

Machine learning based approach for solving atomic structures of nanomaterials combining pair distribution functions with density functional theory

Magnus Kløve^[a], Sanna Sommer^[a], Bo B. Iversen^{[a]}, Bjørk Hammer^{[b]ψ} and Wilke Dononelli^{[c]#}*

AUTHOR ADDRESS

[a] Department of Chemistry and iNano, Aarhus University, 8000 Aarhus, Denmark

[b] Center for Interstellar Catalysis, Department of Physics and Astronomy, Aarhus University, Ny Munkegade 120, Aarhus C 8000, Denmark

[c] MAPEX Center for Materials and Processes, Bremen Center for Computational Materials Science and Hybrid Materials Interfaces Group, Bremen University, 28359 Bremen, Germany

AUTHOR INFORMATION

Corresponding Author

*Bo B. Iversen: bo@chem.au.dk

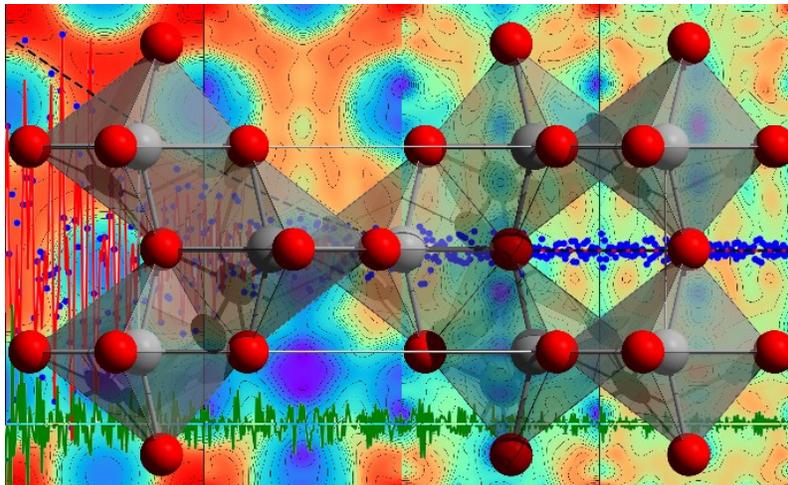
ψ Bjørk Hammer: hammer@phys.au.dk

#Wilke Dononelli: wido@uni-bremen.de

ABSTRACT

Determination of crystal structures of nanocrystalline or amorphous compounds is a great challenge in solid states chemistry and physics. Pair distribution function (PDF) analysis of X-Ray or neutron total scattering data has proven to be a key element in tackling this challenge. However, in most cases a reliable structural motif is needed as starting configuration for structure refinements. Here, we present an algorithm that is able to determine the crystal structure of an unknown compound by means of an on-the-fly trained machine learning model that combines density functional theory (DFT) calculations with comparison of calculated and measured PDFs for global optimization in an artificial landscape. Due to the nature of this landscape, even metastable configurations can be determined.

TOC GRAPHICS



KEYWORDS

Machine learning, Global optimization, Pair-distribution function, DFT, Crystal Structure Prediction, PDF

1. Introduction

Characterization of functional materials at the atomic level plays a crucial role in the advancement of modern society, as the properties of functional materials are intimately related to their atomic structure^[1-3].

In this regard, X-ray and neutron scattering techniques are indispensable tools, since they are direct probes of the atomic arrangement in solids, and the atomic structure of unknown compounds can be routinely elucidated from 3D single-crystal diffraction (SCD) data^[4]. For a powdered sample, however, the 3D structural information condenses onto a 1D diffractogram causing significant peak overlap and loss of directional information. Still, phase retrieval procedures such as the Patterson method and Direct Methods have been applied for solving atomic structures from powder X-ray diffraction (PXRD) experiments analogous to structure solution from SCD data^[5-8]. Global optimization techniques such as genetic algorithms^[9,10], Monte-Carlo optimization^[11-14] and simulated annealing^[11,15] attempt direct-space atomic structure solution from agreement with experimental PXRD data utilizing raw computational power. Hybrid methods such as the works of Wolverton and coworkers on FPASS build upon the same idea of global optimization using a genetic algorithm, however, with the use of density functional theory (DFT) energies to identify low energy structural motifs during the search^[16,17]. In all cases, however, structural solution from PXRD typically requires a homogenous and well-crystalline sample together with high quality data.

Real, functional materials are not necessarily well-crystalline, but are likely to be nanostructured, and so the traditional crystallographic approaches to structure solution fall short. For these materials, the concept of an infinitely repeating unit cell breaks down. This results in peak broadening and thus even

more severe peak overlap in the 1D diffractograms, yet again reducing the amount of unique information available^[18,19].

The pair distribution function (PDF) analysis of X-ray or neutron total scattering data has proven a valuable tool for studying nanomaterials^[20–23]. Still, most PDF analyses are guided by *a priori* structural information and are carried out as real-space counterparts to Rietveld refinements. Recent efforts by the Billinge group on the cloud platform, *PDFitc.org*, have provided the PDF community with real-space matching software for identifying unknown phases in PDF patterns through a structural database search^[24,25]. While such software can aid real-space analyses if the structure has been previously determined, determining the atomic structures of truly unknown nanocrystalline phases from PDF data remains yet a pending task for the community in the absence of efficient and versatile procedures.

Large-box global optimization like Reverse Monte-Carlo (RMC) modelling of PDFs has long been applied to study the atomic structure of amorphous materials, where no unique structural solution exists, but where the statistical average of the ensemble can provide structural insight^[26,27]. However, the RMC algorithm is inherently inefficient due to its unguided, random-guessing approach, often leading to search trapping in local minima.

The “Liga” method originally showcased to solve the structure of single-element, non-periodic clusters like C_{60} from experimental PDFs^[28] has also been extended to solve atomic structures of multiple-element crystalline samples from experimental PDFs^[29]. While successful in most cases, the algorithm has only been tested on well-crystalline powder samples, where traditional reciprocal space analysis could be more intuitively applied.

In this letter, we present an efficient machine-learning enhanced global optimization algorithm for solving atomic structures from experimental PDFs of materials on several length scales including nanomaterials, obtained with only short temporal resolution. By combining the global optimization method GOFEE by Bisbo and Hammer^[30–32] with the DiffPy-CMI software^[33] for PDF modeling, the algorithm performs a global optimization in an on-the-fly, machine learning trained surrogate R-factor landscape, thus solving the atomic structure from the PDF. Combining DFT calculations with PDF modelling the machine learning model can be trained in a combined artificial R-factor/DFT landscape to automatically determine the structure of metastable phases correctly.

2. Combining R-factor landscape and DFT Energy for global optimization.

In the original work of Bisbo and Hammer, the GOFEE algorithm was used to optimize structures to their global minimum in the potential energy surface (PES) calculated using density functional theory (DFT). For a detailed description of the GOFEE method, readers are referred to the original work by Bisbo and Hammer^[30–32]. In this work, we want to present an alternative approach to a pure DFT search, where the global optimization is additionally guided based on agreement between a calculated and an experimentally measured PDF through the R-factor in Eq. (1). A machine learning potential is trained on-the-fly on this comparison function and performs local optimizations of new candidate structures in this artificial (R-factor) landscape in combination with the DFT PES (see Eq. (2)), eventually optimizing the structures to their global minimum after several iterations of the algorithm.

$$R_w = \sqrt{\frac{\sum_i w_i (G_{i,obs} - G_{i,calc})^2}{\sum_i w_i G_{i,obs}^2}} \quad (1)$$

In Eq. (1), the R-factor is defined as the deviation between the calculated PDF, $G_{i,calc}$, and the observed PDF, $G_{i,obs}$, at each data point i with the weights, w_i , defined as $1/\sigma_i^2$, with σ_i being the uncertainty of $G_{i,obs}$. In the underlying work, these R-factors are determined using the DiffPy CMI software^[33].

The fitness F of new candidate structures within the GOFEE framework is calculated by combining the R-factor with prefactor α , with the total energy of the system E_{DFT} calculated at DFT level of theory with prefactor β as shown in Eq. (2).

$$F = -\alpha \cdot (1 - R_w) + \beta \cdot \frac{E_{DFT}}{1 \text{ eV}} \quad (2)$$

Eq. (2) allows different mixing factors for R_w and the total DFT energy to be used as the fitness function in the global optimization run. The two extrema correspond to setting either $\alpha = 0$ or $\beta = 0$. In the first case, the global optimization is only based on the DFT energy, in the latter the search is only guided based on agreement with the PDF through R_w . Note, a more straight forward definition of F could be given with prefactor λ instead of α as shown in Eq. (3).

$$F = \lambda \cdot R_w + \beta \cdot \frac{E_{DFT}}{1 \text{ eV}} \quad (3)$$

For preparation of the underlying manuscript we used Eq. (2) only for technical reasons, since all structures showing a positive value of F can automatically be sorted out by the algorithm.

3. Results and Discussion

3.1 Finding the crystal structure of 6 nm Anatase nanoparticles

To demonstrate the ability of the algorithm to find the crystal structure of an unknown compound and to analyze the influence of the prefactors α and β , we globally optimize the atomic structure of 6 nm Anatase (TiO_2) nanoparticles from a PDF obtained with only 3 s of X-ray exposure. Each run is initialized by creating four different random structures by placing 4 Ti and 8 O atoms at random non-overlapping positions in the simulation cell. Each of the four initial candidates is then locally optimized for 10 BFGS steps. After the initialization and local optimization, 1500 GOFEE cycles are used to guarantee convergence of the algorithm. Details on the algorithm can be found in the methods section and for details on GOFEE in general in the original work by Bisbo and Hammer^[30–32].

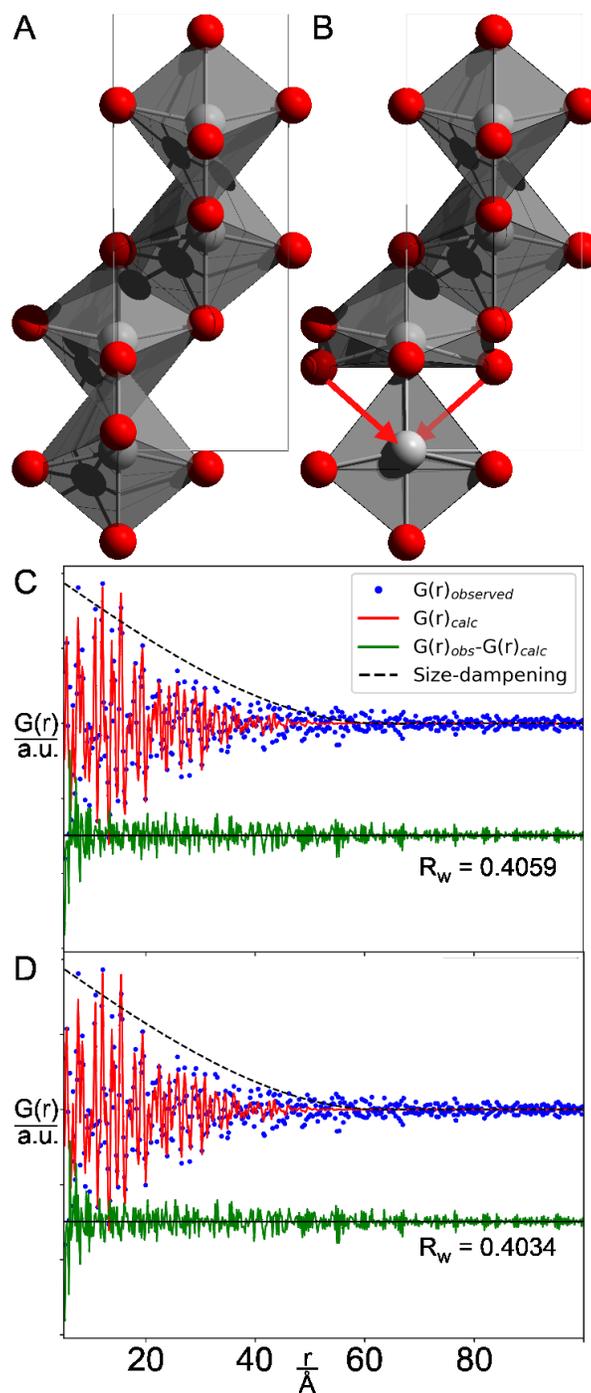


Figure 1: Result of global optimization based on R-factor compared to reference structure as found by DFT. A: reference structure; B: resulting structure based on a global optimization only based on R-factor ($\beta = 0$). Both structures have been postprocessed by local optimization based on the R-factor. The red arrows in B indicate the correct position of the displaced O atom with respect to the non-defected crystalline structure as shown in A; C: PDF pattern of perfectly crystalline structure in A; D: PDF pattern of structure B. Grey atoms are Ti and red atoms are O.

Using the R-factor alone ($\beta = 0$), the global optimization yields a best-fit structure where one oxygen atom is displaced with respect to the perfectly ordered crystalline structure (cf. Figure 1 A and B). As shown in Figure 1 C and D, the difference in the corresponding R_w values is only very small (0.025). Nevertheless, such a small difference is enough to force the algorithm to focus on structures that show such a structural defect.

To analyse this effect in a statistical fashion, ten different instances of the same global optimization were started for four different combinations of α and β , respectively. The results are shown in Figure 2 B. GOFEE was able to determine the correct atomic arrangement, i.e. crystal structure as shown in Fig. 2 A, of non-defected TiO_2 from the PDF ($\beta = 0$) with a success rate of 60% (c.f. Figure 2 B red line). The criterion for success is defined as finding a structure in the pool of candidates that measured by comparison of the sum of all bond lengths agrees with a corresponding non-defected crystal structure as shown in Fig. 2A. We found a similar result for CeO_2 as shown in the Supporting Information, although due to higher symmetry of the CeO_2 crystal structure, the success rate was found to be 80%. In contrast to global optimization of the Anatase structure based on the R-factor ($\beta = 0$), a global optimization based entirely on the DFT energy ($\alpha = 0$) gives a 100% success rate already after a few (54) iterations, however, without considering the agreement with experimental data (the R_w value) during the search.

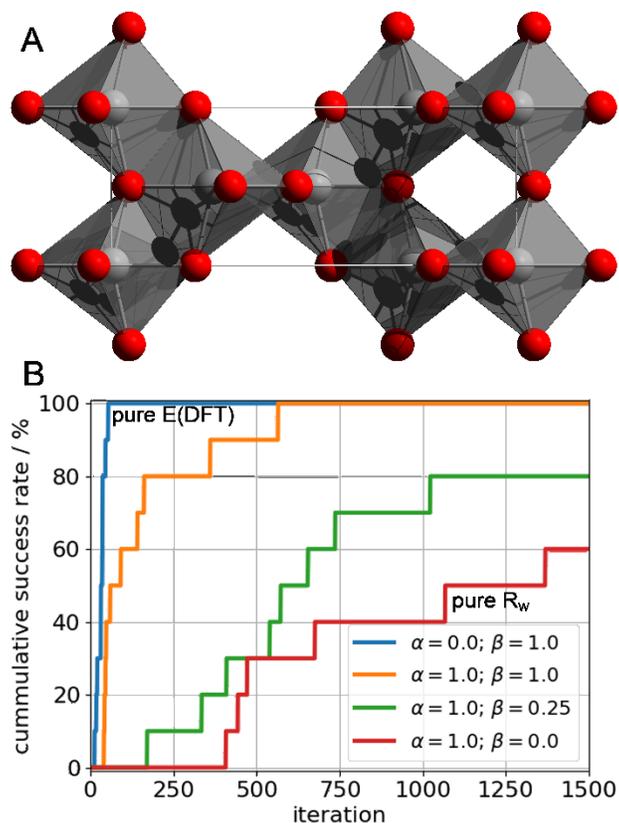


Figure 2: Result of global optimization based on R-factor, DFT energy and combination of both. A: Non-defected reference structure defining a successful run; B: cumulative success rate vs. number of GOFEE iterations for finding a structure similar to structure A based on different combinations of α and β . Grey atoms are titanium and red atoms are oxygen.

Now, the question arises why the global optimization based on the DFT energy is much more efficient than the same search based on the R-factor. The answer to this question is illustrated in Figure 3, where the landscapes for placing the shaded Ti atom at every position in the (100)-plane is shown for four combinations of DFT/PDF weighting. The pure DFT PES shows five maxima (blue) and just three distinct favourable minima (red). In comparison, the corresponding pure R-factor landscape shows multiple different minima. Exploration of all these minima takes much more effort than exploring the few local minima present in the DFT PES. Combining the DFT energy with the R-

factor yields intermediate landscapes with fewer or at least more pronounced local minima (Figure 3) and consequently leads to success rates in between the ones of a pure DFT or R-factor based search (Figure 2B). As a side effect, even for weighting the R-factor stronger than the DFT energy (Figure 2B, green line), the somewhat unphysical features of the pure R-factor based search can be discarded and a structure similar to the perfect, non-defected anatase crystal structure results.

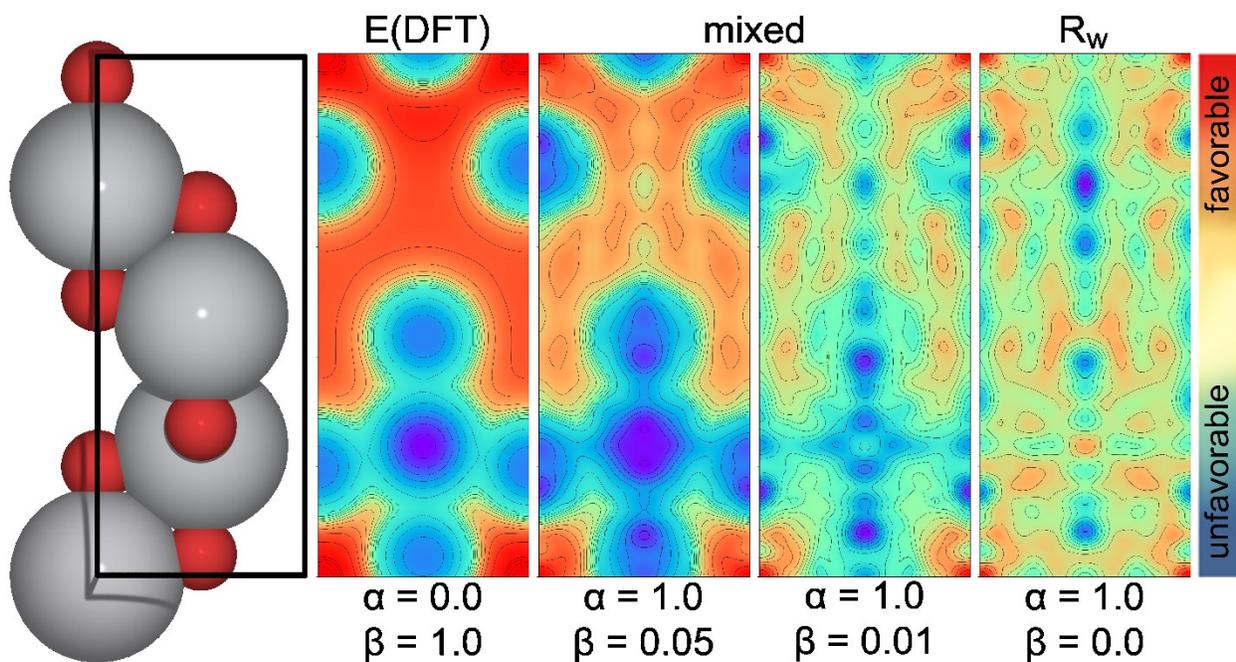


Figure 3: Normalized E(DFT) potential energy landscape, R-factor landscape and combined E(DFT) and R-factor landscapes for Anatase. The transparent Ti atom is placed at all positions in the unit cell for its given correct height. Corresponding DFT Energies and R_w values are calculated for each point. α and β values differ from values used in Fig. 1 for illustrational puposes. Grey atoms are titanium and red atoms are oxygen.

3.2 Studying the defected crystal structure of Gahnite ZnAl_2O_4

The R-factor landscape of solid matter can become quite complex, as demonstrated for Anatase resulting in only 60% success rate in a pure R-factor based search for this rather simple crystal structure. In the case of Gahnite (ZnAl_2O_4), which crystallizes in the more complex spinel structure that features a large unit cell with both octahedrally and tetrahedrally coordinated metal ions, the R-factor landscape seems to be too complex for the GOFEE algorithm to determine a physical meaningful average structure only based on the R-factor even after many iterations of the global optimization algorithm (see Figure S2). In the R-factor optimized structure, some local structural motifs of the metal ions that resemble the octahedral and tetrahedral coordination found in the spinel crystal structure are present. Yet, the overall structure is far from being a realistic representation of the crystal structure. Using a pure DFT PES ($\alpha = 0$), the GOFEE algorithm yields the perfect, non-defected crystal structure as known from databases. Depending e.g. on the functional used for energy evaluations (e.g. a pure GGA like PBE or a hybrid functional like PBE0), some distances and angles between different atoms might differ slightly, however, the overall motif still matches the spinel structure.

Regardless, both the R-factor and DFT optimized structures provide fits to the experimental PDFs of similar quality, and this example therefore clearly shows the limitations of using GOFEE or any other global optimization method weighted only by the R-factor, whenever the structure is more complex than a simple, highly symmetric binary compound.

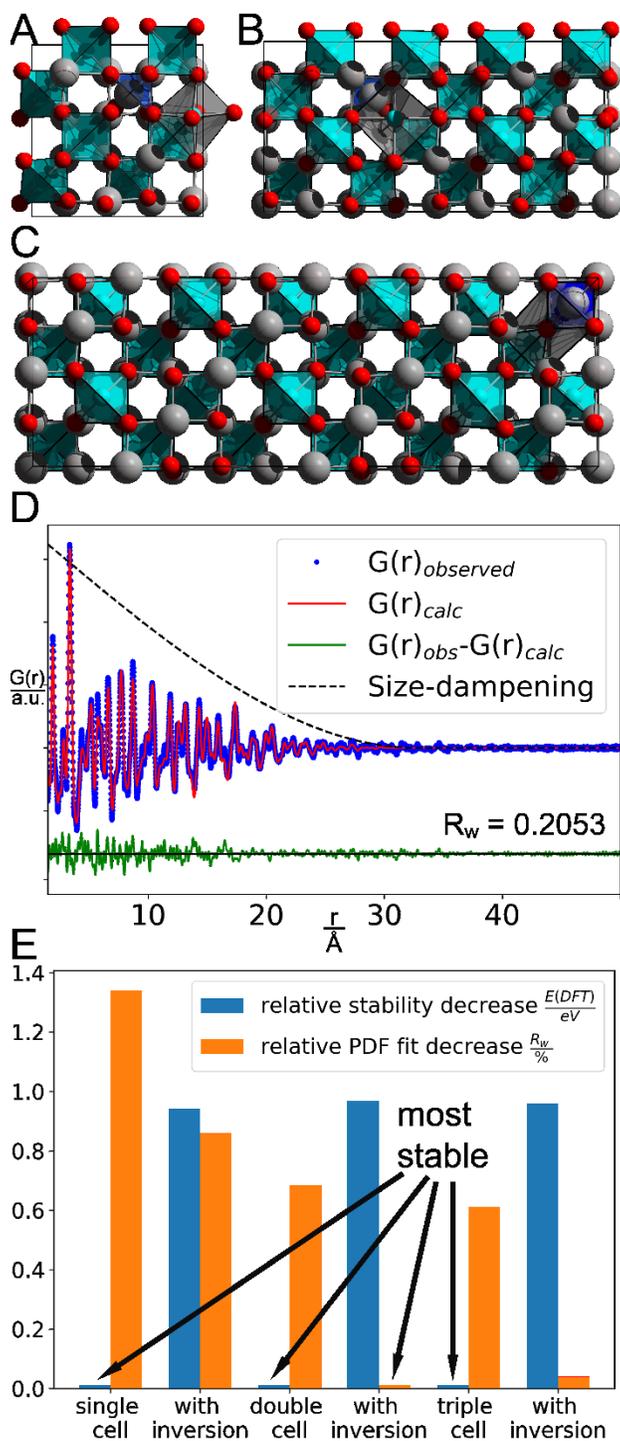


Figure 4: Modeled PDF and structure of Gahnite global optimized based on the R-factor and DFT energy; Inversion in a single unit cell; B: Inversion in a double unit cell; C: Inversion in a triple unit cell; All three models were predicted by GOFEE in the combined DFT PES and R-factor landscape; D: computed PDF of structure A with corresponding R_w value ; E: stability/likelihood comparison of Gahnite unit cells with and without inversion. Grey atoms are aluminum, blue atoms are zinc and red atoms are oxygen.

If on the other hand, the R-factor and DFT energy are combined in the global search, a different structure was found showing an inversion of one Al atom with one Zn atom (Figure 4A). This is a well-known point defect in spinel structures^[34,35] and therefore not too surprising. Nevertheless, the locally optimized (based on the DFT energy) non-defected structure is 0.93 eV more stable than the locally optimized defected structure. On the other hand, the calculated PDF of the defected structure (locally optimized using numerical R-factor forces) shows 0.63 % better agreement to the measured PDF than the calculated PDF of the locally optimized non-defected structure (c.f. Fig. 4 E). Therefore, although the non-defected structure might represent the thermodynamically most stable configuration best, the defected structure found with the combination of DFT energy evaluation and R-factor calculation, leads to the best average representation of the studied Gahnite nanoparticles with respect to the measured PDF pattern.

In the above, we worked with a 1x1x1 cell and hence a high concentration of defects whenever these are present. In an attempt to study the occurrence of defects at lower concentrations, a global optimization of a 2x1x1 Gahnite unit cell was run and similar results were obtained. The search based on the DFT energy leads to a non-defected crystal structure, whereas the search in the combined DFT and R-factor landscape leads to a structure with an inversion of Zn and Al. Comparable to the single unit cell, the non-defected structure is 0.95 eV more stable. On the other hand, again, locally optimizing the resulting structures in the R-factor landscape leads to 0.86 % better fit of the PDF of the defected structure. This is slightly more pronounced than in the case of a single unit cell. Increasing the cell size to a 3x1x1 super cell, the non-defected structure is 0.96 eV more stable than the one showing one inversion comparable to a single unit cell and a 2x1x1 super cell. On the other hand the defected super cell yields now only a 0.72% better fit than the non-defected structure. Therefore, one can estimate that the concentration of point defects is likely lower than one Al/Zn inversion per unit

cell, but bigger than 0.33 inversions per unit cell. Note, however, that in the case of the $3 \times 1 \times 1$ super cell the dimension of the periodically repeated cell almost reaches the size of the nanoparticle itself. Therefore, some additional effects might influence the decrease in the stability difference. Finally it should be noted that we have considered a perfect stoichiometry in the search, whereas non-stoichiometric amounts of the three elements were used in the refinement in the original study by Sommer et al.^[35] from which the data originate. Using non-stoichiometric amounts of any of the three elements in our GOFEE search could potentially result in other structural motifs showing up such as the different point defects found in their work.

4. Conclusion and outlook

We have introduced a global optimization method, which utilizes methods from the field of machine learning to find the crystal structure of unknown compounds only based on the comparison of a modeled to a measured PDF (based on the agreement factor, R_w). This method alone works well for simple crystal structures, but for more complex materials, one needs to combine the R-factor with an energy evaluation of the system thus introducing chemical intuition in the algorithm, e.g. by means of DFT in order to sort out physical non-meaningful structures. In addition, the combined DFT+R-factor corrected landscape may lead to metastable configurations as global minima, whereas a search based on DFT would always try to predict the thermodynamically most stable configuration as the global minimum. Using this method, crystal structures of unknown compounds can be solved fully automatically based on a combination of PDF measurements and DFT calculations within only a few hours to days of computational time. Although only tested with combination of the DFT energy with the R-factor, the method is not limited to combine these two values and create an artificial landscape

for the global optimization. As alternatives classical potentials or force fields could potentially be used in order to force the algorithms to generate physically meaningful structures. We additionally assume that instead of using the R_w factor from a PDF agreement evaluation, other experimental comparisons could be used. One example could be to utilize the R-factor from a comparison of a measured and a calculated PXRD in combination with DFT, or even in a combined DFT+R(PDF)+R(PXRD) landscape evaluation.

5. Methods

5.1 Experimental methods. The TS data on TiO_2 and ZnAl_2O_4 was collected at the P02.1 beamline at PETRA-III (DESY, Hamburg, Germany) with a photon energy of ~ 60 keV ($\lambda = 0.2072 \text{ \AA}$)^[36] on a 2D PerkinElmer XRD 1621 area detector for data acquisition. The TS data on CeO_2 shown in Supporting Information was collected at the P07 beamline at PETRA-III (DESY, Hamburg, Germany) with a photon energy of ~ 100 keV ($\lambda = 0.124 \text{ \AA}$)^[37] on a DECTRIS PILATUS3 X 2M CdTe detector. The 2D detector images were azimuthally integrated with the Dioptas^[38] software and calibrated using either a LaB_6 660b NIST line standard for TiO_2 and ZnAl_2O_4 or the CeO_2 data itself. The data was appropriately background subtracted and subsequently Fourier transformed using PDFgetX3^[39] in the xPDFsuite^[40] software, thus giving the PDF.

The TS data of Anatase (TiO_2) nanoparticles was obtained from an *in situ* synthesis by Søndergaard-Pedersen et al.^[41] An industrial grade TiOSO_4 precursor was diluted with sulfuric acid to a titanium concentration of 1.29(1) M. The solution was injected into a fused silica capillary mounted in a revised version of the *in situ* setup described in detail elsewhere^[42], pressurized to 250 bar and heated to 200 °C. Data collection was carried out with a temporal resolution of 3 s. The data was background

subtracted with a measurement of diluted sulfuric acid and Fourier transformed with a $Q_{\max} = 16.8 \text{ \AA}^{-1}$. Instrumental dampening was accounted for by separately obtaining and refining TS data of a LaB₆ 660b NIST line standard in the PDFgui^[43] software.

The TS data of Gahnite (ZnAl₂O₄) nanoparticles originates from a study by Sanna Sommer et al.^[35], and readers are referred to their paper on the preparation of the nanoparticles. The data was background subtracted with a measurement of an empty capillary and Fourier transformed with a $Q_{\max} = 23.5 \text{ \AA}^{-1}$. Instrumental dampening was accounted for by separately obtaining and refining TS data of a LaB₆ 660b NIST line standard in the PDFgui^[43] software.

The TS data of CeO₂ was background subtracted with a measurement of an empty capillary and Fourier transformed with a $Q_{\max} = 25.0 \text{ \AA}^{-1}$. Instrumental dampening was account for by performing an initial refinement of the data in the TOPAS^[44] v6 software.

5.2 Computational methods.

Local structure optimization

Local geometry optimizations can be performed straightforwardly for DFT calculations using analytical forces as implemented in GPAW^[45]. The exchange-correlation interaction was treated by the generalized gradient approximation using the PBE functional.

In addition, local geometry optimizations can be performed in the artificial R-factor landscape. Instead of using an analytical expression, forces are calculated numerically by displacing the atom positions $\pm 0.001 \text{ \AA}$ into each Cartesian direction and calculating the corresponding R-factor (from Eq. (1)). Then, the BFGS algorithm as implemented in ASE^[46] is used to optimize the structure into its local minimum.

In this work, the global optimization method GOFEE is used. In contrast to the original work by Bisbo and Hammer^[30–32], the machine learning potential is not always exclusively trained on DFT energies, but additionally on the R-Factor (as defined in Eq. (1)) or a combination of the R-factor and DFT energy according to Eq. (2).

The workflow of the code can be divided into five steps. First (I), a user defined number of initial random atomic structures is created, e.g. 10 structures. All structures share a common set of lattice parameters and stoichiometry. Each of these structures are geometrically locally relaxed either using DFT forces or numerical R-factor forces. Based on this data, the initial machine learning potential can be trained. In the second step (II), m new structures are created via random guess or mutation of the previous population ($m =$ number of CPU cores). In the current version of the code, the mutations are either permutation of two non-identical atoms or rattling some of the atoms to new positions. In the third step (III), all new structures' atomic positions are relaxed in the model machine learning potential. Next (IV), the fitness of each of these new structures is evaluated in the machine learning potential. Then a structure, where the machine learning potential either predicts a good fitness (exploit) or has a high uncertainty (explore), is chosen. For this structure, the correct fitness is evaluated. The fitness can either be the DFT potential energy, the R-factor as introduced in Eq. (1) or a combination of both as introduced in Eq. (2). In the last step (V), the chosen new structure and its corresponding fitness is added to training data and the model potential is retrained. The algorithm starts again from step (II) until n iterations has been reached (e.g. $n = 2000$), resulting in hundreds of different local minimum structures. The resulting structures including the global minimum structure were only relaxed in the ML potential during the search. Therefore, the resulting global minimum structure has to be post-processed and has to be locally optimized as described above. If the energetic

stability should be compared, the final structure is relaxed using analytical DFT forces. If the fit to the experiment is important, then numerical R-factor forces are used for the final local relaxation.

Supporting Information

The Supporting Information is available free of charge at:

Experimental section, algorithm description, additional figures (PDF)

Notes

The authors declare no competing financial interests.

Acknowledgement

Funding

We acknowledge financial support through the APF project ‘Materials on Demand’ within the ‘Humans on Mars’ Initiative funded by the Federal State of Bremen and the University of Bremen and support from the the Villum Fonden). In addition, this work was supported by the Danish National Research Foundation through the Center of Excellence “InterCat” (Grant Agreement No. DNRF150)

References

- [1] M. Christensen, N. Lock, J. Overgaard, B. B. Iversen, *J. Am. Chem. Soc.* **2006**, *128*, 15657.
- [2] Z. W. Seh, J. Kibsgaard, C. F. Dickens, I. Chorkendorff, J. K. Nørskov, T. F. Jaramillo, *Science* **2017**, *355*, eaad4998.
- [3] G. J. Snyder, M. Christensen, E. Nishibori, T. Caillat, B. B. Iversen, *Nature Mater* **2004**, *3*, 458.
- [4] C. Giacovazzo, Ed., *Fundamentals of Crystallography*, Oxford University Press, Oxford ; New York, **2011**.
- [5] P. Bénard, M. Louër, D. Louër, *J. Solid State Chem.* **1991**, *94*, 27.
- [6] P. Lightfoot, C. Glidewell, P. G. Bruce, *J. Mater. Chem.* **1992**, *2*, 361.
- [7] A. Le Bail, *Journal of Solid State Chemistry* **1993**, *103*, 287.
- [8] A. Clearfield, L. B. McCusker, P. R. Rudolf, *Inorg. Chem.* **1984**, *23*, 4679.

- [9] G. W. Turner, E. Tedesco, K. D. M. Harris, R. L. Johnston, B. M. Kariuki, *Chemical Physics Letters* **2000**, *321*, 183.
- [10] M. Tremayne, C. C. Seaton, C. Glidewell, *Acta Cryst B* **2002**, *58*, 823.
- [11] Y. G. Andreev, P. Lightfoot, P. G. Bruce, **n.d.**, 12.
- [12] K. D. M. Harris, M. Tremayne, P. Lightfoot, P. G. Bruce, *J. Am. Chem. Soc.* **1994**, *116*, 3543.
- [13] K. D. M. Harris, M. Tremayne, *Chem. Mater.* **1996**, *8*, 2554.
- [14] M. Tremayne, B. M. Kariuki, K. D. M. Harris, K. Shankland, K. S. Knight, *J Appl Crystallogr* **1997**, *30*, 968.
- [15] W. I. F. David, K. Shankland, K. Shankland, N. Shankland, *Chem. Commun.* **1998**, 931.
- [16] L. Ward, K. Michel, C. Wolverton, *Phys. Rev. Materials* **2017**, *1*, 063802.
- [17] B. Meredig, C. Wolverton, *Nature Mater* **2013**, *12*, 123.
- [18] S. Billinge, *Physics* **2010**, *3*, 25.
- [19] S. J. L. Billinge, I. Levin, *Science* **2007**, *316*, 561.
- [20] R. S. Christensen, M. Kløve, M. Roelsgaard, S. Sommer, B. B. Iversen, *Nanoscale* **2021**, *13*, 12711.
- [21] S. Banerjee, C.-H. Liu, J. D. Lee, A. Kovyakh, V. Grasmik, O. Prymak, C. Koenigsmann, H. Liu, L. Wang, A. M. M. Abeykoon, S. S. Wong, M. Epple, C. B. Murray, S. J. L. Billinge, *J. Phys. Chem. C* **2018**, *122*, 29498.
- [22] A. S. Anker, T. L. Christiansen, M. Weber, M. Schmiele, E. Brok, E. T. S. Kjær, P. Juhás, R. Thomas, M. Mehring, K. M. Ø. Jensen, *Angewandte Chemie* **2021**, *133*, 20570.
- [23] E. D. Bøjesen, B. B. Iversen, *CrystEngComm* **2016**, *18*, 8332.
- [24] L. Yang, P. Juhás, M. W. Terban, M. G. Tucker, S. J. L. Billinge, *Acta Crystallogr A Found Adv* **2020**, *76*, 395.
- [25] L. Yang, E. A. Culbertson, N. K. Thomas, H. T. Vuong, E. T. S. Kjær, K. M. Ø. Jensen, M. G. Tucker, S. J. L. Billinge, *Acta Crystallogr A Found Adv* **2021**, *77*, 2.
- [26] D. A. Keen, R. L. McGreevy, *Nature* **1990**, *344*, 423.
- [27] F. Zhang, P. J. Chupas, S. L. A. Lui, J. C. Hanson, W. A. Caliebe, P. L. Lee, S.-W. Chan, *Chem. Mater.* **2007**, *19*, 3118.
- [28] P. Juhás, D. M. Cherba, P. M. Duxbury, W. F. Punch, S. J. L. Billinge, *Nature* **2006**, *440*, 655.
- [29] P. Juhás, L. Granlund, S. R. Gujarathi, P. M. Duxbury, S. J. L. Billinge, *J Appl Crystallogr* **2010**, *43*, 623.
- [30] M. K. Bisbo, B. Hammer, *Phys. Rev. B* **2022**, *105*, 245404.
- [31] M. K. Bisbo, B. Hammer, *Phys. Rev. Lett.* **2020**, *124*, 086102.
- [32] M. K. Bisbo, B. Hammer, *arXiv:2012.15222 [cond-mat, physics:physics]* **2020**.
- [33] P. Juhás, C. L. Farrow, X. Yang, K. R. Knox, S. J. L. Billinge, *Acta Crystallogr. A* **2015**, *71*, 562.
- [34] K. E. Sickafus, J. M. Wills, N. W. Grimes, *Journal of the American Ceramic Society* **1999**, *82*, 3279.
- [35] S. Sommer, E. D. Bøjesen, N. Lock, H. Kasai, J. Skibsted, E. Nishibori, B. B. Iversen, *Dalton Trans.* **2020**, *49*, 13449.
- [36] A.-C. Dippel, H.-P. Liermann, J. T. Delitz, P. Walter, H. Schulte-Schrepping, O. H. Seeck, H. Franz, *J. Synchrotron Radiat.* **2015**, *22*, 675.
- [37] N. Schell, A. King, F. Beckmann, T. Fischer, M. Müller, A. Schreyer, *Materials Science Forum* **2014**, *772*, 57.
- [38] C. Prescher, V. B. Prakapenka, *High Pressure Res.* **2015**, *35*, 223.
- [39] P. Juhás, T. Davis, C. L. Farrow, S. J. L. Billinge, *J. Appl. Crystallogr.* **2013**, *46*, 560.
- [40] X. Yang, P. Juhás, C. L. Farrow, S. J. L. Billinge, *arXiv:1402.3163 [cond-mat.mtrl-sci]* **2015**.
- [41] F. Søndergaard-Pedersen, N. L. N. Broge, J. Yu, M. Roelsgaard, B. B. Iversen, *Chem. Mater.* **2020**, *32*, 5134.
- [42] J. Becker, M. Bremholm, C. Tyrsted, B. Pauw, K. M. Ø. Jensen, J. Eltzholt, M. Christensen, B. B. Iversen, *J. Appl. Crystallogr.* **2010**, *43*, 729.

- [43] C. L. Farrow, P. Juhas, J. W. Liu, D. Bryndin, E. S. Božin, J. Bloch, T. Proffen, S. J. L. Billinge, *J. Phys. Condens. Matter* **2007**, *19*, 335219.
- [44] A. A. Coelho, *J. Appl. Crystallogr.* **2018**, *51*, 210.
- [45] J. Enkovaara, C. Rostgaard, J. J. Mortensen, J. Chen, M. Dulak, L. Ferrighi, J. Gavnholt, C. Glinsvad, V. Haikola, H. A. Hansen, H. H. Kristoffersen, M. Kuisma, A. H. Larsen, L. Lehtovaara, M. Ljungberg, O. Lopez-Acevedo, P. G. Moses, J. Ojanen, T. Olsen, V. Petzold, N. A. Romero, J. Stausholm-Møller, M. Strange, G. A. Tritsarlis, M. Vanin, M. Walter, B. Hammer, H. Häkkinen, G. K. H. Madsen, R. M. Nieminen, J. K. Nørskov, M. Puska, T. T. Rantala, J. Schiøtz, K. S. Thygesen, K. W. Jacobsen, *J. Phys.: Condens. Matter* **2010**, *22*, 253202.
- [46] A. Hjorth Larsen, J. Jørgen Mortensen, J. Blomqvist, I. E. Castelli, R. Christensen, M. Dulak, J. Friis, M. N. Groves, B. Hammer, C. Hargus, E. D. Hermes, P. C. Jennings, P. Bjerre Jensen, J. Kermode, J. R. Kitchin, E. Leonhard Kolsbjerg, J. Kubal, K. Kaasbjerg, S. Lysgaard, J. Bergmann Maronsson, T. Maxson, T. Olsen, L. Pastewka, A. Peterson, C. Rostgaard, J. Schiøtz, O. Schütt, M. Strange, K. S. Thygesen, T. Vegge, L. Vilhelmsen, M. Walter, Z. Zeng, K. W. Jacobsen, *J. Phys.: Condens. Matter* **2017**, *29*, 273002.