ACS OMEGA

Open Access

http://pubs.acs.org/journal/acsodf

Article

Revisiting Aspirin Polymorphic Stability Using a Machine Learning Potential

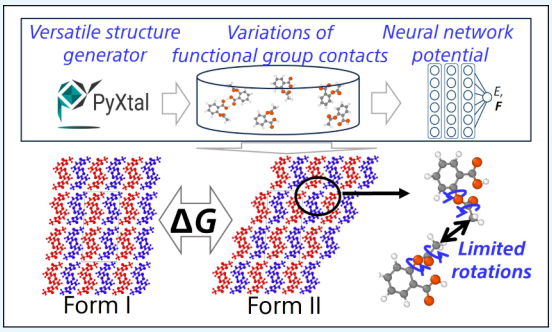
Shinnosuke Hattori* and Qiang Zhu*

Cite This: *ACS Omega* 2024, 9, 36589−36599

Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: In this study, we present a systematic computational investigation to analyze the long-debated free energy stability of two well-known aspirin polymorphs, denoted as Form I and Form II. Specifically, we developed a strategy to collect training configurations covering diverse interatomic interactions between representative functional groups in aspirin crystals. Utilizing a state-of-the-art neural network interatomic potential (NNIP) model, we trained an accurate machine learning potential to simulate aspirin crystal dynamics under finite temperature conditions with ∼0.46 kJ/mol/molecule accuracy. Employing the trained NNIP model, we performed thermodynamic integration to assess the free energy difference between aspirins I and II, accounting for the anharmonic effects in a large supercell consisting of 512 molecules. For the first time, our results convincingly demonstrated that Form I is more stable than Form II at 300 K, ranging from 0.74 to 1.83 kJ/mol/molecule, aligning with experimental observations. Unlike the majority of previous simulations based on (quasi)harmonic approximations in a small super cell, which often found degenerate energies between aspirins I and II, our findings underscore the importance of anharmonic effects in determining polymorphic stability ranking. Furthermore, we proposed the use of the rotational degrees of freedom of methyl and ester/phenyl groups in aspirin crystals as characteristic motions to highlight rotational entropic contribution that favors the stability of Form I. From the structural perspective, we also found that the subtle free energy difference can be used to explain the distinct thermal expansion responses as observed in both experimental and simulation data. Beyond the aspirin polymorphism, we anticipate that such entropy-driven stabilization can be broadly applicable to many other organic systems, suggesting that our approach holds great promise for stability studies in small-molecule drug design.



INTRODUCTION

The stability of polymorphs, characterized by free energy differences, remains a critical issue in the study of organic crystals.[1,2] To make reliable predictions, one aims for a lattice energy precision well below the threshold of chemical accuracy (4 kJ/mol, about 43 meV), which is necessary to distinguish between competing polymorphs. The stability of polymorphs is essential for many applications, including the control of drug solubility,[3] new drug and materials development,[4,5] and the advancement of computational chemistry methods.[6−10] The identification of new polymorphs that undergo phase selection under external environmental influences such as temperature and pressure is crucial for understanding synthesis and deposition processes.

In recent years, aspirin's polymorphism has been the subject of intensive studies due to the popularity of aspirin in practical applications,[11−16] as well as contrasting outcomes observed between experimental results and computational models.[7,8,17] Form I is by far the most easily selected polymorph of aspirin crystals, whereas Form II, although structurally very similar to I, requires experimental ingenuity to synthesize it. From the perspective of crystal packing, both forms exhibit a layered structure characterized by densely packed aspirin dimers at the (100) plane,[9,10] as illustrated in Figure 1. Despite their structural resemblance, the primary distinctions between the two forms are how the layers are stacked and the orientation of aspirin dimer pairs within each layer. In both forms, there exist two types of molecular orientations (colored in red and blue in Figure 1). In Form I, the red and blue dimers are alternatively replicated along the (001) direction, whereas Form II can be considered as a shear-slip modulation of Form I, with a lateral shift of each aspirin dimer layer along the (100)[001] direction, resulting in an alternative layer stacking arrangement. Consequently, these subtle yet non-negligible differences in the packing and orientation of aspirin dimers between Forms I and II lead to

Received: May 21, 2024
Revised: July 26, 2024
Accepted: July 30, 2024
Published: August 19, 2024

https://doi.org/10.1021/acsomega.4c04782
ACS Omega 2024, 9, 36589−36599

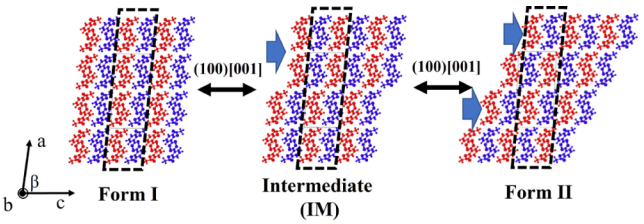


Figure 1. Schematic representations of aspirin Forms I and II, as well as the intermediate structure viewed from the projection of the ac plane. For convenience, a schematic transition path between I and II via the shear deformation along the (100)[001] direction is also depicted. In all structures, molecules are colored in red and blue to distinguish their orientations.

distinct thermodynamic behaviors, underscoring the complexity of polymorph stability in pharmaceutical compounds.

Despite the fact that Form I has been well-known for over a century, the successful synthesis of Form II has been reported only in the past two decades.[6,11−14] Such a contrast naturally raises a prevailing question in the field of aspirin research: why is Form I more commonly preferred under the standard experimental conditions? In recent years, multiple experiments have shown that Form II can be induced through a variety of external conditions, including the cooling rate, choice of solvent,[13] pressure,[14] indentation,[15,18] and laser irradiation.[19] These observations underscore the necessity for a thorough evaluation of the thermodynamic properties of aspirin crystals, pointing toward the influence of external factors on polymorph stability. Such insights are crucial for advancing our understanding of aspirin polymorphism and may have broader implications for the study of other pharmaceutical compounds.

To elucidate the observed difference in experiments, substantial computational efforts has been undertaken to quantify the energy differences between aspirin I and II.[7−10,17,20] In these studies, the interlayer hydrogen bonding interactions of methyl groups were often considered as the main differences between Forms I and II.[8,13,20] To date, most state-of-the-art electronic structure methods (e.g., RI_MP2,[7] MP2C,[17] PBE+MBD,[8] B86bPBE-XDM[9]) have found a lattice energy difference of less than 0.3 kJ/mol between the two polymorphs. When zero-point vibrational energies are included, the energy different decreases by 0.1 kJ/mol.[9] Given that the energy difference is comparable to a thermal vibration at room temperature, incorporating finite temperature effects becomes crucial. Indeed, a computational study based on the harmonic phonon approximation using PBE+MBD found that Form I's stability is enhanced by a unique low-frequency vibration mode around at 37 cm^{-1} related to a mixture of motions of whole molecules with concerted motions of methyl groups from different molecules in and out of phase with one another.[8] However, several more recent studies based on different methods have reported nearly indistinguishable free energy differences between two polymorphs, after considering harmonic, quasiharmonic, or anharmonic phonon contributions on a small super cell.[9,20] Even with the same PBE+MBD functional, it was found that both forms I and II possess similar vibrational modes at around 35 cm^{-1}.[21] On the contrary, the experimental evidence supporting the existence of low-frequency phonon modes in both forms remains contentious.[20,22] The disparity naturally leads to a speculation that Form I is favored due to other factors.

Considering an organic crystal under room temperature conditions, it is reasonable to expect that the anharmonic effects may become pronounced in determining the polymorph's free energy ranking. A standard approach to evaluate the free energy difference of two solid polymorphs is thermodynamic integration (TI)[23] based on molecular dynamics (MD) simulation in conjunction with the DFT method.[24] Although such a DFT-MD method has been employed to predict aspirin's Raman spectrum within a small supercell,[21] the cost of DFT-MD calculations on the large supercell remains heavy and limits the scope of comprehensive thermodynamic analysis of aspirin crystals, which cannot be surpassed by similar approaches, especially for free energy evaluations.

Encouragingly, the recent development of neural network interatomic potentials (NNIPs) has achieved significant progress in balancing accuracy and computational efficiency.[25−29] These NNIPs handle E(3) equivariants and enhance accuracy even when training data are sparse, by preserving identity through symmetric operations in 3D space. Among them, the Allegro package has demonstrated remarkable scalability and successfully simulated large systems with over 100 000 atoms with the aid of large GPU arrays.[30,31] This breakthrough encouraged us to explore the use of NNIPs to study the stability of aspirin polymorphs via high-fidelity MD simulations. Our initial step involved the development of accurate interatomic potentials tailored for aspirin polymorphs. Following this, we utilized Allegro to train an NNIP model to achieve an optimal balance between simulation accuracy and computational efficiency for organic crystals. Finally, we employed TI techniques to calculate the free energy differences between Forms I and II and quantified the contributions that are related to anharmonic effects. As elucidated in subsequent sections, our findings reveal that the developed framework not only is effective for this specific study but also offers a generalizable approach for the analysis of other organic crystals.

RESULTS AND DISCUSSIONS

NNIP Model Training and Validation. To obtain an accurate NNIP model, we began by preparing three sets of training data as follows. First, we performed a series of MD simulations using the general-purpose Allegro model that were pretrained on the SPICE[32] data set for the known aspirin forms (including Forms I and II and another recently identified metastable Form IV[16]) at 300 K for a duration of 10 ps. From these MD trajectories, 1200 representative structures (400 structures per polymorph) were selected from data set A that represents the three most important energy basins around the experimentally identified polymorphs. Second, to explore more crystal configurations that are inaccessible through direct room temperature MD simulations of the known aspirin polymorphs, we developed an automated computational pipeline to harness low-energy hypothetical crystal structures using the PyXtal code[33] based on an evolutionary crystal structure prediction (CSP) algorithm.[34−36] We note that a similar workflow based on PyXtal has been demonstrated in a recent study on GaN,[37] and the use of PyXtal was also integrated for automated NNIP training of inorganic materials[38] and crystal structure prediction.[39] In this work, we performed a systematic crystal packing search on 10 common space groups with one aspirin molecule in the asymmetric unit ($Z' = 1$) within 200 generations and 100 structures for each generation, based on the classical General Amber Force Field (GAFF).[40] Furthermore, we removed the duplicate structures from the identified CSP

https://doi.org/10.1021/acsomega.4c04782
ACS Omega 2024, 9, 36589−36599

structures and refined their energy ranking using the ANI-2x potential,[41] resulting in 4387 low-energy CSP structures in data set B to represent a diverse set of energy minima. Finally, we also chose the 30 lowest-energy hypothetical CSP structures to run 5 ps MD simulations and sampled a total of 1200 configurations (40 configurations per MD simulation). These configurations were labeled as data set C to represent 30 low-energy basins in the potential energy surface of the aspirin crystal.

In parallel, we performed additional MD simulation at 350 K for the intermediate structures (IM) as depicted in Figure 1 at 350 K. This results in 200 configurations for data set D, which aims to explore the transition states between aspirin I and II. Unlike the aforementioned data sets, this data set will not be seen during training but will be used solely during the validation stage.

It is well-known that the quality of the machine learning model crucially depends on the representation capability of training data. In computational chemistry, the primary molecular interactions in a crystal are generally believed to be contributed by the interactions between the representative functional groups. In the case of aspirin crystals, the most important functional groups are carboxylic acid (Ca), methyl group (Me), and phenyl group (Ph). Therefore, we proposed the use of distance distributions between function groups within the aspirin crystals to infer the coverage of our training data. Figure 2 shows the comparison of distributions of three most

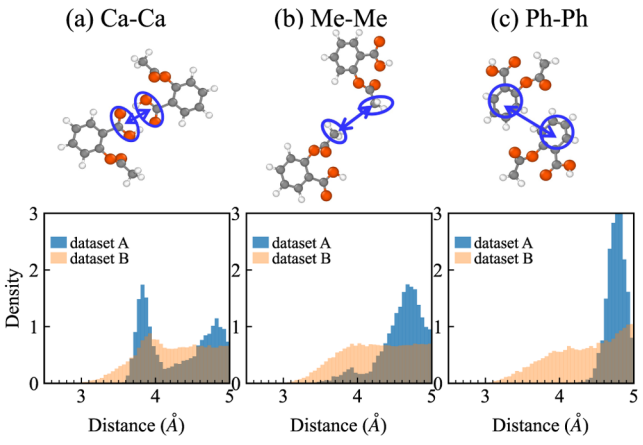


Figure 2. Representative intermolecular contacts between functional groups, including (a) carboxylic acid−carboxylic acid, (b) methyl−methyl, and (c) phenyl−phenyl. The bottom panels show the distance distributions for each contact for data sets A (blue) from MD simulation and B (orange) from crystal structure prediction.

important functional interactions (including Ca−Ca, Me-Me, and Ph-Ph) from both data sets A and B. Clearly, Figure 2 suggests that data set A, derived from MD simulations, has a narrower distribution peak than the CSP data set B. This quantitative comparison demonstrates that it is necessary to use a more efficient structure sampling approach prior to the NNIP model training.

For each configuration in the collected data sets, we performed single-point energy calculation to obtain the reference energy, forces, and stress tensors, with the combination r2SCAN functional and D4 correction for the van der Waals (vdW) correction.[42] Within the framework of the Allegro code,[28] we trained the NNIP models with different choices of training data and strategies (see more details about the choices of hyperparameters in the Experimental Section).

Here, we trained four different models from (i) to (iv) using data sets A, B, and C with or without SPICE pretraining. The model (i) was trained on data set A only, (ii) was trained on data set A after pretraining on the SPICE data set, (iii) was trained on data sets A, B, and C, and (iv) was trained on data sets A, B, and C after pretraining on the SPICE data set. The accuracy of the root-mean-square error (RMSE) was evaluated using a test data set (10% of the total) separated from the training data set and is summarized in Table 1.

As shown in Table 1, we found that model (iv) delivered the most favorable outcomes, achieving the RMSEs of 0.12 kJ/mol/molecule for energy, 3.92 kJ/mol/Å for force, and 0.06 GPa for stress tensors within the validation data set D. On the other hand, the model with data set A without pretraining has the RMSE values of 0.39 kJ/mol/molecule for energy, 5.27 kJ/mol/Å for force, and 0.03 GPa for stress tensors. Our results demonstrate that developing a tailored NNIP for aspirin only can achieve a considerably better accuracy as compared to the commonly used general-purpose force field.

To further clarify the accuracy of our NNIP model, we performed two additional validations. First, we tested the temperature dependence of cell parameters for both I and II by running a series of 4 × 4 × 4 supercell NPT MD simulations lasting for 100 ps at a range of temperatures from 70 to 350 K. In each simulation, the averaged cell parameters for the last 10 ps were used to determine the equilibrium geometry at the given temperature. For the purpose of comparison, we repeated the same procedures to obtain the temperature-dependent lattice parameters by using the classical GAFF model. The results of NNIP and GAFF, and available experimental results are shown in Figure 3. As compared to the GAFF model, the NNIP results tend to achieve better agreement with the experimental data. Importantly, in the direction which has more sensitive temperature dependence (see Figure 3a,d), the GAFF model significantly overestimates the thermal expansion, while NNIP remains to yield accurate results as compared to the experimental values.

Second, we tested whether the NNIP model can faithfully reproduce the transition pathway between aspirins I and II, which is crucial for the subsequent free energy calculation. Following the recent works,[9,10] we constructed the transition

Table 1. Accuracy of the NNIP Model on Different Datasets and Pre-training Strategies; the Models are (i), (ii), (iii) and (iv)[a]

	energy RMSE (kJ/mol/molecule)				force RMSE (kJ/mol/Å)				stress RMSE (GPa)			
	A	B	C	D	A	B	C	D	A	B	C	D
(i)	0.12	8.23	3.27	0.39	3.00	25.28	11.11	5.27	0.02	0.19	0.13	0.03
(ii)	0.10	25.12	3.03	0.26	2.75	24.02	90.91	4.63	0.02	0.31	0.10	0.03
(iii)	0.96	1.74	0.63	0.29	4.18	4.19	2.39	4.19	0.08	0.09	0.07	0.08
(iv)	0.67	1.46	0.81	0.12	4.02	3.89	2.41	3.92	0.07	0.08	0.06	0.06

[a]The accuracy is evaluated by the RMSEs for the datasets A, B, C and D.

https://doi.org/10.1021/acsomega.4c04782
ACS Omega 2024, 9, 36589−36599

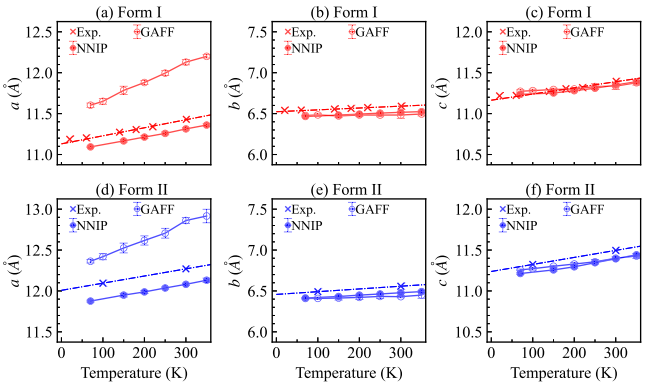


Figure 3. Comparison of lattice constants of aspirin crystals between experimental data (cross markers), NNIP (solid circles), and GAFF (open circles) results for Forms I (the upper panel from a to c) and II (the lower panel from d to f) sampled from a 100 ns MD simulation of the 4 × 4 × 4 supercell structures. The experimental data were taken from the Cambridge Crystallographic Data Centre (CCDC),[43] including ACSALA01,[44] ACSALA03, ACSALA04, ACSALA05, ASCALA06, and ACSALA08[45] for Form I, and ACSALA13[11] and ACSALA17[46] for Form II.

path by applying a constant strain rate (0.1 Å/ps) on the xz component of a 2 × 2 × 2 supercell of Form I in a MD simulation with the NNIP model as implemented in the LAMMPS code.[47] ϵ_{xz} refers to the shear strain component in the xz plane of the crystal lattice. In the context of the aspirin polymorphic transition between Forms I and II, ϵ_{xz} represents the key structural parameter that characterizes the shear deformation along the (100)[001] direction. This shear transformation is the primary mechanism by which Form I transitions to Form II, as illustrated in Figure 1. By applying a constant strain rate to the ϵ_{xz} component during our MD simulations, we were able to generate a series of intermediate structures along the transition path between the two polymorphs. From this MD simulation trajectory, we selected the snapshots every 50 ps as initial structures to run subsequent calculations for sampling their thermodynamic quantities. For each snapshot structure, we performed MD simulation with two stages. First, we performed a constrained NPT MD simulation by fixing the strain of ϵ_{xz} for 10 ps to obtain a relaxed configuration at the given ϵ_{xz} condition. Then, another 10 ps NVT MD calculation was performed to gain the averaged thermodynamic quantities. In addition, the same calculations were repeated using GAFF and DFT-MD simulations for a comparison.

Figure 4 displays the NNIP, GAFF, and DFT results regarding the profiles of energy (U) and σ_{xz} stress along the transition path from I to II. Clearly, both U and σ_{xz} profiles generated by NNIP accurately reproduce the DFT results (see Figure 4a,c). As a comparison, the GAFF also captures the general trend; however, the errors are notably larger than the NNIP results. These validations guarantee the use of NNIP in studying the free energy difference between Forms I and II within the framework of TI to be discussed in the following section.

Free Energy Difference Calculation. In order to determine the free energy difference between aspirin I and II, we employed the thermodynamic integration approach in conjunction with MD simulations[48] by following a recent work on NiTi alloys.[49] In the context of TI, we chosen cos β as the integrating variable (λ) and computed the Helmholtz free energy difference $\Delta F_{I\to II}$ via integrating the transition path from I to II as follows.

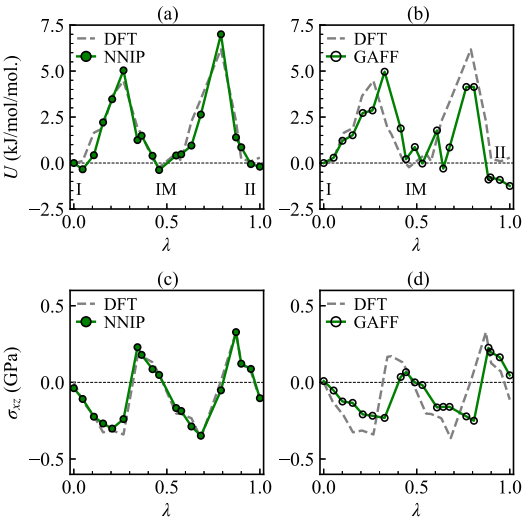


Figure 4. Computed potential energy (a, b) and σ_{xz} stress values (c, d) between aspirin I−II transition. In all subplots, we used a normalized variable λ based on $\Delta\cos\beta$ along the transition path to represent the x-coordinates. For NNIP and GAFF, the DFT results are superimposed with dashed lines for comparison. All values are averaged from the results of MD simulations with 2 × 2 × 2 supercells at 300 K.

$$\Delta G_{I\to II} \approx \sum_{k=0}^{n} V_k \left[\langle \boldsymbol{\sigma}_k \rangle \boldsymbol{\Omega}_k^{-T} : \frac{\partial \boldsymbol{\Omega}_k}{\partial \lambda_k} \right] \Delta\lambda_k \quad (1)$$

where n is the number of sampling points along the λ-path, and $\boldsymbol{\Omega}_k$, V_k, and σ_k denote the cell matrix, the cell volume, and the stress tensor in each sampled state. The derivative of $\boldsymbol{\Omega}_k$ can be obtained by numerical derivation of fitting function along the λ-path, and $\Delta\lambda_k$ indicates the difference in λ between sampling points $k-1$ and k. Furthermore, the Gibbs free energy difference $\Delta G_{I\to II}$ can be approximated as $\Delta F_{I\to II}$ by omitting the variation of the pV term due to its negligible contribution.

In this work, we tried TI calculations with a series of supercell models up to 2 × 4 × 16 (i.e., at most 512 molecules). In particular, we choose to focus on varying the replication on the c-axis (up to 185.1 Å) in order to minimize the fluctuation of stress applied in the (100)[001] direction (σ_{xz}) during the sampling of high energy intermediate configuration, thus reducing the error in the free energy calculation (see details in the Experimental Section). In order to minimize the error caused by changes in the λ-path, we consider taking as many sampling points as possible. Ultimately, for all calculations, more than 70 sampling points are implemented for different λ values between Form I and Form II.

Figure 5 summarizes the free energy differences from the TI calculations between Forms I and II using the NNIP and GAFF models. As shown in Figure 5a, the TI simulations using the NNIP model at 300 K suggest that Form I has a lower free energy than Form II ($\Delta G_{I\to II}$ = 1.83 kJ/mol/molecule), despite the nearly zero difference in the internal energy ($\Delta U_{I\to II}$). In addition, we considered the errors of $\Delta G_{I\to II}$ in the TI simulation due to the fluctuation of stress tensors. In a 2 × 4 × 16 supercell at 300 K, the estimated error is about 0.46 kJ/mol/molecule, which is notably smaller than $\Delta G_{I\to II}$, thus confirming that Form I should be more stable regardless of numerical errors. Furthermore, Figure 5c plots the temperature dependence of Forms I and II from 0 to 400 K. In the entire temperature range, it shows that $\Delta G_{I\to II}$ first increases to 2.9 kJ/mol/molecule and then gradually decreases from 150 to 400 K.

https://doi.org/10.1021/acsomega.4c04782
ACS Omega 2024, 9, 36589−36599

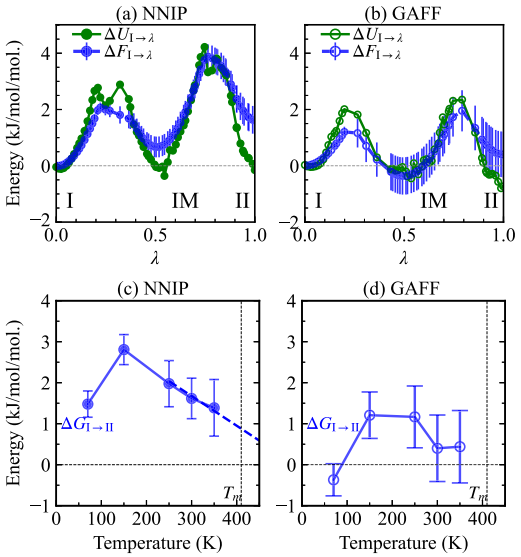


Figure 5. Simulated energy profiles along the path from Form I to Form II at 300 K, with (a) NNIP and (b) GAFF. (c) and (d) display the temperature dependence of the Gibbs free energy difference between Forms I and II (red). Error bars in each subplot indicate the accumulated integrated error throughout the analysis. The vertical black dotted line indicates the melting point of Form I.

Hence, $\Delta G_{I\to II}$ remains positive in the entire temperature range from 0 to 400 K.

For a comparison, the GAFF simulations, as shown in Figure 5b,d, reveal a general trend very similar to the NNIP results. However, the accompanying errors are considerably larger. Thus, it is clear that a quantum-accuracy model is necessary when the free energy difference is near the sub kJ/mol range.

Given that the TI results may strongly depend on the setup of the simulated model size, it is necessary to further investigate the size dependence of our TI simulation. Therefore, we compared TI results by considering a variety of supercell models, as detailed in Table 2. Clearly, the Gibbs free energy difference,

Table 2. Computed Free Energy Differences (kJ/mol/molecule) between Aspirin Forms using NNIP at 300 K[a]

supercell sizes	2 × 4 × 4	2 × 4 × 6	2 × 4 × 8	2 × 4 × 16
$\Delta G_{I\to II}$	4.53 (1.16)	2.99 (0.99)	2.41 (0.84)	1.83 (0.46)
$\Delta G_{I\to IM}$	2.60 (0.68)	1.59 (0.63)	1.49 (0.58)	1.01 (0.34)

[a]Errors in parentheses are shown in the RMSE.

$\Delta G_{I\to II}$, decreases with the model size (in particular, it depends on the length of *c*-axis). Fitting the trend of the decrease with the reciprocal of *c* value, we find the lower bound of $\Delta G_{I\to II}^{min}$ should be around 0.74 kJ/mol/molecule even when extrapolated to infinity, as shown in Figure 6. In addition, the integral error of the TI method, evaluated based on the variance of the pressure at each sampling point is sufficiently reduced when the supercell model increases. Hence, it is safe to conclude that Form I is more consistently stable than Form II at 300 K regardless of the supercell size choices.

Validation with Experimental Data. Experimentally, the free energy difference between Form I and Form II crystalline polymorphs can be estimated from an empirical approach by considering the a melting point difference (10 K) between I and II.[50] Assuming that (i) the free energy difference of two

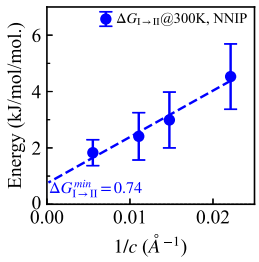


Figure 6. Computed $\Delta G_{I\to II}$ at 300 K as a function of model size (1/*c*).

polymorphs at the melting point is zero, (ii) the temperature dependence of the free energy difference is linear between the melting points, and (iii) the molar heat capacity difference (ΔC_p) is constant, the free energy difference (ΔG) can be approximated using the equation $\Delta G \approx \Delta S_f \times \Delta T$, where ΔS_f is the melting entropy (entropy change associated with melting; fusion entropy) and ΔT is the difference in melting points, 10 K (see details in the Experimental Section). Using a typical melting entropy value for organic compounds of 56.5 J/mol-K[51] (called Walden's rule), the calculated ΔG is approximately 0.56 kJ/mol. This value is comparable to the 1 kJ/mol obtained by extrapolating the data in Figure 5c to around 410 K, which is near the experimental melting point of aspirin I. When the size effect is considered, the calculated result is not significantly different from the figure value. Therefore, we conclude that our free energy difference results from TI simulation is consistent with the previous experimental reports.[12,50]

Atomistic Mechanism Analysis. As discussed earlier, the majority of previous (free) energy calculations,[9,20] based on the either harmonic or quasihamonic phonon approach, indicate that aspirin I and II are nearly indistinguishable under the room temperature conditions. Indeed, we also found that the vibrational density of states (VDOS) of aspirin I and II at 70, 150, and 300 K (see Figure 7) is nearly identical by using the Fourier method from the NNIP-MD simulation. Using the harmonic phonon approximation, the free energy difference between I and II is only 0.03 kJ/mol/molecule at 300 K. However, our TI results based on both the classical GAFF and more accurate NNIP models yield a consistent stability ranking as compared to the experiment. Furthermore, both GAFF and NNIP results (see Figure 5b,d) suggest that $\Delta G_{I\to II}$ first increases while the temperature rises from 0 to 150 K and then has a desceding trend while further increasing the temperature from 150 K to a high temperature limit of around 400 K. This is an indication that some anharmonic molecular motions, which cannot be captured by the harmonic phonon approximation, may have different temperature dependencies in Forms I and II. To understand the physical origin of free energy difference between I and II, we proceeded to analyze the representative molecular motions from MD simulation trajectories. As discussed previously, the majority of molecular interactions in the aspirin crystals may be characterized by the motions of functional groups. Hence, we performed NPT-MD simulations for both Forms I and II at a variety of temperatures to track the motions of representative functions groups. Among them, two characteristic motions were identified and are discussed below.

The first motion is the rotation of methyl groups. It can be seen that the methyl groups are closer together and more likely to interact in Form II as shown in Figure 8c than in Form I as shown in Figure 8b. The middle panel of Figure 8 plots the histogram of the methyl group's orientation at the whole range

https://doi.org/10.1021/acsomega.4c04782
ACS Omega 2024, 9, 36589−36599

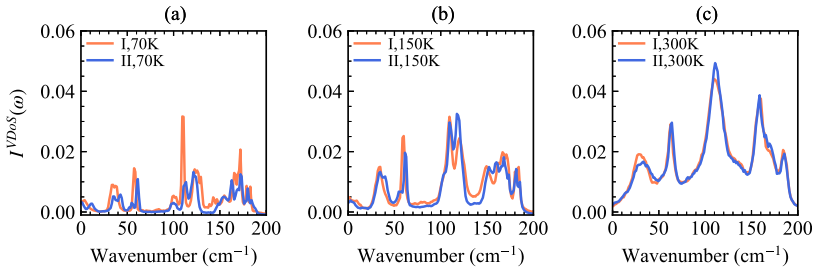


Figure 7. Vibrational density of states from NNIP-MD at (a) 70, (b) 150, and (c) 300 K. Line colors show Form I (orange) and Form II (blue).

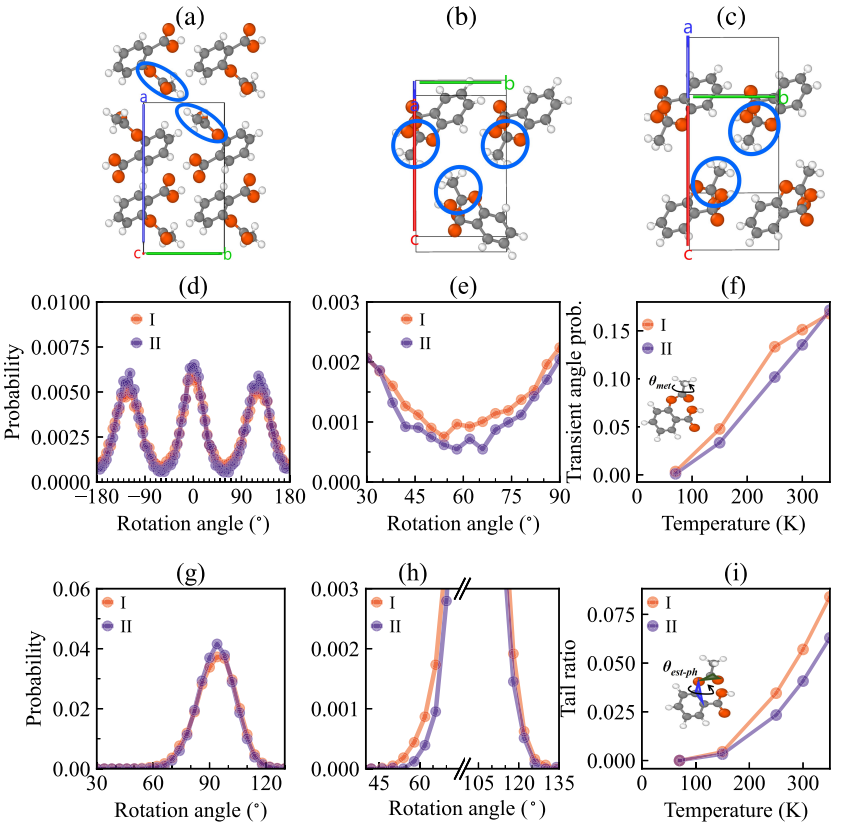


Figure 8. Two representative molecular rotations and their occurring frequencies in Forms I and II from NNIP-MD simulation. (a) The rotation of methyl groups viewed from the *ab* plane in Form I, and (b) and (c) the rotations of ester/phenyl groups viewed from the *bc* plane in Forms I and II, respectively. As for the methyl group, (d) shows the distribution of rotation angles for methyl groups evaluated with at 300 K, (e) shows the low probability (transition) regions centering around 60 degree, and (f) shows the transient probability as a function of temperature. For the ester/phenyl rotations, (g) shows the distribution of rotation angles from NNIP simulation at 300 K, (h) shows the low probability regions on both left and right tails, and (i) plots low transient probability plot as a function of temperature.

in Figure 8d based on the NNIP-MD results. From Figure 8d, it is clear that the methyl group has three preferred orientation distributions centered at around 0°, 120°, or −120°, which correspond to three degenerate low-energy configurations to attach the methyl group to the backbone framework of the aspirin molecule. In the majority of MD simulations, the methyl groups rotate within the three distinct energy basins. However, each methyl group also has the chance to switch from one basin to another via crossing the transient states due to a small rotation barrier. Figure 8e shows the rotational crossover events occurring on one typical transient state at around 60°. Clearly, such events occur more often in aspirin I as compared to those in aspirin II, indicating that I gains more rotational entropy than II. Figure 8f shows the plot of the probability of observing the transient states in the MD simulations under different temperature conditions. Transition states are defined as the ranges of 60 ± 30°, −60 ± 30° and 180 ± 30°, and the ratio of transition states to the total is indicated in Figure 8f. There is a clear trend that Form I generally has more rotational freedoms at low temperatures, indicating that it has a smaller activation barrier. Our results qualitatively agree with the an activation temperature range of 120 to 275 K as found in the experimental solid-state ^{2}H NMR study.[52] When the temperature becomes sufficiently high, such rotation modes become equally frequent in both Forms I and II, and thus, the free energy difference between I and II becomes indistinguishable, as shown in Figure 5b.

Figure 8g−i presents a similar analysis on the relative rotations between the ester and phenyl groups. We found that such rotations are mainly restricted to 90 ± 30° range in both Forms I and II. However, Form II has a sharper peak, indicating a lower crossover probability (i.e., a lower rotational entropy). On the

https://doi.org/10.1021/acsomega.4c04782
ACS Omega 2024, 9, 36589−36599

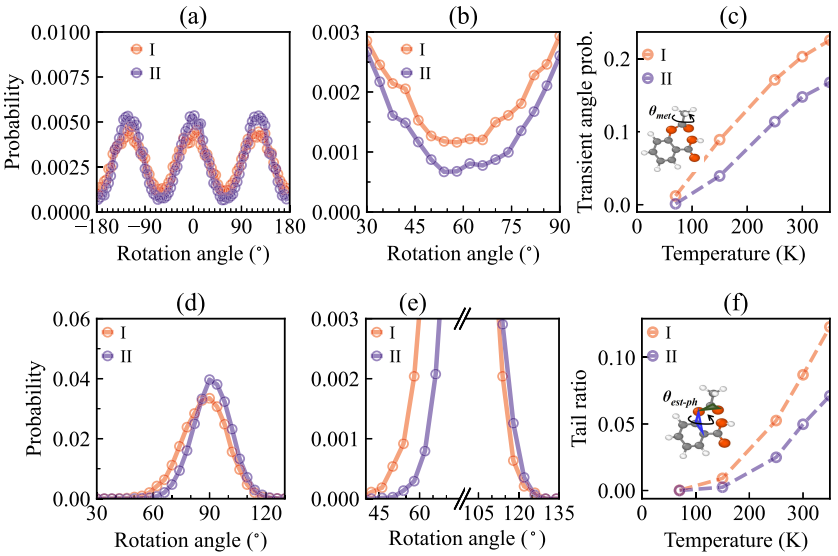


Figure 9. Statistics of molecular rotations derived from the GAFF model. As for the methyl group, (a) shows the distribution of rotation angles for methyl groups evaluated with at 300 K, (b) shows the low probability (transition) regions centering at around 60°, and (c) shows the transient probability as a function of temperature. For ester/phenyl rotations, (d) shows the distribution of the rotation angle from GAFF simulation at 300 K, (e) shows the low probability regions on both left and right tails, and (f) plots low transient probability plot as a function of temperature.

contrary, the distribution on Form I has a longer tail, indicating a higher level of rotational activity, as shown in Figure 8h. Here, the tail states are defined as those with angles less than 70° or greater than 110°. Figure 8i shows the percentages of crossover events in both forms as a function of temperature, suggesting that Form I has a greater rotational entropy than Form II at the low temperature region.

While the distinct kinetics of rotational transitions between energy minima contributes to the entropic stabilization of Form I, it is important to acknowledge that other factors may also play a role in the observed entropy difference. One potential source could be variations in the degrees of freedom associated with the interfacial configurations upon which the rotation of methyl groups occurs. Such interfacial configurations, potentially involving anharmonic lattice dynamics, could lead to differences in the accessible space between Forms I and II. These structural variations could result in changes in the accessible configurational space, thereby affecting the overall entropy. A comprehensive understanding of the polymorphic stability in aspirin crystals may require a more in-depth investigation, involving advanced sampling techniques and theoretical approaches, to rigorously quantify the various entropic contributions arising molecular rotations/vibrations, interfacial slip, lattice change, and configurational degrees of freedom.

Additionally, we repeated the analysis on the GAFF-MD results and observed a similar trend, as shown in Figure 9, confirming the robustness of our analysis. Given the above motion analysis, we conclude that crossover molecular rotations of functional groups in Form I has a lower activation barrier as compared to that in Form II. Importantly, such crossover events can not be well treated within the framework of harmonic approximation. This may explain why the previous studies fail to find that Form I has a lower free energy by considering harmonic vibrations only.

Alternative Interpretation from the Structural Aspect. When the temperature becomes sufficiently high and such rotation modes become equally probable for both Forms I and II, the rotational entropy difference becomes negligible. Under such conditions, the free energy difference is expected to converge to zero. This phenomenon can also be understood from a structural aspect. As shown in Figure 1, the structural difference between Forms I and II mainly lies in the slips in the *ac* plane, while *b*-axis is always perpendicular to the *ac* plane and invariant to the choice of basis vectors. Figure 10 plots the

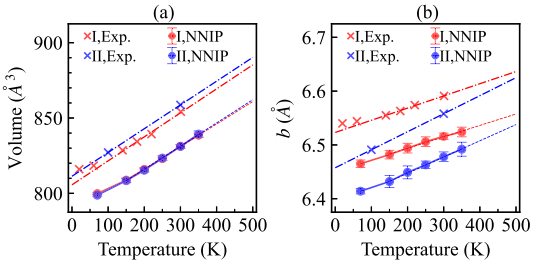


Figure 10. Temperature dependencies of unit cell volume (a) and *b*-axis length (b) in both Forms I and II.

evolution of cell volumes and *b* values as a function of temperature for both experimental and simulated data, respectively. It can be seen that the volumes (Figure 10a) for both Forms I and II are nearly identical. However, the difference in *b* is clearly more distinct. Thus, we can count *b* as a main characteristic variable to distinguish Forms I and II. Indeed, Figure 10b found that Form II's cell parameter *b* is consistently lower than that of Form I. This indicates that a larger *b* value in Form I is clearly more beneficial to promote the rotation of methyl and ester groups, thus leading to a higher entropy. With the increase of the temperature, the *b* value of Form II expands more rapidly than that of Form I and eventually leads to the closure of the free energy gap.

Prospects for Reducing the Number of Sampling Points. In this study, the error of the TI method emerged as a limiting issue. The evaluation of free energy requires a large number of samples, and at this time, the evaluation of large systems with supercells is required. Therefore, we discuss the possibility of reducing the error of the TI method itself in order to obtain fewer sampling points. In the present study, the

https://doi.org/10.1021/acsomega.4c04782
ACS Omega 2024, 9, 36589−36599

sampling points between phases were obtained in detail for λ following a continuous cell deformation. Beyond the TI approach, the combination of Bennett's acceptance ratio method[53−55] and the extended ensemble method may be used to obtain enough points on the path by a stochastic approach. In addition, a method that introduces an alchemical pathway rather than a real pathway may also bypass the increased number of samples near the steep barrier between polymorphs present in the potential energy surface.[56] In addition, the choice of numerical integration method is an essential point to consider when aiming for a sufficiently small number of samples.[57] These approaches may be left for future work due to time constraints.

Remarks on the Choice of Force Field Models. Finally, the comparative analysis between NNIP and GAFF highlights the unique value of both approaches. While the GAFF model yields less accurate energetic descriptions, it can still predict the right trend for both ΔG values (Figure 5b,d) and low-probability molecular rotation events (Figure 9). Therefore, the generic force field models such as GAFF remain valuable due to their easy accessibility and low computational cost. For example, they can be used for early stage studies for the screening purposes, or the cases where a truly large scale modeling is necessary. However, in more challenging cases (e.g., the examination of polymorphic stability ranking in this study) when both a fine accuracy (e.g., < 2 kJ/mol) and a large model size (e.g., > 100 molecules) are critical, developing a quantum-accurate NNIP model becomes necessary.

CONCLUSIONS

In summary, we developed a machine-learning driven interatomic potential to precisely model intermolecular interactions within aspirin crystals. Utilizing this model to analyze the polymorphic stability ranking of two aspirin crystal forms, we demonstrate that this approach yields predictions more accurate than those of traditional models in describing the thermal and kinetic profiles for the known aspirin crystals. By employing the thermodynamic integration method that considers anharmonic effects, we determined that Form I consistently exhibits lower Gibbs free energy under the finite temperatures. Furthermore, we found that the Gibbs free energy between the two forms stems from different temperature responses of rotational motions in both Froms I and II. For the first time, our work successfully resolves the discrepancy between the previous experimental observation and modeling results regarding the polymorph stability ranking between aspirin I and II. The success of our findings not only highlights the importance of anharmonic effects in free energy calculation for organic crystals but also demonstrates the power of high-fidelity NNIP models in boosting the accurate atomistic modeling for polymorphic stability of molecular crystals. Finally, the generic training method for accurate target-specific NNIP proposed in this study strongly supports the discovery of new polymorphs by tailoring the target-specific NNIP, facilitating the future rational design of drugs and functional materials.

EXPERIMENTAL SECTION

DFT and MD Simulations. All DFT-MD simulations in this study were performed using the VASP code.[58] MD simulations with NNIP and GAFF models in this study were performed using the LAMMPS code.[47] In all MD simulations, time step Δt was set to 0.5 fs. In the DFT-MD simulations, the electronic states were calculated by the projector augmented wave method with a plane wave cutoff energy of 520 eV.[59] The Brillouin zone was sampled at the Γ-point only. The r2SCAN meta-GGA exchange-correlation functional[42,60,61] was employed along with the DFT-D4 van der Waals correction.[62−64] Convergence criteria for electronic and ionic relaxations were set to 10^{-5} eV and 10^{-4} eV/Å, respectively.

NNIP Modeling. In this work, the Allegro model was built through NequIP[27] that introduced an E(3) symmetry equivariant neural network coded by PyTorch.[65] Our model adopts two layers of 64 tensor features with a $l_{max} = 2$ in full $O(3)$ symmetry. We used a two-body latent multi layer perceptron (MLP) and later latent MLP with hidden dimensions [64, 128, 256, 512] and [512, 512, 512] respectively, both with SiLU nonlinearities. The embedded MLP was a linear projection. For the final edge energy MLP, we used a single hidden layer of dimension 512 and no nonlinearity. All four MLPs were initialized according to a uniform distribution of unit variance. Models were trained with a radial cutoff of 5.0 Å. The loss function was the sum of the per-atom energy, force, and stress of the RMSE, and the weight ratio was 1:1:1 respectively. The Adam optimizer, a gradient-based probabilistic algorithm, was used for parameter updating.

To ensure the representativeness of the training data, we analyzed the distributions of key intermolecular contacts between functional groups (e.g., carboxylic acid−carboxylic acid, methyl−methyl, and phenyl−phenyl) in the collected configurations. As shown in Figure 2, the MD-derived data set (A) covers a narrower range of distances compared to the crystal structure prediction data set (B), indicating the importance of incorporating diverse configurations from various sampling techniques. The convergence of the training process was monitored by tracking the root-mean-square errors (RMSEs) for energy, forces, and stress tensors on a separate validation set, ensuring that the NNIP model accurately captures the potential energy surface of the aspirin crystals. It was observed that the training converged after approximately 800 epochs, at which point the RMSEs remained stable with further training.

Free Energy Calculation from Thermodynamic Integration. In the context of thermodynamic integration, the Helmholtz free energy F difference at a given temperature T between two reference phases (I and II) can be obtained via the integration as follows:[24,66]

$$\begin{aligned}\Delta F_{\mathrm{I}\to\mathrm{II}} &= \int_0^1 \frac{\partial F}{\partial \lambda}\mathrm{d}\lambda \\ &= \int_0^1 \frac{\partial(-k_{\mathrm{B}}T\ln \mathcal{Z})}{\partial \lambda}\mathrm{d}\lambda \\ &= -\int_0^1 \frac{k_{\mathrm{B}}T}{\mathcal{Z}}\frac{\partial \mathcal{Z}}{\partial \lambda}\mathrm{d}\lambda\end{aligned} \tag{2}$$

where λ is a thermodynamic variable along a path that connects I and II, $\mathcal{Z} = \sum_s \exp(-\mathcal{H}/k_{\mathrm{B}}T)$ is the partition function, and k_{B} is the Boltzmann constant.

An equation transformation via a linear transformation of the cell matrix $\mathbf{\Omega}$ generalized by Haskins et al.[66] is presented as follows:

$$\Delta F_{\mathrm{I}\to\mathrm{II}} = -\int_0^1 \frac{k_{\mathrm{B}}T}{\mathcal{Z}}\frac{\partial \mathcal{Z}}{\partial \mathbf{\Omega}} : \frac{\partial \mathbf{\Omega}}{\partial \lambda}\mathrm{d}\lambda \tag{3}$$

where $\frac{\partial \mathcal{Z}}{\partial \mathbf{\Omega}}$ and $\frac{\partial \mathbf{\Omega}}{\partial \lambda}$ are both tensors and ":" denotes the Frobenius inner product (i.e., $\Sigma_{ij}\frac{\partial \mathcal{Z}}{\partial \mathbf{\Omega}_{ij}}\frac{\partial \mathbf{\Omega}}{\partial \lambda}_{ij}$), and $\mathbf{\Omega}$ is the following 3 × 3

https://doi.org/10.1021/acsomega.4c04782
ACS Omega 2024, 9, 36589−36599

matrix as the triclinic simulation box with periodic boundary condition imposed on the entire MD system as a supercell of each crystal form,

$$\boldsymbol{\Omega} = \begin{bmatrix} a_x & b_x & c_x \\ a_y & b_y & c_y \\ a_z & b_z & c_z \end{bmatrix}$$

where, it consists of lattice vectors, $\boldsymbol{a} = (a_x, b_y, a_z)$, $\boldsymbol{b} = (b_x, b_y, b_z)$, $\boldsymbol{c} = (c_x, c_y, c_z)$. λ can be constructed from these cell parameters. In our case, it has been used to convert Aspirin Form I to II by $\lambda = \cos(\beta) = \boldsymbol{a}\cdot\boldsymbol{c}/|\boldsymbol{a}|\cdot|\boldsymbol{b}|$. This free energy expression can be described by the average of Hamiltonian.

$$\Delta F_{\mathrm{I}\to\mathrm{II}} = \int_0^1 \frac{k_\mathrm{B}T}{\mathcal{Z}} \sum_s \frac{1}{k_\mathrm{B}T} \exp[-\mathcal{H}/k_\mathrm{B}T] \frac{\partial \mathcal{H}}{\partial \boldsymbol{\Omega}} : \frac{\partial \boldsymbol{\Omega}}{\partial \lambda} \mathrm{d}\lambda = \int_0^1 \left\langle \frac{\partial \mathcal{H}}{\partial \boldsymbol{\Omega}} \right\rangle : \frac{\partial \boldsymbol{\Omega}}{\partial \lambda} \mathrm{d}\lambda, \tag{4}$$

where $\left\langle \frac{\partial \mathcal{H}}{\partial \boldsymbol{\Omega}} \right\rangle$ is a tensor derived by the differentiating of the Hamiltonian with cell matrix elements. It can be extracted as an ensemble average of the thermodynamic quantities evaluated within the MD simulation. The Hamiltonian $\mathcal{H}$ of the system is defined as the kinetic energy and the potential energy $U(\{\boldsymbol{r}_i\})$, function of the coordinates of all particles in the system, $\{r_i\}$.

$$\mathcal{H} = \sum_{i=1}^{N} \frac{\boldsymbol{p}_i \cdot \boldsymbol{p}_i}{2m_i} + U(\{\boldsymbol{r}_i\}) \tag{5}$$

where $\boldsymbol{p}_i$ are momentum and m_i are mass of particle i. For the derivative transformation, $\{\boldsymbol{r}_\mathrm{i}\}$ and $\{\boldsymbol{p}_\mathrm{i}\}$ are transformed by $\boldsymbol{r}_\mathrm{i} = \boldsymbol{\Omega}\boldsymbol{\rho}_\mathrm{i}$, $\boldsymbol{p}_\mathrm{i} = \boldsymbol{\pi}_\mathrm{i}\boldsymbol{\Omega}^{-1}$, where $\boldsymbol{\rho}_\mathrm{i}$ are reduced coordinates, $\boldsymbol{\pi}_\mathrm{i}$ are transformed momentum, and a superscript "-1" indicates the inverse tensor. Then,

$$\begin{aligned} \mathcal{H} &= \sum_{i=1}^{N} \frac{1}{2m_i} (\boldsymbol{\pi}_i\boldsymbol{\Omega}^{-1})\cdot(\boldsymbol{\pi}_i\boldsymbol{\Omega}^{-1}) + U(\{\boldsymbol{\Omega}\boldsymbol{\rho}_i\}) \\ \frac{\partial \mathcal{H}}{\partial \boldsymbol{\Omega}} &= \sum_{i=1}^{N} \frac{1}{m_i} (\boldsymbol{\pi}_i\boldsymbol{\Omega}^{-1}) \cdot \frac{\partial (\boldsymbol{\pi}_i\boldsymbol{\Omega}^{-1})}{\partial \boldsymbol{\Omega}} + \sum_{i=1}^{N} \frac{\partial U}{\partial (\boldsymbol{\Omega}\boldsymbol{\rho}_i)} \cdot \frac{\partial (\boldsymbol{\Omega}\boldsymbol{\rho}_i)}{\partial \boldsymbol{\Omega}} \\ &= -\sum_{i=1}^{N} \frac{1}{m_i} (\boldsymbol{\pi}_i\boldsymbol{\Omega}^{-1}) \otimes (\boldsymbol{\pi}_i\boldsymbol{\Omega}^{-1})\boldsymbol{\Omega}^{-T} - \sum_{i=1}^{N} \boldsymbol{f}_i \otimes \boldsymbol{\rho}_i \\ &= -V\left[\frac{1}{V}\sum_{i=1}^{N} \frac{1}{m_i}(\boldsymbol{p}_i \otimes \boldsymbol{p}_i) + \frac{1}{V}\sum_{i=1}^{N} (\boldsymbol{f}_i \otimes \boldsymbol{r}_i)\right]\boldsymbol{\Omega}^{-T} \end{aligned} \tag{6}$$

where $\otimes$ is the cross product, $\boldsymbol{f}_i$ is the force on particle i, and V is a volume of system. The brackets in the last equation are consistent with the pressure tensor of the kinetic and virial part. Here, we introduce a stress tensor $\boldsymbol{\sigma}$ with a negative pressure tensor. Then,

$$\frac{\partial \mathcal{H}}{\partial \boldsymbol{\Omega}} = V(\boldsymbol{\sigma}\boldsymbol{\Omega}^{-T}) \tag{7}$$

Finally, the free energy difference can be evaluated numerically by combining eqs 4 and 7 as follows:

$$\Delta F_{\mathrm{I}\to\mathrm{II}} = \int_0^1 \langle V(\boldsymbol{\sigma}\boldsymbol{\Omega}^{-T})\rangle : \frac{\partial \boldsymbol{\Omega}}{\partial \lambda} \mathrm{d}\lambda \approx \sum_{k=0}^{n} V_k \langle \boldsymbol{\sigma}_k \rangle \boldsymbol{\Omega}_k^{-T} : \frac{\partial \boldsymbol{\Omega}_k}{\partial \lambda_k} \Delta\lambda_k \tag{8}$$

where k loops over all MD simulations along the path from Form I to Form II. In each configuration, $\boldsymbol{\Omega}_k$ were extracted from the given configuration along the path, V_k could be calculated from $\boldsymbol{\Omega}_k$. $\boldsymbol{\sigma}_k$ were obtained by averaging all stress tensors from the MD simulations, and $\partial\boldsymbol{\Omega}_k/\partial\lambda_k$ were obtained from cubic spline interpolation.

From eq 8, it is clear that V_k, $\boldsymbol{\Omega}_k$, and $\partial\boldsymbol{\Omega}_k/\partial\lambda_k$ are fixed values for a given path. However, $\boldsymbol{\sigma}_k$ (in particular, the $\boldsymbol{\sigma}_{xz}$ component in the case of aspirin I−II transition) may fluctuate strongly when the configuration is away from the nearby energy minimum, thus leading to a variation of thr ΔF value. Hence, we use the following relation to evaluate the error of $\Delta F_{\mathrm{I}\to\mathrm{II}}$,

$$\mathrm{Error}(\Delta F) = \sum_{k=0}^{n-1} \left(\frac{\Delta\lambda_k}{2}\mathrm{RMSE}(f_k)\right)^2 + \sum_{k=0}^{n-1} \left(\frac{\Delta\lambda_k}{2}\mathrm{RMSE}(f_{k+1})\right)^2 \tag{9}$$

Here, f_k indicates the inside of integration as $f_k = V_k\boldsymbol{\sigma}_k\boldsymbol{\Omega}_k^{-T} : \frac{\partial\boldsymbol{\Omega}_k}{\partial\lambda_k}$, and RMSE($f$) is root-mean-square error of f originated in stress tensor fluctuation.

Free Energy Calculation from Harmonic Phonons. The vibrational density of states (VDOS) was evaluated to analyze the vibrational differences between aspirin Forms I and II as frequency space contributions. To evaluate VDOS, molecular dynamics (MD) simulations with NPT ensembles were performed at 300 K using the NNIP model with 2 × 4 × 16 supercells. The VDOS, calculated by Fourier transform from the velocity autocorrelation function of atoms, is shown in Figure 7. A total of 1000 frames per 25 fs were used as the MD trajectory data. The free energy of harmonic phonon approximation using the VDOS spectrum was calculated using the following equation:[8]

$$F^{\mathrm{VDoS}} = k_\mathrm{B}T \int I^{\mathrm{VDoS}}(\omega) \ln\left[1 - \exp\left(-\frac{\hbar\omega}{k_\mathrm{B}T}\right)\right] \tag{10}$$

Here, $I^{\mathrm{VDoS}}(\omega)$ indicates the intensity of VDoS as ω is the angular frequency of vibration.

Rough Estimation of Free Energy Difference Using Experimental Values of Phase Melting Difference. The derivation of the equation $\Delta G \approx \Delta S_\mathrm{f} \times \Delta T$ is based on several assumptions and approximations. At the melting point (T_m) of each polymorph, the free energy of the solid form equals that of the liquid form: $G_\mathrm{s}(T_m) = G_\mathrm{l}(T_m)$. The free energy difference between the two polymorphs (I and II) at any temperature T can be expressed as

$$\Delta G(T) = [G_\mathrm{I}(T) - G_\mathrm{I}(T_{\mathrm{mI}})] - [G_\mathrm{II}(T) - G_\mathrm{II}(T_{\mathrm{mII}})] \tag{11}$$

Assuming a linear temperature dependence of G between the two melting points: $G(T) \approx G(T_\mathrm{m}) - \Delta S_\mathrm{f}(T_\mathrm{m} - T)$, where ΔS_f is the entropy of fusion. Substituting this into the equation for $\Delta G(T)$:

$$\Delta G(T) \approx [\Delta S_{\mathrm{fI}}(T_{\mathrm{mI}} - T)] - [\Delta S_{\mathrm{fII}}(T_{\mathrm{mII}} - T)] \tag{12}$$

https://doi.org/10.1021/acsomega.4c04782
ACS Omega 2024, 9, 36589−36599

If we assume that the entropies of fusion for both polymorphs are approximately equal ($\Delta S_{fI} \approx \Delta S_{fII} \approx \Delta S_f$), we can simplify

$$\Delta G(T) \approx \Delta S_f(T_{mII} - T_{mI}) = \Delta S_f \times \Delta T \quad (13)$$

This derivation neglects the differences in heat capacities between the solid and liquid forms as well as potential differences in the melting enthalpies of the two polymorphs. These approximations are reasonable when the melting point difference is small, as is the case for aspirin Forms I and II ($\Delta T \approx 10$ K).

AUTHOR INFORMATION

Corresponding Authors

Shinnosuke Hattori − *Advanced Research Laboratory, Research Platform, Sony Group Corporation, Atsugi-shi 243−0014, Japan*; orcid.org/0000-0003-0786-1308; Email: shinnosuke.hattori@sony.com

Qiang Zhu − *Department of Mechanical Engineering and Engineering Science, University of North Carolina at Charlotte, Charlotte, North Carolina 28223, United States*; Email: qzhu8@uncc.edu

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.4c04782

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

Computational resource of the AI Bridging Cloud Infrastructure (ABCI) provided by the National Institute of Advanced Industrial Science and Technology (AIST) was used. Q.Z. thanks Hao Tang from the MIT for the insightful discussion about the technical details of the TI approach.

REFERENCES

(1) Nyman, J.; Day, G. M. Static and lattice vibrational energy differences between polymorphs. *CrystEngcomm* **2015**, *17*, 5154−5165.

(2) Cruz-Cabeza, A. J.; Reutzel-Edens, S. M.; Bernstein, J. Facts and fictions about polymorphism. *Chem. Soc. Rev.* **2015**, *44*, 8619−8635.

(3) Sun, C. C. Microstructure of tablet—pharmaceutical significance, assessment, and engineering. *Pharm. Res.* **2017**, *34*, 918−928.

(4) Corpinot, M. K.; Bučar, D.-K. A Practical Guide to the Design of Molecular Crystals. *Cryst. Growth Des.* **2019**, *19*, 1426−1453.

(5) Zhu, Q.; Hattori, S. Organic crystal structure prediction and its application to materials design. *J. Mater. Res.* **2023**, *38*, 19−36.

(6) Ouvrard, C.; Price, S. L. Toward crystal structure prediction for conformationally flexible molecules: the headaches illustrated by aspirin. *Cryst. Growth Des.* **2004**, *4*, 1119−1127.

(7) Wen, S.; Beran, G. J. Accidental degeneracy in crystalline aspirin: New insights from high-level ab initio calculations. *Cryst. Growth Des.* **2012**, *12*, 2169−2172.

(8) Reilly, A. M.; Tkatchenko, A. Role of dispersion interactions in the polymorphism and entropic stabilization of the aspirin crystal. *Phys. Rev. Lett.* **2014**, *113* (5), 055701.

(9) LeBlanc, L. M.; Otero-de-la Roza, A.; Johnson, E. R. Evaluation of shear-slip transitions in crystalline aspirin by density-functional theory. *Cryst. Growth Des.* **2016**, *16*, 6867−6873.

(10) Vaksler, Y.; Idrissi, A.; Urzhuntseva, V. V.; Shishkina, S. V. Quantum Chemical Modeling of Mechanical Properties of Aspirin Polymorphic Modifications. *Cryst. Growth Des.* **2021**, *21*, 2176−2186.

(11) Vishweshwar, P.; McMahon, J. A.; Oliveira, M.; Peterson, M. L.; Zaworotko, M. J. The predictably elusive form II of aspirin. *J. Am. Chem. Soc.* **2005**, *127*, 16802−16803.

(12) Bond, A. D.; Boese, R.; Desiraju, G. R. On the polymorphism of aspirin: crystalline aspirin as intergrowths of two "polymorphic" domains. *Angew. Chem., Int. Ed.* **2007**, *46*, 618−622.

(13) Bond, A. D.; Boese, R.; Desiraju, G. R. On the polymorphism of aspirin. *Angew. Chem., Int. Ed.* **2007**, *46*, 615−617.

(14) Bond, A. D.; Solanko, K. A.; Parsons, S.; Redder, S.; Boese, R. Single crystals of aspirin form II: crystallisation and stability. *CrystEngcomm* **2011**, *13*, 399−401.

(15) Varughese, S.; Kiran, M. S. R.; Solanko, K. A.; Bond, A. D.; Ramamurty, U.; Desiraju, G. R. Interaction anisotropy and shear instability of aspirin polymorphs established by nanoindentation. *Chem. Sci.* **2011**, *2*, 2236−2242.

(16) Shtukenberg, A. G.; Hu, C. T.; Zhu, Q.; Schmidt, M. U.; Xu, W.; Tan, M.; Kahr, B. The Third Ambient Aspirin Polymorph. *Cryst. Growth Des.* **2017**, *17*, 3562−3566.

(17) Huang, Y.; Shao, Y.; Beran, G. J. Accelerating MP2C dispersion corrections for dimers and molecular crystals *J. Chem. Phys.* **2013**, *138*, 22. .

(18) Ghosh, S.; Reddy, C. M. Elastic and bendable caffeine cocrystals: implications for the design of flexible organic materials. *Angew. Chem., Int. Ed.* **2012**, *51*, 10319−10323.

(19) Tsuri, Y.; Maruyama, M.; Fujimoto, R.; Okada, S.; Adachi, H.; Yoshikawa, H. Y.; Takano, K.; Murakami, S.; Matsumura, H.; Inoue, T.; Tsukamoto, K.; Imanishi, M.; Yoshimura, M.; Mori, Y. Crystallization of aspirin form II by femtosecond laser irradiation. *Appl. Phys. Express* **2019**, *12*, 015507.

(20) Li, Q.; Bond, A. D.; Korter, T. M.; Zeitler, J. A. New Insights into the Crystallographic Disorder in the Polymorphic Forms of Aspirin from Low-Frequency Vibrational Analysis. *Mol. Pharm.* **2022**, *19*, 227−234.

(21) Raimbault, N.; Athavale, V.; Rossi, M. Anharmonic effects in the low-frequency vibrational modes of aspirin and paracetamol crystals. *Phys. Rev. Mater.* **2019**, *3*, 053605.

(22) Laman, N.; Harsha, S. S.; Grischkowsky, D. Narrow-line waveguide terahertz time-domain spectroscopy of aspirin and aspirin precursors. *Appl. Spectrosc.* **2008**, *62*, 319−326.

(23) Frenkel, D.; Smit, B.*Understanding Molecular Simulation: from Algorithms to Applications*, 2nd nd ed.; Academic Press, 2002; Vol. *1*.

(24) Haskins, J. B.; Lawson, J. W. Finite temperature properties of NiTi from first principles simulations: Structure, mechanics, and thermodynamics. *J. Appl. Phys.* **2017**, *121* (20), 205103.

(25) Gasteiger, J.; Groß, J.; Günnemann, S. Directional Message Passing for Molecular Graphs. In *International Conference on Learning Representations*, 2020.

(26) Zeng, J.; Zhang, D.; Lu, D.; Mo, P.; Li, Z.; Chen, Y.; Rynik, M.; Huang, L.; Li, Z.; Shi, S.; et al. DeePMD-kit v2: A software package for deep potential models. *J. Chem. Phys* **2023**, *159* (5), 054801.

(27) 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. *Nat. Commun.* **2022**, *13* (1), 2453.

(28) Musaelian, A.; Batzner, S.; Johansson, A.; Sun, L.; Owen, C. J.; Kornbluth, M.; Kozinsky, B. Learning local equivariant representations for large-scale atomistic dynamics. *Nat. Commun.* **2023**, *14* (1), 579.

(29) Deng, B.; Zhong, P.; Jun, K.; Riebesell, J.; Han, K.; Bartel, C. J.; Ceder, G. CHGNet as a pretrained universal neural network potential for charge-informed atomistic modelling. *Nat. Mach. Intell.* **2023**, *5*, 1031−1041.

(30) Kozinsky, B.; Musaelian, A.; Johansson, A.; Batzner, S. Scaling the Leading Accuracy of Deep Equivariant Models to Biomolecular Simulations of Realistic Size. *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis. New York, NY, USA* **2023**, 1−12.

(31) Ibayashi, H.; Razakh, T. M.; Yang, L.; Linker, T.; Olguin, M.; Hattori, S.; Luo, Y.; Kalia, R. K.; Nakano, A.; Nomura, K.-I.; Vashishta, P Allegro-Legato: Scalable, Fast, and Robust Neural-Network Quantum Molecular Dynamics via Sharpness-Aware Minimization. *High Perform. Comput.* **2023**, *13948*, 223−239.

(32) Eastman, P.; Behara, P. K.; Dotson, D. L.; Galvelis, R.; Herr, J. E.; Horton, J. T.; Mao, Y.; Chodera, J. D.; Pritchard, B. P.; Wang, Y.; De Fabritiis, G.; Markland, T. E. SPICE, A dataset of drug-like molecules and peptides for training machine learning potentials. *Sci. Data* **2023**, *10* (1), 11.
(33) Fredericks, S.; Parrish, K.; Sayre, D.; Zhu, Q. PyXtal: A Python library for crystal structure generation and symmetry analysis. *Comput. Phys. Commun.* **2021**, *261*, 107810.
(34) Lyakhov, A. O.; Oganov, A. R.; Stokes, H. T.; Zhu, Q. New developments in evolutionary structure prediction algorithm USPEX. *Comput. Phys. Commun.* **2013**, *184*, 1172−1182.
(35) Zhu, Q.; Oganov, A. R.; Glass, C. W.; Stokes, H. T. Constrained evolutionary algorithm for structure prediction of molecular crystals: methodology and applications. *Acta Crystallogr., Sect. B: Struct. Sci.* **2012**, *68*, 215−226.
(36) Zhu, Q.; Shtukenberg, A. G.; Carter, D. J.; Yu, T.-Q.; Yang, J.; Chen, M.; Raiteri, P.; Oganov, A. R.; Pokroy, B.; Polishchuk, I.; Bygrave, P. J.; Day, G. M.; Rohl, A. L.; Tuckerman, M. E.; Kahr, B. Resorcinol crystallization from the melt: a new ambient phase and new "riddles. *J. Am. Chem. Soc.* **2016**, *138*, 4881−4889.
(37) Santos-Florez, P. A.; Hattori, S.; Zhu, Q. Bending deformation driven by molecular rotation. *Phys. Rev. Res.* **2023**, *5*, 033185.
(38) Menon, S.; Lysogorskiy, Y.; Knoll, A. L. M.; Leimeroth, N.; Poul, M.; Qamar, M.; Janssen, J.; Mrovec, M.; Rohrer, J.; Albe, K.; Behler, J.; Drautz, R.; Neugebauer, J. From electrons to phase diagrams with classical and machine learning potentials: automated workflows for materials science with pyiron *arXiv* **2024**
(39) Janmohamed, H.; Wolinska, M.; Surana, S.; Pierrot, T.; Walsh, A.; Cully, A. Multi-Objective Quality-Diversity for Crystal Structure Prediction *arXiv* 2024
(40) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157−1174.
(41) 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. *J. Chem. Theory Comput.* **2020**, *16*, 4192−4202.
(42) Ehlert, S.; Huniar, U.; Ning, J.; Furness, J. W.; Sun, J.; Kaplan, A. D.; Perdew, J. P.; Brandenburg, J. G. r2SCAN-D4: Dispersion corrected meta-generalized gradient approximation for general chemical applications. *J. Chem. Phys.* **2021**, *154* (6), 061101.
(43) Groom, C. R.; Bruno, I. J.; Lightfoot, M. P.; Ward, S. C. The Cambridge Structural Database. *Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater.* **2016**, *72*, 171−179.
(44) Kim, Y.; Machida, K.; Taga, T.; Osaki, K. Structure redetermination and packing analysis of aspirin crystal. *Chem. Pharm. Bull.* **1985**, *33*, 2641−2647.
(45) Wilson, C. C. Interesting proton behaviour in molecular structures. Variable temperature neutron diffraction and ab initio study of acetylsalicylic acid: characterising librational motions and comparing protons in different hydrogen bonding potentials. *New J. Chem.* **2002**, *26*, 1733−1739.
(46) Chan, E. J.; Welberry, T. R.; Heerdegen, A. P.; Goossens, D. J. Diffuse scattering study of aspirin forms (I) and (II). *Acta Crystallogr. B* **2010**, *66*, 696−707.
(47) Plimpton, S. Fast parallel algorithms for short-range molecular dynamics. *J. Comput. Phys.* **1995**, *117*, 1−19.
(48) Menon, S.; Lysogorskiy, Y.; Rogal, J.; Drautz, R. Automated free-energy calculation from atomistic simulations. *Phys. Rev. Mater.* **2021**, *5* (10), 103801.
(49) Tang, H.; Zhang, Y.; Li, Q.-J.; Xu, H.; Wang, Y.; Wang, Y.; Li, J. High accuracy neural network interatomic potential for NiTi shape memory alloy. *Acta Mater.* **2022**, *238*, 118217.
(50) Mittal, A.; Malhotra, D.; Jain, P.; Kalia, A.; Shunmugaperumal, T. Studies on Aspirin Crystals Generated by a Modified Vapor Diffusion Method. *AAPS PharmScitech* **2016**, *17*, 988−994.
(51) Gilbert, A. S. Entropy−enthalpy compensation in the fusion of organic molecules: implications for Walden's rule and molecular freedom in the liquid state. *Thermochim. Acta* **1999**, *339*, 131−142.
(52) Kitchin, S. J.; Halstead, T. K. Solid-state2H NMR studies of methyl group dynamics in aspirin and aspirin ·βcyclodextrin. *Appl. Magn. Reson.* **1999**, *17*, 283−300.
(53) Errington, J. R.; Kofke, D. A. Calculation of surface tension via area sampling. *J. Chem. Phys.* **2007**, *127* (17), 174709.
(54) König, G.; Bruckner, S.; Boresch, S. Unorthodox uses of Bennett's acceptance ratio method. *J. Comput. Chem.* **2009**, *30*, 1712−1718.
(55) Matsunaga, Y.; Kamiya, M.; Oshima, H.; Jung, J.; Ito, S.; Sugita, Y. Use of multistate Bennett acceptance ratio method for free-energy calculations from enhanced sampling and free-energy perturbation. *Biophys. Rev.* **2022**, *14*, 1503−1512.
(56) Bruckner, S.; Boresch, S. Efficiency of alchemical free energy simulations. I. A practical comparison of the exponential formula, thermodynamic integration, and Bennett's acceptance ratio method. *J. Comput. Chem.* **2011**, *32*, 1303−1319.
(57) Bruckner, S.; Boresch, S. Efficiency of alchemical free energy simulations. II. Improvements for thermodynamic integration. *J. Comput. Chem.* **2011**, *32*, 1320−1333.
(58) Kresse, G.; Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B: Condens. Matter* **1996**, *54*, 11169−11186.
(59) Kresse, G.; Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B: Condens. Matter* **1999**, *59*, 1758−1775.
(60) Kothakonda, M.; Kaplan, A. D.; Isaacs, E. B.; Bartel, C. J.; Furness, J. W.; Ning, J.; Wolverton, C.; Perdew, J. P.; Sun, J. Testing the r2SCAN Density Functional for the Thermodynamic Stability of Solids with and without a van der Waals Correction. *ACS Mater. Au* **2023**, *3* (2), 102−111.
(61) Kingsbury, R.; Gupta, A. S.; Bartel, C. J.; Munro, J. M.; Dwaraknath, S.; Horton, M.; Persson, K. A. Performance comparison of r2SCAN and SCAN metaGGA density functionals for solid materials via an automated, high-throughput computational workflow. *Phys. Rev. Mater.* **2022**, *6*, 013801.
(62) Caldeweyher, E.; Bannwarth, C.; Grimme, S. Extension of the D3 dispersion coefficient model. *J. Chem. Phys.* **2017**, *147* (3), 034112.
(63) Caldeweyher, E.; Ehlert, S.; Hansen, A.; Neugebauer, H.; Spicher, S.; Bannwarth, C.; Grimme, S. A generally applicable atomic-charge dependent London dispersion correction. *J. Chem. Phys.* **2019**, *150* (15), 154122.
(64) Caldeweyher, E.; Mewes, J.-M.; Ehlert, S.; Grimme, S. Extension and evaluation of the D4 London-dispersion model for periodic systems. *Phys. Chem. Chem. Phys.* **2020**, *22*, 8499−8512.
(65) Paszke, A., et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, 8-14 December 2019*; Curran Associates, Inc., 2019; pp 80248035.
(66) Haskins, J. B.; Thompson, A. E.; Lawson, J. W. Ab initio simulations of phase stability and martensitic transitions in NiTi. *Phys. Rev. B: Condens. Matter* **2016**, *94*, 214110.

https://doi.org/10.1021/acsomega.4c04782
ACS Omega 2024, 9, 36589−36599