Universal Thermodynamic Interatomic Potentials for Crystalline Materials

Juno Nam,[1] Bowen Deng,[1] Xiaochen Du,[1] Luis Barroso-Luque,[2]
Benjamin Kurt Miller,[2,*] and Rafael Gómez-Bombarelli[1,†]

[1] *Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*
[2] *Fundamental AI Research, Meta, San Francisco, CA 94105, USA*[‡]

(Dated: August 17, 2026)

Free energies govern solid-state phase stability, yet computational materials discovery still relies largely on ground-state energies because free energy calculations require ensemble averages. We introduce the *thermodynamic interatomic potential* (TIP), which extends an interatomic potential from its static energy to a thermodynamically consistent Gibbs free energy model, with thermodynamic responses following from temperature and pressure by automatic differentiation. We implement TIP[UMA] using the universal potential UMA, train it on free energies from quasi-harmonic to molecular dynamics fidelity, and calibrate it to higher-resolution calculations or experiment. From a single evaluation, it returns the equation of state of a crystal and locates phase transitions among competing branches, including dynamically stabilized phases. Fine-tuning extends the model to alloy solubility limits and miscibility gaps. TIP makes the free energy as accessible as the potential energy, opening finite-temperature phase stability to high-throughput discovery.

arXiv:2608.14502v1 [cond-mat.mtrl-sci] 14 Aug 2026

I. INTRODUCTION

Phase diagrams, stability fields, and metastability windows in solid-state materials science are determined by free energies at the relevant conditions, not by ground-state energies alone. Nevertheless, computational materials discovery still relies heavily on static density functional theory (DFT) energetics and convex-hull constructions, because large-scale screening workflows were built around ground-state data [1–3]. This creates a persistent imbalance: structure and energy predictions are available at large scale, whereas the temperature- and pressure-dependent free energies remain less available, more computationally expensive, and less standardized [4].

This imbalance arises because a ground-state energy characterizes one relaxed configuration, whereas a free energy averages over a thermodynamic ensemble including entropic contributions. Computing a free energy requires averaging over thermally accessible configurations for each structure and thermodynamic condition. Approximating this ensemble can be inaccurate because vibrational and configurational contributions often determine phase stability, and vibrational free energies alone can shift phase boundaries by amounts comparable to configurational contributions [5, 6]. The vibrational contribution is particularly demanding because it is strongly temperature dependent and deviates increasingly from harmonic behavior as anharmonicity grows. Phase stability depends on small free-energy differences between competing phases and decomposition products, rather than only on formation from the elements [7, 8]. Therefore, these free energies must be resolved precisely, over the temperature and pressure ranges in which the solid phases remain stable or metastable [9, 10].

First-principles thermodynamics has been addressing this burden by coarse-graining free energies into tractable effective models. The cluster expansion [11–13] represents the configurational energy of a crystal on a fixed parent lattice as a sum of effective interactions fitted to first-principles energies, and statistical-mechanical calculations then yield composition–temperature phase behavior [14, 15]. This approach applies when a crystal can be represented as occupations on a parent lattice [16, 17], but its interactions are specific to a chemistry and lattice and must be refitted for a new system. Additionally, the fitted quantity is usually configurational energy, and temperature enters through a separate statistical-mechanical calculation rather than as model inputs.

A separate hierarchy addresses the vibrational free energy of a fixed crystal. Quasi-harmonic lattice dynamics, built on density functional perturbation theory [18] and finite-displacement phonon workflows [19], is the default route to solid-state free energies. Because it absorbs anharmonicity only into the volume dependence of harmonic phonons, it grows fragile in the regimes that govern many functional solids: soft modes, strong phonon renormalization, entropy-stabilized high-temperature phases, and mechanically unstable harmonic references. Self-consistent phonon theories [20], temperature dependent effective potential methods [21], and formalisms for unstable phases [22, 23] address these cases. At higher fidelity, explicit free energy integration evaluates absolute free energies based on molecular dynamics (MD) simulations, where thermodynamic integration proceeds through Frenkel–Ladd switching [24], reversible scaling [25], and alchemical switching [26], realized as nonequilibrium MD workflows [27, 28]. The computational cost increases with the degree of anharmonicity

* bkmi@meta.com
† rafagb@mit.edu
‡ Meta-affiliated authors served solely in an advisory role. All access to the models, datasets, and code, as well as the development and release of the dataset and code, was carried out exclusively by the non-Meta authors or their academic institution.

resolved, and is especially high when energies and forces are evaluated from first principles.

Universal machine learning interatomic potentials (MLIPs) [29–35] reduce this computational cost by reproducing first-principles energies and forces across broad chemical spaces, enabling high-throughput relaxation and large-scale MD without system-specific refitting. However, MLIPs predict the potential energy of individual configurations. Free energy calculations still require ensemble sampling for each structure and condition, even though individual force evaluations are computationally inexpensive.

An alternative is to model free energy directly rather than apply a statistical-mechanical calculation to an energy model. The CALPHAD method [36, 37] represents the Gibbs free energy of each phase as an analytic function of temperature and composition, assessed from experimental and first-principles data. These assessments are semi-empirical, specific to each chemical system, and often proprietary. Descriptor-based Gibbs free energy model [38] predicts temperature-dependent free energies of stoichiometric solids using compact, interpretable composition-level descriptors, but they are limited to near-ambient pressure and lack the structural resolution needed to identify phase transitions. Structure-resolved surrogates have been developed for specific systems, such as a machine-learned metastable phase diagram of carbon [39], but remain confined to a single chemistry or a narrow class of structures. A chemically transferable, branch-resolved Gibbs free-energy model is still needed, which spans harmonic and anharmonic fidelity levels and derives thermodynamic response functions from one differentiable surface.

We therefore introduce the thermodynamic interatomic potential (TIP). Conceptually, TIP integrates three complementary ideas from existing approaches to solid-state thermodynamics (Fig. 1a): it inherits the chemical transferability of MLIPs; represents the thermodynamics of an ordered branch with an efficient surrogate, analogous to the cluster expansion but without lattice- or chemistry-specific parameterization; and adopts the CALPHAD philosophy of modeling free energy directly as a differentiable thermodynamic function. Together, these elements enable equilibrium properties and thermodynamic response functions to be predicted without repeated statistical-mechanical calculations.

For a relaxed ordered bulk crystal, TIP encodes the structure once and returns a differentiable approximation to that branch's Gibbs free energy surface across a range of temperature and pressure in a single evaluation. This confines the statistical-mechanical cost to training, so the free energy of a new crystal within the model's domain follows without further sampling. TIP resolves the vibrational thermodynamics of one ordered branch at a time, and configurational thermodynamics is obtained by sampling ordered branches across composition, extending the model to solid solutions and miscibility gaps. The current model identifies transitions only among the supplied branch candidates, i.e., it does not discover crystal structures. Also, the liquid state is outside the present scope because it has no ordered branch representation. TIP is built on a universal interatomic potential and trained sequentially on quasi-harmonic and MD free energies, and experimental calibration data. We demonstrate prediction of equations of state and high-temperature phase transitions, including phases without stable harmonic references.

II. RESULTS

A. Thermodynamic Interatomic Potentials

TIP predicts the Gibbs free energy of an ordered crystal as a continuous function of temperature and pressure, from its relaxed structure (Fig. 1b). While conventional free energy calculations require thermal ensemble sampling and thermodynamic path integration at every state point, TIP instead evaluates a fitted continuous surface. We enable this via *free energy adaptation*: the absolute Gibbs free energy of a crystal is decomposed into its static (0 K) energy plus a smooth *thermodynamic residual*. A pre-trained MLIP supplies the static relaxed energy U°, so the adapter model learns only the residual $\hat{G}^\delta(\boldsymbol{x}^\circ; T, P)$, giving the predicted Gibbs free energy

$$\hat{G}(\boldsymbol{x}^\circ; T, P) = U^\circ + \hat{G}^\delta(\boldsymbol{x}^\circ; T, P). \tag{1}$$

Learning $\hat{G}^\delta$ rather than $\hat{G}$ reduces the structure-dependent dynamic range of the target and focuses the adapter on finite-temperature contributions. A single relaxed structure then generates an entire free energy surface over (T, P).

The frozen MLIP encodes the structure, and the adapter uses its per-atom features and the thermodynamic variables (T, P) as inputs (Fig. 1c). The thermodynamic backbone is temperature- and pressure-agnostic: from the structure features alone, it predicts the coefficients of a closed-form thermodynamic expression, and (T, P) enter only through the analytic head. The functional forms are drawn from established physical models: the temperature dependence combines a quantum Einstein oscillator vibrational free energy with a low-order polynomial, as in CALPHAD Gibbs energy functions [36, 37], and the pressure–volume relation of each per-atom contribution follows a Murnaghan-like equation of state [40]. Because these forms are analytic, the fitted coefficients carry physical meaning, and derivatives of the free energy, including higher-order responses, remain smooth by construction. Furthermore, automatic differentiation yields thermodynamic conjugate variables such as volume and entropy, and thermodynamic responses, such as heat capacity, bulk modulus, and thermal expansion, from the same $\hat{G}$, ensuring thermodynamic consistency within the model. Because the analytic head assembles the free energy from per-atom contributions,

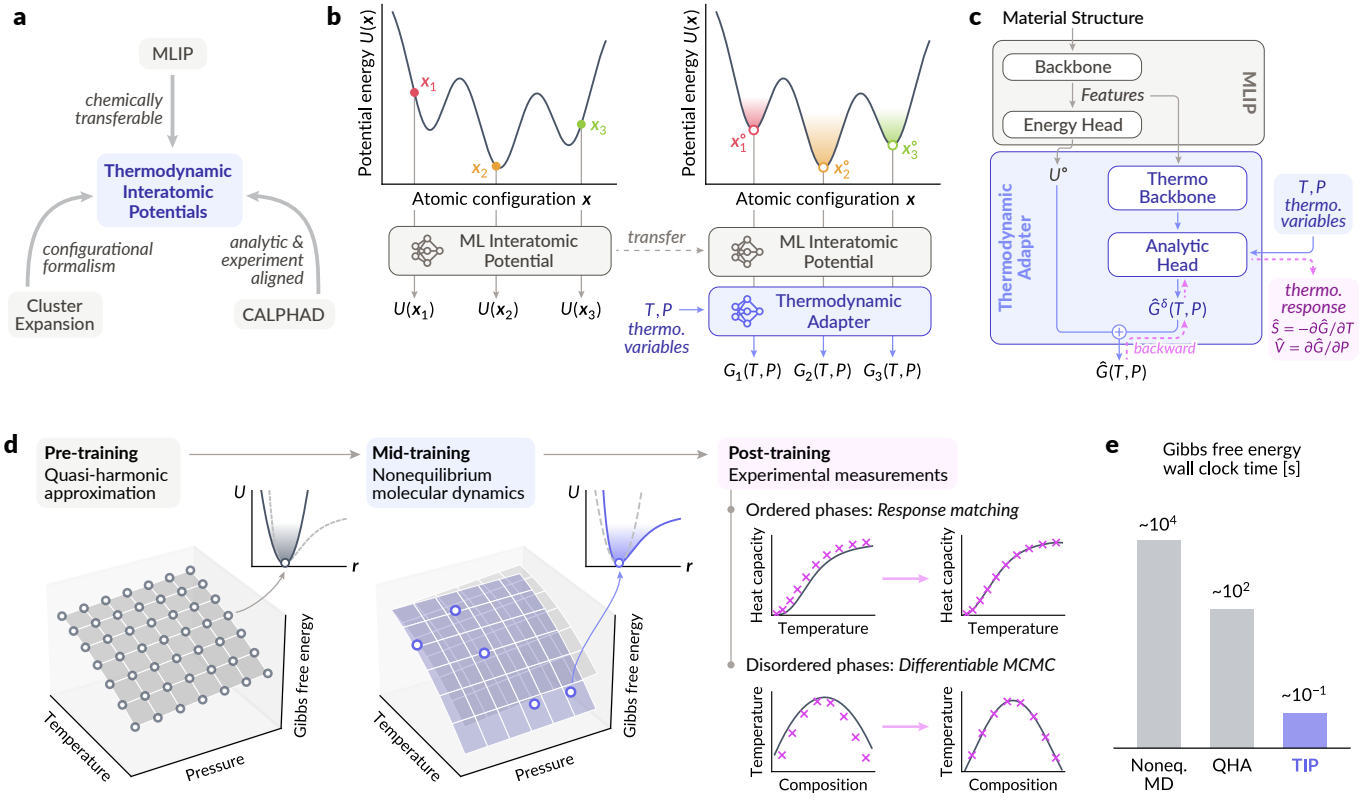

Fig. 1. **Free energy adaptation with a thermodynamic interatomic potential (TIP). (a)** A TIP combines three established routes to solid-state thermodynamics: the chemical transferability of MLIPs, the configurational phase stability formalism of the cluster expansion, and the analytic, experiment-aligned free energy representation of CALPHAD. **(b)** A TIP augments a pre-trained MLIP with a thermodynamic adapter: for each locally relaxed branch representative $\boldsymbol{x}^\circ$, the frozen MLIP encodes $\boldsymbol{x}^\circ$, and the adapter combines those features with (T, P) to predict the branch Gibbs free energy $\hat{G}(T, P)$. **(c)** Architecture: the MLIP backbone's energy head returns the static reference $U^\circ := U(\boldsymbol{x}^\circ)$ and its per-atom features condition the adapter, whose analytic head outputs the thermodynamic residual $\hat{G}^\delta(T, P)$. Their sum defines $\hat{G}$, and its derivatives give the response properties. **(d)** Three-stage training: pre-training on quasi-harmonic (T, P) grids, mid-training on anharmonic absolute free energies from nonequilibrium MD, and post-training against experimental data by response matching for ordered phases (e.g., heat capacity) and differentiable reweighting for disordered phases (e.g., composition–temperature binodals). **(e)** Wall-clock time per Gibbs free energy evaluation on a given $\boldsymbol{x}^\circ$: nonequilibrium MD ($\sim 10^4$ s) and the quasi-harmonic approximation ($\sim 10^2$ s) compared with one TIP[UMA] evaluation ($\sim 10^{-1}$ s), five and three orders of magnitude faster.

model inference requires only the relaxed unit cell, even though the training labels are computed on the large supercells required to converge phonon and MD free energies. TIP is therefore trained to reproduce those free energies from a unit cell input, at a computational cost set by the unit cell rather than the simulation supercell. This reduces calculations requiring hours of explicit sampling or minutes of a quasi-harmonic workflow to seconds of model inference (Fig. 1e).

TIP models one ordered crystalline branch at a time, represented by its relaxed representative $\boldsymbol{x}^\circ$. A branch may be a basin of the potential energy surface, a symmetry-constrained stationary point for a dynamically stabilized phase (Fig. 1b), a distinct polymorph, or a distinct chemical ordering. Each branch has a distinct predicted free energy surface $\hat{G}(\boldsymbol{x}^\circ; T, P)$. Stable phases and the transitions between phases are then resolved by comparing the predicted free energies of the candidate branches. The formal branch construction and its underlying assumptions are given in Supplementary Note A. This way, a TIP realizes the coarse-grained free energy landscape of first-principles phase stability theory [6] in a differentiable, transferable form, made practical by universal MLIPs and improved statistical-mechanical algorithms.

Our instantiation, TIP[UMA], is built on Universal Models for Atoms (UMA) [41], an MLIP trained across chemistries on large DFT datasets. It is trained in three stages of increasing physical fidelity (Fig. 1d). Pre-training fits dense quasi-harmonic free energy grids over (T, P), which are computationally cheap and abundant but approximate each potential energy basin as harmonic. Mid-training corrects toward anharmonic, absolute free energies computed from nonequilibrium MD. Because a TIP exposes the thermodynamic responses as consistent derivatives of a single free energy surface, mea-

sured responses can supervise TIP directly, whereas an MLIP could obtain the same observables only through ensemble reweighting. Post-training exploits this, calibrating against experiment by matching measured responses, such as heat capacity, for ordered phases and through differentiable reweighting of Monte Carlo simulations for disordered phases.

B. Multi-Fidelity Gibbs Free Energy Dataset

Training a structure-conditioned free energy surrogate requires thermodynamic labels at scale, which no single method can provide both efficiently and accurately. We therefore assemble a multi-fidelity dataset that combines abundant, approximate labels with sparse but accurate labels (Fig. 2a). We draw the initial structures from the Materials Project [1], retain metastable entries within 0.2 eV/atom of the convex hull (123,424 structures), and relax each to an ordered branch representative $\boldsymbol{x}^\circ$ with UMA [41]. A representative subset is selected by BIRCH clustering [43] of the UMA structure embeddings, so that computationally expensive labels are assigned to a structurally diverse set of representatives while low-fidelity labels cover all structures.

The low-fidelity signal is the QHA Gibbs free energy $G^{\mathrm{QHA}}(T, P)$, computed for every structure; it is computationally inexpensive and includes volume-dependent harmonic vibrational contributions but neglects anharmonic mode coupling. The high-fidelity signal is the absolute Gibbs free energy $G^{\mathrm{MD}}(T, P)$ from nonequilibrium MD, computed only on representatives by combining a Frenkel–Ladd absolute reference on a fast surrogate potential (Orb-v3 [42], trained on the same materials dataset as UMA) with an alchemical transfer to the accurate target potential; it includes anharmonic effects but is far more demanding. After quality filtering, the dataset comprises over 120,000 QHA Gibbs free energy surfaces and roughly 95,000 MD state points, spanning temperatures from 0 K up to 4000 K and pressures up to 40 GPa. The combined labels cover much of the periodic table (Fig. 2b) and a broad range of the thermodynamic residual G^{δ} over temperature and pressure (Fig. 2c).

C. Validation and Error Analysis

After pre-training on the quasi-harmonic (T, P) grids and mid-training on the anharmonic MD labels, we evaluate TIP[UMA] on a held-out validation set not used in training (Fig. 3). Across the filtered validation set, the model attains a global median absolute Gibbs energy error of 10.3 meV atom^{-1}. For context, the descriptor-based Gibbs free energy model of Bartel *et al.* [38] reported test errors of 46–60 meV atom^{-1}, although the targets and validation datasets differ. The TIP[UMA] error is also comparable to the energy scale governing polymorph competition: experimentally observed inorganic phases have a median energy above the convex hull of approximately 15 meV atom^{-1} [9], while chemistry-dependent metastability windows can extend to tens or hundreds of meV atom^{-1} [10]. The element-resolved errors concentrate in chemically distinct regions (Fig. 3a): the halogens, chalcogens, alkali and alkaline-earth metals, heavy p-block elements, and actinides carry the largest median errors, whereas most transition metals sit at or below the global median. These groups include soft, polarizable, and strongly anharmonic chemistries whose absolute free energies are difficult to converge, and rare chemistries like actinides are sparsely represented in the Materials Project pool.

Error trends with thermodynamic state and the underlying nonequilibrium MD convergence are more systematic (Fig. 3b). Comparing the pre-trained and mid-trained models (light versus dark bars), mid-training on the MD labels lowers the median error about threefold at every temperature, correcting the offset between the quasi-harmonic pre-training and the absolute MD reference. For the mid-trained model, the error grows only weakly with temperature and more strongly at high pressure, and its strongest dependence is on the Frenkel–Ladd switching dissipation of the high-fidelity calculation. At the highest dissipation, the signed median error also becomes positive, indicating systematic bias rather than symmetric scatter. Such high-dissipation labels are only a small fraction of the dataset (Fig. 3b, left). Because the dissipated work measures how far each switching trajectory departs from reversibility, the positive bias at high dissipation reflects the convergence of the free energy labels rather than a failure of the surrogate. The least-converged calculations are removed by a per-leg dissipation filter during curation. These trends indicate that improving label coverage and convergence is more likely to reduce worst-case errors than increasing the model capacity.

D. Ordered Crystalline Phases

For an ordered crystalline branch, the equilibrium response functions are derivatives of the Gibbs free energy, and TIP[UMA] predicts this surface from a single relaxed structure. From this, we evaluate heat capacities, equations of state, polymorphic transitions, and a phase diagram (Fig. 4).

For 255 stoichiometric solids curated from the NIST–JANAF thermochemical tables [44] (Methods), the mid-trained model predicts the standard entropy and heat capacity with median absolute errors of 2.28 and 1.94 J K^{-1} (mol atoms)$^{-1}$ (Fig. 4a). The predicted per-atom heat capacity tracks the NIST–JANAF values [44] closely for Al_2O_3 and CaO over the full temperature range, correctly recovering $\hat{C}_P \to 0$ as $T \to 0$ (Fig. 4b). Because the MD labels do not directly constrain heat capacity, mid-training regularizes C_P toward the frozen quasi-harmonic model, while the analytic Ein-

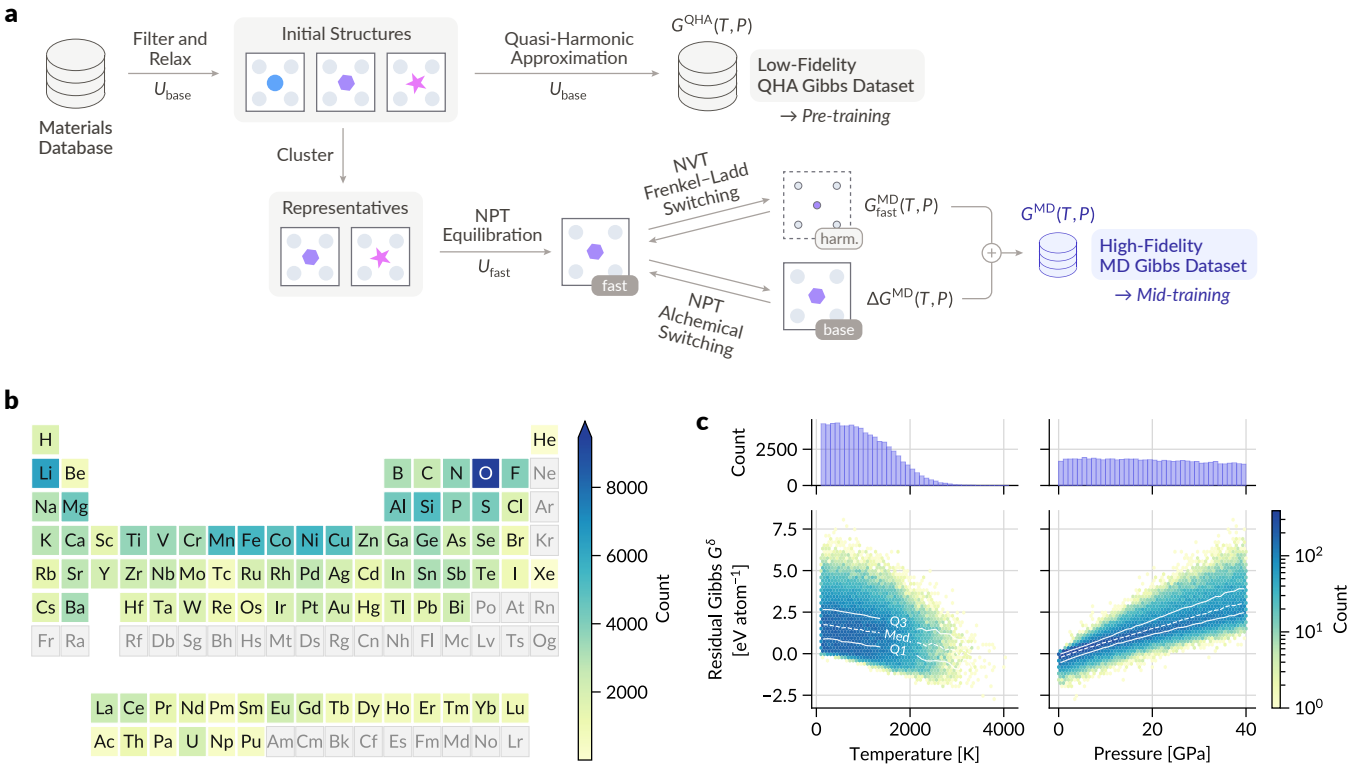

Fig. 2. **Multi-fidelity Gibbs free energy dataset. (a)** Data-generation workflow. Structures from a materials database are filtered and relaxed with the base potential U_{base} (UMA [41]); the quasi-harmonic approximation (QHA) on each yields the low-fidelity QHA Gibbs dataset $G^{\text{QHA}}(T,P)$. Clustering selects representatives, which are equilibrated under NPT dynamics with a fast potential U_{fast} (Orb-v3 [42]) and then processed by NVT Frenkel–Ladd switching [24] (absolute free energy against a harmonic Einstein crystal reference) and NPT alchemical switching (transfer from the fast to the base Hamiltonian); combining the two contributions gives the high-fidelity MD Gibbs dataset $G^{\text{MD}}(T,P)$. **(b)** Elemental coverage of the MD Gibbs dataset: the periodic table is colored by the number of dataset entries containing each element. **(c)** Distribution of the residual Gibbs free energy G^{δ} of the MD Gibbs dataset versus temperature (left) and pressure (right), shown as two-dimensional histograms with the marginal count distributions above; white lines mark the quartiles (Q1, median, Q3).

stein term preserves the low-temperature limit. Mid-training adds an anharmonic free energy correction without replacing the quasi-harmonic low-temperature heat-capacity behavior with the classical MD limit. The larger errors for Cr_2O_3 and Fe_3O_4 arise from magnetic λ-anomalies (the antiferromagnetic Néel and ferrimagnetic Curie transitions) in the measured C_P, which lie outside the vibrational contributions. An additional explicit magnetic or electronic term could capture these contributions (see Discussion). The volumetric response of forsterite Mg_2SiO_4 shows that the predicted compression and thermal expansion reproduce the UMA reference behavior (Fig. 4c,d). Both the UMA QHA reference and the model predictions retain the known volume overestimate of the Perdew–Burke–Ernzerhof (PBE) functional, inherited from its systematic underbinding and a property of the training target rather than the adapter. Post-training against experimental labels later removes it (see below).

We next examine temperature-driven polymorphic transitions involving high-temperature phases that the quasi-harmonic surface does not describe adequately (Fig. 4e–h). These transitions are entropy-driven and require anharmonic treatment: the high-temperature bcc metals are dynamically unstable at 0 K and have no harmonic reference, while for $CaSiO_3$ a harmonic reference exists but omits the anharmonic entropy that sets the transition [20–23]. Consistent with these limitations, the pre-trained model produces no $\Delta G = 0$ crossing for Ti or Sc, and for Hf and $CaSiO_3$ it crosses only far above experiment (near 3200 and 2200 K, against 2016 and 1398 K). After mid-training on nonequilibrium MD free energies, the model predicts three of the four (Sc only in part, below), placing the Ti and Hf α-hcp $\rightarrow$ β-bcc transitions and the $CaSiO_3$ wollastonite $\rightarrow$ pseudowollastonite transition within $\sim$250 K of experiment. Of the four, only Ti has both phases in the MD training set, while Sc and Hf have their high-temperature bcc phase held out and $CaSiO_3$ has neither polymorph. These entropy-driven transitions are recovered even though mid-training updates only a lightweight adapter on a frozen encoder, indicating that the MD labels provide transferable anharmonic corrections. The Sc α-hcp $\rightarrow$ β-bcc transition (Fig. 4h) is recovered only in part: mid-training moves

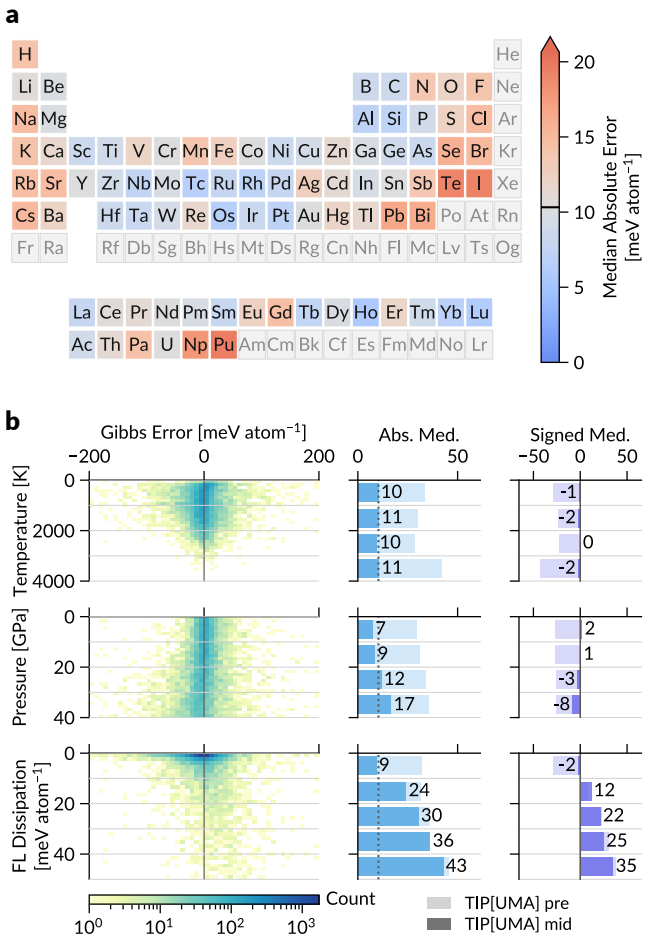

Fig. 3. **Validation error analysis. (a)** Element-wise validation error. Periodic-table map of the median absolute Gibbs error for each element, computed over the filtered validation data points that contain that element; elements represented by fewer than 20 validation state points are shown in gray. The black tick on the colorbar marks the global median absolute error. **(b)** Thermodynamic and dissipation trends. Signed Gibbs error distributions as functions of temperature, pressure, and Frenkel–Ladd (FL) dissipation for the filtered validation set. Left panels show two-dimensional count histograms of the mid-trained model on a logarithmic color scale; right panels show the binned median absolute error and the signed median error (meV atom^{-1}) for the pre-trained (light) and mid-trained (dark) models. The dotted line in the absolute-median panels marks the global median absolute error.

ΔG toward a crossing but does not produce a crossing at the measured transition temperature, showing that the anharmonic correction remains imperfect where high-fidelity training data are sparse.

The pre- and mid-trained models so far inherit the systematic volume overestimate of the PBE reference. Post-training (Fig. 1d) removes this bias, calibrating the model against the Holland–Powell experimental thermodynamic assessment over the Si–Al–Mg–Ca–O chemical space [45]. The model is fitted on the binary and ternary oxides, with the quaternary oxides held out (Methods and Supplementary Note D). The calibration lowers the bulk modulus error against experiment from ∼13% to a few percent, decreases the reaction Gibbs free energy and entropy errors, and removes the ∼5% volume overestimate from the PBE offset (Fig. 4c,d, TIP[UMA] post). The metrics on the held-out quaternary phases improve alongside those on the fitted binary and ternary ones (Fig. 4i), indicating that the representation learned through pre- and mid-training supports transferability, so calibration on a small experimental set refines the surface rather than fitting each chemistry independently. The same calibration also corrects the SiO_2 pressure–temperature phase diagram (Fig. 4j–l): the mid-trained model spuriously stabilizes cristobalite at ambient conditions, and post-training restores quartz as the ambient ground state and the quartz → coesite → stishovite ordering within the supplied candidate set [48].

E. Disordered Solid Solutions

Ordered branches are only part of the phase stability problem: alloy solubility limits and miscibility gaps require configurational disorder in addition to ordered-branch thermodynamics. TIP treats these systems by evaluating branch free energies across composition on a parent lattice, which defines the configurational Hamiltonian used to sample the disordered state. That free energy can then be calibrated against a chosen reference in either of two complementary directions: *bottom-up*, distilling a higher-resolution calculation for a target chemistry into the surrogate, or *top-down*, matching measured phase boundaries.

The dilute solubility of Sc in fcc Al, which depends on the vibrational entropy of dissolution in addition to configurational mixing [49], provides a bottom-up example of chemistry-specific fine-tuning across disordered compositions and configurations (Fig. 5a). At each temperature and $P = 0$, TIP[UMA] evaluates the Gibbs free energies required to compute the dissolution free energy $\Delta G_{\mathrm{sol}}(T)$ relative to the adjacent Al–Al_3Sc convex hull (Eq. (21)). The equilibrium solubility then follows from the ideal-dilute relation $x_{\mathrm{Sc}} \approx \exp(-\Delta G_{\mathrm{sol}}/k_{\mathrm{B}}T)$ (Eq. (22)). Using only 0 K static energies from the baseline UMA potential (no vibrational free energy), the estimate increases with temperature through the ideal-mixing term but remains about an order of magnitude below the measured solubility across 600–900 K. Including branch vibrational free energies increases x_{Sc} by one to nearly two orders of magnitude over that static line, reaching the measured range even from the pre-trained model with no Al–Sc-specific tuning. We then test chemistry-specific distillation by fine-tuning the surrogate on an additional set of UMA quasi-harmonic free energies for diverse fcc Al–Sc structures spanning compositions and configurations using a lightweight low-rank adaptation (LoRA) [50] (Methods and Supplementary Note B). The

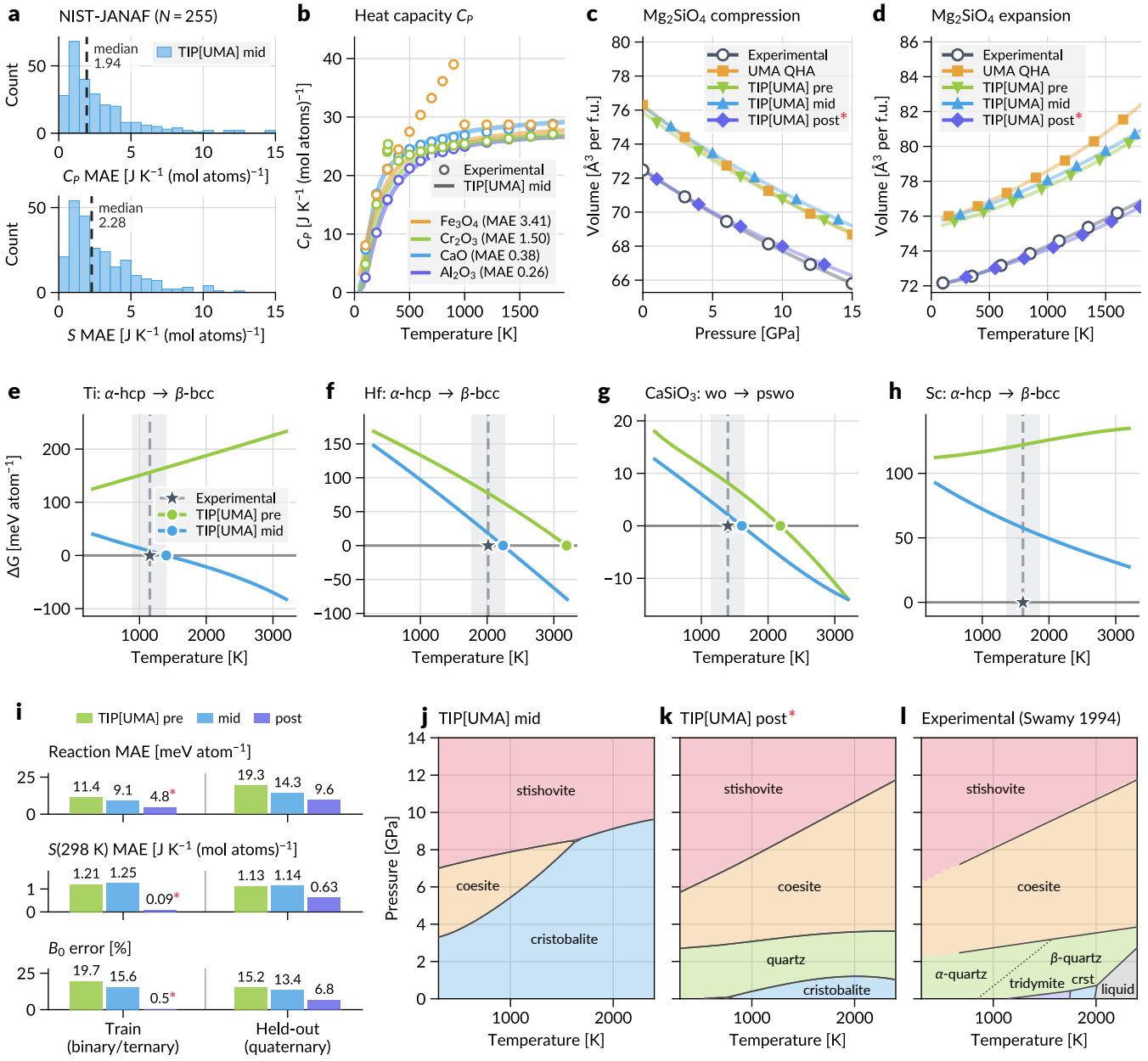

Fig. 4. **Thermodynamics of ordered crystalline phases from a learned free energy surface.** **(a)** Mid-trained mean absolute error in heat capacity C_P and standard entropy $S(298\ \mathrm{K})$ across the NIST–JANAF benchmark of 255 stoichiometric solids, with medians of 1.94 (C_P) and 2.28 (S) J K^{-1} (mol atoms)$^{-1}$. **(b)** Mid-trained heat capacity $C_P(T)$ (solid lines) against NIST–JANAF values [44] (open circles). The larger deviations for Cr_2O_3 and Fe_3O_4 occur near magnetic λ-anomalies (Néel and Curie transitions) outside the model's vibrational scope. **(c)** Isothermal compression $V(P)$ at 300 K and **(d)** thermal expansion $V(T)$ at $P \approx 0$ for forsterite Mg_2SiO_4, comparing experiment (Holland–Powell [45]), the UMA quasi-harmonic reference, and TIP[UMA] across pre-, mid-, and post-training. The quasi-harmonic reference and the pre- and mid-trained models share the known PBE volume overestimate, which post-training removes. **(e–h)** Temperature-driven polymorphic transitions, shown as the per-atom Gibbs difference $\Delta G(T)$ between the high- and low-temperature phases for the pre-trained and mid-trained models: **(e)** Ti, **(f)** Hf, **(g)** $CaSiO_3$, and **(h)** Sc. The $\Delta G = 0$ crossing (circle) is the predicted transition temperature, and the dashed line and star mark experimental transition [46, 47] with the shaded band spanning a ± 250 K tolerance. The pre-trained surface shows no crossing (Ti, Sc) or one far above experiment (Hf, $CaSiO_3$), whereas mid-training recovers the transitions, Sc only in part. **(i)** Accuracy of the pre-, mid-, and post-trained models on the Holland–Powell ordered-oxide benchmark for the reaction Gibbs free energy, standard entropy, and bulk modulus B_0, on a training partition (binary and ternary oxides) and a held-out partition (quaternary oxides). Post-training improves every metric on both, including the held-out quaternaries. **(j–l)** SiO_2 pressure–temperature phase diagram for **(j)** the mid-trained model, **(k)** the post-trained model, and **(l)** the experimentally assessed topology [48]. The mid-trained model spuriously stabilizes cristobalite at ambient conditions, whereas post-training recovers the low-pressure quartz field and the quartz, coesite, stishovite ordering within the supplied candidate set. Red asterisks mark comparisons for which the experimental value was included in calibration.

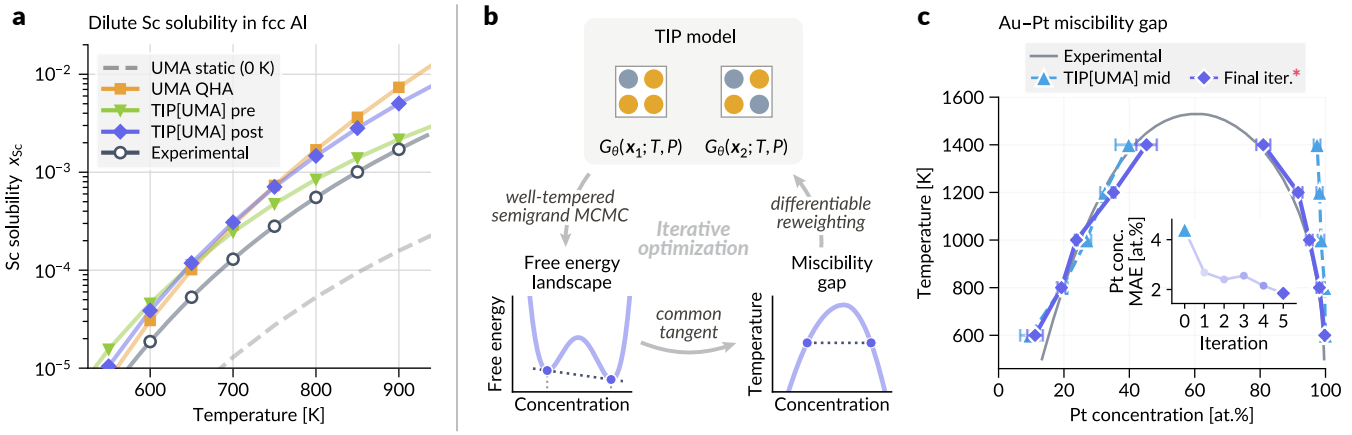

Fig. 5. **Calibration of TIP[UMA] for disordered phases using computational and experimental references.** Lightweight fine-tuning fits the frozen free energy surrogate to a chosen reference and extends it from ordered branches to configurationally disordered solid solutions, preserving thermodynamic consistency. **(a)** Dilute Sc solubility in fcc Al versus temperature, obtained from the dissolution free energy relative to the adjacent Al–Al_3Sc convex hull (Section IV). By accounting for vibrational entropy, TIP[UMA] predicts the measured order of magnitude and tracks the quasi-harmonic UMA reference (UMA QHA), whereas the 0 K static estimate falls about an order of magnitude below the measured value; Al–Sc-focused fine-tuning (pre to post) moves the model toward that reference. The computed reference itself lies above the measured solubility, an offset set by the reference thermodynamics rather than the surrogate. **(b)** Workflow for a disordered phase: TIP[UMA] scores lattice configurations, well-tempered semigrand Monte Carlo samples the mixing free energy, a common-tangent construction yields the miscibility gap, and differentiable reweighting calibrates the surrogate. **(c)** Au–Pt miscibility gap, composition in Pt atomic percent (at.%): iterated differentiable reweighting moves the TIP[UMA] binodal from its initial prediction to the experimental solvus, the Pt-concentration error falling over the iterations (inset). Horizontal bars are the seed spread, one standard deviation of the branch Pt concentration across four independent well-tempered metadynamics seeds. Red asterisks mark comparisons for which the experimental value was included in calibration.

fine-tuned model then more closely reproduces the dedicated Al–Sc quasi-harmonic reference. While the reference overpredicts the measured high-temperature solubility by a factor of three to four, this example evaluates fidelity to the computational reference rather than agreement with experiment.

The Au–Pt miscibility gap provides a top-down example in which the model is calibrated to the experimental solvus [51, 52]. Since TIP[UMA] assigns each lattice configuration a branch-conditioned Gibbs free energy that includes explicit temperature and pressure dependence, it serves as a temperature- and pressure-dependent configurational Hamiltonian analogous to a cluster expansion. Semigrand canonical Monte Carlo [53], accelerated by well-tempered metadynamics [54], therefore samples the mixing free energy across composition directly on the surrogate. A common-tangent construction then yields the miscibility gap (Fig. 5b), and differentiable reweighting [55] of the stored samples calibrates the surrogate toward the measured solvus over successive cycles. To reduce computational cost, the structure is fully relaxed every few steps, the stored samples are reused within each cycle, and the metadynamics bias is warm-started between cycles (Methods and Supplementary Note B). Over these cycles, the predicted binodal moves toward the measured solvus (Fig. 5c), and the error in Pt concentration falls (inset). The same surrogate supports both ordered and disordered thermodynamics and can be specialized against computational or experimental references. In both cases, differentiable calibration refines an effective free energy model against the available reference without fitting a separate configurational Hamiltonian, providing a flexible realization of the coarse-graining program of first-principles phase stability theory [6].

F. Chemical Trends in Predicted Responses

To examine what chemical information its learned atomic contributions encode, we analyze its per-atom contributions. Although it is supervised only on structure-level thermodynamics, its head assembles each Gibbs free energy surface additively from per-atom contributions. This decomposition defines per-atom entropy $\hat{S}_i$ and volume $\hat{V}_i$ that sum to the structure totals ($\hat{S} = \sum_i \hat{S}_i$, $\hat{V} = \sum_i \hat{V}_i$). We therefore test whether these learned descriptors follow established chemical trends (Fig. 6).

Across the periodic table, the predicted per-atom entropy $\hat{S}_i$ is largest for large, soft, polarizable elements such as the heavy alkali metals, Tl, Hg, and the heavy halogens, and smallest for compact covalent and stiff refractory species such as B, C, Si, P, and the early-to-mid transition metals (Fig. 6a). The element-level medians of entropy and volume are strongly correlated, and both generally increase with atomic mass (Fig. 6b). Notable

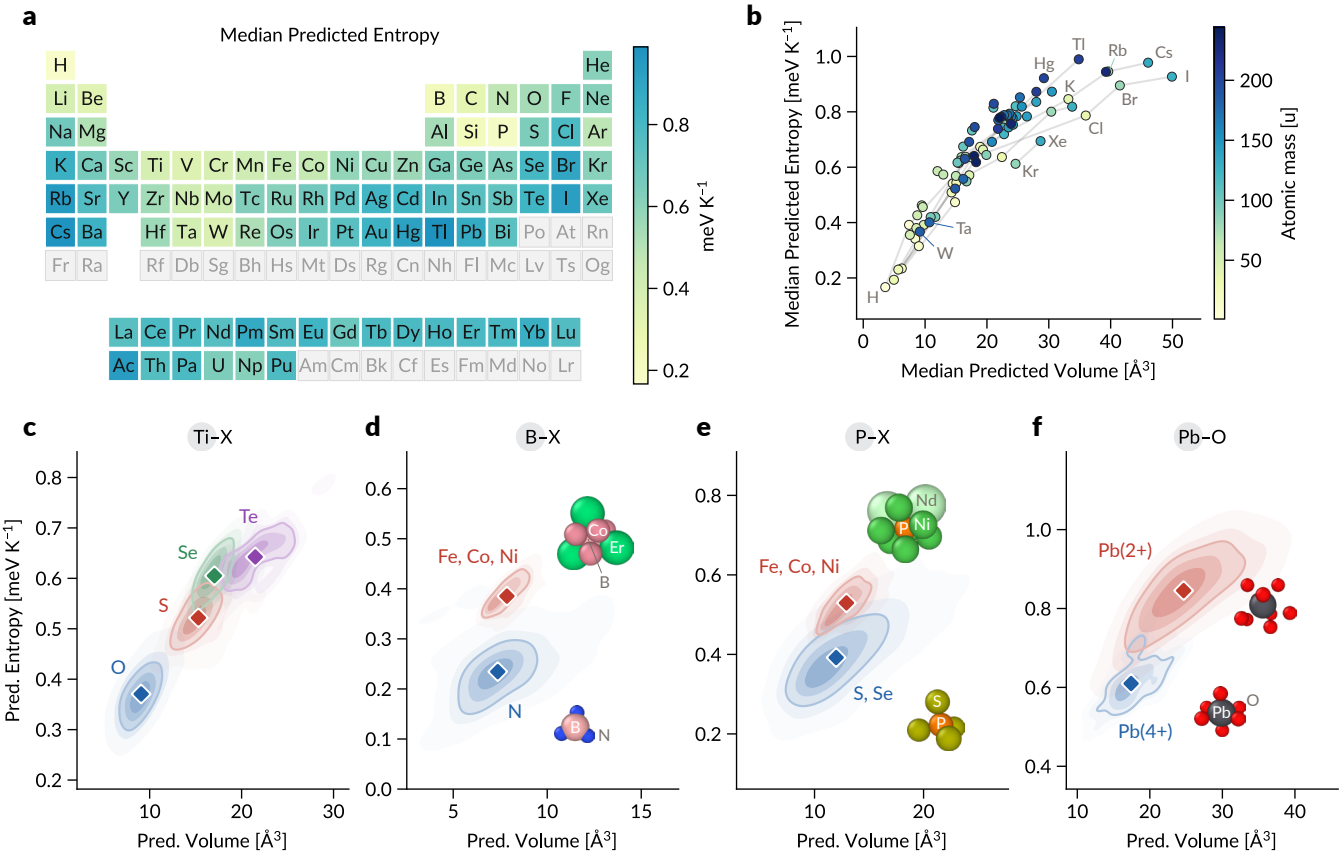

Fig. 6. **Chemical trends in predicted entropy and volume.** Predicted per-atom entropy $\hat{S}$ and volume $\hat{V}$ from TIP[UMA] at $T = 1000$ K and $P = 0$ GPa, across the filtered high-fidelity dataset. **(a)** Periodic table colored by each element's median predicted entropy. **(b)** Element-level median entropy versus median volume, colored by atomic mass, with faint guide lines connecting elements within the same group; the two are strongly correlated. **(c–f)** Site-resolved entropy–volume density maps, each grouping sites by one aspect of local chemistry over all sites of the central element, with diamonds marking the group medians. **(c)** Neighbor identity, Ti–X, for neighbors O, S, Se, and Te. **(d,e)** Bond character: covalent B–N and P–S/P–Se sites versus metallic Fe, Co, and Ni borides and phosphides. **(f)** Formal valence, Pb–O: compact Pb^{4+} versus expanded Pb^{2+}, softened by a stereochemically active $6s^2$ lone pair.

exceptions are Ta and W, which remain compact with low entropy because of their strong bonding, in contrast to softer elements such as I, Cs, Tl, and Hg with similar masses, which occupy larger volumes with higher entropy. Compared with the composition-level descriptor of the Gibbs free energy from Bartel et al. [38], in which atomic mass enters directly through a reduced-mass term alongside volume, TIP[UMA] shows a related but more mixed dependence, with mass dependence mediated mainly by the bonding environment and atomic volume. This pattern is consistent with vibrational entropy increasing as local bonding softens and effective atomic volume grows.

Within an element, much of the variation arises from local environment, and because TIP[UMA] is structure-conditioned and per-atom, it resolves variation that a composition-level descriptor averages away. We group sites by neighbor identity, bond character, and formal valence (Fig. 6c–f). For Ti, predicted volume and entropy increase monotonically as the dominant neighbor becomes heavier and more polarizable from O through S, Se, and Te (Fig. 6c). Bond character separates sites at similar predicted volume: stiff, low-coordination covalent sites, such as threefold B–N or P bound to S and Se, fall well below the soft, high-coordination metallic sites of the corresponding borides and transition-metal phosphides (Fig. 6d,e). This separation is consistent with the established microscopic picture of vibrational entropy: short, stiff, high-frequency bonds carry little entropy, while soft, low-frequency metallic bonding carries much more [5, 6]. Formal valence provides an additional distinction: even when the neighboring chemistry is the same, the response descriptors separate by oxidation state. In Pb–O, compact, near-octahedral Pb^{4+} has lower predicted entropy, while Pb^{2+} has larger predicted volume and entropy, consistent with its stereochemically active $6s^2$ lone pair (Fig. 6f).

Because only the structure-level properties are supervised and atomic decompositions of bulk responses are non-unique, $\hat{S}_i$ and $\hat{V}_i$ should be interpreted as learned descriptors rather than physical observables. Even so,

their agreement with several established trends in local chemistry provides qualitative evidence that the representation captures chemically meaningful structure.

III. DISCUSSION

TIP predicts a branch-conditioned Gibbs free energy surface $G(\boldsymbol{x}^\circ; T, P)$ from a relaxed crystal structure. Automatic differentiation yields volume, entropy, and heat capacity from that surface rather than from separate predictions. The model is trained across fidelities, from quasi-harmonic grids to molecular dynamics free energies, and calibrated against either higher-resolution calculations or experiment. TIP therefore links several approaches for first-principles thermodynamics within one differentiable model.

The present model includes vibrational and configurational contributions but omits magnetic and electronic entropy. The heat capacity anomalies of Cr_2O_3 and Fe_3O_4 (Fig. 4) illustrate this limitation. Including this entropy is a prerequisite for magnetic and mixed-valence solids, where it can qualitatively rearrange the phase diagram [56]. The modular adapter could include these additional residual terms analogous to magnetic models in CALPHAD [57]. Alternatively, charge- and spin-informed per-atom features, for example from CHGNet [33], could be supplied to the encoder. This would offer a complementary route in which the model resolves the magnetic and electronic entropy from electronic-structure-derived features directly. Point defects [58], surfaces [59], and interfaces lie outside the ordered bulk branches. Their free energies scale per defect or per area rather than with bulk atom count. Extending TIP to these objects would require chemical potentials and finite-size or slab treatments. Occupational disorder is resolved here by conditioning each branch on a specific ordering and sampling compositions with semigrand canonical Monte Carlo, so the configurational search remains computationally costly. Supplying site occupancies directly as continuous inputs [26] could reduce or replace explicit configurational sampling and enable efficient alloy thermodynamics.

A second limitation is the quality of the high-fidelity reference. TIP[UMA]'s worst-case error tracks the Frenkel–Ladd switching dissipation of the underlying calculation more strongly than with the tested variables (Fig. 3). Accuracy is therefore set by the quality of the free energy labels, making label convergence and coverage major determinants of accuracy. Improvements should therefore prioritize tighter reversibility of the switching simulations, broader chemical and structural coverage, and extension to higher pressures and disordered structures. The underlying level of theory imposes a separate limit: the PBE underbinding inherited from UMA sets a systematic volume and bulk modulus bias that experimental post-training can correct empirically but cannot remove from the underlying reference. Rebuilding the same construction on a base potential trained at a higher level of theory would address this bias at its source. Finally, extension to liquids requires both new absolute reference states [60] and input representations that do not depend on a relaxed ordered branch. Meeting both requirements would enable full solid–liquid phase diagrams.

Together, these extensions define a broader goal: a differentiable thermodynamic foundation model calibrated to first-principles and experimental data and evaluated at MLIP-like inference compute cost. Such a model would make finite-temperature free energies and response functions routine inputs to materials discovery.

IV. METHODS

A. Branch-conditioned Gibbs free energy

TIP models the Gibbs free energy of an ordered branch labeled by its relaxed reference structure $\boldsymbol{x}^\circ$. Each $\boldsymbol{x}^\circ$ is obtained by a symmetry-constrained relaxation of the static potential energy surface. The branch-conditioned isothermal-isobaric (NPT) partition function restricts configuration space to the basin $\mathcal{B}(\boldsymbol{x}^\circ)$ that relaxes onto $\boldsymbol{x}^\circ$, and the resulting $G(\boldsymbol{x}^\circ; T, P)$ marginalizes vibrational and cell fluctuations within the branch. The formal branch map, the partition function construction, and the three theoretical assumptions (branch representability, finite-(T, P) branch stability, and fixed-branch scope) that make this target well defined are given in Supplementary Note A.

For modeling, we use the static/thermodynamic split of Eq. (1), with $U^\circ := U(\boldsymbol{x}^\circ)$ the static energy of the branch representative and $\hat{G}^\delta$ absorbing all finite-(T, P) contributions. The zero-point energy is excluded from the labels (Eq. (8)), to maintain consistency with the classical MD reference, which carries no zero-point contribution. This convention also avoids an ill-defined harmonic zero-point energy for dynamically unstable branches with imaginary modes. First-order equilibrium responses follow as branch-conditioned thermodynamic derivatives,

$$S = -\left(\frac{\partial G}{\partial T}\right)_P, \qquad V = \left(\frac{\partial G}{\partial P}\right)_T, \tag{2}$$

and the second-order responses (heat capacity, isothermal bulk modulus, and volumetric thermal expansion) follow from higher derivatives of the same G,

$$C_P = -T\left(\frac{\partial^2 G}{\partial T^2}\right)_P, \tag{3}$$

$$B = -V\left(\frac{\partial^2 G}{\partial P^2}\right)_T^{-1}, \tag{4}$$

$$\alpha = \frac{1}{V}\frac{\partial^2 G}{\partial T\,\partial P}. \tag{5}$$

B. Gibbs free energy model

We parametrize the thermodynamic residual atomwise to respect extensivity,

$$\hat{G}^{\delta}(\boldsymbol{x}^{\circ};\, T, P) = \sum_{i=1}^{N} \hat{g}_i^{\delta}(\boldsymbol{h}_i;\, T, P), \tag{6}$$

where $\boldsymbol{h}_i$ is the per-atom latent feature produced by the thermodynamic backbone introduced in the main text (Fig. 1c). The backbone is an equivariant PaiNN encoder [61] that takes the frozen MLIP's scalar (l=0) and vector (l=1) per-atom features, and refines them by equivariant message passing over the local geometry. The encoder receives neither temperature nor pressure, so (T, P) enter only at the head; we use a trunk with $r_{\max} = 6$ Å, $d = 64$ latent features, 16 radial basis functions, and 3 message-passing layers. Each per-atom residual is written so that the volume is exactly the pressure derivative,

$$\hat{g}_i^{\delta}(\boldsymbol{h}_i;\, T, P) = \hat{g}_i^{(0)}(\boldsymbol{h}_i;\, T) + \int_0^P \hat{V}_i(\boldsymbol{h}_i;\, T, p)\, \mathrm{d}p, \tag{7}$$

so that $\hat{V} = \partial \hat{G}^{\delta}/\partial P = \sum_i \hat{V}_i$ holds by construction, while $\hat{S} = -\partial \hat{G}/\partial T$ follows by automatic differentiation in temperature.

The zero-pressure reference branch combines an Einstein oscillator vibrational term with a low-order polynomial in a normalized temperature $\tau := T/T_{\max}$ (with $T_{\max} = 4000$ K a fixed global scale),

$$\hat{g}_i^{(0)}(\boldsymbol{h}_i;\, T) = c_0 + c_1\tau + c_2\tau^2 + c_3\tau^3 + 3k_{\mathrm{B}}T \ln\!\big(1 - e^{-\theta_{\mathrm{E}}/T}\big), \tag{8}$$

with the coefficients $\{c_k\}$ and the Einstein temperature $\theta_{\mathrm{E}} = T_{\max}[\mathrm{softplus}(\tilde{\theta}) + \theta_{\min}]$ predicted from $\boldsymbol{h}_i$. This five-coefficient form (c_0–c_3 and θ_{E}) is the model used in pre-training. Mid-training augments this branch with two gated anharmonic terms, adding the coefficients $c_{\log}$ and c_{inv} for seven in total,

$$\hat{g}_i^{(0),\mathrm{mid}}(\boldsymbol{h}_i;\, T) = \hat{g}_i^{(0)}(\boldsymbol{h}_i;\, T) + w(\tau)\left[c_{\log}\,\tau\ln\tau + \frac{c_{\mathrm{inv}}}{\tau}\right], \tag{9}$$

where sigmoidal gate $w(\tau) = (\tau/\tau_{\mathrm{E}})^p/[1 + (\tau/\tau_{\mathrm{E}})^p]$, with $\tau_{\mathrm{E}} := \theta_{\mathrm{E}}/T_{\max}$ and $p = 4$. This confines the $\tau\ln\tau$ and $1/\tau$ corrections to temperatures above θ_{E}, so they represent anharmonic behavior beyond the quasi-harmonic reference while leaving the low-temperature branch intact. Each atomic volume contribution follows a Murnaghan-like pressure response [40] whose per-atom parameters are softplus-transformed quadratic functions of τ,

$$V_{0,i}(\tau) = \mathrm{softplus}\big(\textstyle\sum_{k=0}^{2} a_k^{V_0}\tau^k\big) + V_{0,\min}, \tag{10}$$

$$B_{0,i}(\tau) = \mathrm{softplus}\big(\textstyle\sum_{k=0}^{2} a_k^{B_0}\tau^k + B_{\mathrm{off}}\big) + B_{0,\min}, \tag{11}$$

$$B'_{0,i}(\tau) = 1 + \mathrm{softplus}\big(\textstyle\sum_{k=0}^{2} a_k^{B'_0}\tau^k\big) + B'_{0,\min}, \tag{12}$$

the additive floors enforce $V_{0,i} > 0$, $B_{0,i} > 0$, $B'_{0,i} > 1$, and $\theta_{\mathrm{E}} > 0$. An offset initializes the initial bulk modulus near $B_{\mathrm{off}} = 100$ GPa. This parametrization gives the closed-form atomwise volume and analytic pressure integral

$$\hat{V}_i(\tau, P) = V_{0,i}(\tau)\, x_i^{-1/B'_{0,i}(\tau)}, \tag{13}$$

$$\int_0^P \hat{V}_i(\tau, p)\,\mathrm{d}p = \frac{V_{0,i}(\tau)\,B_{0,i}(\tau)}{B'_{0,i}(\tau) - 1}\big(x_i^{(B'_{0,i}(\tau)-1)/B'_{0,i}(\tau)} - 1\big), \tag{14}$$

with $x_i := 1 + B'_{0,i}(\tau)\,P/B_{0,i}(\tau)$. The head uses the Murnaghan-like form because its pressure integral Eq. (14) is closed-form and yields an analytic Gibbs free energy, whereas the Vinet form used for the quasi-harmonic fit [62] requires a numerical root solve. An ablation comparing the two forms gave comparable accuracy.

C. Multi-fidelity dataset

Source curation and splits. The Materials Project pool [1] is filtered to metastable entries within 0.2 eV/atom of the convex hull (123,424 structures) and each structure is relaxed with UMA (`UMA-S-1.1`, `omat` task) [41] to its static branch representative via ASE [63]. A representative subset of 26,226 structures is obtained by BIRCH clustering [43] of the UMA embeddings under a Euclidean radius threshold of 0.15, split 90/10 into 23,603 training and 2,623 validation representatives. The remaining structures are mapped to the nearest representative in embedding space, and the low-fidelity stream is sampled with inverse-cluster-size weights.

Low-fidelity quasi-harmonic labels. The low-fidelity signal is the quasi-harmonic approximation [18, 19]. At each cell volume, the Helmholtz free energy is the sum of the quantum harmonic oscillator free energies over the Brillouin-zone phonon branches. Imaginary-frequency modes, whose harmonic free energy is ill-defined, are excluded from this sum. Phonons are computed by finite displacements of 0.01 Å in an approximately isotropic supercell of minimum side length 12 Å with Phonopy [19]. Thermal properties are then evaluated from 0 K to $T_m + 500$ K in 5 K steps, where T_m is a predicted melting point [64] read from the Materials Project metadata. Repeating this over a 12-point isotropic strain grid ($\eta \in [-0.08, 0.03]$) traces the free energy against volume. A Legendre transform at fixed pressure and a Vinet equation-of-state fit [62] then yield the branch-

conditioned Gibbs free energy, with the zero-point contribution subtracted. The governing equations are given in Supplementary Note B. For training, one (T, P) point is sampled per structure, with T drawn uniformly over the same 0 to $T_m + 500$ K range used for the grid and $P \sim \mathcal{U}([0, 40]$ GPa). Samples with equation-of-state RMS residual above 5 meV/atom or a negative fitted bulk modulus are rejected, leaving 122,913 accepted surfaces (111,402 training / 11,511 validation).

High-fidelity molecular dynamics labels. The high-fidelity signal is the absolute Gibbs free energy from nonequilibrium MD, computed only on representatives. Each representative is relaxed with a fast, non-equivariant Orb-v3 potential (`orb-v3-conservative-inf-omat`) [42] and sampled at several independent state points (∼4 per material on average), each an independent draw of $T \sim \mathcal{U}([100$ K, $T_m + 500$ K$])$ and $P \sim \mathcal{U}([0, 40]$ GPa). The absolute Gibbs free energy is assembled as

$$G^{\mathrm{UMA}} = G^{\mathrm{Orb}}_{\mathrm{FL}} + \Delta G^{\mathrm{Orb}\to\mathrm{UMA}}, \tag{15}$$

a Frenkel–Ladd reference on the Orb-v3 Hamiltonian followed by an alchemical transfer to the equivariant UMA target at the same (T, P), obtained in the three steps below.

Step 1: NPT equilibration. At each state point, an approximately isotropic ∼1,000-atom supercell is equilibrated under NPT for ∼10 ps with a 1 fs timestep, an isotropic Martyna–Tobias–Klein barostat [65], and a Nosé–Hoover chain thermostat [66] (100 fs coupling).

Step 2: Frenkel–Ladd switching. The absolute Helmholtz free energy of the Orb-v3 crystal is obtained by Frenkel–Ladd thermodynamic integration [24] at the fixed equilibrated cell (NVT, Langevin thermostat). A mixed Hamiltonian $U(\boldsymbol{r}; \lambda) = (1 - \lambda)\, U(\boldsymbol{r}) + \lambda\, U_{\mathrm{harm}}(\boldsymbol{r})$ connects the physical potential to an Einstein crystal reference U_{harm}, whose per-atom spring constants $k_i = 3k_{\mathrm{B}}T/\langle \Delta r_i^2 \rangle$ are estimated from a 10 ps mean-squared-displacement run. The coupling is then driven along a smooth polynomial schedule $\lambda(t)$ with zero slope at both endpoints [26, 67]: forward over 25 ps, re-equilibrated in the harmonic reference for 5 ps, then backward over 25 ps. Symmetrizing the accumulated nonequilibrium work W^s over the forward and backward legs [25] cancels the leading dissipation,

$$\Delta F = \tfrac{1}{2}\big(\mathbb{E}_{\mathbb{P}_{0\to1}}[W^s] - \mathbb{E}_{\mathbb{P}_{1\to0}}[W^s]\big), \tag{16}$$

and the absolute branch-conditioned Gibbs free energy follows by adding the closed-form Einstein reference F_{harm}, the PV term at the equilibrated cell, and a center-of-mass correction F_{COM},

$$G_{\mathrm{FL}}(\boldsymbol{x}^\circ; T, P) = F_{\mathrm{harm}}(T) - \Delta F + PV + F_{\mathrm{COM}}(T). \tag{17}$$

The reference free energy and the center-of-mass correction are given in Supplementary Note B.

Step 3: Alchemical switching. To avoid a full Frenkel–Ladd calculation on the computationally expensive UMA potential, we transfer the absolute free energy from the Orb-v3 Hamiltonian to UMA at the same state $(\boldsymbol{x}^\circ; T, P)$. The mixed Hamiltonian $U(\boldsymbol{r}; \lambda) = (1 - \lambda)\, U^{\mathrm{Orb}} + \lambda\, U^{\mathrm{UMA}}$ is driven under NPT along the same schedule $\lambda(t)$: forward over 5 ps, re-equilibrated at UMA for 1 ps, then backward over 5 ps. Symmetrizing the two legs as in Eq. (16) yields the Gibbs free energy difference $\Delta G^{\mathrm{Orb}\to\mathrm{UMA}}$, and the absolute free energy transfers by the triangle relation of Eq. (15). Because Orb-v3 and UMA are trained on overlapping density functional theory data, this single alchemical step is a small perturbation. The general two-Hamiltonian derivation is given in Supplementary Note B.

Quality filtering and dataset. Each state point must pass four filters that reject unconverged or non-crystalline runs. The equilibrated cell edge must stay within a factor of two of its 0 K value (scale factor in $[0.5, 2.0]$), excluding collapse or sublimation. The atoms' time-averaged displacement from their lattice sites must not exceed 1.0 Å, removing configurations that melt or reconstruct. The NPT cell volume must reach equilibrium within 8,000 of its 10,000 steps, leaving a sufficient production window. The residual switching dissipation, half the sum of the forward and backward nonequilibrium work, must stay below 0.05 eV/atom for both the Frenkel–Ladd and alchemical switches. After filtering, 95,308 state points remain (85,774 training / 9,534 validation; 22,894 / 2,548 unique materials). Complete dataset sizes are tabulated in Supplementary Note C. Because T_m only sets the sampling ceiling, state points above it are retained as the metastable continuation of the crystalline branch.

D. Multi-fidelity training

Pre-training. Both training stages supervise the thermodynamic residual $G^\delta = G - U^\circ$ rather than the absolute G. Pre-training trains the model from scratch with a graph-normalized L_1 loss on the residual and its differentiated responses,

$$\mathcal{L}_{\mathrm{pre}} = \mathcal{L}_G + \lambda_V\, \mathcal{L}_V + \lambda_S\, \mathcal{L}_S + \lambda_{C_P}\, \mathcal{L}_{C_P} + \lambda_B\, \mathcal{L}_B + \lambda_\alpha\, \mathcal{L}_\alpha, \tag{18}$$

where $\mathcal{L}_G$ matches the QHA Gibbs residual, and the remaining terms match its volume, entropy, heat capacity, isothermal bulk modulus, and thermal expansion. Hence, the full first- and second-order (T, P) structure of G is supervised. The response labels come from the QHA equation-of-state fit and finite-temperature stencils. Extensive terms are normalized per atom. Per-sample response losses are capped to prevent rare pathological states from dominating optimization (loss weights in Supplementary Note C). Optimization uses Adam [68] (learning rate 6×10^{-4}, per-GPU batch size 32, gradient clipping at norm 10) over ∼125,000 optimization steps.

Mid-training. Mid-training initializes from the pre-

trained checkpoint and applies a LoRA update [50] to the linear layers of the thermodynamic backbone and head, with all other weights frozen. The update has rank 8, scaling factor 16, and no dropout, adding ∼44,000 trainable parameters. Mid-training activates the augmented Einstein basis of Eq. (9) by adding two anharmonic coefficients initialized to zero. The training target becomes the MD Gibbs residual, with the UMA-relaxed structure kept as input. Because the MD labels provide G and the equilibrated cell volume but no higher derivatives, the loss reduces to $\mathcal{L}_G$ matching the MD Gibbs residual, $\mathcal{L}_V$ matching the MD equilibrium volume ($\lambda_V = 0.05$), and a heat-capacity regularizer ($\lambda_{C_P}^{\mathrm{reg}} = 0.005$),

$$\mathcal{L}_{\mathrm{mid}} = \mathcal{L}_G + \lambda_V \mathcal{L}_V + \lambda_{C_P}^{\mathrm{reg}} \mathcal{L}_{C_P}^{\mathrm{reg}}. \tag{19}$$

The regularizer $\mathcal{L}_{C_P}^{\mathrm{reg}}$ is a smooth-L_1 penalty on the deviation of the predicted per-atom heat capacity from the frozen pre-trained value, with the tolerance set to 1 J K^{-1} (mol atoms)$^{-1}$. It ties the temperature dependence of the entropy and heat capacity to the well-behaved pre-trained model, while permitting an absolute anharmonic correction. Optimization uses Adam at learning rate 2×10^{-4}, and the training runs for approximately ∼87,000 steps.

E. Experimental alignment for ordered phases

NIST–JANAF heat-capacity and entropy benchmark. The heat-capacity and entropy accuracy of Fig. 4a is measured against the NIST–JANAF thermochemical tables [44], after restriction to solid phases. The noble gases and diatomic gases, whose JANAF records are gaseous element reference states rather than solids, are excluded. For each remaining species, we use data from 100 K to the first melting or vaporization transition, hard-capped at 4000 K. On the resulting set of 255 solids, we report the per-material mean absolute error between the recorded and predicted C_P and S.

Holland–Powell calibration. For ordered phases, the post-training stage calibrates against the `ds62` internally consistent thermodynamic assessment [45] using a predefined set of Si–Al–Mg–Ca–O minerals (the training/held-out phase split and the reaction set are tabulated in Supplementary Note D). The absolute Gibbs free energy inherits a PBE-level offset from the QHA and MD stages, so the objective matches only relative stabilities and response derivatives rather than absolute G. It is a smooth-L_1 objective combining phase-equilibrium, single-phase anchor, and reference-curve terms,

$$\begin{aligned}\mathcal{L}_{\mathrm{post}} = \mathcal{L}_{\mathrm{rxn}} + \lambda_{\mathrm{poly}}\mathcal{L}_{\mathrm{poly}} + \lambda_{B_0}\mathcal{L}_{B_0} + \lambda_{S_0}\mathcal{L}_{S_0} + \lambda_{V_0}\mathcal{L}_{V_0} \\ + \lambda_V\mathcal{L}_V + \lambda_{C_P}\mathcal{L}_{C_P} + \lambda_{\mathrm{shape}}\mathcal{L}_{\mathrm{shape}} + \lambda_{\mathrm{reg}}\mathcal{L}_{\mathrm{reg}}.\end{aligned} \tag{20}$$

The dominant terms $\mathcal{L}_{\mathrm{rxn}}$ and $\mathcal{L}_{\mathrm{poly}}$ match the experimentally assessed reaction and polymorph Gibbs differences $\Delta G(T, P)$ across the preset. The anchors $\mathcal{L}_{B_0}$, $\mathcal{L}_{S_0}$, and $\mathcal{L}_{V_0}$ match the `ds62` single-phase bulk modulus, standard entropy, and molar volume at 298.15 K and 1 bar through the differentiated responses of Eqs. (2) and (4). On the training phases, $\mathcal{L}_V$ and $\mathcal{L}_{C_P}$ match the `ds62` reference $V(T, P)$ and $C_P(T)$ curves, and $\mathcal{L}_{\mathrm{shape}}$ matches the phase-centered Gibbs surface $\hat{G}(T, P) - \hat{G}(T_0, P_0)$. The final term $\mathcal{L}_{\mathrm{reg}}$ is an L_2 penalty toward the pre-calibration weights. The residual bulk-modulus error against experiment is a further PBE bias shared by both earlier stages. This stage therefore fine-tunes the full thermodynamic adapter (the encoder-attached heads together with the equation-of-state parameters), allowing the equation of state parameters to adjust. The per-term weights are listed in Supplementary Note D. Optimization uses Adam (learning rate 1×10^{-4}, gradient clipping at norm 5) for 350 steps.

SiO_2 pressure–temperature phase diagram. The diagram (Fig. 4j–l) is constructed from the calibrated surface by evaluating $\hat{G}(T, P)$ for the five polymorphs on a grid in temperature (300–2400 K) and pressure (0–14 GPa) and taking the minimum-G phase at each grid point. Zero contours of pairwise Gibbs free energy differences define the phase boundaries.

F. Disordered solid solutions

For a disordered solid solution, we evaluate ordered branches across composition and calibrate them against a reference: bottom-up by fine-tuning on a higher-resolution computation, or top-down by iteratively reweighting sampled configurations onto a measured phase boundary. Both freeze the encoder backbone and update only a lightweight adapter: the bottom-up fit trains the LoRA layers, and the top-down calibration additionally updates the analytic thermodynamic heads.

Dilute solubility (bottom-up). The dilute solubility of Sc in fcc Al is determined by the branch free energies of pure Al, the dilute Sc-substituted solid solution, and the Al_3Sc ($L1_2$) compound. The per-Sc free energy of dissolving one Al_3Sc unit into a 256-site fcc Al is

$$\Delta G_{\mathrm{sol}} = G(\mathrm{Al}_{255}\mathrm{Sc}) - G(\mathrm{Al}_3\mathrm{Sc}) - 252\, G(\mathrm{Al}), \tag{21}$$

with $G(\mathrm{Al})$ the per-atom Gibbs free energy of pure fcc Al, and the dilute limit follows from the ideal-dilute equilibrium

$$\frac{x_{\mathrm{Sc}}}{1 - x_{\mathrm{Sc}}} = \exp\left(-\frac{\Delta G_{\mathrm{sol}}}{k_{\mathrm{B}}T}\right). \tag{22}$$

The bottom-up calibration fine-tunes the LoRA layers on a dedicated Al–Sc quasi-harmonic dataset so this dilute solubility approaches the quasi-harmonic UMA reference; the vibrational entropy of dissolution, which the 0 K static estimate omits, sets the solubility's order of magnitude [6, 49]. This dataset densely samples the fcc Al–Sc composition axis, a chemically disordered region under-

represented in pre-training. The dataset contains 491 substitutional supercells of up to 64 atoms spanning the full range from pure Al to pure Sc across 65 distinct Sc fractions. The orderings at each composition are selected for pair-correlation diversity. Each structure carries a UMA quasi-harmonic Gibbs free energy computed by the low-fidelity procedure (Supplementary Note B). Starting from the pre-trained checkpoint, the rank-8, scaling-16 adapter matches these labels with the Gibbs, volume, entropy, and heat-capacity terms of the pre-training loss Eq. (18) ($\lambda_V = 0.1$, $\lambda_S = 0.005$, $\lambda_{C_P} = 0.001$), optimized with Adam (learning rate 1×10^{-4}, batch size 16, gradient clipping at norm 10) for 100 epochs.

Semigrand canonical Monte Carlo. Equilibrium compositions within an ordered branch are sampled by semigrand canonical Monte Carlo (SGCMC) [53] on a fixed parent lattice, accelerated by well-tempered metadynamics [54]. At fixed species assignment $\boldsymbol{z}$, the atomic positions and cell are relaxed onto the branch representative, so vibrational and cell free energy contributions enter only through the surrogate G_θ, while identity-swap moves $z_i \to z'$ are accepted with probability

$$P_{\mathrm{acc}}(z_i \to z') = \min\bigl(1,\, e^{-\beta[\Delta G_\theta + \mu_{z_i} - \mu_{z'} + \Delta V_{\mathrm{bias}}]}\bigr), \quad (23)$$

where ΔG_θ is the change in the surrogate branch-conditioned Gibbs free energy of the swap and ΔV_{bias} is the accompanying change in the metadynamics bias. Full geometry relaxation is performed at every 10 steps when a configuration is stored as a training or reweighting sample (after a 40% burn-in). Intermediate steps score each swap with a single surrogate evaluation on the running configuration, which enhances the sampling efficiency. Each simulation runs 20,000 steps. Each step attempts a composition-changing identity move or a composition-conserving swap with equal probability. Metadynamics deposits a bias V_{bias} on the composition collective variable $\boldsymbol{c}(\boldsymbol{z})$ (0.05 eV Gaussians every 25 steps, bias factor $\gamma = 10$); in the long-time limit $V_{\mathrm{bias}} \to -(1 - 1/\gamma)\, G_{\mathrm{SGC}}(\boldsymbol{c}) + \mathrm{const}$, so the equilibrium concentration distribution

$$\pi_{\mathrm{SGC}}(\boldsymbol{c}) \propto \exp\bigl[-\beta\, G_{\mathrm{SGC}}(\boldsymbol{c};\, \boldsymbol{x}^\circ, T, P, \boldsymbol{\mu})\bigr] \quad (24)$$

is recovered by reweighting [69], and sweeping the chemical potentials $\boldsymbol{\mu}$ traces the branch-conditioned phase boundary in composition space. The parent-lattice assumptions are detailed in Supplementary Note B.

Differentiable reweighting. The top-down calibration matches a measured phase boundary, the Au–Pt solvus [51, 52], by differentiable reweighting [55] of the stored SGCMC and metadynamics samples. Differentiable reweighting re-scores the recorded configurations under updated adapter parameters without re-running the sampler. At each state point, the coexistence chemical potential is determined by equating the semigrand free energies of the two basins (the common-tangent condition). The adapter parameters then minimize a weighted error between the predicted and experimental binodal concentrations, regularized by an energy-drift penalty that keeps the importance weights valid and an L_2 penalty toward the pre-calibration weights. The binodal concentrations depend on the parameters through the reweighted samples and through the implicitly differentiated coexistence root, which makes this regression loss differentiable in the adapter parameters, and optimization uses Adam (learning rate 5×10^{-6}, gradient clipping at norm 5) for 160 gradient steps per cycle. We repeat the calibration cycle five times: within each cycle the stored samples are reused by reweighting, and between cycles the reference simulation is re-run at the updated parameters, warm-starting its metadynamics bias from the previous cycle to speed convergence. The regression loss, the reweighting estimator, the coexistence root, and its implicit gradient are given in Supplementary Note B.

CODE AVAILABILITY

The dataset and code to reproduce the results will be made publicly available upon acceptance of this manuscript.

ACKNOWLEDGMENTS

The authors thank X. Fu for the discussions that inspired the conception of this project, and M. Schebek, P. Holderrieth, Y. Du, M. Cheng, K. Sheriff, Z. W. Ulissi, M. Gao, C. L. Zitnick, B. M. Wood, and others from the FAIR Chemistry team for helpful scientific disscussions. J.N. acknowledges support from the Mathworks Fellowship. B.D. would like to acknowledge the funding support by Shell Inc. X.D. acknowledges funding from Amazon as part of the MIT Climate and Sustainability Consortium (MCSC). This research used resources of the National Energy Research Scientific Computing Center (NERSC), a Department of Energy Office of Science User Facility using NERSC award ALCC-ERCAP-m5068.

COMPETING INTERESTS

The authors declare no competing interests.

[1] 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 Mater. **1**, 011002 (2013).
[2] S. Kirklin, J. E. Saal, B. Meredig, A. Thompson, J. W. Doak, M. Aykol, S. Rühl, and C. Wolverton, The Open Quantum Materials Database (OQMD): assessing the accuracy of DFT formation energies, npj Comput. Mater. **1**, 15010 (2015).
[3] W. Sun, C. J. Bartel, E. Arca, S. R. Bauers, B. E. Matthews, B. Orvañanos, B.-R. Chen, M. F. Toney, L. T. Schelhas, W. Tumas, J. Tate, A. Zakutayev, S. Lany, A. M. Holder, and G. Ceder, A map of the inorganic ternary metal nitrides, Nat. Mater. **18**, 732 (2019).
[4] K. Tolborg, J. Klarbring, A. M. Ganose, and A. Walsh, Free energy predictions for crystal stability and synthesisability, Digital Discovery **1**, 586 (2022).
[5] G. Garbulsky and G. Ceder, Contribution of the vibrational free energy to phase stability in substitutional alloys: Methods and trends, Phys. Rev. B **53**, 8993 (1996).
[6] A. Van De Walle and G. Ceder, The effect of lattice vibrations on substitutional alloy thermodynamics, Rev. Mod. Phys. **74**, 11 (2002).
[7] C. J. Bartel, A. W. Weimer, S. Lany, C. B. Musgrave, and A. M. Holder, The role of decomposition reactions in assessing first-principles predictions of solid stability, npj Comput. Mater. **5**, 4 (2019).
[8] C. J. Bartel, A. Trewartha, Q. Wang, A. Dunn, A. Jain, and G. Ceder, A critical examination of compound stability predictions from machine-learned formation energies, npj Comput. Mater. **6**, 97 (2020).
[9] W. Sun, S. T. Dacek, S. P. Ong, G. Hautier, A. Jain, W. D. Richards, A. C. Gamst, K. A. Persson, and G. Ceder, The thermodynamic scale of inorganic crystalline metastability, Sci. Adv. **2**, e1600225 (2016).
[10] M. Aykol, S. S. Dwaraknath, W. Sun, and K. A. Persson, Thermodynamic limit for synthesis of metastable inorganic materials, Sci. Adv. **4**, eaaq0148 (2018).
[11] J. M. Sanchez, F. Ducastelle, and D. Gratias, Generalized cluster description of multicomponent systems, Physica A Stat. Mech. Appl. **128**, 334 (1984).
[12] J. W. D. Connolly and A. R. Williams, Density-functional theory applied to phase transformations in transition-metal alloys, Phys. Rev. B **27**, 5169 (1983).
[13] P. Zhong, T. Chen, L. Barroso-Luque, F. Xie, and G. Ceder, An $\ell_0\ell_2$-norm regularized regression model for construction of robust cluster expansions in multicomponent systems, Phys. Rev. B **106**, 024203 (2022).
[14] M. Asta, D. de Fontaine, M. van Schilfgaarde, M. Sluiter, and M. Methfessel, First-principles phase-stability study of fcc alloys in the Ti-Al system, Phys. Rev. B **46**, 5055 (1992).
[15] D. De Fontaine, Cluster approach to order-disorder transformations in alloys, in *Solid state physics*, Vol. 47 (Elsevier, 1994) pp. 33–176.
[16] G. Ceder, A. Van der Ven, C. Marianetti, and D. Morgan, First-principles alloy theory in oxides, Model. Simul. Mater. Sci. Eng. **8**, 311 (2000).
[17] A. van de Walle and G. Ceder, Automating first-principles phase diagram calculations, J. Phase Equilib. **23**, 348 (2002).
[18] S. Baroni, S. De Gironcoli, A. Dal Corso, and P. Giannozzi, Phonons and related crystal properties from density-functional perturbation theory, Rev. Mod. Phys. **73**, 515 (2001).
[19] A. Togo and I. Tanaka, First principles phonon calculations in materials science, Scr. Mater. **108**, 1 (2015).
[20] P. Souvatzis, O. Eriksson, M. Katsnelson, and S. Rudin, Entropy driven stabilization of energetically unstable crystal structures explained from first principles theory, Phys. Rev. Lett. **100**, 095901 (2008).
[21] O. Hellman, P. Steneteg, I. A. Abrikosov, and S. I. Simak, Temperature dependent effective potential method for accurate free energy calculations of solids, Phys. Rev. B **87**, 104111 (2013).
[22] J. C. Thomas and A. V. d. Ven, Finite-temperature properties of strongly anharmonic and mechanically unstable crystal phases from first principles, Phys. Rev. B **88**, 214111 (2013).
[23] A. Van De Walle, Q. Hong, S. Kadkhodaei, and R. Sun, The free energy of mechanically unstable phases, Nat. Commun. **6**, 7559 (2015).
[24] D. Frenkel and A. J. Ladd, New Monte Carlo method to compute the free energy of arbitrary solids. application to the fcc and hcp phases of hard spheres, J. Chem. Phys. **81**, 3188 (1984).
[25] M. de Koning, A. Antonelli, and S. Yip, Optimized free-energy evaluation using a single reversible-scaling simulation, Phys. Rev. Lett. **83**, 3973 (1999).
[26] J. Nam, J. Peng, and R. Gómez-Bombarelli, Interpolation and differentiation of alchemical degrees of freedom in machine learning interatomic potentials, Nat. Commun. **16**, 4350 (2025).
[27] S. Menon, Y. Lysogorskiy, J. Rogal, and R. Drautz, Automated free-energy calculation from atomistic simulations, Phys. Rev. Mater. **5**, 103801 (2021).
[28] B. Cheng and M. Ceriotti, Computing the absolute Gibbs free energy in atomistic simulations: Applications to defects in solids, Phys. Rev. B **97**, 054102 (2018).
[29] V. L. Deringer, M. A. Caro, and G. Csányi, Machine learning interatomic potentials as emerging tools for materials science, Adv. Mater. **31**, 1902765 (2019).
[30] J. Behler and M. Parrinello, Generalized neural-network representation of high-dimensional potential-energy surfaces, Phys. Rev. Lett. **98**, 146401 (2007).
[31] A. P. Bartók, M. C. Payne, R. Kondor, and G. Csányi, Gaussian approximation potentials: The accuracy of quantum mechanics, without the electrons, Phys. Rev. Lett. **104**, 136403 (2010).
[32] C. Chen and S. P. Ong, A universal graph deep learning interatomic potential for the periodic table, Nat. Comput. Sci. **2**, 718 (2022).
[33] 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 modelling, Nat. Mach. Intell. **5**, 1031 (2023).
[34] I. Batatia, P. Benner, Y. Chiang, A. M. Elena, D. P. Kovács, J. Riebesell, X. R. Advincula, M. Asta, M. Avaylon, W. J. Baldwin, *et al.*, A foundation model for atomistic materials chemistry, J. Chem. Phys. **163**, 184110 (2025).
[35] A. Mazitov, F. Bigi, M. Kellner, P. Pegolo, D. Tisi, G. Fraux, S. Pozdnyakov, P. Loche, and M. Ceriotti, PET-MAD as a lightweight universal interatomic potential for advanced materials modeling, Nat. Commun. **16**, 10653 (2025).
[36] L. Kaufman and H. Bernstein, *Computer Calculation of Phase Diagrams: With Special Reference to Refractory Metals* (Academic Press, New York, 1970).

[37] H. Lukas, S. G. Fries, and B. Sundman, *Computational thermodynamics: the Calphad method* (Cambridge University Press, 2007).

[38] C. J. Bartel, S. L. Millican, A. M. Deml, J. R. Rumptz, W. Tumas, A. W. Weimer, S. Lany, V. Stevanović, C. B. Musgrave, and A. M. Holder, Physical descriptor for the Gibbs energy of inorganic crystalline solids and temperature-dependent materials chemistry, Nat. Commun. **9**, 4168 (2018).

[39] S. Srinivasan, R. Batra, D. Luo, T. Loeffler, S. Manna, H. Chan, L. Yang, W. Yang, J. Wen, P. Darancet, and S. K. R. S. Sankaranarayanan, Machine learning the metastable phase diagram of covalently bonded carbon, Nat. Commun. **13**, 3251 (2022).

[40] F. D. Murnaghan, The compressibility of media under extreme pressures, Proc. Natl. Acad. Sci. U. S. A. **30**, 244 (1944).

[41] B. M. Wood, M. Dzamba, X. Fu, M. Gao, M. Shuaibi, L. Barroso-Luque, K. Abdelmaqsoud, V. Gharakhanyan, J. R. Kitchin, D. S. Levine, K. Michel, A. Sriram, T. Cohen, A. Das, A. Rizvi, S. J. Sahoo, Z. W. Ulissi, and C. L. Zitnick, UMA: A family of universal models for atoms, in *The Thirty-ninth Annual Conference on Neural Information Processing Systems* (2025).

[42] B. Rhodes, S. Vandenhaute, V. Šimkus, J. Gin, J. Godwin, T. Duignan, and M. Neumann, Orb-v3: atomistic simulation at scale (2025), arXiv:2504.06231 [cond-mat.mtrl-sci].

[43] T. Zhang, R. Ramakrishnan, and M. Livny, BIRCH: an efficient data clustering method for very large databases, SIGMOD Rec. **25**, 103 (1996).

[44] M. W. Chase, *NIST-JANAF Thermochemical Tables*, 4th ed., Journal of Physical Chemistry Reference Data, Monograph No. 9 (American Chemical Society and American Institute of Physics for the National Institute of Standards and Technology, 1998).

[45] T. J. B. Holland and R. Powell, An improved and extended internally consistent thermodynamic dataset for phases of petrological interest, involving a new equation of state for solids, J. Metamorph. Geol. **29**, 333 (2011).

[46] A. T. Dinsdale, SGTE data for pure elements, Calphad **15**, 317 (1991).

[47] P. Richet, R. A. Robie, and B. S. Hemingway, Thermodynamic properties of wollastonite, pseudowollastonite and CaSiO3 glass and liquid, Eur. J. Mineral. **3**, 475 (1991).

[48] V. Swamy, S. K. Saxena, B. Sundman, and J. Zhang, A thermodynamic assessment of silica phase diagram, J. Geophys. Res. Solid Earth **99**, 11787 (1994).

[49] V. Ozoliņš and M. Asta, Large vibrational effects upon calculated phase boundaries in Al-Sc, Phys. Rev. Lett. **86**, 448 (2001).

[50] E. J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, and W. Chen, LoRA: Low-rank adaptation of large language models, in *International Conference on Learning Representations* (2022).

[51] A. S. Darling, R. A. Mintern, and J. C. Chaston, The gold-platinum system, J. Inst. Met. **81**, 125 (1952).

[52] A. Münster and K. Sagel, Separation curve and critical point of the system gold-platinum, Z. Phys. Chem. **23**, 415 (1960), in German.

[53] B. Sadigh, P. Erhart, A. Stukowski, A. Caro, E. Martinez, and L. Zepeda-Ruiz, Scalable parallel Monte Carlo algorithm for atomistic simulations of precipitation in alloys, Phys. Rev. B **85**, 184203 (2012).

[54] A. Barducci, G. Bussi, and M. Parrinello, Well-tempered metadynamics: a smoothly converging and tunable free-energy method, Phys. Rev. Lett. **100**, 020603 (2008).

[55] S. Thaler and J. Zavadlav, Learning neural network potentials from experimental data via Differentiable Trajectory Reweighting, Nat. Commun. **12**, 6884 (2021).

[56] F. Zhou, T. Maxisch, and G. Ceder, Configurational electronic entropy and the phase diagram of mixed-valence oxides: The case of Li_xFePO_4, Phys. Rev. Lett. **97**, 155704 (2006).

[57] M. Hillert and M. Jarl, A model for alloying in ferromagnetic metals, Calphad **2**, 227 (1978).

[58] I. Mosquera-Lois, S. R. Kavanagh, J. Klarbring, K. Tolborg, and A. Walsh, Imperfections are not 0 K: free energy of point defects in crystals, Chem. Soc. Rev. **52**, 5812 (2023).

[59] X. Du, J. K. Damewood, J. R. Lunger, R. Millan, B. Yildiz, L. Li, and R. Gómez-Bombarelli, Machine-learning-accelerated simulations to enable automatic surface reconstruction, Nat. Comput. Sci. **3**, 1034 (2023).

[60] R. Paula Leite, R. Freitas, R. Azevedo, and M. de Koning, The Uhlenbeck-Ford model: Exact virial coefficients and application as a reference system in fluid-phase free-energy calculations, J. Chem. Phys. **145**, 194101 (2016).

[61] K. Schütt, O. Unke, and M. Gastegger, Equivariant message passing for the prediction of tensorial properties and molecular spectra, in *International conference on machine learning* (PMLR, 2021) pp. 9377–9388.

[62] P. Vinet, J. R. Smith, J. Ferrante, and J. H. Rose, Temperature effects on the universal equation of state of solids, Phys. Rev. B **35**, 1945 (1987).

[63] A. Hjorth Larsen, J. Jørgen Mortensen, J. Blomqvist, I. E. Castelli, R. Christensen, M. Dułak, J. Friis, M. N. Groves, B. Hammer, C. Hargus, *et al.*, The atomic simulation environment—a Python library for working with atoms, J. Phys.: Condens. Matter **29**, 273002 (2017).

[64] Q.-J. Hong, S. V. Ushakov, A. van de Walle, and A. Navrotsky, Melting temperature prediction using a graph neural network model: From ancient minerals to new materials, Proc. Natl. Acad. Sci. U. S. A. **119**, e2209630119 (2022).

[65] G. J. Martyna, D. J. Tobias, and M. L. Klein, Constant pressure molecular dynamics algorithms, J. Chem. Phys. **101**, 4177 (1994).

[66] G. J. Martyna, M. L. Klein, and M. Tuckerman, Nosé–Hoover chains: The canonical ensemble via continuous dynamics, J. Chem. Phys. **97**, 2635 (1992).

[67] M. de Koning and A. Antonelli, Einstein crystal as a reference system in free energy estimation using adiabatic switching, Phys. Rev. E **53**, 465 (1996).

[68] D. P. Kingma and J. Ba, Adam: A method for stochastic optimization (2014), arXiv:1412.6980 [cs.LG].

[69] P. Tiwary and M. Parrinello, A time-independent free energy estimator for metadynamics, J. Phys. Chem. B **119**, 736 (2015).

Supplementary Information for Universal Thermodynamic Interatomic Potentials for Crystalline Materials

Juno Nam,[1] Bowen Deng,[1] Xiaochen Du,[1] Luis Barroso-Luque,[2]
Benjamin Kurt Miller,[2,*] and Rafael Gómez-Bombarelli[1,†]

[1] *Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*
[2] *Fundamental AI Research, Meta, San Francisco, CA 94105, USA*

(Dated: August 17, 2026)

CONTENTS

A. Theoretical Formalism and Statistical Basis 2
1. Notation 2
2. Assumptions 2
3. Branch-Conditioned Gibbs Free Energy 3
4. Statistical Basis of the Learned Free Energy 3

B. Derivations of the Free Energy and Sampling Methods 4
1. Quasi-Harmonic Approximation 4
2. Frenkel–Ladd Switching 4
3. Alchemical Switching 5
4. Semigrand Canonical Monte Carlo 6
5. Differentiable Reweighting 7

C. Training Details 8
1. Multi-Fidelity Training 8
2. Analytic Head Constants and Numerical Evaluation 8

D. Holland–Powell Calibration and Benchmark Details 8

References 10

* bkmi@meta.com

† rafagb@mit.edu

A. THEORETICAL FORMALISM AND STATISTICAL BASIS

Table S1. **Summary of the notation.**

Symbol	Description
N	Number of atoms in a crystal
$\boldsymbol{x}$	Crystal structure $(\boldsymbol{L}, \boldsymbol{z}, [\boldsymbol{r}])$
$\boldsymbol{x}^\circ$	Branch representative (relaxed structure)
$\boldsymbol{L}$	Cell matrix ($\mathrm{GL}^+(3,\mathbb{R})$)
V	Cell volume ($\det \boldsymbol{L} > 0$)
$\boldsymbol{z}$	Atomic species in a crystal ($\mathcal{A}^N$)
$\boldsymbol{m}$	Atomic masses ($\mathbb{R}^N_{>0}$)
$\boldsymbol{r}$	Cartesian atomic positions ($\mathbb{R}^{3\times N}$)
$\boldsymbol{c}$	Composition on a parent lattice
O°	Relaxed/reference observable O
O^*	Equilibrium (ensemble) average of O
$\hat{O}$	Predicted observable O
P	External pressure
T	External temperature
β	Inverse temperature $1/k_\mathrm{B}T$
U	Potential energy (of a configuration)
U°	Static reference energy $U(\boldsymbol{x}^\circ)$
F	Helmholtz free energy (NVT)
G	Gibbs free energy (NPT)
G^δ	Thermal residual ($G - U^\circ$)
C_P	Isobaric heat capacity
V_0	Equilibrium volume at $P = 0$
B_0	Isothermal bulk modulus at $P = 0$
α	Volumetric thermal expansion coefficient
$S(298\ \mathrm{K})$	Standard entropy at 298 K
λ	Switching (coupling) coordinate ($[0,1]$)
γ	Well-tempered metadynamics bias factor
μ_z	Chemical potential of species z
$\Delta\mu$	Semigrand chemical potential difference
θ	TIP parameters

1. Notation

Table S1 summarizes the notation used throughout this work. A periodic crystal of N atoms is represented by the triple

$$\boldsymbol{x} = (\boldsymbol{L}, \boldsymbol{z}, [\boldsymbol{r}]) \in \mathcal{X}_N, \tag{S1}$$

where $\boldsymbol{L} \in \mathrm{GL}^+(3,\mathbb{R})$ is the cell matrix whose columns are the lattice vectors. The species $\boldsymbol{z} = (z_1, \ldots, z_N) \in \mathcal{A}^N$ are drawn from the periodic table $\mathcal{A}$ (a finite set of chemical elements). The symbol $[\boldsymbol{r}]$ denotes the periodic Cartesian configuration of the N atoms, i.e., the equivalence class of $\boldsymbol{r} \in \mathbb{R}^{3\times N}$ modulo lattice translations. We fix the residual O(3) gauge associated with global rotations and reflections by restricting $\boldsymbol{L}$ to upper-triangular form with positive diagonal entries.

The cell matrix induces the Bravais lattice $\boldsymbol{L}\mathbb{Z}^3$ with fundamental-cell volume $V := \det \boldsymbol{L} > 0$, and each periodic position $[\boldsymbol{r}_i]$ has a unique representative in that domain.

For a structure-dependent observable $O(\boldsymbol{x})$, we write $O^\circ := O(\boldsymbol{x}^\circ)$ for its zero-temperature value at the branch representative, and

$$O^* := \mathbb{E}_{\boldsymbol{x}\sim\pi_{NPT}}[O(\boldsymbol{x})] \tag{S2}$$

for its expectation under the branch-conditioned NPT Gibbs measure $\pi_{NPT}(\boldsymbol{x}; \boldsymbol{x}^\circ, T, P) \propto e^{-\beta\,(U(\boldsymbol{x})+PV)}$ supported on $\mathcal{B}(\boldsymbol{x}^\circ)$ (the branch, defined in the Assumptions below), formally normalized in Eq. (S9).

2. Assumptions

The thermodynamic interatomic potential (TIP) models the Gibbs free energy of an ordered bulk crystal branch indexed by the reference structure $\boldsymbol{x}^\circ$. To make this precise, let

$$\mathcal{X}_N^\circ \subset \mathcal{X}_N \tag{S3}$$

denote the subset of reference (relaxed) configurations. We then introduce a deterministic, idempotent *branch map*

$$\mathfrak{R}_\mathrm{sym} : \mathcal{X}_N \to \mathcal{X}_N^\circ, \qquad \mathfrak{R}_\mathrm{sym} \circ \mathfrak{R}_\mathrm{sym} = \mathfrak{R}_\mathrm{sym}, \tag{S4}$$

that maps each periodic structure $\boldsymbol{x} \in \mathcal{X}_N$ to its branch representative $\mathfrak{R}_\mathrm{sym}(\boldsymbol{x}) \in \mathcal{X}_N^\circ$. In practice, $\mathfrak{R}_\mathrm{sym}$ is implemented by minimizing the base potential energy U (the UMA potential for TIP[UMA]) while constraining the reference space group, so that $\mathfrak{R}_\mathrm{sym}(\boldsymbol{x})$ is a stationary point

$$\nabla_{\boldsymbol{r},\boldsymbol{L}} U\big|_{\mathfrak{R}_\mathrm{sym}(\boldsymbol{x})} \approx 0 \tag{S5}$$

under the imposed constraint. For ordinary phases, the constraint is inactive and $\mathfrak{R}_\mathrm{sym}$ coincides with unconstrained 0 K relaxation. For dynamically stabilized phases such as bcc Ti above the martensitic transition, the constraint is essential because the static high-symmetry archetype is not itself a mechanical minimum. The preimage

$$\mathcal{B}(\boldsymbol{x}^\circ) := \mathfrak{R}_\mathrm{sym}^{-1}(\boldsymbol{x}^\circ) = \{\boldsymbol{x} \in \mathcal{X}_N : \mathfrak{R}_\mathrm{sym}(\boldsymbol{x}) = \boldsymbol{x}^\circ\} \tag{S6}$$

is the *branch* (or basin) associated with $\boldsymbol{x}^\circ$, and the family $\{\mathcal{B}(\boldsymbol{x}^\circ)\}_{\boldsymbol{x}^\circ\in\mathcal{X}_N^\circ}$ partitions $\mathcal{X}_N$ by construction.

The formulation uses three assumptions:

(A1) *Branch representability.* The reference structure $\boldsymbol{x}^\circ \in \mathcal{X}_N^\circ$ is a fixed point of $\mathfrak{R}_\mathrm{sym}$, i.e., $\mathfrak{R}_\mathrm{sym}(\boldsymbol{x}^\circ) = \boldsymbol{x}^\circ$, and a stationary point of U under the symme-

try constraint defining $\mathfrak{R}_{\rm sym}$, but it need not be an unconstrained local minimum.

(A2) *Finite-(T,P) branch stability.* At the queried (T,P), the branch $\mathcal{B}(\boldsymbol{x}^\circ)$ is a stable or metastable NPT basin: the system equilibrates within $\mathcal{B}(\boldsymbol{x}^\circ)$ on the simulation timescale, while escape to other branches is exponentially rare over that timescale.

(A3) *Fixed-branch scope.* A TIP models the *branch-conditioned* Gibbs free energy $G(\boldsymbol{x}^\circ;T,P)$ obtained by restricting the NPT partition function to $\boldsymbol{x}\in\mathcal{B}(\boldsymbol{x}^\circ)$. Competing branches are compared after evaluating G for their respective representatives.

A1 treats $\boldsymbol{x}^\circ$ as a branch label rather than requiring an unconstrained mechanical minimum, admitting dynamically stabilized phases. A2 makes the conditioned ensemble well-defined, so that long-time molecular dynamics (MD) initialized in $\mathcal{B}(\boldsymbol{x}^\circ)$ samples the branch ergodically. A3 limits the target to a single branch, and therefore requires an unambiguous symmetry constraint for $\mathfrak{R}_{\rm sym}$.

3. Branch-Conditioned Gibbs Free Energy

Following the fixed-ordering coarse-graining framework of vibrational alloy thermodynamics [S1], a TIP targets the Gibbs free energy of a crystal conditioned on its branch label $\boldsymbol{x}^\circ\in\mathcal{X}_N^\circ$, rather than the free energy summed over all branches.

At fixed cell $\boldsymbol{L}$, the branch-conditioned canonical (NVT) partition function is

$$Z(\boldsymbol{x}^\circ,\boldsymbol{L};T):=\frac{1}{\sigma(\boldsymbol{z})\,\Lambda_{\boldsymbol{m}}^{3N}}\int_{\mathcal{B}_{\boldsymbol{L}}(\boldsymbol{x}^\circ)}\exp\big[-\beta U(\boldsymbol{x})\big]\,\mathrm{d}\boldsymbol{r}, \tag{S7}$$

where $\beta:=1/k_{\rm B}T$, $\mathcal{B}_{\boldsymbol{L}}(\boldsymbol{x}^\circ):=\mathcal{B}(\boldsymbol{x}^\circ)\cap\{\boldsymbol{x}:\mathrm{cell}(\boldsymbol{x})=\boldsymbol{L}\}$ is the slice of the branch at fixed cell, $\sigma(\boldsymbol{z})$ is the indistinguishability factor for permutations of identical species, and $\Lambda_{\boldsymbol{m}}^{3N}:=\prod_{i=1}^{N}(h^2/2\pi m_i k_{\rm B}T)^{3/2}$ is the species-resolved thermal de Broglie prefactor obtained by integrating out atomic momenta. The associated branch-conditioned Helmholtz free energy is

$$F(\boldsymbol{x}^\circ,\boldsymbol{L};T):=-k_{\rm B}T\ln Z(\boldsymbol{x}^\circ,\boldsymbol{L};T). \tag{S8}$$

The branch-conditioned isothermal-isobaric (NPT) partition function is the Laplace transform of Z over the cell degrees of freedom at fixed pressure,

$$\Delta(\boldsymbol{x}^\circ;T,P):=\int e^{-\beta PV}\,Z(\boldsymbol{x}^\circ,\boldsymbol{L};T)\,\mathrm{d}\boldsymbol{L}, \tag{S9}$$

where the integral runs over cell matrices $\boldsymbol{L}$ consistent with the branch, in the upper-triangular gauge introduced in the Notation subsection, and $V=\det\boldsymbol{L}$. Here $\mathrm{d}\boldsymbol{L}$ denotes a fixed reference measure on cell matrices. Its choice and the discrete multiplicity of equivalent lattice bases enter G only through a subextensive normalization that is negligible per atom and cancels in the branch differences and (T,P)-derivatives used here. The corresponding *branch-conditioned Gibbs free energy* is

$$G(\boldsymbol{x}^\circ;T,P):=-k_{\rm B}T\ln\Delta(\boldsymbol{x}^\circ;T,P). \tag{S10}$$

By construction, $G(\boldsymbol{x}^\circ;T,P)$ is a branchwise quantity, obtained by marginalizing over the vibrational and cell-shape fluctuations within a single branch $\mathcal{B}(\boldsymbol{x}^\circ)$. Summing over the disjoint branches gives the free energy over the full configuration space,

$$G(T,P)=-k_{\rm B}T\ln\sum_{\boldsymbol{x}^\circ\in\mathcal{X}_N^\circ}\exp\big[-\beta\,G(\boldsymbol{x}^\circ;T,P)\big]. \tag{S11}$$

For a chemically disordered system, this branch sum is evaluated by sampling the disjoint branches on a parent lattice, as in the main-text solid solution examples. We carry this out by configurational sampling on a parent lattice, as for the disordered solid solutions of the main text.

Equilibrium observables within the branch are obtained as standard thermodynamic derivatives of $G(\boldsymbol{x}^\circ;T,P)$,

$$S=-\left(\frac{\partial G}{\partial T}\right)_P,\qquad V=\left(\frac{\partial G}{\partial P}\right)_T. \tag{S12}$$

For modeling purposes, we split the target into a static reference and a thermodynamic residual,

$$G(\boldsymbol{x}^\circ;T,P)=U^\circ+G^\delta(\boldsymbol{x}^\circ;T,P), \tag{S13}$$

where $U^\circ:=U(\boldsymbol{x}^\circ)$ is the zero-temperature potential energy of the branch representative and G^δ is the thermodynamic residual that contains all finite-(T,P) contributions.

4. Statistical Basis of the Learned Free Energy

The quasi-harmonic approximation (QHA) and nonequilibrium MD simulation free energies use different nuclear statistics. After removal of the zero-point term, the QHA label in Eq. (S16) is the quantum thermal-occupation free energy $\sum_{\boldsymbol{q}}w_{\boldsymbol{q}}\sum_\nu k_{\rm B}T\ln(1-e^{-\beta\hbar\omega_{\boldsymbol{q}\nu}})$, whereas the Frenkel–Ladd free energy in Eq. (S27) is classical: it is built on the classical Einstein reference of Eq. (S26) and sampled by classical molecular dynamics. The two limits agree as $\beta\hbar\omega\to0$, where $k_{\rm B}T\ln(1-e^{-\beta\hbar\omega})\to k_{\rm B}T\ln(\beta\hbar\omega)$ reproduces the classical form of Eq. (S26). At low temperatures, they differ: the quantum heat capacity approaches zero, whereas the classical harmonic heat capacity approaches $3Nk_{\rm B}$. The analytic head combines these targets by retaining the quantum Einstein term $3\,k_{\rm B}T\ln(1-e^{-\theta_{\rm E}/T})$, while fitting MD-derived anharmonic corrections over the sampled temperature range. Mid-trained TIP[UMA] there-

fore retains nuclear quantum effects (except the zero-point offset) at the harmonic level and incorporates classical anharmonic corrections from MD.

B. DERIVATIONS OF THE FREE ENERGY AND SAMPLING METHODS

1. Quasi-Harmonic Approximation

The QHA [S2, S3] provides the low-fidelity approximation for $G(\boldsymbol{x}^\circ; T, P)$ that retains volume-dependent vibrational physics while neglecting anharmonic mode coupling. We use it as the low-fidelity training signal for TIP[UMA].

For a fixed cell $\boldsymbol{L}$, we expand the potential energy around the cell-constrained relaxed positions $\boldsymbol{r}^\circ(\boldsymbol{L}) := \arg\min_{\boldsymbol{r}} U(\boldsymbol{r}; \boldsymbol{L})$ to second order,

$$U(\boldsymbol{r}; \boldsymbol{L}) \approx U^\circ(\boldsymbol{L}) + \frac{1}{2}\,\Delta\boldsymbol{r}^\top \Phi(\boldsymbol{L})\,\Delta\boldsymbol{r}, \tag{S14}$$

where $\Delta\boldsymbol{r} := \boldsymbol{r} - \boldsymbol{r}^\circ(\boldsymbol{L})$, $U^\circ(\boldsymbol{L}) := U(\boldsymbol{r}^\circ(\boldsymbol{L}); \boldsymbol{L})$ is the cell-constrained static energy, and $\Phi(\boldsymbol{L})$ is the interatomic force-constant matrix at $\boldsymbol{r}^\circ(\boldsymbol{L})$. By lattice periodicity the mass-weighted force constants block-diagonalize over the Brillouin zone: at each wavevector $\boldsymbol{q}$ the dynamical matrix $\widetilde{\Phi}(\boldsymbol{q}; \boldsymbol{L}) \in \mathbb{C}^{3n\times 3n}$ (n atoms per unit cell) has eigenvalues $\omega^2_{\boldsymbol{q}\nu}(\boldsymbol{L})$ indexed by the branch $\nu = 1, \ldots, 3n$.

Within this harmonic ansatz at fixed cell, the canonical partition function factorizes into independent quantum harmonic oscillators, giving the closed-form Helmholtz free energy

$$F_{\mathrm{QHA}}(\boldsymbol{L}, T) = U^\circ(\boldsymbol{L}) + F_{\mathrm{vib}}(\boldsymbol{L}, T), \tag{S15}$$

with the vibrational contribution

$$F_{\mathrm{vib}}(\boldsymbol{L}, T) = \sum_{\boldsymbol{q}} w_{\boldsymbol{q}} \sum_{\nu=1}^{3n} \left[\frac{\hbar\,\omega_{\boldsymbol{q}\nu}(\boldsymbol{L})}{2} + k_{\mathrm{B}} T \ln\!\left(1 - e^{-\beta\hbar\omega_{\boldsymbol{q}\nu}(\boldsymbol{L})}\right)\right], \tag{S16}$$

per unit cell, where the sum runs over a regular mesh of wavevectors $\boldsymbol{q}$ in the first Brillouin zone with weights normalized to $\sum_{\boldsymbol{q}} w_{\boldsymbol{q}} = 1$. The first term is the zero-point energy and the second is the thermal occupation contribution. The cell dependence of $\omega_{\boldsymbol{q}\nu}(\boldsymbol{L})$ accounts for the thermal expansion and Grüneisen-type effects that a fixed-cell harmonic approximation omits. The force constants are obtained from finite displacements in the relaxed supercell and Fourier-interpolated onto the Brillouin-zone mesh with Phonopy [S3]. We discard modes below a small near-zero cutoff, including the three acoustic Γ-point modes and small imaginary modes caused by numerical noise, to avoid singular occupation terms.

The branch-conditioned QHA Gibbs free energy is obtained by minimizing $F(L, T) + PV$ over cells consistent with the branch,

$$G_{\mathrm{QHA}}(\boldsymbol{x}^\circ; T, P) := \min_{\boldsymbol{L}} \left[F_{\mathrm{QHA}}(\boldsymbol{L}, T) + PV\right], \tag{S17}$$

where the minimization runs over cells consistent with the branch $\mathcal{B}(\boldsymbol{x}^\circ)$. In practice we restrict the search to an isotropic-strain family $\boldsymbol{L}(\eta) := (1+\eta)\,\boldsymbol{L}^\circ$ around the branch representative, evaluate F_{QHA} on a strain grid, and fit the resulting $F_{\mathrm{QHA}}(V, T)$ with a Vinet equation of state [S4] to perform the minimization analytically. Subtracting the static reference yields the QHA approximation to the thermodynamic residual,

$$G^{\delta}_{\mathrm{QHA}}(\boldsymbol{x}^\circ; T, P) := G_{\mathrm{QHA}}(\boldsymbol{x}^\circ; T, P) - U^\circ, \tag{S18}$$

which is the regression target for the low-fidelity training signal.

Within the isotropic-strain restriction, QHA is exact for a harmonic vibrational Hamiltonian and captures leading volumetric anharmonicity through $\omega_{\boldsymbol{q}\nu}(\boldsymbol{L})$. It neglects anisotropic cell relaxation and mode-mode coupling, and breaks down when any $\omega^2_{\boldsymbol{q}\nu}(\boldsymbol{L}) < 0$ along the relevant cell trajectory [S5]. For dynamically stabilized branches covered by assumption (A1), the static reference $\boldsymbol{x}^\circ$ is itself unstable under unconstrained relaxation, and QHA must be augmented or replaced by a finite-temperature method such as Frenkel–Ladd switching (Section B 2).

2. Frenkel–Ladd Switching

Frenkel–Ladd switching [S6] computes the absolute Helmholtz free energy of a single crystalline branch by thermodynamic integration along a path that connects the physical Hamiltonian to an analytically tractable harmonic reference. We follow the optimized switching variant of de Koning *et al.* [S7].

Let $U(\boldsymbol{r})$ denote the physical potential energy of the branch evaluated at the cell $\boldsymbol{L}^*$ equilibrated at the target (T, P) under the branch-conditioned NPT ensemble, and define the Einstein crystal reference

$$U_{\mathrm{harm}}(\boldsymbol{r}) := \frac{1}{2}\sum_{i=1}^{N} k_i \left\|\boldsymbol{r}_i - \boldsymbol{r}_i^\circ\right\|^2, \tag{S19}$$

where $\boldsymbol{r}_i^\circ$ are the relaxed atomic positions of $\boldsymbol{x}^\circ$ and k_i are species-resolved spring constants. The mixed Hamiltonian

$$U(\boldsymbol{r}; \lambda) := (1-\lambda)\,U(\boldsymbol{r}) + \lambda\,U_{\mathrm{harm}}(\boldsymbol{r}), \qquad \lambda \in [0, 1], \tag{S20}$$

interpolates between the two endpoints. Letting $\pi_\lambda(\boldsymbol{r}) \propto e^{-\beta U(\boldsymbol{r};\lambda)}$ denote the canonical (NVT) measure under the mixed Hamiltonian at fixed λ, the reversible-work

theorem reads

$$F_{\mathrm{harm}} - F = \int_0^1 \mathbb{E}_{\boldsymbol{r}\sim\pi_\lambda}[U_{\mathrm{harm}}(\boldsymbol{r}) - U(\boldsymbol{r})]\,\mathrm{d}\lambda. \quad \text{(S21)}$$

For a finite-rate switching trajectory in which $\lambda(t)$ is driven from 0 to 1 over a time τ_s, the accumulated nonequilibrium work

$$W^s := \int_0^{\tau_s} \dot{\lambda}(t)\left[U_{\mathrm{harm}}(\boldsymbol{r}(t)) - U(\boldsymbol{r}(t))\right]\mathrm{d}t \quad \text{(S22)}$$

is a path-dependent random variable whose distribution depends on the switching protocol; let $\mathbb{P}_{0\to1}$ and $\mathbb{P}_{1\to0}$ denote the forward and backward switching path measures, each initialized from the corresponding equilibrium endpoint. The driving follows a smooth polynomial schedule [S8, S9] in the reduced time $\tilde{t} = t/\tau_s$,

$$\lambda(\tilde{t}) = \tilde{t}^5\big(70\tilde{t}^4 - 315\tilde{t}^3 + 540\tilde{t}^2 - 420\tilde{t} + 126\big), \quad \text{(S23)}$$

whose slope $\mathrm{d}\lambda/\mathrm{d}\tilde{t} = 630\,\tilde{t}^4(1-\tilde{t})^4$ vanishes at both endpoints and reduces the dissipation relative to a linear ramp. The Jarzynski–Crooks dissipation inequality $\mathbb{E}_{\mathbb{P}_{0\to1}}[W^s] \geq F_{\mathrm{harm}} - F$ holds with equality only in the quasistatic limit. Forward and backward trajectories are combined symmetrically [S7] so that the leading-order dissipation cancels, giving the bidirectional estimator

$$\Delta F = \frac{1}{2}\big(\mathbb{E}_{\mathbb{P}_{0\to1}}[W^s] - \mathbb{E}_{\mathbb{P}_{1\to0}}[W^s]\big) \quad \text{(S24)}$$

and the bidirectional dissipation diagnostic

$$E_{\mathrm{d}} := \frac{1}{2}\big(\mathbb{E}_{\mathbb{P}_{0\to1}}[W^s] + \mathbb{E}_{\mathbb{P}_{1\to0}}[W^s]\big). \quad \text{(S25)}$$

The reference free energy is known in closed form: the classical Einstein crystal of N independent three-dimensional oscillators with frequencies $\omega_i := \sqrt{k_i/m_i}$ has

$$F_{\mathrm{harm}}(T) = 3\,k_{\mathrm{B}}T\sum_{i=1}^{N}\ln\left(\frac{\hbar\omega_i}{k_{\mathrm{B}}T}\right), \quad \text{(S26)}$$

so the absolute Helmholtz free energy of the physical branch follows from Eq. (S24) as

$$F(\boldsymbol{x}^\circ, \boldsymbol{L}^*; T) = F_{\mathrm{harm}}(T) - \Delta F. \quad \text{(S27)}$$

Pinning each atom to its lattice site $\boldsymbol{r}_i^\circ$ in U_{harm} also fixes the center of mass of the reference, a constraint absent from the unconstrained crystal partition function. The corresponding finite-size correction [S6] is

$$F_{\mathrm{COM}}(T) = -k_{\mathrm{B}}T\ln\left[V_{\mathrm{prim}}\left(2\pi k_{\mathrm{B}}T\sum_{i=1}^{N}\frac{\nu_i^2}{k_i}\right)^{-3/2}\right], \quad \text{(S28)}$$

where V_{prim} is the primitive-cell volume and $\nu_i := m_i/\sum_j m_j$ are the mass fractions. Combining the absolute Helmholtz free energy, the $G = F + PV$ conversion at the equilibrated cell, and the COM correction yields the Frenkel–Ladd estimate of the branch-conditioned Gibbs free energy

$$G_{\mathrm{FL}}(\boldsymbol{x}^\circ; T, P) = F_{\mathrm{harm}}(T) - \Delta F + PV + F_{\mathrm{COM}}(T). \quad \text{(S29)}$$

Because this $G = F + PV$ conversion evaluates the Helmholtz free energy at the single equilibrated cell $\boldsymbol{L}^*$ rather than integrating over cell fluctuations, the estimate Eq. (S29) is an approximation to the full branch-conditioned NPT Gibbs free energy Eq. (S10). This is the same cell-fluctuation approximation adopted in the CALPHY workflow [S10]. The neglected cell-fluctuation contribution arises from the few cell degrees of freedom, so it is of order $k_{\mathrm{B}}T$ per simulation cell and subextensive, hence negligible per atom for the ∼1,000-atom supercells used here. It can nonetheless grow for strongly flexible or anharmonic cells.

3. Alchemical Switching

Alchemical switching estimates the Gibbs free energy difference between two potential energy surfaces U^A and U^B at the same thermodynamic state $(\boldsymbol{x}^\circ; T, P)$, without requiring an absolute reference for either. The mixed Hamiltonian

$$U(\boldsymbol{r};\lambda) := (1-\lambda)\,U^A(\boldsymbol{r}) + \lambda\,U^B(\boldsymbol{r}), \qquad \lambda \in [0,1], \quad \text{(S30)}$$

interpolates between the two surfaces at fixed species and temperature, holding species and temperature fixed while allowing the cell variables to respond to (T, P). Both legs are run in the NPT ensemble. The same forward-backward symmetrization of Eq. (S24), with switching path measures $\mathbb{P}_{A\to B}$ and $\mathbb{P}_{B\to A}$ replacing $\mathbb{P}_{0\to1}$ and $\mathbb{P}_{1\to0}$, now yields the Gibbs free energy difference

$$\Delta G^{A\to B}(\boldsymbol{x}^\circ; T, P) := G^B - G^A = \frac{1}{2}\big(\mathbb{E}_{\mathbb{P}_{A\to B}}[W^s] - \mathbb{E}_{\mathbb{P}_{B\to A}}[W^s]\big), \quad \text{(S31)}$$

where each switching work is defined as in Eq. (S22) with the integrand $U_{\mathrm{harm}} - U$ replaced by $U^B - U^A$.

Combining alchemical switching with an absolute Frenkel–Ladd estimate Eq. (S29) performed against a faster potential A transfers the absolute Gibbs free energy onto a target surface B via the triangle relation

$$G^B(\boldsymbol{x}^\circ; T, P) = G^A(\boldsymbol{x}^\circ; T, P) + \Delta G^{A\to B}(\boldsymbol{x}^\circ; T, P). \quad \text{(S32)}$$

We use this with A = Orb-v3 (the faster reference) and B = UMA (the target): an absolute Frenkel–Ladd reference is computed on a faster potential, and a single alchemical step then transfers it onto the target sur-

face, sidestepping the computational cost of a direct Frenkel–Ladd calculation on the latter. Because Orb-v3 and UMA are trained on overlapping density functional theory data, their potential energy surfaces are closely matched, so the dissipation is much smaller than that of the Frenkel–Ladd switching.

4. Semigrand Canonical Monte Carlo

We combine semigrand canonical Monte Carlo (SGCMC) [S11] with well-tempered metadynamics (WT-MetaD) [S12] to sample equilibrium species concentrations on a fixed parent lattice. The parent lattice is a structural template defined by $\boldsymbol{L}^\circ$ and N atomic sites $\boldsymbol{r}^\circ = (\boldsymbol{r}_1^\circ, \ldots, \boldsymbol{r}_N^\circ)$ shared by all species assignments under consideration. For each $\boldsymbol{z} \in \mathcal{A}^N$, the corresponding branch representative is

$$\boldsymbol{x}_{\boldsymbol{z}}^\circ := \mathfrak{R}_{\mathrm{sym}}\big((\boldsymbol{L}^\circ, \boldsymbol{z}, [\boldsymbol{r}^\circ])\big) \in \mathcal{X}_N^\circ, \tag{S33}$$

and the SGCMC ensemble is supported on the parent-lattice union $\bigcup_{\boldsymbol{z}\in\mathcal{A}^N} \mathcal{B}(\boldsymbol{x}_{\boldsymbol{z}}^\circ)$. The Markov chain samples only the discrete site-occupancy variable $\boldsymbol{z}$: the continuous coordinates $(\boldsymbol{r}, \boldsymbol{L})$ are relaxed onto the branch representative and contribute only through the marginalized branch-conditioned Gibbs free energy

$$G(\boldsymbol{z}; T, P) := G(\boldsymbol{x}_{\boldsymbol{z}}^\circ; T, P) \tag{S34}$$

of Eq. (S10). We assume the parent lattice is preserved across all assignments under consideration, i.e., $\mathfrak{R}_{\mathrm{sym}}$ acts site-locally without breaking the shared lattice topology. This excludes vacancies, interstitials, and symmetry-changing reconstructions, which change the site count or lattice topology and lie outside this scope.

Within this formalism, SGCMC samples a Gibbs measure at fixed N, T, P, and chemical potentials $\boldsymbol{\mu} := (\mu_z)_{z\in\mathcal{A}}$, while the assignment $\boldsymbol{z} \in \mathcal{A}^N$ fluctuates; metadynamics promotes sampling across the relevant composition basins so that all phases of interest are visited within a single trajectory.

The SGC ensemble is sampled over the discrete assignment $\boldsymbol{z}$ by *identity swap moves*, with the configurational coordinates $(\boldsymbol{r}, \boldsymbol{L})$ relaxed onto the branch representative: select a site i uniformly at random, propose changing its species $z_i \to z' \in \mathcal{A}$, and accept with probability

$$P_{\mathrm{acc}}(z_i \to z') = \min\Big(1,\, e^{-\beta\,[\Delta G_\theta + \mu_{z_i} - \mu_{z'} + \Delta V_{\mathrm{bias}}]}\Big), \tag{S35}$$

where ΔG_θ is the change in the surrogate branch-conditioned Gibbs free energy G_θ under the swap, evaluated on the running configuration, and ΔV_{bias} is the accompanying change in the well-tempered metadynamics bias Eq. (S37) on the composition collective variable. Only chemical-potential differences are physical, so we fix a reference species $z_0 \in \mathcal{A}$ and parametrize the ensemble by $\Delta\mu_z := \mu_z - \mu_{z_0}$ for $z \neq z_0$. Detailed balance with Eq. (S35) yields, at fixed bias, the biased semigrand measure $\pi_\theta(\boldsymbol{z}) \propto \exp\big[-\beta\big(G_\theta(\boldsymbol{z}; T, P) - \sum_{z\in\mathcal{A}} \mu_z N_z(\boldsymbol{z})\big) - \beta\, V_{\mathrm{bias}}(\boldsymbol{c}(\boldsymbol{z}))\big]$ as the stationary distribution of the Markov chain [S11]; the unbiased semigrand distribution Eq. (S40) is recovered by reweighting.

The acceptance ratio of Eq. (S35) requires only the surrogate Gibbs change ΔG_θ and the bias change of the proposed swap, obtained from a single surrogate evaluation on the running configuration. The full symmetry-constrained relaxation occurs only on logging steps, i.e., when the configurations are retained as training or reweighting samples. Intermediate proposals are not relaxed, which reduces the computational cost of sampling, while all stored configurations are fully relaxed. For intermediate proposals, evaluating G_θ on the running configuration approximates the relaxed branch free energy $G(\boldsymbol{x}_{\boldsymbol{z}}^\circ)$ of Eq. (S34). The acceptance probability depends only on the free energy difference of two configurations that differ by a single swap, for which the relaxation contribution largely cancels, and every stored sample carries the fully relaxed value used in the reweighting of Section B 5. In the production runs, each step attempts a composition-changing identity move or a composition-conserving swap with equal probability. The well-tempered bias of Eq. (S37) deposits 0.05 eV Gaussians every 25 steps with bias factor $\gamma = 10$. Configurations are retained for relaxation and labeling every 10 steps after a 40% burn-in.

To accelerate composition sampling, define the composition collective variable

$$\boldsymbol{c}(\boldsymbol{z}) := \frac{1}{N}\big(N_z(\boldsymbol{z})\big)_{z\in\mathcal{A}}, \qquad N_z(\boldsymbol{z}) := \sum_{i=1}^{N} \mathbf{1}\{z_i = z\}, \tag{S36}$$

which lives on the $(|\mathcal{A}|-1)$-simplex. Well-tempered metadynamics [S12] deposits Gaussian kernels along the CV trajectory,

$$V_{\mathrm{bias}}(\boldsymbol{c}, t) = \sum_{t'\le t} w(t')\, \exp\left[-\frac{\|\boldsymbol{c} - \boldsymbol{c}(t')\|^2}{2\sigma^2}\right], \tag{S37}$$

with deposition heights rescaled by the running bias,

$$w(t) = w_0\, \exp\left[-\frac{V_{\mathrm{bias}}(\boldsymbol{c}(t),\, t)}{(\gamma - 1)\, k_{\mathrm{B}} T}\right]. \tag{S38}$$

The bias factor $\gamma > 1$ sets an effective CV temperature $T_{\mathrm{eff}} := \gamma T$; in the long-time limit the bias converges to

$$V_{\mathrm{bias}}(\boldsymbol{c}) \xrightarrow{t\to\infty} -\left(1 - \frac{1}{\gamma}\right) G_{\mathrm{SGC}}(\boldsymbol{c};\, \boldsymbol{x}^\circ, T, P, \boldsymbol{\mu}) + \mathrm{const}, \tag{S39}$$

where $G_{\mathrm{SGC}}(\boldsymbol{c}; \cdots)$ is the projection of the SGC Gibbs free energy onto the composition CV at fixed branch and external conditions.

The equilibrium concentration distribution at $(T, P, \boldsymbol{\mu})$ on the branch $\mathcal{B}(\boldsymbol{x}^\circ)$ is recovered by reweighting accord-

ing to Eq. (S39),

$$\pi_{\mathrm{SGC}}(\boldsymbol{c};\,\boldsymbol{x}^{\circ},T,P,\boldsymbol{\mu}) \propto \exp\big[-\beta\,G_{\mathrm{SGC}}(\boldsymbol{c};\,\boldsymbol{x}^{\circ},T,P,\boldsymbol{\mu})\big], \tag{S40}$$

and is reweighted from the biased trajectory using the time-independent estimator of Tiwary and Parrinello [S13]. Equilibrium concentrations are identified with the basin-conditioned mean concentrations of π_{SGC}, and sweeping $\boldsymbol{\mu}$ at fixed (T,P) traces out the branch-conditioned phase boundary in composition space.

5. Differentiable Reweighting

The parameters θ of a Gibbs free energy model $G_\theta(\boldsymbol{z};T,P)$ are fine-tuned to reproduce experimental binodal concentrations by differentiable reweighting [S14] of stored SGCMC plus WTMetaD samples. For clarity, we specialize to a binary alloy with species $\mathcal{A}=\{A,B\}$, scalar concentration $c(\boldsymbol{z}) := N_B(\boldsymbol{z})/N$, and chemical potential difference $\Delta\mu := \mu_B - \mu_A$. The basins $b\in\{A,B\}$ label the two coexisting phases (A-rich and B-rich). Generalization to multi-species alloys follows by replacing c with the simplex-valued composition $\boldsymbol{c}$ of Eq. (S36) and $\Delta\mu$ with a vector of chemical-potential differences relative to a reference species.

For each experimental state point $m=(T_m,P_m)$, we run a single reference SGCMC plus WTMetaD simulation at parameters θ_0 and chemical potential $\Delta\mu_{0,m}$ chosen close to coexistence so that both basins are sampled, and store M_m decorrelated samples

$$\mathcal{D}_m := \big\{\boldsymbol{z}_{m,i},\,c_{m,i},\,V_{m,i},\,\kappa_{m,i}\big\}_{i=1}^{M_m}, \tag{S41}$$

where $c_{m,i} := c(\boldsymbol{z}_{m,i})$, $V_{m,i} := V_{\mathrm{bias}}^{(m)}(c_{m,i},t_i)$ is the metadynamics bias of Eq. (S37) at the deposition time t_i, and $\kappa_{m,i}$ is the time-dependent reweighting offset of well-tempered reweighting [S13] (an irrelevant additive constant if a frozen final bias is used). The simulator is not differentiated; gradients flow only through reweighted observables on $\mathcal{D}_m$.

For a candidate $(\theta,\Delta\mu)$, the importance log-weight of sample i relative to the reference is

$$\begin{aligned}\ell_{m,i}(\theta,\Delta\mu) := -\beta_m\,\big[&G_\theta(\boldsymbol{z}_{m,i};\,T_m,P_m)\\ &- G_{\theta_0}(\boldsymbol{z}_{m,i};\,T_m,P_m) - N\,(\Delta\mu-\Delta\mu_{0,m})\,c_{m,i}\big]\\ &+\beta_m\,(V_{m,i}-\kappa_{m,i}),\end{aligned} \tag{S42}$$

with $\beta_m := 1/k_{\mathrm{B}}T_m$. The bracketed term is the semigrand-energy difference between candidate and reference parameters, and the $V_{m,i}-\kappa_{m,i}$ term unbiases the metadynamics bias accumulated during the reference simulation. Normalizing by softmax,

$$w_{m,i}(\theta,\Delta\mu) := \frac{\exp\ell_{m,i}(\theta,\Delta\mu)}{\sum_{j=1}^{M_m}\exp\ell_{m,j}(\theta,\Delta\mu)}, \tag{S43}$$

yields a finite-sample estimator of any semigrand expectation under the candidate measure

$$\widehat{\mathbb{E}}^{(m)}_{\theta,\Delta\mu}[O] := \sum_{i=1}^{M_m} w_{m,i}(\theta,\Delta\mu)\,O(\boldsymbol{z}_{m,i}). \tag{S44}$$

The differentiable observable for binodal calibration is a basin-conditioned mean concentration, not the full free energy profile. Define basin masks $M_{m,b}:[0,1]\to\{0,1\}$ on the concentration axis, with $M_{m,b}(c)=1$ iff c lies in the support of basin b identified from the unbiased reference profile (typically the interval enclosing the corresponding minimum, with the boundary between basins placed near the intervening barrier). The basin-local weights and predicted binodal concentration are

$$w^{(b)}_{m,i}(\theta,\Delta\mu) := \frac{M_{m,b}(c_{m,i})\,\exp\ell_{m,i}(\theta,\Delta\mu)}{\sum_j M_{m,b}(c_{m,j})\,\exp\ell_{m,j}(\theta,\Delta\mu)}, \tag{S45}$$

$$\widehat{c}_{m,b}(\theta,\Delta\mu) := \sum_i w^{(b)}_{m,i}(\theta,\Delta\mu)\,c_{m,i}. \tag{S46}$$

At coexistence, $\Delta\mu$ is the value enforcing equal semigrand free energies of the two basins, so it is not a free parameter. With unnormalized basin partition function

$$Z_{m,b}(\theta,\Delta\mu) := \sum_i M_{m,b}(c_{m,i})\,\exp\ell_{m,i}(\theta,\Delta\mu), \tag{S47}$$

coexistence reads $Z_{m,A}(\theta,\Delta\mu)=Z_{m,B}(\theta,\Delta\mu)$, or equivalently

$$H_m(\theta,\Delta\mu) := \ln Z_{m,A}(\theta,\Delta\mu) - \ln Z_{m,B}(\theta,\Delta\mu) = 0. \tag{S48}$$

For each θ, Eq. (S48) is solved for $\Delta\mu$ as a scalar root-finding problem using the stored samples Eq. (S41) without additional simulation. The condition Eq. (S48) is the "common-tangent construction" for the binodal: the two basins, evaluated at a common candidate $\Delta\mu$, share the same tangent slope, and equality of their semigrand free energies (equivalently $Z_{m,A}=Z_{m,B}$) makes that tangent "touch" both branches of the composition free energy. Denoting the solution by $\Delta\mu^\star_m(\theta)$, the coexistence-conditioned binodal observable is

$$\widehat{c}^{\,\star}_{m,b}(\theta) := \widehat{c}_{m,b}\big(\theta,\,\Delta\mu^\star_m(\theta)\big). \tag{S49}$$

Implicit differentiation of Eq. (S48) yields the gradient of $\Delta\mu^\star_m$ with respect to θ,

$$\frac{\mathrm{d}\Delta\mu^\star_m}{\mathrm{d}\theta} = \frac{\widehat{\mathbb{E}}^{(m,A)}[\nabla_\theta G_\theta] - \widehat{\mathbb{E}}^{(m,B)}[\nabla_\theta G_\theta]}{N\,\big(\widehat{c}^{\,\star}_{m,A}-\widehat{c}^{\,\star}_{m,B}\big)}, \tag{S50}$$

where $\widehat{\mathbb{E}}^{(m,b)}$ denotes the basin-local expectation with weights $w^{(b)}_{m,i}$ at $\Delta\mu=\Delta\mu^\star_m(\theta)$.

Given experimental binodal data $\{c^{\mathrm{exp}}_{m,b}\}$ across state points m and basins $b\in\{A,B\}$, the model parameters

are obtained by minimizing the weighted regression loss

$$\mathcal{L}(\theta) := \lambda_{\mathrm{bin}} \sum_{m} \sum_{b\in\{A,B\}} \left(\widehat{c}^{\star}_{m,b}(\theta) - c^{\mathrm{exp}}_{m,b}\right)^2 + \lambda_{\mathrm{drift}} \sum_{m} \Delta_{\mathrm{drift},m}(\theta)^2 + \lambda_{\mathrm{anchor}} \left\|\theta - \theta_0\right\|_2^2, \quad \text{(S51)}$$

with $c^{\mathrm{exp}}_{m,b}$ the experimental boundary concentrations. The energy-drift term

$$\Delta_{\mathrm{drift},m}(\theta) := \frac{1}{NM_m} \sum_{i=1}^{M_m} \left|G_\theta(\boldsymbol{z}_{m,i}) - G_{\theta_0}(\boldsymbol{z}_{m,i})\right| \quad \text{(S52)}$$

measures the absolute per-atom change in the surrogate branch free energy between the reference parameters θ_0 and the current θ, averaged over the M_m stored samples $\mathcal{D}_m$ of Eq. (S41) at state point m. This term limits the parameter update within each cycle and maintains overlap with the sampled distribution. The final term is an L_2 penalty toward the same reference parameters θ_0. The parameters are optimized by Adam (learning rate 5×10^{-6}, gradient clipping at norm 5) for 160 gradient steps within each cycle, with weights $\lambda_{\mathrm{bin}} = 100$, $\lambda_{\mathrm{drift}} = 300$, and $\lambda_{\mathrm{anchor}} = 3 \times 10^4$ in the deployed Au–Pt calibration. Within each calibration cycle, the stored samples are reused by reweighting without further simulation; between cycles, the reference SGCMC plus WT-MetaD simulation is re-run at the updated parameters and coexistence estimate. Between cycles, the metadynamics bias is initialized from the previous cycle. Because successive parameter updates are small, fewer deposition steps are required to reconverge.

C. TRAINING DETAILS

1. Multi-Fidelity Training

Table S2 lists the weights and per-response caps.

Table S2. **Pre-training loss terms.** Each is a L_1 penalty on the residual values. The extensive quantities are normalized per atom; the intensive responses are not. The cap (in the units shown) bounds each per-sample contribution, and a global gradient-norm clip of 10 applies to the summed loss.

Term	Units	Per atom	λ	Cap
$\mathcal{L}_G$	eV atom^{-1}	yes	1	—
$\mathcal{L}_V$	Å^3 atom^{-1}	yes	0.25	—
$\mathcal{L}_S$	J K^{-1} (mol atoms)$^{-1}$	yes	0.005	200
$\mathcal{L}_{C_P}$	J K^{-1} (mol atoms)$^{-1}$	yes	0.003	200
$\mathcal{L}_B$	GPa	no	0.003	2000
$\mathcal{L}_\alpha$	10^{-5} K^{-1}	no	0.05	200

The trained thermodynamic adapter is small relative to the frozen encoder it augments. The UMA representation (`UMA-S-1.1`) carries $\sim 1.5 \times 10^8$ parameters, of which $\sim 6 \times 10^6$ are active per structure through its mixture-of-experts routing. The trainable thermodynamic adapter has $\sim 2.1 \times 10^5$ parameters. Of these, the analytic coefficient heads account for $\sim 9 \times 10^3$. Mid-training adds a $\sim 4.4 \times 10^4$-parameter LoRA update [S15] (Methods), bringing the adapter to $\sim 2.6 \times 10^5$ parameters in total, of which the LoRA update is about 17%.

2. Analytic Head Constants and Numerical Evaluation

The head uses two fixed global normalization scales, $T_{\max} = 4000$ K and $P_{\max} = 40$ GPa. The reduced temperature $\tau = T/T_{\max}$ and the normalized pressure $P/P_{\max}$ each map their sampled range into $(0, 1]$. Neither normalization scale is material specific, so the predicted surface is a function of structure alone. All head quantities use a fixed unit system: per-atom energies in eV, temperatures in K, per-atom volumes in Å^3, and pressures in GPa. The reference coefficients c_0–c_3, $c_{\log}$, and c_{inv} are then per-atom energies, and θ_{E} is a temperature. The equation-of-state parameters are read as $V_{0,i}$ in Å^3, $B_{0,i}$ in GPa, and $B'_{0,i}$ dimensionless. The polynomial coefficients $a_k^{V_0}$, $a_k^{B_0}$, and $a_k^{B'_0}$ enter their softplus as dimensionless numbers. In the Murnaghan volume integral the pressure enters only through the dimensionless ratio $B'_{0,i}P/B_{0,i}$, and the pressure-volume work, formed in Å^3 GPa, is converted to eV with 1 eV/Å^3 = 160.2 GPa before it is added to the reference branch. All remaining constants are fixed as follows: the normalized Einstein-temperature floor $\theta_{\min} = 10^{-4}$, the temperature floor $\tau_{\min} = 10^{-8}$ (both in units of $T_{\max}$), and the gate exponent $p = 4$ regularize the normalization and vibrational terms. The equation-of-state floors $V_{0,\min} = 10^{-6}$ Å^3, $B_{0,\min} = 10^{-3}$ GPa, and $B'_{0,\min} = 10^{-3}$, together with the Murnaghan argument floor $x_{\min} = 10^{-8}$, keep the equation of state well-defined. The initialization bias $B_{\mathrm{off}} = 100$ GPa is added inside the $B_{0,i}$ softplus so the untrained bulk modulus starts near 100 GPa.

D. HOLLAND–POWELL CALIBRATION AND BENCHMARK DETAILS

The experimental anchoring of ordered phases (Methods) uses the Holland–Powell `ds62` dataset [S16]: a predefined set of Si–Al–Mg–Ca–O crystalline phases (Table S3) together with reactions among them (listed below). For each phase, `ds62` provides the experimentally assessed single-phase properties at 298.15 K and 1 bar (the isothermal bulk modulus B_0, the thermal expansivity α, the standard entropy $S(298\ \mathrm{K})$, and the molar volume V_0). Every reaction and polymorph contributes a Gibbs free energy difference target. The training phases are the binary and ternary oxides above the divider in Table S3. The six quaternary oxides below the divider, to-

gether with the held-out reactions that form them, are excluded from fine-tuning and used only for evaluation. Held-out accuracy therefore measures transfer from binary and ternary training phases to quaternary phases. Phase abbreviations follow Holland and Powell [S16].

Table S3. **Holland–Powell benchmark phases.** Binary and ternary phases used for training are listed above the divider; quaternary phases used only for evaluation are listed below.

Abbrev.	Composition	Mineral
q	SiO_2	quartz
trd	SiO_2	tridymite
crst	SiO_2	cristobalite
coe	SiO_2	coesite
stv	SiO_2	stishovite
per	MgO	periclase
lime	CaO	lime
cor	Al_2O_3	corundum
fo	Mg_2SiO_4	forsterite
mwd	Mg_2SiO_4	wadsleyite
mrw	Mg_2SiO_4	ringwoodite
en	$Mg_2Si_2O_6$	enstatite
mpv	$MgSiO_3$	Mg-perovskite
mak	$MgSiO_3$	akimotoite
cen	$Mg_2Si_2O_6$	clinoenstatite
pren	$Mg_2Si_2O_6$	protoenstatite
lrn	Ca_2SiO_4	larnite
rnk	$Ca_3Si_2O_7$	rankinite
wo	$CaSiO_3$	wollastonite
cpv	$CaSiO_3$	Ca-perovskite
pswo	$CaSiO_3$	pseudowollastonite
wal	$CaSiO_3$	walstromite
cstn	$CaSi_2O_5$	Ca-Si titanite
ky	Al_2SiO_5	kyanite
and	Al_2SiO_5	andalusite
sill	Al_2SiO_5	sillimanite
sp	$MgAl_2O_4$	spinel
mont	$CaMgSiO_4$	monticellite
merw	$Ca_3MgSi_2O_8$	merwinite
ak	$Ca_2MgSi_2O_7$	akermanite
di	$CaMgSi_2O_6$	diopside
py	$Mg_3Al_2Si_3O_{12}$	pyrope
gr	$Ca_3Al_2Si_3O_{12}$	grossular

Reactions are written with the `ds62` phase abbreviations of Table S3.

Training (15):

- 2 per + q → fo
- fo + q → en
- lime + q → wo
- cor + q → ky
- cor + q → and
- cor + q → sill
- 2 lime + q → lrn
- 3 lime + 2 q → rnk
- lime + 2 q → cstn
- per + cor → sp
- cen → en
- pren → en
- pswo → wo
- and → ky
- sill → ky

Polymorph (16):

- trd → q
- crst → q
- coe → q
- stv → coe
- mwd → fo
- mrw → fo
- 2 mpv → en
- 2 mak → en
- cen → en
- pren → en
- cpv → wo
- pswo → wo
- wal → wo
- and → ky
- sill → ky
- sill → and

Held-out (6):

- per + lime + q → mont
- 3 lime + per + 2 q → merw
- 2 lime + per + 2 q → ak
- wo + per + q → di
- 1.5 en + cor → py
- 3 wo + cor → gr

The post-training objective is smooth-L_1 (Huber) throughout, apart from the L_2 weight regularizer $\mathcal{L}_{\mathrm{reg}}$. The reaction, polymorph, and shape terms are normalized per atom. All loss terms are evaluated on the training phases, and $\mathcal{L}_{\mathrm{rxn}}$ and $\mathcal{L}_{\mathrm{poly}}$ run over the 15 training reactions and 16 polymorph pairs above. The single-phase anchors $\mathcal{L}_{B_0}$, $\mathcal{L}_{S_0}$, and $\mathcal{L}_{V_0}$ target the `ds62` bulk modulus, entropy, and molar volume tabulated at 298.15 K and 1 bar. The remaining data terms target the `ds62` apparent Gibbs surface $G^{\mathrm{ds62}}(T, P)$. A global gradient-norm clip of 5 applies to the summed loss. The per-term weights are collected in Table S4.

Table S4. **Post-training loss terms for ordered phases.** Per-term weights λ of the experimental calibration objective (Methods).

Term	Target	λ
$\mathcal{L}_{\mathrm{rxn}}$	reaction ΔG	1
$\mathcal{L}_{\mathrm{poly}}$	polymorph ΔG	0.5
$\mathcal{L}_{B_0}$	B_0	0.003
$\mathcal{L}_{S_0}$	$S(298\ \mathrm{K})$	0.003
$\mathcal{L}_{V_0}$	V_0	0.015
$\mathcal{L}_{V}$	$V(T, P)$	0.05
$\mathcal{L}_{C_P}$	$C_P(T)$	0.3
$\mathcal{L}_{\mathrm{shape}}$	$G(T, P) - G(T_0, P_0)$	0.1
$\mathcal{L}_{\mathrm{reg}}$	pre-calibration weights	0.00001

[S1] A. Van De Walle and G. Ceder, The effect of lattice vibrations on substitutional alloy thermodynamics, Rev. Mod. Phys. **74**, 11 (2002).

[S2] S. Baroni, S. De Gironcoli, A. Dal Corso, and P. Giannozzi, Phonons and related crystal properties from density-functional perturbation theory, Rev. Mod. Phys. **73**, 515 (2001).

[S3] A. Togo and I. Tanaka, First principles phonon calculations in materials science, Scr. Mater. **108**, 1 (2015).

[S4] P. Vinet, J. R. Smith, J. Ferrante, and J. H. Rose, Temperature effects on the universal equation of state of solids, Phys. Rev. B **35**, 1945 (1987).

[S5] A. Van De Walle, Q. Hong, S. Kadkhodaei, and R. Sun, The free energy of mechanically unstable phases, Nat. Commun. **6**, 7559 (2015).

[S6] D. Frenkel and A. J. Ladd, New Monte Carlo method to compute the free energy of arbitrary solids. application to the fcc and hcp phases of hard spheres, J. Chem. Phys. **81**, 3188 (1984).

[S7] M. de Koning, A. Antonelli, and S. Yip, Optimized free-energy evaluation using a single reversible-scaling simulation, Phys. Rev. Lett. **83**, 3973 (1999).

[S8] M. de Koning and A. Antonelli, Einstein crystal as a reference system in free energy estimation using adiabatic switching, Phys. Rev. E **53**, 465 (1996).

[S9] J. Nam, J. Peng, and R. Gómez-Bombarelli, Interpolation and differentiation of alchemical degrees of freedom in machine learning interatomic potentials, Nat. Commun. **16**, 4350 (2025).

[S10] S. Menon, Y. Lysogorskiy, J. Rogal, and R. Drautz, Automated free-energy calculation from atomistic simulations, Phys. Rev. Mater. **5**, 103801 (2021).

[S11] B. Sadigh, P. Erhart, A. Stukowski, A. Caro, E. Martinez, and L. Zepeda-Ruiz, Scalable parallel Monte Carlo algorithm for atomistic simulations of precipitation in alloys, Phys. Rev. B **85**, 184203 (2012).

[S12] A. Barducci, G. Bussi, and M. Parrinello, Well-tempered metadynamics: a smoothly converging and tunable free-energy method, Phys. Rev. Lett. **100**, 020603 (2008).

[S13] P. Tiwary and M. Parrinello, A time-independent free energy estimator for metadynamics, J. Phys. Chem. B **119**, 736 (2015).

[S14] S. Thaler and J. Zavadlav, Learning neural network potentials from experimental data via Differentiable Trajectory Reweighting, Nat. Commun. **12**, 6884 (2021).

[S15] E. J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, and W. Chen, LoRA: Low-rank adaptation of large language models, in *International Conference on Learning Representations* (2022).

[S16] T. J. B. Holland and R. Powell, An improved and extended internally consistent thermodynamic dataset for phases of petrological interest, involving a new equation of state for solids, J. Metamorph. Geol. **29**, 333 (2011).