CHEMISTRY OF MATERIALS

Open Access

pubs.acs.org/cm

Article

Cluster-Based Machine Learning Potentials to Describe Disordered Metal–Organic Frameworks up to the Mesoscale

Published as part of Chemistry of Materials special issue "Machine Learning for Materials Chemistry".

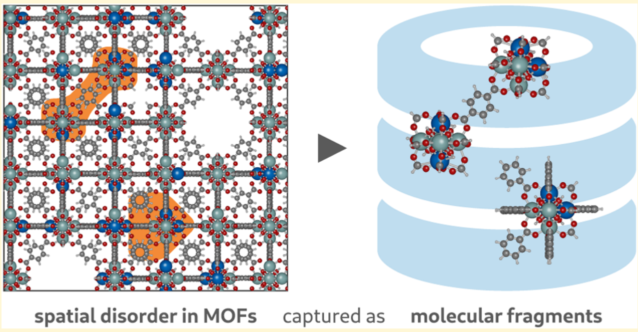
Pieter Dobbelaere, Sander Vandenhaute, and Veronique Van Speybroeck*

Cite This: https://doi.org/10.1021/acs.chemmater.5c00821

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Metal–organic frameworks (MOFs) are highly interesting and tunable materials. By incorporating spatial defects into their atomic structure, MOFs can be finetuned to exhibit precise chemical functionalities, extending their applicability in various technological fields. Defect engineering requires a fundamental understanding of the nature of spatial disorder and consequent changes in material properties, which is currently lacking. We introduce the cluster-based learning methodology, enabling the development of state-of-the-art machine learning potentials (MLPs) from defective systems at any length scale. Our method identifies atomic interactions in bulk structures and extracts local environments as finite molecular fragments to augment the model's training data where needed. We show that cluster-based learning delivers MLPs capable of accurately describing spatial defects in mesoscopic systems with over 20 thousand atoms. Afterward, we select our best model to investigate some major mechanical properties of spatially disordered UiO-66-derived structures, elucidating the influence of defect concentration and composition on material behavior. Our analysis includes large supercell structures, demonstrating that (near-) *ab initio* accuracy is within reach at the mesoscale.



1. INTRODUCTION

Metal–organic frameworks, MOFs, are porous crystalline solids that have evolved into versatile materials with many technological and industrial applications in e.g., heterogeneous catalysis, gas sorption and separation or nanoscopic actuating and sensing.[1−4] Structurally, MOFs comprise a topological lattice made from several secondary building units, namely metal nodes and organic ligands.[5,6] In computational analyses, they are usually treated as well-ordered and pristine molecular systems. However, many recent studies have highlighted that the strength of MOFs lies in their ability to encapsulate atomically precise functions through defects.[7,8] Many enticing MOF properties are heavily influenced or modulated by inhomogeneities in the perfect framework. Such deviations from order exist in every realistic structure, appearing in different forms and over multiple length scales. We find point defects like metal atom substitutions or missing ligand vacancies on the nanoscale, whereas mesoscopic disorder materializes as larger cavities or mesopores, regions of phase coexistence and surface boundaries in finite crystals.[9,10] Understanding how various types and arrangements of spatial disorder enhance or interfere with desired material characteristics, is crucial to exploit this configurational freedom and tailor frameworks to their intended application.[11,12] To unlock the full potential of MOFs through defect engineering, we require computationally efficient and accurate modeling techniques that can describe spatial disorder up to the mesoscale.

Over the past few years, advances in machine learning potentials (MLPs) have initiated a new era for molecular modeling of functional materials.[13−18] These cutting-edge models parametrize a molecular potential energy surface (PES) by learning atomic interactions from underlying quantum mechanical (QM) calculations. They can accurately reproduce their reference level of theory (LOT) at a (comparatively) vanishingly small inference cost, once trained. In simulations, MLPs assume the role of interatomic potential, similar to classical force fields—albeit much more faithful to QM behavior and without enforcing a fixed bonding topology.[19,20] Their main weakness is the notoriously poor ability to describe out-of-dataset structures; a phenomenon aptly named the extrapolation problem.[21−23] Therefore, the accuracy and transferability of any MLP depend vitally on the chemical and configurational space

Received: April 4, 2025
Revised: July 11, 2025
Accepted: July 11, 2025

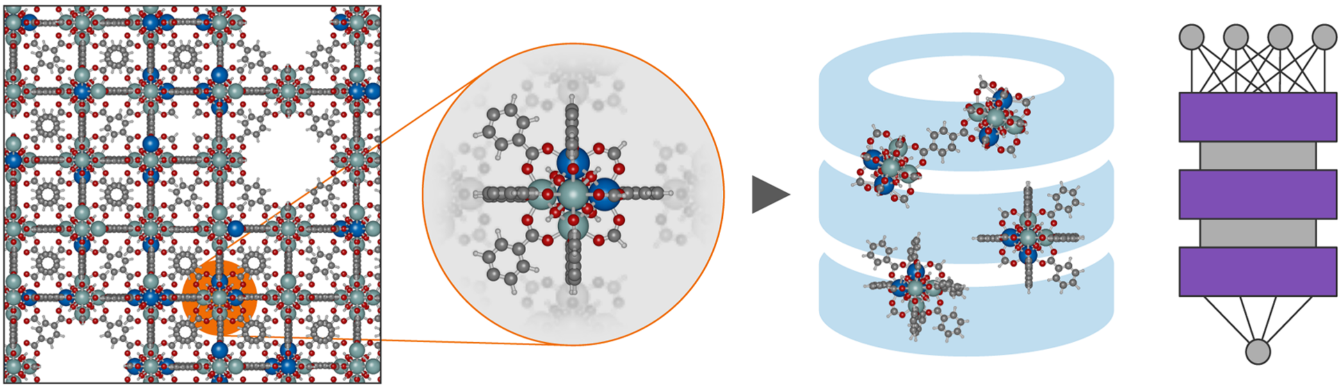


Figure 1. Cluster-based learning. Local chemical environments in bulk frameworks are extracted as finite clusters to capture specific atomic interactions, enabling MLPs to learn spatial disorder at the mesoscale.

covered by its training data. Developing capable potentials for disordered frameworks entails constructing representative data sets, which becomes computationally intractable at the length scales needed to represent disorder. At present, the main hurdle holding back MLP development for defective materials is the cost of *ab initio* data generation. Here, we will address that limitation and create robust neural network models for defective MOFs.

To achieve our goal, we introduce the chemical environment of an atom as the key concept governing molecular interactions. Intuitively, a chemical environment can be understood as the sphere of influence between an atom and its periphery, uniquely encompassing all non-negligible interactions with neighboring atoms and externally applied fields. The principle of nearsightedness of electronic matter—which states that a perturbing potential only causes a finite change in the local electron density, whose magnitude decreases monotonically with increasing distance to said perturbation[24]—presumes a finite radius of electronic interaction and lies at the core of many linearly scaling density functional theory (DFT) implementations.[25,26] Consequently, environments must have a limited spatial extent and can be treated as local molecular properties. Hence, a data set of QM calculations can also be interpreted as a collection of independent environment-interaction pairs. If the data set contains all environments to fully describe every atomic interaction present in a given system of interest, we say it is representative of that system.

In this work, we propose that representative data sets can be constructed for arbitrarily large (and disordered) structures by extracting individual environments in the form of small molecular fragments (see Figure 1). The procedure involves two steps: (i) identify unknown chemical environments using a learned MLP characterization and (ii) separate those environments from the molecular bulk as isolated fragments. This "cluster-based learning" idea circumvents computational bottlenecks and allows us to selectively design the chemical and configurational space spanned by the training data.

For optimal efficiency, we embed our methodology into an active learning (AL) scheme—a machine-learning paradigm where models are iteratively refined by cycling between data acquisition and retraining phases.[27−30] In a molecular modeling context, AL foregoes any expensive *ab initio* sampling conventionally used to generate reference data, making it exceptionally attractive.[31] We recently developed an in-house AL framework, Psiflow, that facilitates the definition and execution of intricate workflows on high-performance computing infrastructures—without direct user intervention, essentially automating MLP training for periodic systems and molecules.[32] Here, we implement a novel cluster-based data acquisition algorithm and provide an interface with existing Psiflow functionality. Two interesting use cases are tackled in this work:

Large structures. Molecular systems are often spatially complex and can only be described in large unit cells. This is especially true in the world of MOFs, e.g., the isoreticular DUT-series (1000−2000 atoms),[33] MIL-100 (∼3000 atoms)[34] or plenty of other hierarchically porous frameworks.[35] To also include defects and disorder, structures must grow even larger. With so many degrees of freedom, *ab initio* evaluations become downright impossible. Deconstructing large cells into smaller fragments decouples their characteristic dimensions from the cost of QM computations. Provided we can define a suitable partitioning into compact clusters, this strategy enables the construction of accurate MLPs for any atomic structure regardless of its inherent size.

Transferable MLPs. Selectively filtering for missing environments can efficiently increase model transferability while reducing data redundancy. Consider, for example, the prototypical MOF UiO-66 and its isoreticular cousin UiO-67. To train an MLP, one could naively collect periodic data for both systems separately. This is wasteful, as both frameworks share most atomic interactions. Instead, we could construct a data set for UiO-66 first, and later extend it with fragments that contain the unique environments of UiO-67. Gathering data incrementally eliminates unnecessary QM evaluations and is especially suited to developing universal models that describe large families of materials.

We should concede one caveat: since our idea involves finite fragments for training, models will only learn short-range interactions. However, reference data can always contain mixed boundary conditions, i.e., some periodic structures—describing long-range phenomena—and molecular clusters—containing local environments. When electrostatics or electron dispersion become dominant energy contributions, MLPs are often enhanced by including physical priors or predicting partial charges, for example.[36] These corrections remain applicable in conjunction with the proposed approach.

In this work, we explain the inner workings of the cluster-based learning procedure; how chemical environments are identified based on an internal MLP representation, how clusters are designed to extract specific interactions and how everything fits into an automated AL workflow. Using UiO-66 as a case study, we investigate how small point defects alter framework interactions in pristine MOFs, show when trained MLPs fail and how to correct spurious behaviors by extending existing data sets. To highlight the general applicability of our methodology, we successfully learn disordered systems at diverse length scales, containing different types and concentrations of spatial disorder (see Figure 1). Finally, with a robust and transferable MLP at hand, we explore the pressure response of various systems in the UiO-66 family and uncover fundamental relations between induced disorder and mechanical resilience in MOFs.

2. METHODS

Below, we discuss the two major components that form our cluster-based learning implementation: uncertainty quantification and cluster extraction (see Figure 2A). First, we characterize chemical environments in sample structures using a learned MLP representation and quantify model uncertainty—a measure of confidence in its predictive accuracy. Excessive uncertainties indicate MLP extrapolation and unreliable inference, showing deficiencies in the training data set. Then, we extract regions of high uncertainty from the bulk material into representative molecular fragments, using an algorithm to design clusters that mimic environments found in the original system. This data acquisition method is automated to easily enable *ab initio* evaluation of bulk interactions. It can deliver state-of-the-art MLPs for arbitrarily large structures at very modest computational expense, without requiring manual intervention.

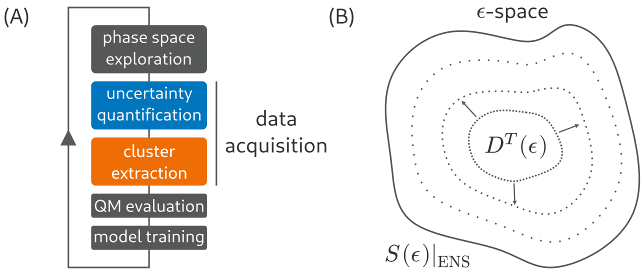


Figure 2. (A) Schematic overview of the active learning workflow. (B) Every cycle $D^T(\epsilon)$ grows in ϵ-space, incrementally approaching the target $S(\epsilon)|_{\mathrm{ENS}}$.

We formalize the intuitive definition of a chemical environment in Section 2.1. A brief overview of the AL workflow is given in Section 2.2. Mathematical details on MLP uncertainty quantification are provided in Section 2.3. Finally, Section 2.4 discusses the intricacies of the cluster extraction procedure.

2.1. Chemical Environments.

Within a molecular structure, we define the chemical environment ϵ_i of atom i as the radius of interaction between this atom and its neighborhood. It encompasses everything the central atom can "feel", such as surrounding atoms and externally applied fields, and is necessarily local by the principle of nearsightedness of electronic matter.[24] Concretely, an environment ϵ_i consists of all (structural) information to uniquely describe the total sum of perceptible influences on atom i. We will drop the subscript when referring to any (generic) environment.

Two isolated structures connected by rotations, translations and inversions are physically identical; they carry the same atomic interactions, although the resulting forces could differ in orientation. Accordingly, these symmetries will project ϵ onto itself. We say that ϵ is invariant for transformations of the Euclidean group E(3), abstracting away directional degrees of freedom. This is analogous to an invariant molecular energy giving rise to equivariant atomic forces, which transform like vectors under E(3).

Simplifying further notation (see Table 1), we denote a molecular system with S, a data set of structures with D, and use $D(\epsilon)$ to explicitly refer to the ϵ contained within D. When sampling in some thermodynamic ensemble, system S has access to a volume of its configuration space in accordance with state variables ENS (e.g., the NPT ensemble with P = 1 bar and T = 300 K). Consequently, S can occupy a related volume in "chemical environment space". $S(\epsilon)|_{\mathrm{ENS}}$ represents all ϵ that appear in any configuration of S under thermodynamic conditions ENS, as schematically shown in Figure 2B. Adopting these conventions, we assume D to be representative of S under ENS if every ϵ_i in $S(\epsilon)|_{\mathrm{ENS}}$ has a similar ϵ_j in D. The notion of similarity between environments will become meaningful in Sections 2.1.1 and 2.1.2.

Table 1. Summary of Used Symbols

ϵ	chemical environment
S	molecular system
D	data set of structures
$\boldsymbol{F}$	atomic force vector
$\mathcal{F}$	atomic feature vector

Computationally, we can describe molecular structures using various LOTs. The properties of every ϵ—i.e., its shape and spatial extent or the relative importance of different interactions—will depend on the chosen computational method. Therefore, we introduce a further specification, ϵ^{lot}, to distinguish the LOT employed. In this work, reference data sets are constructed using DFT and all simulations rely on MLP inference. Correspondingly, we consider both ϵ^{dft} and ϵ^{mlp} as approximations to the true QM ϵ. In practice, the superscript "dft" refers to a specific set of computational parameters (functional, basis set, etc.) and "mlp" points to a particular MLP.

The fundamental ansatz of cluster-based learning is the idea that ϵ from bulk structures can be captured and extracted in finite molecular fragments. Atom i should experience the same total interaction in the designed cluster and the original parent system, for all relevant LOTs

$$(\epsilon_i^{\mathrm{dft}}, \epsilon_i^{\mathrm{mlp}})_{\mathrm{bulk}} = (\epsilon_i^{\mathrm{dft}}, \epsilon_i^{\mathrm{mlp}})_{\mathrm{cluster}} \tag{1}$$

Equation 1 represents the condition of "environment matching", and is a prerequisite for MLPs to learn bulk interactions from a data set of clusters. We can enforce it by defining methods to characterize and compare different ϵ, namely force matching (Section 2.1.1) and feature matching (Section 2.1.2). The latter requires a quantitative representation of ϵ, while the former only involves evaluated force labels.

2.1.1. Force Matching.

Atomic forces are local observables that directly reflect underlying molecular interactions. Disregarding orientation, equal interactions will cause identical forces. If the original structure and extracted cluster are aligned so that corresponding atoms overlap perfectly, forces obey

$$(\epsilon_i)_{\mathrm{bulk}} = (\epsilon_i)_{\mathrm{cluster}} \Rightarrow (\boldsymbol{F}_i)_{\mathrm{bulk}} = (\boldsymbol{F}_i)_{\mathrm{cluster}} \tag{2}$$

where $\boldsymbol{F}_i$ is the total force vector felt by atom i. Equation 2 is not injective; many ϵ give rise to the same $\boldsymbol{F}$. Nevertheless, we will assume the inverse holds too. In other words: if the force on atom i matches between fragment and parent, their environments should be equivalent.

Force matching prescribes a transparent algorithm to find appropriate clusters: (i) evaluate $(\boldsymbol{F}_i)_{\mathrm{bulk}}$, (ii) evaluate $(\boldsymbol{F}_i)_{\mathrm{cluster}}$ for a series of candidate fragments and (iii) find the closest match between (i) and (ii), as that cluster encloses the best approximation of $(\epsilon_i)_{\mathrm{bulk}}$. At large length scales, (*ab initio*) evaluations of the parent structure might no longer be possible. As a workaround, one can extrapolate the evolution of $(\boldsymbol{F}_i)_{\mathrm{cluster}}$ for fragments of increasing size to estimate a limit for $(\boldsymbol{F}_i)_{\mathrm{bulk}}$. An example of force matching for ϵ^{dft} and ϵ^{mlp} is discussed in Section SI.6.

2.1.2. Feature Matching.

During training, neural network MLPs learn to encode the surroundings of an atom into a descriptive feature representation, progressively increasing the level of abstraction throughout several hidden layers. The range of ϵ^{mlp} is limited by the atomic interaction radius r_{max} of the model, determining how far it can "look ahead". This is only an upper bound, as r_{max} can be chosen arbitrarily large, whereas the intrinsic size of ϵ^{mlp} cannot be. To accurately infer molecular interactions, the MLP must discriminate various ϵ^{mlp} by its features. We define a feature descriptor $\mathcal{F}$ constructed from (a subset of) n invariant network nodes—e.g., the

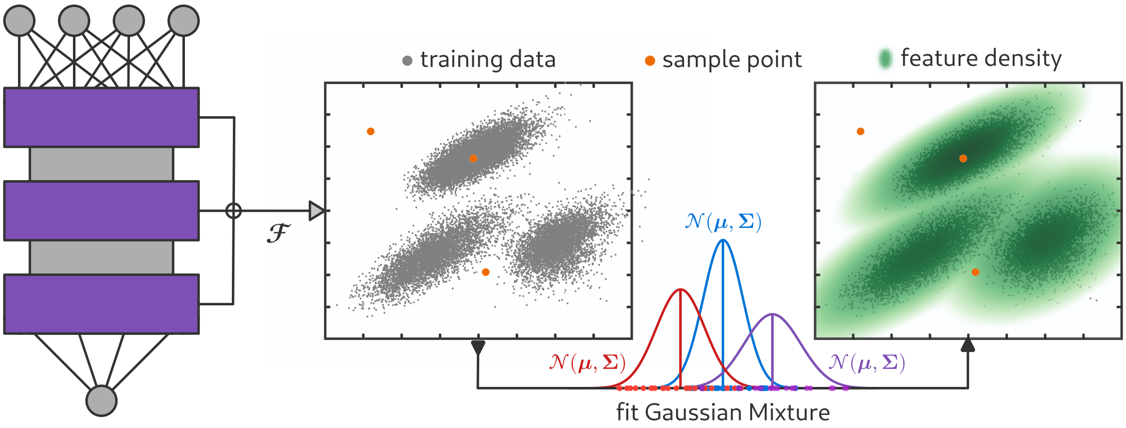


Figure 3. Descriptor $\mathcal{F}$ casts $D^T(\epsilon)$ into feature space. A GMM is fit to represent the density of training points. For any sample ϵ (orange), the density likelihood represents similarity with $D^T(\epsilon)$. In trained MLPs, this likelihood is inversely correlated with model uncertainty.

final hidden layer—to identify all ϵ^{mlp}. These n-dimensional $\mathcal{F}$-vectors span feature space. In analogy to eq 2, the relation between environment and descriptor is given by

$$(\epsilon_i^{\mathrm{mlp}})_{\mathrm{bulk}} = (\epsilon_i^{\mathrm{mlp}})_{\mathrm{cluster}} \Rightarrow (\mathcal{F}_i)_{\mathrm{bulk}} = (\mathcal{F}_i)_{\mathrm{cluster}} \qquad (3)$$

If the MLP architecture and chosen $\mathcal{F}$ are sufficiently expressive, the inverse of eq 3 will also hold.[37,38] Therefore, we can identify $\epsilon_i^{\mathrm{mlp}}$ with a point in feature space and associate differences in $\mathcal{F}$ with a degree of (dis)similarity between ϵ^{mlp}. Note that comparing multiple $\mathcal{F}$ is only meaningful for a single MLP, because altering network weights changes the structure of feature space. In Section 4.1, we show that a trained model does indeed distinguish distinct atomic interactions internally, i.e., that $\mathcal{F}$-vectors embed chemical information.

2.2. Active Learning. We aim to create an accurate MLP for system S under thermodynamic conditions ENS while minimizing computational costs and linear execution time. AL workflows achieve this objective by iteratively expanding a training set D^T (superscript T for training) to be more representative of $S(\epsilon)|_{\mathrm{ENS}}$. The process—coined an AL campaign—consists of several AL cycles and is illustrated in Figure 2A.

Following a one-off initialization step, every cycle involves phase space exploration, data acquisition and evaluation, and model retraining stages. During exploration, the goal is to sample new ϵ from $S(\epsilon)|_{\mathrm{ENS}}$ by probing additional configurations of S, using a model trained in the previous cycle. The resulting structures are analyzed for unknown $\epsilon \notin D^T(\epsilon)$, which are extracted into finite clusters and evaluated at an *ab initio* LOT. These fragments are appended to the existing data set, extending the region of ϵ-space described by $D^T(\epsilon)$ (see Figure 2B). Finally, the MLP is retrained with a newly improved data set, concluding the current cycle. An AL campaign continues for a predefined number of iterations or until some accuracy threshold has been met. Section SI.4 provides concrete descriptions of a full cycle as implemented in this work.

2.3. Uncertainty Quantification. The computational efficiency of AL workflows can be drastically improved through active (as opposed to random) sample selection. These techniques aim to maximize transferability and reduce data redundancy of D^T by identifying out-of-dataset structures. Broadly, they assess MLP uncertainty—whether we expect the model to reproduce atomic interactions correctly—for unlabeled sample configurations. High uncertainty suggests poor inference accuracy and the presence of unfamiliar environments, making it worthwhile to incorporate those ϵ into $D^T(\epsilon)$.

We find a myriad of methods to quantify uncertainty in recent AL literature. Effective estimates for cluster-based learning should not involve *ab initio* calculations, only rely on the current model and data set and provide per-atom predictions. If GPU compute time abounds, query-by-committee approaches have proven general and widely successful in machine learning.[39,40] Gaussian Process models can immediately leverage their built-in predictive uncertainty.[41,42] Other data-driven implementations employ statistical measures computed over kernel or feature embeddings of D^T.[43−46]

Here, we take the latter route, using a learned descriptor $\mathcal{F}$ to convert the abstract $D^T(\epsilon)$ into a concrete set of vectors that describe all interactions of D^T in MLP feature space. After fitting these features with a density distribution, the likelihood of $\mathcal{F}_i$ estimates how similar $\epsilon_i^{\mathrm{mlp}}$ is to existing training data. Low likelihoods imply few nearby data points, indicating out-of-dataset ϵ, or vice versa (see Figure 3). Following model training, we assume that MLP uncertainty and density likelihoods are inversely correlated (see Section SI.7).

Many unsupervised density models are suited for this task. We opt for a Gaussian Mixture Model (GMM),[47] which is a weighted summation of multivariate Gaussians

$$\mathrm{GMM}[D^T(\epsilon), \mathcal{F}] = \sum_m^M a_m \mathcal{N}_n(\boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m) \qquad (4)$$

where n is the dimension of $\mathcal{F}$, M denotes the number of mixture components, a_m is a relative weight factor for each component and $\mathcal{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is an n-dimensional normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. The total distribution is normalized to unity. We utilize the Bayesian information criterion to decide an appropriate value for M.[48]

If the curse of dimensionality[49] complicates density construction, one can resort to techniques such as principal component analysis (PCA)[50] to reduce the dimension of $\mathcal{F}$ to a more manageable n', while retaining most information. This is also useful in visualizations. For increased model flexibility (see Section 2.4), we group $\mathcal{F}_i$ according to the atomic number of atom i and build independent distributions for every element in D^T. If no training data is available for a particular element, matching $\mathcal{F}$ are given an artificial likelihood of zero.

2.4. Cluster Extraction. Molecular fragments are finite clusters designed to encapsulate some ϵ from a larger parent structure. Using the uncertainty approach of Section 2.3, we can scan sample configurations for unknown $\epsilon \notin D^T(\epsilon)$, effectively creating a heatmap of where new interactions are located (see Figure SI.3). If we find a spatially concentrated group of ϵ with high uncertainty, they should be extracted into a new fragment.

The first component of a cluster under design is its core, containing all atoms whose ϵ are to be added to $D^T(\epsilon)$. To fulfill eq 1 (environment matching), we must envelop the chosen core in a suitable mantle of atoms from the original structure, serving as padding to mimic bulk environments. This is achieved using the force matching approach (see Section 2.1.1). At this point, our cluster successfully captures the core ϵ, but includes several dangling bonds at its surface, complicating future *ab initio* evaluations. We form a termination layer to saturate broken bonds - creating new atoms not in the parent system—and form a chemically valid, preferably closed-shell and charge-neutral fragment. In total, the cluster consists of three distinct regions: (i) a "core" realizing eq 1, (ii) a "mantle" violating eq 1 and (iii) a termination layer without a counterpart in the original structure (see Figure 5).

This section only sketches a rough outline, rather than providing a concrete extraction recipe. Defining reasonable clusters requires chemical insight, and depends strongly on the material of interest. We seek to capture a maximal amount of information in each fragment, while keeping computational costs to a minimum. A more thorough discussion of this process, applied to MOFs, is given in Section SI.5. We provide an in-depth example of cluster design through force matching in Section SI.6. Section 4.2 provides a summary.

3. COMPUTATIONAL DETAILS

We present a concise overview of the main computational choices made in this work, and will regularly refer to the SI for more exhaustive explanations.

QM Evaluations. Every *ab initio* evaluation is performed by the GPAW software engine,[51] because it allows both finite and periodic boundary conditions. This way, clusters and periodic structures are evaluated in a consistent way. We work with its finite difference grid formulation, employ the Perdew−Burke−Ernzerhof (PBE) DFT functional[52] with Becke-Johnson D3 dispersion correction[53] and use a basis grid spacing of 0.175 Å. Further details can be found in Section SI.1.

MLPs. We choose NequIP as the fundamental architecture for all MLPs.[54] It is a convolutional neural network that implements E(3)-equivariant message-passing between nearby atoms. Its convolution operation consists of products between geometric tensors of rotation order l, resulting in scalar ($l = 0$), vectorial ($l = 1$) and tensorial ($l > 1$) interactions that can efficiently represent atomic environments. We activate equivariant features by setting the rotation order $l > 0$. In Section 4, model accuracy is used to assess the quality of D^T. Because inference flaws should directly reflect data set deficiencies, we allow MLPs to reach optimal performance and extract maximal information—in practice, until their validation error stops decreasing. A comprehensive specification of network hyperparameters and training setup is given in Section SI.2. Note that our methodology is architecture-agnostic, and other MLP frameworks could be used as well.

Simulations. Molecular dynamics (MD) is performed with either the OpenMM engine[55] or the in-house YAFF software.[56] OpenMM offers more efficient simulation algorithms but is limited in functionality for periodic systems, whereas YAFF implements many ensembles specifically geared toward periodic boundary conditions. Simulation parameters can be found in Section SI.3.

Data Set Generation. Periodic data sets for system S are generated from OpenMM MD in the isobaric−isothermal (NPT) ensemble at 600 K and various pressures, ensuring diversity of ϵ. Structures are selected uniformly within a fixed volume range around the equilibrium volume of S. However, meticulously exploring the molecular PES requires a capable MLP, which—in turn—requires representative training data. In this work, every data set is constructed post-AL, i.e., after we verified the model has become suitably accurate and samples the correct distribution of structural configurations.

Active Learning. In the AL workflow, exploration consists of 500 fs OpenMM walks using applied pressures randomly chosen between −1.5 and 1.5 GPa. We restrict descriptor $\mathcal{F}$ to the final hidden layer of every MLP, which has 8 or 16 dimensions (see Section SI.2) and directly precedes the atomic energy prediction. Density models in feature space are parametrized using the Gaussian mixture implementation and expectation-maximization algorithm of scikit-learn.[57] To design molecular clusters for different types of spatial disorder, we follow the approach illustrated in Section 4.2.

Error Metrics. As seen in Section 4.3, MLP force errors tend to be very localized around regions of disorder. Conventional metrics such as the mean absolute error (MAE) or root-mean-square error (RMSE) lack the sensitivity required to capture this local behavior, while a maximal error is very susceptible to outliers. We propose a different metric. The MAE_{Pk} is a thresholded MAE, computed over error labels larger in magnitude than the kth percentile of absolute errors. The value of k tunes its sensitivity to outlying values. In the limit, $k = 0$ reverts to the standard MAE and $k = 100$ returns the maximal error. We choose $k = 95$ to average over the 5% largest errors and use MAE_{P95} as the prime metric to discuss force accuracy in disordered structures.

Some MLPs make wildly inaccurate energy predictions for out-of-dataset systems. In these instances, the dominating source of error is usually a constant offset and the remaining variance is very small. Because MAE or RMSE statistics fail to separate both error contributions, we will report the mean and standard deviation of ΔE_i

$$\Delta E_{\text{avg}} = \frac{1}{M}\sum_{i}^{M} \Delta E_i, \qquad \Delta E_{\text{std}} = \sqrt{\frac{1}{M}\sum_{i}^{M} (\Delta E_i - \Delta E_{\text{avg}})^2} \tag{5}$$

where ΔE_i is the per-atom energy error for structure i (out of M). For single-system data sets, ΔE_{avg} approximates the inherent shift between the MLP PES and *ab initio* LOT. ΔE_{std} can be interpreted as an offset-corrected energy RMSE. The former metric can be ignored when comparing configurations of S; it is irrelevant for optimizations or MD sampling. On the contrary, ΔE_{avg} and ΔE_{std} are both important for accurate analysis of the relative stability of different systems.

Mechanical Characterization. We investigate the mechanical behavior of system S by deriving static energy-vs-volume (EV) and dynamic pressure-vs-volume (PV) profiles to inspect the impact of defects on its properties. EV curves are computed following the approach outlined in ref 58: perform a series of fixed-volume structure optimizations for a grid of volume points—allowing cell shape and atomic positions to relax—and fit an appropriate equation-of-state (EOS).[59,60] The bulk modulus of S, at zero kelvin, is found from the curvature of the resulting $E(V)$ relation. More details can be found in Section SI.9.2.

PV curves are constructed at a finite temperature and explain the pressure response of S. In the elastic strain regime, we perform NPT MD over a grid of pressures in OpenMM to find the equilibrium volume under applied pressure $\langle V(P_{\text{ext}})\rangle$. In unstable PV regions, we switch to the $(N, V, \boldsymbol{\sigma_a} = \mathbf{0}, \boldsymbol{T})$ ensemble implemented in YAFF,[61] which constrains cell volume but allows its shape to vary freely, to recover the average internal pressure at a specified volume $\langle P_{\text{int}}(V)\rangle$. Under equilibrium conditions, both ensembles should agree and, when combined, describe the complete PV behavior. From a PV curve of S, we can deduce its vacuum equilibrium volume, its bulk modulus and the maximal pressure it can withstand before collapsing. Section SI.9.1 contains an in-depth explanation.

4. RESULTS

Following a theoretical exposition in Section 2, we apply our AL workflow to several spatially disordered MOFs belonging to the UiO type series and answer the following questions: (i) When

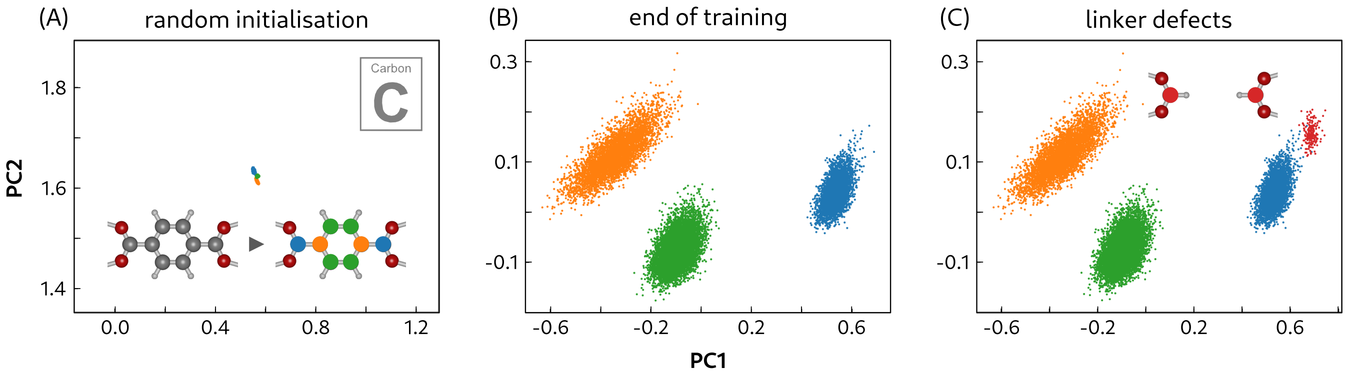


Figure 4. Feature embeddings of C in UiO-66 (D_{pr}) for a randomly initialized model (A) and the fully trained $\mathbf{mlp}_{\text{pr}}$ (B). In (C), $\mathcal{F}$-vectors of linker defects are superimposed on panel (B). Data points are color-coded according to their C-atom type.

do framework defects lead to large MLP inference inaccuracies? (ii) Can we avoid out-of-dataset extrapolation with isolated chemical environments in molecular fragments? (iii) Can cluster-based learning deliver transferable models for a family of disordered MOFs?

We select the prototypical zirconium MOF UiO-66(Zr) as the principal material from which defective structures will be derived. It consists of $Zr_6O_4(OH)_4$ inorganic bricks connected with 12 1,4-benzenedicarboxylate (BDC) ligands, to create a network of coordination bonds adopting the fcu topology.[62] UiO-66 is known for its tolerance to significant defect concentrations and is widely studied in both experimental and computational literature. This provides ample reference data regarding its mechanical behavior and the impact of various expressions of spatial disorder on framework properties.

From experiment, we identify three point defects that appear commonly in as-synthesized UiO-66, or that can be deliberately introduced with specialized synthesis protocols.[12,63,64] Linker defects are missing organic ligands that create small vacancies in the regular topological lattice. A metal substitution occurs when a chemically similar but different metal occupies an ionic atom site—conventionally hafnium or cerium for Zr-MOFs.[65] Lastly, a node defect is a large framework void caused by the absence of a brick and all surrounding linkers. Diffraction measurements show that such defects tend to appear in correlated nanodomains, forming local regions with reo topology.[66]

With these types of spatial disorder, we will construct disordered UiO-66 systems from the nanoscale (Section 4.3) to the mesoscale (Section 4.4) and demonstrate how one can attain high-accuracy MLPs. In Section 4.5, the superior model will be used to describe the mechanical pressure response of a handful of UiO-66 variants, uncovering important defect-property relations. In this work, we always restrict ourselves to periodic representations of bulk materials and do not consider any crystal surface phenomena. First, however, we investigate the nature of MLP feature space (Section 4.1) and discuss the design of molecular clusters (Section 4.2).

The following sections will compare multiple systems, data sets and MLPs. Section SI.10 provides an exhaustive overview for clarity.

4.1. Chemical Interpretation of MLP Feature Space.

In Section 2.3, we rely on a descriptor $\mathcal{F}$ to deduce a metric of model uncertainty for sample ϵ. The underlying assumption is that MLP feature space inherently contains physical or chemical information about the neighborhood of atoms. Here, we will show that this embedded space is indeed informative for atomic environments.

We choose a conventional 456-atom unit cell, containing four bricks and 24 linkers, to represent the pristine UiO-66(Zr) framework and name it S_{pr}. We generate a training (D_{pr}^{T}) and test (D_{pr}) data set, containing 200 and 100 configurations of S_{pr}, respectively. The superscript T will consistently refer to a training set. All structures are sampled uniformly in a volume range of 8500−9700 Å^3. We train our first model on D_{pr}^{T} and label it $\mathbf{mlp}_{\text{pr}}$. Table SI.1 details all important hyperparameters. The final layer of $\mathbf{mlp}_{\text{pr}}$ has eight scalar features. Accordingly, we examine the structure of an eight-dimensional $\mathcal{F}$-space.

We limit this discussion to ϵ and $\mathcal{F}$ of carbon atoms in UiO-66. First, we define a unique PCA reduction by extracting all C $\mathcal{F}$-vectors of $D_{\text{pr}}(\epsilon)$ using $\mathbf{mlp}_{\text{pr}}$. Every data point shown in Figure 4 is projected on its two largest principal components. Note that a 2D representation is less informative than the original 8D features, and is only chosen for visualization purposes.

Figure 4A shows the $\mathcal{F}$-embeddings of $D_{\text{pr}}(\epsilon)$, evaluated using a randomly initialized (i.e., untrained) checkpoint of $\mathbf{mlp}_{\text{pr}}$. The scale of this plot matches Figure 4B, which illustrates $D_{\text{pr}}(\epsilon)$ according to (the final checkpoint of) $\mathbf{mlp}_{\text{pr}}$. Initially, the model projects all ϵ to a small region of $\mathcal{F}$-space. Throughout the training procedure, it learns to spread and separate ϵ to better reproduce the atomic interactions of D_{pr}^{T}. Based on first-neighbor chemical intuition, we find 3 types of C in the BDC linker (see inset in Figure 4A). We color-code every $\mathcal{F}_i$ per carbon type of atom i and observe that $\mathbf{mlp}_{\text{pr}}$ groups ϵ in the same manner. The division is never imposed on the model, hence it must have learned to encode this chemistry in its features.

In the next step, we introduce a defect in S_{pr} by replacing a linker with two formate capping groups. For a set of defective structures, $\mathbf{mlp}_{\text{pr}}$ produces a fourth $\mathcal{F}$-cloud (colored red in Figure 4C), distinct from existing ones, indicating a new type of carbon ϵ. The model has never encountered formate during training, but successfully distinguishes these ϵ from regular BDC carbon atoms, proving its feature space can recognize unknown ϵ.

We stress that MLP $\mathcal{F}$-space is highly nonlinear and model-specific, i.e., different models can give strongly divergent embeddings. As such, attaching concrete chemical meaning to various regions of feature space is not really viable. Nevertheless, qualitative characteristics—like the point cloud grouping in Figure 4—emerge naturally with model optimization and conform to our intuition of different chemical environments.

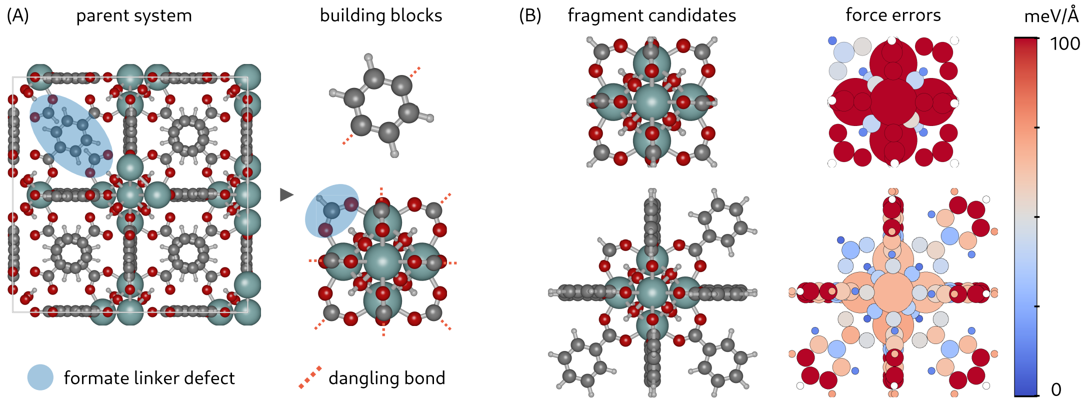


Figure 5. Cluster extraction in UiO-66. (A) Unit cell representation with a single linker defect and its deconstruction into modular building blocks. (B) Force matching for two fragments of the Zr-brick. The color scale represents a per-atom RMSE between *ab initio* $(\boldsymbol{F}_i)_{\mathrm{bulk}}$ and $(\boldsymbol{F}_i)_{\mathrm{cluster}}$.

4.2. Cluster Extraction in UiO-66. Sections SI.5 and SI.6 thoroughly discuss how we design, select and extract finite molecular fragments in disordered UiO-66. The following paragraphs provide a concise summary of the force matching procedure for designing clusters in bulk frameworks. We use the unit cell S_{pr} with a single linker defect as a case study and partition it into compact and loosely connected blocks (see Figure 5A). With our choice, clusters can be constructed easily by recombining adjacent building blocks and terminating dangling bonds with simple H-atoms.

To extract a new ϵ from the "brick" block with one formate group, i.e., adjacent to the linker defect, we devise a suitable mantle of atoms to surround and pad the fragment core. Figure 5B shows two possible constructions: a minimal cluster (top) and a first-neighbor cluster (bottom). Following the force matching approach, we compare $(\boldsymbol{F}_i)_{\mathrm{bulk}}$ and $(\boldsymbol{F}_i)_{\mathrm{cluster}}$ for a set of cells and extracted fragments (see Section SI.6) and observe unacceptably large force discrepancies for the minimal cluster; it is not suited to extract ϵ from the periodic system. On the other hand, the first-neighbor fragment captures bulk forces much more faithfully, while remaining computationally cheap. Therefore, it is selected to approximate $(\epsilon)_{\mathrm{bulk}}$.

Note that this fragment search was not exhaustive, and that our definition of building blocks is only valid for UiO-66-like materials. Other MOFs will require different rules to fragment the framework into separate blocks. However, we now possess a working cluster "blueprint"—a combination of core, mantle and termination—which can be reused across many disordered UiO-66-derived systems, provided no major structural changes occur (e.g., amorphization, severe topological rearrangements). We propose a general design rule: clusters given by a central core surrounded by its directly adjacent blocks provide an adequate approximation of $(\epsilon)_{\mathrm{bulk}}$ for our materials of interest. This hypothesis is validated a posteriori through MLP accuracy metrics.

4.3. Disorder in UiO-66 Unit Cells. As a first case study, we investigate how introducing single point defects alters atomic interactions in UiO-66(Zr). Finding a causal relation with the underlying change in ϵ can establish an informed pathway to construct representative data sets and train transferable MLPs. Starting from S_{pr}, we create three disordered systems by introducing a single linker defect (ld), a single hafnium substitution (hf) or a single node defect (reo). Although literature offers different hypotheses regarding the termination of bricks in the absence of coordination-bonded ligands,[66,67] we will consistently terminate linker defects with formate groups. The resulting periodic structures are depicted in Figure 6A and are referred to as S_{ld}, S_{hf}, and S_{reo}. Table 2 summarizes some naming abbreviations commonly used in the following sections.

Table 2. Summary of Used Abbreviations

pr	pristine or defect-free
ld	linker defect
hf	hafnium substitution
reo	reo node defect

Analogous to S_{pr}, we generate periodic training and test data sets for every disordered UiO-66 variant (D_{ld}^{T} and D_{ld}, D_{hf}^{T} and D_{hf}, D_{reo}^{T} and D_{reo}). Our baseline model, $\mathbf{mlp}_{\mathrm{pr}}$, performs excellently for D_{pr} with $\Delta E_{\mathrm{std}} < 0.4$ meV/atom and force RMSE < 26 meV/Å. However, it fails to correctly reproduce molecular interactions in the neighborhood of spatial defects for any other test set (see Figures 6C and SI.10 or Tables SI.4 and SI.5). Notably, model errors in framework regions free from disorder remain very low. The localized nature of these extreme force errors points to the existence of new $\epsilon \notin D_{\mathrm{pr}}^{T}(\epsilon)$. In UiO-66, the qualitative "area of effect" (AOE) of a linker defect, i.e., how many atoms feel the missing linker, is relatively small, followed by a hafnium substitution and a node defect, affecting almost the entire unit cell.

To address deficiencies in $D_{\mathrm{pr}}^{T}(\epsilon)$ and improve the transferability of $\mathbf{mlp}_{\mathrm{pr}}$, we employ the cluster-based learning methodology. The design principles of Section 2.4 readily identify three suitable molecular fragments that isolate and extract new ϵ from S_{ld}, S_{hf}, and S_{reo} (see Figure 6A). We initiate an AL campaign with $\mathbf{mlp}_{\mathrm{pr}}$ and the three disordered systems as learning targets. Every cycle, 50 walks are performed per system and 150 different cluster conformations are gathered to retrain the MLP. After several iterations, outlying force errors largely vanish and the new model successfully learns to describe simple defects. In total, we collect roughly 3000 fragments with 150–200 atoms each, following the examples in Figure 6A. As points of reference, we also train MLPs for each of the disordered unit cells: $\mathbf{mlp}_{\mathrm{ld}}$ on D_{ld}^{T}, $\mathbf{mlp}_{\mathrm{hf}}$ on D_{hf}^{T} and $\mathbf{mlp}_{\mathrm{reo}}$ on D_{reo}^{T}. These "periodic" models—trained solely on periodic structures—will be compared with "cluster" MLPs, which include nonperiodic training data as well. We refer to Table SI.9 for a full overview of all models and their training sets.

We construct a learning curve for D_{ld}, D_{hf}, and D_{reo} to systematically study model accuracy with varying amounts of cluster training data in Figure 6B. First, all extracted fragments

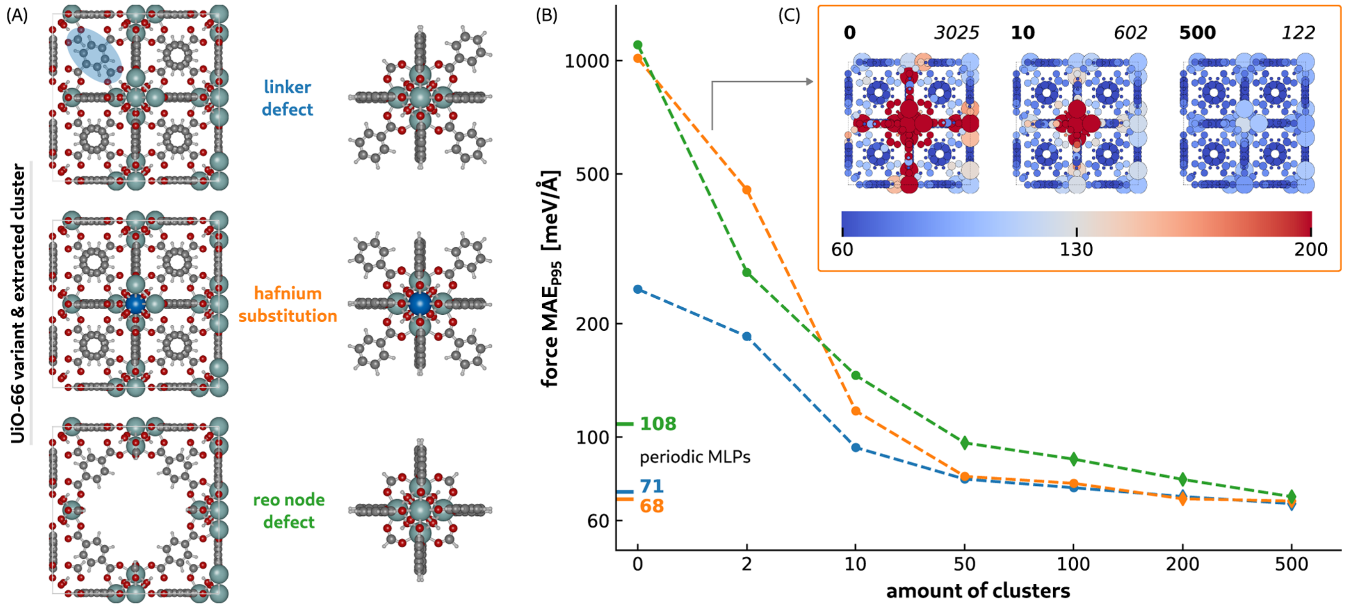


Figure 6. Learning point defects in a UiO-66 unit cell. (A) An overview of S_{ld}, S_{hf}, and S_{reo} and a matching molecular fragment to capture new ϵ. (B) Learning curves showing force MAE$_{P95}$ versus N for D_{ld}, D_{hf}, and D_{reo} (see main text). When a curve undercuts its periodic counterpart, the marker changes from a dot to a diamond (see vertical axis). (C) A per-atom visualization of force MAE$_{P95}$ for snapshots of the D_{hf} learning curve. Above each cell, N is indicated in bold, and the Hf atom force error is in cursive.

from the AL campaign are amassed into a database, from which we sample random subsets of size N (between 2 and 500) for ld, hf and reo defects separately. Every cluster subset is combined with D^{T}_{pr} to retrain **mlp**$_{\mathrm{pr}}$, resulting in a single data point. Figure 6B shows the full learning curves and plots force MAE$_{P95}$ (see Section 3) on a logarithmic scale versus N—the number of clusters added to D^{T}_{pr}. All curves start from **mlp**$_{\mathrm{pr}}$ (N = 0) and gradually incorporate more clusters of the corresponding defect type. Improvements in model accuracy are a direct consequence of the newly included ϵ. The MAE$_{P95}$ metrics of periodic models (**mlp**$_{\mathrm{ld}}$, **mlp**$_{\mathrm{hf}}$, and **mlp**$_{\mathrm{reo}}$) are indicated on the vertical axis of Figure 6B with colored dashes.

We observe that **mlp**$_{\mathrm{reo}}$ performs strikingly worse than **mlp**$_{\mathrm{ld}}$ and **mlp**$_{\mathrm{hf}}$ on their respective data set. Describing the large void created by a reo defect correctly might be intrinsically more difficult.[15] Node defects have a large AOE: many ϵ contribute to the measured error. S_{reo} also contains significantly fewer atoms, making D^{T}_{reo} the smallest data set. This is counterbalanced by the fact that most ϵ in S_{reo} contain useful information about the missing node (compare with the AOE of a linker defect in S_{ld}). Increasing network complexity and r_{max} results in a more performant model for the same D^{T}_{reo}. However, we use model error as a proxy to compare and find flaws in $D^{T}(\epsilon)$. Therefore, we must eliminate extraneous variables and keep the NequIP hyperparameter configuration fixed for all models.

In Figure 6B, **mlp**$_{\mathrm{pr}}$ is markedly more inaccurate for D_{hf} and D_{reo} than for D_{ld}. Large errors for D_{hf} are unsurprising; the model has never encountered the element Hf and will guess randomly in its vicinity. The difference between D_{reo} and D_{ld} can be understood from their relative AOE, i.e., missing linkers are much more local. Every learning curve quickly surpasses its matching periodic MLP—after 50 clusters for S_{reo} and after 200 clusters for S_{ld} and S_{hf}. This proves that the fragments of Figure 6A properly capture relevant ϵ from their parent systems. We observe the biggest improvements in accuracy for low N (<50). Including additional clusters leads to diminishing returns, a pattern that is expected in machine learning.[68]

Figure 6C provides a per-atom visualization of force MAE$_{P95}$ for D_{hf} along three points of its learning curve, $N \in \{0, 10, 500\}$ (indicated in bold). At N = 0, large force errors coalesce in a sizable sphere centered on the Hf substitution. As more training clusters are added, the sphere steadily shrinks in radius and error magnitudes decrease. MAE$_{P95}$ values for the single Hf atom (reported in cursive) drop from a massive 3025 meV/Å to just 122 meV/Å. Still, the average MAE$_{P95}$ for Zr atoms is only 97 meV/Å, owing to the Hf/Zr ratio in the final data set. Similar observations hold for linker and node defects (see Figure SI.10 and Table SI.10 for average error values).

From these results, we conclude that adapting an existing MLP, **mlp**$_{\mathrm{pr}}$, with a modest amount of clusters is a viable alternative to training new periodic models from scratch. Moreover, cluster-based learning can lower the *ab initio* computational cost of data set generation significantly—roughly by a factor of 5 in this case—without forfeiting model accuracy. This advantage will amplify as periodic cells grow larger or the level of theory becomes more expensive.

Up to now, we have only evaluated models on their specific defect type (ld, hf, and reo). In Section SI.8, we cross-validate MLPs and test data sets to uncover relations between different kinds of spatial disorder. We find that force errors are more sensitive to MLP extrapolation than energy errors, and should be preferred when trying to identify missing interactions. In inference, "cluster" models for S_{ld} and S_{reo} perform very similarly, indicating a strong correspondence in ϵ for linker and reo defects. The most transferable model, named **mlp**$^{c}_{\mathrm{mix}}$, is obtained by combining all three cluster types (see Table SI.9). A superscript c indicates the MLP is trained using clusters, alongside the basic D^{T}_{pr} data set. Energy errors generally decompose into small ΔE_{std} and enormous ΔE_{avg} values. For most models, the absolute energy scale is clearly wrong, although relative energy differences are captured adequately. Only **mlp**$^{c}_{\mathrm{mix}}$ manages to accurately predict absolute energies for all test sets, owing to the compositional variety of its training data.

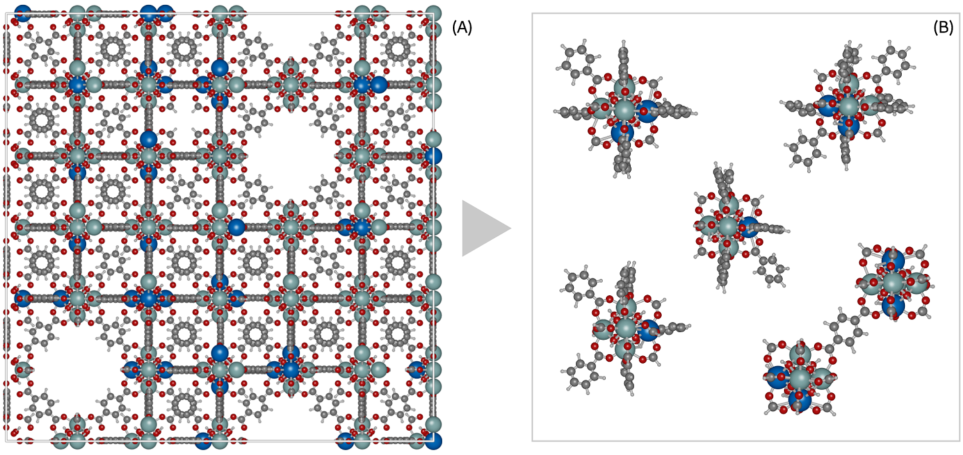


Figure 7. Spatial disorder in UiO-66 at the mesoscale. (A) An example 3 × 3 × 1 supercell similar to S_{sup}. (B) A handful of clusters extracted from S_{sup}.

4.4. Disorder in a UiO-66 Supercell.

In Section 4.3, we investigated spatial disorder as individual point defects in UiO-66 unit cells. So far, periodic correlations heavily limited the available configurational freedom of atoms and disorder, and hence the explorable ϵ-space. To more closely approximate realistic frameworks, an ensemble of missing linkers, metal substitutions and node defects should be considered. We generate a highly disordered UiO-66 system, starting from a pristine 4 × 4 × 4 supercell (29,184 atoms), by randomly removing 20% of linkers, replacing 20% of Zr atoms with Hf and creating node defects from 10% of bricks and surrounding ligands. Dangling bonds are saturated by hydrogen to form formate groups (analogous to Section 4.3) and unconnected building blocks are discarded from the remaining framework. The resulting structure contains 22,052 atoms and will be labeled S_{sup}. Since the distribution of disorder is completely random and the concentration of defects is very substantial, we expect to find a wide variety of previously unexplored environments, such as fully mixed Zr−Hf bricks and mesoporous channels caused by adjacent reo cavities. A 3 × 3 × 1 example cell, identically constructed to S_{sup}, is shown in Figure 7A.

We commence an AL campaign to assemble representative data for S_{sup}. The initial training set remains D^{T}_{pr}—i.e., the pristine 456-atom unit cell—to ensure consistency with previous models. However, we expand the MLP architecture of $\mathbf{mlp}_{\mathrm{pr}}$ toward more internal parameters and an increased r_{max} (see Section SI.2), anticipating a vastly enlarged and more complex $S_{\mathrm{sup}}(\epsilon)|_{\mathrm{ENS}}$. In every AL iteration, we perform 2 MD simulations and collect roughly 150 fragments, extracted following the design rules from Section 4.2. Figure 7B depicts a handful of examples, showing significantly more compositional and structural variety than Figure 6A. After 14 AL rounds, gathering about 2000 clusters of 150 atoms on average, we select a subset of 1500 structures (in analogy with $\mathbf{mlp}^{c}_{\mathrm{mix}}$, see Table SI.9), retrain the AL model and name it $\mathbf{mlp}^{c}_{\mathrm{sup}}$.

Because *ab initio* calculations of S_{sup} are computationally infeasible, we resort to our unit cell test sets to evaluate model performance. Tables 3 and SI.11 show virtually identical energy errors between $\mathbf{mlp}^{c}_{\mathrm{mix}}$ and $\mathbf{mlp}^{c}_{\mathrm{sup}}$; we appear to have reached the accuracy limit for energy predictions with our DFT/MLP configuration. In contrast, force RMSE and MAE$_{P95}$ improve by roughly 13−24% depending on the data set and metric. By construction, both models differ in hyperparameters and training data. To isolate the effects of each difference, we train a final MLP with the architecture of one model ($\mathbf{mlp}^{c}_{\mathrm{mix}}$) and the data set of the other ($\mathbf{mlp}^{c}_{\mathrm{sup}}$). It still surpasses $\mathbf{mlp}^{c}_{\mathrm{mix}}$ in force inference, although relative improvement shrinks to 0−9% (see $\mathbf{mlp}^{c}_{\mathrm{sup*}}$ in Table SI.11). We conclude that the superior accuracy of $\mathbf{mlp}^{c}_{\mathrm{sup}}$ is caused in part by a more expressive data set, but mostly by a larger network size. Our best model for D_{ld}, D_{hf}, and D_{reo} is derived from clusters that were not extracted from configurations of S_{ld}, S_{hf} or S_{reo}. The most representative data set contains the greatest diversity in ϵ, regardless of the fragments' parent system (compare Figures 6A and 7B).

Table 3. Validation Metrics of $\mathbf{mlp}^{c}_{\mathrm{sup}}$ for Every Test Dataset from Section 4.3

	D_{pr}	D_{ld}	D_{hf}	D_{reo}
ΔE_{avg}[a]	−0.1	0.2	0.0	0.4
ΔE_{std}[a]	0.3	0.5	0.4	0.6
RMSE[b]	19.8	18.9	19.1	20.0
MAE$_{P95}$[b]	56.0	53.6	53.7	55.3

[a] [meV/atom]. [b] [meV/Å].

Validation metrics on simple point defects might not generalize to arbitrarily disordered frameworks. As a substitute for S_{sup}, we construct a cluster test set D_{cl} containing 500 fragments, newly sampled from MD simulations. In Table 4, $\mathbf{mlp}^{c}_{\mathrm{sup}}$ outperforms $\mathbf{mlp}^{c}_{\mathrm{mix}}$ in all error statistics, and more convincingly than in Table 3. Most surprising, however, is the robustness of $\mathbf{mlp}^{c}_{\mathrm{mix}}$ for (out-of-dataset) clusters. This model never learned interactions between defects during training—e.g., multiple hf and ld defects in a brick—yet remains impressively accurate on D_{cl}. A posteriori, we discover that $\mathbf{mlp}^{c}_{\mathrm{mix}}$ can describe the PES of S_{sup} decently well, or equivalently, that all essential ϵ in S_{sup} could be learned with clusters from S_{ld}, S_{hf}, and S_{reo}. Based on isolated Hf substitutions, it inferred that Zr and Hf serve a similar role in UiO-66. Nevertheless, force RMSE values for Zr and Hf atoms are 44.5

Table 4. Validation Metrics of $\mathbf{mlp}^{c}_{\mathrm{mix}}$ and $\mathbf{mlp}^{c}_{\mathrm{sup}}$ for D_{cl}

	$\mathbf{mlp}^{c}_{\mathrm{mix}}$	$\mathbf{mlp}^{c}_{\mathrm{sup}}$
ΔE_{avg}[a]	0.3	0.0
ΔE_{std}[a]	0.5	0.3
RMSE[b]	27.6	19.1
MAE$_{P95}$[b]	83.2	52.8

[a] [meV/atom]. [b] [meV/Å].

and 68.7 meV/Å for $\mathbf{mlp}^{c}_{\mathrm{mix}}$, and 28.9 and 29.3 meV/Å for $\mathbf{mlp}^{c}_{\mathrm{sup}}$, indicating that (interactions of) metal ions limit overall accuracy in $\mathbf{mlp}^{c}_{\mathrm{mix}}$. Note that D_{cl} error metrics are only indicative of performance for the original periodic system S_{sup}. Molecular clusters should capture interactions from their parent, but this is difficult to verify at the mesoscale.

Combining conclusions from Sections 4.3 and 4.4, we summarize: for an MLP and training set D^T, (i) defects or interactions wholly absent from D^T cause huge local force errors (Figure 6B), (ii) combinations of defects are likely not problematic if every type of disorder is contained in D^T separately (Table 4), and (iii) the most accurate and transferable model is trained from the most extensive $D^T(\epsilon)$. Constructing such $D^T(\epsilon)$ can be challenging and requires trial-and-error or hands-on experience with the material of interest. Our cluster-based learning algorithm abstracts the required know-how and delivers representative data sets for arbitrarily disordered systems.

4.5. Mechanical Properties of Disordered UiO-66 Species.

Creating performant atomic potentials is only the first step in uncovering structure−property relations of materials. In this section, we will employ $\mathbf{mlp}^{c}_{\mathrm{sup}}$ to investigate the mechanical behavior of several spatially disordered UiO-66 species, which also serves as a first order validation to show reliable MLP predictions of derived properties. Starting from the pristine UiO-66(Zr) unit cell, S_{pr}, we systematically incorporate higher concentrations of disorder. S_{ld} is created by introducing a single linker defect (see Section 4.3). We form cells with an average brick coordination number of 11 by removing two ligands. Rogge et al. showed that the crystal symmetry of UiO-66 allows for seven physically distinct configurations of a double linker defect (see Figure SI.8); each with its own set of properties.[69] We will not fixate on an in-depth comparison between these variants, hence, refer to them collectively as "ld-2" systems ($S^{1-7}_{\mathrm{ld}-2}$). Removing a third ligand would generate a combinatorially exploding number of new structures. Instead, we will focus on three framework topologies recently observed in transmission electron microscopy experiments (see Figure SI.9).[70] A bcu network, S_{bcu}, is obtained by removing all linkers in planes perpendicular to a chosen cell axis (X, Y, or Z), leaving four 8-connected bricks and 16 linkers. We make S_{reo}, with three 8-connected bricks and 12 linkers, using a reo-type node defect. Superimposing the defects of S_{bcu} and S_{reo} results in the scu topology, S_{scu}, characterized by one 8-connected and two 4-connected bricks held together with 8 linkers. Finally, we include a pristine UiO-66(Hf) unit cell as a reference point for hafnium-substituted materials ($S^{\mathrm{hf}}_{\mathrm{pr}}$).

Figure 8 shows the pressure-vs-volume behavior at room temperature (300 K) of all aforementioned systems, computed with $\mathbf{mlp}^{c}_{\mathrm{sup}}$ following the procedure in Section 3. Table 5 reports the bulk modulus K at equilibrium volume V_0 and PV maximum P_{max} for every curve. P_{max} is the maximal pressure a system can withstand before collapsing into an unstable PV branch. It coincides with a significant drop in internal symmetry for MOFs and is sometimes called the loss-of-crystallinity pressure.[69] At lower cell volumes, we expect to recover another stable branch corresponding with the compression of an amorphous framework. Note that the interpretation of K under anisotropic strains is not straightforward. Nevertheless, it forms a starting point to compare the mechanical properties of our systems with earlier computational and experimental predictions.

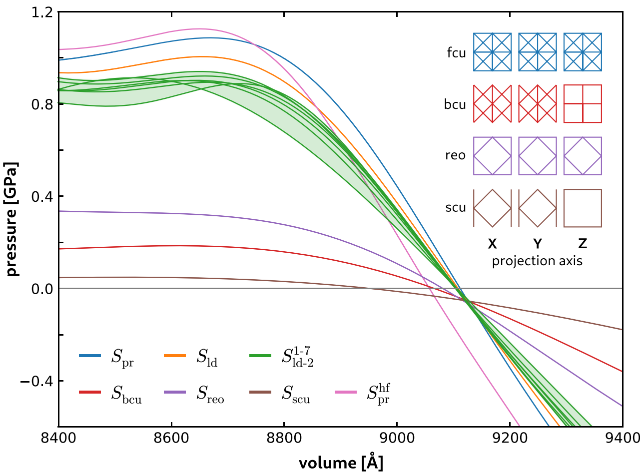


Figure 8. Pressure-vs-volume curves computed for a diverse set of UiO-66-derived systems (see main text). Frameworks with double linker defects ($S^{1-7}_{\mathrm{ld}-2}$) are merged in green. The inset shows structural representations of S_{pr}, S_{bcu}, S_{reo}, and S_{scu}, coarse-graining linkers to lines and omitting bricks. The 3D networks are projected along principal cell axes to highlight differences in topology and linker connectivity.

Table 5. Bulk Moduli and Loss-of-Crystallinity Pressures Derived from Figure 8[a]

[GPa]	K	P_{max}
S_{pr}	37	1.09
S_{ld}	32	1.01
$S^{1-7}_{\mathrm{ld}-2}$	23−30	0.89−0.94
S_{bcu}	8	0.19
S_{reo}	13	>0.34
S_{scu}	2	0.05
$S^{\mathrm{hf}}_{\mathrm{pr}}$	36	1.13

[a] For $S^{1-7}_{\mathrm{ld}-2}$, the reported values represent an aggregated range.

For S_{pr}, we find a bulk modulus of 37 GPA, which is in good agreement with static *ab initio* predictions of 41−42 GPa,[71,72] and in even better agreement with an experimental study that found 37.9 GPa using in situ synchrotron X-ray powder diffraction.[73] In the same experiment, V_0 was measured at 9009 Å^3, which we overshoot by roughly 100 Å^3 (±1%)—a known consequence of the PBE functional approximation.[74] Our simulations suggest that pristine UiO-66 will collapse under hydrostatic pressures above 1.1 GPa. From literature, one expects mechanical resilience to decrease when incorporating (linker) defects into MOFs.[75−77] However, interpolating consistent quantitative results rather than qualitative trends across various sources is difficult. Empirical amorphization seems particularly troublesome in this regard, as a measurable analogue of P_{max} is hard to define. Figure 8 predicts a modest drop in both K and P_{max} for S_{ld}. We find bulk moduli between 23 and 30 GPa and loss-of-crystallinity pressures around 0.89−0.94 GPa for the various $S^{1-7}_{\mathrm{ld}-2}$ systems, emphasizing that material characteristics are governed by the concentration as well as the distribution of spatial disorder. In Section SI.9.3, we compare our findings with earlier work by Rogge et al.,[69] which employs system-specific force fields parametrized through QuickFF,[78] for the systems discussed so far. We consistently predict larger bulk moduli and smaller loss-of-crystallinity pressures, but recover a robust linear relation for K between both LOTs.

Alterations in framework topology induce drastic changes in mechanical properties. Figure 8 shows a clear reduction in V_0 and K following the order S_{pr} (*fcu*) → S_{reo} → S_{bcu} → S_{scu}, which seems surprising at first, given that S_{bcu} retains more building blocks of the original fcu cell than S_{reo}. To explain this behavior, the inset in Figure 8 illustrates a schematic depiction of each topology by projecting its nodes and ligands along every cell axis. The atomistic models for S_{bcu}, S_{reo} and S_{scu} are compared with their complete 3D topological networks in Figure SI.9. These structural representations show that the asymmetric removal of ligands creates a weak crystal axis in S_{bcu} and S_{scu}, i.e., they will compress more easily in the XY-plane and elongate in the Z-direction under hydrostatic pressure. On the contrary, S_{reo} is symmetric in all three major axes, meaning it has no preferred direction of strain. In terms of P_{max}, both S_{bcu} and S_{scu} show a slight maximum in the considered volume range, whereas S_{reo} does not; its PV curve keeps rising steadily as the cell shrinks to volumes where MLP accuracy can no longer be assumed. Moreover, it exhibits (at least) two linear stable branches separated by a transitionary region between 8700 and 9000 Å^3, potentially indicating the existence of multiple stable phases. At present, we can only estimate a lower bound on P_{max} for S_{reo} (>340 MPa). In Section SI.9.4, we construct EV profiles for all topologies to uncover major structural changes that occur under compression. Our analysis indicates that cells of S_{pr} and S_{reo} compress through a collective rotation of building blocks and buckling of ligands, whereas S_{bcu} and S_{scu} undergo a limited reorientation of coordination bonds as a shearing strain. These distinct deformation mechanisms can explain the differences in mechanical behavior of Figure 8.

We conclude with UiO-66(Hf), S_{pr}^{hf}, for which DFT calculations predict a bulk modulus of 39.5 GPa and experimental measurements find V_0 and K estimates of 8906 Å^3 and 37 GPa, respectively.[71,72] Both values are marginally smaller than those for the Zr framework; a trend our simulations reproduce. Additionally, P_{max} is only slightly larger. We expect the mechanical properties of Zr–Hf UiO-66 mixtures to not deviate strongly from those observed for single-metal frameworks, which was already concluded in a recent force field study.[65]

4.5.1. Mechanical Behavior of Supercells. In the discussion above, we restricted MOF systems to unit cell dimensions. However, we ultimately aim to describe realistic and disordered frameworks at the mesoscale. Here, we consider three 4 × 4 × 4 supercells of UiO-66 containing over 20k atoms, distinguished by their concentration of spatial disorder: an upscaled version of S_{pr} ($S_{pr}^{4\times4\times4}$), S_{sup}, and a third system with roughly half the defects of S_{sup}—named S_{sup}*. Disorder is introduced at random. We adopt a simple characterization scheme (c_{hf}−c_{brick}−c_{linker}), where c_{hf} denotes the fraction of metal sites occupied by Hf atoms, and c_{brick} and c_{linker} represent the ratio of missing bricks and linkers compared to a defect-free system. With this convention, $S_{pr}^{4\times4\times4}$, S_{sup}* and S_{sup} correspond with (0−0−0), (0.10−0.04−0.17) and (0.21−0.13−0.40) supercells, respectively.

Because simulating PV profiles becomes quite costly at these length scales, Figure 9 shows EV curves for every system, evaluated with $\mathbf{mlp}_{sup}^{c}$ and baseline shifted to remove energy offsets (see Section 3). As the degree of disorder increases, average brick coordination numbers lower from 12 to 10.4 and 8.3. Hand-in-hand, the associated bulk moduli drop 3-fold from 38.1 to 12.5 GPa, and we notice a limited amount of cell contraction (<0.2%), aligning with our earlier findings in unit cells (see Figure 8). For reference, EV profiles for S_{pr}, S_{reo}, S_{bcu}, and S_{scu} find K values of 38.1, 17.7, 10.7, and 4.9 GPa. We observe that S_{reo}, a (0−0.25−0.50) unit cell, has more node defects and a lower coordination number than any supercell considered, yet manages a remarkable resistance to compression. This impressive stability has been attributed to the correlated nature of defects in the reo topology.[76]

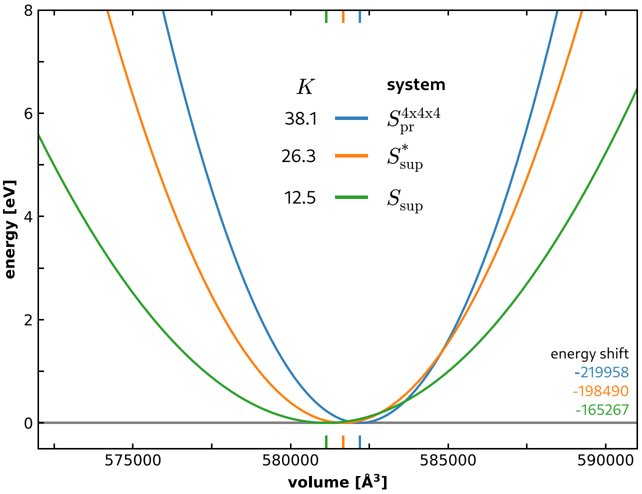


Figure 9. Energy-vs-volume curves for 4 × 4 × 4 UiO-66 supercells with varying degrees of spatial disorder (see main text), along with the corresponding energy baseline shifts, optimal cell volumes and bulk moduli (in GPa).

Note that our characterization says nothing about the distribution of spatial disorder. For statistically representative EV profiles, we would need to average over an appropriate ensemble of disordered MOFs with a fixed defect concentration (c_{hf}−c_{brick}−c_{linker}). While this is overtly out-of-scope in a proof-of-concept work, we have shown that cluster-based learning provides the toolbox necessary to tackle such investigations.

5. DISCUSSION

In this closing section, we reflect on the advantages and shortcomings of our methodology and discuss potential extensions and future research avenues.

Cluster-based learning enables MLP training for molecular systems at any length scale. It is almost fully automated, can be seamlessly integrated into modular AL workflows and delivers transferable models. Discrepancies between MLPs will inevitably propagate into differences of (mechanical) behavior between systems. The ability to describe all systems of interest with a single PES eliminates this source of inconsistency, allowing an apples-to-apples comparison of properties. Currently, defining appropriate clusters requires a fair amount of trial and error (see Section SI.6). In time, a set of base rules will be established that provides a workable initial guess for fragments, to be finetuned for specific use cases. Our implementation deconstructs frameworks like the UiO-66 series into discrete building blocks; but should be adapted for more complex topologies, such as winerack MOFs or even other classes of materials. Concerning uncertainty quantification, we will experiment with new combinations of feature descriptors $\mathcal{F}$ and density models (GMM, as of now) to strengthen the relation between ϵ-likelihood and MLP inference error.

To explore the behavior of realistic frameworks with spatial disorder, large molecular structures with a suitable concentration and composition of defects are needed. Experimentally, one cannot always probe the distribution of disorder, and we commonly assume that point defects are scattered homogeneously throughout the material. Many exceptions exist, however, e.g., correlated reo-defect nanodomains in UiO-66.[9,63,66] In Sections 4.4 and 4.5, we built test systems by randomly introducing point defects. While this naive approach suffices to collect a large variety of ϵ for model training, it will not generate representative structures resembling synthesized crystallites. As a first improvement, we could use Monte Carlo methods or the quasi-chemical approximation to create defective systems based on energetic and entropic grounds.[79−81] Besides structural defects, MOF topology is also known to greatly impact mechanical properties.[82] Given that a handful of building blocks can be assembled into entire families of different frameworks, we believe our method would be ideally suited to high-throughput screening studies that explore the topological landscape of MOFs.

Cluster-based learning is most powerful when exorbitant computational costs prohibit *ab initio* evaluations for a chosen system. Even high-performance computing infrastructures struggle with the computational requirements posed by post-Hartree−Fock methods or DFT functionals higher up on Jacob's ladder for all but the smallest systems. When studying MOFs, we usually resort to the GGA functionals. These approximations cannot describe London dispersion forces and only crudely capture electron correlations, leading to significant underbinding and deviations from real-world properties.[74] With clusters as training data, we can afford *ab initio* calculations at otherwise inaccessible LOTs, effectively bypassing the quantum scaling limit. In particular, Δ-learning workflows, in which MLPs are trained to predict a low-cost LOT and a higher order correction (e.g., PBE to MP2), could benefit from this approach and potentially reach chemical accuracy for MOFs.[83] A second accessible application is the study of populated frameworks. Molecular fragments can isolate guest molecules and their immediate surroundings from the bulk MOF, enabling MLPs to learn and describe diffusion or adsorption processes.

6. CONCLUSIONS

This work explores the fundamental relations between mechanical properties and spatial disorder for a series of UiO-66(Zr)-derived frameworks. We introduce the cluster-based learning methodology to develop robust MLPs at extended length scales. It identifies unknown chemical environments in sample structures through MLP feature space. These are extracted from the molecular bulk as compact fragments to extend the chemical space of atomic interactions covered by the model's training data.

We use this method to learn various point defects in small unit cells. Our investigation shows how to predict MLP extrapolation errors, how different types of spatial disorder are related and how to construct representative data sets that outperform conventionally trained models in accuracy and cost-effectiveness. We employ our procedure to successfully train a performant model from a strongly disordered 4 × 4 × 4 supercell containing over 20 thousand atoms. The major takeaway is that a greater variety of chemical environments in the training set delivers more accurate and transferable MLPs.

Using our leading model, we probe the pressure-versus-volume behavior of pristine UiO-66, disordered cells with up to two linker defects and three experimentally observed framework topologies. Finally, we extend our analysis to disordered supercells at the mesoscale, examining energy-versus-volume characteristics. These simulations highlight the impact of the concentration, composition and correlated nature of spatial disorder on framework properties.

We have shown that cluster-based learning enables the development of highly authentic MLPs by evaluating and learning atomic interactions in small clusters. Afterward, these models can be applied on larger disordered systems, unlocking the study of MOFs and spatial disorder at mesoscopic length scales, potentially including external surfaces.

■ ASSOCIATED CONTENT

Data Availability Statement

Data sets, trained models and further data are made available through a Zenodo repository at DOI: 10.5281/zenodo.14846185.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.chemmater.5c00821.

> *Ab initio* calculations in GPAW; NequIP architecture and training setup; molecular dynamics; active learning; cluster extraction; example force matching; MLP uncertainty and force errors; MLP cross-validation; mechanical characterization; overview of systems, data sets, and MLPs; and MLP accuracy and test metrics (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Veronique Van Speybroeck − *Center for Molecular Modeling, Ghent University, Technologiepark 46, 9052 Zwijnaarde, Belgium;* orcid.org/0000-0003-2206-178X;
Email: veronique.vanspeybroeck@ugent.be

Authors

Pieter Dobbelaere − *Center for Molecular Modeling, Ghent University, Technologiepark 46, 9052 Zwijnaarde, Belgium*
Sander Vandenhaute − *Center for Molecular Modeling, Ghent University, Technologiepark 46, 9052 Zwijnaarde, Belgium*

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.chemmater.5c00821

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

V.V.S. acknowledges funding from the Research Board of Ghent University (BOF) and the Fund for Scientific Research—Flanders (FWO) (grant nr: G0A8123N). P.D. and S.V. wish to thank the Fund for Scientific Research—Flanders (FWO) for aspirant doctoral fellowships (grant nos. 11O2125N and 11H6821N). The computational resources and services used in this work were provided by the VSC (Flemish Supercomputer Center), funded by the Research Foundation—Flanders (FWO) and the Flemish Government department EWI.

■ REFERENCES

(1) Lei, J.; Qian, R.; Ling, P.; Cui, L.; Ju, H. Design and sensing applications of metal−organic framework composites. *TrAC, Trends Anal. Chem.* **2014**, *58*, 71−78.

(2) Burtch, N. C.; Heinen, J.; Bennett, T. D.; Dubbeldam, D.; Allendorf, M. D. Mechanical Properties in Metal-Organic Frameworks: Emerging Opportunities and Challenges for Device Functionality and Technological Applications. *Adv. Mater.* **2018**, *30*, No. 1704124.

(3) Rogge, S. M. J.; Bavykina, A.; Hajek, J.; Garcia, H.; Olivos-Suarez, A. I.; Sepúlveda-Escribano, A.; Vimont, A.; Clet, G.; Bazin, P.; Kapteijn, F.; Daturi, M.; Ramos-Fernandez, E. V.; i Xamena, F. X. L.; Van Speybroeck, V.; Gascon, J. Metal-organic and covalent organic

frameworks as single-site catalysts. *Chem. Soc. Rev.* **2017**, *46*, 3134−3184.

(4) Temmerman, W.; Goeminne, R.; Rawat, K. S.; Van Speybroeck, V. Computational Modeling of Reticular Materials: The Past, the Present, and the Future. *Adv. Mater.* **2024**, No. 2412005.

(5) Yaghi, O. M.; O'Keeffe, M.; Ockwig, N. W.; Chae, H. K.; Eddaoudi, M.; Kim, J. Reticular synthesis and the design of new materials. *Nature* **2003**, *423*, 705−714.

(6) Furukawa, H.; Cordova, K. E.; O'Keeffe, M.; Yaghi, O. M. The Chemistry and Applications of Metal-Organic Frameworks. *Science* **2013**, *341*, No. 1230444.

(7) Bennett, T. D.; Cheetham, A. K.; Fuchs, A. H.; Coudert, F.-X. Interplay between defects, disorder and flexibility in metal−organic frameworks. *Nat. Chem.* **2017**, *9*, 11−16.

(8) Dissegna, S.; Epp, K.; Heinz, W. R.; Kieslich, G.; Fischer, R. A. Defective Metal-Organic Frameworks. *Adv. Mater.* **2018**, *30*, No. 1704501.

(9) Cheetham, A. K.; Bennett, T. D.; Coudert, F.-X.; Goodwin, A. L. Defects and disorder in metal organic frameworks. *Dalton Trans.* **2016**, *45*, 4113−4126.

(10) Dai, S.; Simms, C.; Patriarche, G.; Daturi, M.; Tissot, A.; Parac-Vogt, T. N.; Serre, C. Highly defective ultra-small tetravalent MOF nanocrystals. *Nat. Commun.* **2024**, *15*, No. 3434.

(11) Feng, Y.; Chen, Q.; Jiang, M.; Yao, J. Tailoring the Properties of UiO-66 through Defect Engineering: A Review. *Ind. Eng. Chem. Res.* **2019**, *58*, 17646−17659.

(12) Xiang, W.; Zhang, Y.; Chen, Y.; Liu, C.-j.; Tu, X. Synthesis, characterization and application of defective metal−organic frameworks: current status and perspectives. *J. Mater. Chem. A* **2020**, *8*, 21526−21546.

(13) Behler, J. Perspective: Machine learning potentials for atomistic simulations. *J. Chem. Phys.* **2016**, *145*, No. 170901.

(14) Mueller, T.; Hernandez, A.; Wang, C. Machine learning for interatomic potential models. *J. Chem. Phys.* **2020**, *152*, No. 050902.

(15) Behler, J.; Csányi, G. Machine learning potentials for extended systems: a perspective. *Eur. Phys. J. B* **2021**, *94*, No. 142.

(16) Friederich, P.; Häse, F.; Proppe, J.; Aspuru-Guzik, A. Machine-learned potentials for next-generation matter simulations. *Nat. Mater.* **2021**, *20*, 750−761.

(17) Morrow, J. D.; Gardner, J. L. A.; Deringer, V. L. How to validate machine-learned interatomic potentials. *J. Chem. Phys.* **2023**, *158*, No. 121501.

(18) Domina, M.; Patil, U.; Cobelli, M.; Sanvito, S. Cluster expansion constructed over Jacobi-Legendre polynomials for accurate force fields. *Phys. Rev. B* **2023**, *108*, No. 094102.

(19) Jia, W.; Wang, H.; Chen, M.; Lu, D.; Lin, L.; Car, R.; Weinan, E.; Zhang, L. In *Pushing the Limit of Molecular Dynamics with Ab Initio Accuracy to 100 Million Atoms with Machine Learning*, SC20: International Conference for High Performance Computing, Networking, Storage and Analysis; IEEE, 2020; pp 1−14.

(20) Xie, S. R.; Rupp, M.; Hennig, R. G. Ultra-fast interpretable machine-learning potentials. *npj Comput. Mater.* **2023**, *9*, No. 162.

(21) Kosanovich, K.; Gurumoorthy, A.; Sinzinger, E.; Piovoso, M. In *Improving the Extrapolation Capability of Neural Networks*, Proceedings of the 1996 IEEE International Symposium on Intelligent Control; IEEE, 1996; pp 390−395.

(22) Mishin, Y. Machine-learning interatomic potentials for materials science. *Acta Mater.* **2021**, *214*, No. 116980.

(23) Mahmoud, C. B.; El-Machachi, Z.; Gierczak, K. A.; Gardner, J. L. A.; Deringer, V. L. Assessing zero-shot generalisation behaviour in graph-neural-network interatomic potentials, arXiv:2502.21317. arXiv.org e-Print archive, 2025. https://arxiv.org/abs/2502.21317. (accessed March 31, 2025).

(24) Prodan, E.; Kohn, W. Nearsightedness of electronic matter. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 11635−11638.

(25) Ratcliff, L. E.; Dawson, W.; Fisicaro, G.; Caliste, D.; Mohr, S.; Degomme, A.; Videau, B.; Cristiglio, V.; Stella, M.; D'Alessandro, M.; Goedecker, S.; Nakajima, T.; Deutsch, T.; Genovese, L. Flexibilities of wavelets as a computational basis set for large-scale electronic structure calculations. *J. Chem. Phys.* **2020**, *152*, No. 194110.

(26) Prentice, J. C. A.; Aarons, J.; Womack, J.; et al. ONETEP linear-scaling density functional theory program. *J. Chem. Phys.* **2020**, *152*, No. 174111.

(27) Settles, B. *Active Learning Literature Survey*; Technical Report: 2009, p 1648.

(28) Jinnouchi, R.; Miwa, K.; Karsai, F.; Kresse, G.; Asahi, R. On-the-Fly Active Learning of Interatomic Potentials for Large-Scale Atomistic Simulations. *J. Phys. Chem. Lett.* **2020**, *11*, 6946−6955.

(29) Young, T. A.; Johnston-Wood, T.; Deringer, V. L.; Duarte, F. A transferable active-learning strategy for reactive molecular force fields. *Chem. Sci.* **2021**, *12*, 10944−10955.

(30) Sharma, A.; Sanvito, S. Quantum-accurate machine learning potentials for metal−organic frameworks using temperature driven active learning. *npj Comput. Mater.* **2024**, *10*, No. 237.

(31) Ang, S. J.; Wang, W.; Schwalbe-Koda, D.; Axelrod, S.; Gómez-Bombarelli, R. Active learning accelerates ab initio molecular dynamics on reactive energy surfaces. *Chem* **2021**, *7*, 738−751.

(32) Vandenhaute, S.; Cools-Ceuppens, M.; DeKeyser, S.; Verstraelen, T.; Van Speybroeck, V. Machine learning potentials for metal−organic frameworks using an incremental learning approach. *npj Comput. Mater.* **2023**, *9*, No. 19.

(33) Ying, P.; Zhang, J.; Zhong, Z. Pressure-induced phase transition of isoreticular MOFs: Mechanical instability due to ligand buckling. *Microporous Mesoporous Mater.* **2021**, *312*, No. 110765.

(34) Donà, L.; Brandenburg, J. G.; Bush, I. J.; Civalleri, B. Cost-effective composite methods for large-scale solid-state calculations. *Faraday Discuss.* **2020**, *224*, 292−308.

(35) Feng, L.; Wang, K.-Y.; Lv, X.-L.; Yan, T.-H.; Zhou, H.-C. Hierarchically porous metal−organic frameworks: synthetic strategies and applications. *Natl. Sci. Rev.* **2020**, *7*, 1743−1758.

(36) Anstine, D. M.; Isayev, O. Machine Learning Interatomic Potentials and Long-Range Physics. *J. Phys. Chem. A* **2023**, *127*, 2417−2431.

(37) Musil, F.; Grisafi, A.; Bartók, A. P.; Ortner, C.; Csányi, G.; Ceriotti, M. Physics-Inspired Structural Representations for Molecules and Materials. *Chem. Rev.* **2021**, *121*, 9759−9815.

(38) Domina, M.; Sanvito, S. A general formalism for machine-learning models based on multipolar-spherical harmonics, arXiv:2503.09618. arXiv.org e-Print archive. 2025 http://arxiv.org/abs/2503.09618. (accessed May 28, 2025).

(39) Zhang, L.; Lin, D.-Y.; Wang, H.; Car, R.; E, W. Active learning of uniformly accurate interatomic potentials for materials simulation. *Phys. Rev. Mater.* **2019**, *3*, No. 023804.

(40) Smith, J. S.; Nebgen, B.; Lubbers, N.; Isayev, O.; Roitberg, A. E. Less is more: Sampling chemical space with active learning. *J. Chem. Phys.* **2018**, *148*, No. 241733.

(41) Zhai, Y.; Caruso, A.; Gao, S.; Paesani, F. Active learning of many-body configuration space: Application to the Cs+-water MB-nrg potential energy function as a case study. *J. Chem. Phys.* **2020**, *152*, No. 144103.

(42) Vandermause, J.; Torrisi, S. B.; Batzner, S.; Xie, Y.; Sun, L.; Kolpak, A. M.; Kozinsky, B. On-the-fly active learning of interpretable Bayesian force fields for atomistic rare events. *npj Comput. Mater.* **2020**, *6*, No. 20.

(43) O'Neill, J.; Delany, S. J.; MacNamee, B. *Advances in Computational Intelligence Systems*; Springer International Publishing: Cham, 2017; Vol. *513*, pp 375−386.

(44) Zaverkin, V.; Holzmüller, D.; Steinwart, I.; Kästner, J. Exploring chemical and conformational spaces by batch mode deep active learning. *Digital Discovery* **2022**, *1*, 605−620.

(45) Tan, A. R.; Dietschreit, J. C. B.; Gomez-Bombarelli, R. Enhanced sampling of robust molecular datasets with uncertainty-based collective variables, arXiv:2402.03753. arXiv.org e-Print archive, 2024 https://arxiv.org/abs/2402.03753. (accessed May 8, 2024).

(46) Zhu, A.; Batzner, S.; Musaelian, A.; Kozinsky, B. Fast uncertainty estimates in deep learning interatomic potentials. *J. Chem. Phys.* **2023**, *158*, No. 164111.

(47) Moitra, A. *Algorithmic Aspects of Machine Learning*; Cambridge University Press, 2018; pp 107−131.

(48) Stoica, P.; Selen, Y. Model-order selection. *IEEE Signal Process. Mag.* **2004**, *21*, 36−47.

(49) Altman, N.; Krzywinski, M. The curse(s) of dimensionality. *Nat. Methods* **2018**, *15*, 399−400.

(50) Jolliffe, I. T.; Cadima, J. Principal component analysis: a review and recent developments. *Philos. Trans. R. Soc. A* **2016**, *374*, No. 20150202.

(51) Mortensen, J. J.; Hansen, L. B.; Jacobsen, K. W. Real-space grid implementation of the projector augmented wave method. *Phys. Rev. B* **2005**, *71*, No. 035109.

(52) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, No. 3865.

(53) Johnson, E. R.; Becke, A. D. A post-Hartree-Fock model of intermolecular interactions: Inclusion of higher-order corrections. *J. Chem. Phys.* **2006**, *124*, No. 174104.

(54) Batzner, S.; Musaelian, A.; Sun, L.; Geiger, M.; Mailoa, J. P.; Kornbluth, M.; Molinari, N.; Smidt, T. E.; Kozinsky, B. E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nat. Commun.* **2022**, *13*, No. 2453.

(55) Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L.-P.; Simmonett, A. C.; Harrigan, M. P.; Stern, C. D.; Wiewiora, R. P.; Brooks, B. R.; Pande, V. S. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Comput. Biol.* **2017**, *13*, No. e1005659.

(56) Verstraelen, T.; Vanduyfhuys, L.; Vandenbrande, S.; Rogge, S. Yaff, yet another force field. 2013.

(57) Pedregosa, F. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825−2830.

(58) Vanpoucke, D. E. P.; Lejaeghere, K.; Van Speybroeck, V.; Waroquier, M.; Ghysels, A. Mechanical Properties from Periodic Plane Wave Quantum Mechanical Codes: The Challenge of the Flexible Nanoporous MIL-47(V) Framework. *J. Phys. Chem. C* **2015**, *119*, 23752−23766.

(59) Birch, F. Finite Elastic Strain of Cubic Crystals. *Phys. Rev.* **1947**, *71*, No. 809.

(60) Vinet, P.; Ferrante, J.; Rose, J. H.; Smith, J. R. Compressibility of solids. *J. Geophys. Res.: Solid Earth* **1987**, *92*, 9319−9325.

(61) Rogge, S.; Vanduyfhuys, L.; Ghysels, A.; Waroquier, M.; Verstraelen, T.; Maurin, G.; Van Speybroeck, V. A Comparison of Barostats for the Mechanical Characterization of Metal-Organic Frameworks. *J. Chem. Theory Comput.* **2015**, *11*, 5583−5597.

(62) Cavka, J. H.; Jakobsen, S.; Olsbye, U.; Guillou, N.; Lamberti, C.; Bordiga, S.; Lillerud, K. P. A New Zirconium Inorganic Building Brick Forming Metal Organic Frameworks with Exceptional Stability. *J. Am. Chem. Soc.* **2008**, *130*, 13850−13851.

(63) Meekel, E. G.; Goodwin, A. L. Correlated disorder in metal−organic frameworks. *CrystEngComm* **2021**, *23*, 2915−2922.

(64) Cao, Y.; Mi, X.; Li, X.; Wang, B. Defect Engineering in Metal-Organic Frameworks as Futuristic Options for Purification of Pollutants in an Aqueous Environment. *Front. Chem.* **2021**, *9*, No. 673738.

(65) Rogge, S. M. J.; Yot, P. G.; Jacobsen, J.; Muniz-Miranda, F.; Vandenbrande, S.; Gosch, J.; Ortiz, V.; Collings, I. E.; Devautour-Vinot, S.; Maurin, G.; Stock, N.; Van Speybroeck, V. Charting the Metal-Dependent High-Pressure Stability of Bimetallic UiO-66 Materials. *ACS Mater. Lett.* **2020**, *2*, 438−445.

(66) Cliffe, M. J.; Wan, W.; Zou, X.; Chater, P. A.; Kleppe, A. K.; Tucker, M. G.; Wilhelm, H.; Funnell, N. P.; Coudert, F.-X.; Goodwin, A. L. Correlated defect nanoregions in a metal−organic framework. *Nat. Commun.* **2014**, *5*, No. 4176.

(67) Vandichel, M.; Hajek, J.; Vermoortele, F.; Waroquier, M.; De Vos, D. E.; Van Speybroeck, V. Active site engineering in UiO-66 type metal−organic frameworks by intentional creation of defects: a theoretical rationalization. *CrystEngComm* **2015**, *17*, 395−406.

(68) Frey, N. C.; Soklaski, R.; Axelrod, S.; Samsi, S.; Gómez-Bombarelli, R.; Coley, C. W.; Gadepally, V. Neural scaling of deep chemical models. *Nat. Mach. Intell.* **2023**, *5*, 1297−1305.

(69) Rogge, S. M. J.; Wieme, J.; Vanduyfhuys, L.; Vandenbrande, S.; Maurin, G.; Verstraelen, T.; Waroquier, M.; Van Speybroeck, V. Thermodynamic Insight in the High-Pressure Behavior of UiO-66: Effect of Linker Defects and Linker Expansion. *Chem. Mater.* **2016**, *28*, 5721−5732.

(70) Liu, L.; Chen, Z.; Wang, J.; Zhang, D.; Zhu, Y.; Ling, S.; Huang, K.-W.; Belmabkhout, Y.; Adil, K.; Zhang, Y.; Slater, B.; Eddaoudi, M.; Han, Y. Imaging defects and their evolution in a metal−organic framework at sub-unit-cell resolution. *Nat. Chem.* **2019**, *11*, 622−628.

(71) Wu, H.; Yildirim, T.; Zhou, W. Exceptional Mechanical Stability of Highly Porous Zirconium Metal-Organic Framework UiO-66 and Its Important Implications. *J. Phys. Chem. Lett.* **2013**, *4*, 925−930.

(72) Redfern, L. R.; Ducamp, M.; Wasson, M. C.; Robison, L.; Son, F. A.; Coudert, F.-X.; Farha, O. K. Isolating the Role of the Node-Linker Bond in the Compression of UiO-66 Metal-Organic Frameworks. *Chem. Mater.* **2020**, *32*, 5864−5871.

(73) Redfern, L. R.; Robison, L.; Wasson, M. C.; Goswami, S.; Lyu, J.; Islamoglu, T.; Chapman, K. W.; Farha, O. K. Porosity Dependence of Compression and Lattice Rigidity in Metal-Organic Framework Series. *J. Am. Chem. Soc.* **2019**, *141*, 4365−4371.

(74) Lejaeghere, K.; Van Speybroeck, V.; Van Oost, G.; Cottenier, S. Error Estimates for Solid-State Density-Functional Theory Predictions: An Overview by Means of the Ground-State Elemental Crystals. *Crit. Rev. Solid State Mater. Sci.* **2014**, *39*, 1−24.

(75) Dissegna, S.; Vervoorts, P.; Hobday, C. L.; Düren, T.; Daisenberger, D.; Smith, A. J.; Fischer, R. A.; Kieslich, G. Tuning the Mechanical Response of Metal−Organic Frameworks by Defect Engineering. *J. Am. Chem. Soc.* **2018**, *140*, 11581−11584.

(76) Thornton, A. W.; Babarao, R.; Jain, A.; Trousselet, F.; Coudert, F.-X. Defects in metal−organic frameworks: a compromise between adsorption and stability? *Dalton Trans.* **2016**, *45*, 4352−4359.

(77) Vervoorts, P.; Stebani, J.; Méndez, A. S. J.; Kieslich, G. Structural Chemistry of Metal-Organic Frameworks under Hydrostatic Pressures. *ACS Mater. Lett.* **2021**, *3*, 1635−1651.

(78) Vanduyfhuys, L.; Vandenbrande, S.; Wieme, J.; Waroquier, M.; Verstraelen, T.; Van Speybroeck, V. Extension of the QuickFF force field protocol for an improved accuracy of structural, vibrational, mechanical and thermal properties of metal−organic frameworks. *J. Comput. Chem.* **2018**, *39*, 999−1011.

(79) Liu, X.; Zhang, J.; Yin, J.; Bi, S.; Eisenbach, M.; Wang, Y. Monte Carlo simulation of order-disorder transition in refractory high entropy alloys: A data-driven approach. *Comput. Mater. Sci.* **2021**, *187*, No. 110135.

(80) Brivio, F.; Caetano, C.; Walsh, A. Thermodynamic Origin of Photoinstability in the CH3NH3Pb(I1-xBrx)3 Hybrid Halide Perovskite Alloy. *J. Phys. Chem. Lett.* **2016**, *7*, 1083−1087.

(81) Sher, A.; Van Schilfgaarde, M.; Chen, A.-B.; Chen, W. Quasichemical approximation in binary alloys. *Phys. Rev. B* **1987**, *36*, No. 4279.

(82) Moghadam, P. Z.; Rogge, S. M.; Li, A.; Chow, C.-M.; Wieme, J.; Moharrami, N.; Aragones-Anglada, M.; Conduit, G.; Gomez-Gualdron, D. A.; Van Speybroeck, V.; Fairen-Jimenez, D. Structure-Mechanical Stability Relations of Metal-Organic Frameworks via Machine Learning. *Matter* **2019**, *1*, 219−234.

(83) Dral, P. O.; Owens, A.; Dral, A.; Csányi, G. Hierarchical machine learning of potential energy surfaces. *J. Chem. Phys.* **2020**, *152*, No. 204110.