>, Bart Roose<sup>②</sup>, Alexander J. Sneyd<sup>①</sup>, Jooyoung Sung<sup>④</sup>, Judith L. MacManus-Driscoll<sup>③</sup>, Samuel D. Stranks<sup>②</sup>, Sascha Feldmann<sup>⑤</sup>✉ & Akshay Rao<sup>①</sup>✉

Photoinduced spin–charge interconversion in semiconductors with spin–orbit coupling could provide a route to optically addressable spintronics without the use of external magnetic fields. However, in structurally disordered polycrystalline semiconductors, which are being widely explored for device applications, the presence and role of spin-associated charge currents remains unclear. Here, using femtosecond circular-polarization-resolved pump–probe microscopy on polycrystalline halide perovskite thin films, we observe the photoinduced ultrafast formation of spin domains on the micrometre scale formed through lateral spin currents. Micrometre-scale variations in the intensity of optical second-harmonic generation and vertical piezoresponse suggest that the spin-domain formation is driven by the presence of strong local inversion symmetry breaking via structural disorder. We propose that this leads to spatially varying Rashba-like spin textures that drive spin-momentum-locked currents, leading to local spin accumulation. Ultrafast spin-domain formation in polycrystalline halide perovskite films provides an optically addressable platform for nanoscale spin-device physics.

# How can we better train MLIPs without starting from scratch each time?

Recall that early in the training, ML potentials are often terribly unreliable



# The MLIP for halide perovskites was fine-tuned from the recently trained MACE-MP0 model

Cornell University

We gratefully acknowledge support from the Simons Foundation, member institutions, and all contributors. [Donate](#)

arXiv > physics > arXiv:2401.00096

Search... All fields Help | Advanced Search

Physics > Chemical Physics

[Submitted on 29 Dec 2023 (v1), last revised 1 Mar 2024 (this version, v2)]

## A foundation model for atomistic materials chemistry

Ilyes Batatia, Philipp Benner, Yuan Chiang, Alin M. Elena, Dávid P. Kovács, Janosh Riebesell, Xavier R. Advincula, Mark Asta, Matthew Avaylon, William J. Baldwin, Fabian Berger, Noam Bernstein, Arghya Bhownik, Samuel M. Blau, Vlad Cărare, James P. Darby, Sandip De, Flaviano Della Pia, Volker L. Deringer, Rokas Elijošius, Zakariya El-Machachi, Fabio Falcioni, Edvin Fako, Andrea C. Ferrari, Annalena Genreith-Schriever, Janine George, Rhys E. A. Goodall, Clare P. Grey, Petr Grigorev, Shuang Han, Will Handley, Hendrik H. Heenen, Kersti Hermansson, Christian Holm, Jad Jaafar, Stephan Hofmann, Konstantin S. Jakob, Hyunwook Jung, Venkat Kapil, Aaron D. Kaplan, Nima Karimitari, James R. Kermode, Namu Kroupa, Jolla Kullgren, Matthew C. Kuner, Domantas Kuryla, Guoda Liepuoniute, Johannes T. Margraf, Ioan-Bogdan Magdău, Angelos Michaelides, J. Harry Moore, Aakash A. Naik, Samuel P. Niblett, Sam Walton Norwood, Niamh O'Neill, Christoph Ortner, Kristin A. Persson, Karsten Reuter, Andrew S. Rosen, Lars L. Schaaf, Christoph Schran, Benjamin X. Shi, Eric Sivonxay, Tamás K. Stenczel, Viktor Svahn, Christopher Sutton, Thomas D. Swinburne, Jules Tilly, Cas van der Oord, Eszter Varga-Umbrich, Tejs Vegge, Martin Vondrák, Yangshuai Wang, William C. Witt, Fabian Zills, Gábor Csányi

Machine-learned force fields have transformed the atomistic modelling of materials by enabling simulations of ab initio quality on unprecedented time and length scales. However, they are currently limited by: (i) the significant computational and human effort that must go into development and validation of potentials for each particular system of interest; and (ii) a general lack of transferability from one chemical system to the next. Here, using the state-of-the-art MACE architecture we introduce a single general-purpose ML model, trained on a public database of 150k inorganic crystals, that is capable of running stable molecular dynamics on molecules and materials. We demonstrate the power of the MACE-MP0 model – and its qualitative and at times quantitative accuracy – on a diverse set problems in the physical sciences, including the properties of solids, liquids, gases, chemical reactions, interfaces and even the dynamics of a small protein. The model can be applied out of the box and as a starting or "foundation model" for any atomistic system of interest and is thus a step towards democratising the revolution of ML force fields by lowering the barriers to entry.

Comments: 119 pages, 63 figures, 37MB PDF  
Subjects: Chemical Physics (physics.chem-ph); Materials Science (cond-mat.mtrl-sci)  
Cite as: arXiv:2401.00096 [physics.chem-ph]  
(or arXiv:2401.00096v2 [physics.chem-ph] for this version)

Access Paper:

- [View PDF](#)
- [TeX Source](#)
- [Other Formats](#)

view license

Current browse context:  
**physics.chem-ph**  
< prev | next >  
new | recent | 2401

Change to browse by:  
**cond-mat**  
**cond-mat.mtrl-sci**  
**physics**

References & Citations

- [NASA ADS](#)
- [Google Scholar](#)
- [Semantic Scholar](#)

[1 blog link](#) (what is this?)

[Export BibTeX Citation](#)

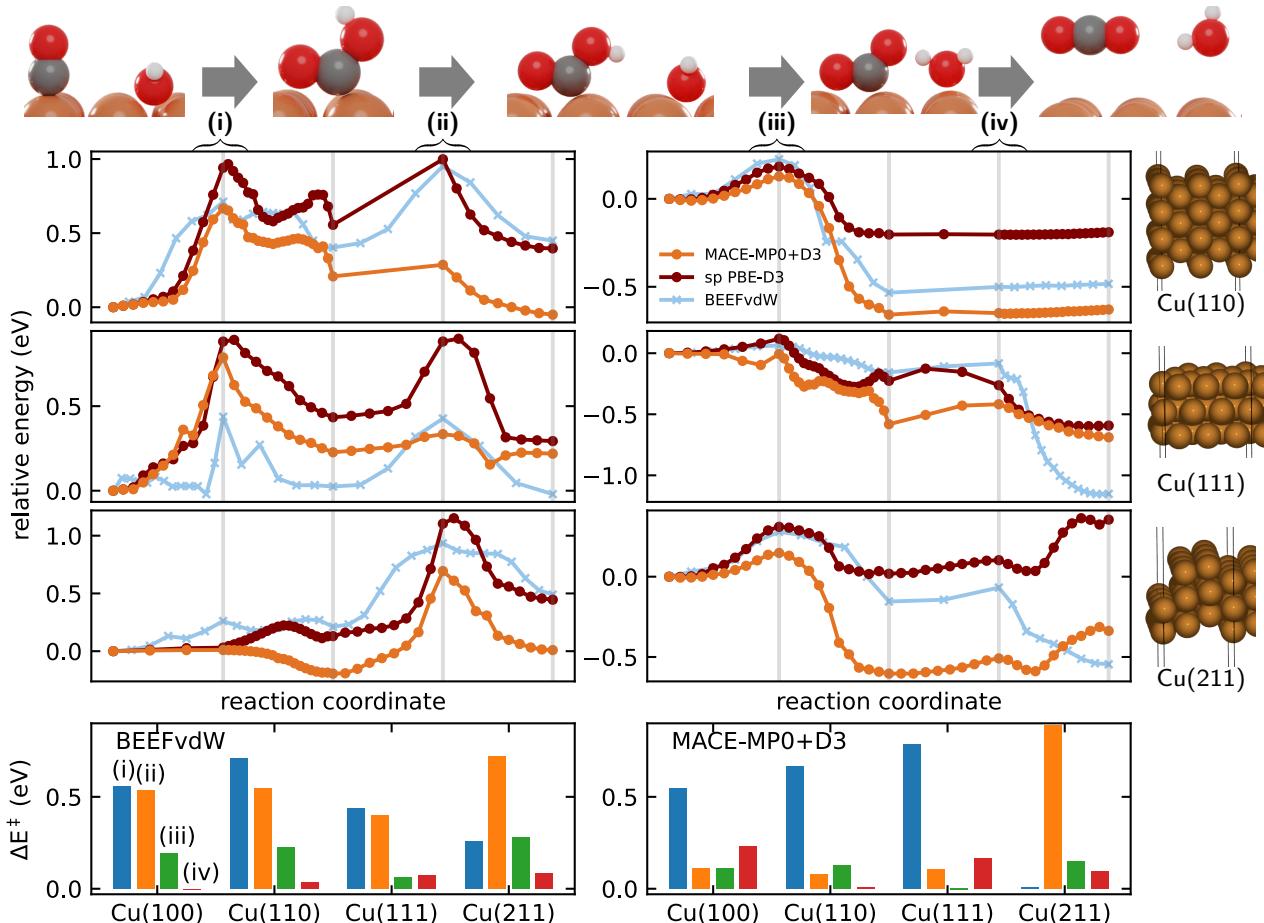
Bookmark

Ilyes Batatia, Gábor Csányi and co-workers “A foundation model for atomistic materials chemistry”

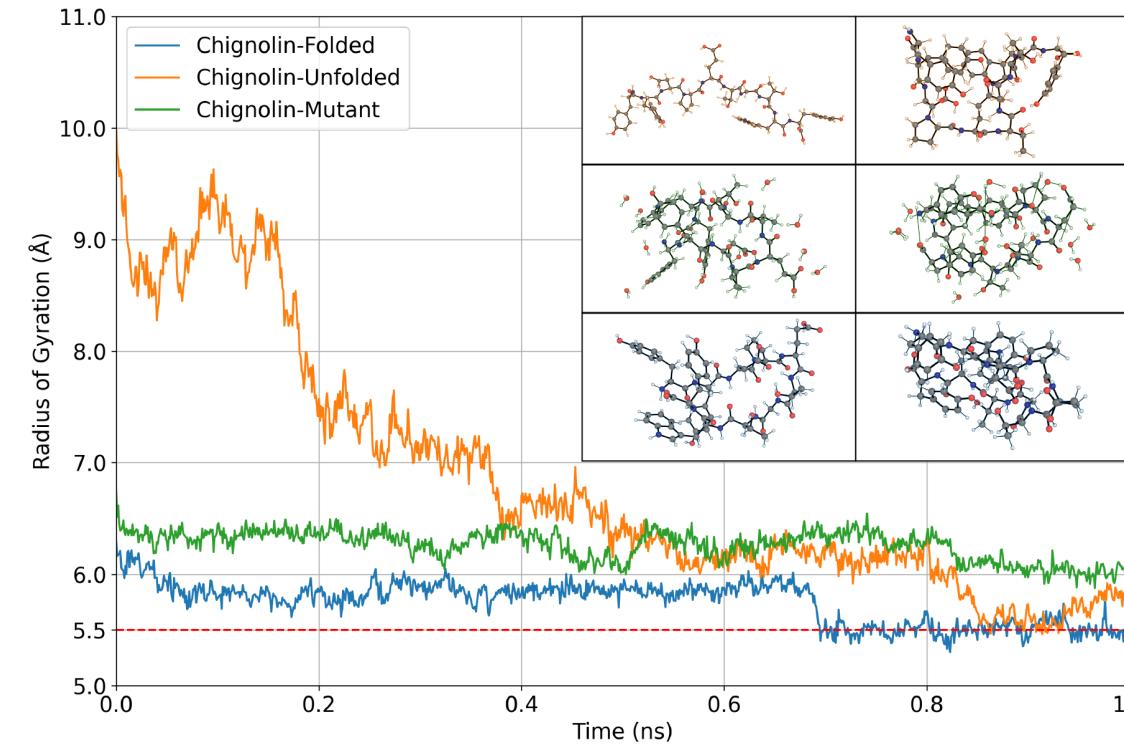
# The MACE-MP0 model can generalize to tasks well outside of the training set

MACE architecture trained on 150,000 geometry relaxations from the Materials Project

Co oxidation on Cu

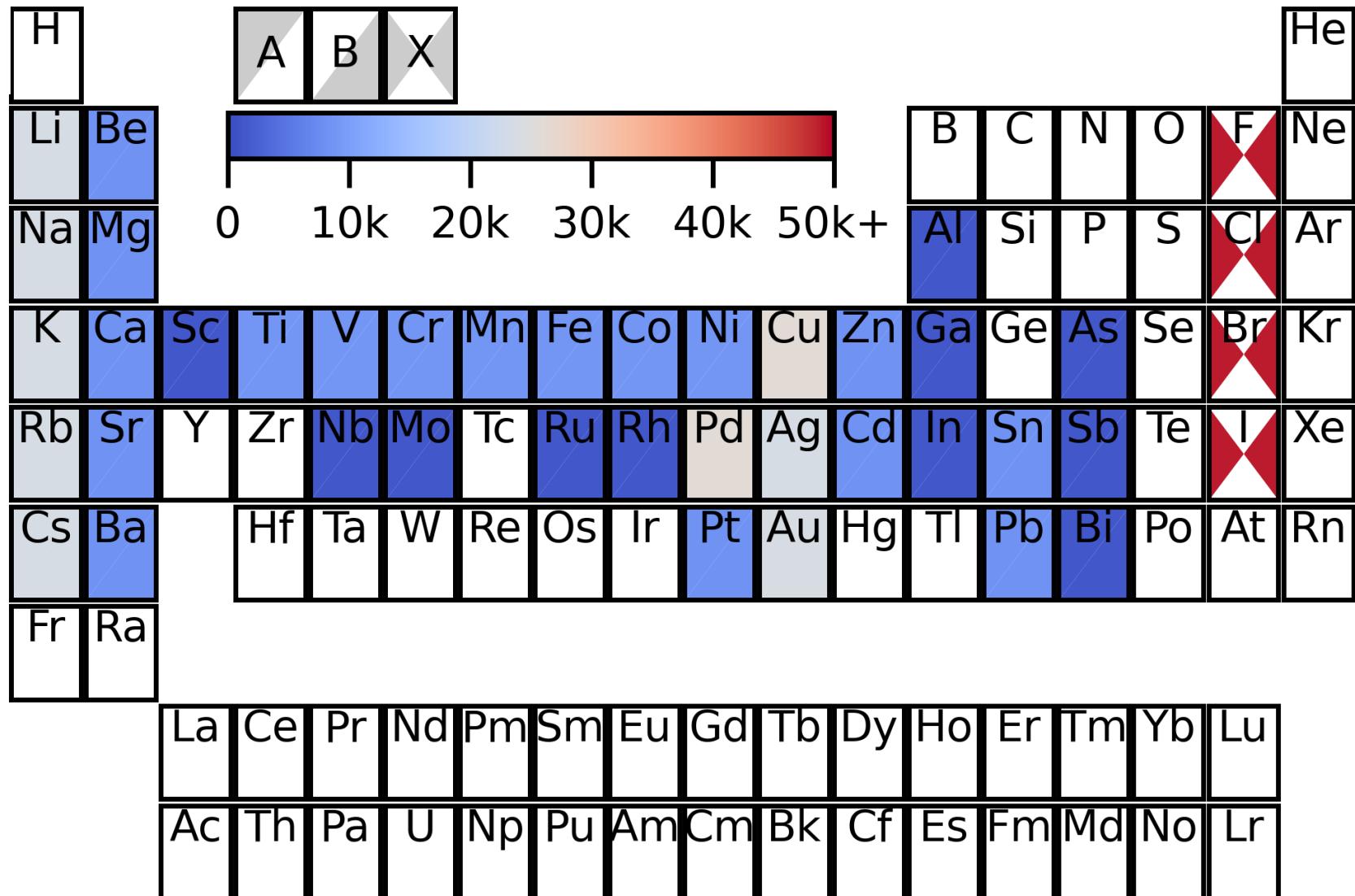
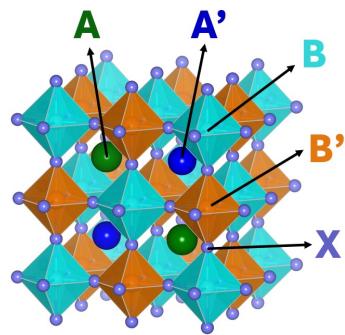


Folding of 10 residue Chignolin



## Goal: train an MLIP to screen all AA'BB'X<sub>6</sub> compounds

## 195,968 AA'BB'X6 compounds based on charge balancing constraints



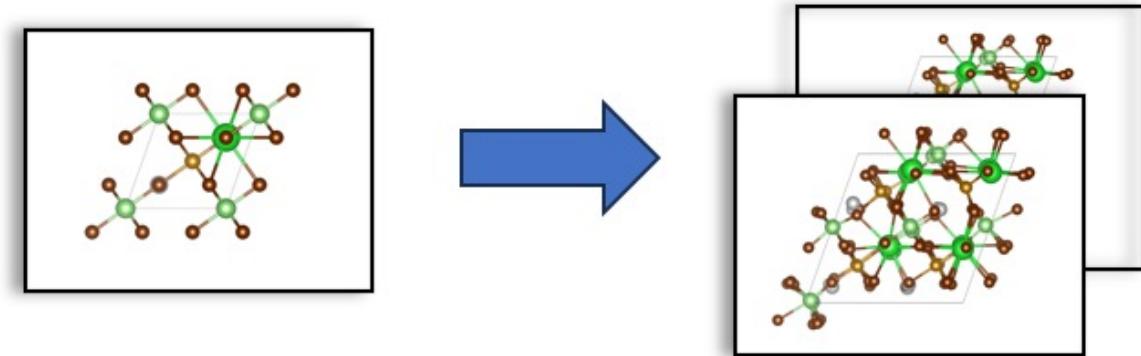
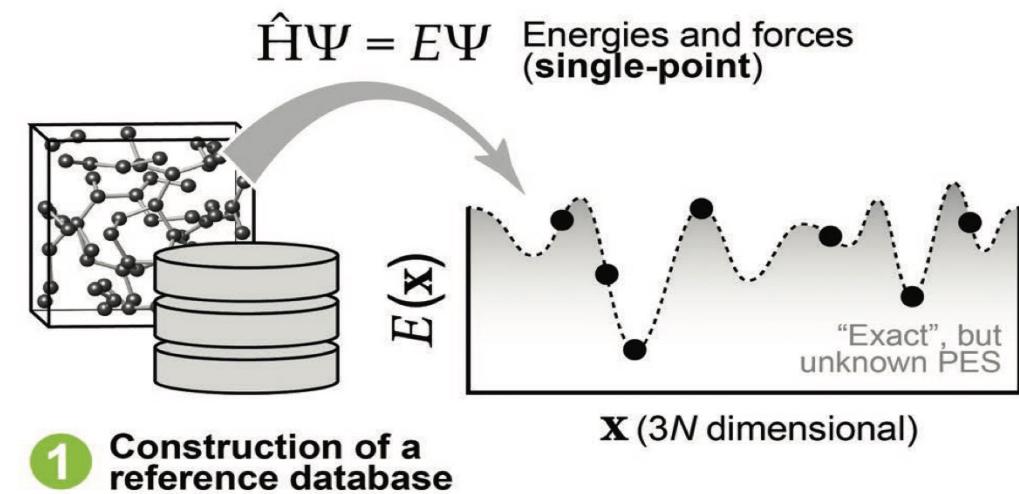
# Training set construction

Initialize training from the MACE-MP0 model

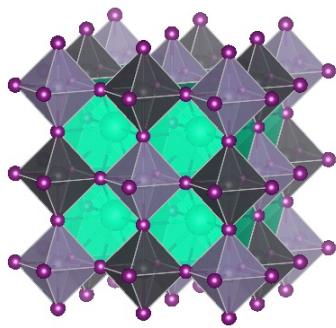
Randomly selected 340 AA'BB'X<sub>6</sub> compositions  
out of ~200K

For each composition, randomly perturbed all  
atoms in an 80-atom supercell

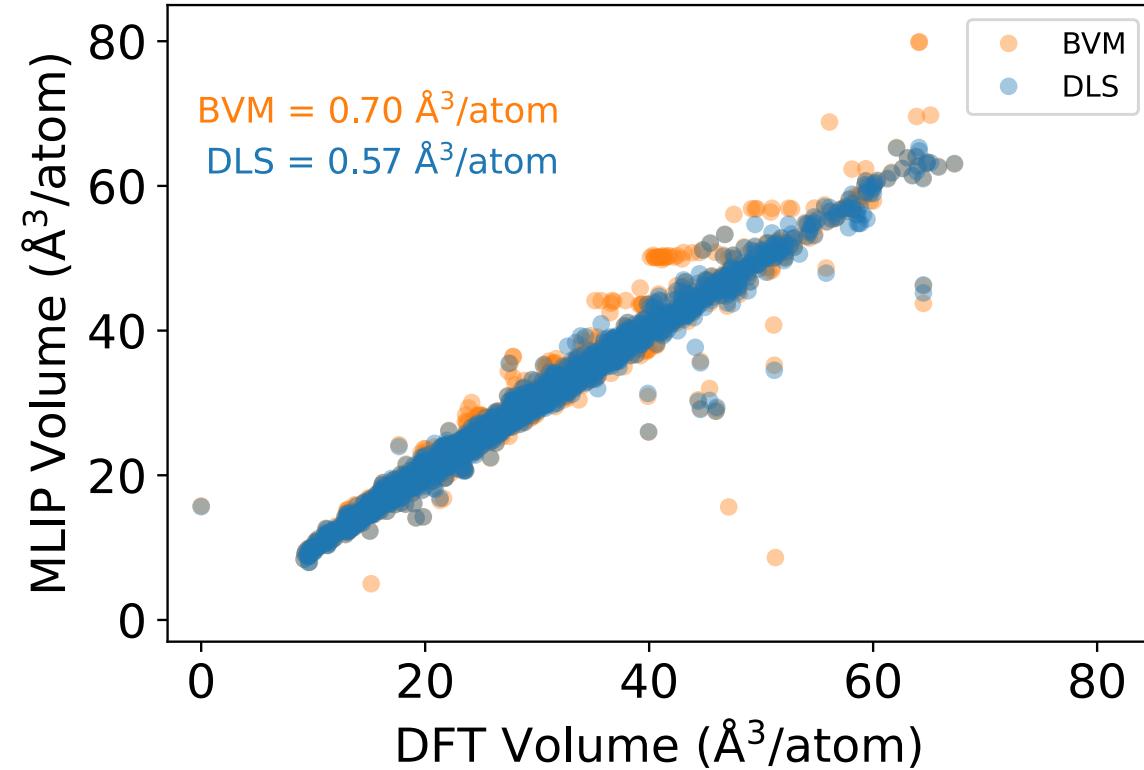
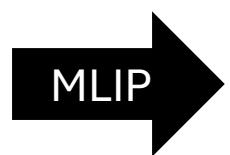
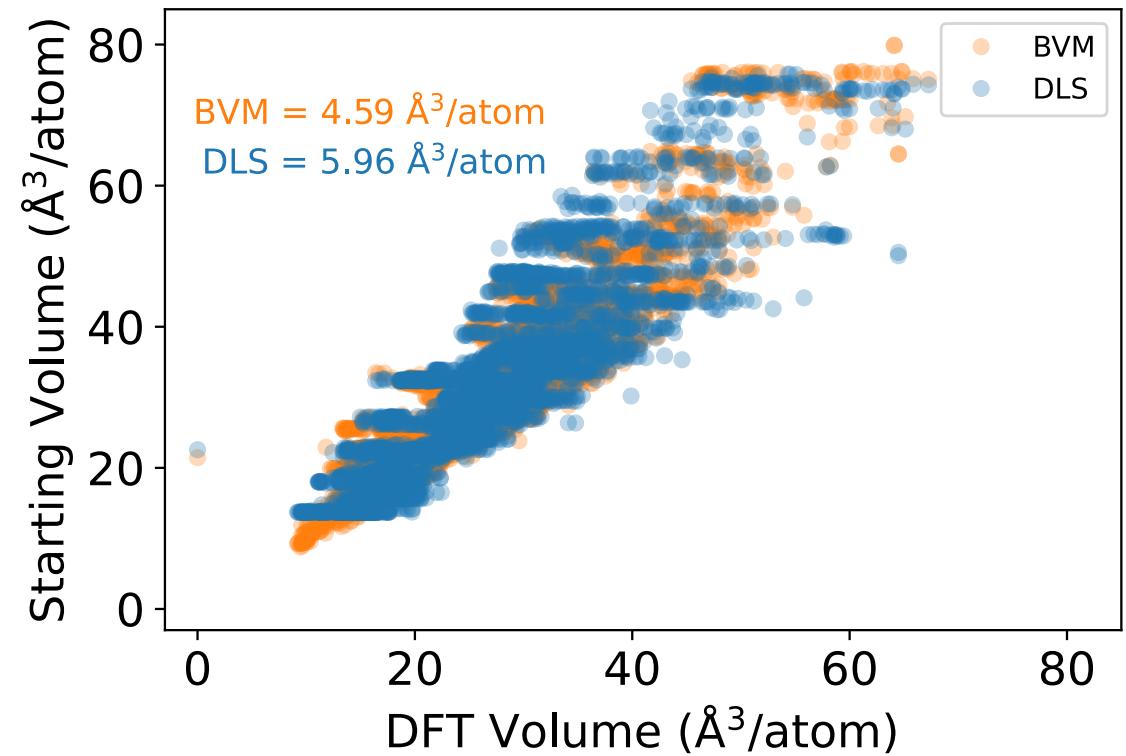
Total training set size: 2770 samples



# MLIP-optimized BVM / DLS starting structures compared well to the DFT optimized structures



MLIP-optimized 4400 unseen compositions

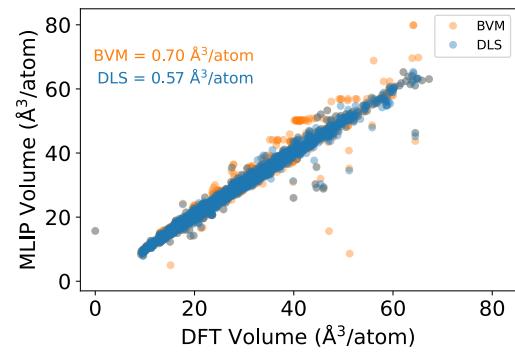
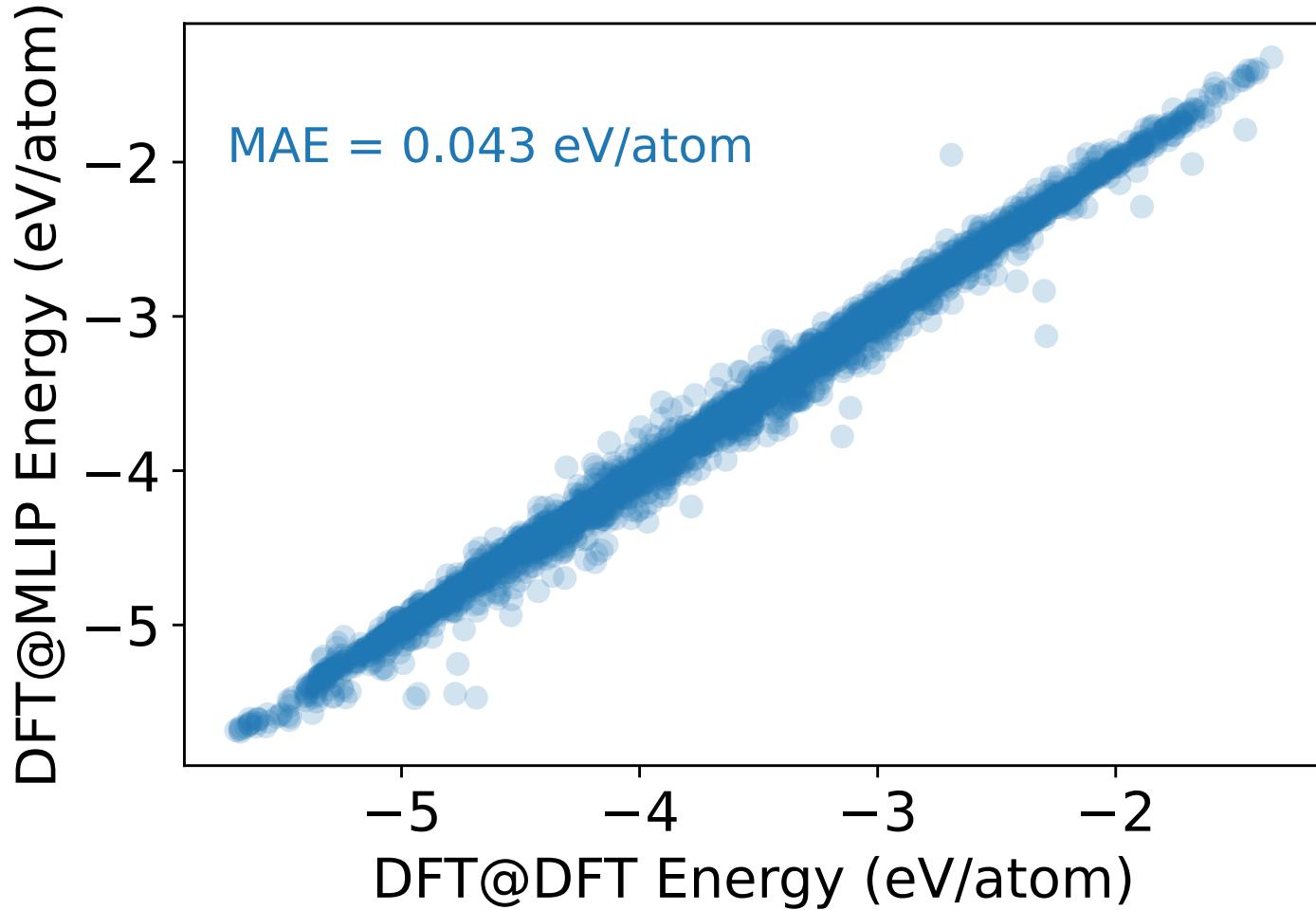


BVM: RJ Morelock et al., J. Chem. Theory Comput. 18, 3257-3267 (2022)

DLS: G Hautier, et al. J. Am. Chem. Soc. 50, 656–663 (2011)

# DFT-single point of the MLIP-optimized structure indicates an MAE = 40 meV/atom

1 kcal/mol = 43.4 meV/atom



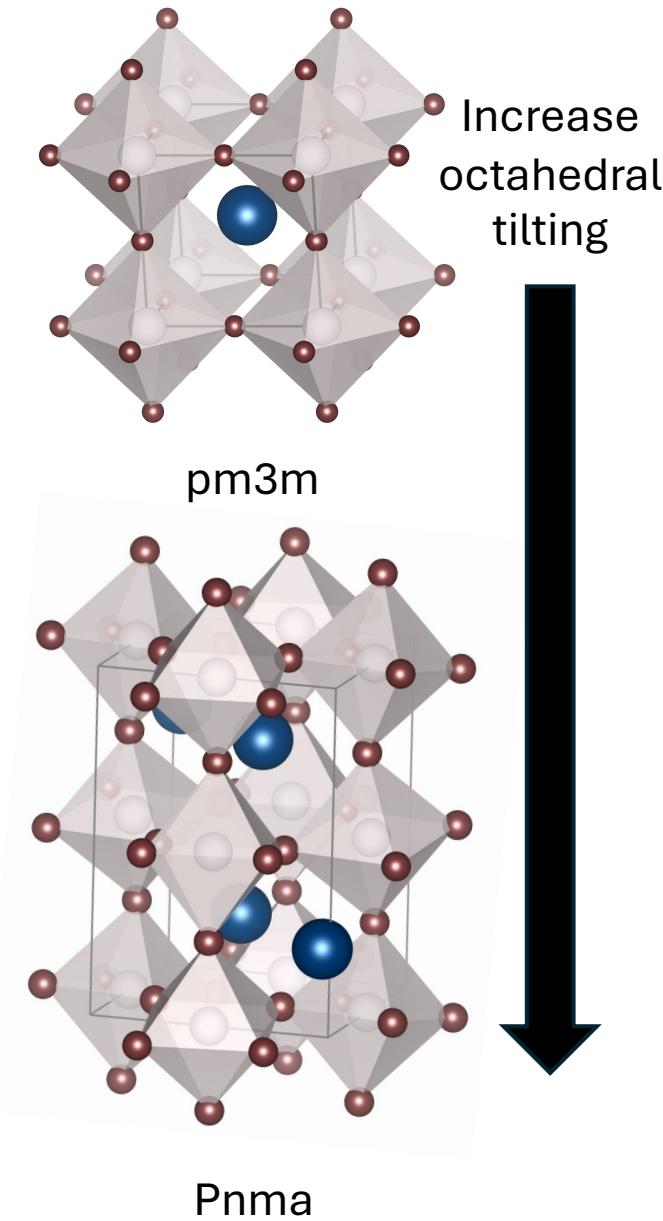
Need to determine the relevant structure  
to accurately calculate the material properties

The cubic phase is the prototypical perovskite structure  
used in calculations<sup>1,2</sup>

The cubic structure can distort to 23 different perovskite  
phases (referred to as Glazer tilts)

A. M. Glazer, The classification of tilted octahedra in  
perovskites. *Acta Cryst. B* 28, 3384–3392 (1972)

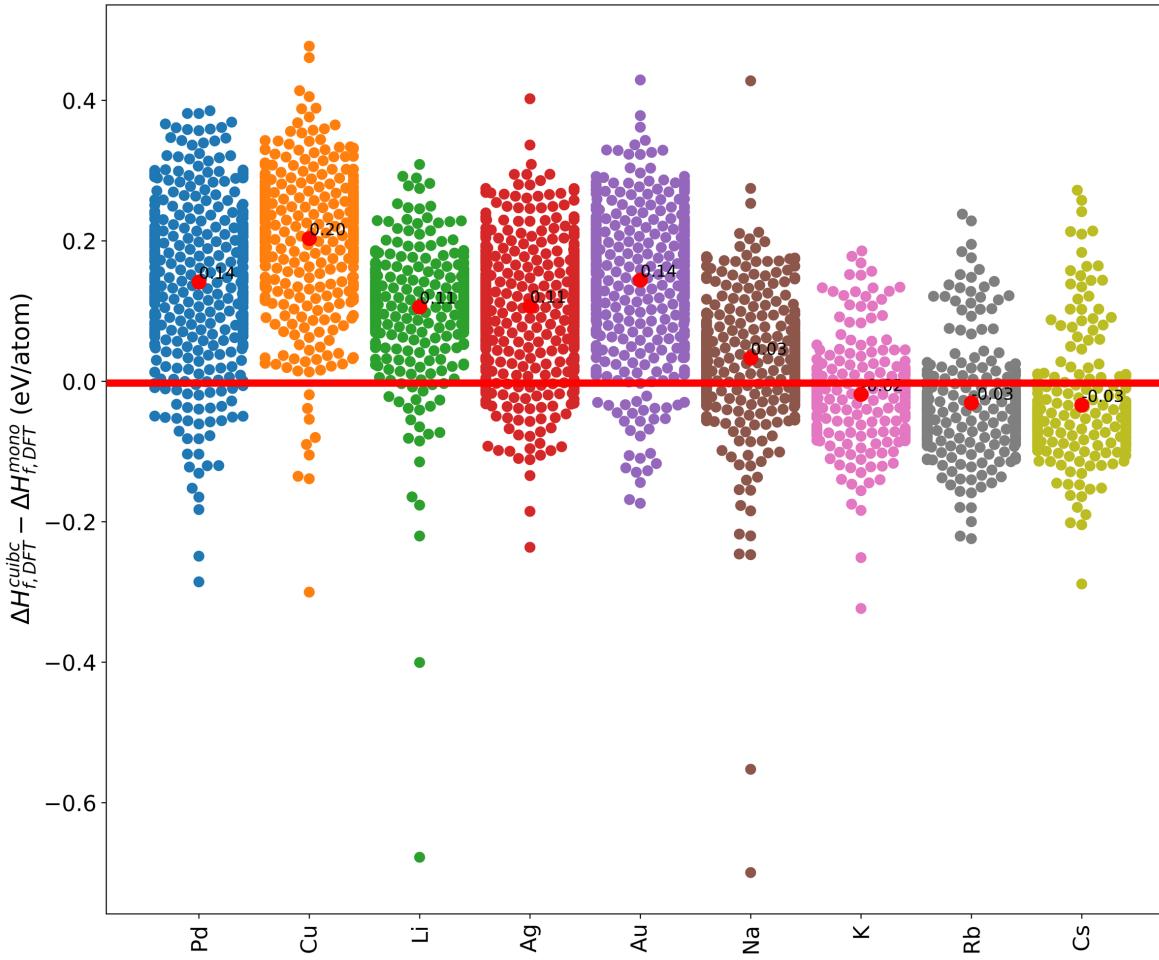
Various factors lead to distortions (such differences in  
ionic radii) but these other phases are often neglected  
in high-throughput screening tasks



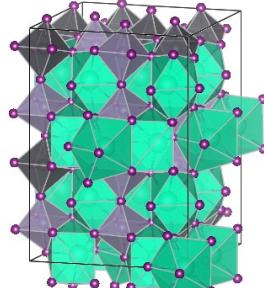
# Identification of trends in the CsA'BB'X<sub>6</sub> space: cubic is typically higher energy

- Evaluation of 23730 CsA'BB'X<sub>6</sub>
- Performed a single point DFT calculation at the MLIP structure

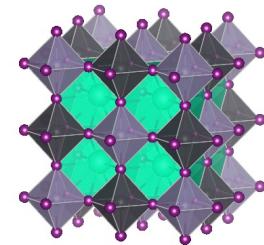
MAE between DFT and MLIP  
= 0.055 eV/atom



mono is more stable

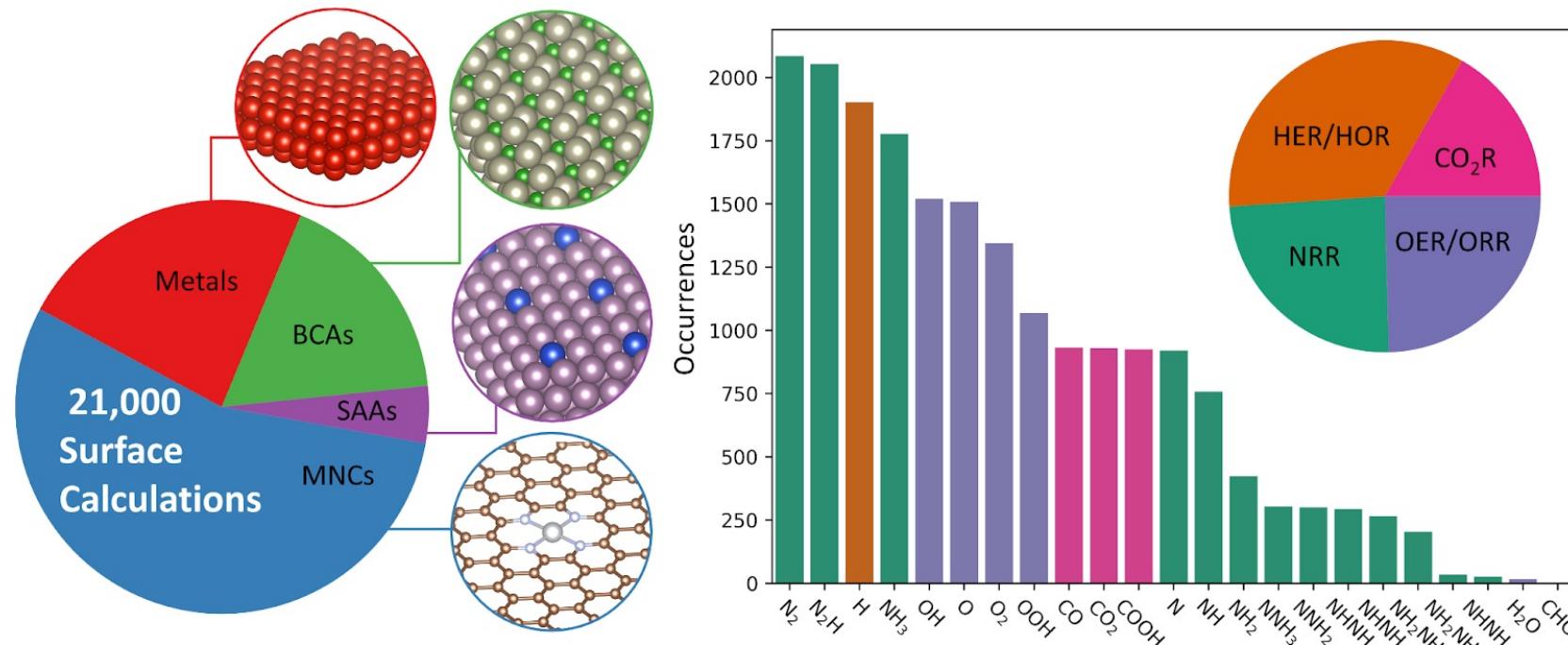


Cubic is more stable



# BEAST DB: Grand-Canonical Database of Electrocatalyst Properties

Beyond-DFT  
Electrochemistry  
with Accelerated and  
Solvated  
Techniques



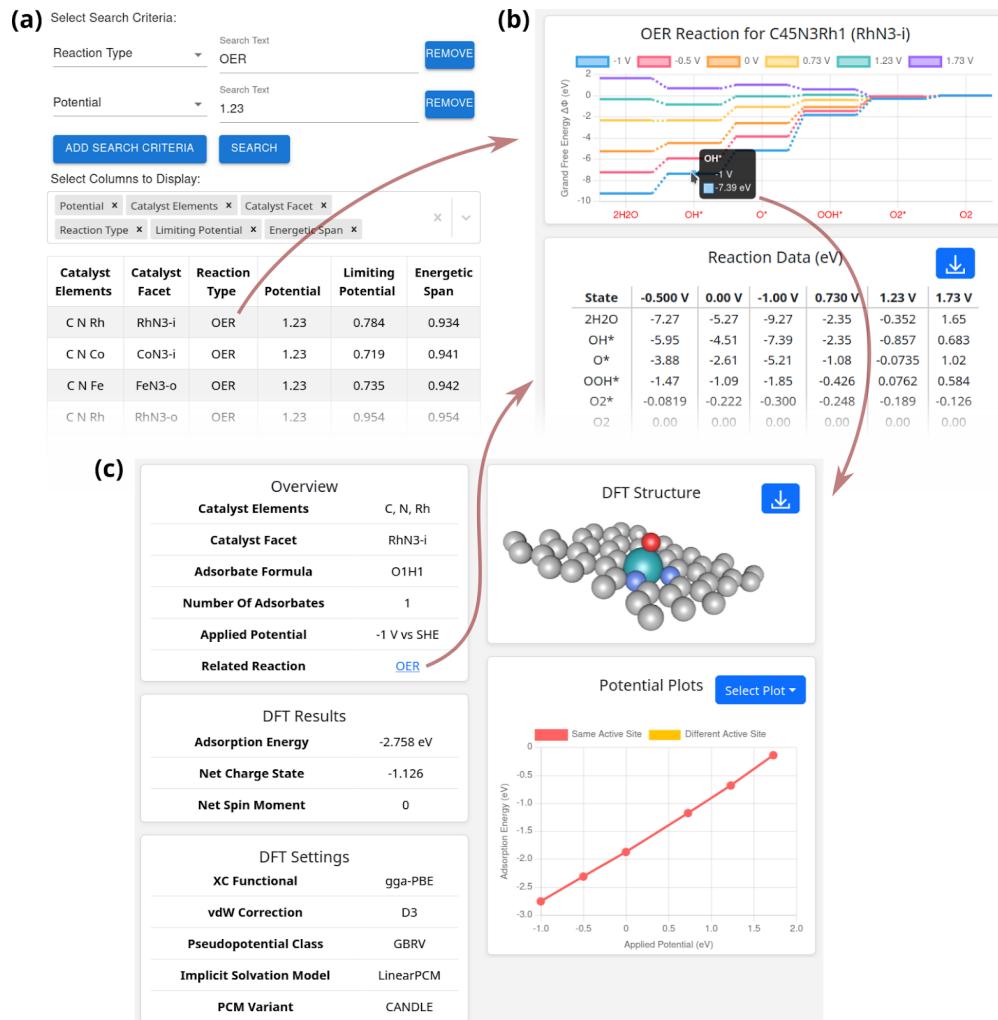
BeastDB: <https://beastdb.nrel.gov>



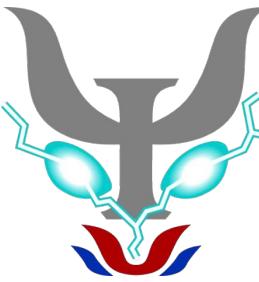
Office of  
Science

Cooper Tezak, Jacob Clary, Sophie Gerits, Joshua Quinton, Benjamin Rich, Nicholas Singstock, Abdulaziz Alherz, Taylor Aubry, Struan Clark, Rachel Hurst, Mauro Del Ben, Christopher Sutton, Ravishankar Sundaraman, Charles Musgrave, Derek Vigil-Fowler, "BEAST DB: Grand-Canonical Database of Electrocatalyst Properties" (2024).

# BEAST DB: Grand-Canonical Database of Electrocatalyst Properties



Beyond-DFT  
Electrochemistry  
with Accelerated and  
Solvated  
Techniques

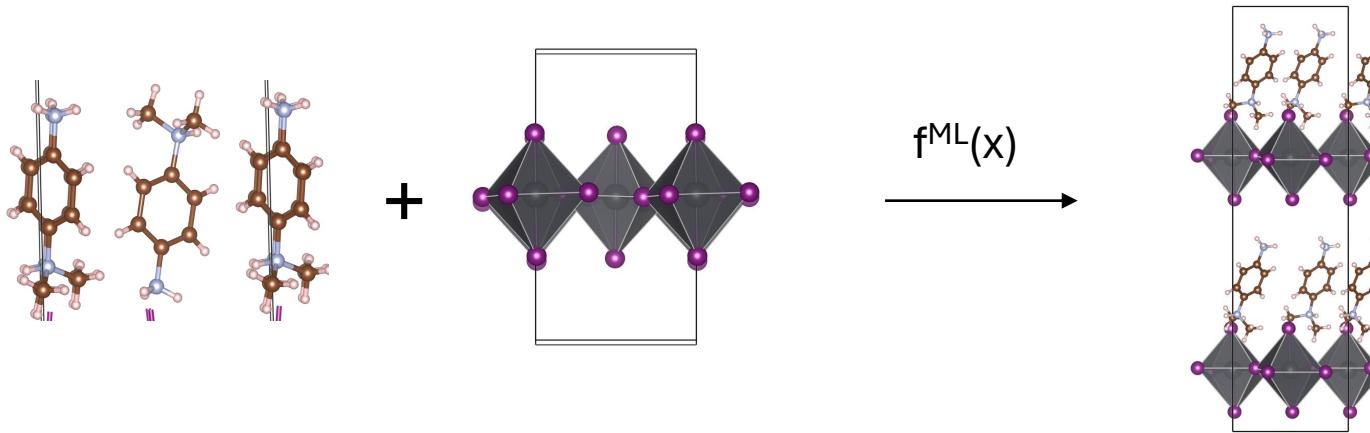


BeastDB: <https://beastdb.nrel.gov>



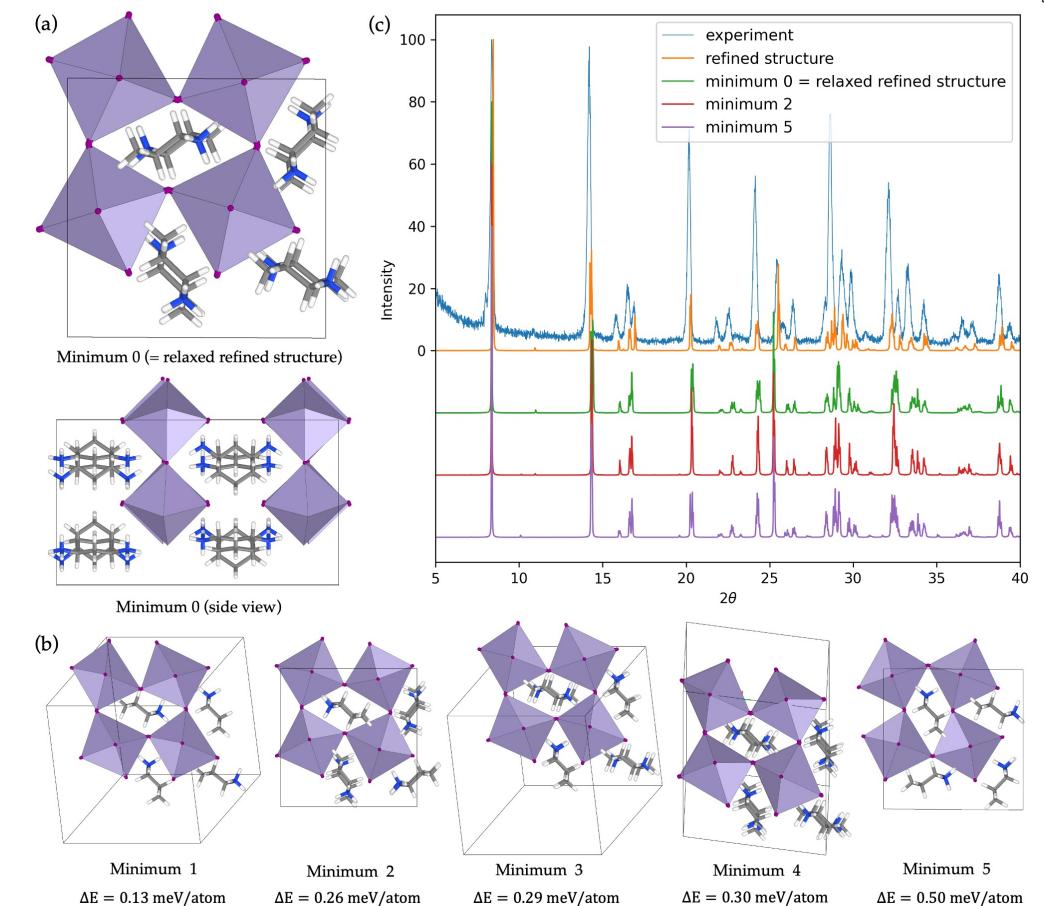
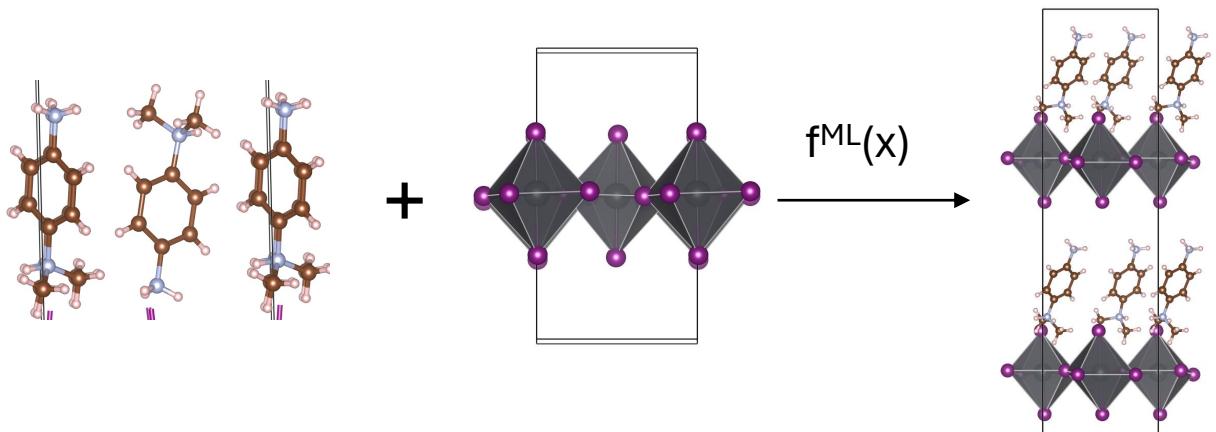
# Summary...

- Trained an accurate ML potentials to perform structure prediction for arbitrary combinations of organic and inorganic units



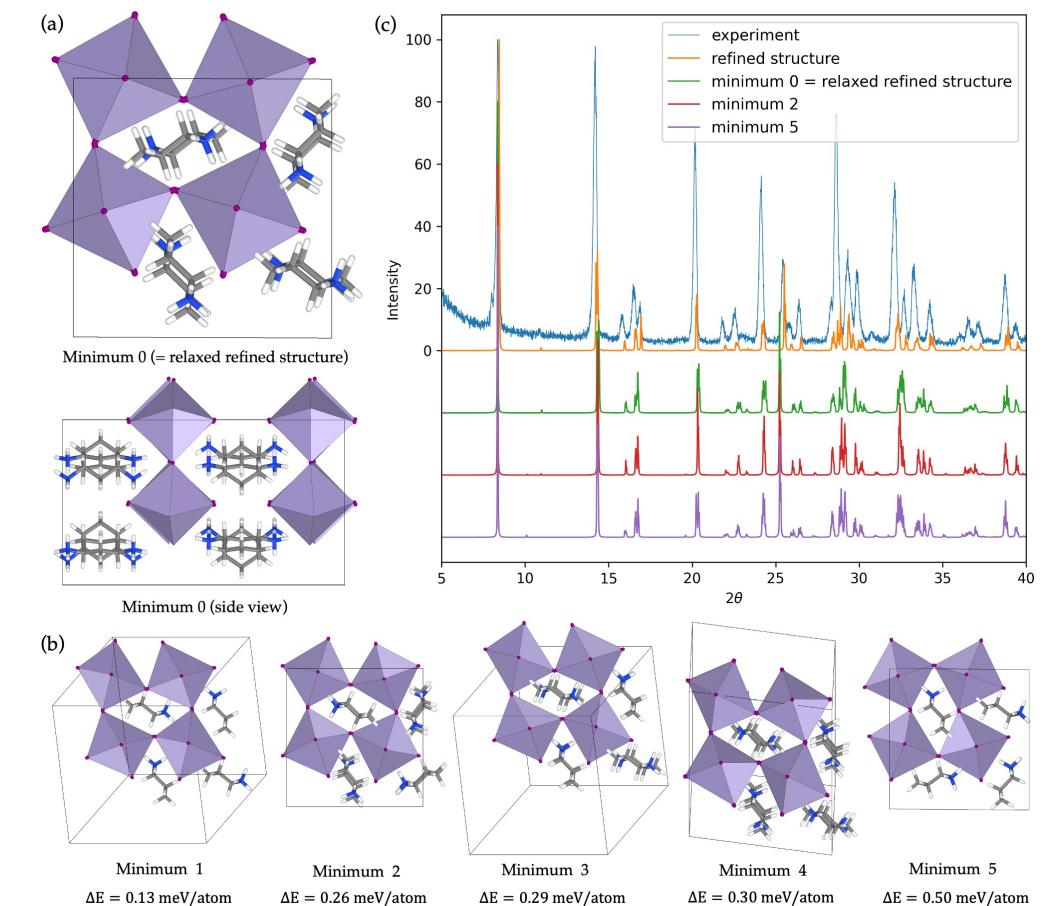
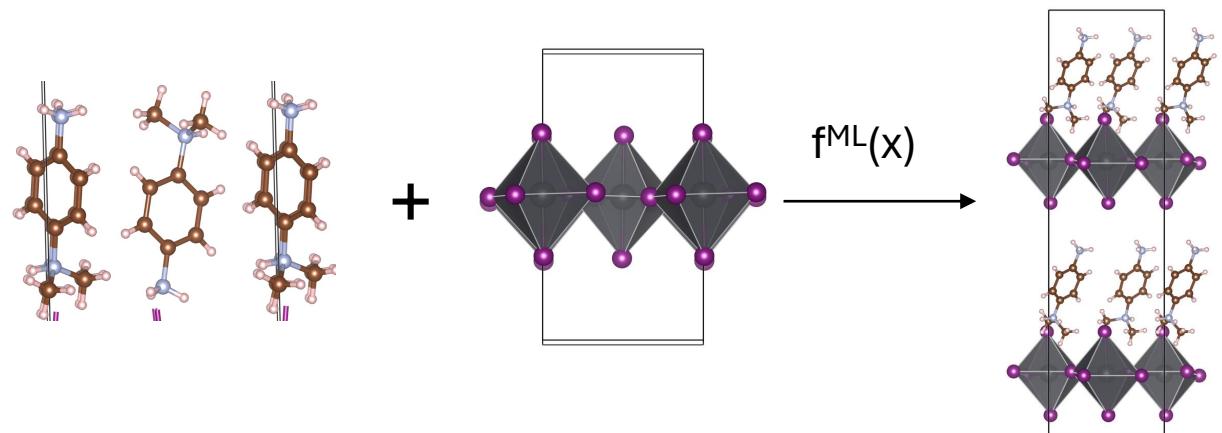
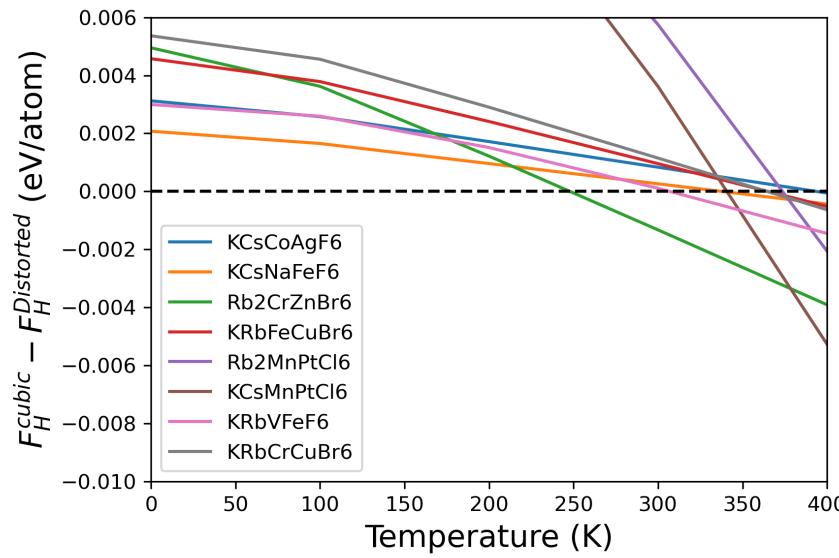
# Summary...

- Trained an accurate ML potentials to perform structure prediction for arbitrary combinations of organic and inorganic units
- Our results were experimentally validated by experiment
- Our results were experimentally validated by experiment



# Summary...

- Trained an accurate ML potentials to perform structure prediction for arbitrary combinations of organic and inorganic units
- Our results were experimentally validated by experiment
- Examined the phase transition in AA'BB'X<sub>6</sub> compounds



# Thank you for your attention!

Will Baldwin (Cambridge)



Gabor  
Csanyi

Prof. Gábor Csányi (Cambridge)

Dr. Josh Kennedy (ARFL/RX)

Dr. Nima Karimitari Bare (USC)

Dr. Zach Bare (USC)

Will  
Baldwin



## Acknowledgement of funding sources:



U.S. DEPARTMENT OF  
**ENERGY**

Office of  
Science

DE-SC0022247



U.S. DEPARTMENT  
OF ENERGY  
SOLAR AWARDEE™

DE-EE0009515



20IOE044