ROYAL SOCIETY
OF CHEMISTRY

Check for updates

Powder X-ray diffraction assisted evolutionary algorithm for crystal structure prediction†

Stefano Racioppi, [ID] [a] Alberto Otero-de-la-Roza, [ID] [b] Samad Hajinazar [ID] [a] and Eva Zurek [ID] *[a]

Experimentally obtained powder X-ray diffraction (PXRD) patterns can be difficult to solve, precluding the full characterization of materials, pharmaceuticals, and geological compounds. Herein, we propose a method based upon a multi-objective evolutionary search that uses both a structure's enthalpy and similarity to a reference PXRD pattern (constituted by a list of peak positions and their intensities) to facilitate structure solution of inorganic systems. Because the similarity index is computed for locally optimized cells that are subsequently distorted to find the best match with the reference, this process transcends both computational (*e.g.*, choice of theoretical method, and 0 K approximation) and experimental (*e.g.*, external stimuli, and metastability) limitations. We illustrate how the proposed methodology can be employed to successfully uncover complex crystal structures by applying it to a range of test cases, including inorganic minerals, elements ramp-compressed to extreme conditions, and molecular crystals. The results demonstrate that our approach not only improves the accuracy of structure prediction, but also significantly reduces the time required to achieve reliable solutions, thus providing a powerful tool for the advancement of materials science and related fields.

1 Introduction

The crystal structure of a compound is key for predicting and rationalizing its properties.[1–3] Therefore, crystal structure determination is one of the bedrocks upon which chemistry, materials science, physics, as well as earth and planetary science is based. Indeed, common to all of these fields is the need for characterizing the structure of the chemical system using various spectroscopies. While methods such as Raman, Infra-Red or Nuclear Magnetic Resonance Spectroscopy provide information that can indirectly deduce the structural motifs present, only diffraction is directly related to the atomic positions. Various diffraction techniques are available, varying according to the scattering source (X-rays, neutrons or electrons) or from the nature of the sample (powder, single crystal or even liquid). Diffraction from single crystals is the gold standard, but in practice it can be difficult or impossible to obtain single crystals of adequate quality and size to achieve a reliable structure solution. Therefore, the possibility of obtaining structural information from microcrystalline powder-like samples becomes important.[4] However, unlike diffraction from single crystals, powder X-ray diffraction (PXRD) is typically not sensitive enough to provide information about the positions of light elements such as hydrogen and lithium (if combined with other heavier elements), nor can it differentiate between elements with similar mass numbers. Though neutron diffraction is sensitive to the location of light elements it requires large samples. Thus, PXRD is the most commonly used tool to deduce the structural information of battery materials, superconductors, minerals found in the deep Earth, pharmaceutical drugs, and more.[1,5–7]

When a good quality PXRD pattern is in-hand, it is relatively easy to retrieve information on the size of the unit cell, but a structure solution with refinement of the atomic positions remains, to date, a challenging procedure. The inherent limitation of PXRD lies in its projection of three-dimensional diffraction data onto a one-dimensional scale when measuring powder samples, often resulting in peak overlap.[4,8] To perform such refinements, various techniques[9–11] are available, from those developed by Rietveld[12] or Le Bail[13] to modellings based on reverse Monte Carlo,[4,14] genetic[15,16] or machine learning[17,18] algorithms. Nonetheless, crystal structure solution from PXRD data remains a grand challenge in crystallography. Further complicating structural characterization is the presence of mixed-phases, sample peaks that are obscured by ones originating from the experimental apparatus, noisy background of the diffractogram, and preferred orientation of the micro-crystallites. These situations are common, for example, when compounds are synthesized for the first time, matter is compressed within diamond anvil cells or in dynamic (shock or

ªDepartment of Chemistry, State University of New York at Buffalo, Buffalo, New York 14260-3000, USA. E-mail: ezurek@buffalo.edu

ᵇDepartmento de Química Física y Analítica and MALTA Consolider Team, Facultad de Química, Universidad de Oviedo, 33006 Oviedo, Spain

† Electronic supplementary information (ESI) available. See DOI: https://doi.org/10.1039/d4dd00269e

ramp) compression experiments, or simply due to the morphology of the crystallites. Because of these difficulties, theoretical calculations have become useful tools to assist structure solution given a PXRD pattern. We also mention a promising route, which has been applied to organic molecular crystals, where structures are determined directly from chemical shifts, aided by machine learning.[19,20]

Another strategy, popular especially in the high-pressure field, is based upon crystal structure prediction (CSP) algorithms, which aim to locate the most stable atomic configuration for a user-defined chemical composition at a given pressure and at zero temperature. Some of the most popular techniques include random or evolutionary searches, particle-swarm optimization, and Monte-Carlo or molecular dynamics based algorithms.[21] In the family of evolutionary algorithms, a fitness is assigned to each DFT-optimized structure, and this fitness is related to the structure's likelihood to be chosen as a parent for the next generation.[22] In a traditional evolutionary search, the fitness is obtained from the energy (or enthalpy) of the system relative to (a subset) of those that have been optimized. This fitness is crucial in driving the algorithm towards promising regions of the energy landscape in the search for thermodynamically stable structures.

However, not all of the compounds that are predicted to be the most stable are necessarily those that are experimentally observed. When compared to PXRD diffractograms, this discrepancy can be attributed to several factors, including the numerous approximations involved in the computations (*e.g.*, choice of level of theory, pseudopotential, and the neglect of finite temperature contributions), as well as variations in synthetic and experimental conditions.[23] As a result, achieving the closest match with the experimental data often necessitates screening many metastable phases, especially in the case of polymorphism, where the differences in energy between them may be minimal.[24,25] This laborious manual screening process carries the risk of overlooking the optimal matching structure amid the hundreds or even thousands of predicted structures.

To circumvent this challenge and potentially steer the structure search towards a better match a constrained algorithm that focuses the search on structures with particular features (Bravais lattices, space groups, coordination numbers) has been proposed.[26] However, a guiding CSP algorithm, which employs both the experimental PXRD and the DFT-calculated energy and structure simultaneously, and in an equal footing, could prove even more beneficial.

Perhaps the most intricate method proposed to date is the first-principle-assisted structure solution (FPASS) technique,[27,28] which combines DFT calculations with experimental XRD data and statistical symmetry information in a genetic algorithm for structure determination. Similar methods have followed during the years, retaining the philosophy of combining information gleaned from diffraction data (lattice parameter, symmetry, stoichiometry, *etc.*) to reduce the space of a structure search with DFT optimization.[25,29] A similarity index calculated between experimental and simulated PXRD patterns has been exploited in particle swarm optimization (PSO),[30] utilizing a weighted cross-correlation function to re-evaluate the velocity

of each structure in the crystal structure search. With this technique, the simulated phases that best match the experimental PXRD pattern lead the PSO search. This methodology was shown to aid the prediction of the ground state phases of ZnO and TiO_2,[30] however the authors did not report if their discovery was accelerated compared to a standard energy-only-search, or if this algorithm could aid in the recognition of metastable phases. Non-CSP-based approaches have also been proposed, such as molecular dynamics simulations biased by experimental diffraction data.[31] For molecular organic crystals fast dispersion-corrected DFT optimizations have been proposed and used to improve the fit with the experimental PXRD patterns.[32]

In the present study, we outline our approach for enhancing CSP by leveraging PXRD data through a synergistic combination of the XtalOpt evolutionary algorithm[22,33] and the variable-cell Gaussian powder-based similarity index (*VC-GPWDF*)[34] implemented in the critic2 program.[35] Confusingly, despite the label "similarity index", this and similar methods actually calculate the *dissimilarity* between two patterns. By using the multi-objective search capability embedded in XtalOpt,[36] we illustrate the CSP search is able to accelerate the structural recognition of both experimental and simulated PXRD patterns. This strategy goes beyond the aforementioned methodologies developed to assist CSP using crystallographic and diffraction information.[27,30] Specifically, it overcomes many of the challenges that result from the comparison of experimental diffraction data collected at finite temperature and pressure with *in silico* patterns calculated for geometry-optimized structures at 0 K. Therefore, this technique becomes particularly advantageous when the reference PXRD diffractogram diverges from the computed ground state structure of a specific stoichiometry due to experimental conditions (pressure, hydrostaticity, temperature, *etc.*) or because of theoretical limitations; or for finding metastable phases.

2   Computational approach

2.1   Multi-objective search

The foundation of our newly proposed technique is based upon multi-objective global optimization,[37] as implemented in the XtalOpt code version 13.0.[36] In this extension of the XtalOpt evolutionary algorithm, the fitness of an individual structure can be based upon multiple objectives, including a structure's energy or enthalpy, as well as other user-specified features. After locally relaxing a structure, XtalOpt automatically calls the external codes specified by the user to compute the desired target properties, whose values are employed in conjunction with the enthalpy to calculate the corresponding multi-objective fitness. In the present work, the similarity in the PXRD pattern of a structure compared to that of a reference (S) is chosen as an objective to be minimized, while the enthalpy (H) is simultaneously minimized (eqn (1)). With S_s and H_s representing the numerical value of the similarity index and enthalpy of structure s, respectively, the fitness is defined through the following weighted sum:

$$f_s = w\left(\frac{S_{max} - S_s}{S_{max} - S_{min}}\right) + (1 - w)\left(\frac{H_{max} - H_s}{H_{max} - H_{min}}\right). \qquad (1)$$

Here, w is the weight assigned to the PXRD similarity objective, and A_{min} and A_{max} represent the minimum and maximum value of the objective $\{A = S, H\}$ for the pool of structures. The weights of the objective are constrained to be a real number between 0 and 1, and their sum must equal 1 for the calculation of the fitness, f_s. This fitness measure, subsequently, is used by XtalOpt to evaluate the suitability of candidate structures for the selection of the parent pool, from which new structures are produced by applying various evolutionary operations, as described more fully in ref. 33.

2.2 PXRD-assisted crystal structure prediction

We introduce here the *XtalOpt-VC-GPWDF* coupled technique, whose schematic workflow is illustrated in Fig. 1, to conduct PXRD similarity tests during the execution of the crystal structure search. In this multi-objective search, the energy or enthalpy (H_s in eqn (1)) is obtained from any external optimizer of periodic systems (herein, we employ the Vienna *Ab initio* Simulation Package, VASP, see also Computational details).[38] The similarity of a structure's simulated PXRD pattern with that of a reference (S_s in eqn (1)) is obtained using the newly developed variable-cell Gaussian powder-based similarity index (*VC-GPWDF*),[34] a modified version of de Gelder's similarity index,[39] that ranges from 0 for identical structures to 1 for maximum dissimilarity. Analogous to other methods,[40–42] the similarity is evaluated between a reference PXRD diffractogram, which can be either experimental or computer-generated, and a second diffractogram computed from one of the XtalOpt predicted structures by critic2,[35] which also handles the *VC-GPWDF* similarity index calculation. The reference diffractogram is input as an external list of values containing the 2θ angle of diffraction, and its corresponding relative intensity $[2\theta; I]$, and then parsed by critic2. The list does not need to be continuous, nor does it need to cover the whole PXRD diffractogram range. In fact, a short list of a few specific indexed peaks, or just fragments of a PXRD diffractogram, are valid inputs as well.

This aspect facilitates the PXRD-assisted crystal structure search, and is particularly useful when multiple phases are present in the sample, or when the noise is large in the experimental data.

The initial pool of structures can be generated internally using the RandSpg algorithm,[43] or externally (*e.g.*, by PyXtal[44]), with the structures subsequently being read in as "seeds". Local geometry relaxations can be performed with any external code for periodic systems. The similarity index, which is used to determine the fitness of the offspring, is calculated by critic2 using the optimized geometries of the structures generated by XtalOpt and the input $[2\theta; I]$ list. The use of the similarity index in the calculation of the fitness will steer the evolutionary algorithm to favor those structures that better represent the experimental PXRD as structures for the next generation. In our implementation the nature of the (dis)similarity does not alter the evolutionary operators used or their probability, though one can imagine that could.

In the original version of de Gelder's method, a similarity index is obtained by integrating numerically the product of the powder patterns of both crystals. It assumes a value of zero if the two patterns are the same (indicating that the crystals are identical) and a value of one for completely non-overlapping patterns. In the modified versions of de Gelder's method (*GPWDF* and *VC-GPWDF*) the similarity index is calculated analytically, rather than using numerical integration, from the list of reflection angles and intensities.[34] This approach is more efficient compared to de Gelder's, since it allows for the computation of analytical derivatives with respect to the structural parameters, enabling the easy minimization of the similarity index as a function of structural parameters like the lattice strain – which is the basis of the *VC-GPWDF* approach.

The *VC-GPWDF* method is exhaustively described in ref. 34; herein we provide a brief summary. In *VC-GPWDF*, the reference powder pattern is first pre-processed to extract the reflection angle and intensity pairs. This pre-processing step helps in circumventing the difficulties caused when comparing the pattern of calculated structures with experimental patterns that contain artifacts such as noise, varying peak shapes and widths, and a background contribution (in the present work, the user inputs a list of $[2\theta; I]$, as previously described). Given the experimental reflection angles and intensities, *VC-GPWDF* seeks lattice deformations (strains) of the input structure that maximize agreement with the experimental pattern. This is done by carrying out a global minimization of the *GPWDF* score as a function of the six lattice strains in the input structure, up to a maximum strain chosen by the user (in the present work, we employed the default values of 10% maximum deformation over the cell length and 5° for the angles). By doing this, *VC-GPWDF* can measure crystal similarity between structures that are not entirely comparable due to cell deformations because of, for instance, being determined under different experimental conditions or, in the case of structures calculated with DFT, due to missing vibrational effects. The final *VC-GPWDF* score is the smallest *GPWDF* among all deformations, and is equal to zero for identical crystals (after deformation). *VC-GPWDF* also produces the cell deformation of the input structure that
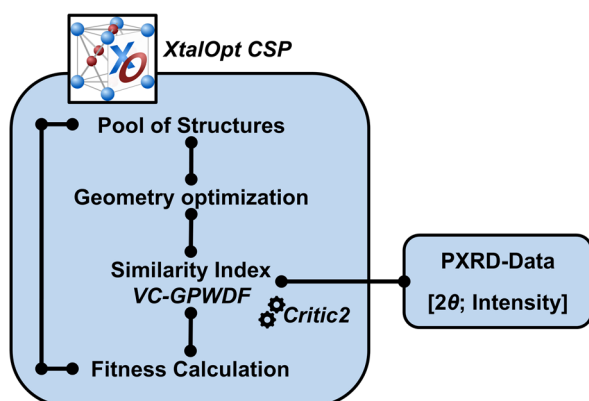

Fig. 1 Schematic workflow for conducting the PXRD-assisted crystal structure search with the combined *XtalOpt-VC-GPWDF* method.

maximizes agreement with the target pattern. Compared to other programs used to perform structure determination using cross-correlation functions, like the Fit with Deviating Lattice parameters (FIDEL),[45] *VC-GPWDF* has two advantages: (i) it performs a global optimization, and therefore is unlikely to get caught in a local minimum, and (ii) it uses an analytical version of de Gelder's index, enabling fast local optimizations. The whole *VC-GPWDF* procedure has very low computational cost, and can take seconds/minutes per structure on a desktop computer.

Let us now illustrate the power of this new technique for three inorganic systems: (1) brookite, a metastable polymorph of TiO_2; (2) sodium ramp-compressed to hundreds of GPa of pressure; and (3) vaterite, a natural polytypic structure of calcium carbonate. Each system will be introduced and discussed in-depth in their specific sections, while the complete computational details are reported at the end of this work. These systems were chosen to limit the computational cost while demonstrating the advantages provided by our method, since the number of local minima scales exponentially with the degrees of freedom, $3N + 3$ for a 3D crystal.[46] In fact, the number of atoms in our CSP searches is not greater than 60 per unit cell, reached by $CaCO_3$ with $Z = 12$ (see Computational details). However, the PXRD-assisted strategy could, in principle, be applied to larger systems (depending on the available computational resources, and on the method used for local relaxation), for which the acceleration would be even more evident compared to random or unassisted searches.

3 Results and discussion

3.1 TiO₂ – brookite

TiO_2 naturally exists in three different polymorphs at ambient conditions: anatase ($I4_1/amd$), brookite ($Pbca$) and rutile ($P4_2/mnm$) with 4, 8 and 2 formula units (FUs), respectively, in their conventional unit cells (Fig. 2). These polymorphs have been frequently used as benchmarks for CSP methods and related computational models.[22,26,30,43] At the PBE level of theory, we predict anatase as the ground state, followed by brookite ($\Delta E = 13.5$ meV per atom) and rutile ($\Delta E = 26.7$ meV per atom), in-line with previous DFT calculations.[26] Some classic interatomic potentials developed for this system, however, predict different stability orderings.[26]

A typical CSP search performed on TiO_2 will almost certainly locate the ground state phase, in this case, anatase (within the PBE approximation and ensuring that the FUs considered in the search include multiples of 4). While a standard CSP search is also likely to discover metastable rutile, owing to its high-symmetry and small unit cell, brookite may be difficult to find due to its metastability and low symmetry. Indeed, in an earlier study, a regular CSP search on TiO_2 with 8 FU found anatase, rutile and brookite as the 187th, 559th and 1141st crystals optimized, respectively.[26] A constrained search, where the parent pool was restricted to those structures that possessed an orthorhombic Bravais lattice only, accelerated the discovery of brookite (as the 345th structure). Another way in which brookite might be found is by steering an evolutionary algorithm with
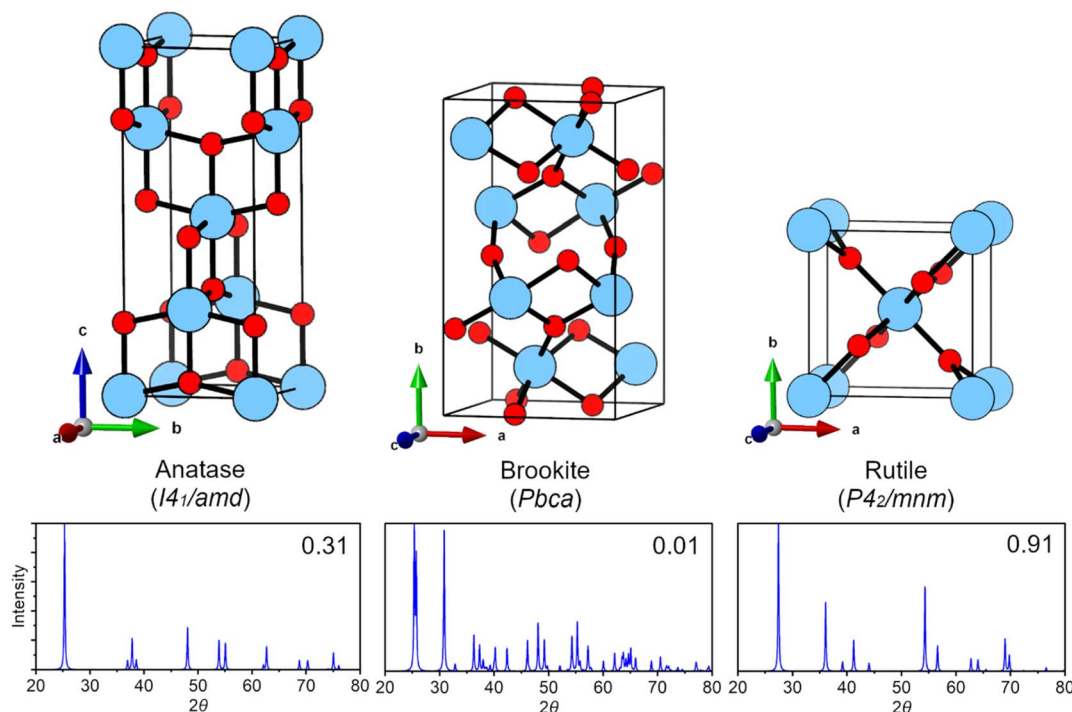

Fig. 2 Polymorphs of TiO₂ and their simulated powder X-ray diffraction (PXRD) patterns. Conventional unit cells of the three natural polymorphs of TiO₂ at ambient conditions: anatase ($I4_1/amd$, 4 FU), brookite ($Pbca$, 8 FU) and rutile ($P4_2/mnm$, 2 FU). Titanium atoms are blue, oxygen red. Below, we show the PXRD pattern generated from the three phases ($\lambda = 1.54056$ Å, which corresponds to the wavelength of Cu Kα radiation) and the corresponding similarity index calculated by *VC-GPWDF* using the PXRD-pattern of brookite as reference. The patterns shown are for the smallest similarity index that can be obtained by varying the unit cell parameters.

additional information, such as PXRD data, as we illustrate below.

In the bottom panel of Fig. 2, the PXRD pattern of the geometry-optimized structures of the three aforementioned polymorphs of titanium dioxide are plotted. The similarity to brookite, where the reference PXRD data was generated with Mercury 2022.3.0 (ref. 47) in the range from 1° to 120° in 2θ with a step of 0.1°, from the experimental structure collected by Meagher and Lager at room temperature (ICSD = 36408; lattice parameters in Å: $a = 5.138$, $b = 9.174$, $c = 5.449$)[48] is also provided. Despite the fact that the similarity index (or dissimilarity) is low, 0.01, the DFT-optimized cell parameters are quite different from that of the reference: $a = 5.192$ Å, $b = 9.274$ Å, $c = 5.509$ Å. In fact, if we calculated the similarity index using the DFT-optimized structure directly, as proposed in other methodologies (see Introduction), a value of 0.17 would be obtained (to be compared to 0.43 for anatase and 0.93 for rutile). Because *VC-GPWDF* performs unit cell deformations, which include varying the DFT-optimized cell parameters during the comparison of the reference and trial structures, it becomes possible to retrieve a nearly zero similarity index for the correct structure. The cell parameters of the relaxed structure post-refinement with *VC-GPWDF* are varied to $a = 5.140$ Å, $b = 9.171$ Å and $c = 5.447$ Å, which almost exactly coincides with the reference structure.

Now that we have described how the *VC-GPWDF* method can modulate a DFT-optimized structure, so that it provides the best match with the reference, let us examine the optimal similarity it provides between brookite and the two higher symmetry polymorphs. The main difference between the diffractogram of orthorhombic brookite with the one computed for the two tetragonal phases is the number of peaks, which is much larger in the less symmetric case. Moreover, brookite presents three intense peaks at low angles, two of them very close to each other (25.3°, 25.7° and 30.8° in 2θ degree, respectively), while in the tetragonal phases, only one high intensity peak is found. One of the reasons why rutile has a larger similarity index to brookite than anatase, is that in anatase the intense peak is in a range of 2θ similar to where the first doublet of peaks in brookite is found, while in rutile it is at a higher angle (~27°).

Now, let's put the *VC-GPWDF* similarity index in action with XtalOpt to predict the metastable, low symmetry, phase of TiO_2, brookite. To begin, two *single-objective* (classic enthalpy based) CSP runs with 8 FU in the cell (24 atoms) were performed as reference tests generating a total of 1000 structures each (Table 1). In both searches, brookite was not found, while the more symmetric anatase and rutile were generated, in-line with previous studies where 1100+ structures were optimized to find the orthorhombic phase.[26] Coupling XtalOpt with the *VC-GPWDF* algorithm to perform the *multi-objective* search, it was possible to find brookite in shorter evolutionary searches (see Table 1). However, this success appeared to be sensitive to the fitness weight parameter connected to the PXRD data (eqn (1)). In this test, brookite was successfully found using $w \geq 0.6$ prompting us to analyze how the choice of the weight influences the fitness of the three polymorphs of TiO_2 (Section S1†). The fitness is related to the probability that a structure has for being

Table 1 Benchmark tests on TiO_2-brookite using *single-* or *multi-objective* crystal structure prediction runs with 8 formula units. The number of total structures-per-run (# structures), the weight assigned to the powder X-ray diffraction similarity objective (w), and the output result of the search (if brookite was found or not), are reported[a]

	Single-objective		Multi-objective		
Run	1	2	1	2	3
# Structures	1000	1000	500	500	500
w	0.0	0.0	0.3	0.6	0.9
Brookite	*No*	*No*	*No*	*Yes*	*Yes*

[a] If the *Pbca* brookite phase was fortuitously generated in the initial pool by RandSpg, the run was repeated.

chosen as a parent in the evolutionary search, but there are other factors, including the symmetry and the types of lattices in the initial pool, as well as the random parameters chosen during the course of the CSP search, which also influence a structure's discovery. Though the fitness of brookite was higher than that of anatase already using a weight of 0.1, this polymorph was not discovered in our short CSP search even when a weight of 0.3 was used. While it is probable that brookite could be found in fewer structures than in a regular search (~1100) with this weight, increasing the weight to 0.6 hastens its appearance by more than a factor of two.

In the past, when the multi-objective PXRD search was not available, it was suggested that constraining the breeding pool to structures whose (sub)lattice was consistent with a particular Bravais lattice or space group (potentially deduced from a diffractogram) could be employed for unveiling the structure of a synthesized compound.[49] However, as shown in the following two examples, the PXRD search proposed herein is preferred since it accounts for possible variations in the crystal lattice (see below, *Na in ramp-compression*), and constraining a CSP search with an indexed unit cell might even be counter-productive in some rare cases (see below, *The tricky case of vaterite*).

3.2 Na in ramp-compression experiments

High-quality, and perhaps already indexed PXRD data can surely increase the success of the *multi-objective* search strategy that we describe above. However, data collected at extreme conditions, such as in dynamic or ramp compression experiments that explore the chemistry of the interiors of planets or high-energy-density quantum matter,[50] often require substantial support from theory for their interpretation. In fact, the data obtained in these experiments is obscured by noisy background, mostly sourced by the hot plasma ablated by the sample target during the laser irradiation,[51] jeopardizing the indexing of weak reflection's peaks, which might be covered by the background. Therefore, the comparison with PXRD data simulated from theoretical structures is often necessary to identify a phase in shock experiments. Despite this synergistic approach, the structural determination of new phases measured at extreme conditions still remains a great scientific challenge.[52]

A noisy background is not the only challenge in the structural solution from shock and ramp compression experiments. In fact, the kinetics in dynamic compression experiments, together with the uniaxial orientation of the shock front can alter the expected (theoretical) P–T path followed in a phase diagram, greatly diverging from the ideal thermodynamic path at low temperature, and leading to unexpected phase transitions, or even decompositions.[53] Yet, these variables are nothing but additional coordinates of the phase diagram of a compound that must be explored to understand the behavior of matter at extreme conditions.[50] From the theoretical point of view, this means that the system must be simulated with techniques beyond the standard 0 K DFT approximation, including quasi-harmonic phonons to account for the thermal volume expansion,[54] or by performing expensive molecular dynamics simulations.[55] Moreover, the computational reproduction of an anisotropic (e.g., uniaxial) compression can be a challenging task even for simple unary systems.[56] Therefore, the possibility to perform volume-cell modulations on-the-fly, and emulate all the effects that contribute to the divergence from an ideal compression experiment, can become useful for a rapid, but meaningful, interpretation of the experimental data. Below, we illustrate the power of this approach on Na, which assumes the iconic $hP4$ insulating electride phase, observed for the first time in diamond anvil cells at ~200 GPa,[57] where the electronic structure could be rationalized using a multi-center bonding scheme.[58]

Studying Na to pressures above 200 GPa, conditions that are accessible mostly with dynamic compression techniques, is currently of great interest as the findings will address important questions for theory and for condensed matter physics.[59] Therefore, recent laser-driven ramp compression experiments where sodium was squeezed to a nearly 7-fold increase in density at a pressure of 500 GPa (and a temperature of ~1500–3000 K) were performed.[59] In situ XRD revealed a series of peaks obtained at the highest pressures that could be indexed as the $hP4$ phase, but the peaks observed between 242–292 GPa were not consistent with $hP4$ and were instead interpreted as either the $cI16$ structure (previously observed between 108–120 GPa (ref. 60)) or an $R\bar{3}m$ phase. A following theoretical study,[61] however, revealed that both $cI16$ and $R\bar{3}m$ were not dynamically stable at the experimentally attained P–T conditions. In 0 K CSP searches $hP4$ emerges as by far the most stable phase at the pressures attained in experiment, but a subset of seven systems was also found, and computed to be preferred at high temperatures within the quasi-harmonic approximation.[61] Unfortunately, none of their diffraction patterns and densities were fully consistent with the experiments.

To identify a Na structure that could have been created using ramp compression, we carried out a VC-GPWDF assisted multi-objective CSP search for Na at 315 GPa using the experimental data published in ref. 59, and a weight of 0.7. In this case, the reference list of reflection values is constituted by only six indexed peaks (six [2θ; I] pairs), also to avoid contamination from the high noise over the experimental 2θ range (Fig. 3a). This search found that the phase producing the best similarity index was actually $hP4$, the expected ground state at these
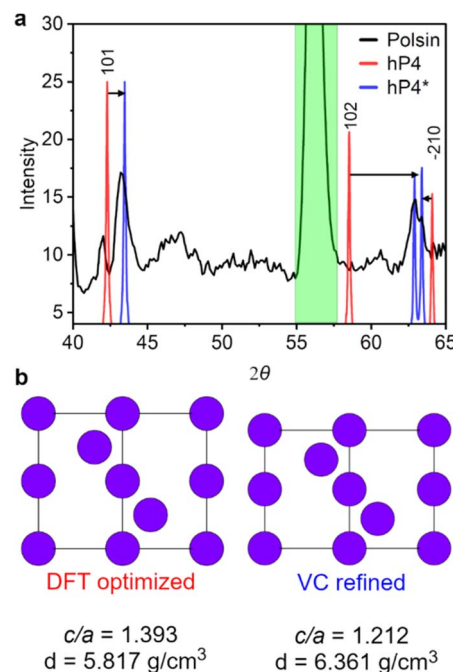

Fig. 3 Simulated XRD patterns and experimental lineouts of Na structures under pressure. (a) Section of the diffraction pattern measured by Polsin et al.[59] (black line) together with the diffraction lines calculated from the DFT-optimized (red line) and refined (blue line) structures of Na-$hP4$ ($\lambda = 1.481$ Å, see ref. 59). The similarity index changed from 0.991 ($hP4$) to 0.086 ($hP4$*) upon refinement. The calibration peak in the diffraction data is shaded in green. (b) View of the unit cell of Na-$hP4$ along the (110) plane, as optimized by DFT ($hP4$), and after the volume-cell refinement with VC-GPWDF ($hP4$*).

conditions. So why wasn't it previously identified by either experiment or by theory? The answer stems from the severe distortions the structure seems to undergo during the ramp-compression, which cannot be emulated by the standard 0 K DFT optimization or molecular dynamics with isotropic pressure, but that is easily revealed by the cell-variation routine in VC-GPWDF.

In Fig. 3 we illustrate the structure of the $hP4$ phase as it emerges from the DFT optimization and post-refinement with VC-GPWDF ($hP4$*), coupled with the diffraction peaks that these two phases yield overlaid on the XRD data collected by Polsin et al.[59] The main structural difference between $hP4$ and $hP4$* is the extra anisotropic compression along the c-axis (Fig. 3b) estimated to be equal to ~425 GPa (DFT stress value) and compared to the ~370 GPa obtained along the a- and b-axes, causing the lowered c/a ratio and the increased density of the crystal. The subsequent effect of this distortion on the calculated diffraction lines is, on one hand, to move the (101) to higher 2θ angles, and on the other hand to almost merge the reflections from the (102) and the (2–10) planes (Fig. 3a), matching with the doublet peaks experimentally observed at ~63°.

In Na-$hP4$* the density increases up to 6.3–6.4 g cm^{-3}, which is relatively high compared to what is expected from 0 K DFT calculations on $hP4$.[57] However, the pressure versus density curve of sodium can deviate quite substantially from the ideal

trend, and produce very different results depending on the experimental conditions.[59] For example, at 300 GPa, the density of sodium can be estimated as being ~3.5 g cm^{-3} following the Sesame principal Hugoniot, or being ~5.9–6.0 g cm^{-3} if extrapolated using static compression data from the FCC and BCC phases.[59] Moreover, hP4* is calculated to be 199 meV per atom higher in enthalpy than the fully relaxed hP4 structure (while retaining all real phonons, see Fig. S1†), which is nonetheless accessible based on the estimated temperature in Polsin's experiment (~2200 K).[59] The weak peak at ~42° that was previously indexed as the hP4 (101), was suggested to indicate that multiple phases coexisted, resulting from pressure and temperature gradients present in the sample, in-line with prior interpretations.[59,61]

Na-hP4* is a distorted structure that cannot be obtained with classic DFT geometry optimizations at high-pressure or molecular dynamics simulations, since in both cases, the system would evolve towards the most stable (relaxed) configuration. Instead, Na-hP4* mirrors the effects of anisotropic/uniaxial compression along the c-axis and the thermal expansion, which are extrapolated and accessed thanks to the iterative refinement over the experimental data. Though, it is not possible to unequivocally identify the phase observed by Polsin et al.[59] as hP4*, it is worth noting that the PXRD assisted CSP with XtalOpt-VC-GPWDF could access interesting new alternatives for the interpretation of challenging data collected at extreme conditions.[52,62]

3.3 The tricky case of vaterite

Among the biogenic minerals, calcium carbonate (CaCO$_3$) is arguably the most abundant. From the three known anhydrous crystalline polymorphs of CaCO$_3$, calcite, aragonite and vaterite, the latter is the least stable, but still commonly found in nature.[63] Surprisingly, despite the nearly 100 year debate on its crystal structure, an apparently satisfying solution was proposed only very recently.[64] Specifically, it was suggested that

vaterite should be regarded as a polytypic structure, a specific type of polymorphism built up by a stacking of almost identical layers, which differ in their stacking sequence. This has made vaterite a very challenging system to solve, even combining several experimental techniques,[64] and almost impossible with computational methods alone. In fact, though vaterite is a relatively simple mineral (composition-wise), it cannot be solved solely with standard CSP methods, even with the possibility of using supercells. Therefore, what can a structure search do to support the solution of such challenging systems? This is what we will try to understand with this last example using our new methodology.

The *multi-objective* evolutionary search coupled with *VC-GPWDF* is obviously limited by the type of PXRD data used. In this case, it was possible to retrieve two extensive lists of peaks [2θ; I] from the studies performed by Le Bail et al.[65] and by DuPont et al.,[66] and two CSP runs using a weight of 0.7 were carried out using one, or the other, as a reference. In Fig. 4, we plot the energy difference (relative to calcite) *vs.* similarity index of the phases output by the CSP runs, focusing on those identified as good matches (similarity index < 0.1).

Using the data indexed by Le Bail et al.[65] (Fig. 4a), it is not surprising to see that the proposed Ama2 phase was predicted by our PXRD assisted-CSP as the best match. However, using this list of peaks, our search also found the Pnma structure proposed by Meyer,[67] which is ~35 meV per atom more stable than the Ama2 phase proposed by Le Bail, as well as the $P2_12_12_1$ phase proposed by DeMichelis[68] (which is isoenergetic to Pnma). Using the second set of data, collected by DuPont et al.[66] (Fig. 4b), yields different results. The structures proposed previously by Le Bail[65] (Ama2), Meyer[67] (Pnma) and DeMichelis[68] ($P2_12_12_1$) were still found. Notice that the similarity index of the recurrent structures changes from one data set to another, but it is consistently very low (<0.1).

Moreover, the two assisted searches have also generated new structures having excellent similarity index and a low energy
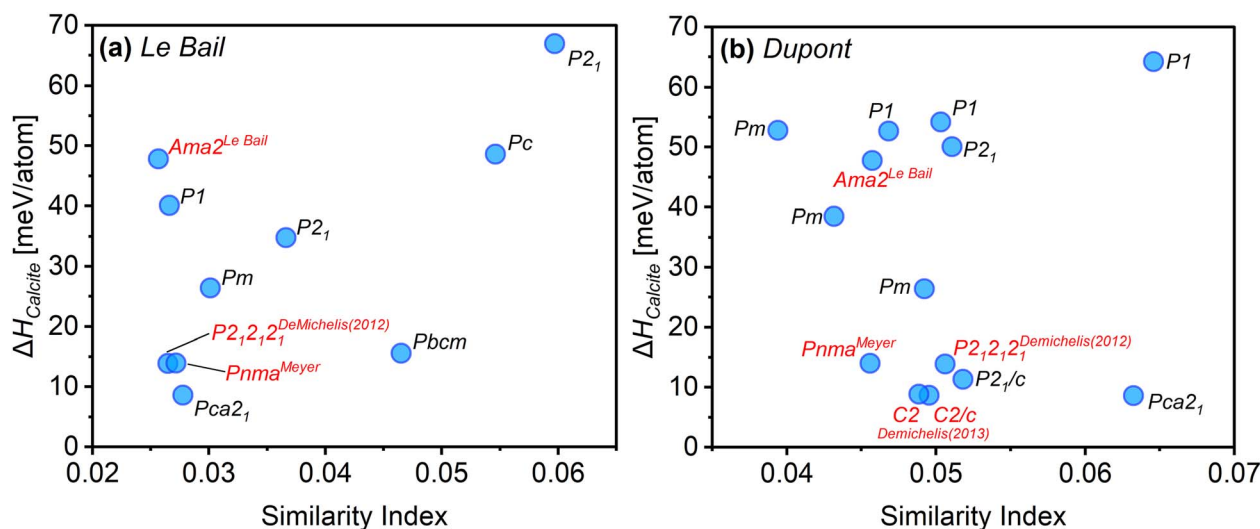

Fig. 4 Plot of the properties of the structures predicted with *XtalOpt-VC-GPWDF* for vaterite. The relative energies (using calcite as a reference) *versus* similarity index generated using (a) Le Bail's[65] and (b) DuPont's[66] experimental powder X-ray diffractograms.

(only ~10 meV per atom above calcite), and not proposed in past works, which are commented on in Section S4.† The most interesting result obtained with DuPont's data is probably the prediction of both the $C2$ and $C2/c$ structures, previously predicted in another work by DeMichelis *et al.*,[69] which differ by the specific orientation of the carbonate group along the stacking direction of the layers. These two structures are extremely important, since they form the sub-set of phases composing the polytypic crystal structure recently proposed.[64]

These results show how the coupled *XtalOpt-VC-GPWDF* algorithm can support the solution of complicated crystal structures such as vaterite, without the need of performing crystal structure searches on supercells as large as the real crystal, which would be computationally inaccessible. Our method was able to generate, almost on-the-fly, most of the crystal structures proposed for vaterite in past theoretical and experimental works, including those forming the polytypic structure, and ranking them by energy and similarity with the experimental PXRD. As we have postulated, even by generating the correct metastable crystal structure, it would have been impossible to thoroughly solve the case of vaterite. However, our new methodology was able to provide all the building blocks necessary to construct the polytypic model that solves the intricate crystal structure of vaterite, supporting crystallographers dealing with such challenging systems.

4 Conclusions

We have introduced a powder X-ray diffraction-assisted crystal structure prediction method that employs both the enthalpy of a structure and its similarity index, as compared to that of a reference X-ray diffraction pattern, in an equal footing. This technique has been implemented within the open-source evolutionary algorithm code, XtalOpt. The similarity index is calculated using *VC-GPWDF*, a modified version of de Gelder's similarity index, which assesses the overlap between diffraction patterns through a cross-correlation function upon iterative distortions of the unit cells. Either complete or partial PXRD patterns can be used to assist the crystal structure predictions, which can depend on the availability and quality of the experimental data. This similarity index is then used to determine the value of the fitness in XtalOpt's multi-objective global optimization process. Our method is shown to be optimal for identifying metastable phases, facilitating the identification of polymorphs in inorganic samples, and aiding in the analysis of structures distorted by the extreme conditions created in shock and ramp compression experiments. Moreover, it is also effective in identifying challenging structures such as polytypic systems. Thanks to the cell-variation procedure, it is likely that a target structure might also be found from geometries optimized at low levels of theory (*e.g.*, DFTB, machine-learned potentials, low energy cut-offs or sparse *k*-point meshes), even in the case where the energy order is not correct, since for the PXRD only the structural information is relevant. However, care would need to be taken so the computational parameters employed yield sufficiently accurate forces to ensure reasonable optimizations. We believe that the coupled *XtalOpt-VC-GPWDF*

tool will be highly beneficial for crystallographers, chemists, materials scientists and geochemists for the solution of challenging structures at ambient and extreme conditions.

5 Methods

5.1 Computational details

The open-source evolutionary algorithm XtalOpt[33,36] version 13.0 was employed for crystal structure prediction, using the multi-objective fitness measure. The parameters employed in XtalOpt are thoroughly described in ref. 22. The initial generation consisted of random symmetric structures that were created by the RandSpg algorithm,[43] except in the case of $CaCO_3$, where the initial generation was created externally with PyXtal[44] then imported as seeds, using Ca atoms and CO_3 trigonal planar units. PyXtal was employed for this purpose, because the generation of symmetric crystals containing molecular fragments for the initial pool of structures is not implemented in XtalOpt. The number of initial structures was equal to 50 in all cases. We believe that the first generation of structures could be improved using automated classifications,[70] by generating a more accurate initial pool of structures, focusing on the most probable space groups identified by the machine learning engine, a possibility that we will explore in future works. The number of formula units (FUs) was set equal to 8 in the case of TiO_2 to automatically cover the FU of all the natural polymorphs, *i.e.* anatase (4 FU), brookite (8 FU) and rutile (2 FU); 4, 6, 8, 12, 20, 24 and 32 in Na; and 4, 6, 8 and 12 in $CaCO_3$. A sum of the atomic radii scaled by a factor of 0.7 was used to determine the shortest distances allowed between pairs of atoms. Duplicate structures were identified and removed from the breeding pool using the XtalComp algorithm.[71] For the TiO_2-brookite test, the total number of generated structures could vary from 500 to 1000 (see Section TiO_2 – brookite). For the tests performed on high-pressure Na and $CaCO_3$, the total number of generated structures per run was equal to 1000. Each structure search followed a multi-step strategy, with three subsequent optimizations with increased level of accuracy, plus a final accurate single point (see below).

Geometry optimizations and electronic structure calculations were performed using Density Functional Theory (DFT) with the Vienna *Ab initio* Simulation Package (VASP), version 6.4.2.[38] The PBE[72] exchange–correlation functional was employed. The projector augmented wave (PAW) method[73] was used to treat the core states in combination with a plane-wave basis set with an energy cutoff of 500 eV. The O $2s^2 2p^4$ (PAW_PBE O_s), Ti $3d^3 4s^1$ (PAW_PBE Ti), Na $2p^6 3s^1$ (PAW_PBE Na_pv), Ca $3p^6 4s^2$ (PAW_PBE Ca_pv) and the C $2s^2 2p^2$ (PAW_PBE C_s) states were treated explicitly. The *k*-point meshes were generated using the Γ-centered Monkhorst–Pack scheme,[74] and the number of divisions along each reciprocal lattice vector was selected so that the product of this number with the real lattice constant was greater than or equal to a given value. The values of 20, 25 and 30 Å were used for the three subsequent optimization steps in the crystal structure search of TiO_2 and $CaCO_3$, then a value of 50 Å was used for the final single point. In the case of sodium, a value of 40 Å was used at

each optimization step, and one of 50 Å for the final single point. The accuracy of the energy convergence was set to increase from 10^{-3} to 10^{-5} eV for the optimizations, and to 10^{-6} for the final single point on the structures for which the norms of all the forces calculated during the relaxations were smaller than 10^{-3} eV Å^{-1}. A Gaussian smearing was used at each optimization step, and for each system with a sigma of 0.02 eV. The tetrahedron method was adopted in the last single point.[75]

CSP methods employ metaheuristic algorithms, which do not always guarantee that the optimal solution, or set of optimal solutions are found. Therefore, in addition to varying the parameters employed in a standard CSP search,[22] a user may need to perform searches where the weights in the objective function are varied (as in Section 3.1), or explore a range of potential solutions identified by plotting the enthalpies of the optimized structures *versus* their similarity index (as in Section 3.3) to find the most likely candidate structure.

Data availability

The data supporting the article has been included as part of the ESI.† The crystal structure prediction code *XtalOpt* is open source and it can be downloaded at this link https://xtalopt.github.io/. The code *VASP*, used for the density functional theory calculations, is commercial, and needs to be purchased. The code *critic2*, used to calculate the similarity index, is open source, and it can be downloaded at this link https://aoterodelaroza.github.io/critic2/. The *script* used to guide the crystal structure prediction performed by XtalOpt with the experimental data is provided in the ESI† of the manuscript, together with a step-by-step procedure for its use. The refined Na-*hP4** structure and new *Pca*2$_1$ and *P*2$_1$/*c* CaCO$_3$ phases are provided in a separate folder as *cif-files*.

Author contributions

The project was conceptualized by E. Z. and S. R.; data curation and formal analysis was performed by S. R.; the funding acquisition was carried out by E. Z. and A. O. R.; the methodology was developed and implemented by S. R., S. H. and A. O. R. The responsibility of the project administration was done by E. Z. and S. R.; while resources were provided by E. Z., who is also the supervisor of the project. S. R, S. H. and A. O. R. took care of the software development, while S. R. was responsible for the validation of the methodology. All authors contributed in writing and reviewing the draft. We also thank Dr Xiaoyu Wang for the valuable discussions.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

Notes and references

1 S. Datta and D. J. W. Grant, *Nat. Rev. Drug Discovery*, 2004, **3**, 42–57.

2 M. D. Allendorf, C. a. Bauer, R. K. Bhakta and R. J. T. Houk, *Chem. Soc. Rev.*, 2009, **38**, 1330–1352.

3 M. De Jong, W. Chen, T. Angsten, A. Jain, R. Notestine, A. Gamst, M. Sluiter, C. K. Ande, S. Van Der Zwaag, J. J. Plata, C. Toher, S. Curtarolo, G. Ceder, K. A. Persson and M. Asta, *Sci. Data*, 2015, **2**, 150009.

4 K. D. M. Harris, M. Tremayne and B. M. Kariuki, *Angew. Chem., Int. Ed.*, 2001, **40**, 1626–1651.

5 E. Fomina, E. Kozlov and S. Ivashevskaja, *Powder Diffr.*, 2019, **34**, S59–S70.

6 T. Duffy, N. Madhusudhan and K. K. M. Lee, in *Treatise on Geophysics*, Elsevier B.V., 2nd edn, 2015, vol. 2, pp. 149–178.

7 J. A. Flores-Livas, L. Boeri, A. Sanna, G. Profeta, R. Arita and M. Eremets, *Phys. Rep.*, 2020, **856**, 1–78.

8 A. Altomare, *IUCrJ*, 2022, **9**, 403–405.

9 R. Černý, *Crystals*, 2017, **7**, 142.

10 R. Černý and V. Favre-Nicolin, *Z. Kristallogr.*, 2007, **222**, 105–113.

11 W. I. F. David and K. Shankland, *Acta Crystallogr., Sect. A: Found. Crystallogr.*, 2008, **64**, 52–64.

12 H. M. Rietveld, *J. Appl. Crystallogr.*, 1969, **2**, 65–71.

13 A. Le Bail, *Powder Diffr.*, 2005, **20**, 316–326.

14 K. D. M. Harris and M. Tremayne, *Chem. Mater.*, 1996, **8**, 2554–2570.

15 B. M. Kariuki, H. Serrano-González, R. L. Johnston and K. D. M. Harris, *Chem. Phys. Lett.*, 1997, **280**, 189–195.

16 K. D. M. Harris, *Comput. Mater. Sci.*, 2009, **45**, 16–20.

17 J. Schuetzke, A. Benedix, R. Mikut and M. Reischl, *IUCrJ*, 2021, **8**, 408–420.

18 E. A. Riesel, T. Mackey, H. Nilforoshan, M. Xu, C. K. Badding, A. B. Altman, J. Leskovec and D. E. Freedman, *J. Am. Chem. Soc.*, 2024, **146**, 30340–30348.

19 M. Cordova and L. Emsley, *J. Am. Chem. Soc.*, 2023, **145**, 16109–16117.

20 M. Balodis, M. Cordova, A. Hofstetter, G. M. Day and L. Emsley, *J. Am. Chem. Soc.*, 2022, **144**, 7215–7223.

21 A. R. Oganov, C. J. Pickard, Q. Zhu and R. J. Needs, *Nat. Rev. Mater.*, 2019, **4**, 331–348.

22 Z. Falls, P. Avery, X. Wang, K. P. Hilleke and E. Zurek, *J. Phys. Chem. C*, 2021, **125**, 1601–1620.

23 M. A. Neumann, F. J. J. Leusen and J. Kendrick, *Angew. Chem., Int. Ed.*, 2008, **47**, 2427–2430.

24 A. G. Shtukenberg, Q. Zhu, D. J. Carter, L. Vogt, J. Hoja, E. Schneider, H. Song, B. Pokroy, I. Polishchuk, A. Tkatchenko, A. R. Oganov, A. L. Rohl, M. E. Tuckerman and B. Kahr, *Chem. Sci.*, 2017, **8**, 4926–4940.

25 E. J. Chan, A. G. Shtukenberg, M. E. Tuckerman and B. Kahr, *Cryst. Growth Des.*, 2021, **21**, 5544–5557.

26 B. Wang, K. P. Hilleke, S. Hajinazar, G. Frapper and E. Zurek, *J. Chem. Theory Comput.*, 2023, **19**, 7960–7971.

27 B. Meredig and C. Wolverton, *Nat. Mater.*, 2013, **12**, 123–127.

28 L. Ward, K. Michel and C. Wolverton, *Phys. Rev. Mater.*, 2017, **1**, 063802.

29 H. Ling, J. Montoya, L. Hung and M. Aykol, *Comput. Mater. Sci.*, 2022, **214**, 111687.

30 P. Gao, Q. Tong, J. Lv, Y. Wang and Y. Ma, *Comput. Phys. Commun.*, 2017, **213**, 40–45.

31 N. Tsujimoto, D. Adachi, R. Akashi, S. Todo and S. Tsuneyuki, *Phys. Rev. Mater.*, 2018, **2**, 053801.

32 J. Van De Streek and M. A. Neumann, *Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater.*, 2014, **70**, 1020–1032.

33 D. C. Lonie and E. Zurek, *Comput. Phys. Commun.*, 2011, **182**, 372–387.

34 A. Otero-De-La-Roza, *J. Appl. Crystallogr.*, 2024, **57**, 1401–1414.

35 A. Otero-De-La-Roza, E. R. Johnson and V. Luaña, *Comput. Phys. Commun.*, 2014, **185**, 1007–1018.

36 S. Hajinazar and E. Zurek, *Comput. Phys. Commun.*, 2024, **304**, 109306.

37 K. Deb, Multi-objective Optimisation Using Evolutionary Algorithms: An Introduction, *Multi-objective Evolutionary Optimisation for Product Design and Manufacturing*, ed. L. Wang, A. H. C. Ng and K. Deb, Springer, London, 2011, pp. 3–34.

38 G. Kresse and J. Hafner, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1994, **49**, 14251–14269.

39 R. De Gelder, R. Wehrens and J. A. Hageman, *J. Comput. Chem.*, 2001, **22**, 273–289.

40 R. A. Mayo, K. M. Marczenko and E. R. Johnson, *Chem. Sci.*, 2023, **14**, 4777–4785.

41 R. A. Mayo and E. R. Johnson, *CrystEngComm*, 2021, **23**, 7118–7131.

42 R. A. Mayo, A. Otero-De-la-Roza and E. R. Johnson, *CrystEngComm*, 2022, **24**, 8326–8338.

43 P. Avery and E. Zurek, *Comput. Phys. Commun.*, 2017, **213**, 208–216.

44 S. Fredericks, K. Parrish, D. Sayre and Q. Zhu, *Comput. Phys. Commun.*, 2021, **261**, 1–10.

45 S. Habermehl, P. Mörschel, P. Eisenbrandt, S. M. Hammer and M. U. Schmidt, *Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater.*, 2014, **70**, 347–359.

46 F. H. Stillinger, *Phys. Rev. E: Stat. Phys., Plasmas, Fluids, Relat. Interdisc. Top.*, 1999, **59**, 48–51.

47 C. F. MacRae, I. Sovago, S. J. Cottrell, P. T. A. Galek, P. McCabe, E. Pidcock, M. Platings, G. P. Shields, J. S. Stevens, M. Towler and P. A. Wood, *J. Appl. Crystallogr.*, 2020, **53**, 226–235.

48 E. P. Meagher and G. A. Lager, *Can. Mineral.*, 1979, **17**, 77–85.

49 K. P. Hilleke, X. Wang, D. Luo, N. Geng, B. Wang and E. Zurek, *Phys. Rev. B*, 2023, 014511.

50 C. S. Yoo, *Matter Radiat. Extremes*, 2020, **5**, 018202.

51 J. R. Rygg, J. H. Eggert, A. E. Lazicki, F. Coppari, J. A. Hawreliak, D. G. Hicks, R. F. Smith, C. M. Sorce, T. M. Uphaus, B. Yaakobi and G. W. Collins, *Rev. Sci. Instrum.*, 2012, **83**, 113904.

52 M. G. Gorman, S. Elatresh, A. Lazicki, M. M. E. Cormier, S. A. Bonev, D. McGonegle, R. Briggs, A. L. Coleman, S. D. Rothman, L. Peacock, J. V. Bernier, F. Coppari, D. G. Braun, J. R. Rygg, D. E. Fratanduono, R. Hoffmann, G. W. Collins, J. S. Wark, R. F. Smith, J. H. Eggert and M. I. McMahon, *Nat. Phys.*, 2022, **18**, 1307–1311.

53 A. Vailionis, E. G. Gamaly, V. Mizeikis, W. Yang, A. V. Rode and S. Juodkazis, *Nat. Commun.*, 2011, **2**, 445.

54 J. Wu, F. González-Cataldo and B. Militzer, *Phys. Rev. B*, 2021, **104**, 014103.

55 S. Hamel, L. X. Benedict, P. M. Celliers, M. A. Barrios, T. R. Boehly, G. W. Collins, T. Döppner, J. H. Eggert, D. R. Farley, D. G. Hicks, J. L. Kline, A. Lazicki, S. Lepape, A. J. MacKinnon, J. D. Moody, H. F. Robey, E. Schwegler and P. A. Sterne, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2012, **86**, 094113.

56 L. He, D. Polsin, S. Zhang, G. W. Collins and N. Abdolrahim, *Sci. Rep.*, 2022, **12**, 18954.

57 Y. Ma, M. Eremets, A. R. Oganov, Y. Xie, I. Trojan, S. Medvedev, A. O. Lyakhov, M. Valle and V. Prakapenka, *Nature*, 2009, **458**, 182–185.

58 S. Racioppi, C. V. Storm, M. I. McMahon and E. Zurek, *Angew. Chem., Int. Ed.*, 2023, **27**, e202310802.

59 D. N. Polsin, A. Lazicki, X. Gong, S. J. Burns, F. Coppari, L. E. Hansen, B. J. Henderson, M. F. Huff, M. I. McMahon, M. Millot, R. Paul, R. F. Smith, J. H. Eggert, G. W. Collins and J. R. Rygg, *Nat. Commun.*, 2022, **13**, 2534.

60 M. I. McMahon, E. Gregoryanz, L. F. Lundegaard, I. Loa, C. Guillaume, R. J. Nelmes, A. K. Kleppe, M. Amboage, H. Wilhelm and A. P. Jephcoat, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 17297–17299.

61 B. Wang, K. P. Hilleke, X. Wang, D. N. Polsin and E. Zurek, *Phys. Rev. B*, 2023, **107**, 184101.

62 X. Gong, D. N. Polsin, R. Paul, B. J. Henderson, J. H. Eggert, F. Coppari, R. F. Smith, J. R. Rygg and G. W. Collins, *Phys. Rev. Lett.*, 2023, **130**, 76101.

63 J. D. C. McConnell, *Mineral. Mag. J. Mineral. Soc.*, 1960, **32**, 535–544.

64 X. San, J. Hu, M. Chen, H. Niu, P. J. M. Smeets, C. D. Malliakas, J. Deng, K. Koo, R. dos Reis, V. P. Dravid and X. Hu, *Nat. Commun.*, 2023, **14**, 7858.

65 A. Le Bail, S. Ouhenia and D. Chateigner, *Powder Diffr.*, 2011, **26**, 16–21.

66 L. Dupont, F. Portemer and M. Figlarz, *J. Mater. Chem.*, 1997, **7**, 797–800.

67 H. J. Meyer, *Angew. Chem., Int. Ed.*, 1959, **71**, 678.

68 R. Demichelis, P. Raiteri, J. D. Gale and R. Dovesi, *CrystEngComm*, 2012, **14**, 44–47.

69 R. Demichelis, P. Raiteri, J. D. Gale and R. Dovesi, *Cryst. Growth Des.*, 2013, **13**, 2247–2251.

70 J. E. Salgado, S. Lerman, Z. Du, C. Xu and N. Abdolrahim, *npj Comput. Mater.*, 2023, **9**, 214.

71 D. C. Lonie and E. Zurek, *Comput. Phys. Commun.*, 2012, **183**, 690–697.

72 J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, 1996, **77**, 3865–3868.

73 P. E. Blöchl, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1994, **50**, 17953–17979.

74 H. J. Monkhorst and J. D. Pack, *Phys. Rev. B: Solid State*, 1976, **13**, 5188–5192.

75 P. E. Blöchl, O. Jepsen and O. K. Andersen, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1994, **49**, 16223–16233.