Supplementary Information:

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] *Chalmers University of Technology, Department of Physics, 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*

Contents

Supplementary Notes **3**

S1. Energy, force, and stress calculations for the NEP89 model 3

S2. The nep.in input file for training NEP89 4

S3. Details on descriptor-space subsampling 5

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

S4. Energy shifting between the different datasets 6
S5. The formation energy calculations of MP-ternary and GNoME datasets for various foundation models 7
S6. Binding energy calculations for S66 dimer set 8
S7. Lattice energy calculations for DMC-ICE13 dataset 9
S8. Calculation of the formation energy of iron vacancy clusters and the adhesion energy of hydrogen atoms in iron anopores 10
S9. Phonon property calculations for various foundation models 11
S10. Bulk and Shear Moduli calculations for various foundation models . . . 12
Supplementary Table **13**
Table S1: The dataset used for training NEP89 13
Supplementary Figures **14**
Fig. S1 Comparison of the recalculated DFT reference values of typical structures in MPtrj database 14
Fig. S2 RMSE and MAE values of NEP89 model for different training subsets 15
Fig. S3 Parity plots of formation energies for MP-ternary and GNoME datasets 16
Fig. S4 A detailed comparison of S66 dimers binding energies calculated by different foundation models 17
Fig. S5 Comparison of lattice energies (E_{latt}) of the DMC-ICE13 dataset . . 18
Fig. S6 Formation energies of iron vacancy clusters and adhesion energy of hydrogen atoms in iron nanopores 19
Fig. S7 Comparison of the maximum and minimum phonon band frequencies for 97 materials 20
Fig. S8 Parity plots of the bulk modulus and shear modulus calculated by different foundation models 21
Fig. S9 Snapshots of ion diffusion in water and activation energies 22
Fig. S10 Dissolution of NaCl at 400 K predicted by various potentials 23
Fig. S11 Potential energy curves and extrapolated glass transition temperature of PdCuNiP metallic glass predicted by various potentials 24
Supplementary References **25**

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^i_ν ($\nu = 1, 2, \cdots, N_{\rm des}$). This function is expressed as

$$U_i = \sum_{\mu=1}^{N_{\rm neu}} w^{(1)}_\mu \tanh\left(\sum_{\nu=1}^{N_{\rm des}} w^{(0)}_{\mu\nu} q^i_\nu - b^{(0)}_\mu\right) - b^{(1)}.$$

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 [1]. 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 [2] 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 [3] is included as a background term [4]. 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 [2] 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 [5] and a wide range of structures predicted by GNoME [6], 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 [7], CHGNet [8], and M3GNet [9]. 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 [10] and CALORINE [11] packages. The predicted formation energies from each model were then compared against DFT reference values from Ref. [12], 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 [13]. 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. [7]) were computed using the Perdew-Burke-Ernzerhof (PBE) functional with D3 dispersion correction and Becke-Johnson damping [14, 15], 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 [16], 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. [7]. 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. [17]. 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 [18]. 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. [7], which were randomly selected from the PhononDB database [19] and previously used to evaluate the performance of MACE-MP-0 [7]. We employed the finite-displacement method implemented in Phonopy [19] 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 [19]. The reliability of the DFT reference values in the PhononDB database is discussed in Ref. [7]. 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 [5]. 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 [7].

Specifically, we computed the Voigt-Reuss-Hill (VRH) averaged bulk and shear moduli [20] from stress-strain relations using different foundation models, following the computational protocols described in Refs. [7, 21]. All calculations were carried out using the elasticity module of the MatCalc package [22], 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 [5] 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 [23]	89	395469	PBE+U	VASP	Add D3(BJ)
MPtrj [8]	89	56987	PBE+U	VASP	Add D3(BJ), discard energy
SPICE [24]	15	24142	ωB97M-D3(BJ)/def2-TZVPPD	PSI4	Shift energy
ANI-1xnr [25]	4	18723	BLYP-D3(0)/TZV2P	CP2K	Shift energy
SSE-ABACUS [26]	27	12300	PBE	ABACUS	Add D3(BJ), Shift energy
SSE-VASP [26]	15	9024	PBEsol	VASP	Add D3(BJ), Shift energy
Protein [27]	5	6842	revPBE-D3(BJ)/def2-TZVP	PSI4	Shift energy
UNEP-v1 [12]	16	5597	PBE	VASP	Add D3(BJ)
CH [28]	2	4582	PBE	VASP	Add D3(BJ)
CHONPS (This work)	6	3049	PBE/DZVP-MOLOPT-SR-GTH	CP2K	Add D3(BJ), Shift energy
Water [29]	2	926	MB-pol	MBX	Shift energy
Total	89	537641	–	–	–

Supplementary Figures

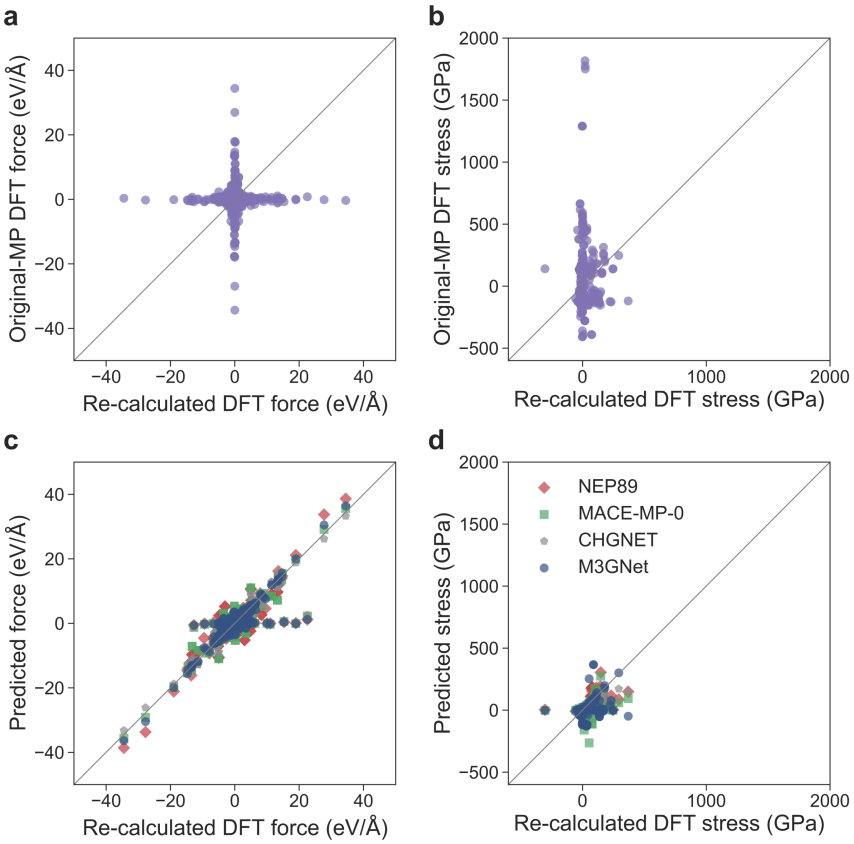

Figure S1: Comparison of the recalculated DFT reference values of typical structures in MPtrj database [30]. **(a,b)** Parity plots comparing the recalculated force and stress of different configurations with original MPtrj references [30]. **(c,d)** Parity plots of force and stress between recomputed DFT results and predictions from various foundation models. We used the MACE-MP-0 model [7] 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 [5]. 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.

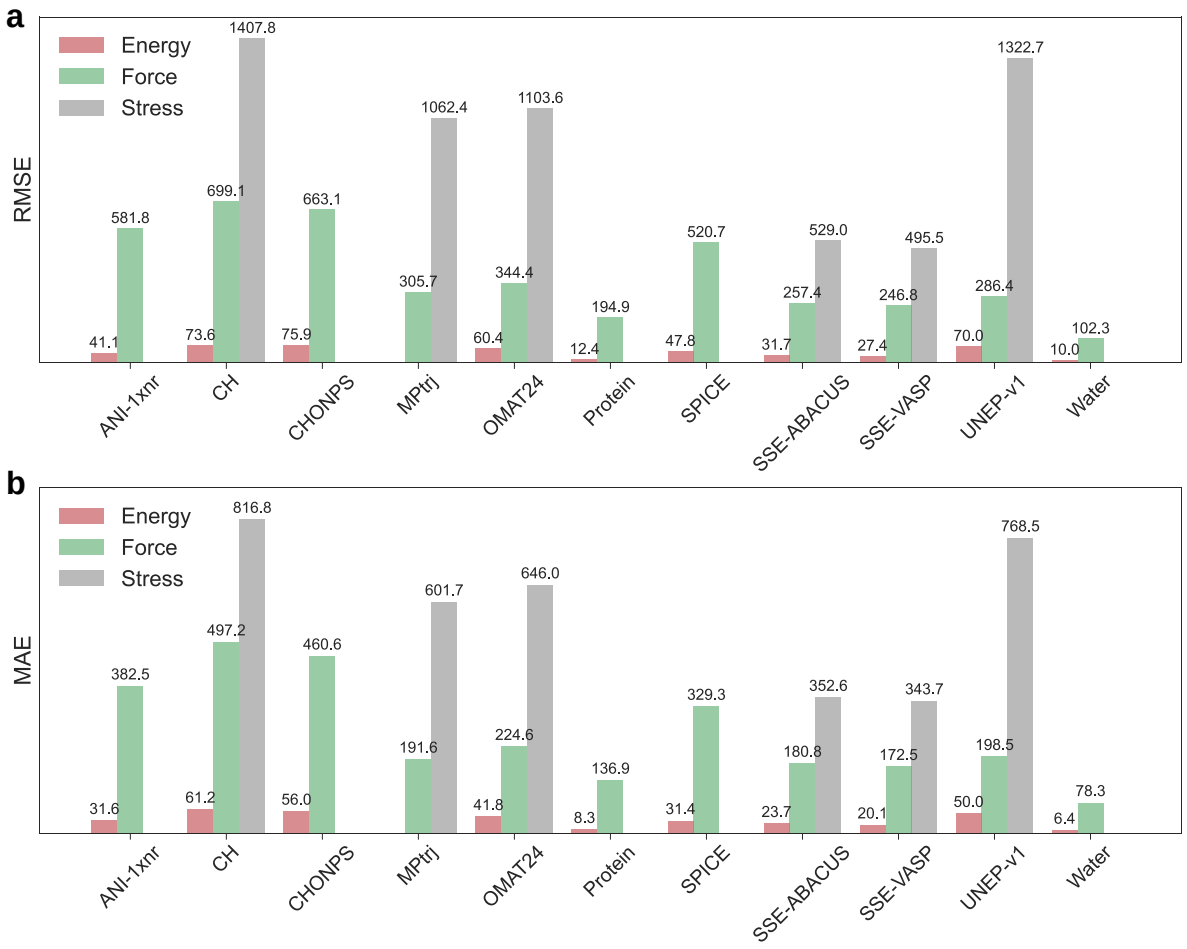

Figure S2: **(a)** RMSE and **(b)** MAE values of NEP89 model for different training subsets (see Table S1), with units of meV/atom for energy, meV/Å for force, and MPa for stress. The RMSE and MAE of the reference energies for the MPtrj dataset [30] are not presented, as they are incompatible with the reference energies in the OMAT24 dataset [23] 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.

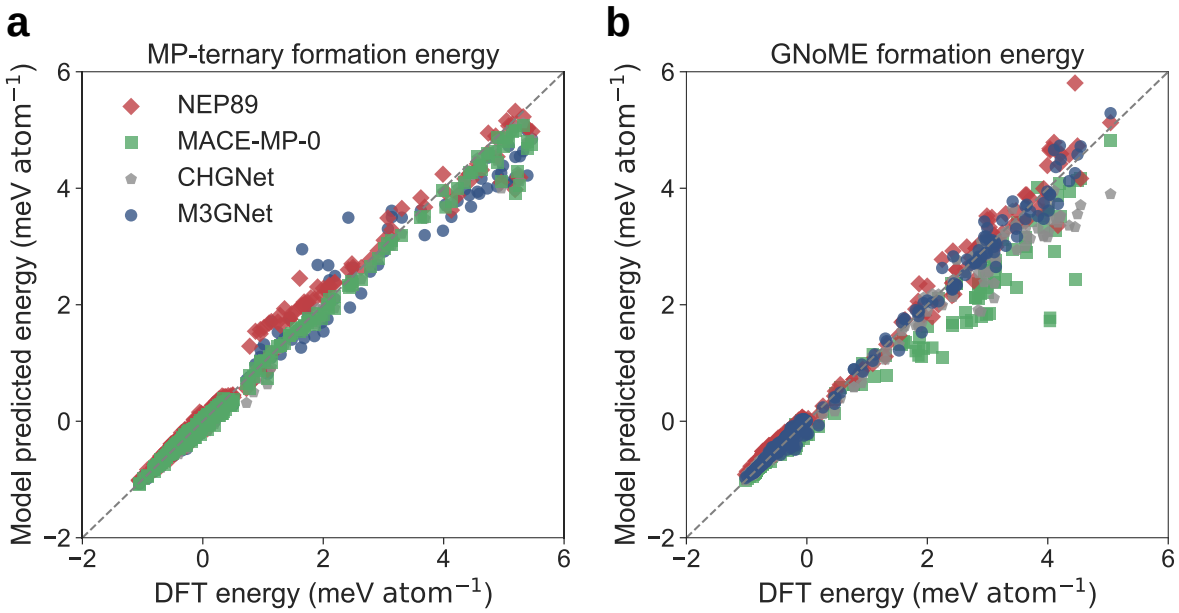

Figure S3: **(a,b)** Parity plots of formation energies (γ) comparing DFT reference data with predictions from NEP89 (this work), MACE-MP-0 [7], CHGNet [8], and M3GNet [9], for the structures sampled from the Materials Project (MP-ternary) [5] and the GNoME dataset [6].

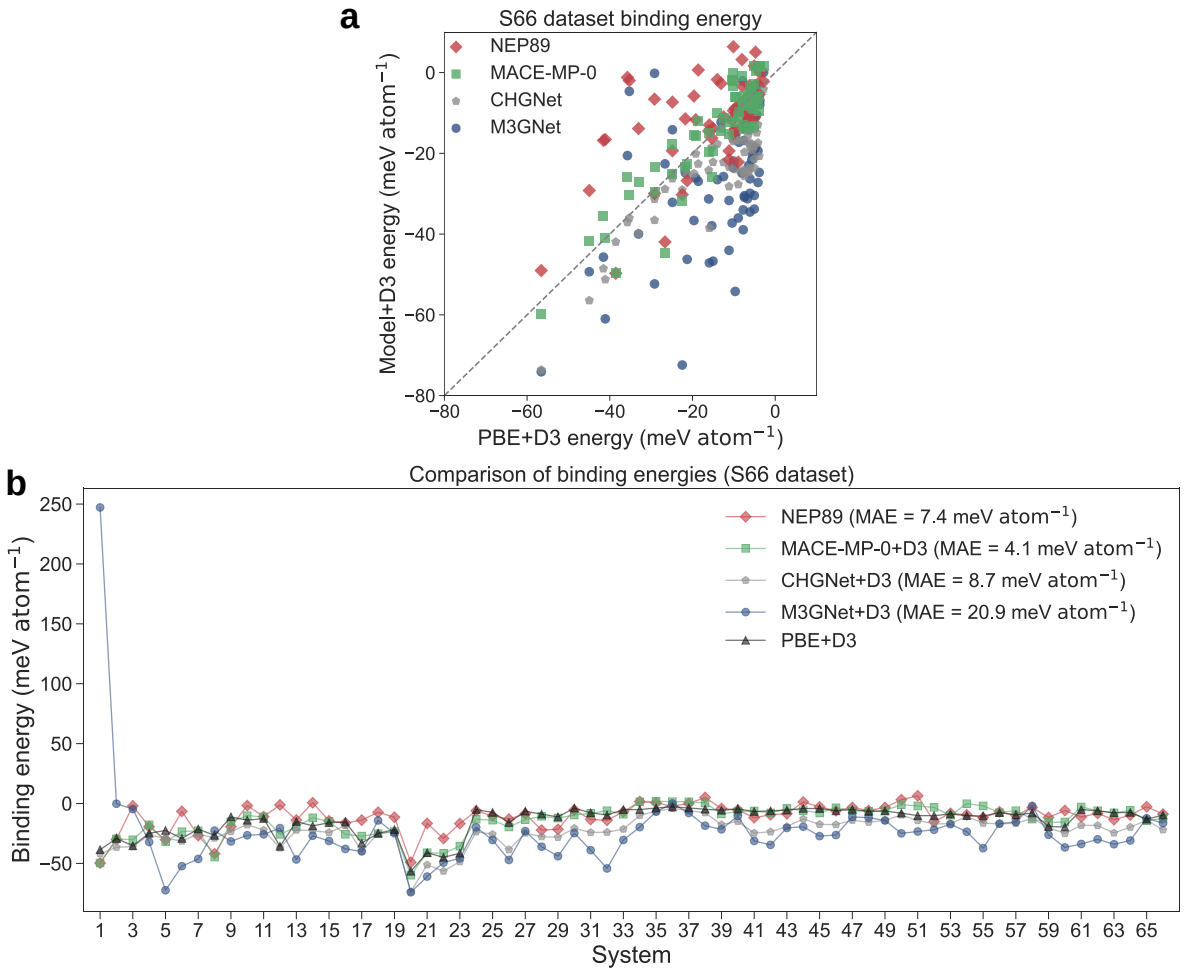

Figure S4: A detailed comparison of S66 dimers binding energies calculated by different foundation models with DFT (PBE+D3) reference values [7]. **(a)** Parity plot for the binding energies (E_b) of S66 dimer set [13], comparing predictions from various foundation models. The NEP89 results presented 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.

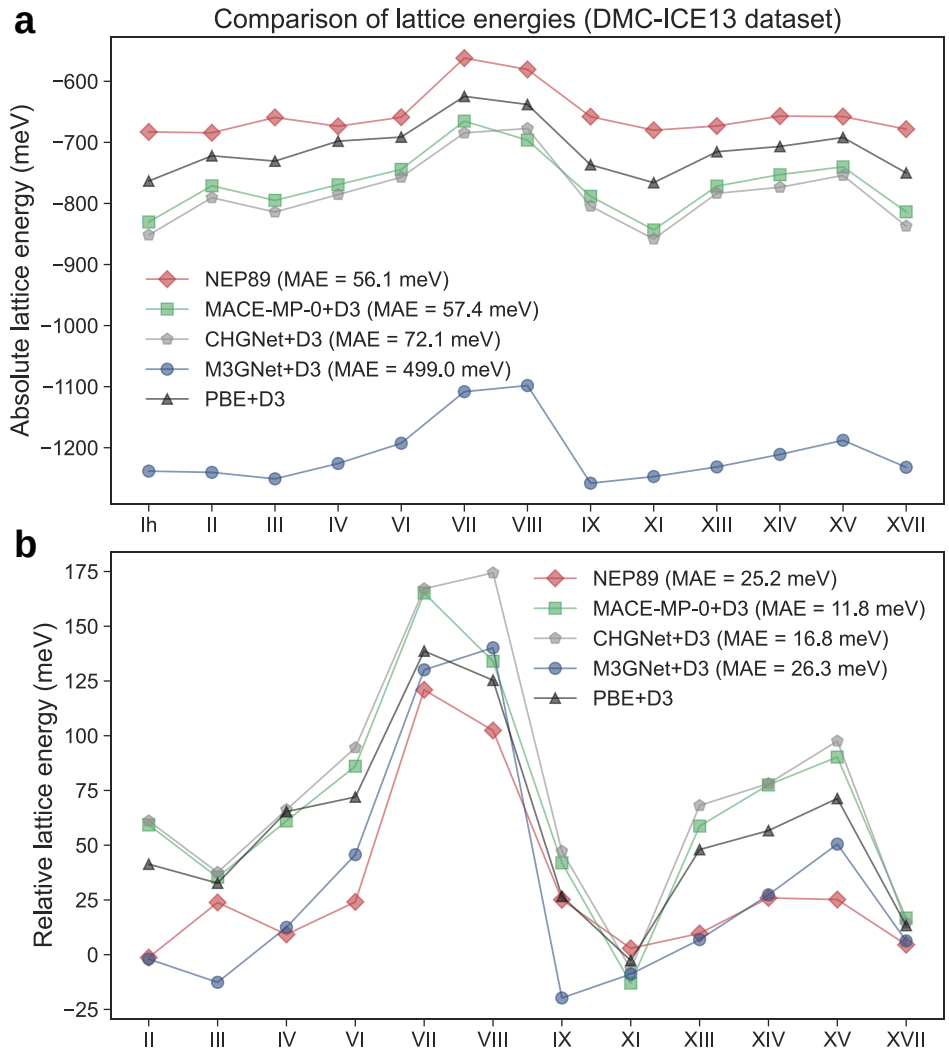

Figure S5: Comparison of lattice energies (E_{latt}) of the DMC-ICE13 dataset [16] calculated by different foundation models with DFT reference values (PBE+D3) [7]. The NEP89 results presented here do not include an explicit D3 correction. **(a)** Absolute lattice energy, defined as the energy per molecule of each crystalline phase relative to the gas phase. **(b)** Relative lattice energy, defined as the lattice energy relative to that of ice Ih.

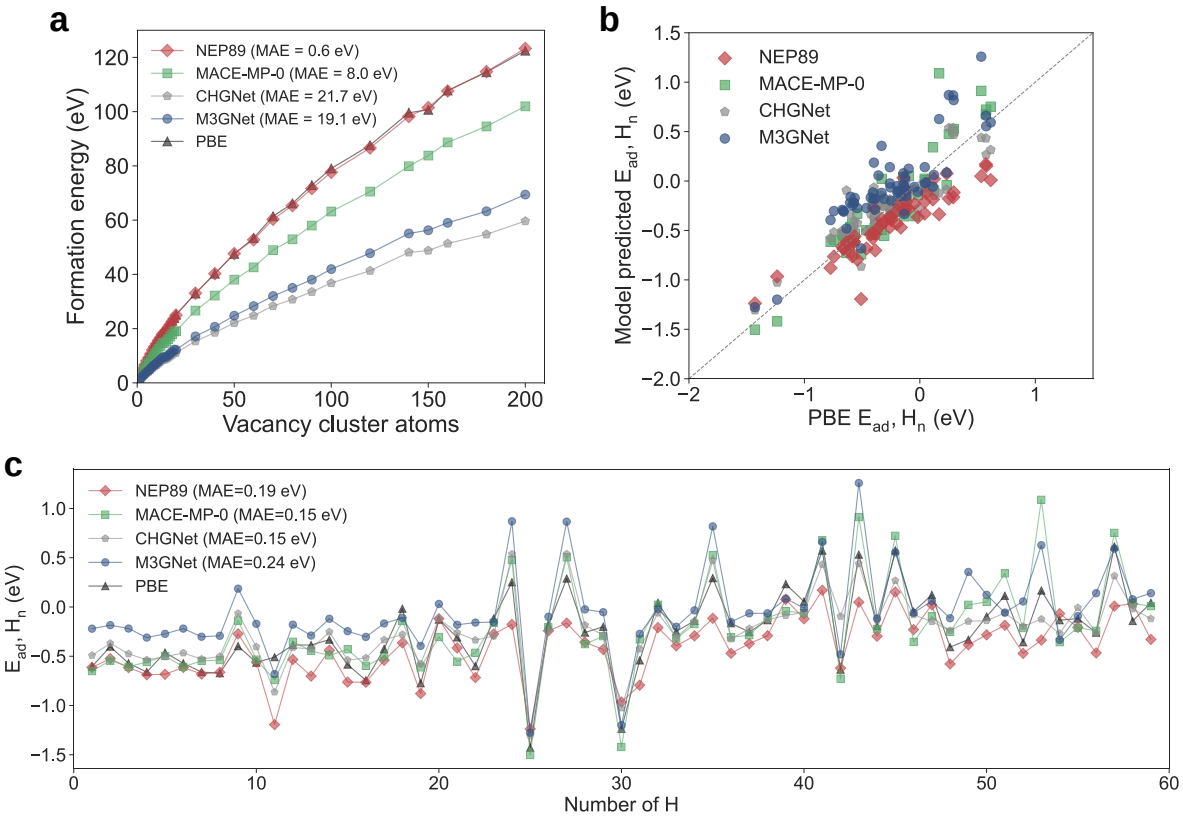

Figure S6: **(a)** Formation energies of iron vacancy clusters of different sizes predicted by various foundation models compared to DFT calculations. **(b)** Parity plot of the adhesion energy of hydrogen atoms in iron nanopores predicted by different foundation models and DFT. **(c)** Variation of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

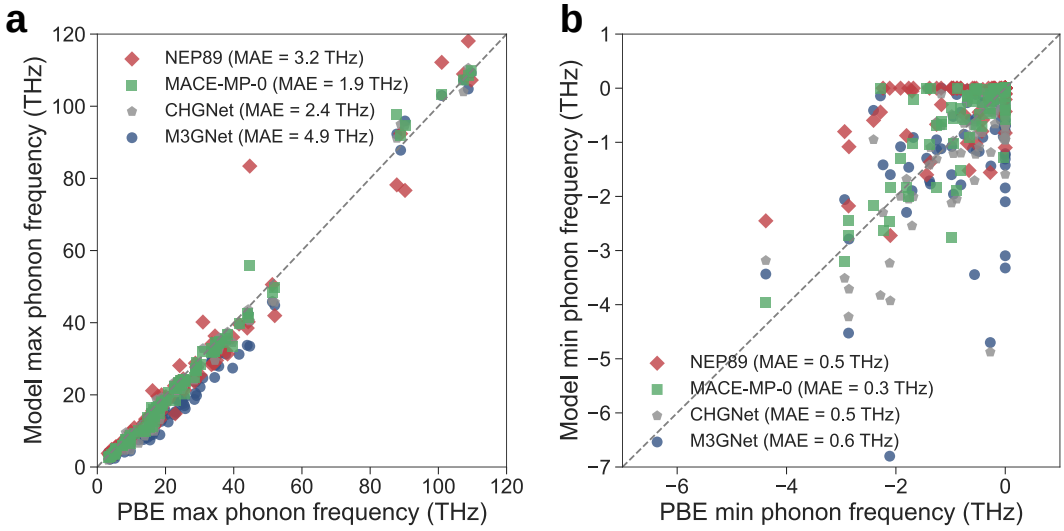

Figure S7: Comparison of the **(a)** maximum and **(b)** minimum phonon band frequencies for 97 materials calculated by different foundation models and by DFT (PBE).

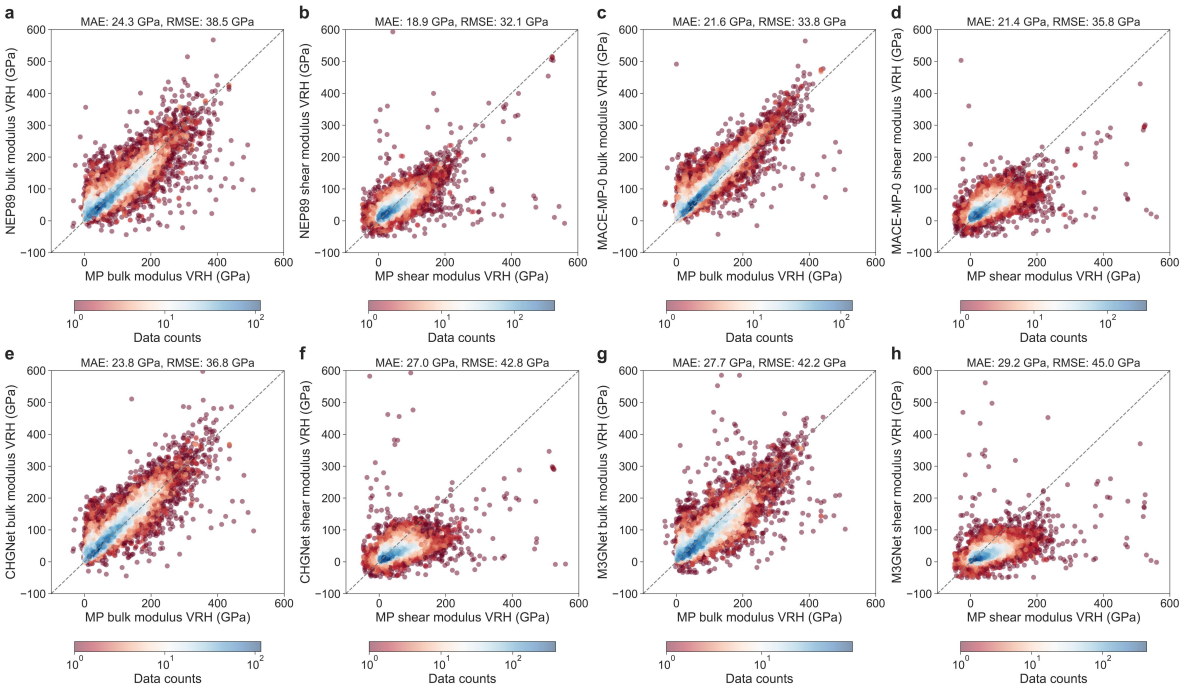

Figure S8: Parity plots of the bulk modulus and shear modulus calculated by different foundation models. The MAE and RMSE relative to the DFT reference values are labeled above each subplot. Different subplots show the predictions of bulk and shear moduli for over 11,000 materials by various foundation 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 [20] averaged bulk and shear moduli.

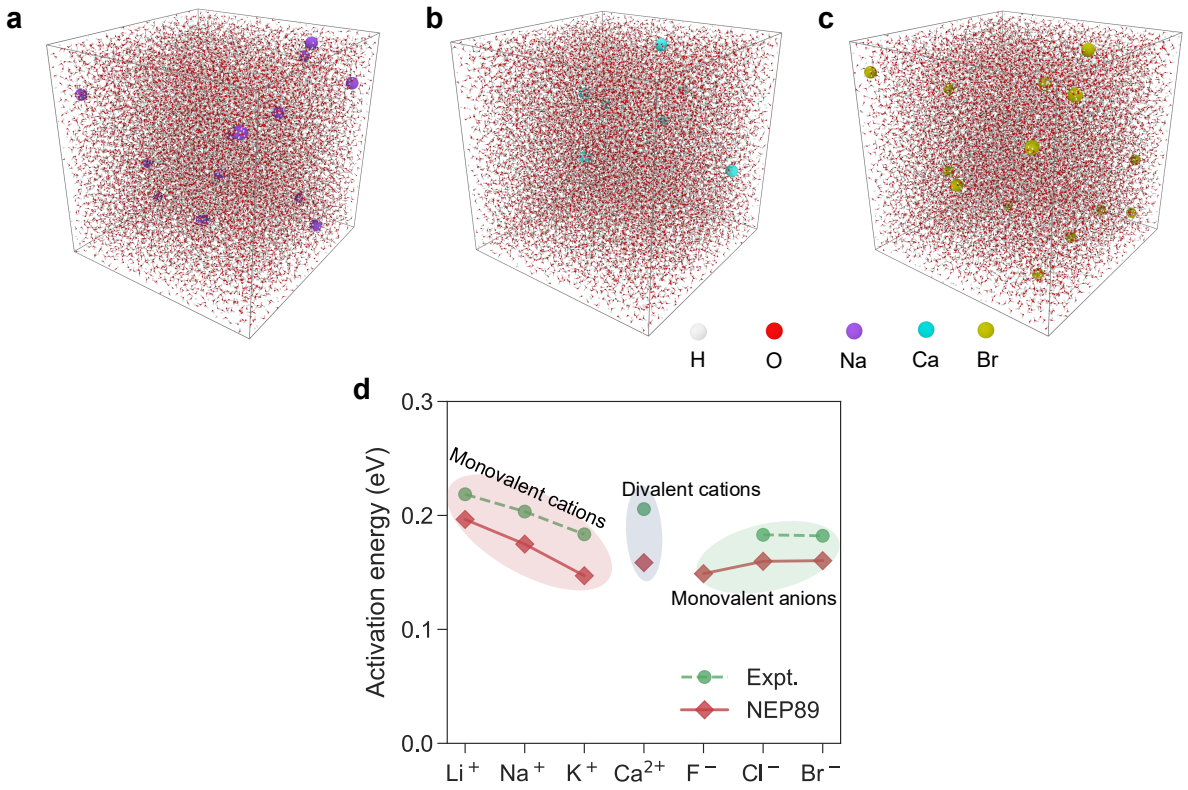

Figure S9: (**a**–**c**) Atomic snapshots of the random distribution of one of the monovalent cations (**a**), the divalent cations (**b**), and one of the monovalent anions (**c**) in water. (**d**) Predicted activation energies for ion diffusion in water at 300 K using NEP89, as well as experimental values at 291 K from Refs. [31, 32]. Note that the experimental data contains no record for F^-.

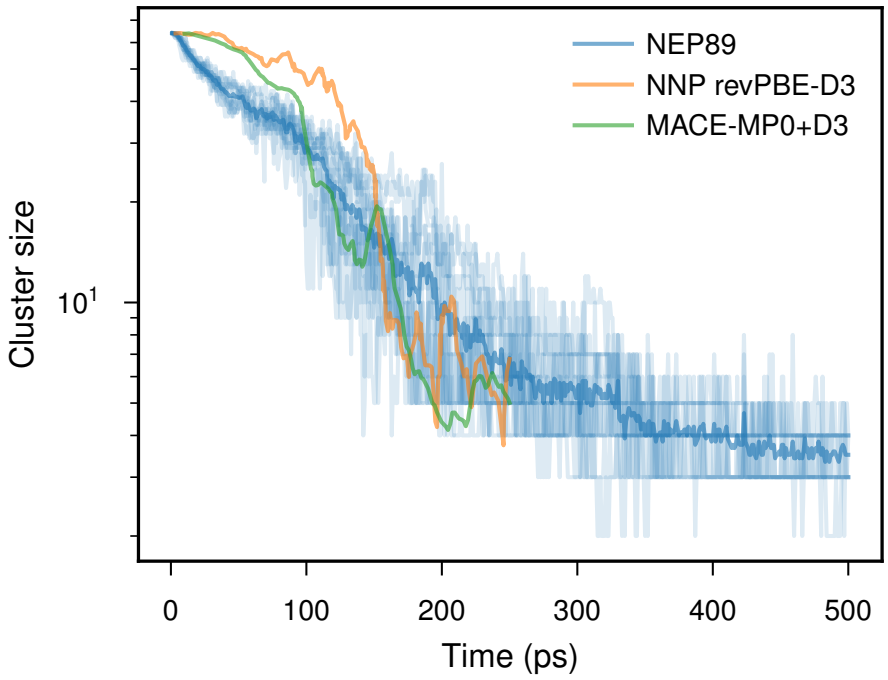

Figure S10: Dissolution kinetics of NaCl crystallite at 400 K predicted using NEP89, as well as other potentials [7]. The average cluster size (solid blue line) and individual runs (faint blue lines) illustrate that thermodynamic averages and kinetics with statistical significance are easily accessible using NEP89, due to the computational efficiency.

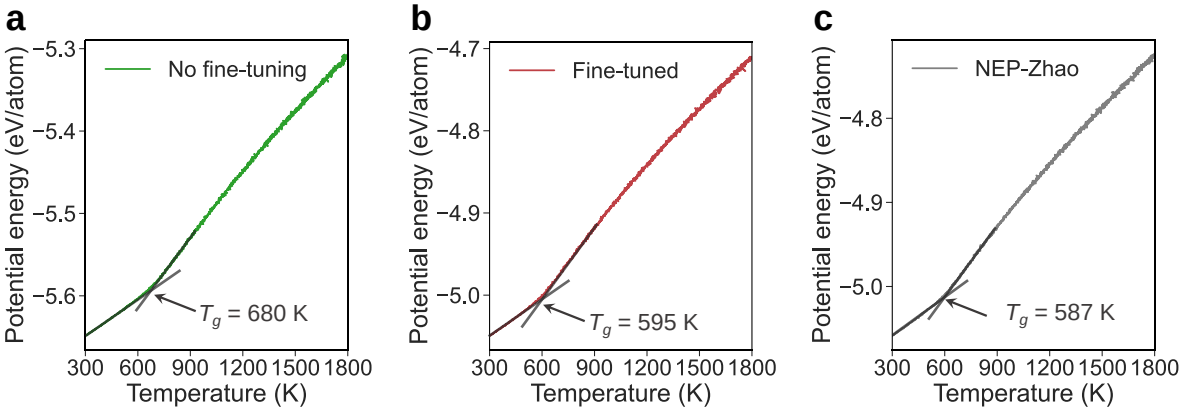

Figure S11: Potential energy curves and extrapolated glass transition temperature (T_g) of $Pd_{42.5}Cu_{30}Ni_{7.5}P_{20}$ metallic glass predicted by **(a)** no fine-tuning, **(b)** after NEP89 fine-tuning, and **(c)** NEP-Zhao [33].

Supplementary References

[1] 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).

[2] 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.

[3] 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.

[4] 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).

[5] 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).

[6] 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).

[7] 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] .

[8] 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).

[9] C. Chen and S. P. Ong, A universal graph deep learning interatomic potential for the periodic table, Nature Computational Science **2**, 718 (2022).

[10] J. Liu, GPUMD-Wizard: A python package for generating and evaluating machine learning potentials, (2024).

[11] 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).

[12] 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).

[13] 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).

[14] 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).

[15] 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).

[16] 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).

[17] 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).

[18] 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).

[19] A. Togo, First-principles phonon calculations with Phonopy and Phono3py, Journal of the Physical Society of Japan **92**, 012001 (2023).

[20] R. Hill, The Elastic Behaviour of a Crystalline Aggregate, Proceedings of the Physical Society. Section A **65**, 349 (1952).

[21] 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).

[22] 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).

[23] 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).

[24] 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).

[25] 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).

[26] 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).

[27] 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).

[28] 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).

[29] 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).

[30] B. Deng, Materials project trajectory (mptrj) dataset, figshare (2023).

[31] L. G. Longsworth, Temperature Dependence of Diffusion in Aqueous Solutions, The Journal of Physical Chemistry **58**, 770 (1954).

[32] *International Critical Tables*, Vol. VI (McGraw-Hill Book Co., Inc., New York, NY, 1926) p. 230.

[33] 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).