NEP89: Universal neuroevolution potential for inorganic and organic materials across 89 elements

Ting Liang,[1,2,*] Ke Xu,[1,2,*] Eric Lindgren,[3] Zherui Chen,[4,5] Rui Zhao,[6] Jiahui Liu,[7] Esmée Berger,[3] Benrui Tang,[1] Bohan Zhang,[1] Yanzhou Wang,[8] Keke Song,[9] Penghua Ying,[10] Nan Xu,[11] Haikuan Dong,[1] Shunda Chen,[12,†] Paul Erhart,[3,13,‡] Zheyong Fan,[1,§] Tapio Ala-Nissila,[8,14] and Jianbin Xu[2,¶]

[1] *College of Physical Science and Technology, Bohai University, Jinzhou 121013, P. R. China*
[2] *Department of Electronic Engineering and Materials Science and Technology Research Center, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong SAR, 999077, P. R. China*
[3] *Department of Physics, Chalmers University of Technology, 41926 Gothenburg, Sweden*
[4] *Future Technology School, Shenzhen Technology University, Shenzhen 518118, P. R. China*
[5] *College of Applied Sciences, Shenzhen University, Shenzhen 518060, P. R. China*
[6] *School of Mechanical and Electrical Engineering, Xinyu University, Xinyu 338004, P. R. China*
[7] *Beijing Advanced Innovation Center for Materials Genome Engineering, University of Science and Technology Beijing, Beijing 100083, P. R. China*
[8] *MSP group, QTF Centre of Excellence, Department of Applied Physics, P.O. Box 15600, Aalto University, FI-00076 Aalto, Espoo, Finland*
[9] *College of Physics and Information Engineering, Fuzhou University, Fuzhou 350108, China*
[10] *Department of Physical Chemistry, School of Chemistry, Tel Aviv University, Tel Aviv, 6997801, Israel*
[11] *College of Chemical and Biological Engineering, Zhejiang University, Hangzhou 310058, P. R. China*
[12] *Department of Civil and Environmental Engineering, George Washington University, Washington, DC 20052, USA*
[13] *Wallenberg Initiative Materials Science for Sustainability, Chalmers University of Technology, 41926 Gothenburg, Sweden*
[14] *Interdisciplinary Centre for Mathematical Modelling, Department of Mathematical Sciences, Loughborough University, Loughborough, Leicestershire LE11 3TU, UK*

(Dated: June 11, 2025)

While machine-learned interatomic potentials offer near-quantum-mechanical accuracy for atomistic simulations, many are material-specific or computationally intensive, limiting their broader use. Here we introduce NEP89, a foundation model based on neuroevolution potential architecture, delivering empirical-potential-like speed and high accuracy across 89 elements. A compact yet comprehensive training dataset covering inorganic and organic materials was curated through descriptor-space subsampling and iterative refinement across multiple datasets. NEP89 achieves competitive accuracy compared to representative foundation models while being three to four orders of magnitude more computationally efficient, enabling previously impractical large-scale atomistic simulations of inorganic and organic systems. In addition to its out-of-the-box applicability to diverse scenarios, including million-atom-scale compression of compositionally complex alloys, ion diffusion in solid-state electrolytes and water, rocksalt dissolution, methane combustion, and protein-ligand dynamics, NEP89 also supports fine-tuning for rapid adaptation to user-specific applications, such as mechanical, thermal, structural, and spectral properties of two-dimensional materials, metallic glasses, and organic crystals.

I. INTRODUCTION

Large-scale atomistic molecular dynamics (MD) simulations are invaluable tools for elucidating the intricate properties of complex materials. However, the fidelity of these simulations critically relies on the accuracy of the underlying interatomic potentials (or force fields). *Ab initio* molecular dynamics (AIMD) simulations offer the desired level of fidelity, but their computational cost is prohibitive in large-scale MD simulations of millions of atoms and more.

In recent years, artificial intelligence has been successfully applied to construct accurate yet efficient machine-learned interatomic potential models [1, 2], significantly advancing the field of atomistic simulations [3–7]. Early models were developed for specific materials and even specific properties of a given material. More recent developments have revealed a trend toward building more general-purpose potentials, from a single element [8] to a set of metals [9], to tens of elements [10, 11], and to almost the entire periodic table [12–17]. The last category of potential models, although encompassing many species, remains incomplete, as they were mostly constructed based on inorganic materials. Moreover, their high computational cost severely limits their practical

* These authors contributed equally to this work.
† phychensd@gmail.com
‡ erhart@chalmers.se
§ brucenju@gmail.com
¶ jbxu@ee.cuhk.edu.hk

applicability, particularly for large-scale or long-timescale simulations.

Here, to overcome these limitations, we introduce NEP89, a foundation model that enables high-fidelity simulations of both inorganic and organic materials across 89 elements, while achieving a computational speed several orders of magnitude faster than previous universal interatomic potential models. Our model is based on the state-of-the-art neuroevolution potential (NEP) approach [9, 18, 19], which combines high accuracy with efficiency. This approach has recently been demonstrated to achieve a general-purpose interatomic potential for 16 metals and their alloys [9].

One major obstacle is the lack of a readily usable training dataset that comprehensively incorporates both inorganic and organic materials. The lack of such a dataset is presumably the reason why previous works have only trained universal potential models either for inorganic [12–17] or organic materials [20–23], or both, but with multiple training tasks instead of a single unified one [11]. A crucial contribution in the present work is the curation of existing databases of separate sets of materials to ensure they have consistent target data for training a unified interatomic potential for both inorganic and organic materials. Starting from a sub-sampled dataset of OMAT24 [24], we iteratively train NEP89 and build the corresponding training dataset, incorporating structures from many other public datasets [9, 13, 20–22, 25–27]. Moreover, we add the D3 dispersion correction [28, 29] to some datasets such that all of them have proper dispersion interactions. Furthermore, the relative energies between different datasets are also optimized during the training process, which is a prerequisite for constructing a unified single-task model. The NEP89 model achieves competitive accuracy in predicting both static and dynamic properties. It also enables large-scale atomistic simulations of both inorganic and organic systems out of the box, while supporting rapid and cost-effective fine-tuning across a broad range of applications.

II. RESULTS

The architecture of the NEP89 model and its fine-tuning. The NEP89 model is built on the NEP approach [9, 18, 19]. We begin by briefly introducing the NEP architecture and then discuss the extensions that enable the fine-tuning of NEP89. The NEP approach is a many-body neural network potential that uses an atom-decomposed total energy and atom-centered descriptors constructed with Chebyshev and Legendre polynomials, along with an advanced training algorithm that incorporates regularization and employs the separable natural evolution strategy [30]. The latter evolves a set of mean and variance values for the trainable parameters that also form the basis for fine-tuning. The descriptors are defined within a specified cutoff radius to ensure that the computational cost scales linearly with the number of atoms. The input layer neurons of the neural network correspond to the descriptors, while the output layer neuron represents the site energy U_i of the central atom i. Force and virial stress are calculated via analytical derivatives of the energy (see Supp. Note S1 for details). Atomic species are encoded in the expansion coefficients of a set of radial functions. For each pair of species, there is an independent set of $N_{\rm ec}$ expansion coefficients optimized during training. Additionally, each species has its own independent set of $N_{\rm nn}$ trainable weight and bias parameters within the neural network. Therefore, for a system with $N_{\rm spe}$ species, there are $N_{\rm spe}^2 N_{\rm ec} + N_{\rm spe} N_{\rm nn}$ trainable parameters (refer to Supp. Note S2 for the hyperparameters used when training NEP89).

Thanks to the combinatorial architecture of NEP, the trained NEP89 model can be conveniently fine-tuned into smaller models tailored to subsets of species using a new training dataset. During the fine-tuning process, it is crucial to maintain the descriptor normalization. To this end, for the targeted species, the mean and variance of their trainable parameters, used by the separable natural evolution strategy, are extracted from NEP89 and reused during continued training. Fine-tuning typically requires only a small fraction of the training steps compared to training from scratch to achieve optimal results. Additionally, to prevent the model from undesirably losing memory, the variance values of the descriptor parameters are set to zero, ensuring that only the neural network parameters are updated during fine-tuning.

Iterative training of NEP89. To train a universal potential model across 89 species for both inorganic and organic materials, the first task is to construct a training set that is diverse, reliable, and consistent. So far, no single available dataset meets these three conditions simultaneously. Fortunately, there are already many individual datasets, and we aim to combine them to form a consistent dataset with enhanced diversity and reliability. A notable one is the OMat24 dataset [24] containing over 110 million structures for inorganic bulk materials. To supplement the OMat24 dataset, we also include structures from other important datasets, including the MPtrj dataset for relaxation trajectories of inorganic bulk materials across 89 elements [13], the SPICE dataset for drug-like small molecules, peptides, and solvated amino acids [21], the ANI-1xnr dataset for reactive organics [22], the SSE-ABACUS and SSE-VASP datasets for solid-state electrolytes [25], the solvated protein fragments (Protein for short) dataset [20], the UNEP-v1 dataset for 16 metals and their alloys [9], the CH dataset for general and reactive CH systems [26], the water dataset with liquid, gas, and vapor-liquid phases [27], and a reactive CHONPS dataset we created. More details regarding these datasets are provided in Supp. Table S1. The successful training of NEP89 was not accomplished by providing all the aforementioned datasets to the optimizer simultaneously, but rather through an iterative training strategy, as described below.

Initially, we note that these datasets are inconsistent in their incorporation of dispersion interactions. While some datasets (SPICE, Protein, ANI-1xnr) include the D3 dispersion [28, 29], we added this dispersion interaction to the remaining datasets for uniformity. The water dataset is unique, as it was labeled using the MB-pol method [31], delivering accuracy close to CCSD(T). Although not all datasets are calculated at the same level of quantum-mechanical theory, the differences in force and stress are minor and can be disregarded in training the NEP89 model. On the other hand, these datasets have significantly different energy references for a given species, which hinders the development of a unified model. To address this challenge, we initially focused exclusively on the energies in the OMAT24 dataset, disregarding reference energies from all other datasets.

Another challenge in training a unified model based on extensive public datasets is managing the vast number of training structures. A good training dataset does not necessarily require a large number of structures but rather a diverse and well-balanced selection. To address this, we first created a randomly sub-sampled dataset from the OMAT24 database to train an initial model. Using this model, we then selected a subset of the OMAT24 database with uniformly distributed descriptor values (see Supp. Note S3 for details) to train an improved model. We then performed model training and data curation for more than ten iterations. During each iteration, the current model predicted unseen structures across all datasets. Structures with high force and stress errors were identified, and a subset with evenly distributed descriptor values was sampled and added to the training dataset. This refined dataset was then used to train the next model in the sequence.

An essential aspect of dataset refinement is identifying inaccuracies in the reference data. Several datasets, such as OMAT24 and MPtrj, contain erroneous entries or unconverged density functional theory (DFT) calculations. To mitigate the impact of noisy reference data, prior studies have employed the Huber loss function [12–14, 16, 17]. Our approach involves iteratively removing outliers from the training dataset, thereby gradually reducing their influence on the resulting models. In Supp. Figure S1, we show results for 113 typical structures in the MPtrj dataset identified as outliers during our training, where the original DFT reference values were found to be inaccurate, while predictions by NEP89 closely aligned with our recalculated reference values.

Through this iterative training and refinement process, we developed a diverse and balanced training dataset, which is visualized in the reduced descriptor space spanned by the two principal components in Figure 1a. However, the model obtained in this process was exclusively trained on the OMAT24 dataset in terms of energy, introducing a possible bias toward this dataset. To ensure balanced training across different datasets, we accordingly adjusted the reference energies for the other datasets. These adjustments minimized differences between the model predictions and the shifted reference energies (see Supp. Note S4 for details). We excluded the reference energies for the MPtrj dataset from training, since they were incompatible with those in the OMAT24 dataset due to differing treatments of the DFT+U technique. We then performed the energy adjustment for a few iterations and trained the final model, which we refer to as the NEP89 model. The parity plot for the energies (Figure 1b) indicates that all the considered reference energies are indeed quite consistent after our adjustment. The parity plot for the forces (Figure 1c) also indicates a well-trained model for the combined dataset. (Supp. Figure S2 shows error metrics for energy, force, and stress for the individual datasets.)

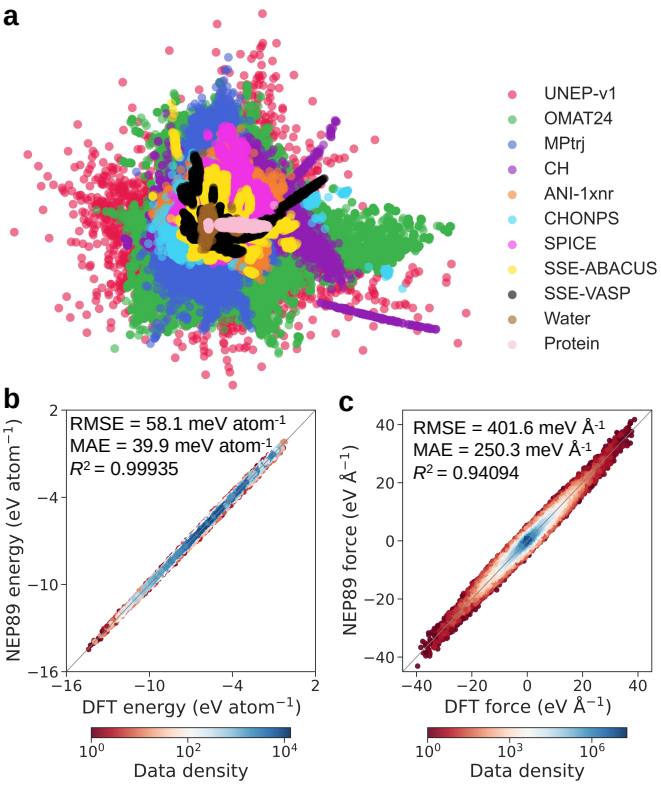

FIG. 1. **Dataset composition for NEP89 and overall training accuracy.** (**a**) Distribution of the NEP89 datasets in the reduced descriptor space spanned by the first two principal components. The individual datasets were subsampled from the UNEP-v1 [9], OMAT24 [24], MPtrj [13], CH [26], ANI-1xnr [22], SPICE [21], SSE-ABACUS [25], SSE-VASP [25], water [27], and Protein [20] datasets, or newly prepared in this work (CHONPS). (**b**,**c**) Parity plots for energy and force comparing NEP89 predictions with reference values (DFT or high-level quantum chemistry). The color intensity visualizes the distribution and density of the NEP89 datasets.

Benchmarks on static properties. Figure 2 presents a number of benchmarks on static properties predicted by NEP89 and various foundation models, including MACE-MP-0 (medium version) [16], CHGNet [13] (version 0.3.0), and M3GNet [12] (version 2021.2.8-DIRECT-

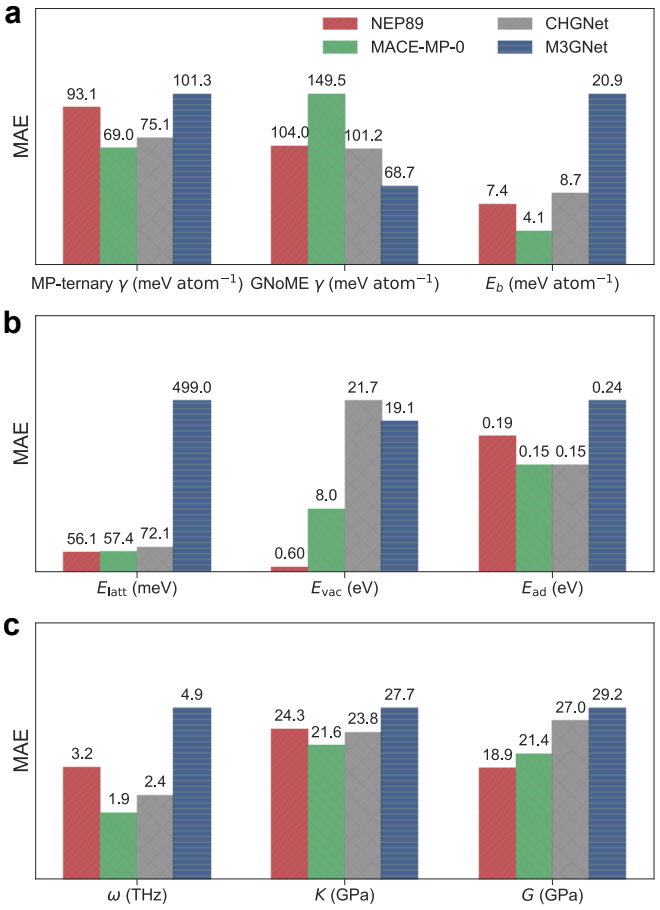

FIG. 2. **Various benchmarks on static properties for validating the foundation models.** Mean absolute errors (MAEs) of several foundation models, NEP89 (this work), MACE-MP-0 [16], CHGNet [13], and M3GNet [12], with respect to DFT reference data for the nine evaluated properties. (**a**) Formation energies (γ) for structures sampled from the Materials Project (MP-ternary) [32] and the GNoME dataset [14], and binding energies (E_b) of the S66 dimer set [33] with DFT (PBE+D3) references [16]. (**b**) Lattice energies (E_latt) for the DMC-ICE13 dataset [34] with DFT references [16], formation energies (E_vac) of iron vacancy clusters of different sizes, and adhesion energy (E_ad) of hydrogen atoms in iron nanopores with DFT (PBE). (**c**) Highest phonon band frequencies (ω) of 97 materials with DFT (PBE) references from the PhononDB database [35], and bulk (K) and shear (G) moduli with DFT references from Materials Project [32] covering more than 11,000 materials.

PES); see Methods and Supp. Figures S3–S8 for calculation details. Overall, NEP89 demonstrates comparable accuracy to other foundation models, based on mean absolute errors across nine evaluated properties relative to DFT data. For the formation energies (γ) of the ternary structures sampled from the Materials Project (MP-ternary) [32], MACE-MP-0 and CHGNet are slightly more accurate, with NEP89 ranking third, despite not being trained on the MPtrj energies. For the 2-component to 5-component structures from the GNoME dataset [14], on which none of the models were trained, MACE-MP-0 shows the worst accuracy, while NEP89 performs very close to CHGNet as the second-best model. For the binding energies (E_b) of the S66 dimer set [33], NEP89 ranks as the second-best model, even though it does not require a separate D3 correction as for the other models. NEP89 demonstrates superior accuracy for lattice energies (E_latt) for the DMC-ICE13 dataset [16, 34] and formation energies (E_vac) of iron vacancy clusters of different sizes, and comparable accuracy for adhesion energy (E_ad) of hydrogen atoms in iron nanopores (see Supp. Note S8 for details on our iron vacancy cluster and nanopore datasets). For the highest phonon band frequencies (ω) for 97 materials from the PhononDB database [35], NEP89 achieves the third-best performance. For more than 11 000 materials from Materials Project [32], NEP89 demonstrates comparable accuracy in predicting the bulk modulus (K) and the best accuracy in predicting the shear modulus (G).

Evaluation of computational performance. Since the primary motivation for developing machine-learned interatomic potentials is to extend the spatiotemporal scales of ab initio calculations, computational efficiency, in terms of both memory usage and computational speed, is a crucial factor. As shown in Figure 3a–b, NEP89 consistently outperforms other foundation models in terms of speed and memory usage across a wide range of system sizes, using a 20-element high-entropy alloy as a representative example (see Methods for details). On the 24-GB RTX 4090 and 80-GB H800 GPUs considered here, NEP89 can simulate up to 8×10^6 atoms at a speed of 5×10^6 atom-steps per second and 1.5×10^7 atoms at a speed of 7×10^6 atom-steps per second, respectively. Overall, NEP89 extends the accessible spatiotemporal scale of atomistic simulations by at least three orders of magnitude compared to other foundation models, while maintaining competitive accuracy for both static properties (as shown above) and dynamic properties (as we will discuss next).

Benchmarks of dynamical properties. To further evaluate the performance of NEP89 in atomistic MD simulations, we compare dynamical properties of three typical systems with different structural phases, including amorphous carbon, liquid water, and three solid-state electrolytes, obtained using NEP89 and other foundation models (see Methods for computational details). Figure 3c shows the bonding statistics of amorphous carbon, comparing experimental data with predictions from NEP89 and other foundation models. NEP89 achieves comparable accuracy to other foundation models and aligns reasonably well with experimental results. Figure 3d–f shows a comparative analysis of the density, radial distribution function, and self-diffusion coefficient of liquid water as obtained from experiments, NEP89, and other foundation models. The results show that the NEP89 predictions closely match exper-

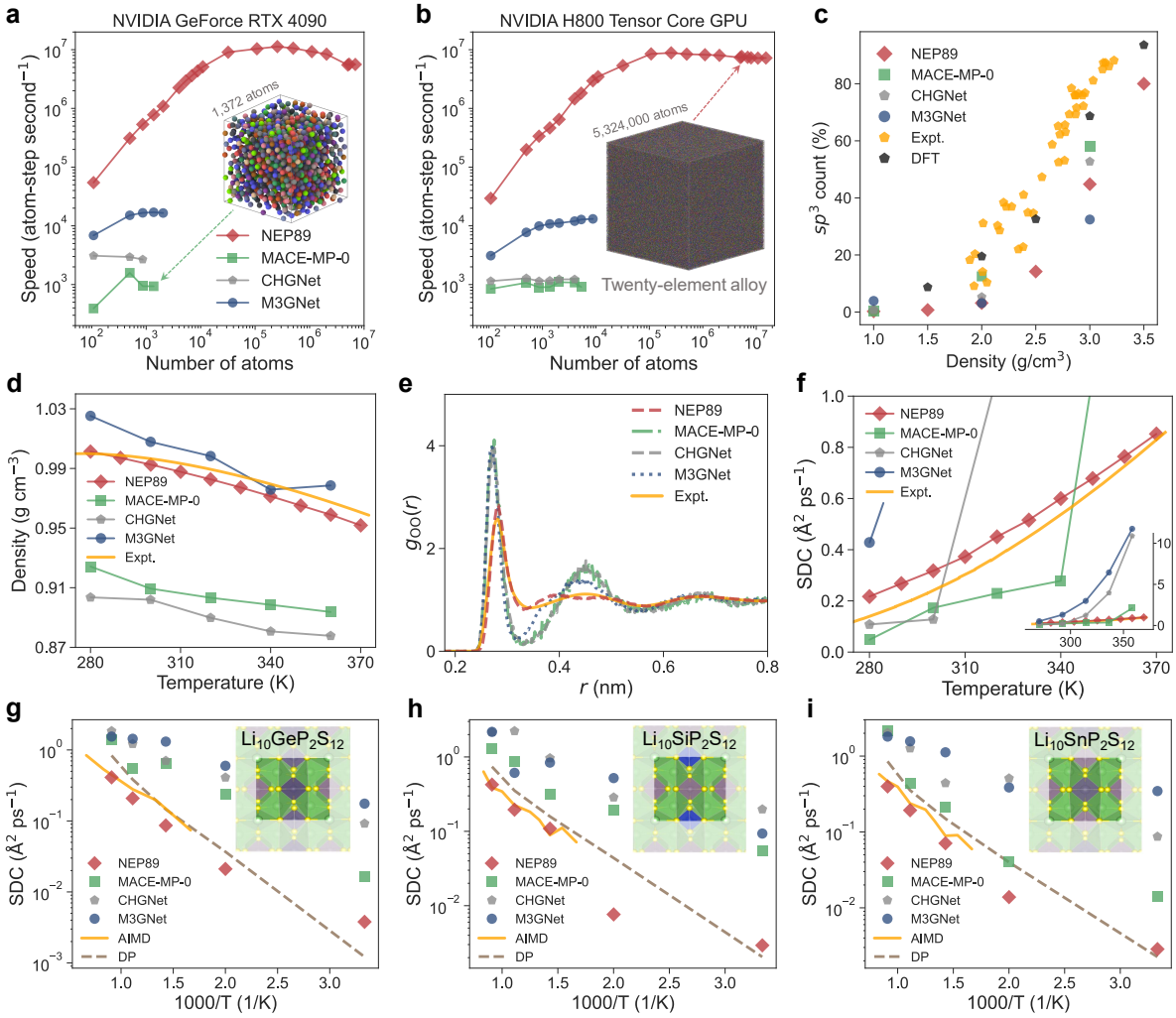

FIG. 3. **Computational performance and benchmarks of dynamical properties**. (**a**) and (**b**) show the benchmark results of computational efficiency of NEP89 and various foundation models in twenty-element alloys. Insets show representative atomic snapshots of the alloy model used in the MD simulations. NEP89 achieves about a three-order-of-magnitude improvement in computational efficiency compared to the other foundation models. (**c**) Comparison of experimental and predicted bonding statistics of amorphous carbon. (**d**–**f**) Experimental and predicted density (**d**), radial distribution function (**e**), and self-diffusion coefficient (SDC) (**f**) of liquid water simulated using NEP89 and other foundation models. (**g**-**i**) Experimental and calculated lithium-ion SDC in (**g**) $Li_{10}GeP_2S_{12}$, (**h**) $Li_{10}SiP_2S_{12}$, and (**i**) $Li_{10}SnP_2S_{12}$ solid-state electrolytes using NEP89, three other foundation models, a specialized DP model, and AIMD. Overall, predictions from NEP89 show good agreement with experimental data or AIMD results.

imental data and that NEP89 outperforms the other foundation models, particularly in reproducing the density (Figure 3d) and the oxygen-oxygen radial distribution function (Figure 3e). Next, Figure 3g–i present a comparative analysis of calculated versus experimental lithium-ion self-diffusion coefficients for three thiophosphate solid-state electrolytes: $Li_{10}GeP_2S_{12}$, $Li_{10}SiP_2S_{12}$, and $Li_{10}SnP_2S_{12}$, spanning temperatures from 300 to 1100 K. Notably, NEP89 achieves excellent agreement with AIMD simulations and a specialized force field [25], while outperforming other foundation models across the full temperature range.

Out-of-the-box large-scale MD simulations. After demonstrating the capabilities of NEP89 in dynamical simulations using small systems designed for benchmarking purposes, we next showcase the out-of-the-box applicability of NEP89 for large-scale atomistic MD simulations of inorganic and organic materials, which are typically impractical for other foundation models. The examples include a million-atom-scale compression simulation of the complex multicomponent $Cu_{0.7}Mo_{25.0}Ta_{29.6}V_{17.0}W_{27.7}$ (CuMoTaVW) alloy, calculations of ion diffusivities in water, a study of rocksalt dissolution kinetics, a simulation of methane combustion, and a dynamical simulation of a protein-ligand system (see Methods for details).

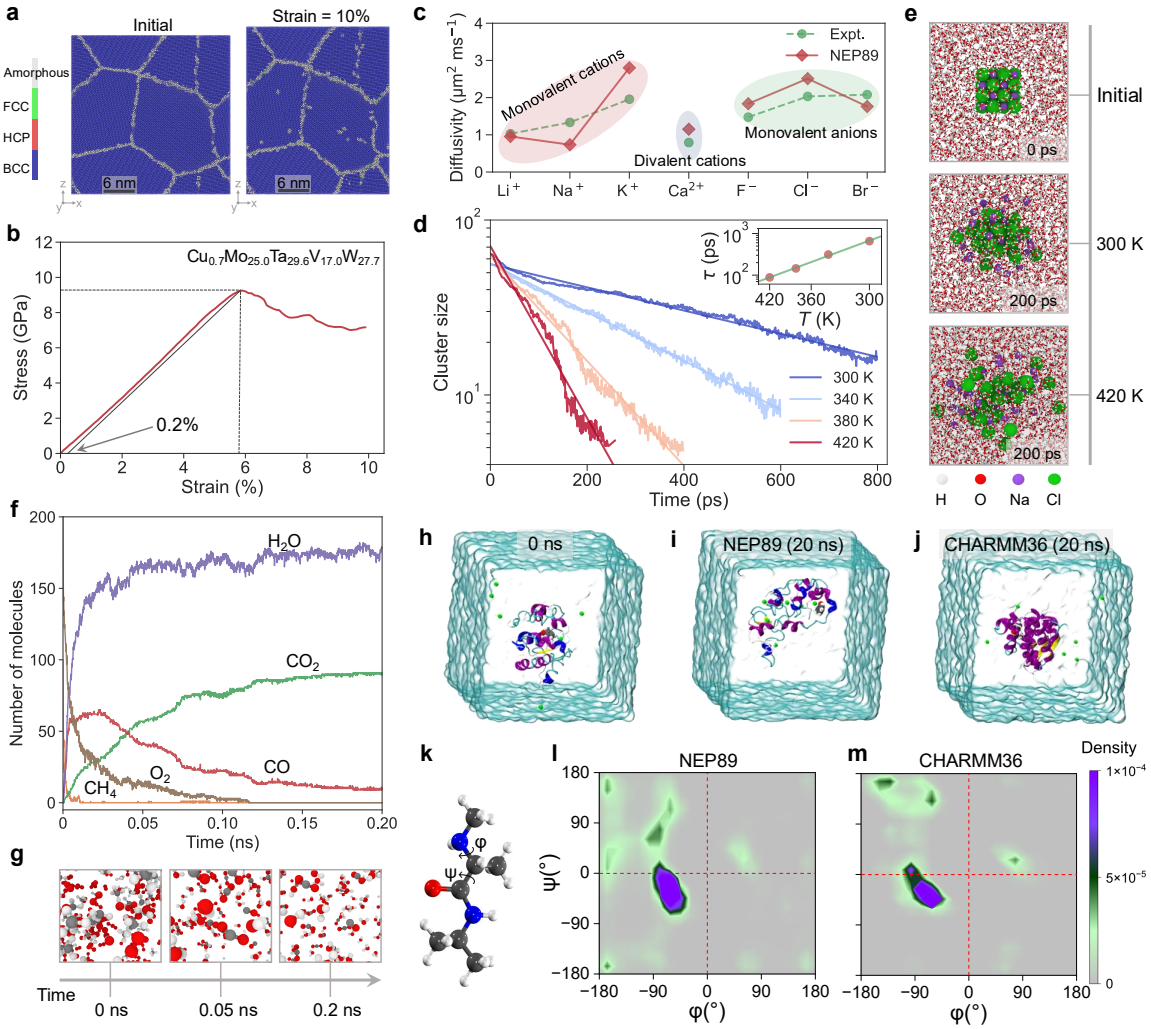

FIG. 4. **NEP89 applications in inorganic and organic systems.** (**a**) Snapshots of atomic configurations of $Cu_{0.7}Mo_{25.0}Ta_{29.6}V_{17.0}W_{27.7}$ alloy at different stages during MD simulations (1 221 240 atoms) using NEP89. The left panel shows the initial configuration; the right panel shows the configuration under 10% strain, colored by common neighbor analysis. Dark blue, red, and green indicate atoms in body-centered cubic, hexagon close-packed, and face-centered cubic environments, respectively; off-white indicates atoms in amorphous environments. (**b**) Stress-strain curve and yield strength predicted using NEP89. The dashed line indicates the 0.2% offset method used to determine the yield strength. The NEP89 prediction is in good agreement with experimentally measured yield strength ($\sim$ 10.0 GPa) [36]. (**c**) Diffusivity of various ions in water at 300 K predicted by NEP89, compared to experimental values from Ref. [37]. (**d**) Dissolution of NaCl nanocrystal in water at different temperatures calculated by NEP89. The dissolution rate τ is extracted from the power law fits (straight lines) at each temperature, and collected in the inset, illustrating that NEP89 captures the expected Arrhenius behavior of the dissolution process. (**e**) Atomic snapshots of the spatial distribution of NaCl during the dissolution in water at different temperatures. (**f**) Time evolution of O_2, CH_4, and major products (CO, CO_2, H_2O) during a methane combustion simulation. (**g**) Atomic snapshots of initial reactants, intermediate species, and final products. The results are consistent with those of Zhang *et al.* [22], showing that NEP89 qualitatively captures the reactive chemistry of methane combustion. (**h**) Initial protein-ligand model; (**i, j**) Configurations obtained after 20 ns of MD simulations using NEP89 **(i)** and CHARMM36 **(j)**. (**k**) Schematic diagram of the backbone rotation angle (ϕ and ψ) used in the Ramachandran diagram. (**l, m**) Ramachandran diagram showing the distribution of ϕ and ψ angles from simulations using NEP89 (**l**) and CHARMM36 (**m**), demonstrating that NEP89 qualitatively captures key aspects of protein-ligand interactions and overall protein structure.

For the CuMoTaVW alloy, common neighbor analysis (Figure 4a) shows that NEP89 successfully maintains structural stability during isothermal equilibration and compression. The polycrystalline body-centered cubic structure remains intact throughout the deformation process, with no observable phase transformation. The yield strength and yield strain calculated by NEP89 are 9.2 GPa and 5.8%, respectively (Figure 4b), which closely reproduce the experimentally measured yield strength ($\sim$10.0 GPa) and yield strain ($\sim$6%) [36]. This quan-

titative agreement illustrates the ability of NEP89 to accurately capture atomic interactions in complex multi-component alloy systems.

Moreover, the diffusivities of various monovalent and divalent ions in water predicted by the NEP89 model (Figure 4c), and their corresponding activation energies (Supp. Figure S9), are in good agreement with experiments [37, 38]. In combination with computational efficiency, this enables the prediction of, for example, the dissolution kinetics of a small salt crystallite in water (Figure 4d,e, Supp. Figure S10). The expected Arrhenius behavior of the dissolution process is captured out of the box with NEP89 (Figure 4d, inset).

The methane combustion simulation (Figure 4f) shows that NEP89 generates primary reaction products and product distributions that are remarkably similar to those obtained with the ANI-1xnr model [22], indicating the capability of NEP89 to capture the underlying reaction physics and mechanisms. Atomic snapshots during combustion (Figure 4g) confirm that NEP89 also provides reasonable predictions of the reaction chemistry. Furthermore, NEP89 reproduces the near-exponential decay profiles of CH_4 and O_2 concentrations originally observed in ANI-1xnr simulations [22], demonstrating its robust performance in modeling organic reactions.

For the protein-ligand case, we simulate the dynamics of a protein (T4 lysozyme L99A/M102Q, 3HTB) in complex with a ligand (2-propylphenol, JZ4) using NEP89 and compare the results with those obtained using the specialized protein-optimized CHARMM36 force field. As shown in Figure 4h–j, both models achieve the stable binding of the ligand JZ4 to the protein 3HTB in long-term simulations. The binding energy calculated by NEP89 is $-1.64\,\mathrm{eV}$, while the reference value from CHARMM36 is $-1.24\,\mathrm{eV}$, suggesting that NEP89 can reasonably describe the protein-ligand interactions. In addition, we calculated the Ramachandran diagram of the protein obtained by the two potentials, which can be used to analyze whether the conformation of the protein model conforms to the rules of stereochemistry. As shown in the Figure 4l,m, the total allowed areas calculated by NEP89 and CHARMM36 are 63.1% and 90.47%, respectively. While NEP89 shows a lower proportion of allowed areas compared to the specialized CHARMM36 force field, both Ramachandran diagrams [39] show that the highest probability of protein configuration is around the $(-60°, -30°)$ area, with a higher proportion of α area (bottom left area in Figure 4l,m) than that of β area (top left area in Figure 4l,m). This indicates that NEP89 can qualitatively describe the overall protein structure. We also calculated the number of protein-water hydrogen bonds, obtaining 220 for NEP89 and 375 for CHARMM36, which are of the same order of magnitude. Overall, these results demonstrate that NEP89 can qualitatively capture key aspects of protein-ligand interactions and protein structure, even though it is not specifically optimized for these protein systems. Further improvements are expected through fine-tuning, which we will introduce next.

Fine-tuning applications. For some specific applications, the accuracy of NEP89 may not be sufficient. In such cases, one can still fully leverage NEP89 to quickly obtain a reliable model through a simple but effective active learning process. We can first use NEP89 to perform atomistic simulations at our target conditions, generating abundant trajectories. By sampling structures from the trajectories and performing single-point DFT calculations, we can create a training dataset to fine-tune NEP89 into a special-purpose model for the material and application in question. This process is considerably more efficient than sampling structures using AIMD. In the following, we present four case studies, including mechanical properties of the 2D material $MoSi_2N_4$, thermal and structural properties of $Pd_{42.5}Cu_{30}Ni_{7.5}P_{20}$ metallic glass, thermal conductivity of monolayer MoS_2, and simulated inelastic neutron scattering spectra for crystalline benzene (see Methods for fine-tuning and simulation details).

$MoSi_2N_4$ has been successfully synthesized via chemical vapor deposition in recent years and exhibited promising properties [40]. Here, we showcase the fine-tuning of NEP89 specifically for $MoSi_2N_4$ with only 100 additional configurations. The out-of-the-box predictions by NEP89 for the fracture strength and the Young's modulus are $(272 \pm 1)\,\mathrm{GPa}$ and $(34.2 \pm 0.4)\,\mathrm{GPa}$, respectively. The fine-tuned model on the other hand yields $(463 \pm 4)\,\mathrm{GPa}$ and $(45.7 \pm 0.2)\,\mathrm{GPa}$, respectively, which is in significantly better agreement with the experimental values of $(491 \pm 139)\,\mathrm{GPa}$ and $(66 \pm 18)\,\mathrm{GPa}$, respectively (Figure 5a–d).

The $Pd_{42.5}Ni_{7.5}Cu_{30}P_{20}$ alloy is among the best-known bulk metallic glasses in terms of glass-forming ability [48]. Here, we demonstrate that fine-tuning NEP89 with only around 200 additional configurations yields excellent performance. After fine-tuning, the computed glass transition temperature T_g matches closely with both the prediction from a specialized NEP model [41] and experimental data [42], as shown in Figure 5e–f. Moreover, the icosahedral coordination, which describes the short-range structure, is quantitatively improved after fine-tuning, achieving results comparable to a specialized NEP model (Figure 5g) [41].

Next, we demonstrate that fine-tuning the NEP89 model with only 104 additional configurations enables accurate prediction of the thermal conductivity of the monolayer MoS_2. Figure 5h shows the evolution of the predicted thermal conductivity using the homogeneous nonequilibrium MD method [49] as a function of the number of fine-tuning steps. As the latter increases, the predicted thermal conductivity gradually rises and eventually converges to the reference values. Notably, after 10 000 fine-tuning steps, the model yields thermal conductivity values that closely match those reported using a customized NEP potential [43] as well as the Boltzmann transport equation approach using force constants

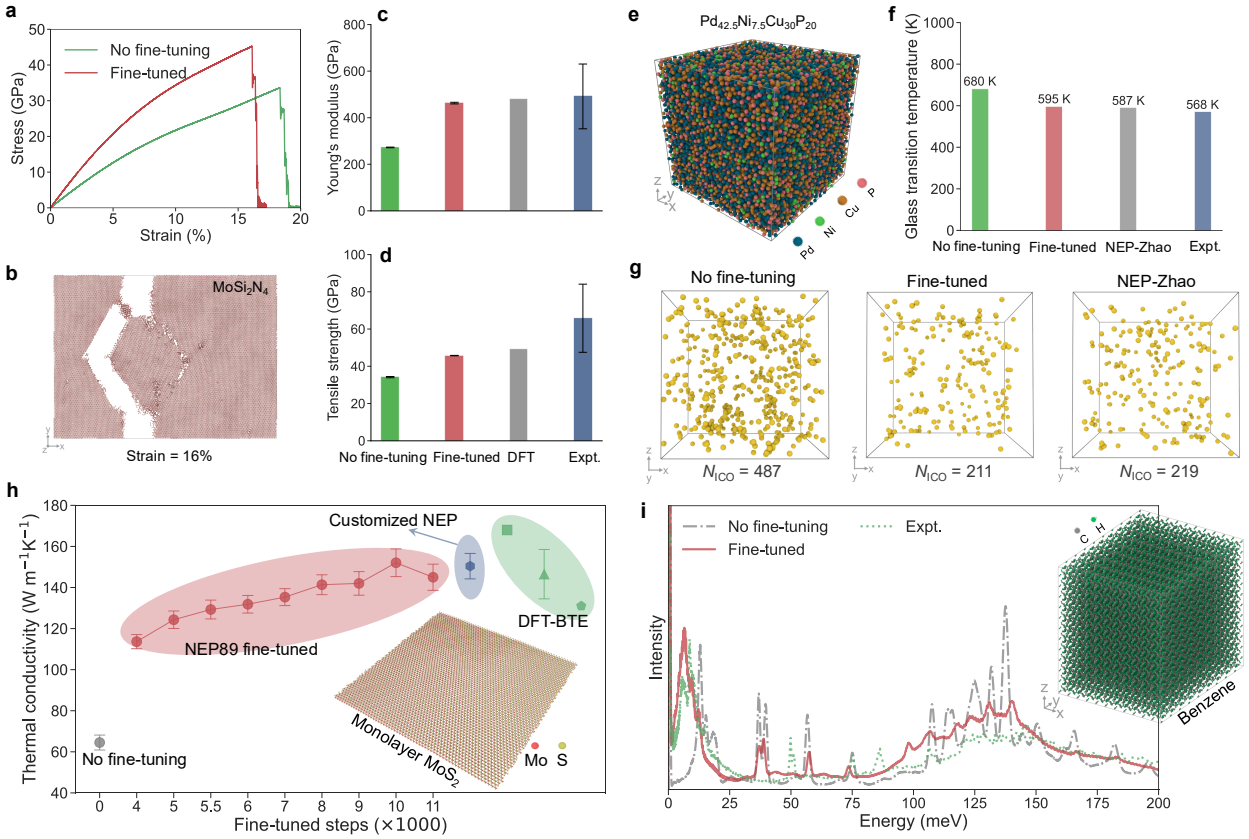

FIG. 5. **Example applications enabled by fine-tuning NEP89. (a–d)** Mechanical properties of monolayer $MoSi_2N_4$. (**a**) Stress-strain curves averaged over five MD simulations at 300 K. (**b**) Atomic snapshot at 16% strain after fine-tuning, showing fracture. (**c,d**) Comparison of the Young's modulus and tensile strength before and after fine-tuning, alongside DFT calculations and experimental measurements [40]. Error bars for the MD results represent the standard deviation across five simulations. **(e–g)** Thermal and structural properties of $Pd_{42.5}Cu_{30}Ni_{7.5}P_{20}$ metallic glass. **(e)** Atomic snapshots of the quenched metallic glass structure generated by a fine-tuned model based on NEP89. **(f)** Comparison of T_g values before and after fine-tuning, with results obtained using the NEP model by Zhao *et al.* [41] and experimental data [42]. **(g)** Spatial distribution and number (N_{ICO}) of icosahedral structures (colored in gold for central atoms) in a quenched glass predicted by different potentials. **(h)** Thermal conductivity of monolayer MoS_2. Results are compared with predictions from a customized NEP model [43] and Boltzmann transport equation (BTE) calculations using DFT-derived force constants [44–46]. **(i)** Simulated inelastic neutron scattering spectra for crystalline benzene at 127 K using NEP89 and a fine-tuned model, compared with experimental data [47].

obtained via DFT calculations [44–46].

Finally, we showcase the prediction of scattering experiments based on the workflow outlined in [47] (see Methods for details), employing NEP89 and a fine-tuned version trained on 120 additional structures obtained via active learning. While the simulated spectrum from NEP89 is overall in qualitative agreement with experimental data [47] (Figure 5i), the peaks in the spectrum, which correspond to phonon modes, deviate from their experimental counterpart by about 10 meV. After fine-tuning, the predicted mode frequencies agree even better with the experimental spectrum, barring a general red shift of the broad feature at 100 meV, which can be attributed to the vdW-DF-cx exchange correlation functional used for the DFT calculations of the structures for fine-tuning. Moreover, the relative peak intensities are in better agreement with experiments as well.

III. DISCUSSION

In summary, we have introduced NEP89, a highly efficient and accurate foundation model for large-scale atomistic simulations of both inorganic and organic materials across 89 elements. To the best of our knowledge, no existing machine-learned interatomic potential model offers this combination of elemental and chemical breadth with inference speeds comparable to typical empirical potentials. NEP89 represents a significant milestone toward the long-standing goal of a unified computational framework that delivers near-ab-initio accuracy and empirical-potential-like computational efficiency across diverse material systems.

To achieve this, we employed an iterative training strategy incorporating descriptor-space subsampling, targeted data curation, and reference energy adjustments. The resulting training dataset strikes a balance

between diversity and compactness, supporting reliable predictions across a wide range of static and dynamic properties, and remains flexible for future extension.

Comprehensive benchmarks highlight the competitive performance of NEP89 in static properties, including formation energies, binding energies, adhesion energies, lattice energies, vacancy formation energies, phonon frequencies, and elastic moduli. NEP89 also demonstrates robust predictions in dynamic properties, such as bonding statistics in amorphous carbon, thermodynamic properties of liquid water, and lithium-ion transport in solid-state electrolytes. NEP89 often matches or outperforms representative foundation models, exhibiting good agreement with experimental results and robustness across diverse applications.

Beyond benchmarks on small-scale systems, we demonstrated the out-of-the-box capability of NEP89 in large-scale MD simulations, enabled by its computational efficiency, which is comparable to that of empirical potentials. From large-scale simulations of complex alloy systems and solution chemistry to modeling methane combustion and protein-ligand dynamics, NEP89 showcases versatility across a wide range of inorganic and organic systems. Notably, NEP89 achieves quantitative agreement with experiments on yield strength and ion diffusivities in water, reproduces the expected Arrhenius behavior in rocksalt dissolution kinetics, and matches specialized model predictions for combustion reaction dynamics. In addition, it qualitatively captures key structural features and interactions in protein systems. These results highlight the ability of NEP89 to model complex systems with high fidelity, even without domain-specific optimization.

The fine-tuning capabilities of NEP89 provide a practical pathway for rapidly developing specialized models from limited training data. Researchers can efficiently customize NEP89 for specific materials or conditions, as demonstrated in case studies on $MoSi_2N_4$ mechanical properties, thermal and structural properties of $Pd_{42.5}Cu_{30}Ni_{7.5}P_{20}$ metallic glass, thermal conductivity of monolayer MoS_2, and neutron scattering simulations of crystalline benzene. Fine-tuning yields substantially improved agreement with results from quantum-mechanical calculations and experiments, demonstrating the utility of NEP89 as both a versatile out-of-the-box and an adaptable foundation model.

While the demonstrated out-of-the-box and fine-tuning applications are not exhaustive, they provide a solid foundation for future explorations in various application domains. Such efforts will further characterize the capabilities of NEP89 and identify potential areas for improvement. We expect that with ongoing extension and refinement, the out-of-the-box performance of NEP89 will continue to improve. Additionally, we anticipate the development of numerous domain-specific, medium-sized models fine-tuned from NEP89, enabling tailored solutions for a wide range of applications.

IV. METHODS

Benchmark calculations for static properties. Static properties were calculated with the help of the GPUMD-WIZARD [50], ASE [51], CALORINE [52], PHONOPY [35], and MATCALC [53] packages. Detailed computational procedures can be found in Supp. Notes S5–S10 . Atomic snapshots were generated using OVITO [54].

Computational performance evaluation. We performed MD simulations with 1000 steps on a 20-element equiatomic alloy composed of Mn, Cr, Fe, Co, Ni, Cu, Ag, W, Mo, Nb, Al, Cd, Sn, Pb, Bi, Zn, Ge, Si, Sb, and Mg [55] to evaluate the computational efficiency of different foundation models. The initial structure of the alloy is face-centered cubic with randomly distributed atoms and a lattice constant of 3.7 Å. Simulations were conducted on systems ranging from the smallest unit cell to progressively larger supercells until memory limitations were reached.

MD simulation of amorphous carbon. A system with 512 atoms in a diamond structure with a given density undergoes an initial rapid melting process at 9000 K for 25 ps, followed by a relaxation at 5000 K for 5 ps. This is followed by a rapid quenching from 5000 K down to 1000 K in 0.4 ps, with further relaxation stages at 1000 K for 5 ps and subsequently at 300 K for 5 ps. The time step for integration is 0.5 fs.

MD simulation of water. A system with 699 atoms with an initial density of $1.0\,\mathrm{g\,cm^{-3}}$ was used in the MD simulations of water. The system was equilibrated in the isothermal-isobaric ensemble at a given temperature and a pressure of 1 bar for 100 ps and then in the canonical ensemble for 50 ps, followed by a production run in the microcanonical ensemble for 25 ps. The time step for integration was 0.5 fs. For each temperature, five independent runs were performed.

Lithium-ion diffusion in solid-state electrolytes. MD simulations were performed on systems with 900 atoms for each of the three thiophosphate solid-state electrolytes: $Li_{10}GeP_2S_{12}$, $Li_{10}SiP_2S_{12}$, and $Li_{10}SnP_2S_{12}$. Each system was equilibrated in the isothermal-isobaric ensemble at a given temperature and a pressure of 1 bar for 50 ps and then in the canonical ensemble for 100 ps, followed by a production run in the microcanonical ensemble for 200 ps. The time step for integration is 0.5 fs. For each material and each temperature, three independent runs were performed.

MD simulation of CuMoTaVW alloy. The initial configuration comprises a body-centered cubic polycrystalline structure with 10 grains (average grain size about 12 nm), forming a 27 nm × 27 nm × 27 nm simulation box containing 1 221 240 atoms. Using a hybrid MC/MD method, we performed 4 million Monte

Carlo attempts to obtain a low-energy configuration. The system was then equilibrated at 300 K for 100 ps under the isothermal-isobaric ensemble. Subsequently, uniaxial compression was applied along the x-axis at an engineering strain rate of 2×10^8 s^{-1}, up to a maximum strain of 10%. Yield strength is determined using the commonly adopted 0.2% offset method, which identifies the point where the stress-strain curve intersects a line parallel to the linear (elastic) region, offset by 0.2% strain.

Ion diffusivity in water. To compute the diffusivity of Li^+, Na^+, K^+, Ca^{2+}, F^-, Cl^-, and Br^-, 14 (7) monovalent (divalent) ions along with the corresponding number of OH^- or H^+ ions were randomly mixed with 14 000 H_2O molecules at an initial density of 1.0 g cm^{-3}. The systems were equilibrated in the isothermal-isobaric ensemble at a given temperature and a pressure of 0 bar for 200 ps, followed by a production run in the micro-canonical ensemble for 400 ps using a time step of 0.2 fs. During the latter run, the mean-square displacement was recorded, from which the diffusivity was calculated.

Dissolution kinetics of NaCl in water. A nanocrystal comprising 32 NaCl units was placed in a cell surrounded by 14 000 H_2O molecules at a density of 1.0 g cm^{-3}. The system was then evolved in the canonical ensemble at a given temperature for up to 2 ns using a time step of 0.2 fs. The size of the largest cluster was tracked over time, where the clusters were defined by connectivity using a maximum Na–Cl bond cutoff of 4 Å calculated using OVITO [54]. 10 runs were carried out at each temperature, and the crystal size was averaged over these runs. The dissolution rate for each temperature was determined by fitting the time dependence of the cluster size to a simple exponential law.

MD simulation of methane combustion. CH_4 and O_2 molecules were mixed into a reactive system containing 100 CH_4 and 200 O_2 molecules, enclosed in a cubic cell of dimensions 37.5 Å × 37.5 Å × 37.5 Å, corresponding to a density of 0.25 g cm^{-3}. Each trajectory began with isothermal equilibration at 300 K; subsequently, the temperature was ramped to trigger combustion. The timestep for integration is 0.1 fs.

MD simulation of protein dynamics. Dynamics of a protein (T4 lysozyme L99A/M102Q, 3HTB) in complex with a ligand (2-propylphenol, JZ4) was simulated using both GROMACS [56] with the CHARMM36 [57] (for 3HTB) and CGenFF [58] (for JZ4), as well as GPUMD with NEP89. The system consists of 33 917 atoms, and the simulation was performed for 20 ns in the isothermal-isobaric ensemble using a timestep of 0.5 fs. The last 15 ns were used for data analysis.

Fine-tuning for $MoSi_2N_4$ and MD simulations. Using the NEP89 model, we conducted tensile loading MD simulations of $MoSi_2N_4$ using an orthorhombic cell with 210 atoms at 300 K and evenly sampled 100 structures from the trajectory. Then we performed single-point DFT-D3(BJ) calculations using the VASP package [59, 60] with the projected augmented wave method and the generalized gradient approximation, along with the Perdew-Burke-Ernzerhof (PBE) functional [61]. A vacuum layer of 20 Å was used to model the two-dimensional systems. The plane-wave truncation energy was set to 500 eV, and a Γ-centered k-point grid with $2 \times 2 \times 1$ divisions was used. Using the new training dataset, we trained 5000 steps to obtain the fine-tuned model for $MoSi_2N_4$. Using both the NEP89 and fine-tuned models, we performed tensile loading simulations for the monolayer $MoSi_2N_4$, using an orthorhombic cell with 44 800 atoms. The engineering strain rate in the herringbone direction was 8.6×10^7 s^{-1}. The thickness of the monolayer was taken as 10.7 Å in calculating the volume, following earlier literature [40]. We conducted five independent tensile loading simulations and averaged them to report the stress-strain curves and standard deviations.

Fine-tuning for PdCuNiP metallic glass and MD simulations. To perform fine-tuning for the $Pd_{42.5}Cu_{30}Ni_{7.5}P_{20}$ metallic glass, we carried out quenching simulations for three specific compositions: $Pd_{40}Cu_{30}Ni_{10}P_{20}$, $Pd_{40}Ni_{40}P_{20}$, and $Pd_{42.5}Cu_{30}Ni_{7.5}P_{20}$ using the NEP89 model. The simulation protocol involved constructing a 108-atom supercell equilibrated at 1800 K in the NPT ensemble for 1 ns, followed by rapid quenching to 300 K at a cooling rate of 5×10^{10} K s^{-1}. The structure was then equilibrated at 300 K for 1 ns. From the resulting trajectory, 209 configurations were sampled. Single-point calculations were performed using VASP [59, 60] with the PBE functional [61], a Γ-centered k-point grid with a spacing of 0.2 Å^{-1}, a plane wave energy cutoff of 600 eV, and a threshold for the self-consistency loop of 10^{-6} eV. After labeling the 209 structures with DFT reference data, the NEP89 model was fine-tuned for 5000 steps. For the calculation of the glass transition temperature (T_g) and short-range order analysis of $Pd_{42.5}Cu_{30}Ni_{7.5}P_{20}$, a system containing 32 000 atoms randomly arranged in a cubic box with dimensions of 7.6 nm × 7.6 nm × 7.6 nm was equilibrated in the NPT ensemble (zero pressure) at 1800 K for 1 ns with a timestep of 1 fs, followed by rapid cooling to 300 K at a rate of 5×10^{10} K s^{-1}, and further equilibrated at 300 K for 1 ns (see Supp. Figure S11 for potential energy curves and T_g). The polyhedral template matching method [62] with a root-mean-square deviation cutoff of 0.12 was employed to identify local structural motifs such as icosahedral clusters and other short-range order structures.

Fine-tuning for MoS_2 and thermal conductivity calculations. Using the NEP89 model, we performed a 6 ns NPT simulation of MoS_2 in an orthorhombic box

containing 90 atoms at 300 K, and applied the farthest point sampling method to sample 104 configurations from the trajectory. Single-point DFT-D3(BJ) calculations were performed with the VASP package [59, 60] with the PBE functional [61], a k-space with a spacing of 0.15 Å^{-1}, a threshold for the self-consistency loop of 10^{-6} eV, and a plane wave energy cutoff of 520 eV. The dataset with 104 structures was used to fine-tune the NEP89 model, yielding models for monolayer MoS_2 with different fine-tuned steps. Using the homogeneous nonequilibrium MD method [49], we calculated the thermal conductivity of a MoS_2 model containing 11 484 atoms. Ten independent 10 ns simulations were conducted, and the standard error was calculated from the results.

Fine-tuning for crystalline benzene and inelastic neutron scattering calculations. Energies and forces for the active-learning training structures were evaluated using the vdW-DF-cx exchange-correlation functional [63]. Crystalline benzene was simulated using both the NEP89 and fine-tuned models in a 57 024-atom supercell at 127 K, with a timestep of 0.5 fs and positions saved every 3 fs. Path-integral MD simulations with 32 beads ensured accurate cell volumes at low temperature [64]. After equilibration, a 1 ns microcanonical production run was performed. The dynamic structure factor $S(\boldsymbol{q}, \omega)$ was computed from the resulting trajectories following Ref. [47], briefly summarized here. $S(\boldsymbol{q}, \omega)$ was computed using DYNASOR [65] for 2116 random $\boldsymbol{q}$-points up to $|\boldsymbol{q}| = 14\ \text{Å}^{-1}$. Results were smoothed with a Gaussian of width 0.01 Å^{-1} and averaged over spherical $|\boldsymbol{q}|$ shells to obtain $S(q, \omega)$. $S(q, \omega)$ was then weighted for neutron probes using species-specific scattering lengths and convolved with the kinematic constraint and resolution function of the TOSCA spectrometer at the ISIS Neutron and Muon Source, UK. Finally, the weighted $S(q, \omega)$ was integrated to yield the spectrum $S(\omega) = \int S(q, \omega) \mathrm{d}q$. The TOSCA resolution function, implemented via EUPHONIC [66] and ResINS [67], is based on the AbINS module in Mantid [68].

ACKNOWLEDGMENTS

Z. F. was supported by the National Science and Technology Advanced Materials Major Program of China (No. 2024ZD0606900). T. L., K. X., and J.-B. X. acknowledge support from the National Key R&D Project from the Ministry of Science and Technology of China (No. 2022YFA1203100), RGC Grants (No. 14220022, JLFS/E-402/24), and the CUHK PhD Studentship and Postdoctoral Fellowship. E. L., E. B., and P. E. gratefully acknowledge funding from the Swedish Research Council (Nos. 2020-04935 and 2021-05072), the Swedish Foundation for Strategic Research via the SwedNESS graduate school (GSn15-0008) as well as computational resources provided by the National Academic Infrastructure for Supercomputing in Sweden at NSC, PDC, and C3SE partially funded by the Swedish Research Council through grant agreement No. 2022-06725, as well as the Berzelius resource provided by the Knut and Alice Wallenberg Foundation at NSC. T. A-N. and Y. W. have been supported in part by the Academy of Finland through its QTF Center of Excellence program (project no. 312298) and European Union – NextGenerationEU instrument grant 353298. Computational resources by the CSC IT Centre for Finland and the Aalto Science-IT are also gratefully acknowledged.

Data availability: The NEP89 model has been made available as part of the GPUMD 4.0 distribution. The associated training data will be released upon the publication of this paper.

Code availability: The source code for GPUMD (version 4.0) is available at the Zenodo repository https://doi.org/10.5281/zenodo.15299684 [69] and the GitHub repository https://github.com/brucefan1983/GPUMD/releases/tag/v4.0.

Declaration of competing interest:

The authors declare that they have no competing interests.

[1] J. Behler and M. Parrinello, Generalized neural-network representation of high-dimensional potential-energy surfaces, Phys. Rev. Lett. **98**, 146401 (2007).
[2] A. P. Bartók, M. C. Payne, R. Kondor, and G. Csányi, Gaussian approximation potentials: The accuracy of quantum mechanics, without the electrons, Physical Review Letters **104**, 136403 (2010).
[3] J. Behler, Perspective: Machine learning potentials for atomistic simulations, The Journal of Chemical Physics **145**, 170901 (2016).
[4] F. Noé, A. Tkatchenko, K.-R. Müller, and C. Clementi, Machine learning for molecular simulation, Annual Review of Physical Chemistry **71**, 361 (2020).
[5] O. T. Unke, S. Chmiela, H. E. Sauceda, M. Gastegger, I. Poltavsky, K. T. Schütt, A. Tkatchenko, and K.-R. Müller, Machine learning force fields, Chemical Reviews **121**, 10142 (2021).
[6] V. L. Deringer, M. A. Caro, and G. Csányi, Machine learning interatomic potentials as emerging tools for materials science, Advanced Materials **31**, 1902765 (2019).
[7] Y. Mishin, Machine-learning interatomic potentials for materials science, Acta Materialia **214**, 116980 (2021).
[8] A. P. Bartók, J. Kermode, N. Bernstein, and G. Csányi, Machine Learning a General-Purpose Interatomic Poten-

tial for Silicon, Phys. Rev. X **8**, 041048 (2018).
[9] K. Song, R. Zhao, J. Liu, Y. Wang, E. Lindgren, Y. Wang, S. Chen, K. Xu, T. Liang, P. Ying, N. Xu, Z. Zhao, J. Shi, J. Wang, S. Lyu, Z. Zeng, S. Liang, H. Dong, L. Sun, Y. Chen, Z. Zhang, W. Guo, P. Qian, J. Sun, P. Erhart, T. Ala-Nissila, Y. Su, and Z. Fan, General-purpose machine-learned potential for 16 elemental metals and their alloys, Nature Communications **15**, 10208 (2024).
[10] S. Takamoto, C. Shinagawa, D. Motoki, K. Nakago, W. Li, I. Kurata, T. Watanabe, Y. Yayama, H. Iriguchi, Y. Asano, *et al.*, Towards universal neural network potential for material discovery applicable to arbitrary combination of 45 elements, Nature Communications **13**, 1 (2022).
[11] D. Zhang, X. Liu, X. Zhang, C. Zhang, C. Cai, H. Bi, Y. Du, X. Qin, J. Huang, B. Li, Y. Shan, J. Zeng, Y. Zhang, S. Liu, Y. Li, J. Chang, X. Wang, S. Zhou, J. Liu, X. Luo, Z. Wang, W. Jiang, J. Wu, Y. Yang, J. Yang, M. Yang, F.-Q. Gong, L. Zhang, M. Shi, F.-Z. Dai, D. M. York, S. Liu, T. Zhu, Z. Zhong, J. Lv, J. Cheng, W. Jia, M. Chen, G. Ke, W. E, L. Zhang, and H. Wang, DPA-2: a large atomic model as a multi-task learner, npj Computational Materials **10**, 293 (2024).
[12] C. Chen and S. P. Ong, A universal graph deep learning interatomic potential for the periodic table, Nature Computational Science **2**, 718 (2022).
[13] B. Deng, P. Zhong, K. Jun, J. Riebesell, K. Han, C. J. Bartel, and G. Ceder, CHGNet as a pretrained universal neural network potential for charge-informed atomistic modeling, Nature Machine Intelligence **5**, 1031 (2023).
[14] A. Merchant, S. Batzner, S. S. Schoenholz, M. Aykol, G. Cheon, and E. D. Cubuk, Scaling deep learning for materials discovery, Nature **624**, 80 (2023).
[15] F. Xie, T. Lu, S. Meng, and M. Liu, GPTFF: A high-accuracy out-of-the-box universal AI force field for arbitrary inorganic materials, Science Bulletin **69**, 3525 (2024).
[16] I. Batatia, P. Benner, Y. Chiang, A. M. Elena, D. P. Kovács, J. Riebesell, X. R. Advincula, M. Asta, M. Avaylon, W. J. Baldwin, F. Berger, N. Bernstein, A. Bhowmik, S. M. Blau, V. Cărare, J. P. Darby, S. De, F. D. Pia, V. L. Deringer, R. Elijošius, Z. El-Machachi, F. Falcioni, E. Fako, A. C. Ferrari, A. Genreith-Schriever, J. George, R. E. A. Goodall, C. P. Grey, P. Grigorev, S. Han, W. Handley, H. H. Heenen, K. Hermansson, C. Holm, J. Jaafar, S. Hofmann, K. S. Jakob, H. Jung, V. Kapil, A. D. Kaplan, N. Karimitari, J. R. Kermode, N. Kroupa, J. Kullgren, M. C. Kuner, D. Kuryla, G. Liepuoniute, J. T. Margraf, I.-B. Magdău, A. Michaelides, J. H. Moore, A. A. Naik, S. P. Niblett, S. W. Norwood, N. O'Neill, C. Ortner, K. A. Persson, K. Reuter, A. S. Rosen, L. L. Schaaf, C. Schran, B. X. Shi, E. Sivonxay, T. K. Stenczel, V. Svahn, C. Sutton, T. D. Swinburne, J. Tilly, C. van der Oord, E. Varga-Umbrich, T. Vegge, M. Vondrák, Y. Wang, W. C. Witt, F. Zills, and G. Csányi, A foundation model for atomistic materials chemistry, (2024), arXiv:2401.00096 [physics.chem-ph].
[17] H. Yang, C. Hu, Y. Zhou, X. Liu, Y. Shi, J. Li, G. Li, Z. Chen, S. Chen, C. Zeni, M. Horton, R. Pinsler, A. Fowler, D. Zügner, T. Xie, J. Smith, L. Sun, Q. Wang, L. Kong, C. Liu, H. Hao, and Z. Lu, Mattersim: A deep learning atomistic model across elements, temperatures and pressures, (2024), arXiv:2405.04967 [cond-mat.mtrl-sci].
[18] Z. Fan, Z. Zeng, C. Zhang, Y. Wang, K. Song, H. Dong, Y. Chen, and T. Ala-Nissila, Neuroevolution machine learning potentials: Combining high accuracy and low cost in atomistic simulations and application to heat transport, Phys. Rev. B **104**, 104309 (2021).
[19] Z. Fan, Y. Wang, P. Ying, K. Song, J. Wang, Y. Wang, Z. Zeng, K. Xu, E. Lindgren, J. M. Rahm, A. J. Gabourie, J. Liu, H. Dong, J. Wu, Y. Chen, Z. Zhong, J. Sun, P. Erhart, Y. Su, and T. Ala-Nissila, GPUMD: A package for constructing accurate machine-learned potentials and performing highly efficient atomistic simulations, The Journal of Chemical Physics **157**, 114801 (2022).
[20] O. T. Unke, M. Stöhr, S. Ganscha, T. Unterthiner, H. Maennel, S. Kashubin, D. Ahlin, M. Gastegger, L. M. Sandonas, J. T. Berryman, A. Tkatchenko, and K.-R. Müller, Biomolecular dynamics with machine-learned quantum-mechanical force fields trained on diverse chemical fragments, Science Advances **10**, eadn4397 (2024).
[21] P. Eastman, P. K. Behara, D. L. Dotson, R. Galvelis, J. E. Herr, J. T. Horton, Y. Mao, J. D. Chodera, B. P. Pritchard, Y. Wang, G. De Fabritiis, and T. E. Markland, Spice, a dataset of drug-like molecules and peptides for training machine learning potentials, Scientific Data **10**, 11 (2023).
[22] S. Zhang, M. Z. Makoś, R. B. Jadrich, E. Kraka, K. Barros, B. T. Nebgen, S. Tretiak, O. Isayev, N. Lubbers, R. A. Messerly, and J. S. Smith, Exploring the frontiers of condensed-phase chemistry with a general reactive machine learning potential, Nature Chemistry **16**, 727 (2024).
[23] D. P. Kovács, J. H. Moore, N. J. Browning, I. Batatia, J. T. Horton, Y. Pu, V. Kapil, W. C. Witt, I.-B. Magdău, D. J. Cole, and G. Csányi, MACE-OFF: Short-Range Transferable Machine Learning Force Fields for Organic Molecules, Journal of the American Chemical Society **147**, 17598 (2025).
[24] L. Barroso-Luque, M. Shuaibi, X. Fu, B. M. Wood, M. Dzamba, M. Gao, A. Rizvi, C. L. Zitnick, and Z. W. Ulissi, Open materials 2024 (OMat24) inorganic materials dataset and models, (2024).
[25] R. Wang, M. Guo, Y. Gao, X. Wang, Y. Zhang, B. Deng, X. Chen, M. Shi, L. Zhang, and Z. Zhong, A pre-trained deep potential model for sulfide solid electrolytes with broad coverage and high accuracy, (2024).
[26] R. Ibragimova, M. S. Kuklin, T. Zarrouk, and M. A. Caro, Unifying the description of hydrocarbons and hydrogenated carbon materials with a chemically reactive machine learning interatomic potential, Chemistry of Materials **37**, 1094 (2025).
[27] Y. Zhai, A. Caruso, S. L. Bore, Z. Luo, and F. Paesani, A "short blanket" dilemma for a state-of-the-art neural network potential for water: Reproducing experimental properties or the physics of the underlying many-body interactions? The Journal of Chemical Physics **158**, 084111 (2023).
[28] S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu, The Journal of Chemical Physics **132**, 154104 (2010).
[29] S. Grimme, S. Ehrlich, and L. Goerigk, Effect of the

damping function in dispersion corrected density functional theory, Journal of Computational Chemistry **32**, 1456 (2011).
[30] T. Schaul, T. Glasmachers, and J. Schmidhuber, in *Proceedings of the 13th Annual Conference on Genetic and Evolutionary Computation*, GECCO '11 (Association for Computing Machinery, New York, NY, USA, 2011) pp. 845–852.
[31] V. Babin, C. Leforestier, and F. Paesani, Development of a "first principles" water potential with flexible monomers: Dimer potential energy surface, VRT spectrum, and second virial coefficient, Journal of Chemical Theory and Computation **9**, 5395 (2013).
[32] A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, and K. A. Persson, Commentary: The Materials Project: A materials genome approach to accelerating materials innovation, APL Materials **1**, 011002 (2013).
[33] J. Rezac, K. E. Riley, and P. Hobza, S66: A well-balanced database of benchmark interaction energies relevant to biomolecular structures, Journal of Chemical Theory and Computation **7**, 2427 (2011).
[34] F. Della Pia, A. Zen, D. Alfè, and A. Michaelides, Dmc-ice13: Ambient and high pressure polymorphs of ice from diffusion monte carlo and density functional theory, The Journal of Chemical Physics **157**, 134701 (2022).
[35] A. Togo, First-principles phonon calculations with Phonopy and Phono3py, Journal of the Physical Society of Japan **92**, 012001 (2023).
[36] S. Alvi, D. M. Jarzabek, M. G. Kohan, D. Hedman, P. Jenczyk, M. M. Natile, A. Vomiero, and F. Akhtar, Synthesis and mechanical characterization of a cumotawv high-entropy film by magnetron sputtering, ACS Applied Materials & Interfaces **12**, 21070 (2020).
[37] P. Vanýsek, Ionic Conductivity and Diffusion at Infinite Dilution, in *CRC Handbook of Chemistry and Physics*, edited by D. R. Lide (Taylor and Francis, Boca Raton, FL, 2006) pp. 5–76 to 5–78.
[38] L. G. Longsworth, Temperature Dependence of Diffusion in Aqueous Solutions, The Journal of Physical Chemistry **58**, 770 (1954).
[39] G. Ramachandran, C. Ramakrishnan, and V. Sasisekharan, Stereochemistry of polypeptide chain configurations, Journal of Molecular Biology **7**, 95 (1963).
[40] Y.-L. Hong, Z. Liu, L. Wang, T. Zhou, W. Ma, C. Xu, S. Feng, L. Chen, M.-L. Chen, D.-M. Sun, X.-Q. Chen, H.-M. Cheng, and W. Ren, Chemical vapor deposition of layered two-dimensional $MoSi_2N_4$ materials, Science **369**, 670 (2020).
[41] R. Zhao, S. Wang, Z. Kong, Y. Xu, K. Fu, P. Peng, and C. Wu, Development of a neuroevolution machine learning potential of Pd-Cu-Ni-P alloys, Materials & Design **231**, 112012 (2023).
[42] O. Haruyama, Thermodynamic approach to free volume kinetics during isothermal relaxation in bulk Pd-Cu-Ni-P_{20} glasses, Intermetallics **15**, 659 (2007), advanced Intermetallic Alloys and Bulk Metallic Glasses.
[43] W. Jiang, H. Bu, T. Liang, P. Ying, Z. Fan, J. Xu, and W. Ouyang, Accurate Modeling of Interfacial Thermal Transport in van der Waals Heterostructures via Hybrid Machine Learning and Registry-Dependent Potentials, (2025), arXiv:2505.00376 [physics.comp-ph].
[44] W. Bao, G. Chen, Z. Wang, and D. Tang, Bilateral phonon transport modulation of Bi-layer TMDCs (MX_2, M=Mo, W; X=S), International Journal of Thermal Sciences **179**, 107669 (2022).
[45] X. Gu, B. Li, and R. Yang, Layer thickness-dependent phonon properties and thermal conductivity of MoS_2, Journal of Applied Physics **119**, 085106 (2016).
[46] A. Cepellotti, G. Fugallo, L. Paulatto, M. Lazzeri, F. Mauri, and N. Marzari, Phonon hydrodynamics in two-dimensional materials, Nature Communications **6**, 1 (2015).
[47] E. Lindgren, A. J. Jackson, E. Fransson, E. Berger, S. Rudić, G. Škoro, R. Turanyi, S. Mukhopadhyay, and P. Erhart, Predicting neutron experiments from first principles: A workflow powered by machine learning, (2025), arXiv:2504.19352 [cond-mat.mtrl-sci].
[48] S. Hosokawa, J.-F. Bérar, N. Boudet, W.-C. Pilgrim, L. Pusztai, S. Hiroi, S. Kohara, H. Kato, H. E. Fischer, and A. Zeidler, Relationship between atomic structure and excellent glass forming ability in $Pd_{42.5}Ni_{7.5}Cu_{30}P_{20}$ metallic glass, Journal of Non-Crystalline Solids **596**, 121868 (2022).
[49] Z. Fan, H. Dong, A. Harju, and T. Ala-Nissila, Homogeneous nonequilibrium molecular dynamics method for heat transport and spectral decomposition with many-body potentials, Phys. Rev. B **99**, 064308 (2019).
[50] J. Liu, GPUMD-Wizard: A python package for generating and evaluating machine learning potentials, (2024).
[51] A. Hjorth Larsen, J. Jørgen Mortensen, J. Blomqvist, I. E. Castelli, R. Christensen, M. Dułak, J. Friis, M. N. Groves, B. Hammer, C. Hargus, E. D. Hermes, P. C. Jennings, P. Bjerre Jensen, J. Kermode, J. R. Kitchin, E. Leonhard Kolsbjerg, J. Kubal, K. Kaasbjerg, S. Lysgaard, J. Bergmann Maronsson, T. Maxson, T. Olsen, L. Pastewka, A. Peterson, C. Rostgaard, J. Schiøtz, O. Schütt, M. Strange, K. S. Thygesen, T. Vegge, L. Vilhelmsen, M. Walter, Z. Zeng, and K. W. Jacobsen, The atomic simulation environment–a Python library for working with atoms, Journal of Physics: Condensed Matter **29**, 273002 (2017).
[52] E. Lindgren, M. Rahm, E. Fransson, F. Eriksson, N. Österbacka, Z. Fan, and P. Erhart, calorine: A python package for constructing and sampling neuroevolution potential models, Journal of Open Source Software **9**, 6264 (2024).
[53] R. Liu, E. Liu, J. Riebesell, J. Qi, S. P. Ong, and T. W. Ko, MatCalc: A Python library for calculating materials properties from the potential energy surface (PES), `https://github.com/materialsvirtuallab/matcalc` (2024).
[54] A. Stukowski, Visualization and analysis of atomistic simulation data with OVITO–the Open Visualization Tool, Modelling and Simulation in Materials Science and Engineering **18**, 015012 (2009).
[55] B. Cantor, I. Chang, P. Knight, and A. Vincent, Microstructural development in equiatomic multicomponent alloys, Materials Science and Engineering: A **375-377**, 213 (2004).
[56] H. Berendsen, D. van der Spoel, and R. van Drunen, Gromacs: A message-passing parallel molecular dynamics implementation, Computer Physics Communications **91**, 43 (1995).
[57] J. Huang, S. Rauscher, G. Nawrocki, T. Ran, M. Feig, B. L. de Groot, H. Grubmüller, and A. D. MacKerell, CHARMM36m: an improved force field for folded and

intrinsically disordered proteins, Nature Methods **14**, 71 (2017).
[58] I. Soteras Gutiérrez, F.-Y. Lin, K. Vanommeslaeghe, J. A. Lemkul, K. A. Armacost, C. L. Brooks, and A. D. MacKerell, Parametrization of halogen bonds in the charmm general force field: Improved treatment of ligand–protein interactions, Bioorganic & Medicinal Chemistry **24**, 4812 (2016).
[59] P. E. Blöchl, Projector augmented-wave method, Phys. Rev. B **50**, 17953 (1994).
[60] G. Kresse and J. Furthmüller, Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set, Phys. Rev. B **54**, 11169 (1996).
[61] J. P. Perdew, K. Burke, and M. Ernzerhof, Generalized gradient approximation made simple, Phys. Rev. Lett. **77**, 3865 (1996).
[62] P. M. Larsen, S. Schmidt, and J. Schiøtz, Robust structural identification via polyhedral template matching, Modelling and Simulation in Materials Science and Engineering **24**, 055007 (2016).
[63] M. Dion, H. Rydberg, E. Schröder, D. C. Langreth, and B. I. Lundqvist, Van der waals density functional for general geometries, Physical Review Letters **92**, 246401 (2004).
[64] P. Ying, W. Zhou, L. Svensson, E. Berger, E. Fransson, F. Eriksson, K. Xu, T. Liang, J. Xu, B. Song, S. Chen, P. Erhart, and Z. Fan, Highly efficient path-integral molecular dynamics simulations with GPUMD using neuroevolution potentials: Case studies on thermal properties of materials, The Journal of Chemical Physics **162**, 064109 (2025).
[65] E. Fransson, M. Slabanja, P. Erhart, and G. Wahnström, Dynasor—A Tool for Extracting Dynamical Structure Factors and Current Correlation Functions from Molecular Dynamics Simulations, Advanced Theory and Simulations **4**, 2000240 (2021).
[66] R. Fair, A. Jackson, D. Voneshen, D. Jochym, D. Le, K. Refson, and T. Perring, Euphonic: Inelastic neutron scattering simulations from force constants and visualization tools for phonon properties, Journal of Applied Crystallography **55**, 1689 (2022).
[67] R. Turanyi, A. Jackson, and J. Wilkins, Pace-neutrons/resins: Python library for resolution functions of inelastic neutron scattering instruments, (2025), (accessed 2025-04-10).
[68] O. Arnold, J. Bilheux, J. Borreguero, A. Buts, S. Campbell, L. Chapon, M. Doucet, N. Draper, R. Ferraz Leal, M. Gigg, V. Lynch, A. Markvardsen, D. Mikkelson, R. Mikkelson, R. Miller, K. Palmen, P. Parker, G. Passos, T. Perring, P. Peterson, S. Ren, M. Reuter, A. Savici, J. Taylor, R. Taylor, R. Tolchenov, W. Zhou, and J. Zikovsky, Mantid—data analysis and visualization package for neutron scattering and µSR experiments, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **764**, 156 (2014).
[69] Z. Fan, brucefan1983/GPUMD: GPUMD-v4.0, (2025).
[70] J. F. Ziegler and J. P. Biersack, The stopping and range of ions in matter, in *Treatise on Heavy-Ion Science: Volume 6: Astrophysics, Chemistry, and Condensed Matter*, edited by D. A. Bromley (Springer US, Boston, MA, 1985) pp. 93–129.
[71] J. Liu, J. Byggmästar, Z. Fan, P. Qian, and Y. Su, Large-scale machine-learning molecular dynamics simulation of primary radiation damage in tungsten, Phys. Rev. B **108**, 054312 (2023).
[72] J. Hou, Y.-W. You, X.-S. Kong, J. Song, and C. Liu, Accurate prediction of vacancy cluster structures and energetics in bcc transition metals, Acta Materialia **211**, 116860 (2021).
[73] J. Hou, X. Kong, X. Wu, J. Song, and C. S. Liu, Predictive model of hydrogen trapping and bubbling in nanovoids in bcc metals, Nature Materials **18**, 833 (2019).
[74] R. Hill, The Elastic Behaviour of a Crystalline Aggregate, Proceedings of the Physical Society. Section A **65**, 349 (1952).
[75] M. De Jong, W. Chen, T. Angsten, A. Jain, R. Notestine, A. Gamst, M. Sluiter, K. C. Ande, J. J. Plata, C. Toher, S. Curtarolo, G. Ceder, K. A. Persson, and M. Asta, Charting the complete elastic properties of inorganic crystalline compounds, Scientific Data **2**, 1 (2015).
[76] B. Deng, Materials project trajectory (mptrj) dataset, figshare (2023).
[77] *International Critical Tables*, Vol. VI (McGraw-Hill Book Co., Inc., New York, NY, 1926) p. 230.

Supplementary Information

NEP89: Universal neuroevolution potential for inorganic and organic materials across 89 elements

SUPPLEMENTARY NOTES

Supplementary Note S1: Energy, force, and stress calculations for the NEP89 model

In our universal neuroevolution potential model across 89 elements (NEP89) model, the total potential energy is a sum of the site energies, $U = \sum_i U_i$, where the site energy U_i for a given atom i is a function of an abstract descriptor vector $\mathbf{q}^i$ with a number of components q_ν^i ($\nu = 1, 2, \cdots, N_{\rm des}$). This function is expressed as

$$U_i = \sum_{\mu=1}^{N_{\rm neu}} w_\mu^{(1)} \tanh\left(\sum_{\nu=1}^{N_{\rm des}} w_{\mu\nu}^{(0)} q_\nu^i - b_\mu^{(0)}\right) - b^{(1)}.$$

Here, $\tanh(x)$ is the activation function, $w^{(0)}$ are the weight parameters connecting the input layer (with dimension $N_{\rm des}$) and the hidden layer (with dimension $N_{\rm neu}$), $w^{(1)}$ represents the weight parameters connecting the hidden layer and the output layer (the site energy), $b^{(0)}$ represents the bias parameters in the hidden layer, and $b^{(1)}$ is the bias parameter in the output layer. The descriptor vector consists of radial and angular components as detailed previously [19]. The force acting on an atom i can be derived to be $\mathbf{F}_i = \sum_{j\neq i} \left(\partial U_i/\partial \mathbf{r}_{ij} - \partial U_j/\partial \mathbf{r}_{ji}\right)$, where $\mathbf{r}_{ij} = \mathbf{r}_j - \mathbf{r}_i$. The virial tensor for the whole system can be derived to be $\mathbf{W} = \sum_i \sum_{j\neq i} \mathbf{r}_{ij} \otimes \partial U_j/\partial \mathbf{r}_{ji}$. The stress tensor is defined as $\mathbf{W}/V$, where V represents the volume of the structure. The trainable parameters $\mathbf{z}$ are optimized using the natural evolution strategy [30] to minimize the loss function $L = \Delta U + \Delta F + \Delta W$, where ΔU, ΔF, and ΔW denote the root-mean-square errors (RMSEs) of energy, force, and virial, respectively, between the predicted and reference values. To prevent overfitting and undesirable increases in model parameters, $\mathcal{L}_2$ regularization is applied. Additionally, to ensure physically meaningful behavior at short interatomic distances, the short-range Ziegler-Biersack-Littmark potential [70] is included as a background term [71]. The hyperparameters used for training the NEP89 model are detailed in Supplementary Note S2.

Supplementary Note S2: The nep.in input file for training NEP89

We have used the following inputs in the `nep.in` file of the GPUMD code to train the NEP89 model:

```
type 89 H He Li Be B C N O F Ne Na Mg Al Si P S Cl Ar K Ca
        Sc Ti V Cr Mn Fe Co Ni Cu Zn Ga Ge As Se Br Kr
        Rb Sr Y Zr Nb Mo Tc Ru Rh Pd Ag Cd In Sn Sb Te I Xe
        Cs Ba La Ce Pr Nd Pm Sm Eu Gd Tb Dy Ho Er Tm Yb Lu Hf Ta W
        Re Os Ir Pt Au Hg Tl Pb Bi Ac Th Pa U Np Pu
version     4
zbl         2
cutoff      6  5
n_max       4  4
basis_size  8  8
l_max       4  2  1
neuron      80
lambda_1    0
lambda_e    1
lambda_f    1
lambda_v    2
batch       90000
population  60
generation  1000000
```

Supplementary Note S3: Details on descriptor-space subsampling

To enable subsampling in the descriptor space, we first computed the descriptor components for each structure in the database targeted for subsampling. For each descriptor component, the value of a structure was defined as the mean of the component values across all atoms in that structure. The process began with the random selection of one structure from the database. Subsequently, we iterated through the remaining structures, adding a new structure to the selection only if its distance from all previously selected structures exceeded a predefined threshold, d_{min}, which was set to 0.05 in our case. The distance between two descriptors was quantified as the Euclidean distance calculated from their components. This methodology ensured that the selected structures maintained a minimum separation of d_{min} in the descriptor space.

Supplementary Note S4: Energy shifting between the different datasets

After training the NEP model using a dataset that excluded all reference energies except those from the subsampled OMAT24 dataset, we predicted energies for all other datasets. Then we applied the separable natural evolution strategy [30] to determine the energy shift values for relevant species in other datasets. These shifts were optimized to minimize the difference between the predicted and reference energy values. For a given dataset with $N_{\rm str}$ structures, the predicted and reference energies for a structure a are denoted as $U_a^{\rm pre}$ and $U_a^{\rm ref}$, respectively. The shifted reference energy is expressed as $U_a^{\rm ref} + N_a^I \Delta^I$, where N_a^I represents the number of species I in the structure a, and Δ^I is the energy shift for species I to be determined. The loss function to be minimized is defined as $L = \sum_a \left(U_a^{\rm ref} + N_a^I \Delta^I - U_a^{\rm pre}\right)^2 / N_{\rm str}$.

Supplementary Note S5: The formation energy calculations of MP-ternary and GNoME datasets for various foundation models

We employed two publicly available datasets, comprising all relevant ternary structures from the Materials Project database [32] and a wide range of structures predicted by GNoME [14], spanning 2-component to 5-component systems with force components below 80 eV Å^{-1}. These datasets were used to benchmark the accuracy of different foundational models in predicting formation energies.

Specifically, we evaluated four models: NEP89 (this work), MACE-MP-0 [16], CHGNet [13], and M3GNet [12]. For each model, formation energies were computed with reference to the most stable allotrope of each constituent species. All the formation energy calculations were carried out within an integrated environment combining the GPUMD-WIZARD [50] and CALORINE [52] packages. The predicted formation energies from each model were then compared against DFT reference values from Ref. [9], as illustrated in Fig. S3. The mean absolute errors (MAEs) for each model are summarized in Fig. 2 of the main text.

Supplementary Note S6: Binding energy calculations for S66 dimer set

To benchmark the performance of various foundation models, we evaluate their predictions of binding energies on the widely used S66 dimer set [33]. S66 comprises 66 reference equilibrium geometries of molecular complexes designed to represent the most common types of noncovalent interactions found in biomolecules, while maintaining a balanced representation of dispersion and electrostatic contributions.

The DFT reference binding energies (obtained from Ref. [16]) were computed using the Perdew-Burke-Ernzerhof (PBE) functional with D3 dispersion correction and Becke-Johnson damping [28, 29], with a plane-wave energy cutoff of 520 eV. The predicted binding energies of molecular dimers in S66, computed using different foundation models (NEP89, MACE-MP-0, CHGNet, and M3GNet), are shown in Fig. S4. The MAEs for each model are summarized in Fig. 2 of the main text.

Supplementary Note S7: Lattice energy calculations for DMC-ICE13 dataset

Ice is one of the most important and intriguing molecular crystals, exhibiting a rich and continuously evolving phase diagram. Here, we evaluate the performance of various foundation models by computing both the absolute and relative lattice energies from the DMC-ICE13 dataset [34], where the relative values are referenced to hexagonal ice Ih. The DMC-ICE13 dataset comprises lattice energies of 13 distinct ice polymorphs, capturing the full structural complexity observed in ambient and high-pressure molecular ice phases. Such benchmarking offers insight into the capacity of different foundation models to capture the subtle and diverse intermolecular interactions governing molecular ice polymorphism. The absolute and relative lattice energies computed by different foundation models are presented in Fig. S5. Here, the DFT reference values were obtained using the same computational protocol described in Supplementary Note S6, and were extracted from Ref. [16]. The MAEs of absolute lattice energies are summarized in Fig. 2 of the main text.

Supplementary Note S8: Calculation of the formation energy of iron vacancy clusters and the adhesion energy of hydrogen atoms in iron anopores

Understanding the structure and energetics of vacancy clusters is fundamental to describing defect evolution in metals. We compute the formation energies of vacancy clusters of arbitrary size in body-centered cubic iron using different foundation models. The computational methodology follows the procedure outlined in Ref. [72]. Fig. S6a compares the predicted formation energies from different foundation models (NEP89, MACE-MP-0, CHGNet, and M3GNet) with reference values obtained from DFT.

In addition, the interaction between hydrogen and nanoscale voids is a critical factor in hydrogen-induced damage in structural materials [73]. We evaluate the adsorption energies of varying numbers of hydrogen atoms in iron nanovoids using different foundation models, with comparisons to DFT reference values shown in Fig. S6c–d.

All DFT reference data were calculated using the PBE exchange-correlation functional. The MAEs for the formation and adsorption energies across different foundation models are summarized in Fig. 2 of the main text.

Supplementary Note S9: Phonon property calculations for various foundation models

In this section, we assess the phonon properties of a diverse set of materials using different foundation models, which require accurate force predictions to ensure physically meaningful lattice vibrational characteristics.

To facilitate direct comparison, we adopted a benchmark set of 97 materials introduced in Ref. [16], which were randomly selected from the PhononDB database [35] and previously used to evaluate the performance of MACE-MP-0 [16]. We employed the finite-displacement method implemented in Phonopy [35] to compute the phonon modes—including both the highest and lowest phonon frequencies—for all 97 compounds, and compared the results to DFT reference data (at the PBE level) reported in the PhononDB database [35]. The reliability of the DFT reference values in the PhononDB database is discussed in Ref. [16]. To minimize discrepancies between the foundation model predictions and the PhononDB results, we used the same supercell structures as those in the database and removed the non-analytical corrections arising from Born effective charges.

Fig. S7 presents parity plots comparing the highest and lowest phonon mode frequencies predicted by different foundation models against the DFT reference values. The MAEs for the highest phonon mode frequencies are reported in Fig. 2 of the main text.

Supplementary Note S10: Bulk and Shear Moduli calculations for various foundation models

We conducted a systematic benchmark of the elastic properties predicted by different foundation models across more than 11,000 materials stored in the Materials Project database [32]. A foundation model that accurately captures the potential energy surface should be able to well reproduce elastic properties such as the bulk and shear moduli, which depend on the second derivatives of energy with respect to strain [16].

Specifically, we computed the Voigt-Reuss-Hill (VRH) averaged bulk and shear moduli [74] from stress-strain relations using different foundation models, following the computational protocols described in Refs. [16, 75]. All calculations were carried out using the elasticity module of the MatCalc package [53], and thus all predictions are based solely on equilibrium, bulk crystalline structures.

To exclude potentially unphysical values arising from DFT errors, we filtered out entries in the Materials Project database [32] with VRH-averaged bulk or shear moduli ≥ 600 GPa or ≤ -50 GPa, resulting in a curated dataset of 11,626 structures.

Fig. S8 presents parity plots comparing the bulk and shear moduli predicted by different foundation models with the DFT (PBE-level) reference values, along with corresponding MAE and RMSE values. The MAEs for each model are summarized in Fig. 2 of the main text.

SUPPLEMENTARY TABLE

Supplementary Table S1: The dataset used for training NEP89

Table S1. The dataset used to train the NEP89 model.

Source of dataset	N^{spe}	N^{str}	Quantum-mechanical level	Software	Our curation
OMAT24 [24]	89	395469	PBE+U	VASP	Add D3(BJ)
MPtrj [13]	89	56987	PBE+U	VASP	Add D3(BJ), discard energy
SPICE [21]	15	24142	ωB97M-D3(BJ)/def2-TZVPPD	PSI4	Shift energy
ANI-1xnr [22]	4	18723	BLYP-D3(0)/TZV2P	CP2K	Shift energy
SSE-ABACUS [25]	27	12300	PBE	ABACUS	Add D3(BJ), discard energy
SSE-VASP [25]	15	9024	PBEsol	VASP	Add D3(BJ), discard energy
Protein [20]	5	6842	revPBE-D3(BJ)/def2-TZVP	PSI4	Shift energy
UNEP-v1 [9]	16	5597	PBE	VASP	Add D3(BJ)
CH [26]	2	4582	PBE	VASP	Add D3(BJ)
CHONPS (This work)	6	3049	PBE/DZVP-MOLOPT-SR-GTH	CP2K	Add D3(BJ), discard energy
Water [27]	2	926	MB-pol	MBX	Shift energy
Total	89	537641	–	–	–

SUPPLEMENTARY FIGURES

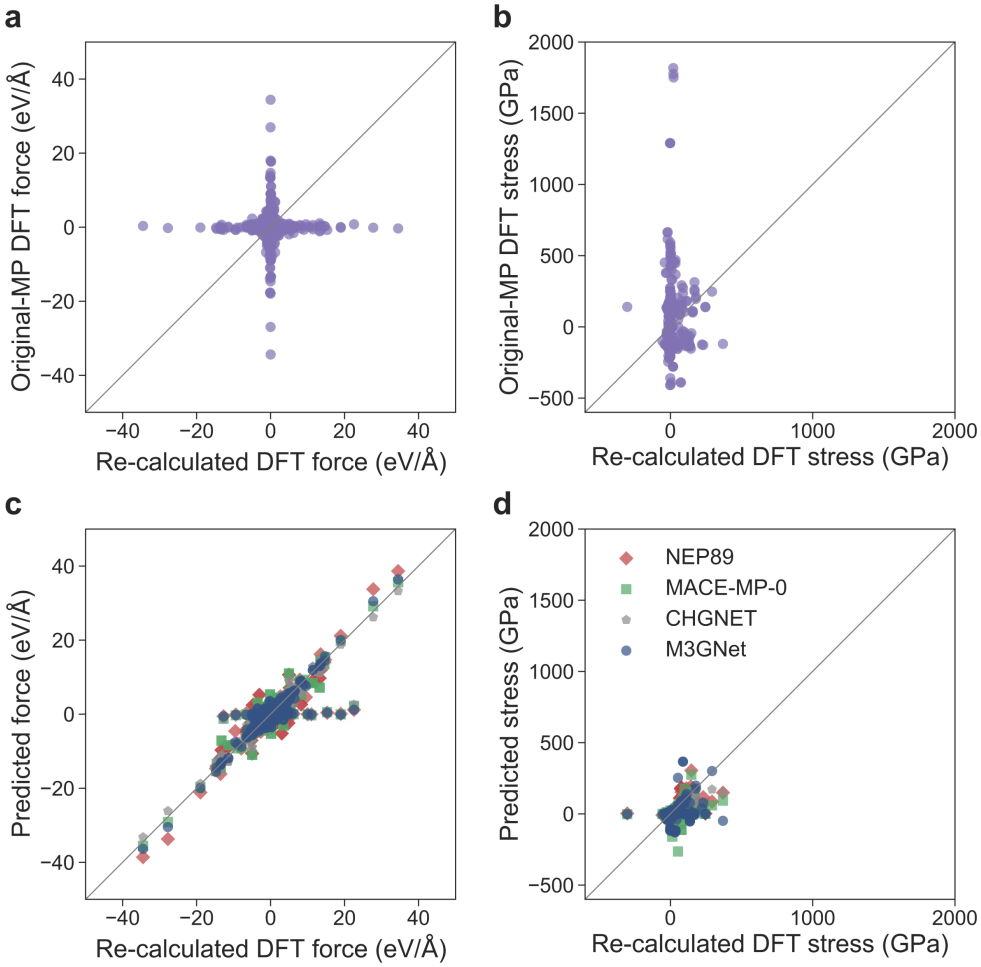

Figure S1. Comparison of the recalculated DFT reference values of typical structures in MPtrj database [76]. **(a,b)** Parity plots comparing the recalculated force and stress of different configurations with original MPtrj references [76]. **(c,d)** Parity plots of force and stress between recomputed DFT results and predictions from various foundation models. We used the MACE-MP-0 model [16] as a predictor, and structures with predicted stresses exceeding the DFT reference values of 100 GPa were selected. Accordingly, we selected 113 configurations for stress outliers. We performed DFT self-consistent calculations on these structures using the original computational inputs provided by the Material Project [32]. It can be observed from the figure that the original MPtrj reference values are unreliable, whereas predictions from different foundation models closely match the recalculated DFT references. It is worth noting that the 113 structures were excluded from the NEP89 training set, yet the model yielded reliable predictions, underscoring the effectiveness of our data screening process.

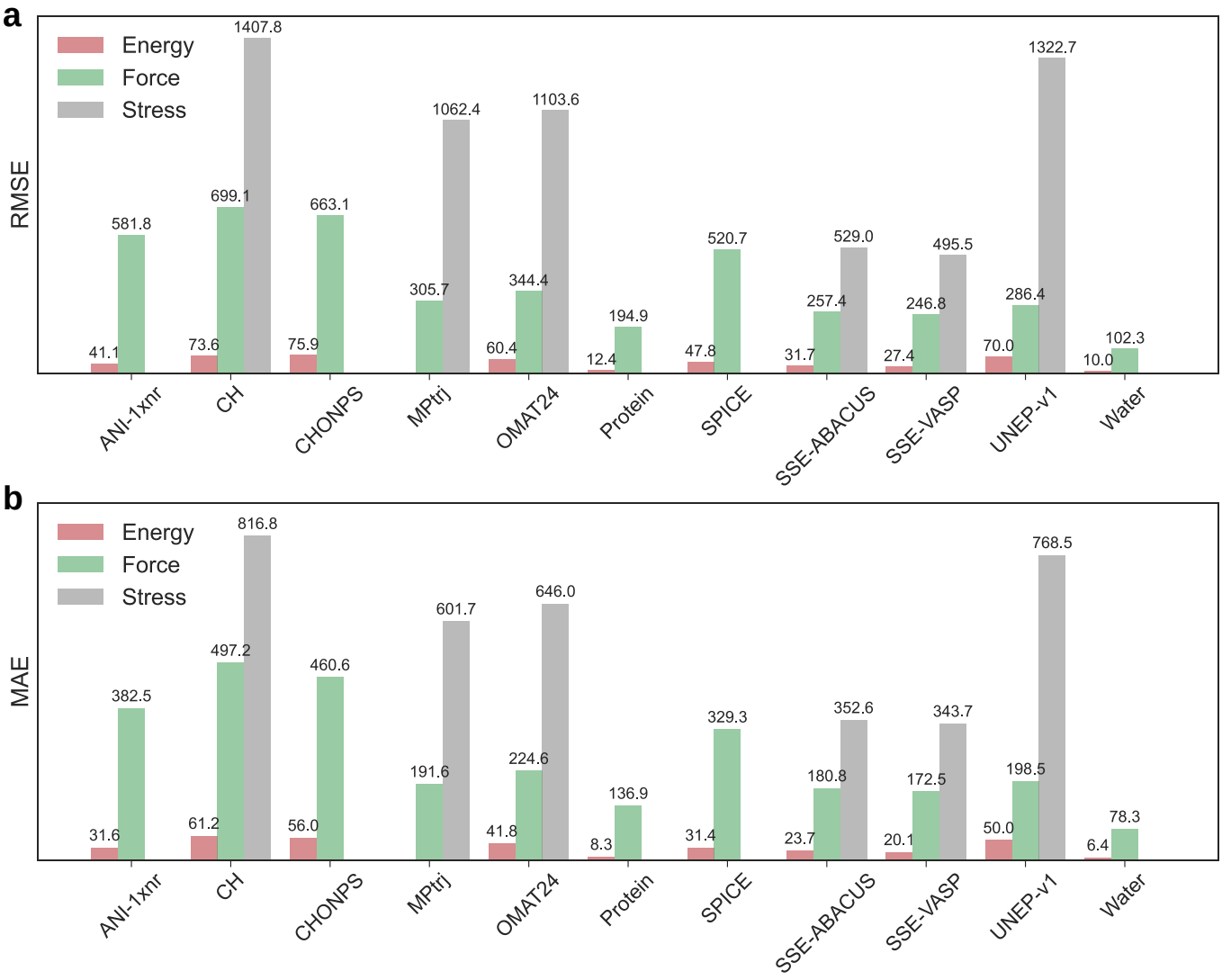

Figure S2. **(a)** RMSE and **(b)** MAE values of NEP89 model for different training subsets (see Table S1), with units of meV/atom for energy, meV/Å for force, and MPa for stress. The RMSE and MAE of the reference energies for the MPtrj dataset [76] are not presented, as they are incompatible with the reference energies in the OMAT24 dataset [24] due to differing treatments using the DFT+U technique. For some other subsets, RMSE and MAE for stress are not available due to the absence of DFT/high-level-quantum-chemistry reference values.

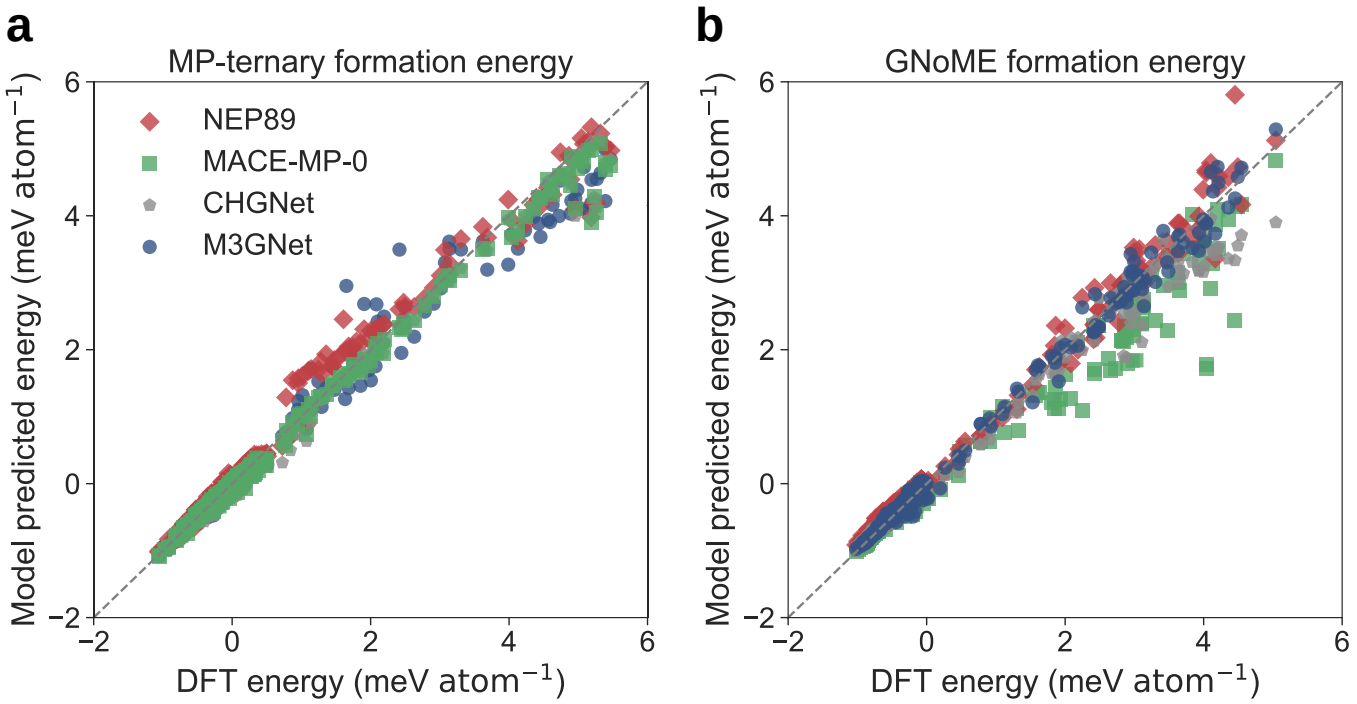

Figure S3. **(a,b)** Parity plots of formation energies (γ) comparing DFT reference data with predictions from NEP89 (this work), MACE-MP-0 [16], CHGNet [13], and M3GNet [12], for the structures sampled from the Materials Project (MP-ternary) [32] and the GNoME dataset [14].

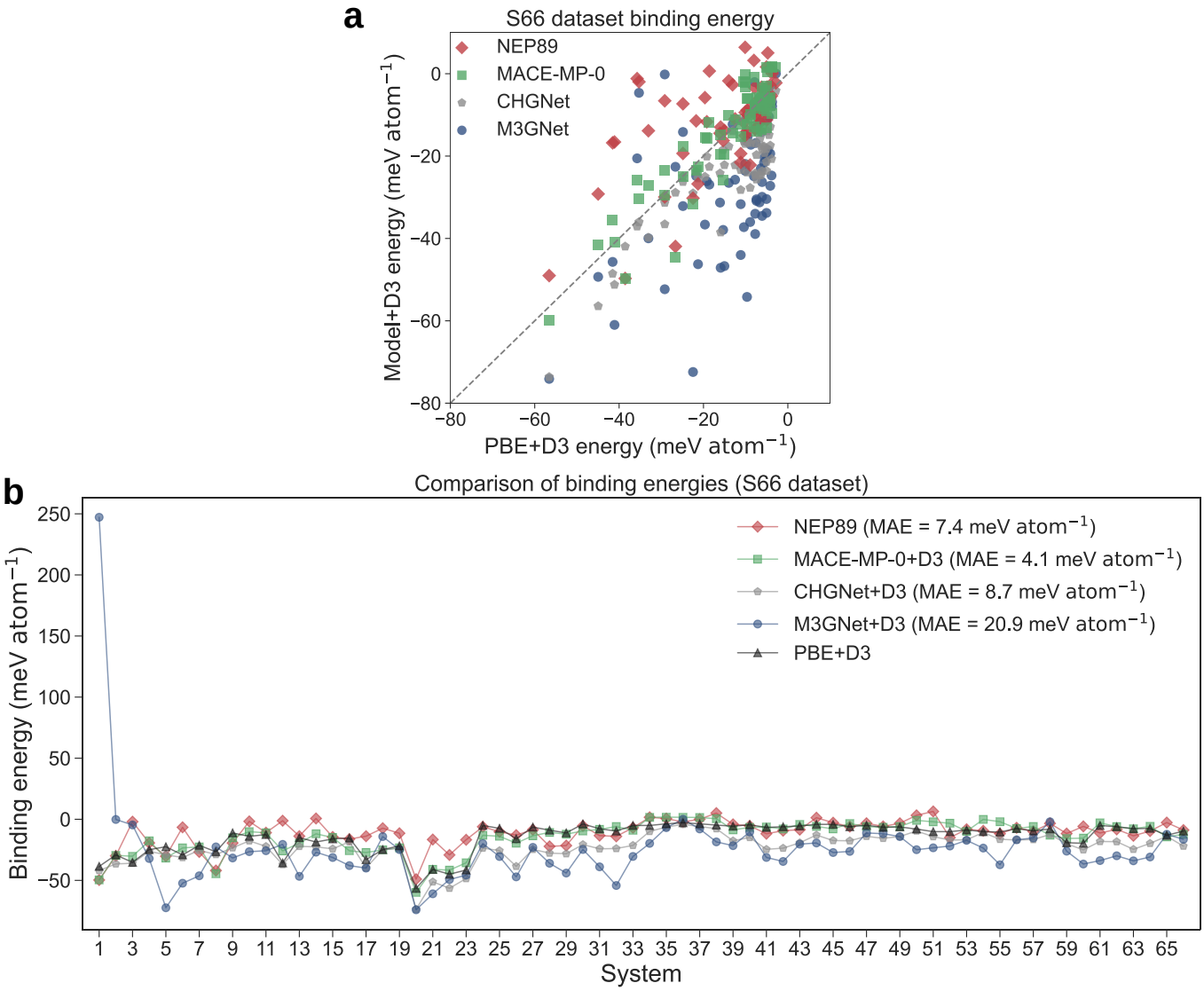

Figure S4. A detailed comparison of S66 dimers binding energies calculated by different foundation models with DFT (PBE+D3) reference values [16]. **(a)** Parity plot for the binding energies (E_b) of S66 dimer set [33], comparing predictions from various foundation models. The NEP89 results presented here do not include an explicit D3 correction. **(b)** A system-by-system comparison of binding energies predicted by various foundation models in the S66 dimer set.

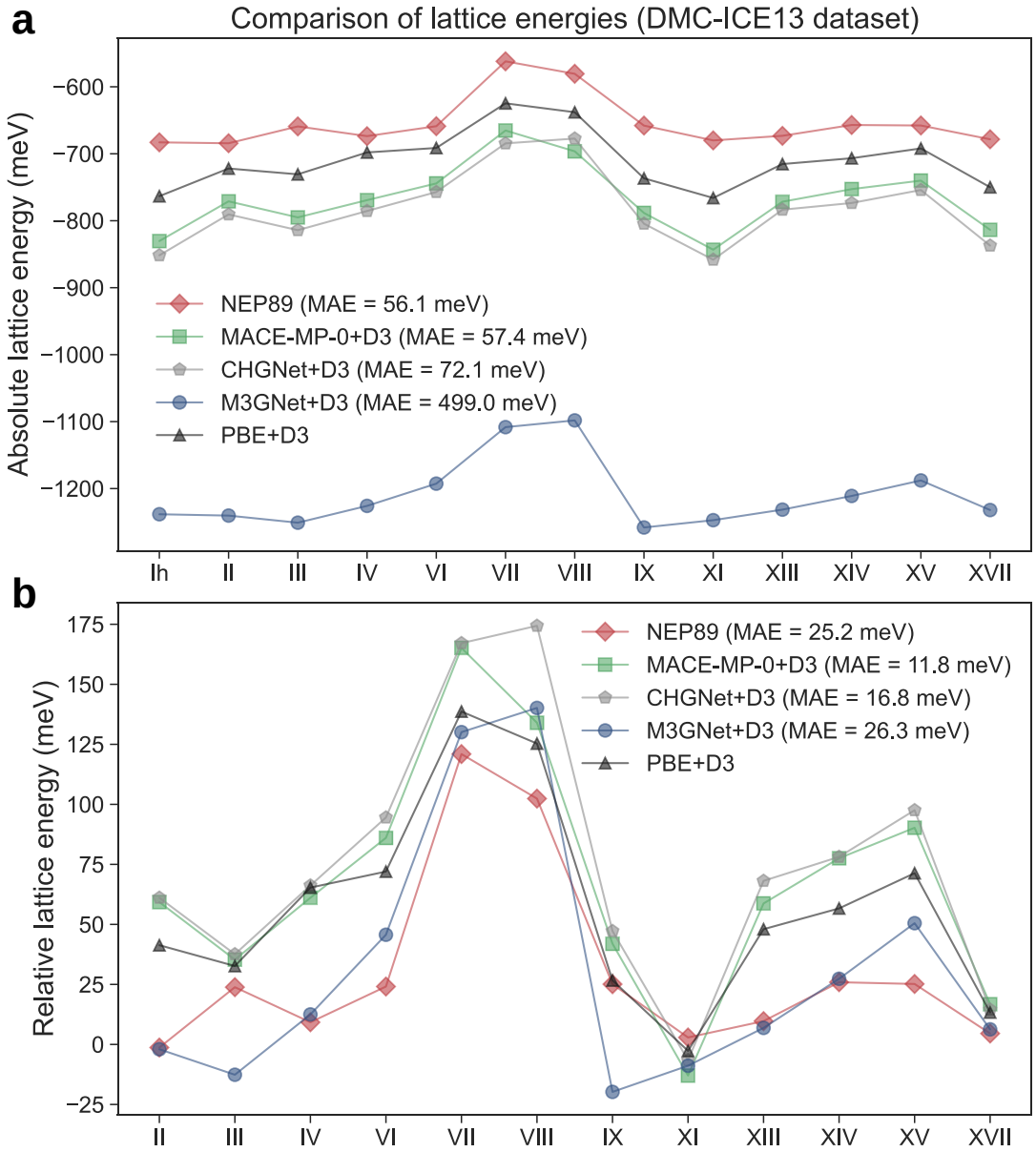

Figure S5. Comparison of lattice energies (E_{latt}) of the DMC-ICE13 dataset [34] calculated by different foundation models with DFT reference values (PBE+D3) [16]. The NEP89 results presented here do not include an explicit D3 correction. **(a)** Absolute lattice energy, defined as the energy per molecule of each crystalline phase relative to the gas phase. **(b)** Relative lattice energy, defined as the lattice energy relative to that of ice Ih.

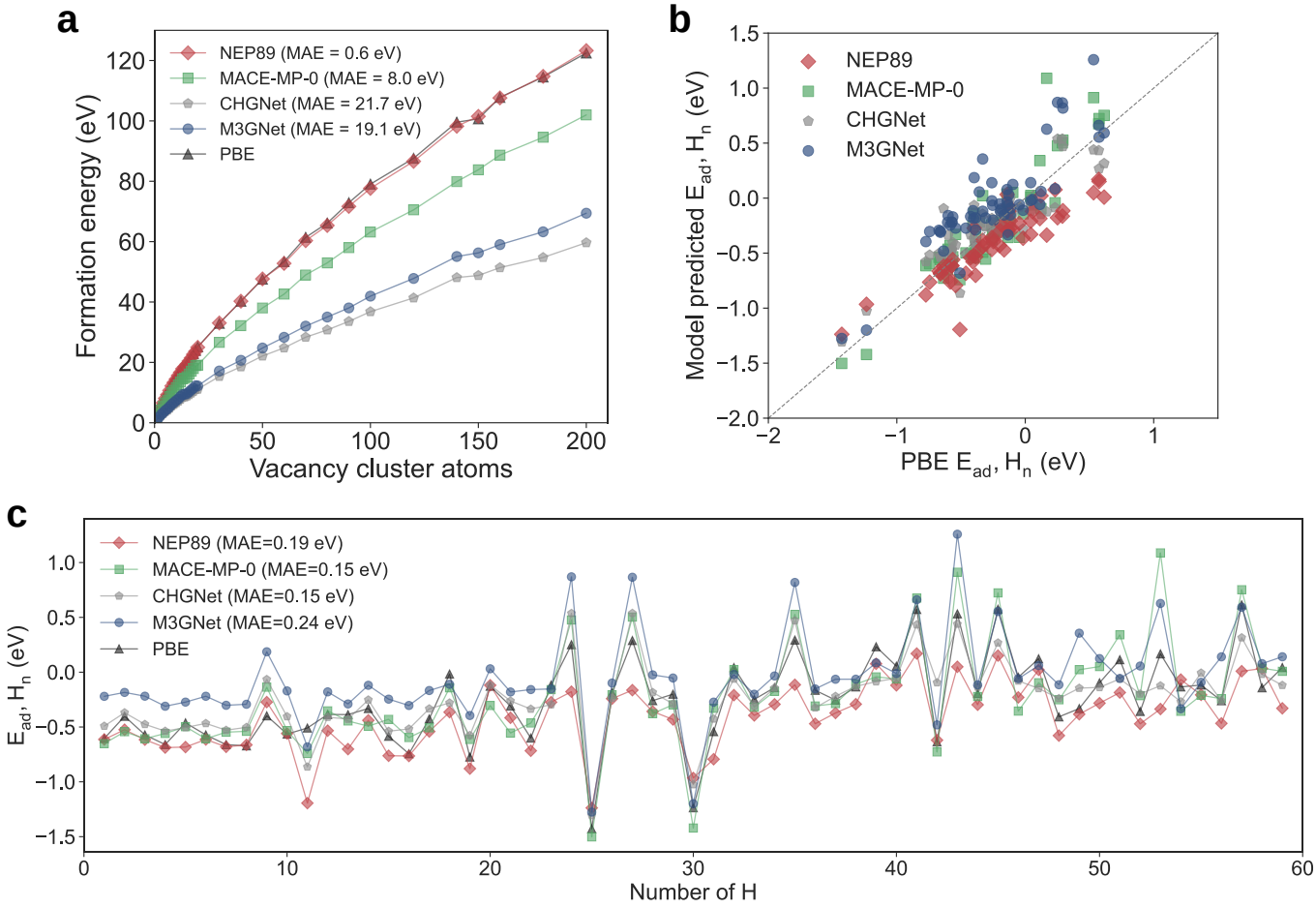

Figure S6. **(a)** Formation energies of iron vacancy clusters of different sizes predicted by various foundation models compared to DFT calculations. **(b)** Parity plot of the adhesion energy of hydrogen atoms in iron nanopores predicted by different foundation models and DFT. **(c)** Variation of adhesion energy with respect to the number of hydrogen atoms in iron nanopores predicted by different foundation models.

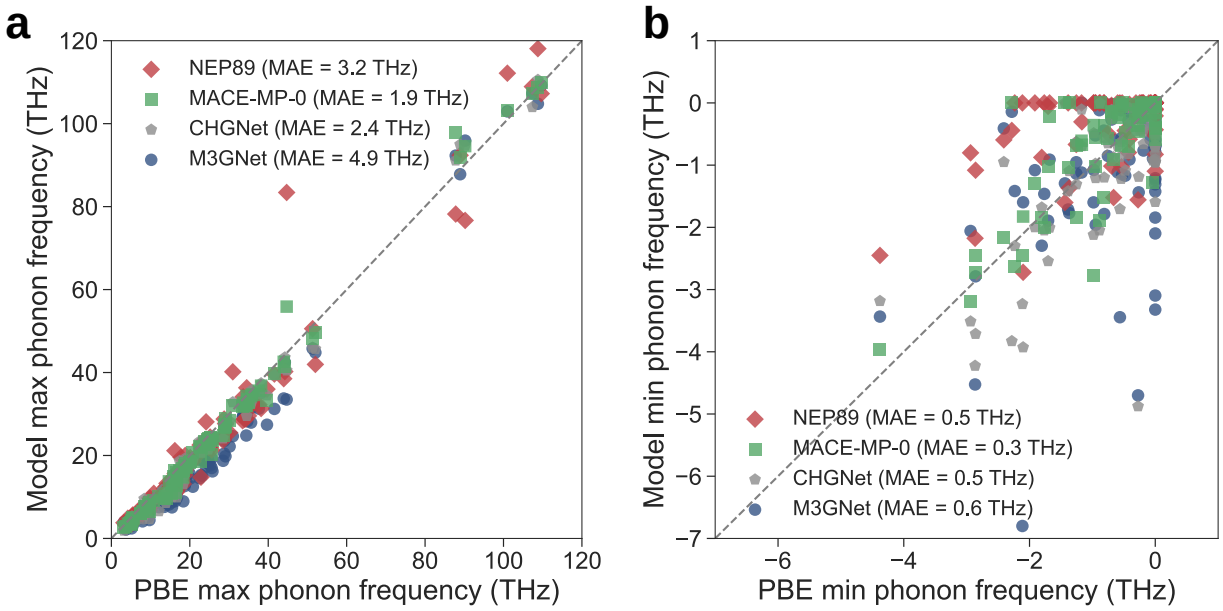

Figure S7. Comparison of the **(a)** maximum and **(b)** minimum phonon band frequencies for 97 materials calculated by different foundation models and by DFT (PBE).

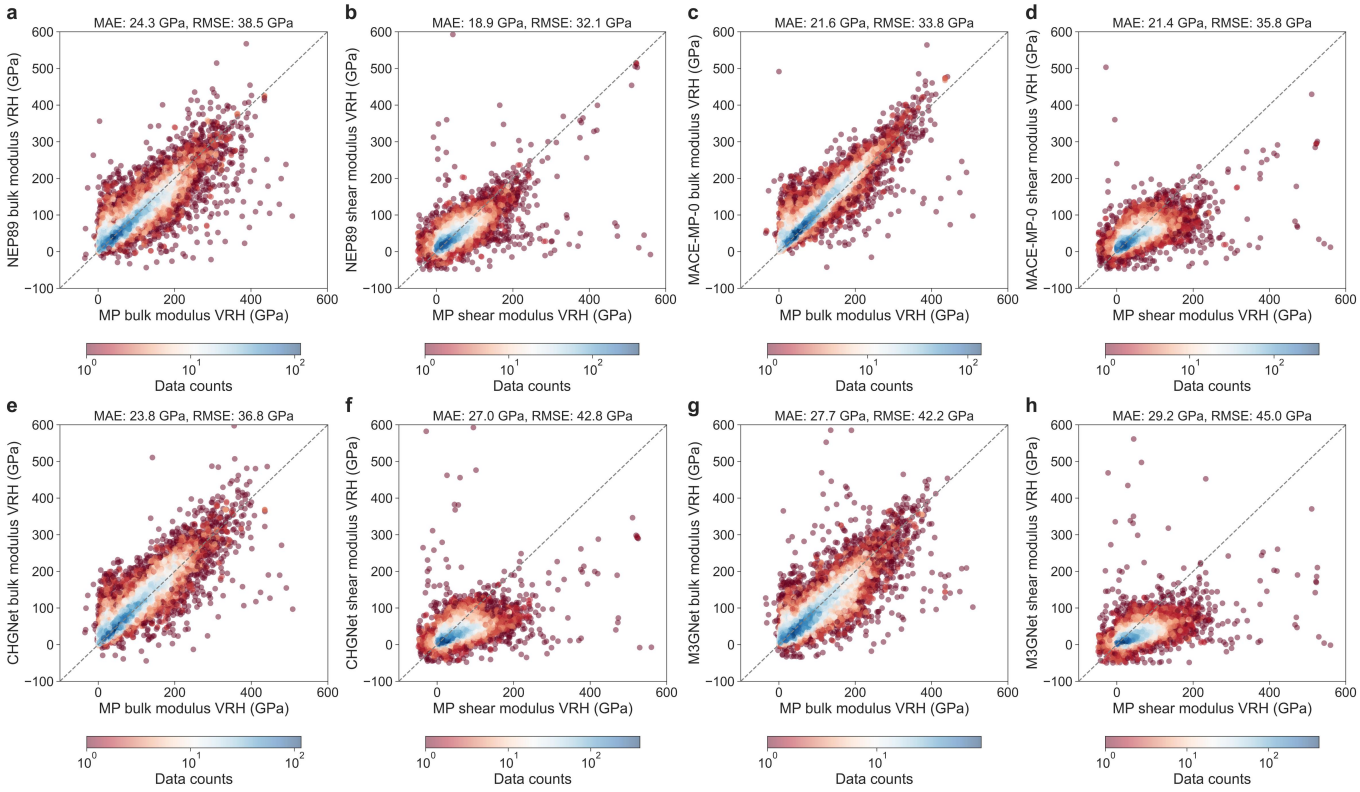

Figure S8. Parity plots of the bulk modulus and shear modulus calculated by different foundation models. The MAE and RMSE relative to the DFT reference values are labeled above each subplot. Different subplots show the predictions of bulk and shear moduli for over 11,000 materials by various foundation models. Color intensity indicates the distribution and density of the modulus data. Predictions considered unphysical ($\geq$ 600 GPa or ≤ -50 GPa) were excluded from the plots for each base model, and MAE and RMSE values were computed accordingly. VRH in the plots refers to the Voigt-Reuss-Hill [74] averaged bulk and shear moduli.

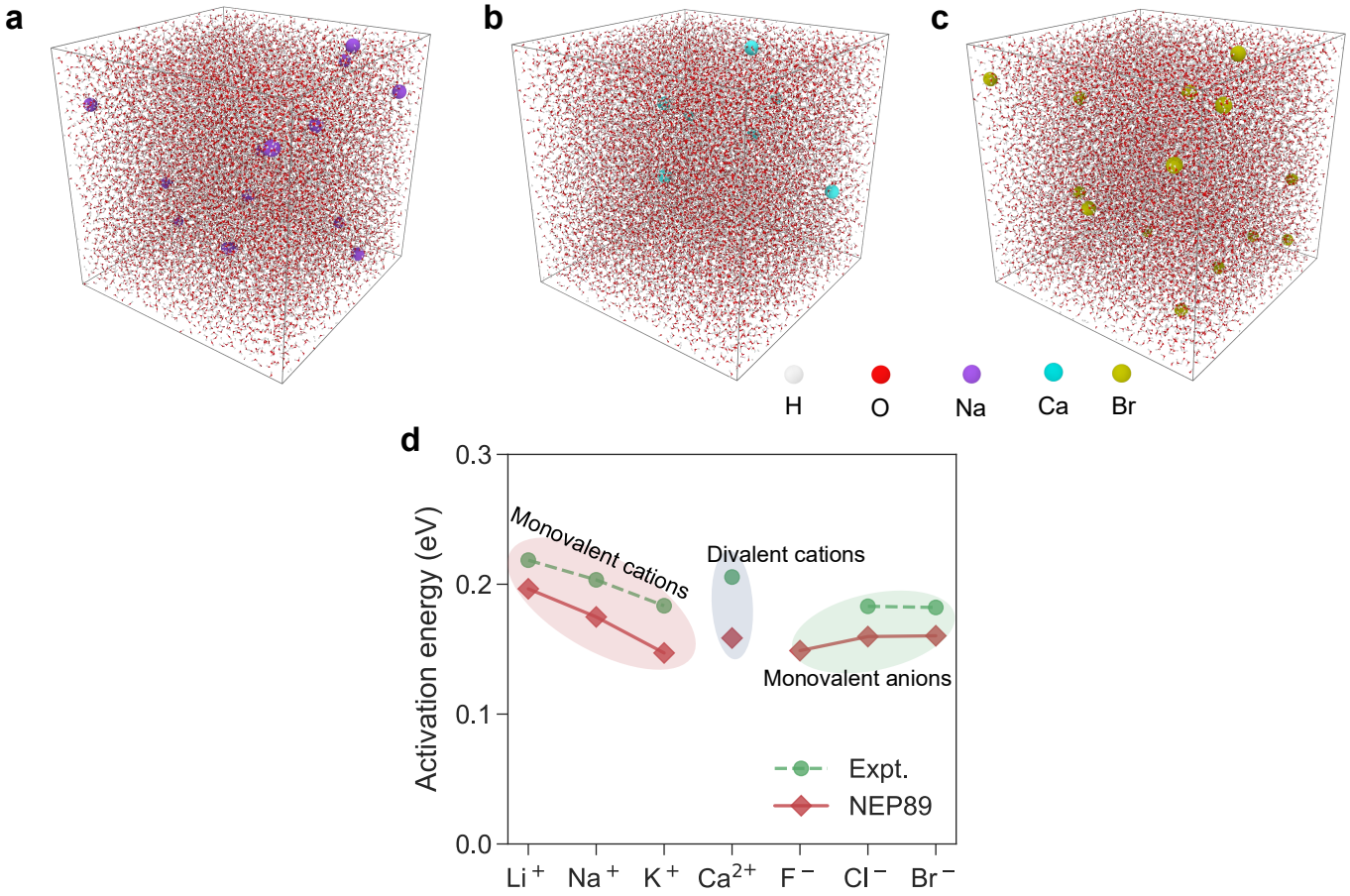

Figure S9. (**a**–**c**) Atomic snapshots of the random distribution of one of the monovalent cations (**a**), the divalent cations (**b**), and one of the monovalent anions (**c**) in water. (**d**) Predicted activation energies for ion diffusion in water at 300 K using NEP89, as well as experimental values at 291 K from Refs. [38, 77]. Note that the experimental data contains no record for F^-.

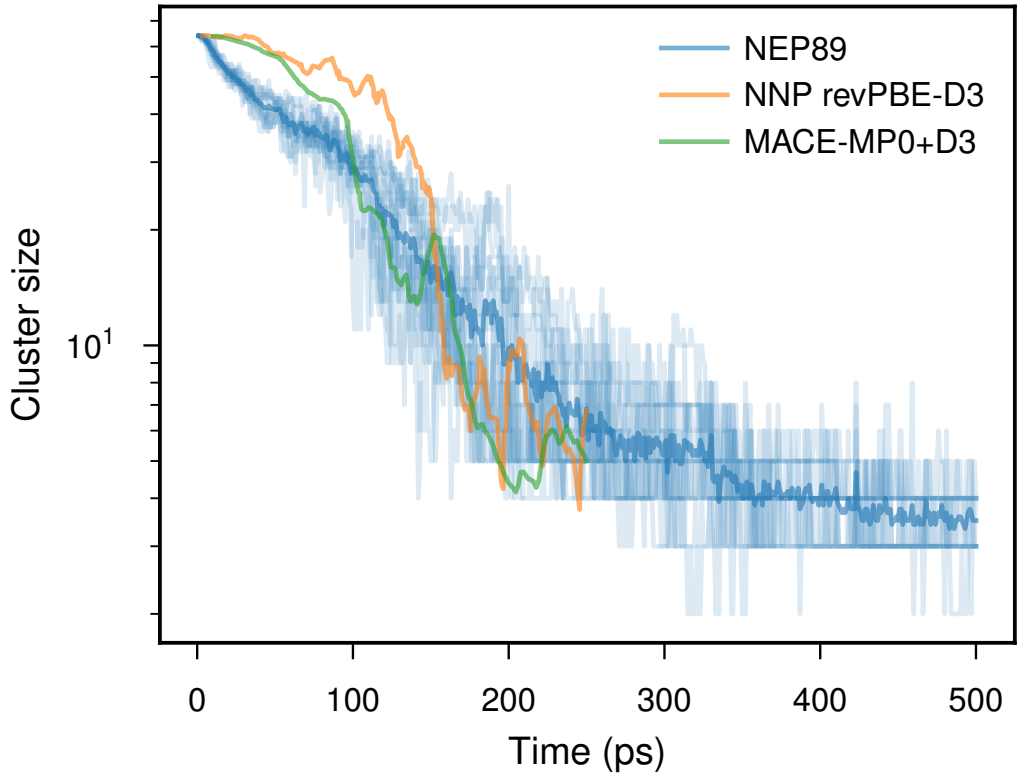

Figure S10. Dissolution kinetics of NaCl crystallite at 400 K predicted using NEP89, as well as other potentials [16]. The average cluster size (solid blue line) and individual runs (faint blue lines) illustrate that thermodynamic averages and kinetics with statistical significance are easily accessible using NEP89, due to the computational efficiency.

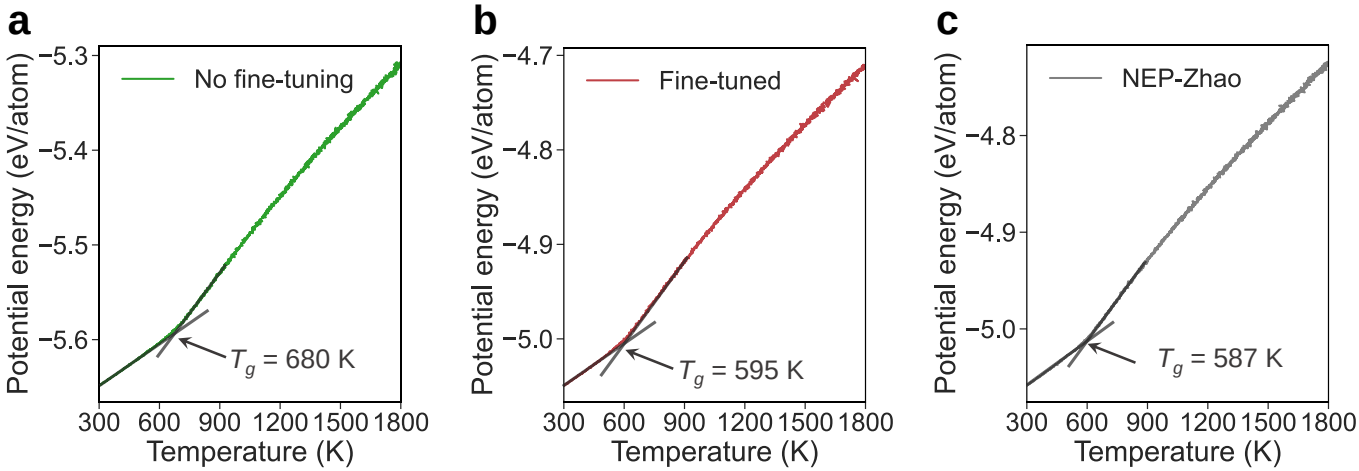

Figure S11. Potential energy curves and extrapolated glass transition temperature (T_g) of $Pd_{42.5}Cu_{30}Ni_{7.5}P_{20}$ metallic glass predicted by **(a)** no fine-tuning, **(b)** after NEP89 fine-tuning, and **(c)** NEP-Zhao [41].