

Thermosalient phase transitions from machine learning interatomic potential

Bruno Mladineo and Ivor Lončarić*

Ruder Bošković Institute, Bijenička 54, Zagreb, Croatia

E-mail: ivor.loncaric@gmail.com

Abstract

We developed an accurate machine learning interatomic potential for the thermosalient molecular crystal N-2-Propylidene-4-hydroxybenzohydrazide. This crystal exhibits one of the largest mechanical responses during its thermosalient phase transition. Leveraging the speed of our developed potential, we performed Gibbs free energy calculations that successfully predict phase transitions in good agreement with experimental observations. Additionally, our model accurately captures the phenomenon of negative linear thermal expansion preceding the thermosalient phase transition. We show that the energy barrier exists at phase transition temperature and that this energy is purely elastic elucidating the physical reasons for the thermosalient effect.

Polymorphism is a common occurrence in molecular crystals, influencing their physico-chemical properties with important consequences in the production and development of pharmaceuticals, dyes, or explosives. This makes modeling polymorphs and their transformations of great importance. Modeling is usually based either on first-principles quantum methods such as density functional theory (DFT) or (tailor-made) classical forcefield. The former brings the needed accuracy at a considerable cost, while the latter is orders of magnitude faster but significantly less reliable. In recent years machine learning interatomic potentials

(MLIPs) have arisen with the promise to provide both accuracy and speed.¹ However, so far their use has been limited for molecular crystals. First, the computational cost of DFT associated with large unit cells of molecular crystals is a limiting factor for the widespread use of MLIPs in modeling molecular crystals.² Second, the lion’s share of MLIP architectures are short-ranged, while long-range interactions, such as electrostatics and dispersion, are inherent in molecular crystals,³ posing a challenge for the application of MLIPs. Despite these challenges, certain successes have been made in specific applications, such as accurately modeling the potential of pentacene and azapentacene with sub-kJmol⁻¹ accuracy⁴ or predicting thermodynamic stability in agreement with experiments for various molecular crystals.^{5,6} Notably, these studies suggest that some properties of specific molecular crystals can be accurately modeled with short-ranged MLIPs.

In this communication, we construct an MLIP to model a particularly interesting polymorphic phase transition called thermosalient effect or colloquially jumping crystals. The thermosalient effect is a rare phenomenon during temperature-induced polymorphic phase transitions characterized by a fast and anisotropic change in lattice parameters that makes the visually attractive mechanical response of crystals jumping.^{7,8} In general, molecular crystals that exhibit a mechanical response on a macroscopic scale when stimulated thermally (and/or photostimulated) are a useful platform for designing actuators, artificial muscles, biomimetic kinematic devices, as well as heat and light sensors. Crystalline materials, if used as actuators, offer a key advantage over soft materials due to their rapid response, typically occurring in less than a millisecond.

Out of a few known thermosalient molecular crystals, we focus on N²-2-Propylidene-4-hydroxybenzohydrazide which stands out as it is a current record-holder for the distance it can jump and at the same time exhibits the largest uniaxial negative thermal expansions reported so far.⁹⁻¹¹ This system has three known polymorphs. In conventional synthesis, form I is obtained. Upon heating to ~ 420 K, there is an irreversible thermosalient transition to form II preceded by extremely large uniaxial negative thermal expansions and large biaxial

positive thermal expansion. Form II is the stable polymorph at higher temperatures. Upon cooling, at ~ 350 K there is a reversible thermosalient transition to form III which is the stable polymorph at low temperatures. Both form II and form III also show uniaxial negative thermal expansion.

The reasons for the thermosalient effect are still not fully understood, impairing the possibility of prediction of thermosalient systems. Several first-principles (DFT) modeling studies shed light on atomistic processes in some thermosalient systems such as phonon vibrations that facilitate the transition.^{9,12,13} However, due to the size of the unit cell, more detailed DFT calculations are prohibitively expensive. For example, N-2-Propylidene-4hydroxybenzohydrazide has a unit cell of four molecules each with 26 atoms.

The training dataset for our MLIP is based on DFT calculations that have been performed using Vienna ab-initio simulation package, VASP.6.3.0^{14,15} employing r²SCAN exchange-correlation functional.¹⁶ A complete input file for calculations is available in supporting information.

The initial dataset was created via normal mode sampling¹⁷ of three different polymorphs and isolated molecules, with different temperatures for each form. Different temperatures were selected to further diversify the dataset with form I at 500 K, form II at 300K, form III at 400K, and the isolated molecule at 200K. We used the phonon module of Atomic Simulation Environment (ASE)¹⁸ with the existing MLIP ANI-2x¹⁹ from the TorchANI package.²⁰ On top of the potential, we used the DFT-D4 correction with parameters corresponding to the baseline functional. Obtained phonon modes were filtered to remove modes (structures) that are degenerate (too similar to each other), or that had negative or too low frequencies resulting in non-physical structures. Figure S1 shows the distribution of selected phonon modes. In total, we selected 706 structures for the initial database.

To improve the robustness and accuracy of the potential we expanded the dataset via an active learning scheme. To that end, we performed molecular dynamics (MD) simulations with conserved particle number (N), pressure (P), and temperature (T), the isother-

mal-isobaric ensemble (NPT), as implemented in ASE. The NPT MDs of the system are run at slowly increasing temperatures (see Figs. S2 and S3), with 1 fs timestep, the characteristic timescale of the thermostat set to 10 fs, and the constant in the barostat differential equation set to 1. Estimated uncertainty was calculated every 30 MD steps. We selected 100-200 structures for which the uncertainty was the highest, recalculated them using DFT, and added them to the dataset. We used the query by committee (QBC), an ensemble-based approach successfully implemented in other works,²¹ as an uncertainty estimator. We have used both the energy uncertainty and the maximum force uncertainty from all the atoms in a structure. The energy uncertainty estimate can be viewed as a global estimate of uncertainty of the entire structure, while the force uncertainty, as defined, is a local estimate of uncertainty. To get the most out of active learning schemes it is important to maximise variation in the models of the ensemble. To do so the training of the models was done with different initial weights, and also with different training/validation dataset splits. In every generation, eight models were trained, six (best performing on the validation set) models were used for the uncertainty estimate and two worst-performing models per generation were dropped. In total, we did six generations of active learning. The final dataset (5th GEN) has a total of 1507 structures and mean absolute error (MAE)/ root mean square error (RMSE) of 0.45/0.56 meV/atom. In Fig. 1 we show the accuracy evaluation of the best model of GEN0/GEN5 compared to DFT data from the final dataset, illustrating the improvement in accuracy and reduction in outliers achieved as a result of the active learning process.

We chose NequIP, an E(3)-equivariant message-passing graph neural network, as our MLIP architecture because of its capacity to achieve state-of-the-art accuracy across a wide range of molecules and materials while maintaining exceptional data efficiency.²² Several parameters of the network can be used to increase the accuracy by increasing the complexity, but this decreases the speed of the model inference. It is critical to find a balance between complexity and speed and we approached this task via hyperparameter optimization. Hyperparameter optimization was done on the number of features (N_F) and cutoff radius (r_c).

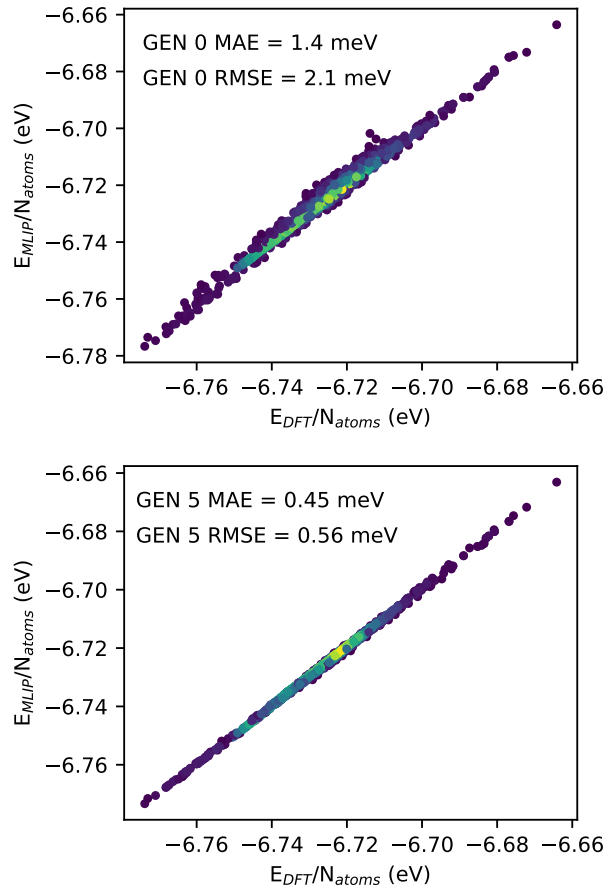


Figure 1: Comparison of MLIP and DFT energy predictions. Top panel: Generation 0 model. Bottom panel: Generation 5 model.

The first specifies the multiplicity of irreducible representations (features) and positively correlates with the complexity of the model and negatively with the speed of inference (prediction). r_c specifies a maximum distance for the message to pass from one atom to another in one message-passing iteration and describes the size of the local environment. A larger r_c can improve accuracy as the network looks at a bigger description space, but slows down the model. All other model parameters were left default and can be found in supporting data.

In Fig. S4 we show the impact of N_F and r_c on the accuracy of the model. Changing the N_F between 32 and 64 has a limited impact on model accuracy, therefore we can infer that the model is complex enough with 32 features. Increasing this value would only lead to a slower model while increasing the danger of overfitting. We control for overfitting by evaluating

the model during fitting on a validation dataset, where we use a 90/10 training/validation dataset split. The fitting process is stopped after the loss on the validation dataset stops decreasing for 100 epochs, and the model selected for use is the best-performing one on the validation dataset from all epochs. r_c specifies the maximum distance a message will pass from one atom to another and can be interpreted as a distance cap for interaction a model can capture, but because of message passing the effective range of local environments is multiple of r_c , allowing for use of a much lower r_c than might be considered physically relevant for an organic molecular crystal. In a molecular crystal atoms inside a molecule are close to each other and the message in the message passing travels without interruption from atom to atom, but atoms from neighboring molecules can be "invisible" if at least a pair of atoms from neighboring molecules are not close enough to "bridge the gap", i.e. transferring the message from one molecule to the other. In effect, there is a minimum r_c for our system so that the model can learn the intermolecular interactions. As can be seen in figure [S4] for values of $r_c = 3.5 \text{ \AA}$ and lower accuracy are progressively worse, while for values of 4 \AA and higher, the accuracy is similar. From this, we conclude that the value of $r_c = 4 \text{ \AA}$ is the minimal value required to get the intermolecular interaction information while the higher values do not contribute to a significant improvement of accuracy while contributing massively to the slowdown of the model. This is understandable considering the amount of information encoded in the local environment effectively scales with the volume or $\sim r_c^3$.

All simulations using the ML model were performed in ASE. To account for van der Waals forces, on top of the ML potential, we used the DFT-D4 correction with parameters corresponding to r²SCAN.²³ Since the three-body term in DFT-D4 is computationally expensive for large supercells used in this work but its effect is rather small for molecular crystals,²⁴ its contribution was neglected. To calculate the phonons in harmonic approximation (HA) and corresponding vibrational free energies we used the Phonopy package.^{25,26} We used $3 \times 3 \times 3$ supercells and the finite difference displacements of 0.01 \AA . Once the force constants were obtained, we used $17 \times 17 \times 17$ Monkhorst-Pack q-point sampling to obtain all

phonon-related properties. High-temperature form II cannot be freely relaxed by our model without conversion to form III (as such calculation corresponds to $T=0$). Therefore, we have fixed its volume to the experimental value for the HA calculations.

All three experimentally known forms have a similar harmonic phonon density of states (DOS) as shown in Fig. S8. Only small differences exist in the low frequencies, where the phonon density of states of form II rises the fastest which makes its entropic contribution to Helmholtz free energy the largest, followed by form I. As shown in Fig. S9, enthalpy is on the other hand the smallest for form III, followed by form I, and then form II. This makes the Helmholtz free energy as a function of the temperature of the three forms to cross, inducing the observed phase transitions.

To include the effect of thermal expansion, we have calculated Gibbs free energy within quasi-harmonic approximation (QHA)²⁷ as shown in Fig. 2. In the development of our

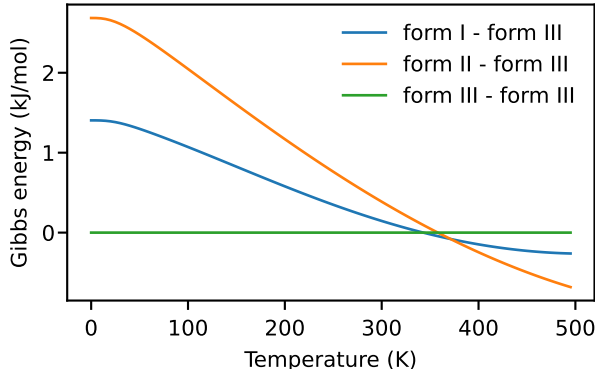


Figure 2: Gibbs free energy per unit cell for the three forms of the crystal calculated in QHA with the full range of the energy-volume curve used for the fit of EOS.

workflow for the QHA calculations, we explored multiple strategies to construct the crucial energy-volume curve. Initially, we used a conventional approach that involves simultaneous minimization of atomic positions and unit cell parameters at a fixed isotropic pressure to acquire data points along the energy-volume curve. However, as shown in Fig. S5, pressure induces phase transitions which prevent the use of such an approach. Next, we moved to the simultaneous optimization of atomic positions and unit cell parameters while maintaining

a constant volume. To avoid phase transitions as volume is changed, we found that the most reliable workflow is as follows. We started with the experimental structure of each form. We relaxed the structure keeping the volume fixed. We continued with a systematic sweep, incrementally increasing or decreasing the volume by 5 \AA^3 from the previous step and relaxing the structure with a fixed volume. We performed the procedure until we reached our desired upper or lower limit for volume.

We fit the energy-volume data to the Birch-Murnaghan equation of state (EOS). Given the complexity and extremely anisotropic thermal expansion of our system, the energy-volume curve is not as smooth as for typical materials. This means that in some situations the EOS fit does not correspond to the minima of the energy-volume data and depends on the range of considered volumes. For this reason, we have used two different EOS fits that can be seen in Fig. 3 for form III and in Figs. S6 and S7 for forms I and II. In one case, we

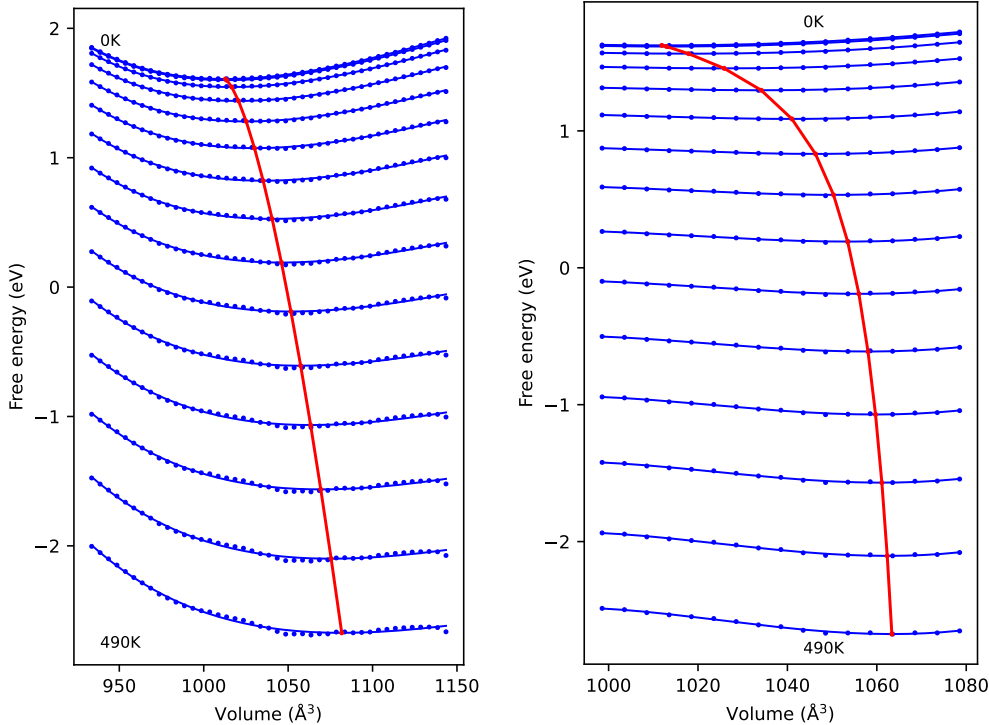


Figure 3: The QHA calculation for form III where the exploration of the energy-volume curve starts from the experimental volume. The top panel covers a full range of volumes from 940 \AA^3 to 1140 \AA^3 , bottom panel is the constrained range (1000 to 1080 \AA^3) to fit the EOS.

use the full range of volumes, and in the other case, we constrain the range of volumes so that EOS is fitted only on points close to the minimum. Although the qualitative conclusions are similar, this choice has an impact on the results. For example, the impact of this difference on free energy can be seen by comparing Fig. 2 for the full range of volumes to Fig. S10 for the constrained range of volumes. Figures in the main text correspond to the EOS fit for the whole range of volumes.

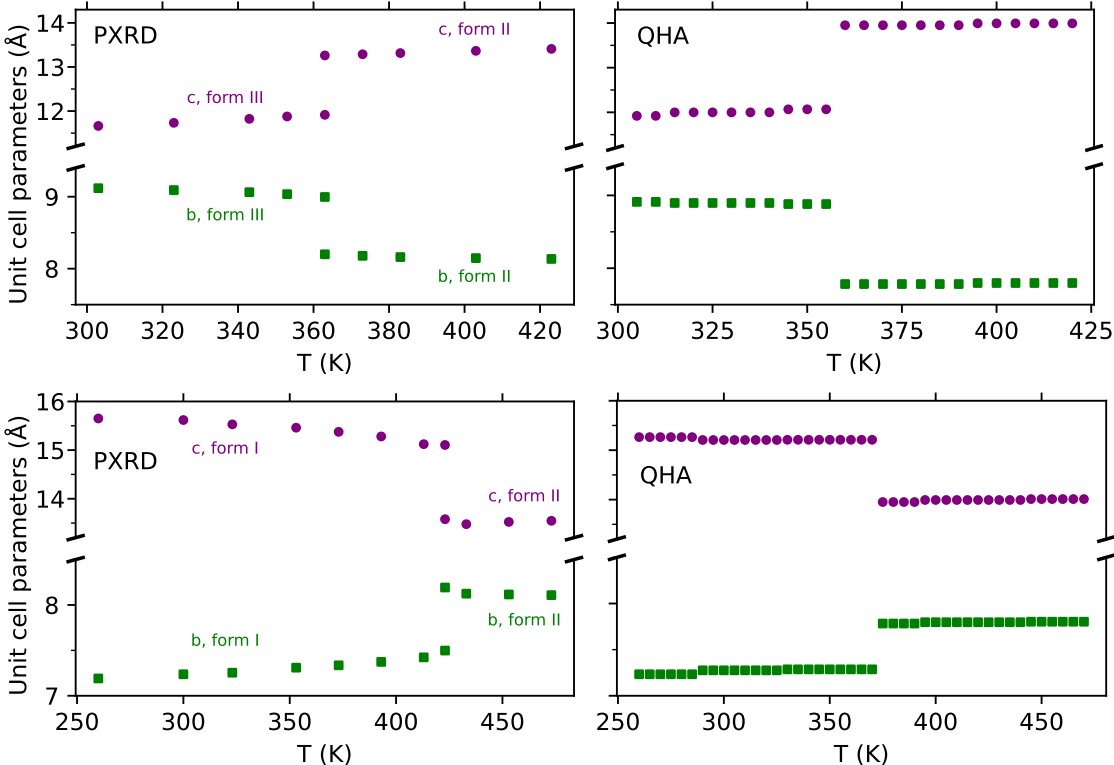


Figure 4: Change in the unit cell parameters during a phase transition between form III and form II (top panels) and form I and form II (bottom panels). The left panels show experimental PXRD values,⁹ and the right panels show QHA calculations using the MLIP.

Qualitatively, Gibbs free energies that include anharmonic corrections on the level of QHA are similar to Helmholtz free energies from HA. However, predicted phase transition temperatures are higher in QHA than in HA. In the following, we list values obtained with EOS fitted to the whole range of volumes while values corresponding to the constrained volume range are listed in parentheses. Form I has an enthalpy of 1.41 (1.02) kJ/mol higher than form III and 1.28 (1.96) kJ/mol lower than form II. As form II has the highest entropy,

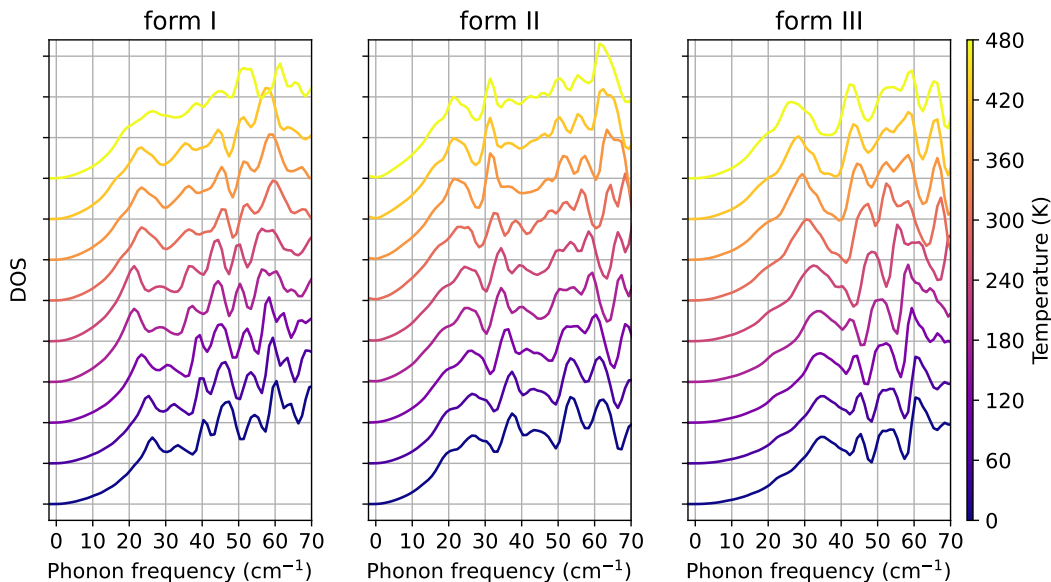


Figure 5: Temperature dependence of phonon DOS for forms I, II and III.

at the temperature of 374 (310) K, Gibbs free energy becomes lower for form II than for form I and phase transition takes place which can be well connected with experimentally observed phase transitions. The experimental phase transition temperature from differential scanning calorimetry (DSC) is 423 K.⁹ On cooling form II, due to the difference in enthalpy, Gibbs free energy of form III becomes the lowest at $T=360$ (325) K, and form III is the most stable at low temperatures. The experimental phase transition temperature from form II to form III from DSC is 351 K in the cooling run and 359 K in the heating run, in excellent agreement with the prediction. Upon further heating and cooling, there is a reversible transformation between forms II and III in agreement with experiments. Considering that state-of-the-art DFT calculations with currently best performing functional can have errors in predicting polymorph phase transition temperature larger than 100 K,²⁸ our results are better than one would expect.

All phase transitions are isomorphous and characterized by large changes in lattice parameters. We extracted a unit cell for each temperature that has the lowest Gibbs free energy and compared it to unit cell parameters as a function of temperature that are measured by

powder x-ray diffraction (PXRD). As can be seen in Fig. 4 our model can reproduce well temperature-dependent PXRD measurements. PXRD values close to the phase transition temperatures are not reliable as phases are mixed at these temperatures. Our MLIP slightly overestimates the c parameter and underestimates the b parameter of a unit cell of form II, while cell parameters of low-temperature forms are well predicted. More importantly, the studied system is characterized by negative linear thermal expansion along one of the axes (c in form I, and b in forms II and III). Our model predicts negative thermal expansion in forms I and III in agreement with experiments, and a small positive thermal expansion in form II.

Thermosalient phase transitions often have been connected to the softening of phonon modes.^{13,29} In Fig. 5, we show low-frequency harmonic phonons as a function of temperature for the three forms. Harmonic phonons have been calculated for the structure that corresponds to the minimum of Gibbs free energy at a given temperature. All phonon modes in all three forms are positive for all temperatures, confirming the dynamical stability of all three forms. As expected, most of the phonon frequencies decrease as the temperature (volume) increases. For form II and III frequencies generally decrease for 5–10 cm^{-1} in the range of temperatures of 0–480 K. Interestingly, for form I, frequencies decrease for temperatures 0–240 K but seem to change to a slightly higher ($\sim 3 \text{ cm}^{-1}$) frequencies for temperatures 300–480 K.

While the above results explain thermodynamical reasons for phase transitions, they do not explain the reasons for the thermosalient phenomenon.

To elucidate it, we have performed climbing-image solid state nudged elastic band (SSNEB) calculations as implemented in Transition State Library for ASE (tsase).^{30–32} As shown in Fig. 6, the minimum energy path between forms I and III (through form II) was calculated on 28 SSNEB images. We have found that the generalized forces in SSNEB almost exclusively correspond to the (scaled) stress. This means that the forces on all atoms along the phase transition path are lower than our structural minimization threshold as shown in

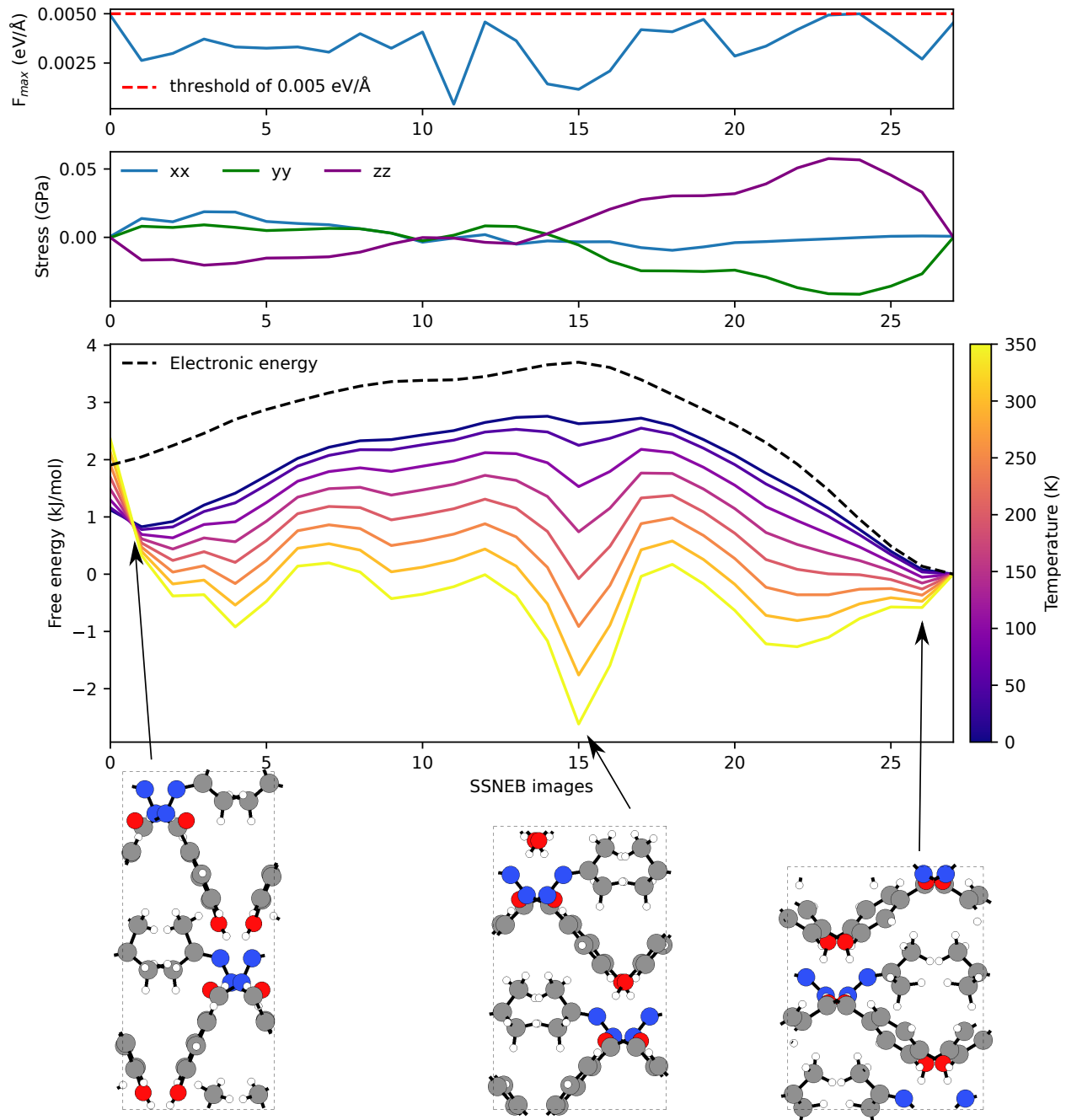


Figure 6: SSNEB results. Top panel: Maximum force on any atom for all SSNEB images. Middle panel: Stress on each unit cell along the minimum energy path. Bottom panel: Helmholtz free energy along the reaction path for different temperatures normalized to the free energy of form III. Crystal structures of Form I (image 1), Form II (image 15) and Form III (image 26) are also shown.

the top panel of Fig. 6. This allowed us to perform HA phonon calculations along the phase transition path to obtain the corresponding free energy as shown in the bottom panel of

Fig. 6. The black dashed line shows the direct output of SSNEB. With zero point energy added, for form I, the most stable image at $T=0$ is not the original one, but the next image. For both form I and form III, as the temperature increases, due to the thermal expansion, the minimum shifts to adjacent images. Image 15, which corresponds to form II, has the highest energy in SSNEB showing again that it is not stable at low temperatures and can convert to low-temperature forms without an energy barrier. However, as the temperature increases its Helmholtz free energy becomes the lowest, and it is the most stable form at high temperatures as discussed above.

Interestingly, at the transition temperature when the free energies of form II and forms I or III are equal, there is a transition energy barrier that amounts to 1.2-1.6 kJ/mol. A transition energy barrier exists also for higher temperatures. That means that additional energy is needed for the phase transition to take place. On the other hand, once the transition state is reached, phase transition can proceed fueled by the energy difference between the transition state and the stable polymorph. As mentioned above, along the SSNEB path forces on atoms are low, below our minimization threshold of 0.005 eV/Å, and SSNEB generalized forces that constitute the transition energy barrier are made of stress of the order of 0.02-0.05 GPa. This means that upon phase transition, the energy released from overcoming the barrier will be in the form of mechanical energy and the existing stress will be relaxed giving rise to the mechanical response that manifests as a thermosalient effect. The kinetics of this process is beyond the scope of this communication but we plan to study it with molecular dynamics using our machine learning potential. Our results explain why the thermosalient transitions are described as non-displacive (martensitic), concerted, first-order phase transformations.³³

In conclusion, we developed the first MLIP for a thermosalient molecular crystal. We choose N-2-Propylidene-4hydroxybenzohydrazide which shows one of the largest mechanical responses during thermosalient phase transition which is preceded by one of the largest negative thermal expansion. Our MLIP has sub-meV/atom accuracy closely resembling the

baseline r²SCAN+D4 DFT calculations. Our Gibbs free energy calculations explain the observed irreversible transition from form I to form II and reversible transitions from form II and form III. The transition temperature for the reversible transformation between forms II and III $T=87$ (52) °C is in remarkable agreement with experiments $T=78 - 86$ °C. Transition from form I to form II is predicted at 101 (37) °C in comparison to the experimental value of 150 °C. Still, this is considered as state-of-the-art accuracy.²⁸ Our MLIP model is also able to predict negative linear expansion for forms I and III in agreement with experimental observations. We provide a first direct calculation of the thermally induced stress that drives the thermosalient effect by calculating temperature-dependent free energies along the transition path. Finally, our openly available MLIP can be used for other types of calculations studying this system to deepen understanding of the thermosalient effect. On our side, we plan to separately publish a detailed study of the dynamics and kinetics of the thermosalient transition. Our deposited DFT dataset can be also used to benchmark more advanced MLIP architectures.

Acknowledgements

We acknowledge fruitful discussions with Željko Skoko and Jasminka Popović. This work has been supported by Croatian Science Foundation under the project UIP-2020-02-5675 and the European Regional Development Fund within the “Center of Excellence for Advanced Materials and Sensing Devices” (Grant No. KK.01.1.1.01.0001).

Data Availability Statement:

The machine learning model and its hyperparameters, training dataset, QHA calculation scripts, and transition path coordinates are openly available in Zenodo at [10.5281/zenodo.11504754](https://doi.org/10.5281/zenodo.11504754).

Supporting Information Available:

Supporting information includes VASP input file and figures on dataset construction (Figs. S1-S3), hyperparameter optimization (Fig. S4), QHA calculations (Figs. S5-S7), HA phonon DOS (Fig. S8), and free energies (Figs. S9-S10) (PDF).

References

- (1) Unke, O. T.; Chmiela, S.; Sauceda, H. E.; Gastegger, M.; Poltavsky, I.; Schütt, K. T.; Tkatchenko, A.; Müller, K.-R. Machine Learning Force Fields. *Chemical Reviews* **2021**, *121*, 10142–10186.
- (2) Clements, R. J.; Dickman, J.; Johal, J.; Martin, J.; Glover, J.; Day, G. M. Roles and opportunities for machine learning in organic molecular crystal structure prediction and its applications. *MRS Bulletin* **2022**, *47*, 1054–1062.
- (3) Hoja, J.; Reilly, A. M.; Tkatchenko, A. First-principles modeling of molecular crystals: structures and stabilities, temperature and pressure. *WIREs Computational Molecular Science* **2017**, *7*, e1294.
- (4) Musil, F.; De, S.; Yang, J.; Campbell, J. E.; Day, G. M.; Ceriotti, M. Machine learning for the structure–energy–property landscapes of molecular crystals. *Chem. Sci.* **2018**, *9*, 1289–1300.
- (5) Kapil, V.; Engel, E. A. A complete description of thermodynamic stabilities of molecular crystals. *Proceedings of the National Academy of Sciences* **2022**, *119*, e2111769119.
- (6) Žugec, I.; Geilhufe, R. M.; Lončarić, I. Global machine learning potentials for molecular crystals. *The Journal of Chemical Physics* **2024**, *160*, 154106.
- (7) Yamanoi, Y.; Omoto, K.; Nakae, T.; Nishio, M. *Soft Crystals: Flexible Response Sys-*

- tems with High Structural Order*; Springer Nature Singapore Singapore, 2023; pp 131–153.
- (8) Commins, P.; Desta, I. T.; Karothu, D. P.; Panda, M. K.; Naumov, P. Crystals on the move: mechanical effects in dynamic solids. *Chem. Commun.* **2016**, *52*, 13941–13954.
- (9) Lončarić, I.; Popović, J.; Despoja, V.; Burazer, S.; Grgicevic, I.; Popovic, D.; Skoko, v. Reversible Thermosalient Effect of N'-2-Propylidene-4-hydroxybenzohydrazide Accompanied by an Immense Negative Compressibility: Structural and Theoretical Arguments Aiming toward the Elucidation of Jumping Phenomenon. *Crystal Growth & Design* **2017**, *17*, 4445–4453, Publisher: American Chemical Society.
- (10) Centore, R.; Jazbinsek, M.; Tuzi, A.; Roviello, A.; Capobianco, A.; Peluso, A. A series of compounds forming polar crystals and showing single-crystal-to-single-crystal transitions between polar phases. *CrystEngComm* **2012**, *14*, 2645–2653.
- (11) Panda, M. K.; Centore, R.; Causà, M.; Tuzi, A.; Borbone, F.; Naumov, P. Strong and anomalous thermal expansion precedes the thermosalient effect in dynamic molecular crystals. *Scientific reports* **2016**, *6*, 29610.
- (12) Zaczek, A. J.; Catalano, L.; Naumov, P.; Korter, T. M. Mapping the polymorphic transformation gateway vibration in crystalline 1,2,4,5-tetrabromobenzene. *Chem. Sci.* **2019**, *10*, 1332–1341.
- (13) Zakharov, B. A.; Michalchuk, A. A.; Morrison, C. A.; Boldyreva, E. V. Anisotropic lattice softening near the structural phase transition in the thermosalient crystal 1, 2, 4, 5-tetrabromobenzene. *Physical Chemistry Chemical Physics* **2018**, *20*, 8523–8532.
- (14) Kresse, G.; Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **1996**, *54*, 11169–11186.

- (15) Kresse, G.; Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B* **1999**, *59*, 1758–1775.
- (16) Furness, J. W.; Kaplan, A. D.; Ning, J.; Perdew, J. P.; Sun, J. Accurate and Numerically Efficient r2SCAN Meta-Generalized Gradient Approximation. *The Journal of Physical Chemistry Letters* **2020**, *11*, 8208–8215.
- (17) Smith, J. S.; Isayev, O.; Roitberg, A. E. ANI-1, A data set of 20 million calculated off-equilibrium conformations for organic molecules. *Scientific Data* **2017**, *4*, 170193, Number: 1 Publisher: Nature Publishing Group.
- (18) Larsen, A. H.; Mortensen, J. J.; Blomqvist, J.; Castelli, I. E.; Christensen, R.; Dułak, M.; Friis, J.; Groves, M. N.; Hammer, B.; Hargus, C., et al. The atomic simulation environment—a Python library for working with atoms. *Journal of Physics: Condensed Matter* **2017**, *29*, 273002.
- (19) Devereux, C.; Smith, J. S.; Huddleston, K. K.; Barros, K.; Zubatyuk, R.; Isayev, O.; Roitberg, A. E. Extending the applicability of the ANI deep learning molecular potential to sulfur and halogens. *Journal of Chemical Theory and Computation* **2020**, *16*, 4192–4202.
- (20) Gao, X.; Ramezanghorbani, F.; Isayev, O.; Smith, J. S.; Roitberg, A. E. TorchANI: A free and open source PyTorch-based deep learning implementation of the ANI neural network potentials. *Journal of chemical information and modeling* **2020**, *60*, 3408–3415.
- (21) Schran, C.; Brezina, K.; Marsalek, O. Committee neural network potentials control generalization errors and enable active learning. *The Journal of Chemical Physics* **2020**, *153*.
- (22) Batzner, S.; Musaelian, A.; Sun, L.; Geiger, M.; Mailoa, J. P.; Kornbluth, M.; Molinari, N.; Smidt, T. E.; Kozinsky, B. E (3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nature communications* **2022**, *13*, 2453.

- (23) Ehlert, S.; Huniar, U.; Ning, J.; Furness, J. W.; Sun, J.; Kaplan, A. D.; Perdew, J. P.; Brandenburg, J. G. r^2 SCAN-D4: Dispersion corrected meta-generalized gradient approximation for general chemical applications. *The Journal of Chemical Physics* **2021**, *154*, 061101.
- (24) Moellmann, J.; Grimme, S. DFT-D3 Study of Some Molecular Crystals. *The Journal of Physical Chemistry C* **2014**, *118*, 7615–7621.
- (25) Togo, A.; Chaput, L.; Tadano, T.; Tanaka, I. Implementation strategies in phonopy and phono3py. *Journal of Physics: Condensed Matter* **2023**,
- (26) Togo, A. First-principles phonon calculations with phonopy and phono3py. *Journal of the Physical Society of Japan* **2023**, *92*, 012001.
- (27) Togo, A.; Chaput, L.; Tanaka, I.; Hug, G. First-principles phonon calculations of thermal expansion in Ti_3SiC_2 , Ti_3AlC_2 , and Ti_3GeC_2 . *Physical Review B* **2010**, *81*, 174301, Publisher: American Physical Society.
- (28) Firaha, D.; Liu, Y. M.; van de Streek, J.; Sasikumar, K.; Dietrich, H.; Helfferich, J.; Aerts, L.; Braun, D. E.; Broo, A.; DiPasquale, A. G., et al. Predicting crystal form stability under real-world conditions. *Nature* **2023**, *623*, 324–328.
- (29) Ko, J.-H.; Lee, K.-S.; Sahoo, S. C.; Naumov, P. Isomorphous phase transition of 1, 2, 4, 5-tetrabromobenzene jumping crystals studied by Brillouin light scattering. *Solid state communications* **2013**, *173*, 46–50.
- (30) Sheppard, D.; Xiao, P.; Chemelewski, W.; Johnson, D. D.; Henkelman, G. A generalized solid-state nudged elastic band method. *The Journal of chemical physics* **2012**, *136*.
- (31) Xiao, P.; Henkelman, G. Communication: From graphite to diamond: Reaction pathways of the phase transition. *The Journal of chemical physics* **2012**, *137*.

- (32) Xiao, P.; Cheng, J.-G.; Zhou, J.-S.; Goodenough, J. B.; Henkelman, G. Mechanism of the CaIrO₃ post-perovskite phase transition under pressure. *Physical Review B* **2013**, *88*, 144102.
- (33) Park, S. K.; Diao, Y. Martensitic transition in molecular crystals for dynamic functional materials. *Chem. Soc. Rev.* **2020**, *49*, 8287–8314.

Supporting Information: Thermosalient phase transitions from machine learning interatomic potential

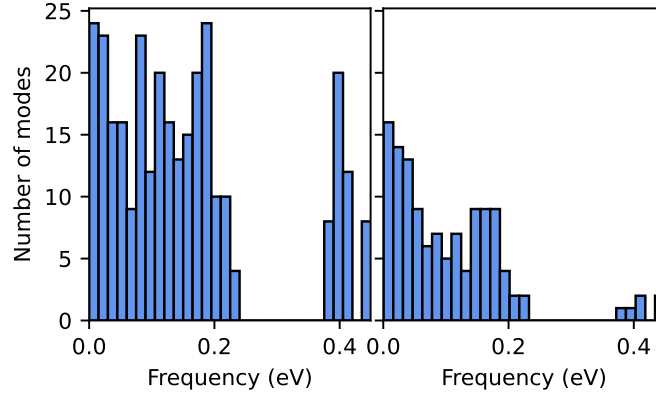


Figure S1: Distribution of vibrational modes for form I. Left: unfiltered distribution; right: filtered distribution of modes selected to be included in the initial dataset.

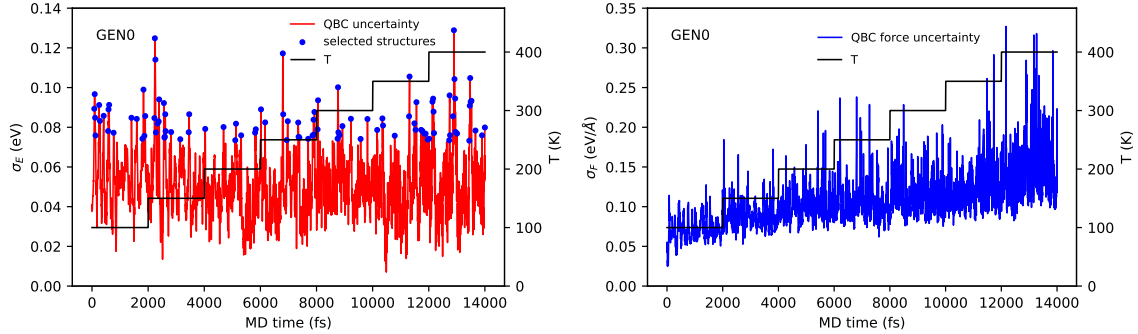


Figure S2: QBC selection of additional training points for GEN0 model. Left panel shows uncertainty in energy and selected structures based on energy criteria, right panel shows uncertainty in forces and selected structures based on force criteria.

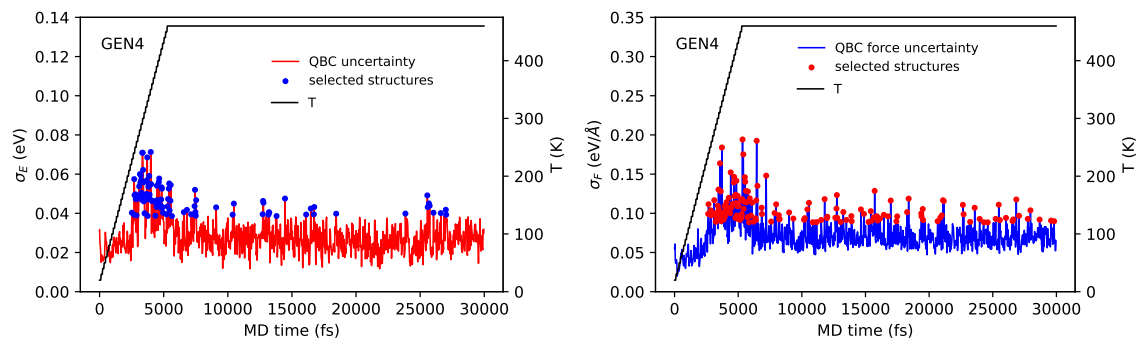


Figure S3: QBC selection of additional training points for GEN4 model. Left panel shows uncertainty in energy and selected structures based on energy criteria, right panel shows uncertainty in forces and selected structures based on force criteria.

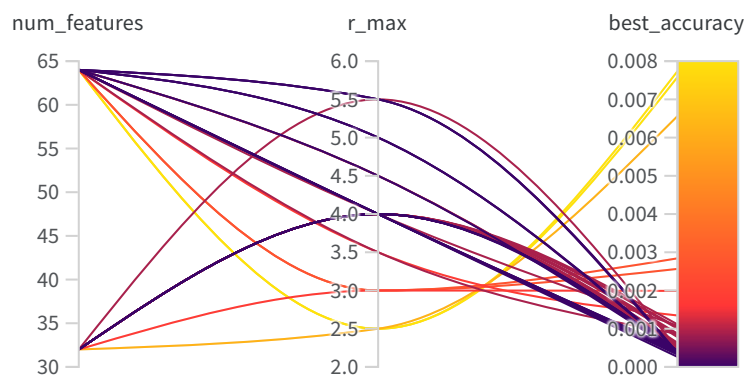


Figure S4: Parallel coordinates plot of hyperparameter optimization showing the influence of number of features N_F (num_features) parameter and cutoff radius r_c (r_max) parameter to final model accuracy (lower best_accuracy means more accurate model). To obtain a good model it is necessary to have cutoff radius of at least $r_c = 4 \text{ \AA}$.

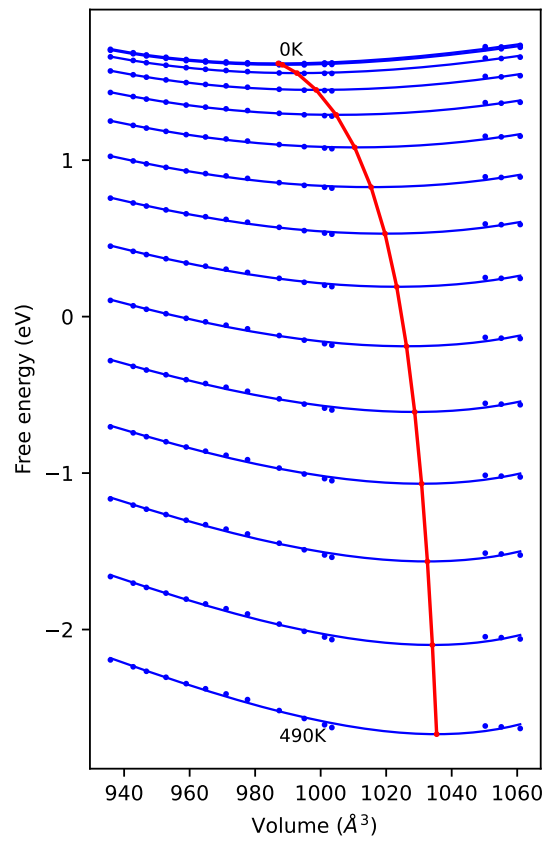


Figure S5: QHA calculation for form I where the exploration of the energy-volume curve was done with minimisation under incrementally changing pressure.

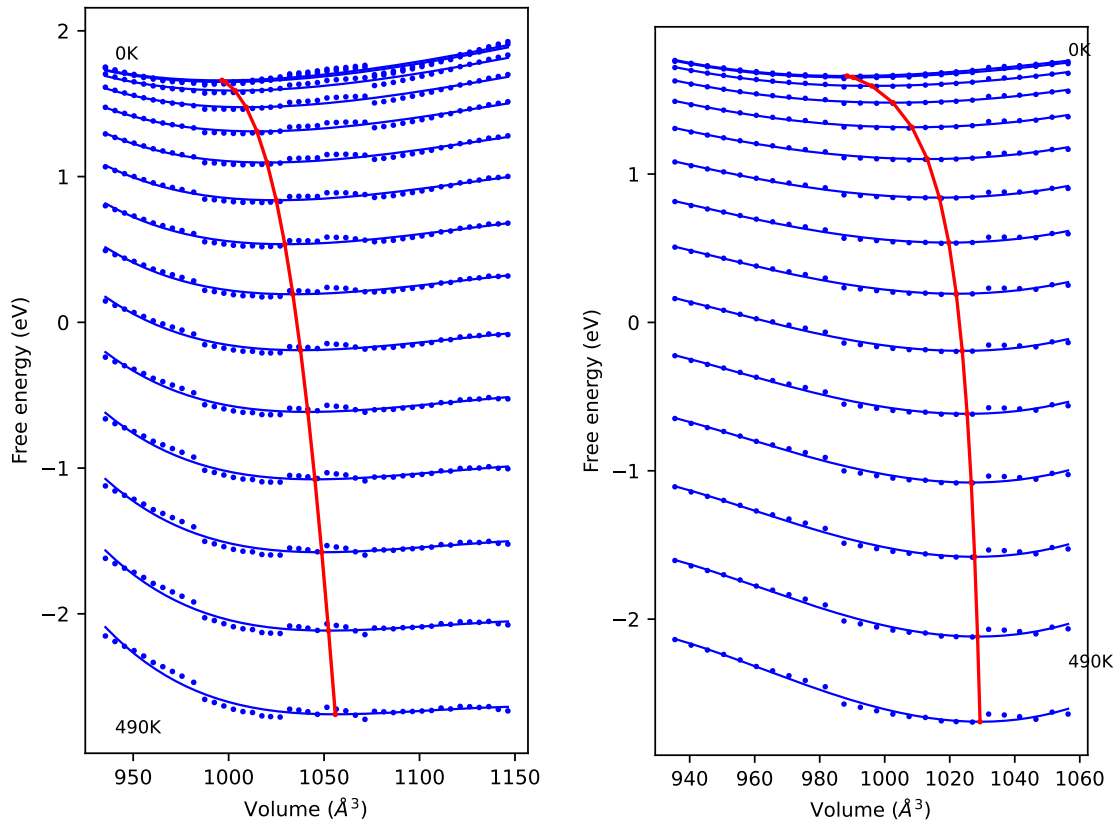


Figure S6: The QHA calculation for form I where the exploration of the energy-volume curve starts from the experimental volume. Left panel covers a full range of volumes from 940 Å to 1140 Å, right panel is the constrained range to fit the EOS close to minima.

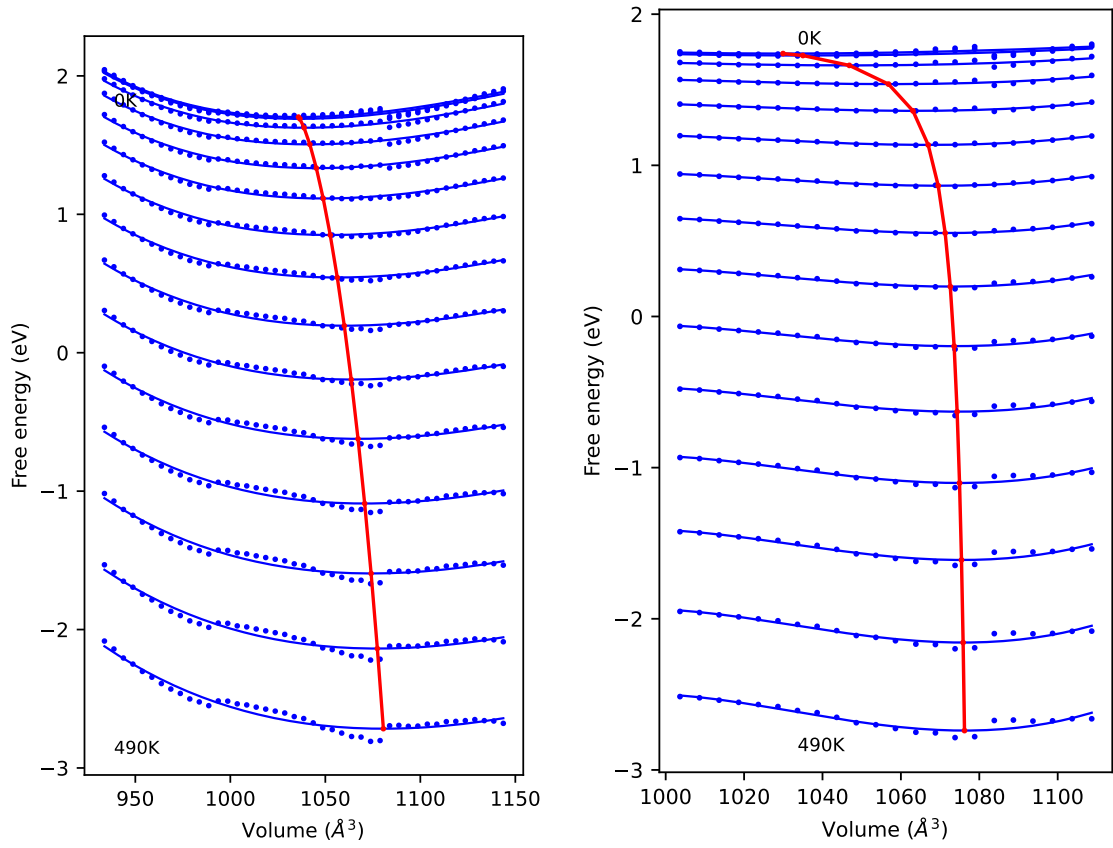


Figure S7: The QHA calculation for form II where the exploration of the energy-volume curve starts from the experimental volume. Left panel covers a full range of volumes from 940 \AA to 1140 \AA , right panel is the constrained range to fit the EOS close to minima.

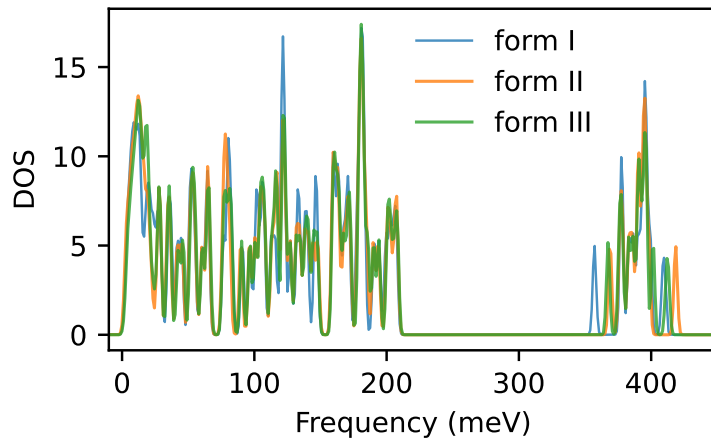


Figure S8: Phonon DOS for the three forms calculated in the HA.

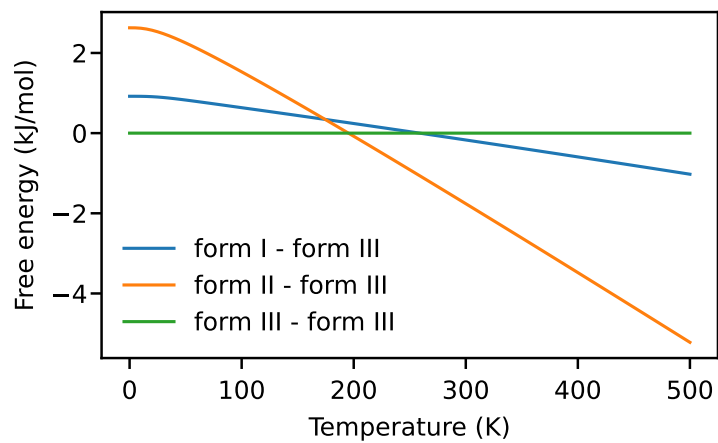


Figure S9: Helmholtz free energy for the three forms calculated with HA.

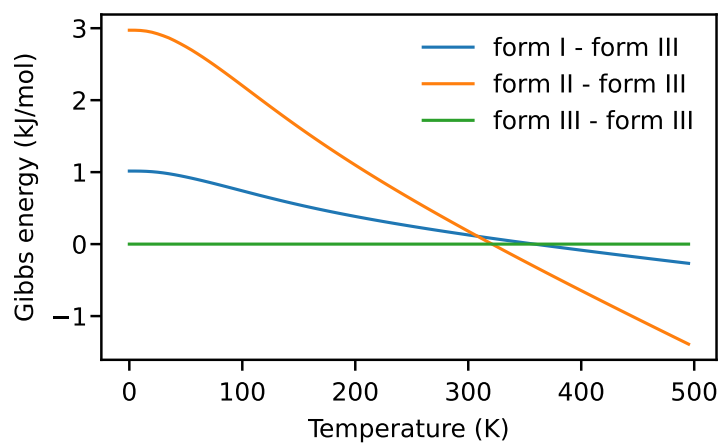


Figure S10: Free energy for the three forms calculated with QHA, where the range of the energy-volume curve has the volume range modified to fit of EOS around the minimum.

VASP input INCAR file for DFT data generation

NCORE = 4

ISMEAR = 0

SIGMA = 0.03

EDIFF = 1E-6

ENCUT = 440

PREC = Accurate

KSPACING = 0.2

METAGGA = R2SCAN

LASPH = .TRUE.

LMIXTAU = .TRUE.

LWAVE = .FALSE.

LCHARG = .FALSE.

TOC Graphic

