Quantifying the Likelihood of Structural Models through a Dynamically Enhanced Powder X-Ray Diffraction Protocol

Sander Borgmans, Sven M. J. Rogge,* Juul S. De Vos, Christian V. Stevens, Pascal Van Der Voort, and Veronique Van Speybroeck*

Abstract: Structurally characterizing new materials is tremendously challenging, especially when single crystal structures are hardly available which is often the case for covalent organic frameworks. Yet, knowledge of the atomic structure is key to establish structure-function relations and enable functional material design. Herein, a new protocol is proposed to unambiguously predict the structure of poorly crystalline materials through a likelihood ordering based on the X-ray diffraction (XRD) pattern. Key of the procedure is the broad set of structures generated from a limited number of building blocks and topologies, which is submitted to operando structural characterization. The dynamic averaging in the latter accounts for the operando conditions and inherent temporal character of experimental measurements, yielding unparalleled agreement with experimental powder XRD patterns. The proposed concept can hence unquestionably identify the structure of experimentally synthesized materials, a crucial step to design next generation functional materials.

Introduction

The design of functional materials is instrumental for advancing technological solutions to tackle pressing societal problems,[1–7] such as combating climate change and creating clean energy pathways. In the last few years, new materials with particular functionalities have been proposed at a very high pace, due to an in-depth microscopic understanding of their functionality and the growing toolbox of synthetic chemists. Within this toolbox lies the concept of reticular design,[8] where judiciously chosen building blocks are assembled in a predesigned topology, giving rise to materials such as metal-organic frameworks (MOFs)[9] and covalent organic frameworks (COFs).[10] Due to their extensive tailorability and attractiveness for a vast range of applications, the structural design of MOFs and COFs has formed a very active field of research.[11–15] Thanks to a meticulous control over synthesis conditions and an ever-expanding versatile set of experimentally available building blocks, it has even become possible to synthesize combinations of building blocks in different structures. To fully explore the functionalities of the enormous amount of potential structures that can be formed from a given set of building blocks, an accurate protocol is needed to systematically derive structural models for these materials at realistic working conditions of temperature and pressure.[16,17] Such structural characterization is a conditio sine qua non for the atomically guided design of functional applications in the fields of catalysis, sorption, light harvesting and more.

Reticular synthesis states that the atomic-level structure of any framework material is defined by a structural model containing three elements:[18] (i) the material's building blocks, (ii) its topology, and (iii) the embedding of the building blocks in this topology. When designing new experimental materials, X-ray crystallography is typically adopted to construct such a structural model. While single-crystal X-ray diffraction (SCXRD) measurements allow for the direct atomic-level determination of a material's bulk structure,[19] it is notoriously difficult to obtain large single crystals for most porous materials such as MOFs and COFs and one usually resorts to powder X-ray diffraction (PXRD) measurements[20] or electron crystallography.[21,22] During a PXRD measurement, however, the 3D reciprocal space is non-injectively projected onto a 1D Bragg angle, making the determination of the true structural model impossible without prior information about the structural model. Given the need for such an ad hoc model—which is absent in SCXRD—PXRD measurements require a method to verify the validity of these proposed structural models. Therefore, as discussed in more detail in Section S1.1, whole powder pattern fitting (WPPF) methods are used to either reject or accept a possible model based on a visual analysis of the experimental and modeled PXRD patterns, and on error measures such as the weighted

[*] S. Borgmans, Dr. S. M. J. Rogge, J. S. De Vos,
Prof. Dr. V. Van Speybroeck
Center for Molecular Modeling (CMM), Ghent University
Technologiepark 46, 9052 Zwijnaarde (Belgium)
E-mail: Sven.Rogge@UGent.be
Veronique.VanSpeybroeck@UGent.be

Prof. Dr. C. V. Stevens
Research Group SynBioC, Department of Green Chemistry and Technology
Faculty of Bioscience Engineering, Ghent University
Campus Coupure, Coupure Links 653, 9000 Gent (Belgium)

Prof. Dr. P. Van Der Voort
Center for Ordered Materials, Organometallics and Catalysis (CO-MOC)
Department of Inorganic and Physical Chemistry, Ghent University
Krijgslaan 281 (S3), 9000 Gent (Belgium)

Supporting information and the ORCID identification number(s) for the author(s) of this article can be found under:
https://doi.org/10.1002/anie.202017153.

profile residual.[23] However, when noisy PXRD patterns are obtained, such as for structures that lack long-range order, extracting any structural information becomes increasingly difficult. As a result, classical methods that rely on the diffraction data to refine initial structural parameters, such as the WPPF methods or other popular top-down methods, fail to accurately discern between multiple possible structural models.[24–28]

To circumvent this ambiguity and ensure that no structural models are overlooked when characterizing challenging functional materials, we here present an automated protocol that fully accounts for the dynamic averaging of structures taking place at operating conditions. The procedure, schematically illustrated in Figure 1, reverses the aforementioned experimental PXRD approach by exhaustively exploring all potential structures starting from the building blocks and all possible topologies, and identifying the most likely structural model upon comparing with an experimental PXRD pattern. To that end, a large set of possible structural models is generated from the provided building blocks, which are assembled in various topologies in the first step of our protocol. In the second step, PXRD patterns are derived, using either a static or a dynamic approach in which the inherent dynamic character of an experimental measurement is introduced. In the third and final step, a heuristic is derived and applied to order the proposed structural models as a function of their likelihood to reproduce the experimental diffraction pattern. Such a systematic computational protocol has the added benefit of replacing the ad hoc visual analysis by a well defined metric, allowing quantitative statements on how well the model performs.

COFs are ideal materials to test our protocol, as their structural characterization is challenging due to often noisy

PXRD patterns, owing to a lack of long-range order and crystallinity in most experimental samples. These promising materials are made up from light, covalently bound, organic constituents, combining low mass densities with a high stability. Potential applications are situated in the field of drug delivery,[29–31] (photo-)catalysis,[32–34] sensing,[35–37] (opto)-electronics[38–40] and gas/liquid storage and separation.[41–43] In general, one can distinguish between 3D COFs, which are completely connected by strong covalent bonds, and 2D COFs, which form layered structures whose stacking order is determined by much weaker forces, such as dispersion and Coulomb interactions.[44] Only a handful of SCXRD patterns are available for COFs, which are limited to the strongly connected 3D COFs.[45] Up to now, it is therefore common practice to propose initial structural COF models by choosing a topology based on the size of each of the building blocks, their point symmetry, and their connectivity with neighboring building blocks (see Figure 2 for representative examples), favoring the most symmetric topology when multiple possibilities arise.[46] However, with over 500 different COFs being synthesized,[47] accurate and systematic characterization techniques are crucial to enable rational functional material design.

Results and Discussion

Material Generation

The general applicability of the in silico protocol is illustrated for materials being built from seven different building blocks, illustrated in Figure 2, giving rise to at least five different experimentally characterized COFs. These
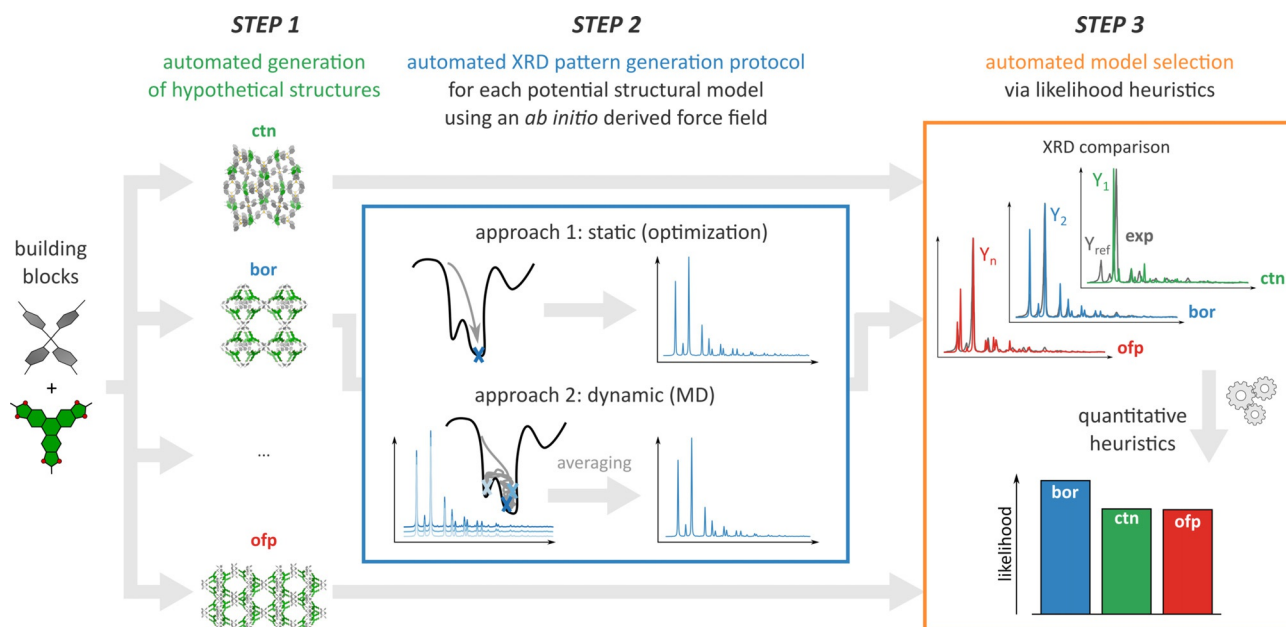

Figure 1. The proposed three-step structural characterization workflow. For each of the hypothetical structural models (generated in step 1), representative PXRD patterns are derived via either a static or a dynamic approach (step 2), which are then ranked according to their likelihood to reproduce the experimental reference PXRD pattern (step 3).
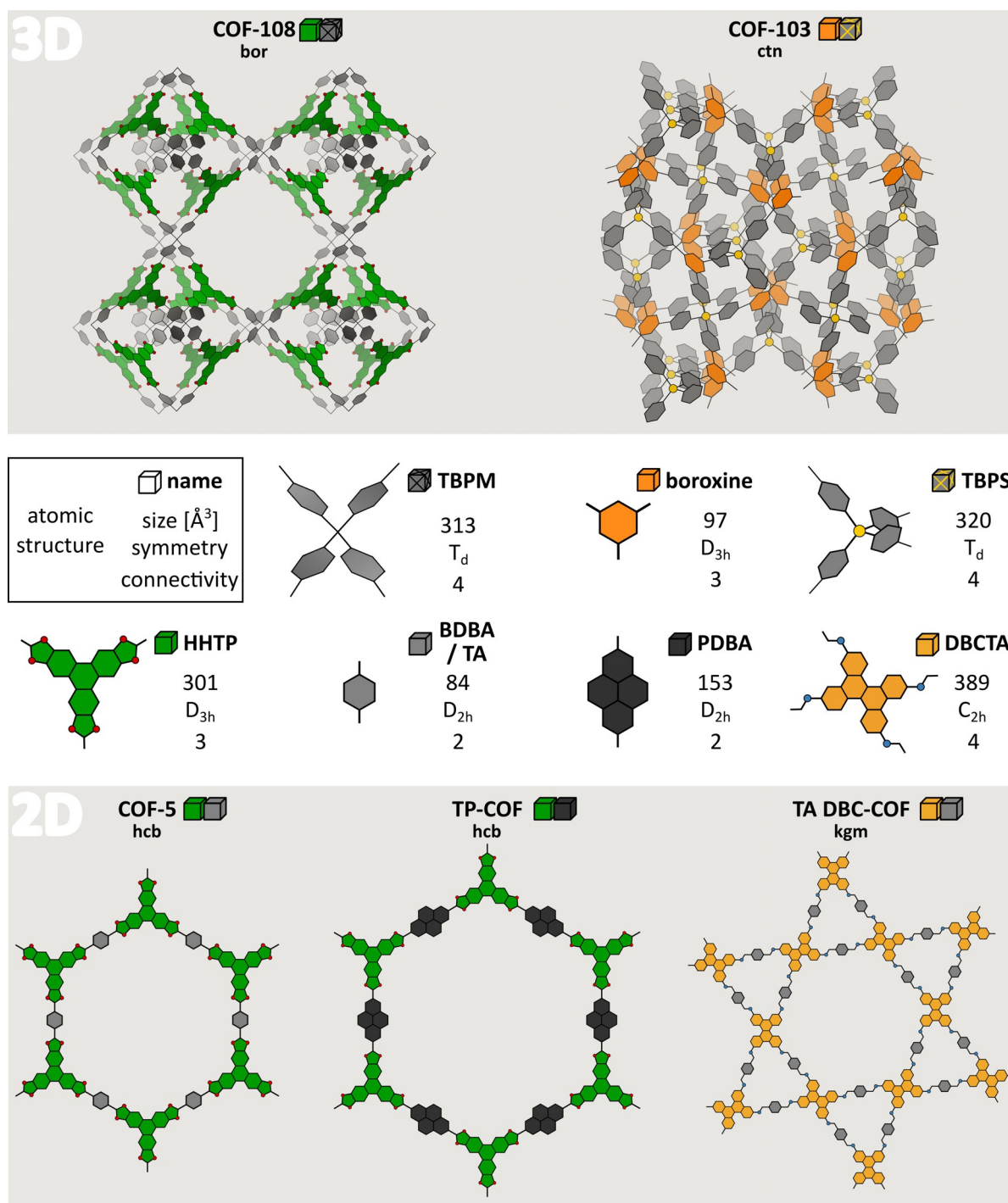
Figure 2. Building blocks and experimentally derived structural models in which they assemble for five experimentally synthesized COFs. The atomic structure and name of each building block (see Table S1 for the abbreviations) are reported, next to their size, symmetry, and connectivity, as expressed by the van der Waals volume, the point group, and number of extension points, respectively.

building blocks were chosen based on their varying sizes, symmetries and connectivities. The two 3D COFs, COF-103 and COF-108, were among the first 3D COFs to have ever been synthesized[48] and exhibit the **ctn** and **bor** topology, respectively. The bottom panel shows the three 2D COFs. Similar to the chosen 3D COFs, COF-5 was one of the first ever synthesized COFs,[10] adopting the **hcb** topology. TP-COF[49,50] was chosen as an isoreticularly expanded equivalent

of COF-5, whereas TA DBC-COF[51] was chosen due to its unique **kgm** topology.

Due to the modular nature of COFs, the seven building blocks of Figure 2 can be assembled in a combinatorial amount of structural models and hence hypothetical COF structures. When considering the five different combinations of two building blocks each that give rise to the five COFs in Figure 2, almost 700 unique structural models can be con-

Table 1: Number of 2D and 3D structural models generated by our in-house protocol (step 1 in Figure 1).

blocks	experimentally synthesized material	# hypothetical structural models	
		2D	3D
⬛ + ⬛	COF-5	42	
⬛ + ⬛	TP-COF	42	
⬛ + ⬛	TA DBC-COF	19	513 (55)
⬛ + ⬛	COF-103		26
⬛ + ⬛	COF-108		26

Block labels correspond to those in Figure 2. For TA DBC-COF the number of structural models with a 3D topology that are retained is reported between brackets.

structed by only looking at the possible topologies for the co-condensation of each building block combination. In Table 1, this number is broken down for each building block combination.

Most building block combinations give rise to only a tenfold of structural models due to the assumed rigidity of the building blocks. However, the building block combination that assembles TA DBC-COF can synthesize in 19 and 513 distinct 2D and 3D topologies, even though only a 2D structural model was resolved experimentally. This huge number of hypothetical models can be attributed to two factors. First, the slight non-planar character of the DBCTA building block accommodates both 2D and 3D topologies, in contrast to all other building blocks considered here. Second, removing the TA building block would preserve the original topology, similar as for any other twofold connected building block. Consequently, the TA DBC-COF can hypothetically synthesize in any of the large subset of topologies that contain only fourfold coordinated building blocks. By filtering out the most strained of these structural models, 19 2D and 55 3D hypothetical models are retained for the TA DBC-COF. For the five building block combinations of Figure 2, step 1 of our protocol hence generates a total of 210 structural models. For each of these models, a fully flexible ab initio based force field was derived using the QuickFF[52,53] protocol (see Section S2) to characterize and validate them with respect to the experimental PXRD patterns.

Automated Model Selection Protocol via Heuristics

An ideal PXRD pattern consists of discrete Bragg reflections, where each peak j is defined by its position θ_j, intensity I_j and shape,[23] as illustrated in Figure 3a. These components are uniquely defined for each structural model; deviations between experimental and representative computational PXRD patterns (see Section S1.2) can hence help localize model inaccuracies through these three components. To assess the correspondence among PXRD patterns, a quantitative measure is required, which, in our protocol, relies on the similarity index S_I and the weighted profile residual R_{wp}. A

detailed explanation and alternative figures of merit are outlined in Section S1.4. To understand how they penalize differences between two patterns, it is instructive to consider their variations for different types of deviations in either peak or trough regions. To this end, Figure 3b visualizes four hypothetical PXRD patterns (colored) that are compared to a given reference profile (in gray). Pattern 1 has a low deviation in both the experimental peak and trough regions, whereas pattern 4 has a high deviation in both regions. Pattern 2 exhibits a high deviation in the peak regions of the experimental PXRD pattern but a low deviation in the trough regions, whereas the opposite is true for pattern 3.

From Figure 3b, it is clear that the similarity index S_I mainly focuses on the overlap of the peaks, but is largely insensitive to deviations in the trough regions. This results in comparable similarity indices for PXRD patterns 1 and 3, and 2 and 4, which differ only in the trough regions. As S_I is normalized to one (see Eq. S1.10), direct comparison between various patterns is possible, in contrast to the R_{wp} factor. However, R_{wp} considers relative instead of absolute differences, as each peak is weighted with an uncertainty factor w_i that is inversely proportional to the reference profile's height. As a result, R_{wp} penalizes the mismatch between the profile intensities, both in peak and trough regions, as is evident when considering the similar R_{wp} factors for patterns 2 and 3. Since the overall relative intensity difference between the reference and the calculated PXRD patterns conveys no physical meaning, we will globally scale each computational pattern to minimize its weighted profile residual when comparing with the full experimental pattern. This will result in a unique R_{wp} value for each pattern, whereas S_I is unaffected as it is scale invariant. Since these figures of merit can vary significantly with the quality of the reference pattern, there is no specific heuristic value at which the agreement is deemed sufficient. However, as the best of two models is defined by a lower value of R_{wp}, ($R_{wp,ideal}=0$) and a higher value of S_I ($S_{I,ideal}=1$), the heuristic will allow for a quantitative ordering of which models are most likely to reproduce the experimental data, resulting in an automated model selection criterion.

Finally, the credibility of any figure of merit is significantly influenced by the presence of background noise in the experimental reference patterns. Especially for poorly crystalline COFs, significant background noise is expected, which originates from amorphous regions, and should be filtered. Clearly, a non-negligible experimental background signal at Bragg angles that correspond to a reflection peak in the computational pattern would artificially improve the figure of merit. As such, eliminating the background profile is essential before employing any PXRD heuristic. This is achieved here using the robust Bayesian analysis method (see Section S1.5),[54] resulting in a protocol that can also be applied if the material contains large amorphous regions. Henceforth, only figures of merit applied on the background filtered profile will be reported, which will be indicated by the "B" subscript.
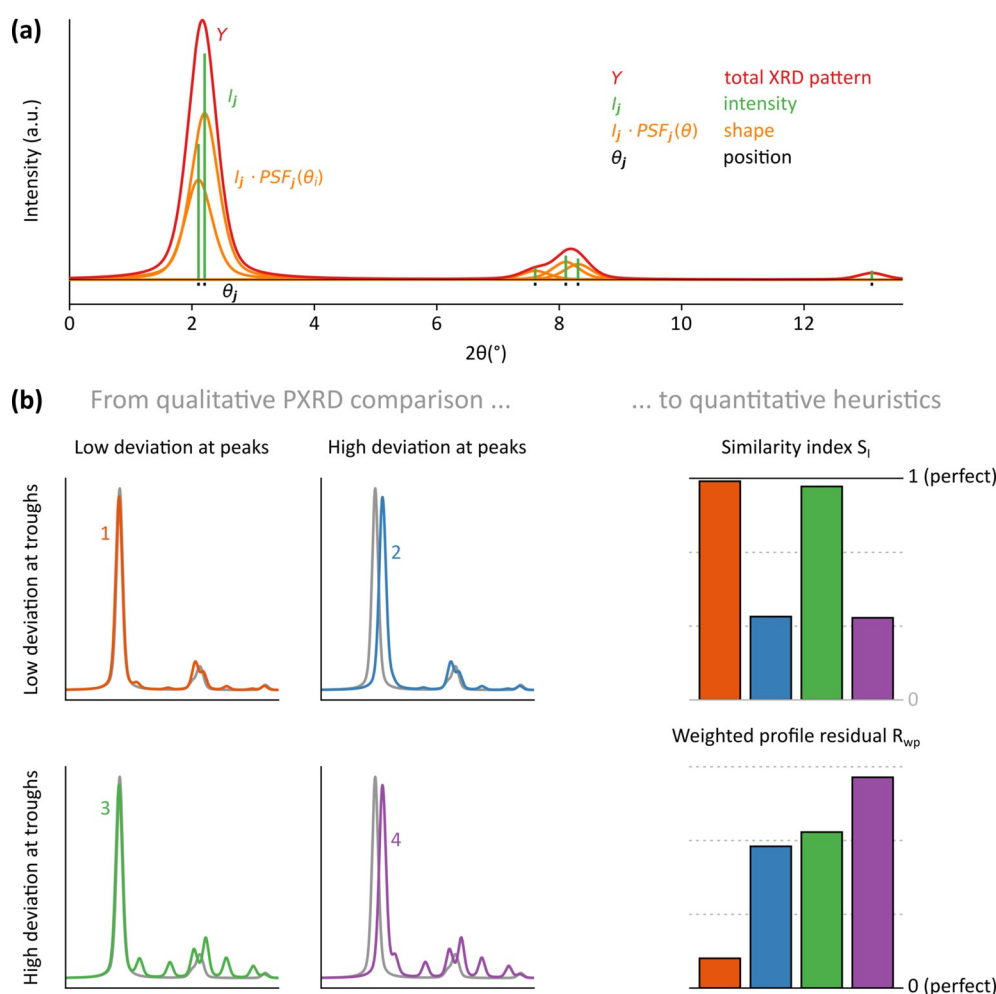
Figure 3. (a) Illustrative PXRD pattern (red) built up from six individual Bragg peaks that are each characterized by their position (black), intensity (green) and shape (orange). (b) Illustration of the figures of merits used in step 3 of the procedure for four hypothetical and representative PXRD patterns.

3D COFs

At first instance, the protocol is applied to the two building block combinations that give rise to the 3D COFs of Figure 2. In the static approach, all potential structural models of Table 1 are optimized towards their equilibrium structure at 0 K and the PXRD patterns of these optimized structures are derived following the procedure outlined in Section S1.3. The resulting likelihood ordering, based on the heuristic value, is visualized in Figure 4a,b, where the reference experimental PXRD patterns were extracted from ref. [48]. As the heuristic is calculated based on the relative difference between two profile intensities, the 2θ range is limited to the measured experimental data. Consequently, any diffraction peak of our hypothetical models that falls outside this experimental 2θ range is not taken into account when calculating the heuristics. As this can artificially increase the heuristic results for these structural models—it is expected that the experimental pattern has no matching peaks outside the experimental 2θ range—the names of these structural models are indicated by an asterisk (*) in Figure 4 and should be considered with care.

The green regions in Figure 4a,b correspond to successfully retrieved experimental structures, whereas the gray regions indicate other likely potential structural models based on the heuristic values. For the 3D-COFs, COF-103 and COF-108, the static approach succeeds in identifying the correct experimentally resolved structural model when taking both the $S_{I,B}$ and $R_{wp,B}$ values into account. The gray zones correspond to structures which might potentially be viable, but it is immediately clear that only one heuristic value is favorable, either $S_{I,B}$ is close to one or $R_{wp,B}$ is close to zero. For COF-103, shown in Figure 4a, an appreciable difference is observed between the best **ctn** topology and the second best **mmm** topology, facilitating the correct model selection. While our protocol also correctly identifies the experimentally reported topology for COF-108 (Figure 4b), the distinction between the best topology (**bor**, indicated in green) and the second best topology (**asn** when considering $R_{wp,B}$ and **ofp** when considering $S_{I,B}$, indicated in gray) is smaller than for COF-103 due to the larger residual experimental background for COF-108 (see Section S1.5). Clearly, the static approach of our protocol is able to correctly identify the most probable structural model for 3D COFs, which are connected
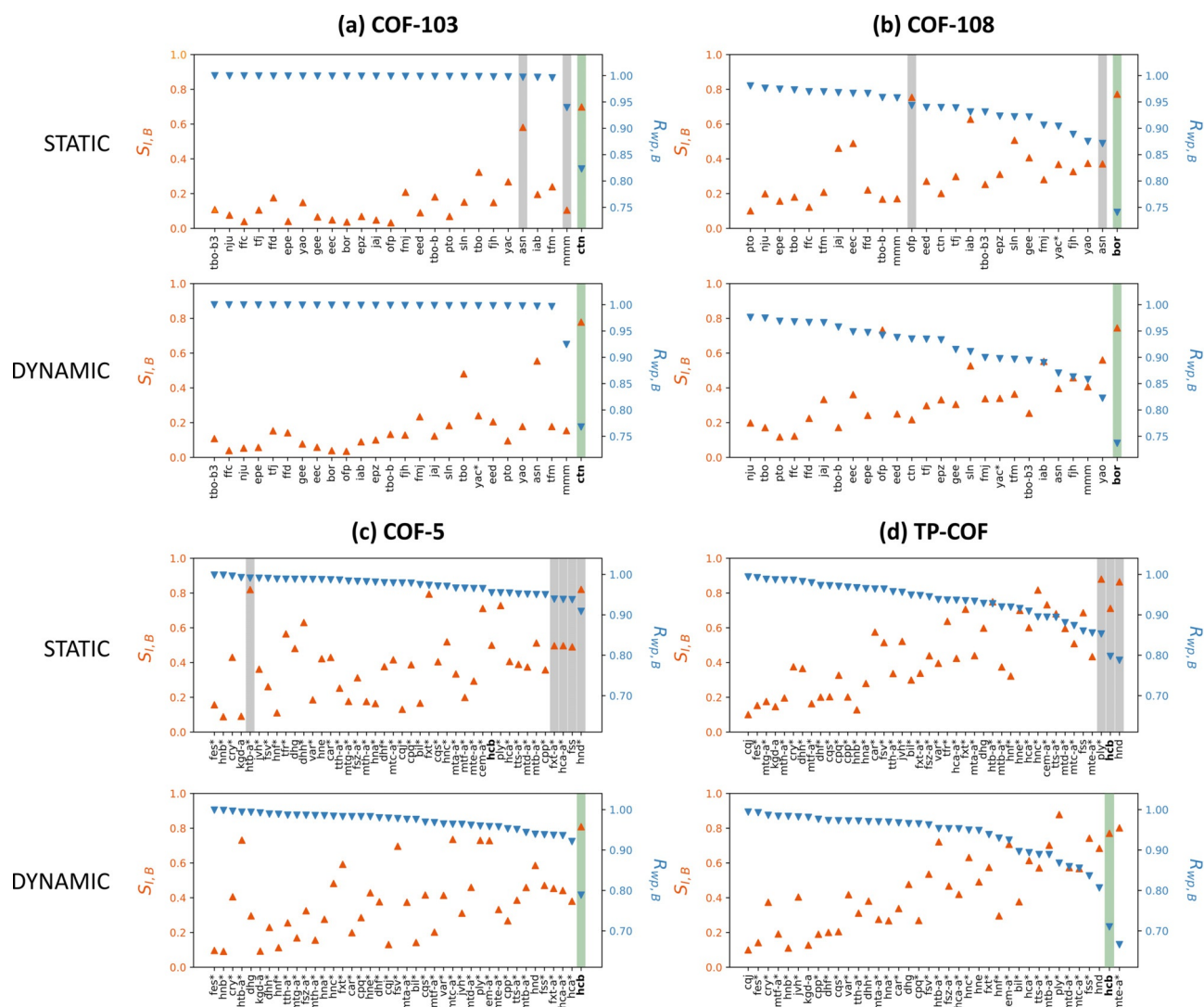
Figure 4. Similarity index $S_{I,B}$ (orange) and weighted profile index $R_{wp,B}$ (blue) ordered as a function of decreasing $R_{wp,B}$ for (a) COF-103, (b) COF-108, (c) COF-5 and (d) TP-COF, without background noise, using either a (top) static or (bottom) dynamic approach. The green regions indicate a successfully retrieved structural model corresponding to the experimentally reported topology (in bold face), whereas the gray regions correspond to favorable heuristic values that are unsuccessful. An asterisk (*) indicates that the heuristic results for that topology are artificially increased due to significant peaks outside the experimental 2θ range, which are not taken into account when comparing the PXRD patterns.

through strong covalent bonds and thus less prone to dynamic effects.

2D COFs

Turning our attention to the 2D COFs, visualized in Figure 4 c,d, our static approach is unable to identify the best structural models for COF-5 and TP-COF, with no distinct outlier, in contrast to the 3D COFs. The **htb-a**, **fxt-a**, **hca-a**, **fss** and **hnd** topologies for COF-5, and the **ply**, **hcb** and **hnd** topologies for TP-COF, indicated in gray, have similar $S_{I,B}$ and/or $R_{wp,B}$ values. However, an optimal heuristic combination is missing for the experimentally reported **hcb** topology of both materials.[10,50] This mismatch originates from the inability of the static approach to account for dynamic averaging at working temperatures, as only one low-energy

structure with a fixed layer inclination is considered (see also Section S5). This limitation is expected to manifest itself for every 2D COF, as their layers are held together by weak, non-covalent forces. However, for TA DBC-COF the static approach does predict the correct **kgm** topology, as in this case the DBCTA building blocks hinder layer shearing due to steric constraints (Figure S18).

To resolve these dynamic effects in structures that have a large intrinsic freedom, characterized by shallow minima on the free energy surface due to the presence of weak interactions, it is necessary to generate PXRD patterns through MD simulations at operando conditions of temperature and pressure (see Figure 1). In this dynamic approach, the PXRD pattern of any structural model is obtained by averaging PXRD patterns collected at different snapshots from the MD trajectory of this structural model. As an example, Figure 5b visualizes the experimental PXRD pat-
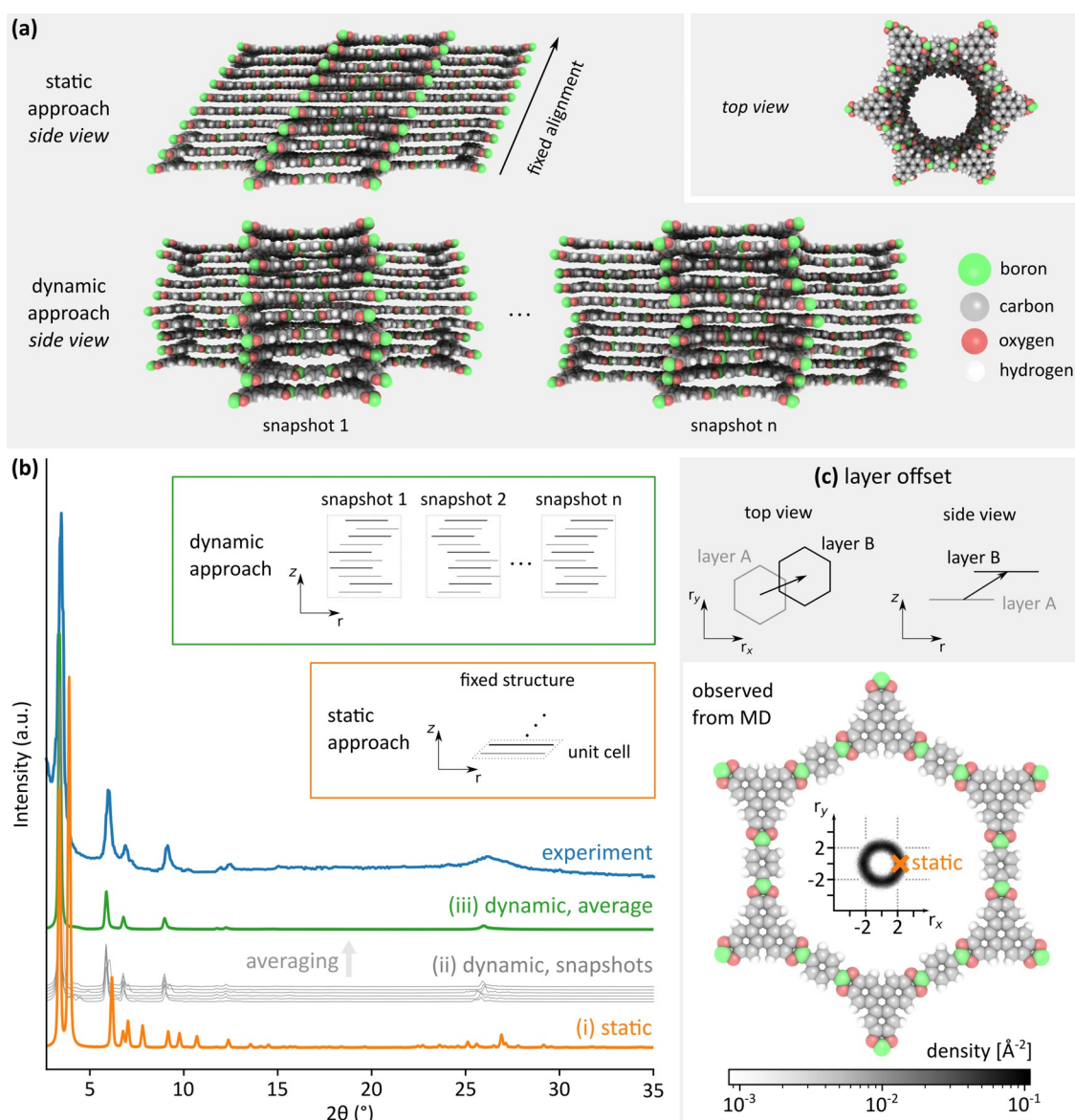
Figure 5. (a) Atomistic representation of the stacking in both the static and dynamic approach, containing two and ten layers, respectively, in their unit cell. (b) Diffraction pattern comparison for COF-5 using the **hcb** topology for the optimized structure (static, orange), six representative snapshots taken from the MD trajectory (gray) and the diffraction pattern obtained by averaging over fifty snapshots of the MD trajectory (dynamic, green). (c) Definition of the layer offset between subsequent layers, showing the variation of this offset during an MD simulation as a heatmap.

tern of COF-5 together with (i) the diffraction pattern of the optimized **hcb** structural model (static approach), (ii) the diffraction patterns corresponding with several snapshots through the MD simulation, and (iii) the diffraction pattern obtained by averaging over the patterns obtained at 50 uniformly distributed MD snapshots, where each snapshot corresponds to an embedding in the **hcb** topology (dynamic approach). Clearly, the time-averaged PXRD pattern of the dynamic approach significantly improves the representation of the true atomic structure during measurement, increasing the similarity between the simulated and experimental PXRD patterns.

Turning our attention to the differences between the computational PXRD patterns in Figure 5b, two observations

can be made. First, compared to the PXRD pattern in the static approach, the reflection peaks occurring in the PXRD patterns of the MD snapshots are either shifted towards lower 2θ values or disappear completely, with the exception of the first peak at $2\theta = 3.4°$. While the shift in peak positions indicates an increase in distance between the diffraction planes at finite temperatures, the complete disappearance of peaks indicates that the static structure is too symmetric, featuring extra diffraction planes compared to the experimental structure. This artificial symmetry in optimized structures is a well-known effect,[55] which is resolved by the MD simulations at finite temperature in our dynamic approach, as demonstrated in Figure 5b. Second, different MD snapshots give slightly different PXRD patterns. These

small deviations are expected, as they correspond to the natural dynamics of the atoms at operando conditions and replace the missing atomic displacement factors in Eq. S1.4. Therefore, the MD-averaged PXRD pattern obtained in the dynamic approach contains, in a natural way, the broadening effects that are absent in the static approach. As a result, Figure 5b demonstrates that our dynamic approach leads to a much better correspondence with the experimental PXRD pattern compared to the static approach.

For 2D COFs, the largest discrepancy between the static and dynamic approach finds its origin in the ability to describe the dynamic character of the layer movement, which is evident from Figure 5a. In a static simulation, the layer offset (see Figure 5c and Section S1.7) has a fixed magnitude and direction, as defined by the cell vectors. Considering two layers in the optimized structure, they are either aligned, yielding an eclipsed stacking (AA), or displaced with respect to each other, yielding serrated (AB) or inclined stacking configurations, where the latter is illustrated in Figure 5a. In contrast, in a dynamic simulation, the layer offset fluctuates in both magnitude and direction during the simulation, as shown in Figure 5a for two snapshots. Moreover, in our dynamic approach, a larger system consisting of ten layers is considered, giving rise to more complex layer behavior. Figure 5c shows the observed layer offsets both in a static and dynamic approach through a heatmap. Dynamically, a range of offsets is found, all with a magnitude around 2.2 Å without any preferential directionality, whereas in the static approach only one single offset is found with the same magnitude but a fixed direction, which is insufficient to account for the dynamics of the system.

Based on the COF-5 case study in Figure 5, one expects the dynamic approach to yield a much better heuristic value compared to the static approach for 2D COFs consisting of weakly bound layers. Figure 4c,d indeed reveals significant alterations when comparing the dynamic with the static approach for 2D COFs. Remarkably, in the dynamic approach, the most likely structural model coincides with the expected topology for COF-5 and, when ignoring the structural models that contain diffraction peaks outside the experimental 2θ range, also for TP-COF. We also adopted the dynamic approach for the 3D COFs and TA DBC-COF, finding no substantial difference with the static approach due to the strong connectivity and steric hindrance in these materials allowing only minimal deviations under finite temperatures and pressures (see Figure 4a,b and Figure S18). These results illustrate the general applicability of our dynamic protocol to predict the correct structural model.

Model Refinement

While the dynamic approach of the in silico protocol correctly predicts the topology for all COFs considered here, the precise embedding of the building blocks inside the topology can still give rise to different structural models. In turn these can lead to slight variations in the cell vectors and can significantly influence the intensities of the XRD patterns (see Section S1.2). Consider for instance TA DBC-COF,

which assembles in the **kgm** topology. Even within this topology, different relative orientations of the imine linkages give rise to three different pore structures and distinct diffraction patterns, as illustrated in Figure 6. In other words, the TA building block can be embedded in three different ways, with the imine linkages of TA pointing in opposite directions (kgm1), or with the imine linkages both pointing in the same direction, which is either bent away (kgm2) or towards (kgm3) the smaller triangular pore. These different embeddings can not all be reached during a regular MD simulation as they correspond to different (meta)stable configurations that are separated by too large energetic barriers.

The heuristic values for the different embeddings reported in this figure demonstrate that the kgm3 model best reproduces the experimental PXRD pattern. Surprisingly, this structural model does not agree with the experimentally proposed model (kgm1), which is illustrated in Figure 2. This experimental model was constructed ad hoc in Material Studio and subsequently refined using the Pawley method,[51] which, however, does not fit the atomic positions, as discussed in Section S1.1. This case study illustrates that, even when the topology is correctly identified, it is instructive to consider whether different embeddings of the building blocks within the topology are possible, at least when multiple (meta)stable configurations exist. From these results it follows that our protocol not only succeeds in finding the optimal structural model for newly synthesized materials, but can also improve upon existing models through refinement of the local geometry and embedding, increasing the similarity between the experimentally observed and calculated diffraction patterns in a quantitative manner.

Conclusion

Crystal structure determination of complex nanoporous materials is key for establishing structure-function relations in material design. However, finding the atomic-level structure of experimentally synthesized materials may become a daunting task. This is certainly true for materials that lack long-range order, such as some COFs, where current techniques struggle to systematically and reproducibly identify the correct structural model as they are highly sensitive to the quality of the experimental data and are prone to confirmation bias. Herein, we presented an in silico protocol that reverses the generally applied strategy of extracting possible models from the observed PXRD peaks. Instead, by generating an exhaustive set of structural models for any given building block combination and generating representative PXRD patterns for each model through a static or a dynamic approach that accounts for the operando experimental conditions, the different structural models are ranked based on intuitive heuristics that unambiguously quantify the correspondence between the computed and experimental PXRD patterns, thereby uniquely identifying the most likely model.

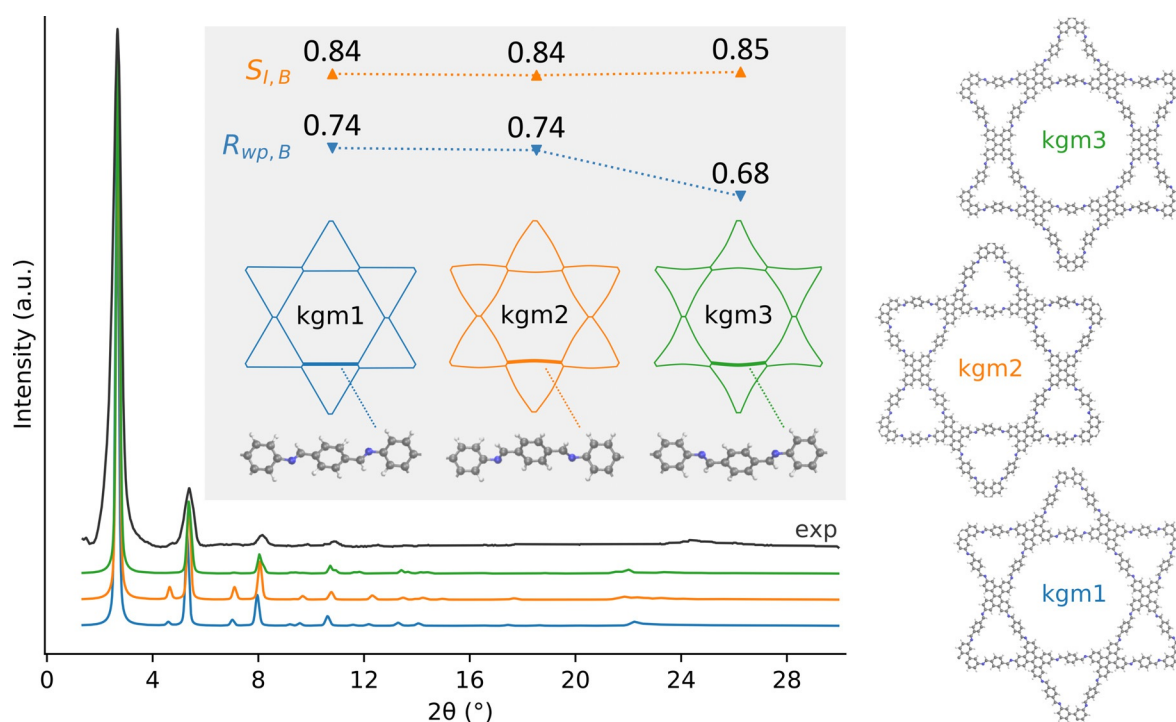The procedure was validated for several well-characterized COFs, but is equally applicable to other modular building

Figure 6. TA DBC-COF model refinement for three distinct structural models with the **kgm** topology but varying orientation of the imine linkage and hence varying embedding, resulting in a different pore geometry and PXRD heuristic value. The PXRD patterns are obtained experimentally[51] (gray, background-filtered) or through our dynamic approach (colored).

block materials, such as MOFs and zeolites, in any topology. Starting from seven building blocks, many structural models were proposed, through all possible topologies. For 3D COFs, a static approach without dynamic averaging proved sufficient considering their characteristic strong covalent bonds. However, for 2D COFs, which show a high degree of flexibility due to the weak interaction between the layers, it was essential to account for different dynamically generated snapshots at experimental temperatures. This approach yields a natural analog to experimental XRD measurements and corrects for an artificially too symmetric structure obtained at 0 K in a static approach. Thus, the in silico protocol proposed here can successfully detect experimental crystal structures, even for 2D COFs. Interestingly, our approach also succeeds in showing finer details of the structure. For the TA DBC-COF, our procedure can distinguish between various structural models that all exhibit the experimental topology but have different orientations of the imine linkages. Such subtleties would easily be overlooked when visually comparing PXRD peak positions.

The presented in silico workflow opens the perspective to systematically screen large databases of materials, to identify their atomic-level structural models, and establish causal structure-function relations. We hope that such detailed knowledge will help computational and experimental material scientists alike by opening the door for a more precise design of challenging modular materials for various applications.

Acknowledgements

Conflict of interest

The authors declare no conflict of interest.

[1] B. M. Connolly, J. P. Mehta, P. Z. Moghadam, A. E. H. Wheatley, D. Fairen-Jimenez, *Curr. Opin. Green Sustainable Chem.* **2018**, *12*, 47–56.
[2] X. Zhao, P. Pachfule, S. Li, T. Langenhahn, M. Ye, G. Tian, J. Schmidt, A. Thomas, *Chem. Mater.* **2019**, *31*, 3274–3280.

[3] K. H. Cho, D. D. Borges, U.-H. Lee, J. S. Lee, J. W. Yoon, S. J. Cho, J. Park, W. Lombardo, D. Moon, A. Sapienza, G. Maurin, J.-S. Chang, *Nat. Commun.* **2020**, *11*, 5112.

[4] H. Zhong, M. Ghorbani-Asl, K. H. Ly, J. Zhang, J. Ge, M. Wang, Z. Liao, D. Makarov, E. Zschech, E. Brunner, I. M. Weidinger, J. Zhang, A. V. Krasheninnikov, S. Kaskel, R. Dong, X. Feng, *Nat. Commun.* **2020**, *11*, 1409.

[5] K. Gottschling, G. Savasci, H. Vignolo-González, S. Schmidt, P. Mauker, T. Banerjee, P. Rovó, C. Ochsenfeld, B. V. Lotsch, *J. Am. Chem. Soc.* **2020**, *142*, 12146–12156.

[6] B. M. Connolly, D. G. Madden, A. E. H. Wheatley, D. Fairen-Jimenez, *J. Am. Chem. Soc.* **2020**, *142*, 8541–8549.

[7] J. Zheng, D. Barpaga, B. A. Trump, M. Shetty, Y. Fan, P. Bhattacharya, J. J. Jenks, C.-Y. Su, C. M. Brown, G. Maurin, B. P. McGrail, R. K. Motkuri, *J. Am. Chem. Soc.* **2020**, *142*, 3002–3012.

[8] O. M. Yaghi, M. O'Keeffe, N. W. Ockwig, H. K. Chae, M. Eddaoudi, J. Kim, *Nature* **2003**, *423*, 705–714.

[9] H. Li, M. Eddaoudi, M. O'Keeffe, O. M. Yaghi, *Nature* **1999**, *402*, 276–279.

[10] A. P. Côté, A. I. Benin, N. W. Ockwig, M. O'Keeffe, A. J. Matzger, O. M. Yaghi, *Science* **2005**, *310*, 1166–1170.

[11] X. Zhao, P. Pachfule, S. Li, T. Langenhahn, M. Ye, C. Schlesiger, S. Praetz, J. Schmidt, A. Thomas, *J. Am. Chem. Soc.* **2019**, *141*, 6623–6630.

[12] K. S. Deeg, D. D. Borges, D. Ongari, N. Rampal, L. Talirz, A. V. Yakutovich, J. M. Huck, B. Smit, *ACS Appl. Mater. Interfaces* **2020**, *12*, 21559–21568.

[13] S. Park, Z. Liao, B. Ibarlucea, H. Qi, H.-H. Lin, D. Becker, J. Melidonie, T. Zhang, H. Sahabudeen, L. Baraban, C.-K. Baek, Z. Zheng, E. Zschech, A. Fery, T. Heine, U. Kaiser, G. Cuniberti, R. Dong, X. Feng, *Angew. Chem. Int. Ed.* **2020**, *59*, 8218–8224; *Angew. Chem.* **2020**, *132*, 8295–8301.

[14] C. Krishnaraj, H. Sekhar Jena, L. Bourda, A. Laemont, P. Pachfule, J. Roeser, C. V. Chandran, S. Borgmans, S. M. J. Rogge, K. Leus, C. V. Stevens, J. A. Martens, V. Van Speybroeck, E. Breynaert, A. Thomas, P. Van Der Voort, *J. Am. Chem. Soc.* **2020**, *142*, 20107–20116.

[15] R. K. Sharma, P. Yadav, M. Yadav, R. Gupta, P. Rana, A. Srivastava, R. Zbořil, R. S. Varma, M. Antonietti, M. B. Gawande, *Mater. Horiz.* **2020**, *7*, 411–454.

[16] J. Keupp, R. Schmid, *Faraday Discuss.* **2018**, *211*, 79–101.

[17] A. L. Goodwin, *Nat. Commun.* **2019**, *10*, 4461.

[18] See Ref. [8].

[19] Y.-B. Zhang, J. Su, H. Furukawa, Y. Yun, F. Gándara, A. Duong, X. Zou, O. M. Yaghi, *J. Am. Chem. Soc.* **2013**, *135*, 16336–16339.

[20] C. S. Diercks, O. M. Yaghi, *Science* **2017**, *355*, eeal1585.

[21] J. Li, J. Sun, *Acc. Chem. Res.* **2017**, *50*, 2737–2745.

[22] J.-H. Dou, M. Q. Arguilla, Y. Luo, J. Li, W. Zhang, L. Sun, J. L. Mancuso, L. Yang, T. Chen, L. R. Parent, G. Skorupskii, N. J. Libretto, C. Sun, M. C. Yang, P. V. Dip, E. J. Brignole, J. T. Miller, J. Kong, C. H. Hendon, J. Sun, M. Dincă, *Nat. Mater.* **2020**, *20*, 222–228.

[23] V. Pecharsky, P. Zavalij, *Fundamentals of Powder Diffraction and Structural Characterization of Materials*, Springer US, **2009**.

[24] A. Le Bail, *Powder Diffr.* **2005**, *20*, 316–326.

[25] G. S. Pawley, *J. Appl. Crystallogr.* **1981**, *14*, 357–361.

[26] A. Le Bail, H. Duroy, J. L. Fourquet, *Mater. Res. Bull.* **1988**, *23*, 447–452.

[27] L. B. McCusker, *Acta Crystallogr. Sect. A* **1991**, *47*, 297–313.

[28] M. Falcioni, M. W. Deem, *J. Chem. Phys.* **1999**, *110*, 1754–1766.

[29] Q. Fang, J. Wang, S. Gu, R. B. Kaspar, Z. Zhuang, J. Zheng, H. Guo, S. Qiu, Y. Yan, *J. Am. Chem. Soc.* **2015**, *137*, 8352–8355.

[30] V. S. Vyas, M. Vishwakarma, I. Moudrakovski, F. Haase, G. Savasci, C. Ochsenfeld, J. P. Spatz, B. V. Lotsch, *Adv. Mater.* **2016**, *28*, 8749–8754.

[31] L. Bai, S. Z. F. Phua, W. Q. Lim, A. Jana, Z. Luo, H. P. Tham, L. Zhao, Q. Gao, Y. Zhao, *Chem. Commun.* **2016**, *52*, 4128–4131.

[32] S. Lin, C. S. Diercks, Y.-B. Zhang, N. Kornienko, E. M. Nichols, Y. Zhao, A. R. Paris, D. Kim, P. Yang, O. M. Yaghi, C. J. Chang, *Science* **2015**, *349*, 1208–1213.

[33] H. Li, Q. Pan, Y. Ma, X. Guan, M. Xue, Q. Fang, Y. Yan, V. Valtchev, S. Qiu, *J. Am. Chem. Soc.* **2016**, *138*, 14783–14788.

[34] G. Lin, H. Ding, R. Chen, Z. Peng, B. Wang, C. Wang, *J. Am. Chem. Soc.* **2017**, *139*, 8705–8709.

[35] G. Das, B. P. Biswal, S. Kandambeth, V. Venkatesh, G. Kaur, M. Addicoat, T. Heine, S. Verma, R. Banerjee, *Chem. Sci.* **2015**, *6*, 3931–3939.

[36] S.-Y. Ding, M. Dong, Y.-W. Wang, Y.-T. Chen, H.-Z. Wang, C.-Y. Su, W. Wang, *J. Am. Chem. Soc.* **2016**, *138*, 3031–3037.

[37] S. Dalapati, E. Jin, M. Addicoat, T. Heine, D. Jiang, *J. Am. Chem. Soc.* **2016**, *138*, 5797–5800.

[38] S. Wan, J. Guo, J. Kim, H. Ihee, D. Jiang, *Angew. Chem. Int. Ed.* **2009**, *48*, 5439–5442; *Angew. Chem.* **2009**, *121*, 5547–5550.

[39] X. Feng, L. Liu, Y. Honsho, A. Saeki, S. Seki, S. Irle, Y. Dong, A. Nagai, D. Jiang, *Angew. Chem. Int. Ed.* **2012**, *51*, 2618–2622; *Angew. Chem.* **2012**, *124*, 2672–2676.

[40] M. Calik, F. Auras, L. M. Salonen, K. Bader, I. Grill, M. Handloser, D. D. Medina, F. Dogru, F. Löbermann, D. Trauner, A. Hartschuh, T. Bein, *J. Am. Chem. Soc.* **2014**, *136*, 17802–17807.

[41] C. J. Doonan, D. J. Tranchemontagne, T. G. Glover, J. R. Hunt, O. M. Yaghi, *Nat. Chem.* **2010**, *2*, 235–238.

[42] L. A. Baldwin, J. W. Crowe, D. A. Pyles, P. L. McGrier, *J. Am. Chem. Soc.* **2016**, *138*, 15134–15137.

[43] X. Zhan, Z. Chen, Q. Zhang, *J. Mater. Chem. A* **2017**, *5*, 14463–14479.

[44] S. B. Alahakoon, S. D. Diwakara, C. M. Thompson, R. A. Smaldone, *Chem. Soc. Rev.* **2020**, *49*, 1344–1356.

[45] T. Ma, E. A. Kapustin, S. X. Yin, L. Liang, Z. Zhou, J. Niu, L.-H. Li, Y. Wang, J. Su, J. Li, X. Wang, W. D. Wang, W. Wang, J. Sun, O. M. Yaghi, *Science* **2018**, *361*, 48–52.

[46] P. J. Waller, F. Gándara, O. M. Yaghi, *Acc. Chem. Res.* **2015**, *48*, 3053–3063.

[47] D. Ongari, A. V. Yakutovich, L. Talirz, B. Smit, *ACS Cent. Sci.* **2019**, *5*, 1663–1675.

[48] H. M. El-Kaderi, J. R. Hunt, J. L. Mendoza-Cortés, A. P. Côté, R. E. Taylor, M. O'Keeffe, O. M. Yaghi, *Science* **2007**, *316*, 268–272.

[49] S. Wan, J. Guo, J. Kim, H. Ihee, D. Jiang, *Angew. Chem. Int. Ed.* **2008**, *47*, 8826–8830; *Angew. Chem.* **2008**, *120*, 8958–8962.

[50] B. J. Smith, N. Hwang, A. D. Chavez, J. L. Novotney, W. R. Dichtel, *Chem. Commun.* **2015**, *51*, 7532–7535.

[51] N. Keller, T. Sick, N. N. Bach, A. Koszalkowski, J. M. Rotter, D. D. Medina, T. Bein, *Nanoscale* **2019**, *11*, 23338–23345.

[52] L. Vanduyfhuys, S. Vandenbrande, T. Verstraelen, R. Schmid, M. Waroquier, V. Van Speybroeck, *J. Comput. Chem.* **2015**, *36*, 1015–1027.

[53] L. Vanduyfhuys, S. Vandenbrande, J. Wieme, M. Waroquier, T. Verstraelen, V. Van Speybroeck, *J. Comput. Chem.* **2018**, *39*, 999–1011.

[54] W. I. F. David, D. S. Sivia, *J. Appl. Crystallogr.* **2001**, *34*, 318–324.

[55] A. M. Pütz, M. W. Terban, S. Bette, F. Haase, R. E. Dinnebier, B. V. Lotsch, *Chem. Sci.* **2020**, *11*, 12647–12654.