npj | computational materials

Article

Published in partnership with the Shanghai Institute of Ceramics of the Chinese Academy of Sciences

https://doi.org/10.1038/s41524-025-01834-9

MP-ALOE: an r^{2}SCAN dataset for universal machine learning interatomic potentials

Check for updates

Matthew C. Kuner[1,2] ✉, Aaron D. Kaplan[2], Kristin A. Persson[1,2], Mark Asta[1,2] & Daryl C. Chrzan[1,2] ✉

We present MP-ALOE, a dataset of nearly 1 million DFT calculations using the accurate r^{2}SCAN meta-generalized gradient approximation. Covering 89 elements, MP-ALOE was created using active learning and primarily consists of off-equilibrium structures. We benchmark a machine learning interatomic potential trained on MP-ALOE, and evaluate its performance on a series of benchmarks, including predicting the thermochemical properties of equilibrium structures; predicting forces of far-from-equilibrium structures; maintaining physical soundness under static extreme deformations; and molecular dynamic stability under extreme temperatures and pressures. MP-ALOE shows strong performance on all of these benchmarks and is made public for the broader community to utilize.

Atomistic simulations are among the most common tools used by computational materials scientists. While ab initio density functional theory (DFT)[1] simulations show impressive accuracy relative to experiment due to adherence to physical constraints[2,3], they are often quite expensive. For solid-state systems specifically, modern plane-wave DFT[4] typically limits users to simulating, at most, hundreds of atoms and timescales on the order of picoseconds. Hence, more efficient methods are needed for calculating quantities that are numerically intensive (e.g., transition state searches) or that require large simulation cells (e.g., diffusion pathways or equilibration of non-crystalline materials).

Classical, empirical force fields for solids, such as the (modified) embedded atom method[5,6], allow users to perform dynamic simulations at low cost. However, the fixed functional forms of such classical force fields limit their accuracy and generalizability to narrow chemical regimes. Additionally, there are no classical force fields for solids that span almost the entire periodic table, nor can they adequately treat stretched bonds encountered in solid-state reactions. In recent years, machine-learning interatomic potentials (MLIPs)[7] have increasingly become a promising alternative. Such potentials allow a researcher to curate a set of ab initio calculations to train an MLIP, which can approximate the potential energy surface (PES) based on the training set alone. MLIPs are often implemented as Graph Neural Networks (GNNs), given that graphs are a natural way to represent the chemical bonding of atoms. Notable examples include refs. 8–12.

The apex of MLIP research would be to create a universal MLIP (UMLIP) that can accurately approximate a given DFT functional across the periodic table. The current generation of UMLIPs, beginning with M3GNet[13] and continuing with, e.g., refs. 13–18 cover most of the periodic table (at least 89 elements) and maintain close-to-linear scaling with the number of simulated atoms, vastly faster than DFT's cubic scaling with the number of *electrons*. However, modern pre-trained models are typically limited in accuracy and transferability: models trained primarily to equilibrium solid-state data tend to underestimate energy and forces in out-of-domain test cases[19]. Further, the accuracy of smaller models tends to exceed the uncertainty of experiment in predicting equilibrium compositional phase stability[20], whereas much larger models are less computationally tractable.

One key challenge in improving UMLIPs is increasing the quality of the underlying DFT data they are trained on. Most current UMLIPs are trained on DFT calculations at the Perdew-Burke-Ernzerhof (PBE) generalized gradient approximation (GGA) level of theory[21], sourced from the Materials Project (MP) database[22,23], the Alexandria database[24,25], and/or the OMat24 dataset[18]. While PBE is often accurate for describing simple *sp*-bonded solids, it struggles to describe weaker bonds present in mixed-/ionic and dispersion-bound solids[3], as well as systems plagued by delocalization errors[2], such as defects[26].

To date, only one solid-state dataset[27] has been calculated at a higher level of approximation, using the r^{2}SCAN meta-GGA[28]. Meta-GGAs typically improve systematically over GGAs like PBE[29,30], reducing their mean absolute errors (MAEs) in solid-state formation enthalpies from ~150 to ~100 meV/atom[31], and often perform comparably to or better than the higher hybrid levels of theory for calculation of equilibrium solid-state properties[31,32].

The MatPES dataset of ref. 27 – the only public r^{2}SCAN dataset for UMLIPs released before this work – showed strong results on a series of near- and off-equilibrium benchmarks, especially given its smaller size relative to other datasets. However, MatPES is still limited to relatively low-energy structures sampled from 300K molecular dynamics (MD) trajectories. Moreover, the chemical compositions (and initial structures for MD) included are exclusively sourced from the compounds in the Materials

[1]Department of Materials Science and Engineering, University of California, Berkeley, CA, USA. [2]Materials Science Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. ✉e-mail: matthewkuner@berkeley.edu; dcchrzan@berkeley.edu

Project. Although MatPES reflects the chemical environments of the Materials Project, MatPES-trained UMLIPs were able to reasonably capture the formation energies of solids whose compositions were not present in the dataset.

MatPES demonstrably samples a wider distribution of interatomic forces and stresses than the Materials Project relaxation trajectories, partly attributable to its source of structures (MD) and partly to its sampling method[33], which attempts to increase dissimilarity of the sampled structures. It is currently unclear how wide a force distribution is needed to ensure general stability for MD, and to stiffen the PES at larger interatomic separations[19]. Approaches such as active learning, applied either to training potentials in a limited chemical regime[34] or to the whole periodic table[17], can aid in systematically improving coverage of PES regions where a UMLIP has a low density of training data. Active learning approaches for UMLIPs currently rely on minimizing model uncertainty either via mathematical quantities derived from the model architecture[34] or from physical property-motivated uncertainty estimators[17].

In this work, we present a new r^{2}SCAN dataset for UMLIPs, which begins to address some of the previously mentioned open questions. This dataset was created via elemental substitution of prototype structures, and augmented using active learning (AL) via query by committee (QBC)[35]. We have named it MP-ALOE (**M**aterials **P**roject - **A**ctive **L**earning of **O**ff **E**quilibrium structures). The MP-ALOE dataset contains a greater sampling of high-energy structures, large magnitude forces, and high pressures than in MatPES. Moreover, given that MP-ALOE and MatPES were calculated using compatible DFT settings, we then perform a series of benchmarks comparing MACE[12] models trained on MP-ALOE, MatPES, and both of these datasets combined. Specifically, we look at predictions for equilibrium properties, off-equilibrium forces, static deformations under extreme hydrostatic pressure, and MD stability under extreme ensemble conditions. We find MP-ALOE- and MatPES-trained MACE models to be competitive in predicting equilibrium energies and off-equilibrium forces. Notably, the MP-ALOE-trained model demonstrates improved stability in MD runs and physicality of the PES under static extreme hydrostatic pressures. By construction, MP-ALOE is completely compatible with the r^{2}SCAN subset of MatPES; a MACE model trained on their union clearly demonstrates the strongest overall performance.

Results

The dataset

We start by detailing the MP-ALOE dataset itself. In total, 909,792 frames of DFT data (from 303,264 structure relaxations) are included in the present work; the workflow for their generation is shown in Fig. 1 (top). The distribution of atom counts by element for MP-ALOE is shown in Fig. 1 (bottom); elemental coverage appears to be quite reasonable. The high representation of oxygen is consistent with experimental databases like the ICSD[36], wherein more than half of the experimental inorganic structures contain oxygen.

The cohesive energies for the MP-ALOE dataset are shown in Fig. 2a, with the MatPES[27] dataset included as a reference. MP-ALOE has a higher mean cohesive energy (−3.65 eV/atom) than MatPES (−4.01 eV/atom); MP-ALOE also has a wider distribution of cohesive energies. This was expected, given that MatPES structures are sampled from 300K MD simulations of (largely) near-stable structures from the Materials Project Database[22,23], whereas MP-ALOE was mostly generated combinatorially from unknown hypothetical structures. 3949 structures (0.4 %) in the database have a positive cohesive energy, indicating instability. These structures can provide meaningful information to an ML potential by better defining regions of the potential energy surface which are energetically unfavorable. However, if desired, dataset users can easily remove these entries by inspecting the `cohesive_energy` key in each training document.

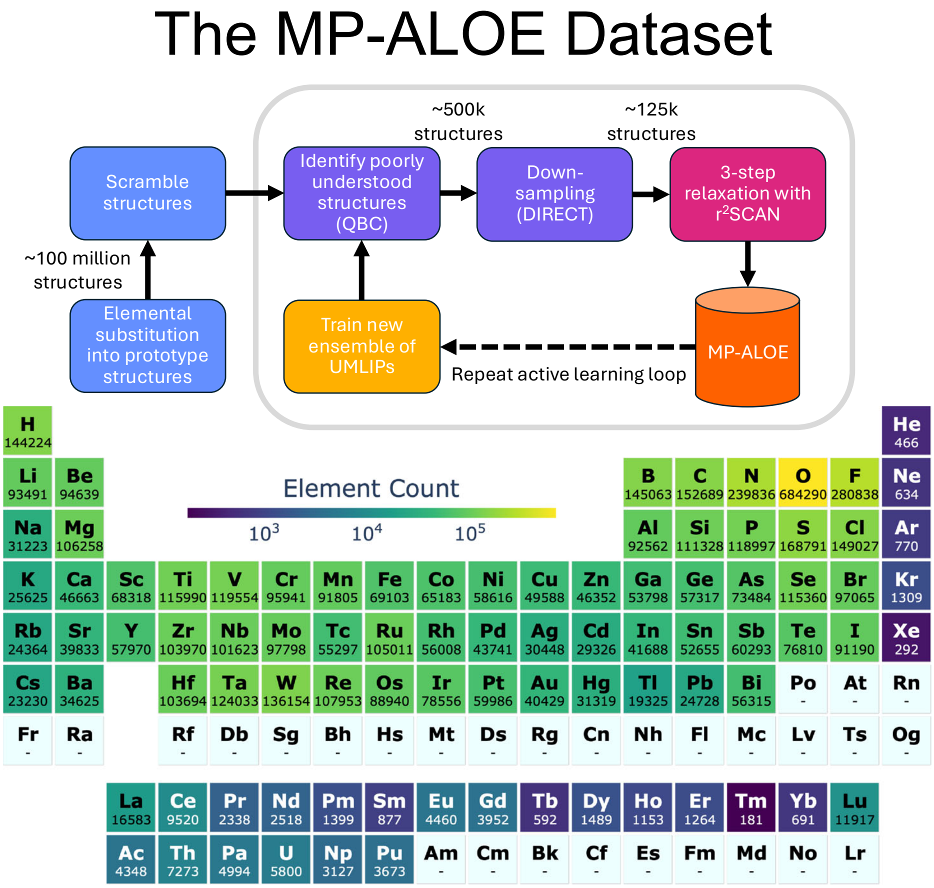

Fig. 1 | Overview of the MP-ALOE dataset. (top) The active learning workflow for generating MP-ALOE. ~100 million structures were generated via elemental substitution into prototype structures. Structures in poorly covered parts of the PES were selected via Query By Committee[35]; these structures were then downsampled using the DIRECT method[33]. The remaining structures were computed using the r^{2}SCAN functional[28]. This process was then repeated. Further details can be found in the Methods section. (bottom) The elemental distribution in MP-ALOE. Counts are the number of *atoms* of a given element. Visualized using the `pymatviz` package[57].

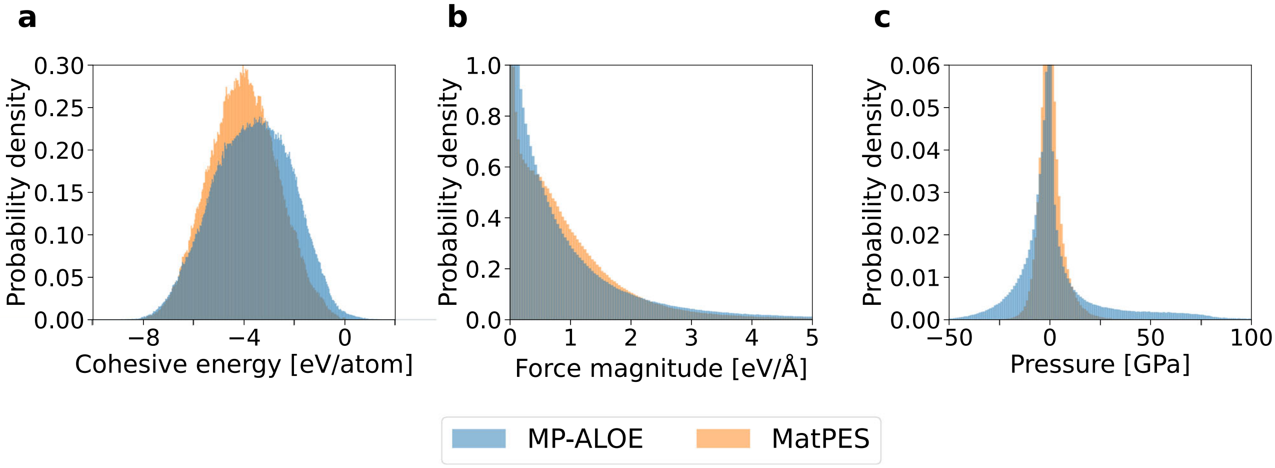

Fig. 2 | Describing the MP-ALOE dataset. Distribution of **a** cohesive energies (eV/atom), **b** interatomic force magnitudes (eV/Å), and **c** pressure as derived from one third of the trace of the DFT stress tensor (GPa) in the MP-ALOE and MatPES datasets (the only two public r^{2}SCAN datasets designed for UMLIPs currently available). A total of 3949 points (0.4%) in the MP-ALOE dataset have a positive cohesive energy; 3432 of which are explained via the presence of noble gases or the presence of two or more unique nonmetals. For the pressures, positive values correspond to compression.

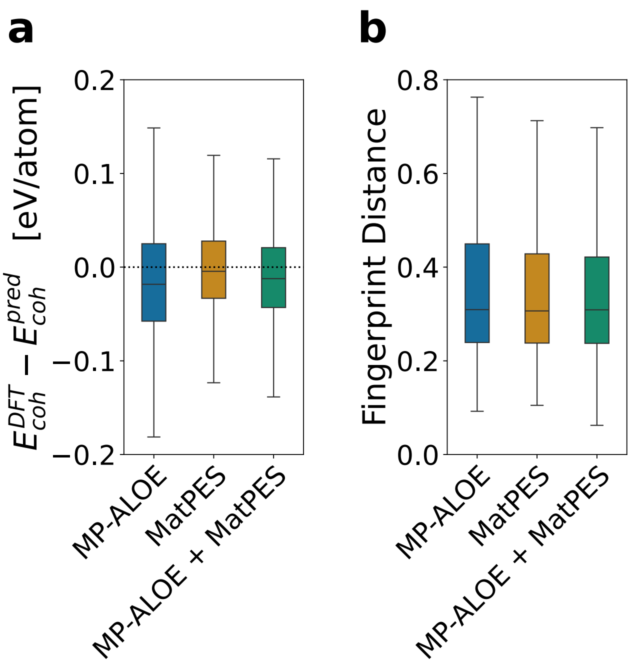

Fig. 3 | Benchmarking performance of UMLIPs at equilibrium. a Predictions of cohesive energies (eV/atom) and **b** fingerprint distances for equilibrium structures sourced from the WBM dataset[37]. Random atomic displacements were applied to DFT-relaxed structures, and then the scrambled structures were re-relaxed using the respective UMLIPs. Fingerprint distance is a measure of structural similarity between the DFT- and UMLIP-relaxed structures; smaller is better. The CrystalNN method used to calculate fingerprint distance is described in ref. 38, and further details of this test are described in ref. 27.

The distribution of forces in MP-ALOE is shown in Fig. 2b, with MatPES as a reference. Overall, the distributions are roughly comparable, with MP-ALOE having a slightly greater mean force (1.03 eV/Å) than MatPES (0.94 eV/Å). MP-ALOE contains a relatively smooth distribution of forces and contains a greater sampling of forces over 2 eV/Å relative to MatPES.

The distribution of pressures in MP-ALOE is shown in Fig. 2c, with MatPES as a reference. MP-ALOE has a significantly broader spread of pressures, with a notable proportion of pressures between −50 and 100 GPa (as opposed to MatPES, which mostly contains pressures between −20 and 30 GPa).

A full discussion of the limitations of the MP-ALOE is included in the Supplementary Information. To summarize, the MP-ALOE dataset is limited to small-cell ordered solids, and therefore, may not describe disordered, non-crystalline, defective, or quasi-non-periodic materials well without additional training data. Those who wish to use any database of bulk crystals to train force fields for, e.g., surfaces, are advised to determine if further training data is needed. By construction, the noble gas elements, actinides, and lanthanides have lower representation in MP-ALOE; the former category is mostly irrelevant for thermochemistry, and the latter two categories have complex electronic structure, which may not be well-captured by DFT, nor by atomistic potentials.

Additionally, further comparisons of MP-ALOE to the OMat24 dataset[18] (a notably diverse dataset calculated at the lower PBE level of theory) are included in the Supplementary Information (Fig. S6).

Benchmarking

Here we compare results between three MACE potentials: a potential trained only on MP-ALOE, a potential trained only on MatPES, and a potential trained on both of these datasets combined (MP-ALOE + MatPES). Details for how MP-ALOE and MatPES were combined can be found in the Methods section. We use the name of a dataset and the potential trained on that dataset interchangeably to increase readability.

Equilibrium benchmarks. Here, we compare the performance of MP-ALOE, MatPES, and the combined dataset potentials on predicting equilibrium properties. Specifically, approximately 1000 structures sourced from the WBM dataset[37] were relaxed using r^{2}SCAN as part of the MatPES preprint[27]. The atomic positions of the DFT-relaxed structures were then randomly perturbed by 0.1 Å and re-relaxed using the UMLIPs listed; then the cohesive energy error and “fingerprint distance” are computed. For the latter, a fingerprint vector is generated by the `CrystalNN` method[38] for both the DFT-relaxed and UMLIP-relaxed structure. The fingerprint distance is calculated as the Euclidean distance between the two fingerprint vectors; a smaller fingerprint distance implies that the DFT-relaxed and UMLIP-relaxed structures are more similar, hence, smaller is better. Further details of the methodology for this benchmark can be found in ref. 27.

For the cohesive energy task (Fig. 3a), the MatPES-only model (MAE of 48 meV/atom, standard deviation σ = 87 meV/atom) performs better than the MP-ALOE-only model (MAE = 64, σ = 99 meV/atom); the combined MP-ALOE + MatPES model (MAE = 51, σ = 90 meV/atom) is quite

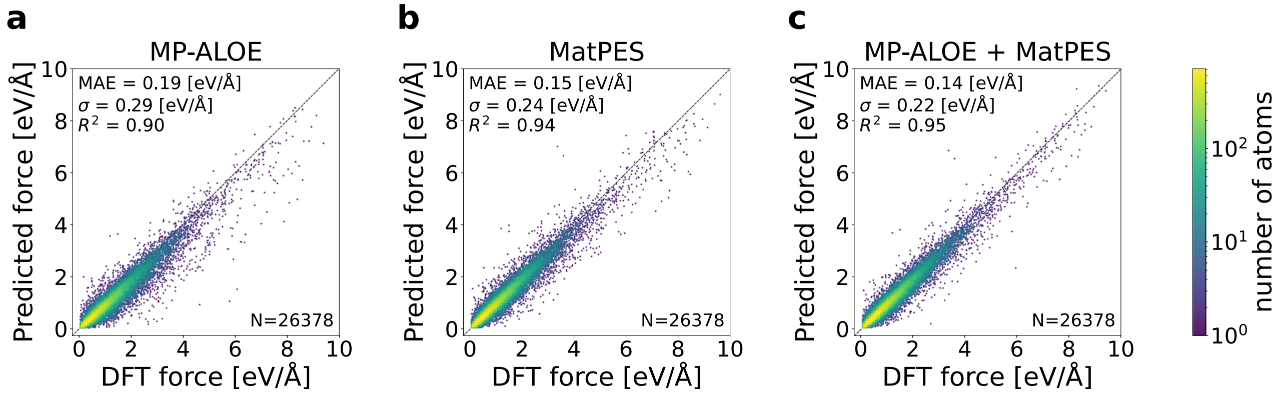

Fig. 4 | Predictions of off-equilibrium forces (eV/Å). Performance of MACE models trained on **a** MP-ALOE, **b** MatPES, and **c** the combined MP-ALOE + MatPES is shown. The DFT data were generated by taking relaxed structures from the WBM dataset[37], making them into supercells, displacing the atoms by 5% of the average interatomic distance, and re-computing the forces using the r^{2}SCAN functional.

comparable to the MatPES-only model. This is likely due to the MatPES dataset containing a greater proportion of near-equilibrium data (as shown in Fig. 2); this is further substantiated by MatPES having a cohesive energy distribution very similar to that of the *explicitly* near-equilibrium MPTrj dataset[14] (see Fig. 2a from ref. 27).

The models are relatively comparable on the fingerprint distance task (Fig. 3b). This implies that, despite energy predictions showing some differences, the relaxed structures are quite comparable for the three models.

One might assume that in moving from PBE to r^{2}SCAN, any gain in accuracy from DFT would also be seen in MLIPs trained to them. Supplementary Tables S3 and S4 validate this assumption in the equilibrium regime, where reliable experimental data are available. Table S3 presents the results of an equation of state test against zero-point corrected experimental data[3], and shows that MLIPs trained to a higher level of DFT approximation, r^{2}SCAN, are more accurate than MLIPs trained to PBE datasets ~10 times their size. Table S4 demonstrates that systematic improvements in formation energy prediction for transuranic compounds from PBE to r^{2}SCAN[29] are also observed in MLIPs trained to them.

Off-Equilibrium Forces. UMLIPs have been shown to have poor performance in predicting far-from-equilibrium forces (with a specific trend of underprediction/“softening”)[19]. To evaluate this, 1000 structures were sampled from the WBM dataset[37]. The selected structures from WBM were then repeated as supercells ranging from 18 to 32 atoms, and the atomic positions of every atom were randomly displaced in a random direction by 5% of the average interatomic distance; this value was selected because it is half of Lindemann’s criterion for melting[39]. Predicted vs. actual DFT forces are shown in Fig. 4. The MatPES-only model performs moderately better than the MP-ALOE-only model, though both seem to overcome the systematic softening reported on previously[19]. Moreover, the combined MP-ALOE + MatPES model achieves marginally higher accuracy than both the MatPES-only and MP-ALOE-only models.

Physicality at extreme deformations. MLIPs are known to have inconsistent performance at large deformations, sometimes leading to unphysically low energies when far from equilibrium[40]; this is especially prevalent at high compressions. To address this, the energy-volume scan (EV-scan) benchmark from `MLIP Arena`[40] was developed. In essence, this task involves a series of *static* calculations of a material that is uniformly strained – no relaxation of the cell nor of the atom positions is permitted. Given that no relaxation is allowed, one can expect a structure’s energy to monotonically increase as it is deformed away from its equilibrium configuration. Indeed, first principles evidence for this behavior at extreme strain can be found in, e.g., refs. 41,42. The EV-scan benchmark quantifies this behavior for a set of 1000 structures from the WBM dataset[37] when subjected to extreme deformations of ± 20% along each lattice direction (all lattice vectors are scaled uniformly). When all three lattice vectors are scaled by the same numeric factor, the energy is expected to have a single local extremum (minimum), hence the derivative should change sign exactly once. This is distinct from the case of, e.g., Bain paths, whereby scaling the lattice vectors by unequal values can produce multiple extrema on the potential energy surface[43].

For this benchmark, cases where the derivative changes sign more than once are counted as failures. Example EV curves for MACE potentials trained on MP-ALOE, MatPES, and both datasets combined are presented in Fig. 5. In these examples, the MatPES-only model predicts unphysical derivative sign changes under compression. Specifically, it predicts the energy to *decrease* significantly below the (DFT-derived) equilibrium volume. Extrapolating to zero internuclear separation, panel (a) would appear to favor zero separation of the nuclei (implosion). Panel (b) demonstrates an unphysical dual local minima, but a repulsive inner wall.

Overall, across the 1000 materials tested, the MP-ALOE-only model has a failure rate of 2.5%, relative to the significantly higher 14.8% failure rate for the MatPES-only model. The combined MP-ALOE + MatPES model obtains the lowest failure rate at 0.8%.

Further analysis and discussion of the EV-scan benchmark can be found in the Supplemental Information.

Molecular dynamics stability. A desirable property of UMLIPs is their relative computational affordability to perform longer molecular dynamics (MD) simulations, and of larger simulation cells. The MD stability benchmark task from `MLIP Arena`[40] attempts to measure this by again subjecting a reasonable structure to extreme environmental conditions. The structures used for this benchmark are taken from the RM24 structures in ref. 40, which were generated as random mixtures of stable materials from the Materials Project using `PackMol`[44]. To obtain reasonable initial structures, they were then relaxed using a Ziegler, Biersack, Littmark (ZBL) potential[45,46]. For the benchmark, a UMLIP is used to perform MD simulations under increasingly extreme environments. In total, 100 structures (with an average of 5-6 elements and 525 atoms) are simulated for each of the following two sub-tasks. Further details on these benchmark tasks can be found in ref. 40.

The first sub-task involves NVT simulations with linear temperature scaling from 300 to 3000 K over 10 ps, with results shown in Fig. 6a. All three datasets perform relatively well on the NVT sub-task. MP-ALOE performs the best with 98.8% of scheduled timesteps completed, compared to MatPES with 94.7% and the combined MP-ALOE + MatPES with 98.2% completion. Here, a given trajectory frame is deemed “valid” simply if the energy is within 100 eV/atom of the initial frame, the maximum kinetic energy for any atom is less than 100 eV. For the NPT simulations, we also demand that the cell volume stay bounded within one-tenth of and ten times the initial volume to be considered valid. The kinetic energy cutoff, corresponding to a

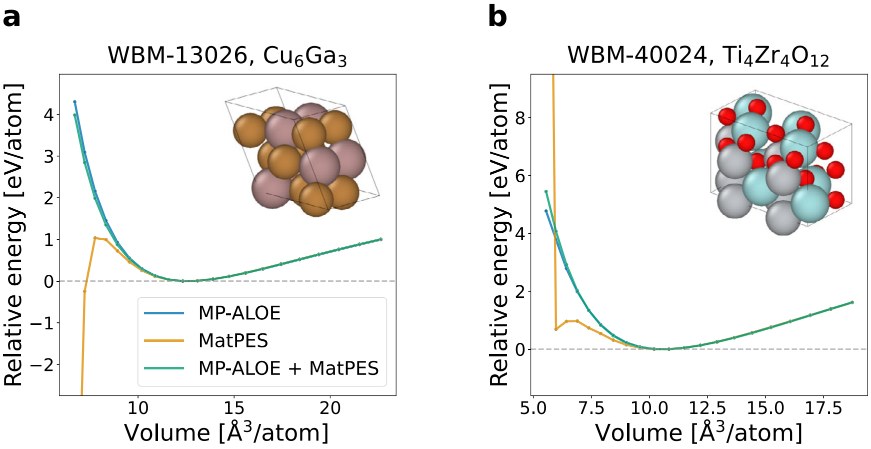

Fig. 5 | Examining the physicality of UMLIPs at extreme uniform static deformations. Performance of MACE models trained on MP-ALOE, MatPES, and the combined MP-ALOE + MatPES in predicting energy-volume curves for two example materials: **a** Cu_6Ga_3 and **b** $Ti_4Zr_4O_{12}$. Each plot contains a series of uniform deformations that are calculated *statically* (i.e., no relaxation is allowed). Hence, there should be a local minimum, and the derivative should change sign only once. A calculation is considered a 'failure' if the sign of the derivative changes more than once. In both of the examples shown, the MP-ALOE-only model and the combined dataset model are successful, but the MatPES-only model fails. Overall, for the 1000 materials examined, MP-ALOE, MatPES, and the combined MP-ALOE + MatPES have a 2.5%, 14.8%, and 0.8% failure percentage, respectively. Relative energy is defined such that the original (center-most) volume is set to zero. The corresponding crystal structure is inlaid in each plot.

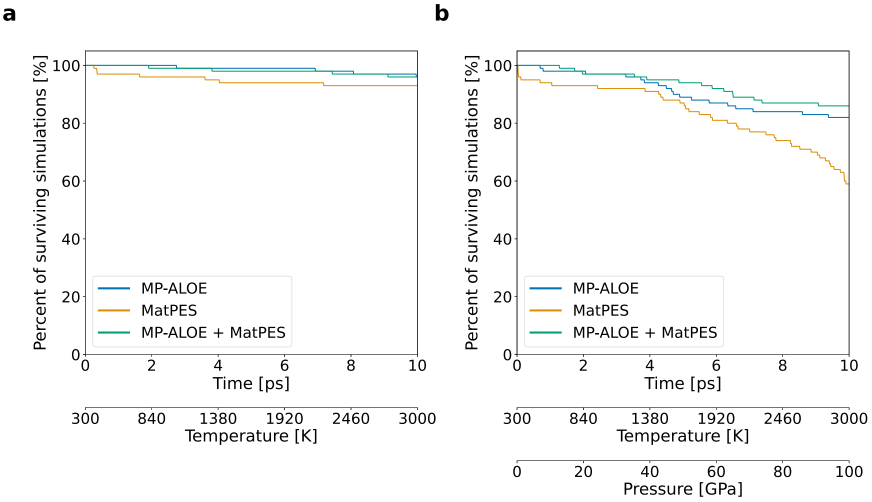

Fig. 6 | Stability during molecular dynamics simulations. **a** 100 NVT simulations with a linear temperature ramp from 300 to 3000 K. **b** 100 NPT simulations with a linear temperature ramp from 300 to 3000 K *and* a linear pressure ramp from 0 to 100 GPa. Tasks and the randomly-mixed structures were sourced from `MLIP Arena`[40]. The ordinate displays the percentage of surviving runs with a "valid" MD trajectory at a given timestep. See the text for a description of "valid".

velocity of 40 km/s for carbon-12, ensures that a simulation does not veer into a ballistic regime. Likewise, the volume criterion, which for a solid amounts to constraining each lattice parameter between roughly half and twice its initial value, ensures that the structures do not fragment (implode or explode).

The second sub-task involves NPT simulations with the same 300 to 3000 K temperature scaling *and* linear pressure scaling from 0 to 100 GPa over 10 ps. In Fig. 6b, the MP-ALOE-only model significantly outperforms the MatPES-only model, completing 90.6% and 83.7% of scheduled timesteps, respectively. This can likely be attributed to the greater sampling of larger pressures in MP-ALOE, as shown in Fig. 2c. The combined MP-ALOE + MatPES-trained model further improves on this, completing 93.2% of scheduled timesteps.

Note that we modify the original NPT benchmark from `MLIP Arena` to reduce the maximum pressure to 100 GPa instead of 500 GPa to better reflect physically attainable pressures. Results for the original benchmark are shown in Supplemental Fig. S7, which demonstrates that the MP-ALOE-only model successfully completes twice as many MD timesteps as the MatPES-only model when the pressure is increased to 500 GPa.

Discussion

We present a new dataset for universal machine learning interatomic potentials (UMLIPs) trained on the accurate r^2SCAN functional[28]. The dataset was constructed via a form of active learning (query by committee[35]) on primarily hypothetical compositions, thus we call it MP-ALOE (**M**aterials **P**roject - **A**ctive **L**earning of **O**ff **E**quilibrium structures). We compare the presented MP-ALOE to MatPES[27] (the only other public r^2SCAN dataset designed specifically for UMLIPs at the time of writing). By construction, both datasets were calculated using directly compatible plane-wave density functional theory (DFT) calculation parameters. To expand the space of possible but realistic chemical environments sampled in the dataset, we performed elemental substitution on a set of prototype structures, emphasizing bonding elements and non-*f*-block elements (which are likely to be poorly described by frozen core pseudopotentials and display less diverse chemical arrangements).

We then trained three MACE potentials[12] on MP-ALOE, MatPES, and the combined MP-ALOE + MatPES datasets. We compared their performance in predicting the cohesive energies of equilibrium structures and the magnitudes of off-equilibrium forces. For these two tasks, all three models perform comparably, however the MatPES-only model tends to predict slightly more accurate cohesive energies, and the combined MP-ALOE+ MatPES model clearly achieves the best performance overall on these three tasks. Importantly, we note that the MP-ALOE- and MatPES-only models perform almost identically in equilibrating structures to a reference configuration, indicating that they sample similar ranges of near-equilibrium interatomic forces.

The third benchmark task examined the physicality of the shapes of the potentials' energy-volume curves at extreme hydrostatic strain. Here, MP-ALOE significantly outperforms MatPES, with 2.5% and 14.8% of $E - V$ curves failing to meeting physicality criteria, respectively. Moreover, the combined MP-ALOE + MatPES shows the best performance overall across all categories.

The final benchmark task demonstrates the true value of MP-ALOE via stability of molecular dynamics (MD) calculations under extreme ensemble conditions. In both NVT runs with the temperature increasing from 300 to 3000 K, and NPT runs with simultaneously increasing temperature and pressure from 0 to 100 GPa, the MP-ALOE-only model significantly outperforms the MatPES-only model. Once more, the combined MP-ALOE + MatPES model demonstrates the highest stability in extreme environments.

In the nascent space of UMLIP research, more systematic methods to improve the quality of the datasets used to train them is highly needed. We have established a systematic and computationally tractable way to perform active learning-reinforced sampling, thereby targeting unexplored sections of the potential energy surface at a tractable computational cost. Our publicly-available MP-ALOE dataset is fully compatible with MatPES, and is recommended for use in training UMLIPs.

Methods

The overarching flow for generating the MP-ALOE dataset is as follows. A link to a GitHub repository containing (simplified) code to reproduce this workflow can be found in the Data Availability section; this does *not* include scripts for running DFT calculations using the `Atomate2` package nor for training `MACE` potentials, which can be found in their respective GitHub repositories.

Generating structures

Most of the structures contained in this dataset were created via the unbiased substitution of elements into prototype structures. Note that, as the probability of sampling an element was uniform, structures that are not conventionally charge balanced could be generated; this is explored and mostly confirmed in Supplemental Table S2. Prototypes contained within the ICSD[36] and Materials Project[22,23] were chosen, wherein the `StructureMatcher` from `pymatgen`[47] was used to remove duplicates. To ensure computational efficiency during the later DFT calculations, the prototypes considered were restricted to be between 2 and 8 atoms. Moreover, to reduce combinatorial explosion when occupying the prototypes with all combinations of elements included in this dataset, only up to ternary structures were considered. The result is 817 prototype structures: 660 binaries and 157 ternaries. When all possible substitutions of the 89 elements contained in MP are enumerated, this totals over 100 million structures to sample from.

The initial lattice parameters for these structures were estimated using the method described in the Supplementary Information. Then, to obtain a broader distribution of forces and stresses, the atomic positions were randomly displaced and the lattice vectors were randomly scaled according to the procedure in Supplementary Information. Supplementary Table S2 demonstrates that this procedure reliably produces structures with minimum nearest neighbor distances greater than the overlap region between pseudopotential cores.

Query By Committee

Query By Committee (QBC)[35] is an established active learning technique that has been applied across a wide range of fields, including interatomic potentials (e.g., ref. 48). Here, we apply the technique to select structures that cover unexplored portions of the PES. Our application of the QBC technique is described as follows.

The generated structures from the previous subsection are fed into an ensemble of interatomic potentials, which predict their properties. If the ensemble 'disagrees' above some heuristically set threshold on the predicted energy, forces, *and/or* stress, the structure is selected for the down-sampling step (detailed in the next subsection). The criterion for disagreement used was when the standard deviation of the ensemble's predicted energies (σ_e), forces (σ_f), or stresses (σ_s) are greater than 100 meV/atom, 100 meV/Å, or 100 meV/Å^3, respectively; these values roughly correspond to the errors of MACE-MP-0 on its own test set. To be explicit, the ensemble of models must disagree on *at least* one of the energies, forces, or stresses for a structure to be selected. To ensure that elements which typically do not bond (i.e., the noble gases and technetium) and elements which are unlikely to be well-described by frozen-core pseudopotentials (i.e., the *f*-block) were not over-represented in the dataset, a heuristic factor was applied to increase the threshold for committee consensus (see the `qbc.py` script in the GitHub repository linked in the Data Availability section for more information on the thresholds used).

For the first iteration of active learning, existing interatomic potentials were used. Namely, the ensemble was comprised of the pre-trained MACE-MP-0[15], CHGNet[14], and M3GNet[13] models. Both to reduce the computational burden of model training and because no existing r^{2}SCAN UMLIP was available when the first active learning cycle was performed, models trained to PBE data were used to select structures for the solely-r^{2}SCAN-based MP-ALOE dataset. These particular three pretrained models were selected because they were the top three performers on the Matbench Discovery benchmark for UMLIPs[20] at the time this project was started. For the remaining iterations of active learning, three MACE models were trained (details for training can be found later in the Methods section). MACE was chosen primarily because it maintains a higher body-order than M3GNet or CHGNet.

This methodology resulted in roughly ~500,000 structures being selected in a given active learning cycle; these 500,000 structures were then downsampled in the next subsection.

Downsampling

One noteworthy drawback of batch-mode active learning is that the batch of data selected for labeling often has informational overlap[49]. To mitigate this, a method for downsampling a diverse subset (from the larger set of identified structures from the QBC) is needed. DIRECT sampling[33] was identified as a suitable method for downsampling, as it has been successfully demonstrated to select diverse samples that can create comparable models with substantially less data. The DIRECT method involves the following steps: (1) all structures are featurized using the M3GNet Formation Energy model[13] (accessed via the `matgl` repository[50]); (2) principal component analysis is then performed for dimensionality reduction; (3) the dimensionally reduced features are then clustered using BIRCH clustering[51]; and (4) a fixed number of structures are selected from each cluster. For a given active learning cycle, the use of DIRECT sampling reduced the number of selected samples from roughly ~500,000 down to roughly ~125,000 structures. These ~125,000 structures are then simulated using DFT, as described in later in the Methods section.

Additional data

The structures generated via the process described above are often far from equilibrium. To augment this aspect of the data generation process, a small amount of additional near-equilibrium data from the Materials Project was also included. Specifically, all structures with up to 3 elements and up to 32 atoms (totaling roughly ~30,000 structures) were re-calculated using the same DFT settings as those used to calculate the structures selected via QBC (described in the next subsection).

Density functional theory calculation details

DFT calculations were performed using the Vienna Ab-Initio Simulation Package (VASP)[52] using projector-augmented wave (PAW) potentials[4,53]. A plane-wave cutoff energy of 680 eV was used, with the KSPACING parameter set to 0.2. Further input parameters used are detailed in the MP24RelaxSet from `pymatgen`.

Calculations were performed in two stages. First, a static calculation using the PBE exchange-correlation functional[21] was performed. Then, the WAVECAR from the static PBE calculation is fed into a relaxation calculation performed using the r^{2}SCAN functional[28]. The second r^{2}SCAN calculation always ran for *three* ionic steps; this was chosen so that initial structures (which were often relatively far from equilibrium) have a chance to equilibriate to a reasonable pressure, less extreme forces, etc. This is especially necessary given the difficulty of estimating the lattice parameter of an unknown material. The authors chose to perform three ionic steps intentionally—this was the number of ionic steps required for 90% of calculations to include both a positive and negative pressure in at least one ionic step for a set of test calculations.

In total, the above-prescribed workflow converged for 82% of structures. All calculations and workflows were performed using the atomate2 package[54], which itself builds upon the fireworks[55], jobflow[56], and pymatgen[47] packages.

Model training

After each iteration of active learning, the r^{2}SCAN data generated was compiled and used to train three MACE potentials[12]. All input parameters for the models trained in this work were (nearly) identical to those used in the 'large' model from the original version of the MACE-MP-0 manuscript on arXiv[15] (now referred to as MACE-MP-0a, which was released with v0.3.6 of the MACE package on GitHub). The only parameters changed were the isolated atom energies, which were recomputed using r^{2}SCAN. Each potential had exactly 5,725,072 parameters. A 90-5-5 train-validation-test split was used, and the models were trained for 100 epochs (which was empirically selected based on when learning appeared to plateau). The only noteworthy difference between the three models within a given active learning iteration was that the training/validation/test splits were randomized, thus creating small variations between the models.

Combining MP-ALOE and MatPES

MP-ALOE and MatPES were calculated using directly compatible DFT parameters. However, both datasets had a small proportion of structures sourced from the Materials Project included, with some MP structures appearing in both datasets. Hence, the datasets cannot simply be merged, as this would lead to data redundancy. MatPES, by construction, contains a greater number of MP-sourced structures, including all of the MP-sourced structures in MP-ALOE. Hence, only the *non-MP* structures in MP-ALOE were merged with *all* r^{2}SCAN calculations within MatPES during training of the "MP-ALOE + MatPES" potential. Python code for combining these two datasets is included in the GitHub repository, linked in the Data Availability section.

Data availability

The MP-ALOE dataset can be downloaded at https://doi.org/10.6084/m9.figshare.29452190. The raw VASP outputs, totaling roughly 50 TB, can be made available upon reasonable request; WAVECAR files were not kept. The trained MACE potentials are also available at the URL above, as is the data for the off-equilibrium forces benchmark. Code used to generate the data can be found at https://github.com/matthewkuner/MP-ALOE. Note that this repository also includes a script for combining MP-ALOE with MatPES.All other materials used in the creation of this work are available upon reasonable request.

Received: 18 July 2025; Accepted: 12 October 2025;
Published online: 21 November 2025

References

1. Kohn, W. & Sham, L. J. Self-consistent equations including exchange and correlation effects. *Phys. Rev.* **140**, A1133–A1138 (1965).
2. Kaplan, A. D., Levy, M. & Perdew, J. P. The predictive power of exact constraints and appropriate norms in density functional theory. *Annu. Rev. Phys. Chem.* **74**, 193–218 (2023).
3. Tran, F., Stelzl, J. & Blaha, P. Rungs 1 to 4 of DFT Jacob's ladder: Extensive test on the lattice constant, bulk modulus, and cohesive energy of solids. *J. Chem. Phys.* **144**, 204120 (2016).
4. Blöchl, P. E. Projector augmented-wave method. *Phys. Rev. B* **50**, 17953–17979 (1994).
5. Daw, M. S. & Baskes, M. I. Embedded-atom method: Derivation and application to impurities, surfaces, and other defects in metals. *Phys. Rev. B* **29**, 6443–6453 (1984).
6. Baskes, M. I. Modified embedded-atom potentials for cubic materials and impurities. *Phys. Rev. B* **46**, 2727–2742 (1992).
7. Behler, J. & Parrinello, M. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys. Rev. Lett.* **98**, 146401 (2007).
8. Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O. & Dahl, G. E. Neural message passing for quantum chemistry. Preprint at http://arxiv.org/abs/1704.01212 (2017).
9. Schütt, K. T. SchNet - A deep learning architecture for molecules and materials. *J. Chem. Phys.* **148**, 241722 (2018).
10. Batzner, S. et al. E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nat. Commun.* **13**, 2453 (2022).
11. Drautz, R. Atomic cluster expansion for accurate and transferable interatomic potentials. *Phys. Rev. B* **99**, 014104 (2019).
12. Batatia, I., Kovacs, D. P., Simm, G., Ortner, C. & Csanyi, G. Mace: Higher order equivariant message passing neural networks for fast and accurate force fields. *Adv. Neural Inf. Process. Syst.* **35**, 11423–11436 (2022).
13. Chen, C. & Ong, S. P. A universal graph deep learning interatomic potential for the periodic table. *Nat. Comput. Sci.* **2**, 718–728 (2022).
14. Deng, B. et al. CHGNet as a pretrained universal neural network potential for charge-informed atomistic modelling. *Nat. Mach. Intell.* **5**, 1031–1041 (2023).
15. Batatia, I. et al. A foundation model for atomistic materials chemistry. Preprint at http://arxiv.org/abs/2401.00096 (2023).
16. Park, Y., Kim, J., Hwang, S. & Han, S. Scalable parallel algorithm for graph neural network interatomic potentials in molecular dynamics simulations. *J. Chem. Theory Comput.* **20**, 4857–4868 (2024).
17. Yang, H. et al. MatterSim: A deep learning atomistic model across elements, temperatures and pressures. Preprint at http://arxiv.org/abs/2405.04967.
18. Barroso-Luque, L. et al. Open Materials 2024 (OMat24) Inorganic Materials Dataset and Models. Preprint at http://arxiv.org/abs/2410.12771 (2024).
19. Deng, B. et al. Systematic softening in universal machine learning interatomic potentials. *npj Comput. Mater.* **11**, 1–9 (2025).
20. Riebesell, J. et al. A framework to evaluate machine learning crystal stability predictions. *Nat. Mach. Intell* **7**, 836–847 (2025).
21. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865–3868 (1996).
22. Jain, A. et al. Commentary: The materials project: A materials genome approach to accelerating materials innovation. *APL Mater.* **1**, 011002 (2013).
23. Horton, M. K. et al. Accelerated data-driven materials science with the Materials Project. *Nat. Mater*. 1–11 (2025).
24. Schmidt, J., Pettersson, L., Verdozzi, C., Botti, S. & Marques, M. A. L. Crystal graph attention networks for the prediction of stable materials. *Sci. Adv.* **7**, eabi7948 (2021).
25. Schmidt, J. et al. Machine-learning-assisted determination of the global zero-temperature phase diagram of materials. *Adv. Mater.* **35**, 2210788 (2023).
26. Nazarov, R., Hickel, T. & Neugebauer, J. Vacancy formation energies in fcc metals: Influence of exchange-correlation functionals and correction schemes. *Phys. Rev. B* **85** (2012).
27. Kaplan, A. D. et al. A foundational potential energy surface dataset for materials. Preprint at http://arxiv.org/abs/2503.04070 (2025).
28. Furness, J. W., Kaplan, A. D., Ning, J., Perdew, J. P. & Sun, J. Accurate and numerically efficient r^{2}SCAN meta-generalized gradient approximation. *J. Phys. Chem. Lett.* **11**, 8208–8215 (2020).
29. Kingsbury, R. et al. Performance comparison of r^{2}SCAN and SCAN metaGGA density functionals for solid materials via an automated, high-throughput computational workflow. *Phys. Rev. Mater.* **6**, 013801 (2022).
30. Swathilakshmi, S., Devi, R. & Sai Gautam, G. Performance of the r^{2}SCAN functional in transition metal oxides. *J. Chem. Theory Comput.* **19**, 4202–4215 (2023).

31. Kothakonda, M. et al. Testing the r^{2}SCAN density functional for the thermodynamic stability of solids with and without a van der Waals correction. *ACS Mater. Au* **3**, 102–111 (2023).
32. Liu, M., Gopakumar, A., Hegde, V. I., He, J. & Wolverton, C. High-throughput hybrid-functional DFT calculations of bandgaps and formation energies and multifidelity learning with uncertainty quantification. *Phys. Rev. Mater.* **8**, 043803 (2024).
33. Qi, J., Ko, T. W., Wood, B. C., Pham, T. A. & Ong, S. P. Robust training of machine learning interatomic potentials with dimensionality reduction and stratified sampling. *npj Comput. Mater.* **10**, 1–11 (2024).
34. Lysogorskiy, Y., Bochkarev, A., Mrovec, M. & Drautz, R. Active learning strategies for atomic cluster expansion models. *Phys. Rev. Mater.* **7**, 043801 (2023).
35. Seung, H. S., Opper, M. & Sompolinsky, H. Query by committee. In *Proceedings of the fifth annual workshop on Computational learning theory*, COLT '92, 287–294 (Association for Computing Machinery, 1992).
36. Zagorac, D., Müller, H., Ruehl, S., Zagorac, J. & Rehme, S. Recent developments in the inorganic crystal structure database: theoretical crystal structure data and related features. *J. Appl. Crystallogr.* **52**, 918–925 (2019).
37. Wang, H.-C., Botti, S. & Marques, M. A. L. Predicting stable crystalline compounds using chemical similarity. *npj Comput. Mater.* **7**, 1–9 (2021).
38. Zimmermann, N. E. R. & Jain, A. Local structure order parameters and site fingerprints for quantification of coordination environment and crystal structure similarity. *RSC Adv.* **10**, 6063–6081 (2020).
39. Lindemann, V. F. A. The calculation of molecular Eigen frequencies. *Phys. Z.* **11**, 609–614 (1910).
40. Chiang, Y. et al. Mlip arena: Advancing fairness and transparency in machine learning interatomic potentials through an open and accessible benchmark platform. https://openreview.net/forum?id=ysKflavYQE (2025).
41. Alchagirov, A. B., Perdew, J. P., Boettger, J. C., Albers, R. C. & Fiolhais, C. Energy and pressure versus volume: Equations of state motivated by the stabilized jellium model. *Phys. Rev. B* **63**, 224115 (2001).
42. Kaplan, A. D., Clark, S. J., Burke, K. & Perdew, J. P. Calculation and interpretation of classical turning surfaces in solids. *npj Comput. Mater.* **7** (2021).
43. Grimvall, G., Magyari-Köpe, B., Ozoliņš, V. & Persson, K. A. Lattice instabilities in metallic elements. *Rev. Mod. Phys.* **84**, 945–986 (2012).
44. Martínez, L., Andrade, R., Birgin, E. G. & Martínez, J. M. PACKMOL: A package for building initial configurations for molecular dynamics simulations. *J. Comput. Chem.* **30**, 2157–2164 (2009).
45. Ziegler, J., Biersack, J. & Littmark, U. Empirical stopping powers for ions in solids 88–100 (1983).
46. Ziegler, J. F. & Biersack, J. P. The stopping and range of ions in matter. In Bromley, D. A. (ed.) *Treatise on Heavy-Ion Science: Volume 6: Astrophysics, Chemistry, and Condensed Matter*, 93–129 (Springer US, 1985).
47. Ong, S. P. et al. Python materials genomics (pymatgen): A robust, open-source Python library for materials analysis. *Comput. Mater. Sci.* **68**, 314–319 (2013).
48. Smith, J. S., Nebgen, B., Lubbers, N., Isayev, O. & Roitberg, A. E. Less is more: Sampling chemical space with active learning. *J. Chem. Phys.* **148**, 241733 (2018).
49. Ren, P. et al. A survey of deep active learning. *ACM Comput. Surv.* **54**, 1–40 (2022).
50. Ko, T. W. et al. Materials Graph Library (MatGL), an open-source graph deep learning library for materials science and chemistry. *npj Comput. Mater.* **11**, 253 (2025).
51. Zhang, T., Ramakrishnan, R. & Livny, M. BIRCH: an efficient data clustering method for very large databases. *SIGMOD Rec.* **25**, 103–114 (1996).
52. Kresse, G. & Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B* **59**, 1758–1775 (1999).
53. Kresse, G. & Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **54**, 11169–11186 (1996).
54. Ganose, A. M. et al. Atomate2: modular workflows for materials science. *Digit. Discov.* **4**, 1944–1973 (2025).
55. Jain, A. et al. FireWorks: a dynamic workflow system designed for high-throughput applications. *Concurr. Comput.* **27**, 5037–5059 (2015).
56. Rosen, A. S. et al. Jobflow: Computational workflows made simple. *J. Open Source Softw.* **9**, 5995 (2024).
57. Riebesell, J., Yang, H., Goodall, R. & Baird, S. G. Pymatviz: visualization toolkit for materials informatics. https://github.com/janosh/pymatviz (2022).

Acknowledgements

This work was intellectually led by the Materials Project program KC23MP, supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, Materials Sciences and Engineering Division under contract No. DE-AC02-05-CH11231. Matthew Kuner was supported in part by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-2146752. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. This research used the Savio computational cluster resource provided by the Berkeley Research Computing program at the University of California, Berkeley (supported by the UC Berkeley Chancellor, Vice Chancellor for Research, and Chief Information Officer). This research used the Lawrencium computational cluster resource provided by the IT Division at the Lawrence Berkeley National Laboratory (Supported by the Director, Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231)This research used resources of the National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility located at Lawrence Berkeley National Laboratory, operated under Contract No. DE-AC02-05CH11231 using NERSC award BES-ERCAP-0022838. The authors appreciate the insightful discussions from Mr. Yuan Chiang, Dr. Shivani Srivastava, Professor Bingqing Cheng, and Professor Mary Scott at UC Berkeley, and from Dr. Anubhav Jain at Lawrence Berkeley National Laboratory.

Author contributions

M.C.K.: conceived of the project, performed the DFT calculations, trained the UMLIPs, performed the benchmark calculations, created visualizations, wrote and edited the manuscript. A.D.K.: performed benchmark calculations, created visualizations, assessed the validity of the DFT calculations, wrote and edited the manuscript. K.A.P.: manuscript editing, provision of computational resources. M.A.: project design, manuscript editing, and provision of computational resources. D.C.C.: project design, manuscript editing, provision of computational resources

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41524-025-01834-9.

Correspondence and requests for materials should be addressed to Matthew C. Kuner or Daryl C. Chrzan.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.