Supplementary Information: Predicting the phase diagram of titanium dioxide with random search and pattern recognition

Aleks Reinhardt,[1] Chris J. Pickard,[2,3] and Bingqing Cheng[1,4,5,a)]

[1)]*Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge, CB2 1EW, United Kingdom*

[2)]*Department of Materials Science & Metallurgy, University of Cambridge, 27 Charles Babbage Road, Cambridge, CB3 0FS, United Kingdom*

[3)]*Advanced Institute for Materials Research, Tohoku University, Sendai, Japan*

[4)]*TCM Group, Cavendish Laboratory, University of Cambridge, J. J. Thomson Avenue, Cambridge, CB3 0HE, United Kingdom*

[5)]*Trinity College, Trinity Street, Cambridge, CB2 1TQ, United Kingdom*

(Dated: 8 May 2020)

CONTENTS

S1 OVERVIEW OF KNOWN TITANIUM DIOXIDE PHASES

Titanium dioxide is one of the most broadly used metal oxides.[1,2] In addition to the most commonly occurring polymorphs of TiO_2 – rutile and anatase –, a large number of both ambient- and high-pressure phases have been observed and hypothesised, including brookite,[3] TiO_2(B),[4] hollandite,[5] columbite,[6] baddeleyite,[7] TiO_2(OI),[8] fluorite,[9] oxygen-distorted fluorite (P$ca2_1$),[10] ramsdellite[11] and cotunnite.[12] Many of the high-pressure phases may have interesting optical properties,[13] and the cotunnite phase has been suggested, somewhat controversially,[14] to be the hardest known oxide.[12] Although phases such as columbite are sometimes described as 'high-pressure', in the sense that they are synthesised at pressures higher than atmospheric pressure, they nevertheless remain metastable when the pressure is reduced, and so can be of practical use even at ambient pressure.[15] Hence it is important to understand not only the limits of thermodynamic stability, but also metastability. An illustration of these previously known phases can be found in Ref. 16.

We use the Matsui–Akaogi (MA) empirical potential[17] of titanium dioxide, which was parameterised to reproduce some physical properties of rutile, anatase, brookite and columbite, such as their unit cell lattice parameters, densities, Ti–O distances, compressibilities and thermal expansivities.[17] In the context of the MA potential, we note that the oxygen-displaced fluorite (P$ca2_1$) structure is not (meta)stable,[16] and the ramsdellite and OI phases are somewhat different from the experimentally reported structures.[16]

S2 NEW PHASES OF TITANIUM DIOXIDE

In the random searches we found a total of 15 additional phases that have low enthalpies or high frequency of occurrence, following the ML pattern recognition schemes described in the main text. Those structures were first converted into a canonical representation (using Avogadro[18]), and then considered in more detail. We illustrate the 15 new structures in Fig. S1. We label them with a standard format `P-a-SYM`, where `P` gives the pressure in gigapascals at which the structure was found in RSS, `a` is a lowercase letter purely used for labelling, and `SYM` gives the space group of the structure.

a)Electronic mail: bc509@cam.ac.uk

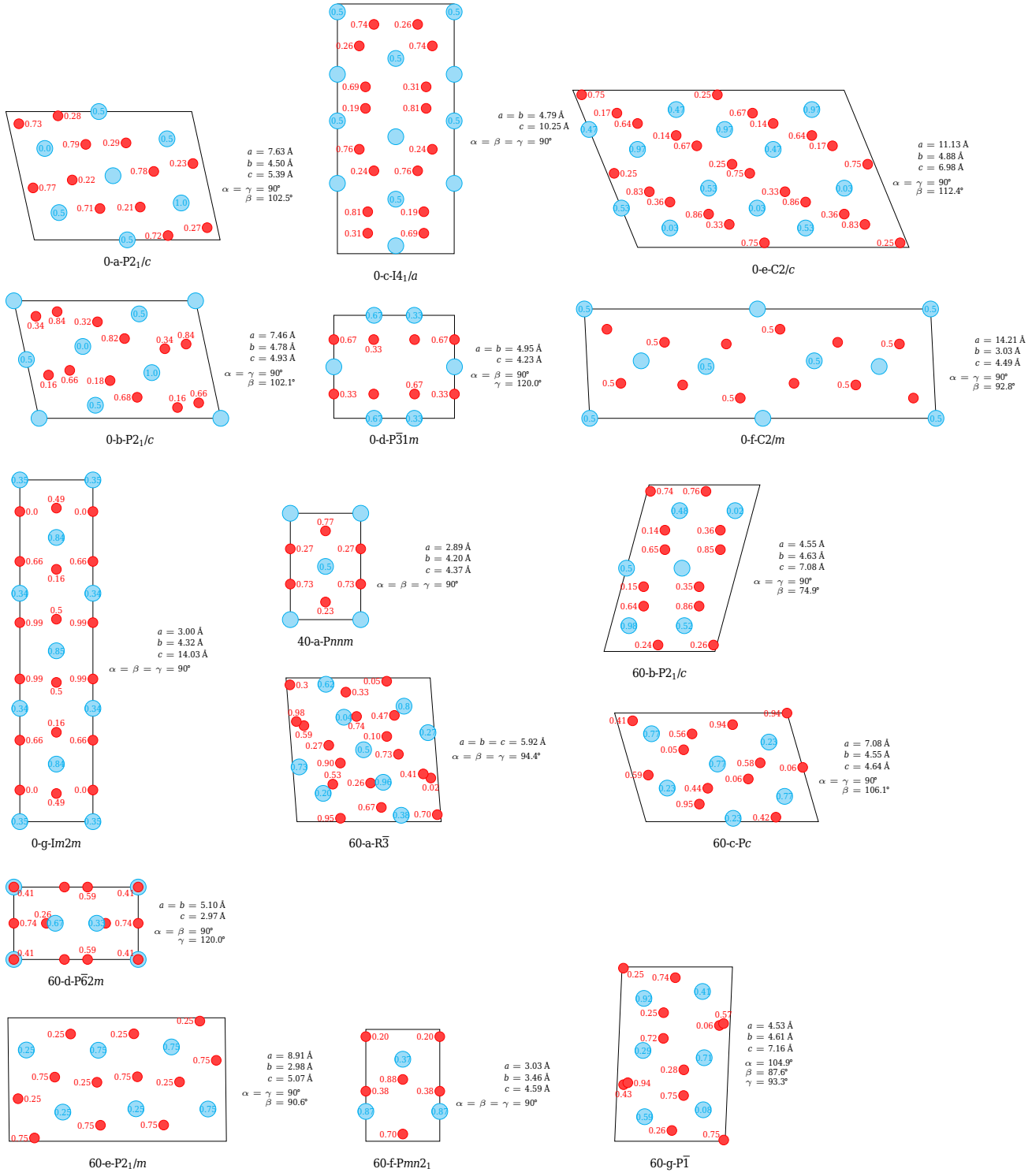

FIG. S1. Plan views of unit cells of the new crystal structures considered, as labelled. Relative heights above the plane are indicated as fractions of the lattice parameter along the b axis; where no height is indicated, this corresponds to a relative height of 0 and 1. Titanium atoms are shown in cyan and oxygen atoms in red. Full crystal structures are available in the supporting data.

In fact, the phase labelled 60-d-P$\bar{6}$2m is the Fe_2P phase of TiO_2 that was previously considered at higher pressures than we have focused on here;[13] however, as we have not previously computed its free energy, we consider it here together with the other 'new' phases.

S3 ENTHALPY OF TITANIUM DIOXIDE PHASES AT 0 K

In Fig. S2, we show the enthalpy of each of the phases considered at 0 K using the MA potential with energy optimisation as well as a machine-learning approach, as discussed in the main text. This figure is analogous to Fig. 3 of the main text, except that it is shown at absolute zero, where entropy plays no role.

By comparing the position of the curves between this figure and Fig. 3 of the main text, we can note that the entropy can significantly affect phase behaviour; for example, the OI analogue phase is entropically disfavoured and so is shifted upwards as the temperature increases.

In Fig. 3(a) of the main text, we report the chemical potentials of different TiO_2 phases at 1600 K. We have chosen to focus on a relatively high temperature because this is where the entropy of the system becomes more important to capture correctly; however, the higher the temperature, the greater the risk that some solid phases will cease to be metastable. For example, the baddeleyite phase and several of the newly considered phases [namely 0-c-I4$_1$/*a*, 0-e-C2/*c*, 60-c-P*c* and 60-e-P2$_1$/*m*] do not appear in Fig. 3(a) of the main text, because they are no longer metastable at such high temperatures. Baddeleyite, for example, spontaneously converts into pyrite at approximately 1150 K when heated at 20 GPa. Although the most stable of the solid phases can remain at least metastable with respect to the liquid phase up to several thousand kelvin, we have here chosen a temperature of 1600 K since the majority of the phases we have so far considered are still metastable, which thus provides a more challenging benchmark for the ML approach.

For most phases shown in Fig. 3(a) and Fig. S2, the curves do not extend throughout the entire pressure range because the polymorphs cease to be metastable with respect to other phases as a function of pressure. In this context, one particularly interesting transition is that between rutile and the new phase 40-a-P*nnm* (see Fig. S1), which are structurally particularly similar. The transition between them appears to be continuous as the system is pressurised: while the space group changes, there does not seem to be any change in density or enthalpy. In a sense, the 40-a-P*nnm* polymorph is thus merely a high-pressure analogue of the rutile phase.

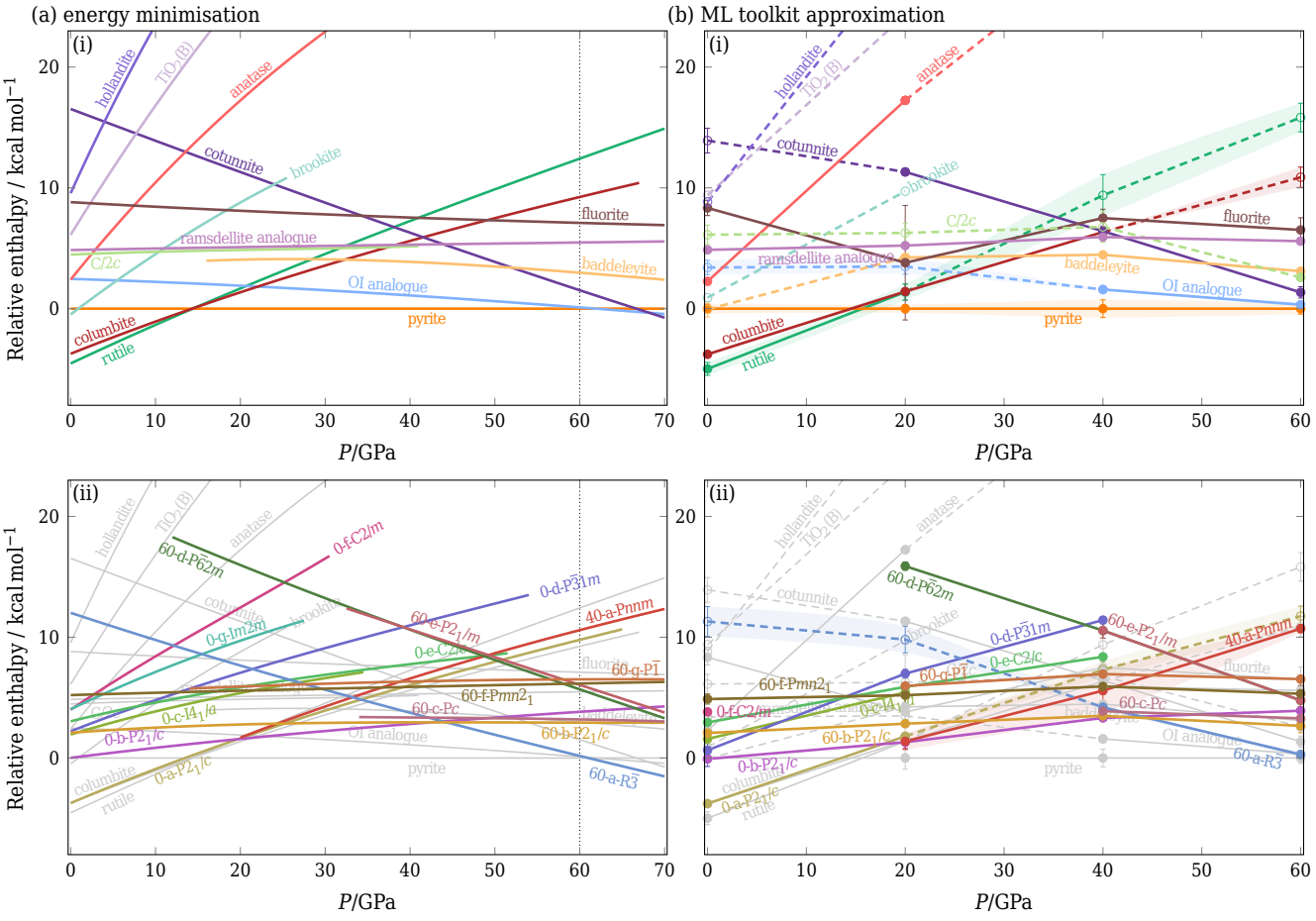

FIG. S2. Enthalpy at 0 K using (a) direct energy minimisation and (b) a machine-learning prediction, expressed per formula unit and relative to the pyrite phase. In each case, in panel (i), only the previously known phases are shown, and in panel (ii), new phases obtained from RSS are shown, with the structures from panel (i) repeated in grey. In (b), the error bar of potentially competitive phases is shown as a coloured band. Where a phase was not obtained at a given pressure from a random search, the marker is hollow and connected to other points by a dashed line. Note that the pressure axis covers different ranges in (a) and (b); dotted lines in (a) indicate the range of panel (b).

In Fig. S3, we show the enthalpy at 0 K from density-functional theory using three different functionals. On the whole, the agreement between the three is remarkably good, with a root mean square deviation of the order of 1.2 kcal mol^{-1}. The data are also consistent with previous DFT-based calculations[19–22] using a variety of functionals, suggesting that the DFT calculation is relatively robust with respect to the choice of functional. None of the newly obtained phases appears to be absolutely favoured in enthalpy; however, a number of phases are certainly competitive and may well be metastable.

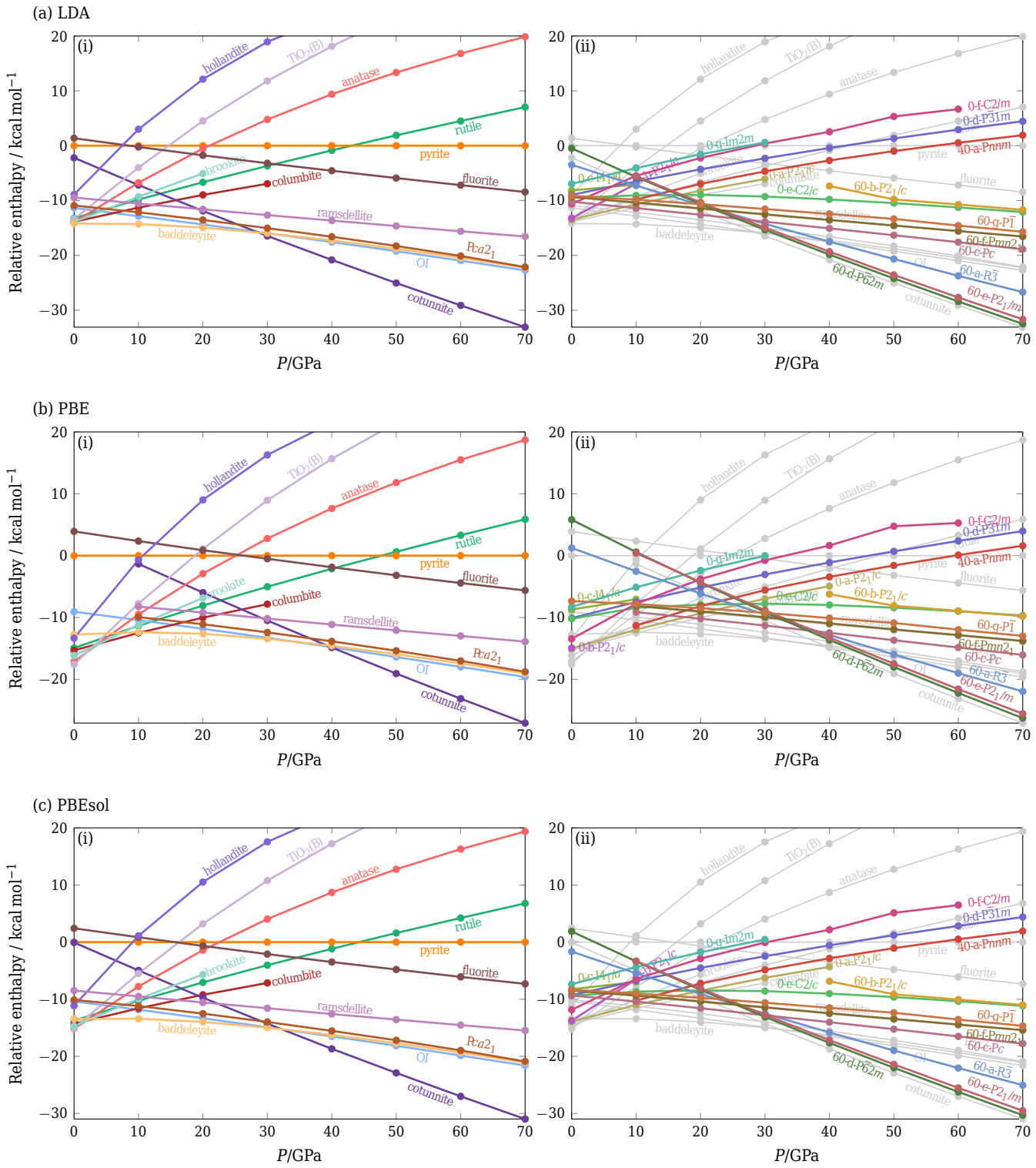

FIG. S3. Enthalpies from density-functional theory at 0 K using the (a) LDA, (b) PBE and (c) PBEsol functionals. In each case, in panel (i), only the previously known phases are shown, and in panel (ii), new phases obtained from RSS are shown, with the structures from panel (i) repeated in grey.

S4 BASIN AND THERMODYNAMIC ENTROPIES

In Fig. 2 of the main text, we compared the entropy estimated from the basin of attraction, $s_\mathrm{b} = k_\mathrm{B} \ln(v)/n$, and the thermodynamic entropy computed from the chemical potential, showing that the basin entropy estimate is slightly too large for low-enthalpy structures. For reference, in Fig. S4(a), we plot exactly the same data as in Fig. 2 of the main text, except that each point is labelled with phase, temperature and pressure information, rather than the enthalpy.

Because the minimisation scheme used in RSS can increase the frequency of finding low enthalpy structures, we fitted aH + constant = $s_\mathrm{b} - S$ over all data points to achieve a better estimate of S. With $a = 8 \times 10^{-5}\,\mathrm{K}^{-1}$, we obtain a Pearson correlation coefficient of $r \approx 0.7$, and we show a comparison between the adjusted entropy $s_\mathrm{b} - aH$ and S in Fig. S4(b).

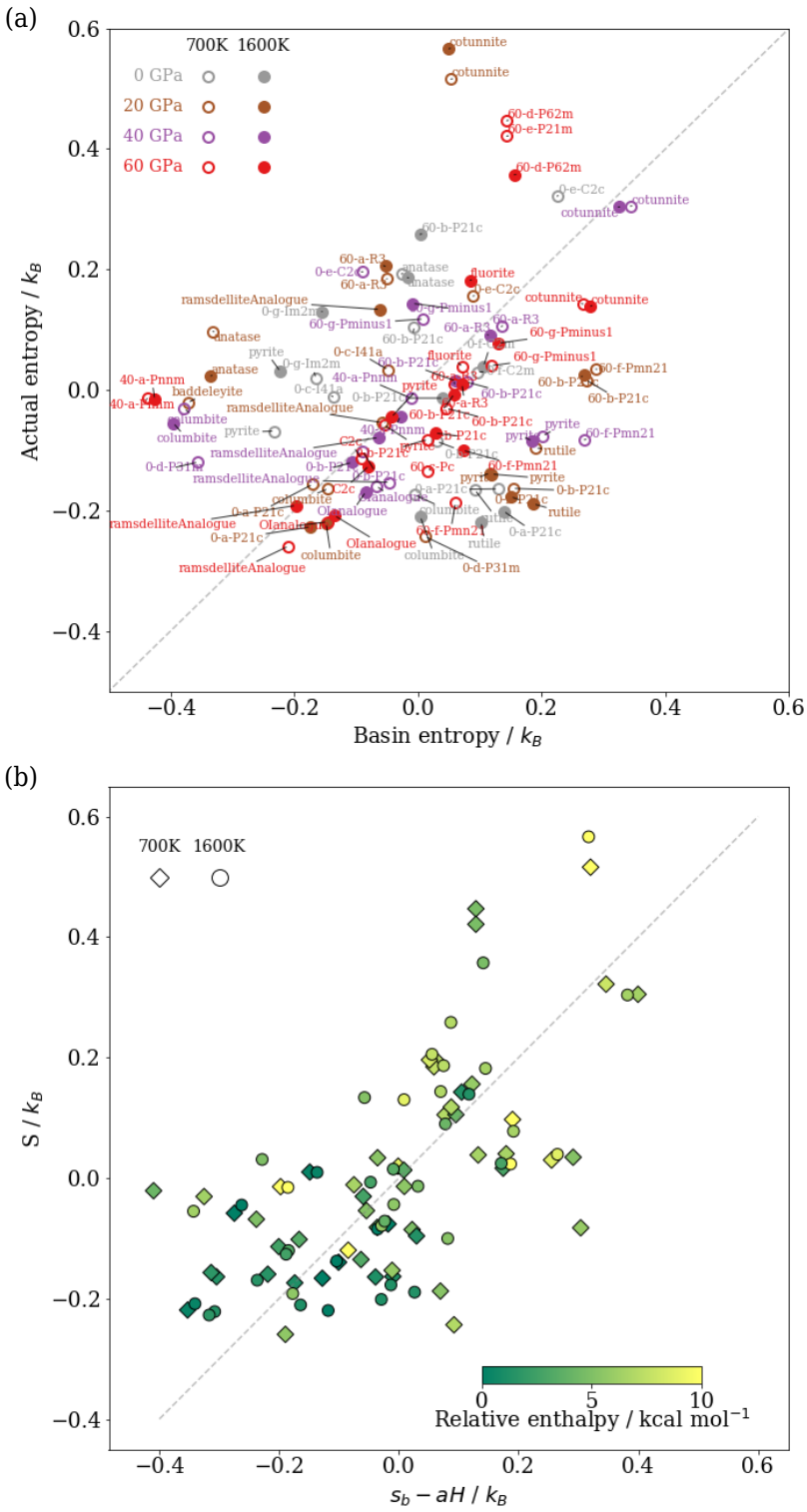

FIG. S4. (a) A comparison between the basin entropy s_b and the thermodynamic entropy S per formula unit of TiO_2. Each data point is labelled by the name of the structure. The colour scheme shows the pressure of the RSS and the free-energy calculations. Data are shown for 700 K (hollow markers) and 1600 K (solid markers). (b) A comparison between the adjusted entropy $s_\mathrm{b} - aH$ and S, where $a = 8 \times 10^{-5}\,\mathrm{K}^{-1}$. As only the relative entropy under a given thermodynamic condition between different phases matters, we centred the data about the origin.

S5 ML PREDICTIONS OF CHEMICAL POTENTIALS

In the main text, we described the kernel ridge regression (KRR) approach. We used KRR to to estimate the enthalpy and basin entropy per formula units of TiO_2 structures. Here we describe the detailed workflow for the KRR fitted on the enthalpy H data for structures generated from RSS at 0 GPa; the estimation of H or S_b at other pressures is essentially the same.

For the total of 3263 structures generated by RSS at 0 GPa, we first randomly partitioned them into a training set (90 % of the structures) and a test set (10 % of the structures). From the training set we select the $M = 200$ representative structures using the farthest-point sampling (FPS) approach. This approach ensures a near uniform and complete coverage of the hyper space of descriptors. In the fitting process of KRR, we set the regularisation matrix $\boldsymbol{\Lambda} = 0.01\boldsymbol{I}_N$, where $\boldsymbol{I}_N$ is an order N identity matrix. The left panel of Fig. S5 shows the outcome of such fitting. The fitting error in RMSE is about 0.8 kcal mol^{-1} in both the training and the test sets, and the correlation coefficients are both 0.99. The right panel shows the learning curve. It can be seen that the learning rate has not saturated; more data points can reduce the error further. Fig. S5 shows the analogous KRR for the prediction of the basin entropy s_b. Notice that we plot the exponential of this entropy in the left panel. The fitting error in RMSE is about $0.2k_B$ in both the training and the test sets, and the correlation coefficients are both 0.77. The ML predictions in Fig. S5 show larger scatter compared with the ones for enthalpy. This may be due to the systematic errors in the estimates of the actual s_b because of the finite sampling and the finite system sizes, or it may be because descriptions of the equilibrium configurations cannot fully capture the basin volume. To estimate the uncertainty of the prediction, we used the subsampling technique.[23] This method involves first creating an ensemble of models by randomly selecting a subset of the training data, and then computing the variance in the predictions of the models. In our case we created 8 models of size $N = 2000$, and compared the variance of the prediction on the test set with the absolute prediction errors of the test set. These procedures are implemented in the Python notebook included in the supporting data.

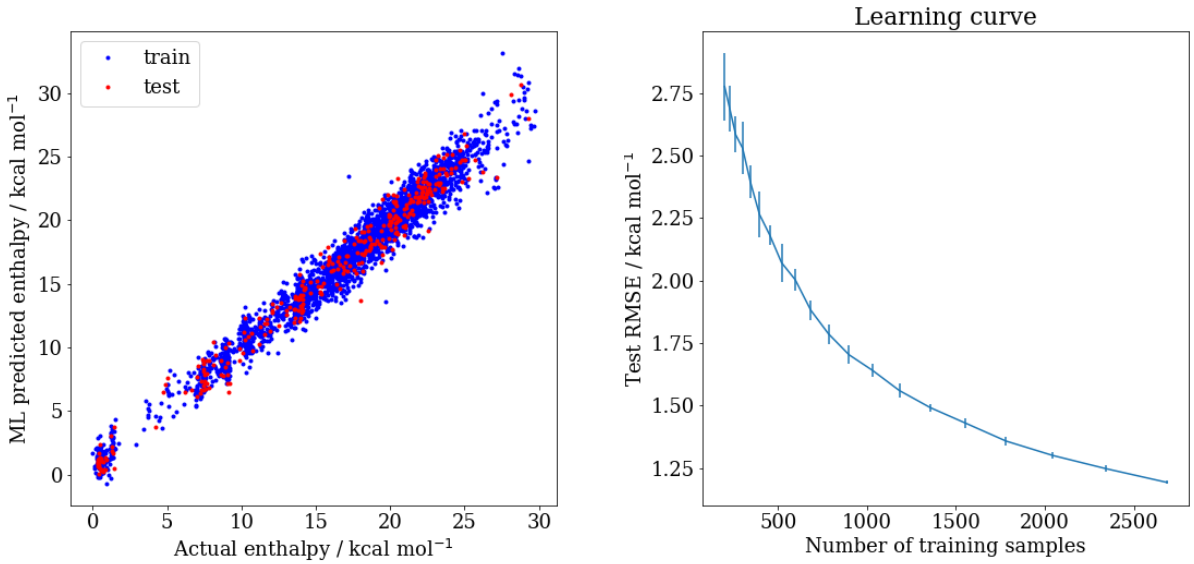

FIG. S5. Left: a comparison between the actual enthalpy and the predicted enthalpy per formula unit of TiO_2 for both the training set and the test set. Right: the learning curve of KRR. We show the evolution of the test root mean square error (RMSE) as the number of training points increases.

After performing the KRR procedures for both enthalpy and basin entropy for each of the four pressures (0, 20, 40, 60 GPa) separately, we estimate the chemical potentials and show these results in Fig. S7. For the results at each pressure, the filled blue markers show the chemical potential at 0 K (i.e. the enthalpy) of all the structures found at that pressure, and the hollow blue markers show the ML prediction for the structures not found. For a given found structure, the error estimate of its enthalpy is given by the standard deviations of the enthalpies of all structures in the cluster to which the structure belongs. For the uncertainty in entropy, we assume a constant value of $0.2k_B$, because that is the value of the root mean squared error in $S - S_b$ per formula unit. After that, we use $h(0\,\mathrm{K}) - T s_b$ to approximate its chemical potential, where h is the enthalpy per formula unit.

For structures that are not found, we first use the ML schemes outlined in the main text to estimate the enthalpy and basin entropy. The error in its enthalpy or entropy is estimated using the standard deviation of KRR models constructed using a subsampling technique.[23] Finally we again use $h(0\,\mathrm{K}) - T s_b$ to estimate the chemical potential.

In order to gauge the quality of the ML approach, we have also computed the standard deviation of the difference in chemical potentials estimated using ML and thermodynamic integration. For the data reported in Fig. 3 of the main text, Spearman's rank correlation coefficients between the two methods were 0.96 for those phases which were found in RSS (indicated by solid lines in Fig. 3) and 0.83 for those where the entropy was also estimated using a ML scheme (indicated by dashed lines in Fig. 3).

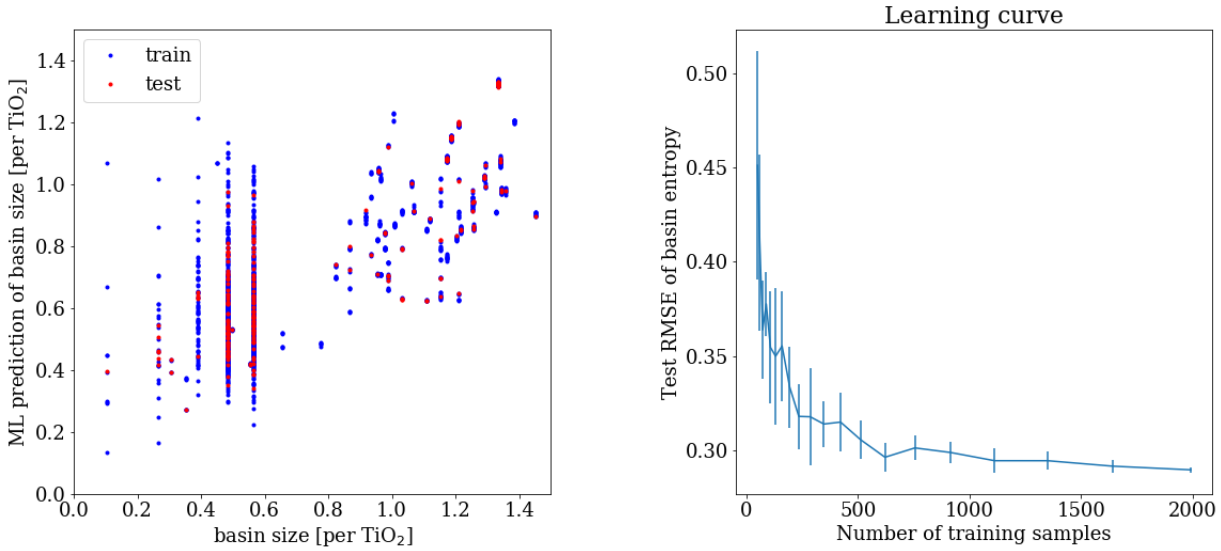

FIG. S6. Left: a comparison between the basin volume (equal to $\exp(s_b/k_B)$) per formula unit of TiO_2 and the predicted values for both the training set and the test set. Right: the learning curve of KRR for the basin entropy s_b, the unit is in k_B. We show the evolution of the test root mean square error (RMSE) as the number of training points increases.

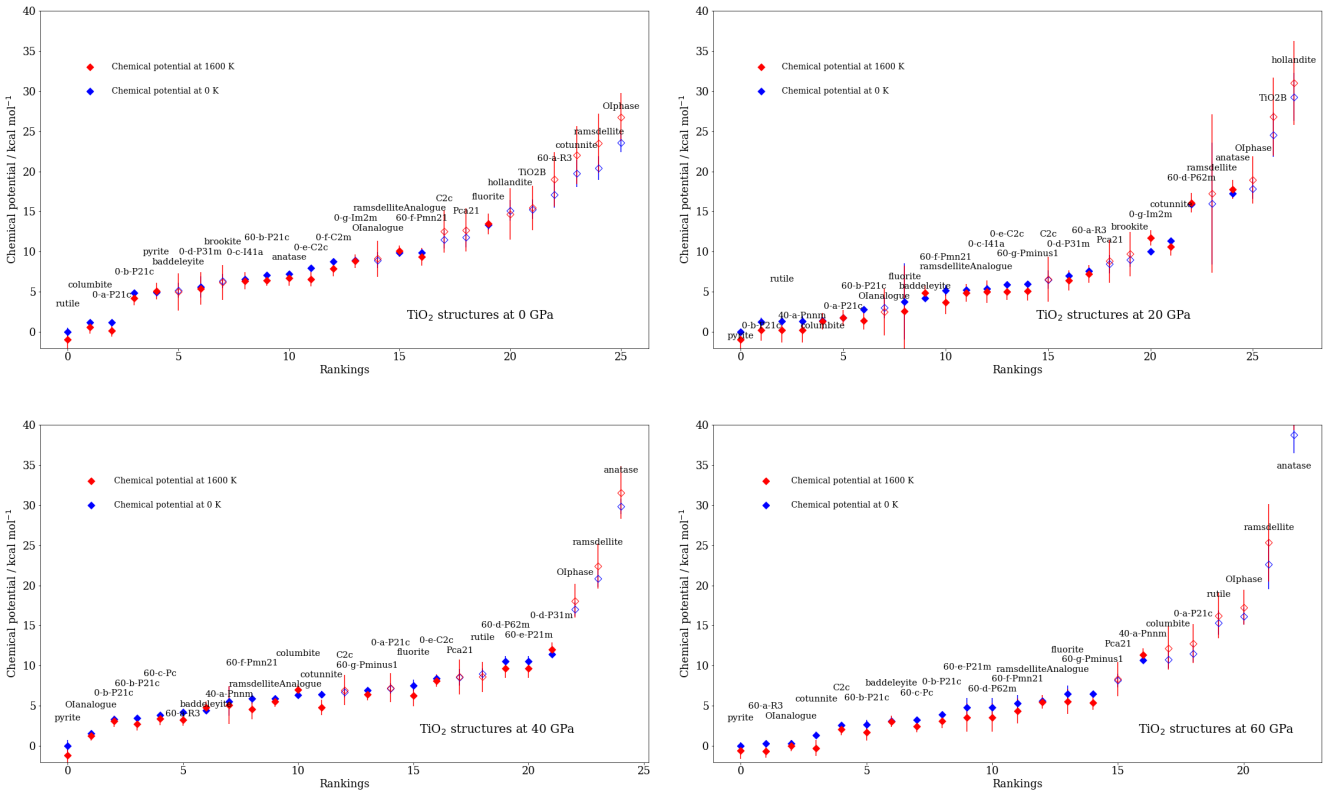

FIG. S7. The chemical potentials at 1600 K (red markers) and at 0 K (blue markers) of the phases at four pressures. For clarity, we ordered the structures at each pressure according their chemical potentials at 0 K (i.e. the enthalpies). Solid markers denote structures found at the particular pressure, and hollow markers mean the converse. If a structure is not found, its entropy and enthalpy are predicted using the KRR ML method.

The standard deviation of the difference between the two averaged over all pressures was 1.1 kcal mol^{-1} for the former and 1.7 kcal mol^{-1} for the latter. These standard deviations are comparable to the error in the predictions themselves (see Fig. S5).

S6 PHASE DIAGRAM OF TITANIUM DIOXIDE

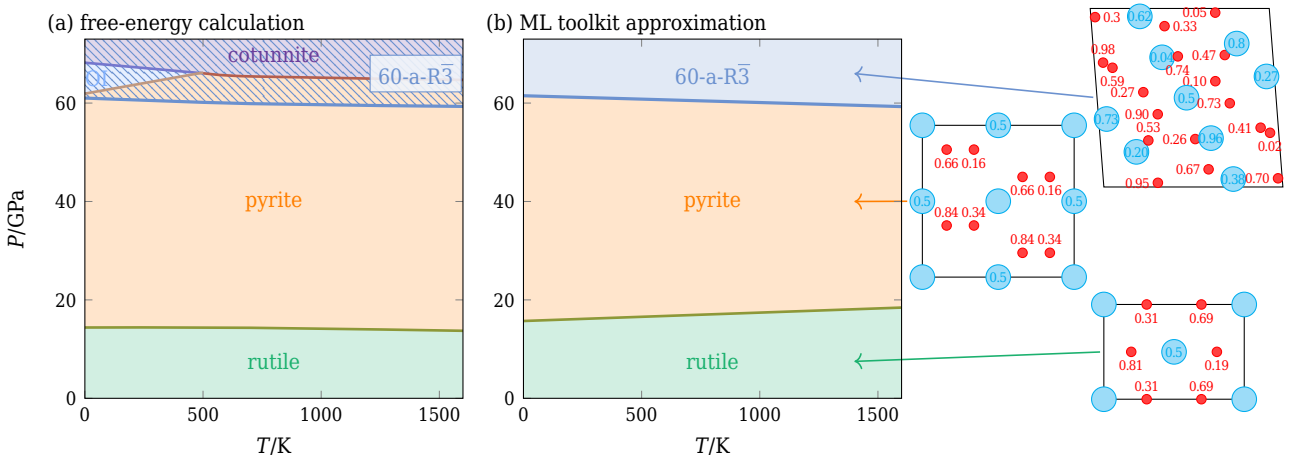

FIG. S8. Pressure–temperature phase diagrams from (a) free-energy calculations and from (b) the ML toolkit approximation. Plan views of the stable crystal structures are also shown for illustration.

In Fig. S8, we estimate the phase diagram of TiO_2 using both free-energy calculations and a machine-learning prediction. The diagram in Fig. S8(a) also shows the previously computed phase diagram (Ref. 16), with a triple point between the pyrite, OI (analogue) and cotunnite phases. Using the new structure obtained from RSS, however, this triple point can be identified to be metastable, and the 60-a-R$\bar{3}$ phase is stable at sufficiently high pressures; this is denoted by the hatched region in the figure. The agreement between the phase diagrams shown in panels (a) and (b) is rather remarkable. However, it should be borne in mind that this result is somewhat fortuitous: the 60-a-R$\bar{3}$ phase is conveniently more stable than the OI and cotunnite phases, and the ML approach would not have correctly predicted the triple point of the pyrite, OI and cotunnite phases. The prediction of the chemical potential of each phase, shown in the main manuscript, is a rather more rigorous test of the approach.

[1] J. H. Braun, A. Baidins, and R. E. Marganski, "TiO_2 pigment technology: A review," Prog. Org. Coat. **20**, 105 (1992).
[2] N. Rahimi, R. A. Pax, and E. M. Gray, "Review of functional titanium oxides. I: TiO_2 and its modifications," Prog. Solid State Chem. **44**, 86 (2016).
[3] E. P. Meagher and G. A. Lager, "Polyhedral thermal expansion in the TiO_2 polymorphs: Refinement of the crystal structures of rutile and brookite at high temperature," Can. Mineral. **17**, 77 (1979).
[4] R. Marchand, L. Brohan, and M. Tournoux, "TiO_2(B) A new form of titanium dioxide and the potassium octatitanate $K_2Ti_8O_{17}$," Mater. Res. Bull. **15**, 1129 (1980).
[5] M. Latroche, L. Brohan, R. Marchand, and M. Tournoux, "New hollandite oxides: TiO_2(H) and $K_{0.06}TiO_2$," J. Solid State Chem. **81**, 78 (1989); T. Sasaki, M. Watanabe, and Y. Fujiki, "Structure of $K_{1.0}Ti_8O_{16}$ and $K_{0.0}Ti_8O_{16}$," Acta Crystallogr. **B49**, 838 (1993).
[6] N. A. Bendeliani, S. V. Popova, and L. F. Vereshchagin, "New modification of titanium dioxide obtained at high pressures," Geochem. Int. **3**, 387 (1966); I. E. Grey, C. Li, I. C. Madsen, and G. Braunshausen, "TiO_2-II. Ambient pressure preparation and structure refinement," Mater. Res. Bull. **23**, 743 (1988).
[7] H. Sato, S. Endo, M. Sugiyama, T. Kikegawa, O. Shimomura, and K. Kusaba, "Baddeleyite-type high-pressure phase of TiO_2," Science **251**, 786 (1991); V. Swamy, L. S. Dubrovinsky, N. A. Dubrovinskaia, F. Langenhorst, A. S. Simionovici, M. Drakopoulos, V. Dmitriev, and H.-P. Weber, "Size effects on the structure and phase transition behavior of baddeleyite TiO_2," Solid State Commun. **134**, 541 (2005).
[8] N. A. Dubrovinskaia, L. S. Dubrovinsky, R. Ahuja, V. B. Prokopenko, V. Dmitriev, H.-P. Weber, J. M. Osorio-Guillen, and B. Johansson, "Experimental and theoretical identification of a new high-pressure TiO_2 polymorph," Phys. Rev. Lett. **87**, 275501 (2001).
[9] M. Mattesini, J. S. de Almeida, L. Dubrovinsky, N. Dubrovinskaia, B. Johansson, and R. Ahuja, "High-pressure and high-temperature synthesis of the cubic TiO_2 polymorph," Phys. Rev. B **70**, 212101 (2004).
[10] X.-F. Zhou, X. Dong, G.-R. Qian, L. Zhang, Y. Tian, and H.-T. Wang, "Unusual compression behavior of TiO_2 polymorphs from first principles," Phys. Rev. B **82**, 060102 (2010).
[11] J. Akimoto, Y. Gotoh, Y. Oosawa, N. Nonose, T. Kumagai, K. Aoki, and H. Takei, "Topotactic oxidation of ramsdellite-type $Li_{0.5}TiO_2$, a new polymorph of titanium dioxide: TiO_2(R)," J. Solid State Chem. **113**, 27 (1994).
[12] L. S. Dubrovinsky, N. A. Dubrovinskaia, V. Swamy, J. Muscat, N. M. Harrison, R. Ahuja, B. Holm, and B. Johansson, "The hardest known oxide," Nature **410**, 653 (2001).
[13] H. Dekura, T. Tsuchiya, Y. Kuwayama, and J. Tsuchiya, "Theoretical and experimental evidence for a new post-cotunnite phase of titanium dioxide with significant optical absorption," Phys. Rev. Lett. **107**, 045701 (2011).
[14] D. Nishio-Hamane, A. Shimizu, R. Nakahira, K. Niwa, A. Sano-Furukawa, T. Okada, T. Yagi, and T. Kikegawa, "The stability and equation of state for the cotunnite phase of TiO_2 up to 70 GPa," Phys. Chem. Miner. **37**, 129 (2010).
[15] M. Y. Manuputty, J. A. H. Dreyer, Y. Sheng, E. J. Bringley, M. L. Botero, J. Akroyd, and M. Kraft, "Polymorphism of nanocrystalline TiO_2 prepared in a stagnation flame: Formation of the TiO_2-II phase," Chem. Sci. **10**, 1342 (2019).
[16] A. Reinhardt, "Phase behavior of empirical potentials of titanium dioxide," J. Chem. Phys. **151**, 064505 (2019).
[17] M. Matsui and M. Akaogi, "Molecular dynamics simulation of the structural and physical properties of the four polymorphs of TiO_2," Mol. Simul. **6**, 239 (1991).
[18] M. D. Hanwell, D. E. Curtis, D. C. Lonie, T. Vandermeersch, E. Zurek, and G. R. Hutchison, "Avogadro: An advanced semantic chemical editor, visualization, and analysis platform," J. Cheminformatics **4**, 17 (2012).
[19] Z.-G. Mei, Y. Wang, S. Shang, and Z.-K. Liu, "First-principles study of the mechanical properties and phase stability of TiO_2," Comput. Mater. Sci. **83**, 114 (2014).
[20] Z. Fu, Y. Liang, S. Wang, and Z. Zhong, "Structural phase transition and mechanical properties of TiO_2 under high pressure," Phys. Status Solidi B **250**, 2206 (2013).
[21] X. G. Ma, P. Liang, L. Miao, S. W. Bie, C. K. Zhang, L. Xu, and J. J. Jiang, "Pressure-induced phase transition and elastic properties of TiO_2 polymorphs," Phys. Status Solidi B **246**, 2132 (2009).
[22] T. Zhu and S.-P. Gao, "The stability, electronic structure, and optical property of TiO_2 polymorphs," J. Phys. Chem. C **118**, 11385 (2014).
[23] F. Musil, M. J. Willatt, M. A. Langovoy, and M. Ceriotti, "Fast and accurate uncertainty estimation in chemical machine learning," J. Chem. Theory Comput. **15**, 906 (2019).