This is an open access article published under an ACS AuthorChoice License, which permits copying and redistribution of the article or any adaptations for non-commercial purposes.

ACS central science

Research Article

Cite This: ACS Cent. Sci. 2019, 5, 1663−1675

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

Building a Consistent and Reproducible Database for Adsorption Evaluation in Covalent−Organic Frameworks

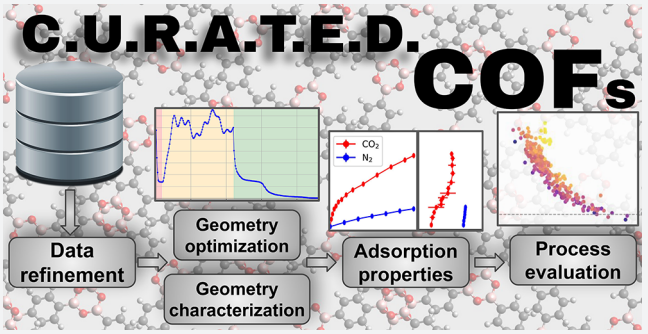
Daniele Ongari,[†] Aliaksandr V. Yakutovich,[†,‡] Leopold Talirz,[†,‡] and Berend Smit*[,†]

[†]Laboratory of Molecular Simulation (LSMO), Institut des Sciences et Ingénierie Chimiques, École Polytechnique Fédérale de Lausanne (EPFL), Rue de l'Industrie 17, Sion, CH-1951 Valais, Switzerland

[‡]Theory and Simulation of Materials (THEOS), Faculté des Sciences et Techniques de l'Ingénieur, École Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland

Supporting Information

ABSTRACT: We present a workflow that traces the path from the bulk structure of a crystalline material to assessing its performance in carbon capture from coal's postcombustion flue gases. This workflow is applied to a database of 324 covalent−organic frameworks (COFs) reported in the literature, to characterize their CO_2 adsorption properties using the following steps: (1) optimization of the crystal structure (atomic positions and unit cell) using density functional theory, (2) fitting atomic point charges based on the electron density, (3) characterizing the pore geometry of the structures before and after optimization, (4) computing carbon dioxide and nitrogen isotherms using grand canonical Monte Carlo simulations with an empirical interaction potential, and finally, (5) assessing the CO_2 parasitic energy via process modeling. The full workflow has been encoded in the Automated Interactive Infrastructure and Database for Computational Science (AiiDA). Both the workflow and the automatically generated provenance graph of our calculations are made available on the Materials Cloud, allowing peers to inspect every input parameter and result along the workflow, download structures and files at intermediate stages, and start their research right from where this work has left off. In particular, our set of CURATED (Clean, Uniform, and Refined with Automatic Tracking from Experimental Database) COFs, having optimized geometry and high-quality DFT-derived point charges, are available for further investigations of gas adsorption properties. We plan to update the database as new COFs are being reported.



INTRODUCTION

Covalent organic frameworks (COFs) are a class of emerging materials composed of covalently bonded organic residues.[1] While classical covalent crystals like diamond, graphene, or carbon nitride could also be considered as COFs, this nomenclature highlights the concept of reticular design, where the precursor molecules, termed *ligands* or *linkers*, are rationally combined into a framework structure. The ancestors of COFs are metal organic frameworks (MOFs), where ligands form coordinate bonds with metal centers. In COFs, however, ligands are directly connected, forming the topology of the structure.

The first two-dimensional microporous COFs, named COF-1 and COF-5, were synthesized by condensation reactions of phenyl diboronic acid in the group of Yaghi.[2] Two years later, the same group employed a similar condensation reaction to obtain the first three-dimensional COFs.[3] To date, hundreds of different COFs have been reported by different research groups.[4,5] This new class of materials attracted the interest of the scientific community for applications ranging from gas adsorption[6] to catalysis[7] and photocatalysis,[8] setting the stage for a possible commercial use in the near future.[9] In addition to the synthesized COFs, new hypothetical COF structures have been designed *in silico*[10−13] and investigated to highlight possible promising frameworks that experimental chemists could target in their synthesis. The total number of these hypothetical structures exceeds 560 000.

Similarly to MOFs, we see a rapid increase in the number of novel COFs that are being reported. Often these materials are synthesized with a particular application in mind, and experimental data are often limited to this application; it is simply not practical to experimentally test a novel material for all possible applications. For such studies a computational screening to identify the most promising candidates is often the most efficient first step. Essential in these computational studies is that all experimental structures are curated in, what is referred to in the literature as, "Computation-Ready Experimental" (CoRE) structures,[14,15] The importance of this curation step cannot be emphasized enough, and as we will argue in this work, in particular for COFs, the lack of a

Received: June 25, 2019
Published: September 26, 2019

 © 2019 American Chemical Society

DOI: 10.1021/acscentsci.9b00619
ACS Cent. Sci. 2019, 5, 1663−1675

systematic approach makes it impossible to reliably compare structures for different applications.

The first public database to present the "CoRE" concept is the CoRE-MOF database, which collects over 5000 structures.[14] This database has been used by many different groups for a large range of studies. The importance of such a database is that all structures are reported in a uniform format (e.g., crystallographic information files, CIFs, with *P*1 symmetry). Indeed, the possibility of easily accessing and computationally analyzing these frameworks allows the screening of thousands of different materials, comparison of their performance for specific applications, and the finding of correlations between geometric properties and performances.[16] An important point is that the CoRE-MOF database not only involves a conversion to a uniform format but also includes a minimal curation, that mainly involves the removal of solvent molecules from the pores of the framework. Solvent molecules are often present in the reported structure of MOFs, and they can be (partially) removed. The idea of the CoRE-MOF database is to arrive at a solvent-free material which serves as a reference for computational modeling. The difficulties of simply removing the solvent molecules from the structure become clear if one realizes that for some MOFs the solvent molecules are essential; without solvent the framework collapses. Another important issue is that from the X-ray data one does not have sufficient information on the position of the H atoms. These are then often added manually or using different software packages, where different protocols would put them on slightly different positions. In addition, depending of the quality of the X-ray data there can be uncertainty in the position of the atoms. The net result of all these small difference is that these screening studies are of limited use and a potential source of artifacts, due to the lack of uniformity between the investigated structures.

Contrary to MOFs, the atomic coordinates of synthesized COFs are not deposited in commonly available standardized databases. Therefore, it is already difficult to precisely estimate how many have been reported so far, and an extensive literature review is necessary to collect these structures: they must be extracted from experimental papers, checked, and converted into a common format for storing atomic coordinates. Recently, Tong et al.[17] published a collection of 280 structures, which were parsed from the experimental literature, and that we used as a starting point to build the database of frameworks of the present study.

It is interesting to note that only a very small fraction of these COFs have been reported in the popular Cambridge Structural Database[18] (CSD). To be accepted into the CSD, a structure has to be determined directly *from X-ray and neutron diffraction studies, and from powder studies using a constrained refinement*.[19] This is the case of the recently synthesized single crystal COF-300, COF-303, LZU-79, and LZU-111.[20] In most of the cases though, the poor long-range crystallinity makes it impossible to resolve the structure directly. The fact that from an experimental point of view it is difficult to obtain sufficiently reliable crystal structures requires experimental groups to increasingly rely on computational techniques to make a model of the material. These proposed structures are subsequently validated by comparison with experimental properties, such as NMR and/or the "DFT pore-size distribution" fitted from gas isotherms.[2] Part of this work was motivated by the notion that, for these types of computational structure creations from scratch, there are no standards available. Depending on the methodology, it is easy to optimize toward different metastable states; different approaches may use different bond lengths, different positions of hydrogen atoms, etc. In addition, these computations often involve geometry optimizations which are performed with very different procedures (e.g., using different *ab initio* density functional theory methods or empirical force fields) that are often poorly documented. A direct use of these structures for molecular simulations therefore becomes a source of artifacts; these seemingly small differences in the way a structure is determined can lead to significant differences in the prediction of, for example, the thermodynamic properties of these materials.

In this work, we make a first step toward a systematic approach, which allows for a standardization. The challenge we aim to address is to develop a protocol that not only allows such a standardization (i.e., all materials are optimized identically and using a reproducible protocol) but also allows for systematic improvements and additions. Our approach aims to address on one hand the needs of synthetic chemists with limited knowledge of computational methods, but that would like to use state of the art methods in an automated protocol to standardize the geometry of their just-synthesized crystal. On the other hand we provide a tool to the computational community to collectively improve these optimization protocols.

In this work, we consider a total of 324 COFs reported experimentally from 2005 to 2018 (we refer to the Materials Cloud, see the Acknowledgements, for the references to the original structures), cleaned from solvent molecules, and with no partial occupations or disorder in the structures. We propose a systematic approach to organize and store these frameworks in a curated database, to track eventual corrections to the structures or additions to the database, and to link each entry back to its original source. The atomic coordinates and the cell dimensions have been optimized using density functional theory (DFT) to give a coherent geometry to all the frameworks. In addition, we have analyzed the results from DFT calculations to spot and correct errors in the structure and to achieve a robust and automated procedure to perform the DFT modeling and the optimization. DFT-derived partial charges of the structures are computed and provided for molecular simulations.

Whereas the optimization of a single structure can be seen as a routine DFT calculation, the optimization of over 300 hundred structures, including both 2- and 3-dimensional topologies, is particularly challenging. For a single structure, one can develop a recipe that is tailor-made and optimized for a given material. For such a large number of structures, the challenge is to build a workchain that is sufficiently general and robust such that any COF structure can be optimized. Moreover, given the number of calculations, one cannot afford to manually check all the outputs and ensure that all the calculations proceeded correctly. Therefore, it is important for the workchain not only to print out just the important values that the user needs to monitor but also to automatically act when a known problem is encountered. Our aim is to shift the attention from the single calculation to the inner logic of the workchain that, performing multiple calculations and tests, can chose automatically the most efficient path that leads to the solution (in this case the optimized geometry), and which can be extended to identify more problems and improve its robustness.

DOI: 10.1021/acscentsci.9b00619
ACS Cent. Sci. 2019, 5, 1663−1675

The important additional complication is that our workchain requires the interaction of multiple codes. Here, we show the advantage of employing a supervision tool that can automatically execute these codes. For this purpose we used the Automated Interactive Infrastructure and Database for Computational Science (AiiDA),[21] that allows the representation of our simulation protocol (the "workchain") in a form of python script and its execution in an automated manner. It also gives a common language for different groups to collaborate in extending and improving the routines we designed.

A novel aspect of this work is not only the development of this workchain but also the full integration between AiiDA and the Materials Cloud. The importance of this integration becomes clear once groups start using our refined structures. For example, for 5% of the structures our workflow did not succeed in obtaining a converged structure because of inconsistencies or errors in the deposited crystal structure. These errors range from missing hydrogen to incorrect labeling of atoms, wrong cell dimensions, etc. The access to the Materials Cloud allows any user to inspect all corrections we have made on the original structure, the outcome of the geometry optimization, and the computed partial charges to be used for molecular simulations. In addition, the reader can find on the Materials Cloud the detailed workchain, i.e., the logic and the input/output files for all programs that have been used to compute and characterize the final structures. These workchains can be downloaded, allowing any researcher to reproduce our results or extend our calculations while maintaining their complete providence. We feel that it is important to create a precedent on the level of transparency and reproducibility that can be obtained.

The other important point we aim to highlight is the potential for future extensions of the entries in our COF database: the number of novel COFs that are published each year is still increasing; novel structures can be added easily, allowing different groups to collectively contribute to extending the number of structures with uniformly refined geometry.

Therefore, in this work we go beyond the "CoRE" approach, presenting the "CURATED" COF database, where the focus is on the full tracking and reproducibility of the operation that are needed to make the database (possibly) error-free, consistent, uniform, and easily upgradable.

Finally, optimizing >300 structures also gives important insights in the state of the art of COF synthesis and in the applicability of the different programs that were used in our workchain. To show an example usage of the structure we obtained, with refined geometry and partial charges, we screened the COFs in our database for their potential application in capturing carbon dioxide from coal's post-combustion gases.

BUILDING OF THE DATABASE

We aim to build a database of COFs structures reported from the literature and with refined geometry. In the following, we will refer to these structures as *CURATED* COFs, an acronym that stands for *Clean, Uniform, and Refined with Automatic Tracking from Experimental Database*. The database includes one set of CURATED COF structures that contains the original frameworks as reported in the literature, with uniform orientations of the layers and minor corrections in the case of typos, chemical errors, or solvent removal ("*CURATED from literature*", 324 COFs), and another set containing the structures that succeeded the DFT optimization ("*CURATED DFT-optimized*", 308 COFs). The whole design of the CURATED COFs repository was conceived to easily correct for possible errors and to be upgraded with more recently reported frameworks.

Structure Labels. In the CURATED database, covalent organic frameworks are labeled using 7-character strings (see Figure 1) that encode information about their structure (charge state and dimensionality) as well as their provenance (publication year, article ID, and structure ID).

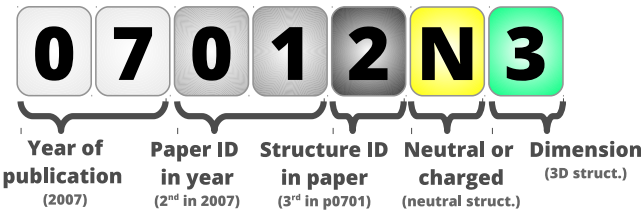


Figure 1. Structure labels used throughout this work. The structure ID is a progressive counter starting from zero and following the order of insertion into the database (not the chronological order of publication). The charge state of a framework may be neutral (N) or charged (C, needing floating counterions to balance its charge). The first four integers are therefore an index for the reference paper, e.g., p0701, where "p" stands for "paper". Note that the counting for the IDs starts form zero.

Therefore, simply looking at a structure's label one can realize in one glance that, for example, COFs 07010N3, 07011N3, 07012N3, and 07013N3 are all 3D, do not contain charge-balancing counterions, and were published in the same paper of 2007 (p0701).[3]

Data Cleaning. In this work, we argue that an important requirement of our database is that the source of the initial structures as well as all the modifications that these structures underwent during the data cleaning process need to be fully tracked. Modifications are necessary to ensure a uniform formatting of the structures and to remove errors in the structure that were detected after a visual inspection or a suspect failure in the DFT optimization.

Our database contains 324 CURATED structures from the literature. Most of them come from the previous collection done by Tong et al.; in particular, we focused on the second version of their CoRE database.[17,22] In the three versions of their database, Tong et al. collected 187,[23] 280,[17] and, only very recently, a third version of the database containing 309 COF structures.[24] Most of these structures were manually parsed from the PDF documents in the Supporting Information of the relative papers, and *some COFs reported without atomic coordinate files are constructed following the experimental information provided in the corresponding synthetic studies.*[23] As mentioned in the introduction, only a few of the known COFs are deposited in the CSD database. Despite the care of the authors in this necessary manual parsing, errors are unavoidable. Indeed, we had to apply ca. 20 corrections to the structures originally provided in the database of Tong et al. As our parsing was also done manually, and is therefore equally prone to errors, we used the Git version-control system to permanently keep track of these modifications.[25] Our repository is built from the 280 structures reported in the second version of Tong et al.'s database.[17]

DOI: 10.1021/acscentsci.9b00619
ACS Cent. Sci. 2019, 5, 1663−1675

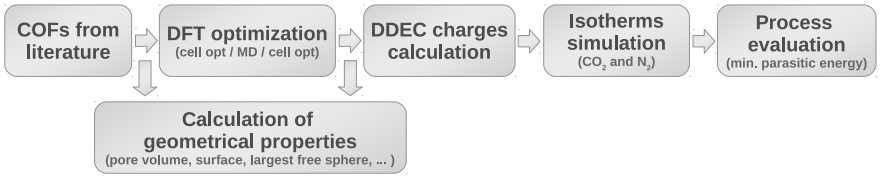


Figure 2. Block diagram of the workflow used in this work.

It is important to realize that most of the errors were found by warnings in the convergence of the DFT optimization and subsequently manually corrected. It was therefore important to monitor this optimization step and double-check the input structure of problematic runs. Problems in the initial structure are typically missing hydrogen atoms, incorrect elements, highly unfavorable protonation states, overlapping atoms (originated by the parsing of partial occupations), and incorrect unit-cell dimensions.

To extend the database, we added new structures from the literature research by querying papers that contain the keywords "covalent−organic framework" or "COF" in the title. For most of the new structures, we had to manually parse the coordinates and apply a few modifications: solvent removal, manual fix of partial occupations, unit cell's realignment for 2D COFs, or correction of typos in the original publication.

It is instructive to discuss a few examples of this detection and correction of typos. In three cases the DFT calculation failed, and from visual inspection we observed unphysical bond lengths. The problem was fixed by keeping the same fractional coordinates for the atoms and rescaling the unit cell dimension to obtain physical C−C bonds in benzene rings (see COFs 18021N3, 18121N3, and 18122N3).[26,27] We see this as an example of a reasonable but ad-hoc fix, which might be challenged in the future. In another case, a significant change in the unit cell was found after optimization, and this was caused by a manual error in the parsing from the SI: the COF with ID 120 in the database of Tong et al.[17] was incorrectly obtained by mixing the unit cell of HAT-NTBA-COF with the atomic coordinates of HAT-NTBCA-COF.[28] The mistake has been corrected in the present database, and the two structures are now reported as 17061N2 and 17060N2, respectively.

Geometry and Cell Optimization. In this database, 85% of the COFs are layered structures. To have consistent representation of all these layered COF structures, the unit cells were chosen to contain two layers for all COFs, and these layers were placed perpendicular to the *z* axis (c dimension). The stacking of two-dimensional COFs is important for the evaluation of adsorption properties, and having only one layer would not allow the geometry to find its energy minimum in a possible AB stacking during the optimization. Out of 274 2D COFs, 253 single-layer structures were found and expanded assuming perfectly eclipsed stacking. We would like to emphasize that one also could opt for 3 or more layers, but this would make the optimization even more CPU-intensive.

For the DFT optimization, it is important to mention that the unit cells of several structures in the database have more than 1000 atoms and volumes beyond 100 000 Å^3. Such large volumes and numbers of atoms pose a challenge for performing density function theory in a high-throughput context, informing our choice for the DFT code. Therefore, our choice of the CP2K package is motivated by its efficient DFT implementation, that exploits the mixed Gaussian and plane waves (GPW) method based on pseudopotentials, and the efficient orbital transformation (OT) method for the optimization of the wave function. Since the OT method is not suited for structures with a vanishing DFT gap, in the case of a final band gap smaller than 0.1 eV (a total of only 44 structures) the calculation was flagged, and different settings were used for the workchain, applying diagonalization and smearing (see the Methods section for the details).

For the unit cell and geometry optimization we designed within AiiDA an advanced workchain that ensures a more robust convergence by performing a "three-stage" optimization. First, a preliminary cell optimization with a fixed unit cell angle is run for a maximum of 20 steps. Then, 100 steps of flexible cell *ab initio* molecular dynamics (AIMD) were performed to give a "shake" to the structure and escape from metastable states. At the end, a final cell optimization is performed without constraints on the unit cell's angles. All the details on the parameters can be found in the reported workchain on the Materials Cloud and are summarized in the Methods section.

In the second step, the AIMD stage was found to be particularly necessary for the optimization of 2-dimensional COFs. We report in the Supporting Information the comparison with a standard direct cell optimization protocol, where we typically observe that, without the AIMD, the geometries become stuck in a more symmetric state with higher energy, typically the perfectly eclipsed AA configuration for 2D COFs.

In the way we designed and applied this workchain, it can almost be seen as a chemical sanity check: if a structure does not pass this stage it most likely violates the collective knowledge of quantum chemistry. This either implies DFT to have some severe inaccuracies in modeling the crystal or, what we found in most of the cases, chemical inconsistencies in the initial structure that need to be fixed manually before the structure can be optimized.

Geometric Properties and Partial Charge Assignment. To characterize the structures before and after the cell optimization, a number of geometric descriptors were evaluated, e.g., pore surface, pore volume, largest free sphere, pore size distribution, etc.

DFT-derived partial charges (DDEC protocol; DDEC, density-derived electrostatic charge) are computed for the optimized structures, utilizing the electron density that was already computed in the previous optimization stage. These partial charges are already a valuable result of our study as it allows for the modeling of the interactions with polar adsorbates such as CO_2 or H_2O.

Adsorption Calculations and Parasitic Energy. As an illustration of the use of these COFs, we screened the materials for their performance to separate CO_2 from flue gases in a case study in which these materials are use to capture carbon from a coal-fired power plant followed by geological sequestration of the captured CO_2. In our model, we mimic a temperature−

DOI: 10.1021/acscentsci.9b00619
ACS Cent. Sci. 2019, 5, 1663−1675

pressure swing process to separate CO_2 from coal post-combustion flue gases and its compression to 150 bar for underground storage. The aim is to find the materials that give the lowest parasitic energy. This parasitic energy is defined as the loss of electricity production caused by separation and compression of 1 kg of CO_2.[29] To evaluate this energy we need to predict the pure component isotherms of CO_2 and N_2 in the COFs at 300 K, as evaluated from grand canonical Monte Carlo (GCMC) simulations.[30] We have developed an AiiDA workchain to compute the minimal CO_2 parasitic energy from the optimized structures, provided with DDEC charges. The full workflow is sketched in Figure 2.

RESULTS AND DISCUSSION

Analysis of Experimental Structures. The database of experimental COFs we are considering contains 324 structures, with 310 being neutral and the remaining 14 being charged: in this second case counterions are necessary to maintain the neutrality of the system, and these charged molecules have been kept in the pore as reported in the reference paper. The counterions found in these structures are dimethylammonium, tetrafluoroborate, and single atom ions, e.g., Li^+, NA^+, F^+, Cl^+, Br^+, and I^+. Figure 3 reports the variety of elements in these

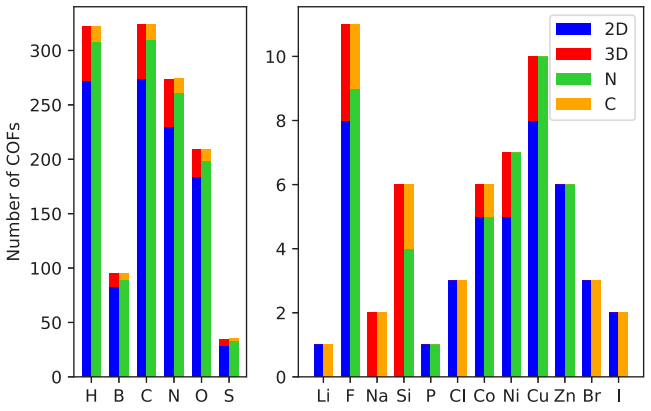


Figure 3. Distribution of elements for the 324 COFs actually present in the database. COFs are also distinguished into 3D or 2D, and neutral (N) or charged (C, that need floating counterions).

crystals, distinguishing between 2D/3D and neutral/charged frameworks. One can note that these COFs contain a number of transition metals, e.g., Co, Ni, Cu, and Zn. However, unlike MOFs, these metals are not part of the connection nodes, but they are embedded in ligands, typically in porphyrins and phthalocyanines.

An important point that we need to discuss is the stacking of 2D COFs, i.e., the arrangement of neighboring layers, sticking together because of noncovalent interactions. Reviewing the experimental literature about the synthesis and characterization of COFs,[2,4,31] we found that it is common to consider just two types of stacking: the eclipsed AA stacking, where the different layers are perfectly superimposed, and the staggered AB stacking, where there is an alternation of even and odd layers with an offset in the *xy* plane. Figure 4 shows as example these two configuration for COF-1 and COF-5.

Lukose et al. found that the minimum energy configuration for these COFs has a slightly tilted stacking.[32] Instead of the perfectly aligned AA and AB, COF-1 and COF-5 were found to have its minimum energy when the layers have an offset of

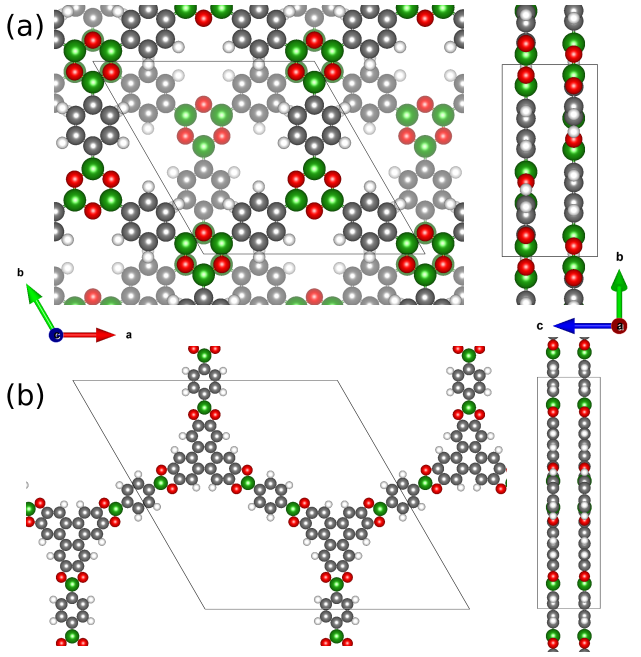


Figure 4. AA and AB stacking for (a) 05000N2 and (b) 05001N2, as reported in the experimental reference,[2] are compared. Conventional names are COF-1 and COF-5, respectively. In that report, the assessment of the presumed stacking was done by generating for both COFs the AA configuration and reasonable AB configurations, and comparing the computed XRD with the experimental PXRD.

ca. 1.4. This misalignment was shown to be hard to distinguish from the experimental PXDR pattern, and it is better to rely on computational methods. Recognizing the correct stacking is indeed a crucial aspect for many applications,[1−4] and particularly for gas adsorption, where a shifted or staggered configuration results in a shrinking of the channel diameter, possibly blocking the channels to the diffusion of gas molecules.

Analyzing the collected 2D COFs we noted that in only a few cases (21) two layers are reported in the unit cell, and it is often not clear the method that was used for the generation of that particular stacking configuration over all possible ones. For the remaining 253 structures only one layer was reported, assuming AA stacking. Therefore, by consistently considering two layers in the conventional unit cell for these 2D COFs, we allow the geometry optimization to explore both the "serrated" configuration (i.e., where odd layers are slightly shifted) and the "inclined" configurations (i.e., where there is a constant offset of the layers, resulting in tilted unit cells). Note that the serrated configuration cannot be obtained when only one layer is considered.

Cell Optimization. We performed the cell optimization for the 310 COFs that do not contain counterions in their pores (i.e., labeled as N, neutral). As we reported in the previous section, most of the problems in the DFT optimization routine were solved after a careful inspection and fixing of the initial structures. The problems that could be effectively attributed to the DFT implementation are related to only two structures: 18081N2 and 18082N2. These two both have cobalt, and they need electron smearing in the SCF: even testing different input parameters (e.g, lower mixing alpha, spin state, etc.) did not lead to a successful optimization. Considering that cobalt's pseudopotential is designed for 17 valence electrons (the

DOI: 10.1021/acscentsci.9b00619
ACS Cent. Sci. 2019, 5, 1663−1675

largest number of electrons in the set we used), this is known to be a challenging element for the SCF diagonalization, and some effort is needed to design a more effective protocol for this particular element. However, AiiDA allowed the full tracking of the problem and reporting of the issue to the CP2K developers.

Given that more than 99% of the structures in the database converged, our three-stage DFT optimization routine is suitable for high-throughput calculations. For comparison, in a previous attempt of a systematic DFT-based geometry optimization of 2612 MOFs, only 879 (33.7%) successfully converged.[33] In that case the unit cell dimension was not optimized, and no check was included for the band gap.

Figure 5 shows the evolution of DFT energy over our three-stage structure optimization workchain for two exemplary structures: 05000N2 (COF-1) and 05001N2 (COF-5).

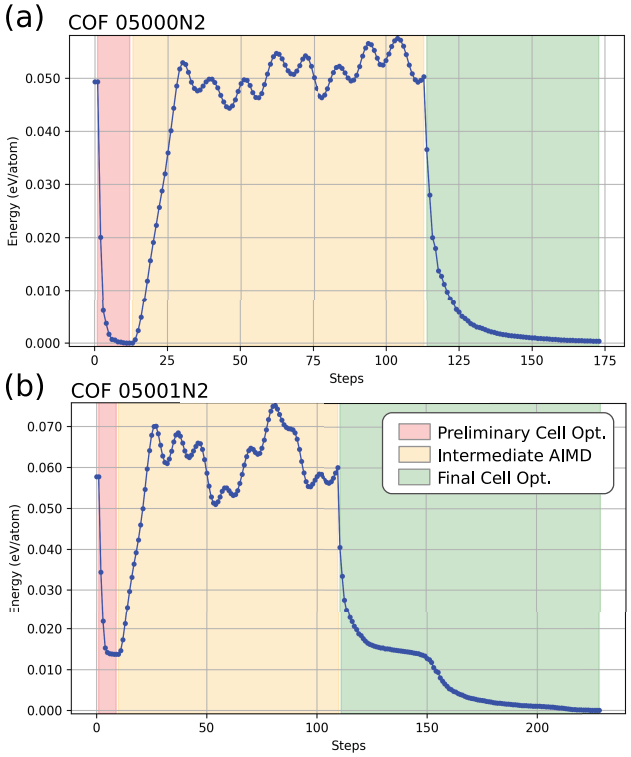


Figure 5. Energy profile during the optimization workchain for COFs (a) 05000N2 and (b) 05001N2. Conventional names are COF-1 and COF-5, and the starting configurations of the layers, as reported,[2] are staggered and eclipsed, respectively. The three colored regions of the plot correspond to the first cell optimization with constrained cell angles and 20 steps maximum (red), the NPT AIMD at 400 K/1 bar for 50 fs, i.e., 100 steps, (yellow), and the final cell optimization without constraints (green). The energy is shifted to assign a 0 value to the minimum.

As one can note for the COF 05000N2 structure,[2] we already find the minimum in the first stage, and there is no apparent need for stages two and three. For COF 05001N2, however, the MD "shaking" of the structure results in it finding a more stable minimum that is 13.7 meV/atom more stable. This final optimization results from a shifting of the perfectly eclipsed layers in the reported structure, coherently with the work of Lukose et al.[32] In the Supporting Information, we show that, without the intermediate AIMD stage, the structure becomes stuck in the first minimum, due to the high initial symmetry where layers are perfectly eclipsed.

To evaluate for the entire COF database the relative importance of the two cell optimization stages, we plotted in the histograms of Figure 6 the difference in energy in the first cell optimization (with constrained angles) and the further drop in the energy for the final cell optimization, showing also the contribution of dispersion interactions. This last is intended as the change in the energy due to the DFT-D3(BJ) terms only.

We observe that in most cases dispersion interactions play the major role in the final cell optimization. In the first cell optimization, where the typical changes in energy are larger by a factor of 10 (cf. the y axis scaling in Figure 6), we mainly observe from the optimization trajectories a rearrangement of the atomic bonds and an increase of the distance between nonbonded fragments when this is set too close in the starting geometry. If we further inspect the statistics of Figure 6, we can observe that, for the first cell optimization, in only one case there was an increase in the energy: this is related to COF 12011N2 and results from a false step of the BFGS minimizer, which did not have the time to relax back in the limit of 20 steps. For as many as 153 COFs (49.68%) the cell optimization converged within the first 20 steps (i.e., at the "preliminary" cell optimization stage), but only 19.61% of these had a negligible energy drop (<1 meV) in the final cell optimization, meaning that still for most of the cases the AIMD helped to find a lower minimum. More detailed statistics are reported in Table 1, distinguishing between two- and three-dimensional COFs.

The difference in the structures before and after the geometry and cell optimization can be appreciated from the change of the geometric properties, as plotted in Figure 7.

We can see that in many cases the DFT optimization leads to lower density, corresponding to a shrinking of the unit cell of the COF. The most evident case is for COF 18120N3 (Figure 8), where the density increases from 0.49 to 1.13 g/cm^3, with a correspondent decrease of the geometric void fraction from 0.72 to 0.39. After the optimization, this structure becomes nonporous (i.e., null N_2-accessible pore volume and surface). By inspecting the change in the atomic positions and the unit cell lengths we can notice that the structure shrunk to reach a more favorable configuration that optimizes the conformation of the ligands and their van der Waals interactions. For this COF, the measured pore volume from the nitrogen uptake is 0.36 cm^3/g,[27] i.e., well below the value computed from the reported structure (accessible N_2 probe occupiable volume equal to 0.96 cm^3/g). We can therefore expect that this structure can shrink after desolvation. For this COF the rigid crystal assumption would not hold to model the adsorption: the reported structure would lead to artificial results.

For the two-dimensional COFs, we observe in most of the cases a shift in the layers (i.e., the "inclined" configuration) that maximizes the noncovalent interactions. This results in a general reduction of the largest free sphere, as shown in Figure 7. We observe that the energy stabilization due to this shift is very low. Considering the 133 layered structures where the geometry converged at the first cell optimization stage, the energy stabilization at the end of the final cell optimization, i.e., due to the tilting, is on average −13.1 meV/atom. This is a relatively low energy, and one may therefore expect that, at finite temperature, there is no unique stacking of these

DOI: 10.1021/acscentsci.9b00619
ACS Cent. Sci. 2019, 5, 1663−1675

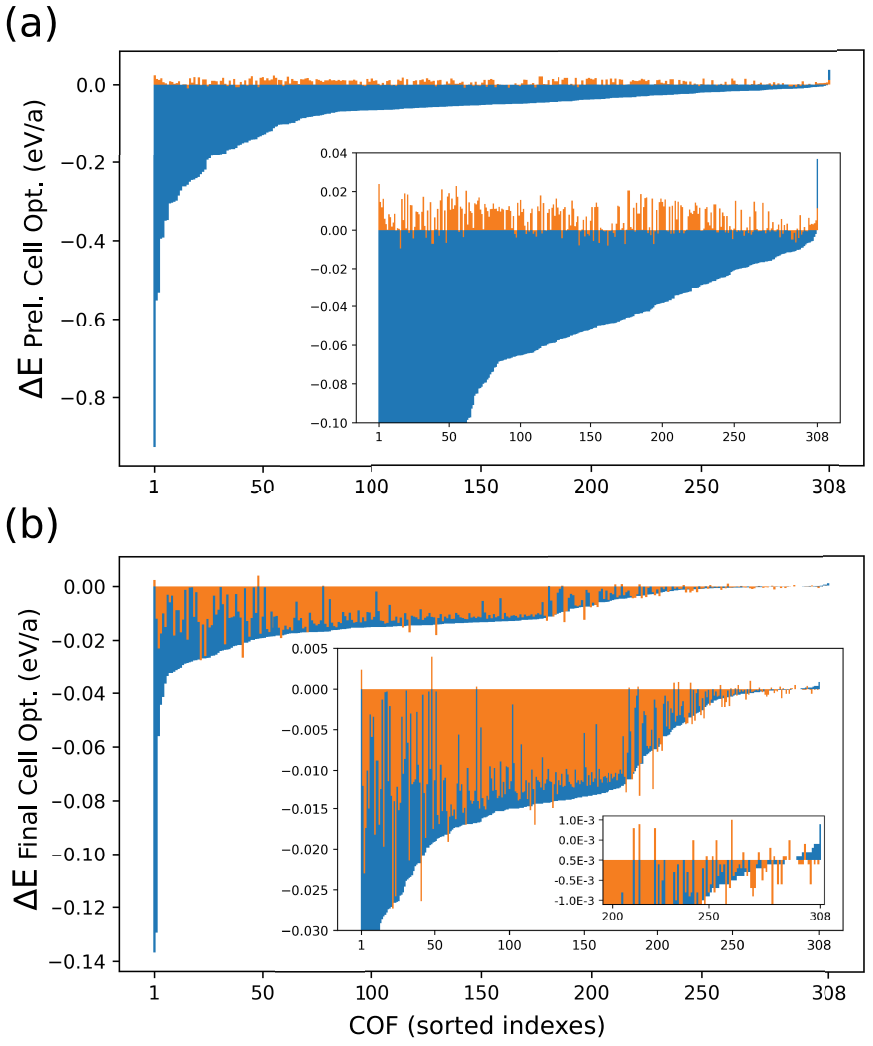


Figure 6. Comparison of the difference in the energy in the (a) first and (b) third cell optimization stages. These two stages are also referred in the text as "preliminary" and "final", respectively. Differences in energies (blue) are sorted for the 308 COFs in the histogram, showing also the relative contribution given by dispersion interactions (orange).

Table 1. Statistics on the Optimization Process of the 308 COFs that Succeeded DFT Optimization

	2D COFs	3D COFs	all COFs
number of optimized structures	261	47	308
converged in preliminary cell opt.	50.96%	42.55%	49.68%
converged, but ΔE final cell opt. > -1.0 meV/atom	12.03%	70.00%	19.61%
avg. steps in preliminary cell opt.	16.3	16.5	16.3
avg. steps in final cell opt.	150.3	113.3	144.7

materials. While our protocol successfully identifies a close-by local minimum for both 2D and 3D structures, for 2D materials, a complete screening of all possible stacking configurations and the averaging over those accessible at finite temperature need to be further explored.

CO_2 Separation Performances. The systematically optimized structures with high-quality charges, that have been obtained for 308 COFs (CURATED DFT-optimized set), allow us to model the interaction with polar molecules, e.g., considering Coulombic attractions, and evaluate the performance of these materials for gas adsorption. As an example, we investigate the use of COFs for the removal of carbon dioxide from coal's postcombustion flue gases.

To highlight the importance of high-quality input structures (i.e., DFT-optimized and with DDEC charges) for the evaluation of CO_2 adsorption, we computed also the empirical Qeq charges for the nonoptimized structures: this is a cheaper protocol that is usually exploited for high-throughput screenings. In Figure 9, we compare the CO_2 Henry coefficient and adsorption energy at infinite dilution (i.e., computed from the Widom insertion method) for the two protocols, where the absolute change in the void fraction is shown by the color of the markers. It is striking to see the impact of both the optimization and the use of accurate partial charges. One can observe at best a weak correlation between the two protocols. The implication of these results is that using the COF-structures as reported combined with a simple charge assignment scheme will potentially result in many false positives and neglecting good-performing materials. This observation is one of the main motivations of this work and illustrates the need to extend our CURATED approach to other databases. More details of the contributions that charges and geometry have to the CO_2 interactions are included in the Supporting Information.

DOI: 10.1021/acscentsci.9b00619
ACS Cent. Sci. 2019, 5, 1663−1675

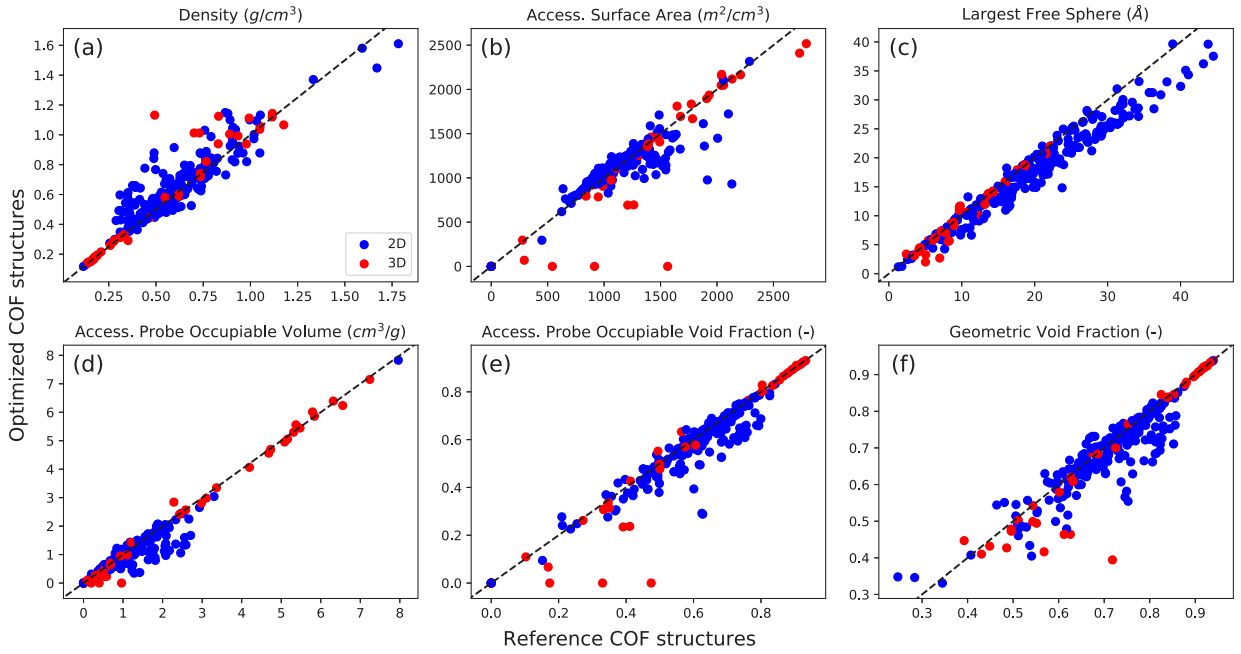


Figure 7. Parity plots that compare the geometric properties of COFs before and after the optimization: (a) density, (b) accessible surface area, (c) largest free sphere, (d) accessible probe occupiable volume, (e) accessible probe occupiable void fraction, and (f) geometric void fraction. To determine the accessibility of the pore volume and surface, a spherical probe with radius 1.86 Å (i.e., the conventional kinetic radius of nitrogen) was considered. The "geometric" void fraction is considered as all of the portion of the pore volume that does not overlap with the atoms.[34] The colors distinguish between 2-dimensional (blue) and 3-dimensional (red) structures.

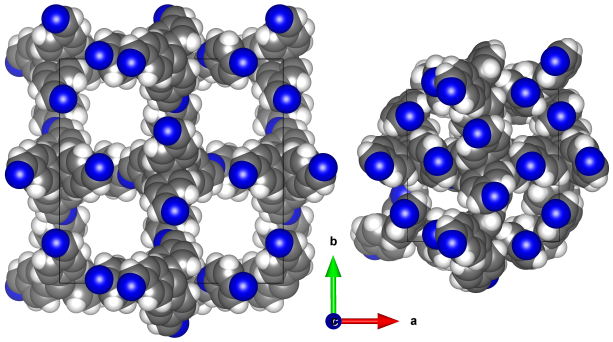


Figure 8. Structure of COF 18120N3 before (left) and after (right) the cell optimization.

With our DFT-optimized set of CURATED COFs with DDEC charges, we are now in a position to evaluate their performance for CO_2 capture and geological storage from a coal-fired power plant. For consistency, we use the same protocol that our group employed in previous studies to evaluate different classes of microporous materials.[29,35] In this protocol we assume a pressure−temperature swing adsorption process to separate CO_2 from a CO_2/N_2 mixture and subsequent compression of purified CO_2 to 150 bar, the requirement for underground storage. The key parameter to assess the performance of this process is the minimal parasitic energy, i.e., the energy that is required to separate and compress 1 kg of CO_2. However, other key parameters are also important for the evaluation: one should aim for high purity of the final high-CO_2 concentration gas, and high working capacity of the materials, to achieve the same productivity with less material or with fewer adsorption/desorption cycles.

Of all the COFs we considered, 12 COFs are nonporous for CO_2 or N_2 (according to the atomic radii definition from UFF, i.e., half of the Lennard-Jones sigma) and need to be excluded from the calculation. In addition, eight structures have inaccessible pores, which need to be blocked to prevent GCMC from growing particles in these narrow pockets. In the present study, we assumed that the COFs are rigid and therefore maintain their stacking upon adsorption. One can envision changes in the stacking upon adsorption. It would be prohibitively expensive to include framework flexibility in our screening structures, but it would be important to study these effects if one further considers top performing structures.

In Figure 10, we compared the simulated CO_2 isotherms for the first seven entries of our CURATED DFT-optimized COF database (i.e., the oldest synthesized) with experimental data.[36]

A notable difference is found for COF 05000N2 (COF-1), which we found to be nonporous under the rigid framework assumption: the small uptake that was measured is possibly the consequence of the motion of the layers that creates interstices where the CO_2 molecule can percolate and occupy the pores. Considering the uptake in the low-pressure range, we note from the isotherms a systematic overestimation of the CO_2−framework interactions, that reflects in higher Henry coefficients. Also, in the 3D COFs 07010N3 and 07011N3 the apparently higher saturation found in simulations may be related with the imperfect crystallinity of the sample or partial desolvation. A more detailed comparison is provided in the Supporting Information. Considering that we carry out a high-throughput approach, we conclude that the agreement with experiments is acceptable.

To compute the minimal parasitic energy of a material, we need as input the pure component isotherms and the heats of adsorption for CO_2 and N_2. The adsorbing bed is assumed to have a void fraction of 0.3, due to the pelletization and packing

DOI: 10.1021/acscentsci.9b00619
ACS Cent. Sci. 2019, 5, 1663−1675

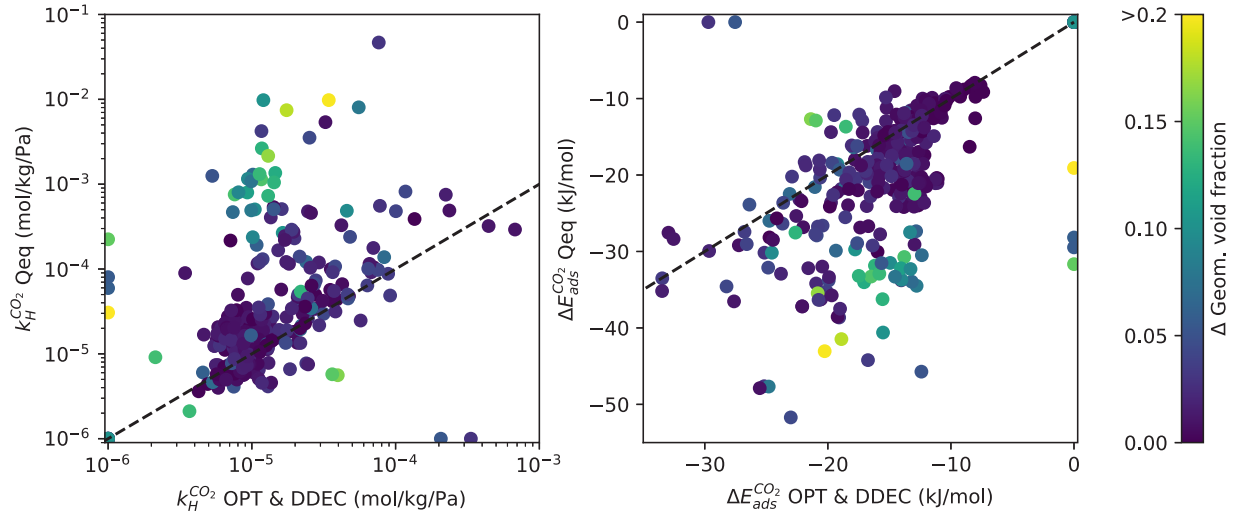


Figure 9. Comparison of the CO_2 Henry coefficient and the adsorption energy at infinite dilution for nonoptimized COFs with Qeq partial charges and DFT-optimized COFs with DDEC charges. Dashed lines show the condition of equal values for both axes. The color of the markers indicates the absolute change of the geometric void fraction after the optimization. Points that lie close to the axes (i.e., null Henry Coefficient and adsorption energy) indicate nonpermeable structures, where all the pores are inaccessible. Blocking spheres have been used in all the calculations to exclude inaccessible pores, as described in the Methods section.

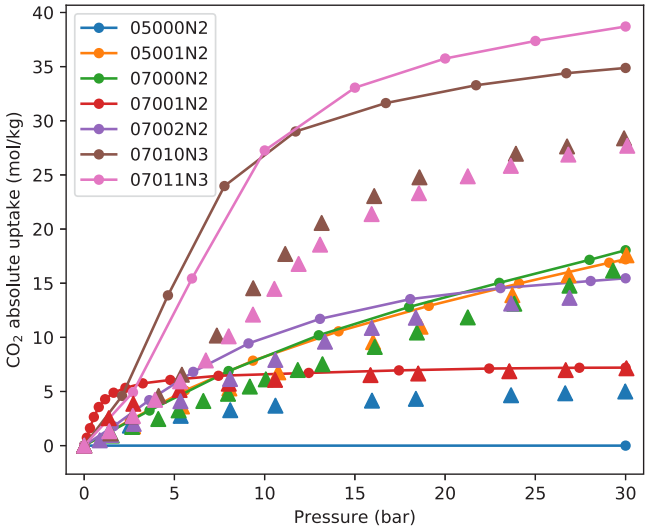


Figure 10. Comparison of simulated (lines) CO_2 isotherms with experimental ones (triangles).[36] The isotherms were obtained at 300 and 298 K, respectively. Note that the simulated uptake of COF 05000N2 is set to zero because, from geometric analysis, no accessible volume was found for CO_2 nor N_2. The conventional names of these COFs are, in the same order as in the legend, COF-1, COF-5, COF-10, COF-6, COF-8, COF-102, and COF-103.

of the crystal. The volumetric working capacity measures the amount of CO_2 that one cubic meter of bed can evacuate between adsorption and desorption.

In Figure 11 the minimal parasitic energy is plotted as a function of the Henry coefficient, the volumetric working capacity, and the final molar purity. In all three comparisons we see a strong correlation. However, for materials with low parasitic energy, i.e., below 1 MJ/kg where the process is potentially more energy efficient than ammine-based technologies, the correlation becomes less evident. One can, for example, select among the materials with low parasitic energies (that is an index for operative costs) the ones with higher working capacity, that would correspond to the need of less adsorbent (and therefore lower capital costs).

The envelope for the minimum parasitic energy versus the Henry coefficient that one can draw in Figure 11 can also be compared with the results obtained for Lin et al. for a database of more than 300 000 hypothetical zeolites and zeolitic imidazolate frameworks (ZIFs),[35] and later by Huck et al., for MOFs and porous polymer networks (PPNs).[29] In the present work, the lowest parasitic energies are obtained in the range of the Henry coefficient between 10^{-4} and 10^{-3} mol/(kg Pa). These results are coherent with the earlier findings on other materials. However, the Henry coefficient for COFs rarely exceeds 10^{-4} mol/(kg Pa), and none were found to exceed 10^{-3} mol/(kg Pa). Our effort consistently extends the "materials genome" for carbon capture and sequestration, now including also COFs. When compared with the best material found in the work of Huck et al.,[29] we observe that the performances of these COFs are still below the best performer MOF-74, for which we get 0.705 MJ per kg of CO_2, with a purity of 0.943 and a volumetric working capacity of 64.87 kg of CO_2 per cubic meter of bed. In total, 14 materials over 60 were indicated by Huck et al. to have a low parasitic energy for coal flue gas (in the range 0.7−0.85 MJ/kg), including 6 cation exchanged zeolites, 5 MOFs, and 3 porous polymers. However, the consideration that COF materials contain only light atoms, being potentially cheap, can be an important criterion of choice, making this class of materials appealing for CO_2 separation.

Since we used AiiDA for the data-tracing of the full workflow, from the reported structure to the performances evaluation, we can now easily inspect every intermediate stage, backtracking the provenance of the final results. We illustrate this for three COFs: 18041N3 having the lowest parasitic energy (0.819 MJ/kg), 13180N3 having the maximum working capacity (30.42 kg/m^3), and the test case 05001N2 (COF-5), which has a relatively high parasitic energy of 2.681 MJ/kg and therefore is not promising for this particular application. For these materials we can inspect in one glance (Figure 12) the

DOI: 10.1021/acscentsci.9b00619
ACS Cent. Sci. 2019, 5, 1663−1675

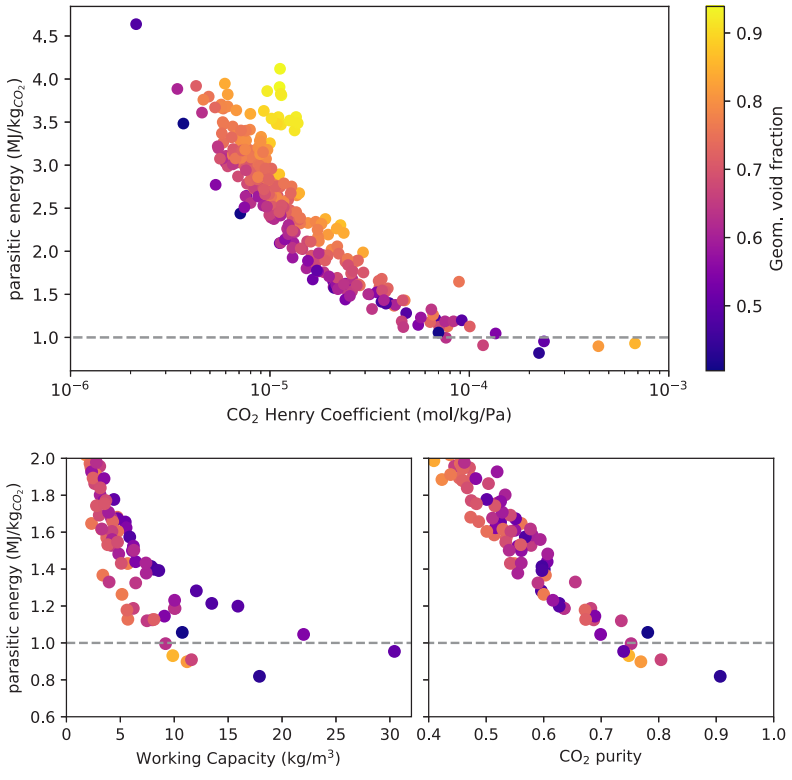


Figure 11. For a set of 296 porous and DFT-optimized COFs the CO_2 parasitic energy is plotted versus the Henry coefficient for CO_2. The dotted line gives the comparable parasitic energy of the amine-based capture process. In the lower plots the parasitic energy is plotted versus the other two main outputs from the process modeling: the CO_2 working capacity per cubic meter of adsorbent bed and the final CO_2 molar purity of the mixture.

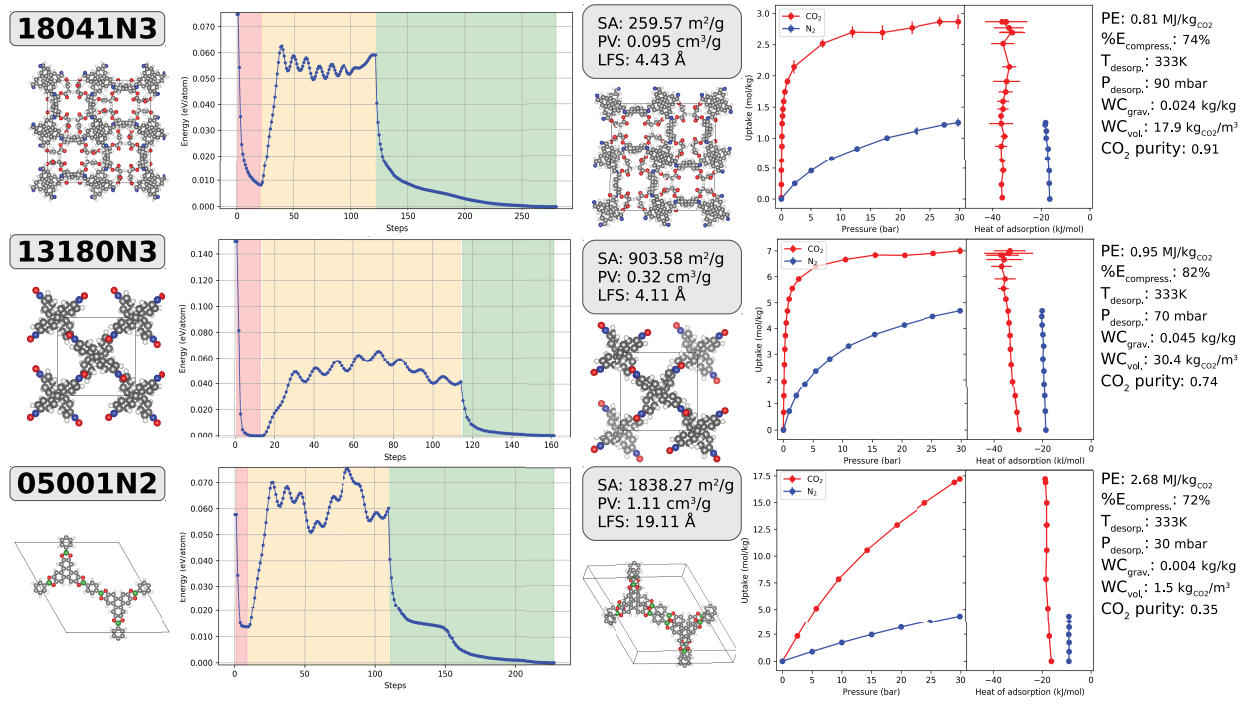


Figure 12. Results obtained from the full workflow are sketched for three COFs: 18041N3, 13180N3, and 05001N2. From the left to the right we can find the initial geometry, the energy profile during the three-stage optimization, the final geometry and its main geometric properties, the CO_2 and N_2 isotherms and heat of adsorption, and finally the results from the process modeling. "SA", "PV", and "LFS" labels indicate, respectively, the nitrogen-accessible surface area, pore volume, and the largest free sphere's diameter. "PE" and "WC" stand for the minimal parasitic energy and the CO_2 working capacity in the adsorbent, respectively.

DOI: 10.1021/acscentsci.9b00619
ACS Cent. Sci. 2019, 5, 1663−1675

energy profile during the optimization, the changes in the structures, the uptake sampling, and the details of the process simulation. The reader can access a similar visualization from the Materials Cloud, in the Discovery section (see Acknowledgments).

It is interesting to compare our results with the original articles, which report the synthesis details for these COFs. For example, it is important to confirm that the experimental structure can be desolvated. This is the case for 18041N3[37] but not for 13180N3,[38] where the COF is reported to lose its crystallinity and porosity after solvent removal. Therefore, to use this material for gas adsorption, the fact that our optimization routine converges to a stable structure suggests that this material can be stable without a solvent; this indicates that it would be worthwhile to investigate different synthesis protocols that may allow for milder conditions for the activation and possibly retain the crystallinity of the sample.

CONCLUSIONS

We presented a systematic way to optimize the structures of covalent−organic frameworks (COFs) using density functional theory. In addition, we have assessed their performance for CO_2 capture as estimated from classical simulations. The optimization revealed some substantial changes, mainly in the layers rearrangement of these materials, and finally provided a set of frameworks that have been obtained from a consistent and reproducible protocol. The Automated Interactive Infrastructure and Database for Computational Science (AiiDA) has been used to achieve the result, and this set of 308 CURATED DFT-optimized COFs is made available, with optimized structure and high-quality partial charges, for further computational investigations. The acronym stands for *Clean, Uniform, and Refined with Automatic Tracking from Experimental Database*. We plan to extend our CURATED set of COFs periodically with the most recently reported frameworks. It can serve as a reference for consulting and for molecular simulations.

We show that the computed adsorption isotherms for CO_2 are in fair agreement with experiments, allowing for a reliable ranking of these materials for carbon capture. The modular design of our workchain allows for future testing of different force fields and settings, to improve the overall match of simulation with experiments: we believe that this will be possible with the parallel effort of building a consistent repository to collect experimental adsorption measurements.

We plan to extend the same concept to other classes of materials, e.g., metal organic frameworks (MOFs) and zeolites, aiming for an extensive database for adsorption properties, useful not only for comparison but also for training machine learning models that can later be used to prescreen new materials from quickly computed geometric properties. We will need to face more complex challenges, e.g., dealing with the DFT modeling of transition metals, or tuning the standard force fields to better describe the interaction with open metal sites in MOFs. In this perspective, AiiDA can serve as a common language for the whole scientific community to collaborate, systematically improving and recombining workchains that need to model more and more complex systems.

METHODS

DFT Calculations. DFT calculations were performed using the Perdew−Burke−Ernzerhof (PBE) exchange-correlation functional[39] with DFT-D3(BJ) dispersion corrections.[40] The Quickstep code of the CP2K package was used,[41] employing GTH pseudopotentials,[42] DZVP-MOLOPT-SR contracted Gaussian basis sets, and an auxiliary plane wave basis set. The plane waves cutoff is set to 600 Ry cutoff, and these are mapped on a 4-level multigrid, with relative cutoff of 50 Ry and progression factor of 3. The orbital transformation (OT) method was used, and if a band gap <0.1 eV was found, the calculation was rerun from scratch using Broyden diagonalization and Thomas−Fermi smearing at 300 K.

For geometry optimization, we considered the atomic positions to be converged once the maximum force on the atoms dropped below 1.0 millihartree/bohr (as well as a root-mean squared value below 0.7). The threshold for the pressure is set to 100 bar for the cell optimization. For the first stage (preliminary cell optimization) the BFGS minimizer is used. For the second stage (AIMD), an NPT_F[43] simulation is performed at 400 K and 1 bar, using the CSVR thermostat[44] and the barostat from Martyna et al.[45] As for the third stage (final cell optimization), the Limited-memory BFGS (L-BFGS) minimizer is employed. Further details on the choice of the convergence threshold and minimizer can be found in the Supporting Information.

The workchain is kept efficient by always restarting the wave functions from the calculation of the previous stage, and using the Always Correct Predictor-Corrector (ASPC)[46] to have a more accurate first guess of coefficients of the wave function at every cell optimization or MD step.

Partial Charges. Density-derived electrostatic charges (DDECs)[47] are evaluated, using the software Chargemol,[48] and feeding the electron density of the optimized structures as computed from CP2K. The version 6 of the protocol, i.e., DDEC-6,[49] is used.

Qeq charges are computed with the same protocol we tested in our previous work:[50,51] the periodic-Qeq (PQeq) calculator egulp[52] was used with GMP parameters. Out of 310 neutral COFs, only one did not converge the Qeq calculation (17131N2).

Geometry-Based Descriptors. Geometric properties are evaluated using the software Zeo++ (version 0.3).[53] The software's default definition of the atomic radii was used. The accessibility of the internal pore volume and surface was assessed using a spherical probe of 1.86 Å, i.e., the conventionally used kinetic radius for nitrogen.

When computing the probe-occupiable accessible pore volume[34] and the blocking spheres[54] to be used for the molecular simulation, we adopted a different set of radii, consistently with the force field of the simulation. The frameworks' atomic diameters were set equal to Lennard-Jones sigma in UFF, and we considered spherical probes with diameters of 3.05 and 3.31 Å for CO_2 (oxygen's sigma in TraPPE) and N_2 (nitrogen's sigma in TraPPE), respectively.

Parasitic Energy Evaluation. CO_2 and N_2 isotherms are computed using the Raspa package,[55] at 300 K, within the range 0.001−30 bar. The TraPPE force field is employed to model the gases,[56] and the dispersion interactions with the framework are computed using Lorentz−Berthelot mixing rules, employing the UFF parametrization[57] for the atoms of the COFs. The pressure points for the sampling are selected using a novel protocol that we describe in the Supporting Information. The uptakes are computed from the lowest pressure, running GCMC for 1000 cycles for initialization and 10 000 for production, and restarting from the final

DOI: 10.1021/acscentsci.9b00619
ACS Cent. Sci. 2019, 5, 1663−1675

configuration, for the next pressure calculation. The heat of adsorption is computed during GCMC using the particle fluctuation method.[58] Blocking spheres are considered, to prevent the insertion of gas molecules in nonaccessible pores. COFs with null probe-occupiable accessible pore volume are considered nonporous, and for these materials the isotherms were not computed.

Finally, we used the in-house code from the work of Huck et al.,[29,59] to compute the optimal parameters for the temperature−pressure swing process, the parasitic energy, the working capacity, and the final purity of the CO_2-rich mixture.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acscentsci.9b00619.

Comparison of the cell optimization protocol with other variants, detailed description of the isotherm sampling protocol, detailed discussion on the contribution of charges and optimization to the CO_2−framework interactions, and a detailed comparison of experimental and simulated isotherms (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: berend.smit@epfl.ch.

ORCID

Daniele Ongari: 0000-0001-6197-2901
Aliaksandr V. Yakutovich: 0000-0002-7496-3075
Leopold Talirz: 0000-0002-1524-5903
Berend Smit: 0000-0003-4653-8562

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The research in this article was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement 666983, MaGic) and by the NCCR MARVEL, funded by the Swiss National Science Foundation. L.T. acknowledges the support of the swissuniversities P-5 "Materials Cloud" project (ID: 182-008). Part of the calculations were enabled by the Swiss National Supercomputing Centre (CSCS), under project ID s761. The authors are grateful to Prof. J. Hutter for the helpful discussions, to the CP2K user forum, and to the AiiDA developers team for the support. Materials Cloud Archive: Additional data are provided free of charge on the Materials Cloud Archive at https://archive.materialscloud.org/2019.0034/v1, including the AiiDA WorkChains used in this work; the AiiDA provenance graph; and the CURATED COF structures in CIF format. The archive record links to a DISCOVER section, where the structures and the main results of this work can be browsed interactively, as well as to an EXPLORE section with the interactive provenance browser (https://www.materialscloud.org/discover/curated-cofs).

■ REFERENCES

(1) Diercks, C. S.; Yaghi, O. M. The atom, the molecule, and the covalent organic frame- work. *Science* **2017**, *355*, No. eaal1585.

(2) Coté, A. P.; Benin, A. I.; Ockwig, N. W.; O'Keeffe, M.; Matzger, A. J.; Yaghi, O. M. Porous, crystalline, covalent organic frameworks. *Science* **2005**, *310*, 1166−1170.

(3) El-Kaderi, H. M.; Hunt, J. R.; Mendoza-Cortes, J. L.; Cote, A. P.; Taylor, R. E.; O'Keeffe, M.; Yaghi, O. M. Designed synthesis of 3D covalent organic frameworks. *Science* **2007**, *316*, 268−272.

(4) Feng, X.; Ding, X.; Jiang, D. Covalent organic frameworks. *Chem. Soc. Rev.* **2012**, *41*, 6010−6022.

(5) Lohse, M. S.; Bein, T. Covalent organic frameworks: structures, synthesis, and appli- cations. *Adv. Funct. Mater.* **2018**, *28*, 1705553.

(6) Olajire, A. A. Recent advances in the synthesis of covalent organic frameworks for CO2 capture. *J. CO2 Util.* **2017**, *17*, 137−161.

(7) Lin, S.; Diercks, C. S.; Zhang, Y.-B.; Kornienko, N.; Nichols, E. M.; Zhao, Y.; Paris, A. R.; Kim, D.; Yang, P.; Yaghi, O. M.; Chang, C. J. Covalent organic frame- works comprising cobalt porphyrins for catalytic CO2 reduction in water. *Science* **2015**, *349*, 1208−1213.

(8) Stegbauer, L.; Schwinghammer, K.; Lotsch, B. V. A hydrazone-based covalent organic framework for photocatalytic hydrogen production. *Chem. Sci.* **2014**, *5*, 2789−2793.

(9) Zhao, W.; Xia, L.; Liu, X. Covalent organic frameworks (COFs): perspectives of indus-trialization. *CrystEngComm* **2018**, *20*, 1613−1634.

(10) Martin, R. L.; Simon, C. M.; Smit, B.; Haranczyk, M. In silico design of porous polymer networks: high-throughput screening for methane storage materials. *J. Am. Chem. Soc.* **2014**, *136*, 5006−5022.

(11) Martin, R. L.; Simon, C. M.; Medasani, B.; Britt, D. K.; Smit, B.; Haranczyk, M. In silico design of three-dimensional porous covalent organic frameworks via known synthesis routes and commercially available species. *J. Phys. Chem. C* **2014**, *118*, 23790−23802.

(12) Mercado, R.; Fu, R.-S.; Yakutovich, A. V.; Talirz, L.; Haranczyk, M.; Smit, B. In silico design of 2D and 3D covalent organic frameworks for methane storage applications. *Chem. Mater.* **2018**, *30*, 5069−5086.

(13) Lan, Y.; Han, X.; Tong, M.; Huang, H.; Yang, Q.; Liu, D.; Zhao, X.; Zhong, C. Materials genomics methods for high-throughput construction of COFs and targeted synthesis. *Nat. Commun.* **2018**, *9*, 1−10.

(14) Chung, Y. G.; Camp, J.; Haranczyk, M.; Sikora, B. J.; Bury, W.; Krungleviciute, V.; Yildirim, T.; Farha, O. K.; Sholl, D. S.; Snurr, R. Q. Computation-ready, experimental metal-organic frameworks: a tool to enable high-throughput screening of nanoporous crystals. *Chem. Mater.* **2014**, *26*, 6185−6192.

(15) Moghadam, P. Z.; Li, A.; Wiggin, S. B.; Tao, A.; Maloney, A. G. P.; Wood, P. A.; Ward, S. C.; Fairen-Jimenez, D. Development of a Cambridge structural database sub- set: a collection of metal-organic frameworks for past, present, and future. *Chem. Mater.* **2017**, *29*, 2618−2625.

(16) Simon, C. M.; Kim, J.; Gomez-Gualdron, D. a.; Camp, J. S.; Chung, Y. G.; Martin, R. L.; Mercado, R.; Deem, M. W.; Gunter, D.; Haranczyk, M.; Sholl, D. S.; Snurr, R. Q.; Smit, B. The materials genome in action: identifying the performance limits for methane storage. *Energy Environ. Sci.* **2015**, *8*, 1190−1199.

(17) Tong, M.; Lan, Y.; Qin, Z.; Zhong, C. Computation-ready, experimental covalent organic framework for methane delivery: screening and material design. *J. Phys. Chem. C* **2018**, *122*, 13009−13016.

(18) Allen, F. H. The Cambridge structural database: a quarter of a million crystal structures and rising. *Acta Crystallogr., Sect. B: Struct. Sci.* **2002**, *58*, 380−388.

(19) The Cambridge Crystallographic Data Centre. ccdc.cam.ac.uk/Community/depositastructure/ (accessed April, 2019).

(20) Ma, T.; Kapustin, E. A.; Yin, S. X.; Liang, L.; Zhou, Z.; Niu, J.; Li, L.-H.; Wang, Y.; Su, J.; Li, J.; Wang, X.; Wang, W. D.; Wang, W.; Sun, J.; Yaghi, O. M. Single-crystal x-ray diffraction structures of covalent organic frameworks. *Science* **2018**, *361*, 48−52.

(21) Pizzi, G.; Cepellotti, A.; Sabatini, R.; Marzari, N.; Kozinsky, B. AiiDA: automated interactive infrastructure and database for computational science. *Comput. Mater. Sci.* **2016**, *111*, 218−230.

DOI: 10.1021/acscentsci.9b00619
ACS Cent. Sci. 2019, 5, 1663−1675

(22) GitHub—core-cof/CoRE-COF-Database. github.com/core-cof/CoRE-COF-Database (accessed November, 2018).
(23) Tong, M.; Lan, Y.; Yang, Q.; Zhong, C. Exploring the structure-property relationships of covalent organic frameworks for noble gas separations. *Chem. Eng. Sci.* **2017**, *168*, 456−464.
(24) Yan, T.; Lan, Y.; Tong, M.; Zhong, C. Screening and design of covalent organic framework membranes for CO2/CH4 separation. *ACS Sustainable Chem. Eng.* **2019**, *7*, 1220.
(25) danieleongari/CURATED-COFs. http://doi.org/10.5281/zenodo.3352138.
(26) Wang, C.; Wang, Y.; Ge, R.; Song, X.; Xing, X.; Jiang, Q.; Lu, H.; Hao, C.; Guo, X.; Gao, Y.; Jiang, D. A 3D covalent organic framework with exceptionally high iodine capture capability. *Chem. - Eur. J.* **2018**, *24*, 585−589.
(27) Guan, X.; Ma, Y.; Li, H.; Yusran, Y.; Xue, M.; Fang, Q.; Yan, Y.; Valtchev, V.; Qiu, S. Fast, ambient temperature and pressure ionothermal synthesis of three-dimensional covalent organic frameworks. *J. Am. Chem. Soc.* **2018**, *140*, 4494−4498.
(28) Xu, S.-q.; Liang, R.-r.; Zhan, T.-g.; Qi, Q.-y.; Zhao, X. Construction of 2D covalent organic frameworks by taking advantage of the variable orientation of imine bonds. *Chem. Commun.* **2017**, *53*, 2431−2434.
(29) Huck, J. M.; Lin, L.-C.; Berger, A. H.; Shahrak, M. N.; Martin, R. L.; Bhown, A. S.; Haranczyk, M.; Reuter, K.; Smit, B. Evaluating different classes of porous materials for carbon capture. *Energy Environ. Sci.* **2014**, *7*, 4132−4146.
(30) Frenkel, D.; Smit, B. Understanding molecular simulation: from algorithms to applications. *Physics Today* **1997**, *50*, 66.
(31) Côté, A. P.; El-Kaderi, H. M.; Furukawa, H.; Hunt, J. R.; Yaghi, O. M. Reticular synthesis of microporous and mesoporous 2D covalent organic frameworks. *J. Am. Chem. Soc.* **2007**, *129*, 12914−12915.
(32) Lukose, B.; Kuc, A.; Heine, T. The structure of layered covalent-organic frameworks. *Chem. - Eur. J.* **2011**, *17*, 2388−2392.
(33) Nazarian, D.; Camp, J. S.; Chung, Y. G.; Snurr, R. Q.; Sholl, D. S. Large-scale refinement of metal-organic framework structures using density Functional Theory. *Chem. Mater.* **2017**, *29*, 2521−2528.
(34) Ongari, D.; Boyd, P. G.; Barthel, S.; Witman, M.; Haranczyk, M.; Smit, B. Accurate characterization of the pore volume in microporous crystalline materials. *Langmuir* **2017**, *33*, 14529−14538.
(35) Lin, L.-C.; Berger, A. H.; Martin, R. L.; Kim, J.; Swisher, J. a.; Jariwala, K.; Rycroft, C. H.; Bhown, A. S.; Deem, M. W.; Haranczyk, M.; Smit, B. In silico screening of carbon-capture materials. *Nat. Mater.* **2012**, *11*, 633−641.
(36) Furukawa, H.; Yaghi, O. M. Storage of hydrogen, methane, and carbon dioxide in highly porous covalent organic frameworks for clean energy applications. *J. Am. Chem. Soc.* **2009**, *131*, 8875−8883.
(37) Lu, Q.; Ma, Y.; Li, H.; Guan, X.; Yusran, Y.; Xue, M.; Fang, Q.; Yan, Y.; Qiu, S.; Valtchev, V. Postsynthetic functionalization of three-dimensional covalent organic frameworks for selective extraction of lanthanide ions. *Angew. Chem., Int. Ed.* **2018**, *57*, 6042−6048.
(38) Beaudoin, D.; Maris, T.; Wuest, J. D. Constructing monocrystalline covalent organic networks by polymerization. *Nat. Chem.* **2013**, *5*, 830−834.
(39) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **1996**, *77*, 3865−3868.
(40) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* **2011**, *32*, 1456−1465.
(41) VandeVondele, J.; Krack, M.; Mohamed, F.; Parrinello, M.; Chassaing, T.; Hutter, J. Quickstep: fast and accurate density functional calculations using a mixed Gaussian and plane waves approach. *Comput. Phys. Commun.* **2005**, *167*, 103−128.
(42) Goedecker, S.; Teter, M. Separable dual-space Gaussian pseudopotentials. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1996**, *54*, 1703−1710.
(43) CP2K user manual. manual.cp2k.org/cp2k-5_1-branch/CP2K_INPUT/MOTION/MD.html (accessed April, 2019).
(44) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*, No. 014101.
(45) Martyna, G. J.; Tobias, D. J.; Klein, M. L. Constant-pressure molecular-dynamics algorithms. *J. Chem. Phys.* **1994**, *101*, 4177−4189.
(46) Kolafa, J. Time-reversible always stable predictor-corrector method for molecular dynamics of polarizable molecules. *J. Comput. Chem.* **2004**, *25*, 335−342.
(47) Manz, T. A.; Sholl, D. S. Chemically meaningful atomic charges that reproduce the electrostatic potential in periodic and nonperiodic materials. *J. Chem. Theory Comput.* **2010**, *6*, 2455−2468.
(48) chargemol, release 09_26_2017. sourceforge.net/projects/ddec/ (accessed August, 2018).
(49) Manz, T. A.; Limas, N. G. Introducing DDEC6 atomic population analysis: part 1. charge partitioning theory and methodology. *RSC Adv.* **2016**, *6*, 47771−47801.
(50) Rappe, A. K.; Goddard, W. A. Charge equilibration for molecular dynamics simulations. *J. Phys. Chem.* **1991**, *95*, 3358−3363.
(51) Ongari, D.; Boyd, P. G.; Kadioglu, O.; Mace, A. K.; Keskin, S.; Smit, B. Evaluating charge equilibration methods to generate electrostatic fields in nanoporous materials. *J. Chem. Theory Comput.* **2019**, *15*, 382−401.
(52) danieleongari/egulp. http://doi.org/10.5281/zenodo.3352156.
(53) Willems, T. F.; Rycroft, C. H.; Kazi, M.; Meza, J. C.; Haranczyk, M. Algorithms and tools for high-throughput geometry-based analysis of crystalline porous materials. *Microporous Mesoporous Mater.* **2012**, *149*, 134−141.
(54) Haranczyk, M.; Sethian, J. A. Automatic structure analysis in high-throughput characterization of porous materials. *J. Chem. Theory Comput.* **2010**, *6*, 3472−3480.
(55) Dubbeldam, D.; Calero, S.; Ellis, D. E.; Snurr, R. Q. RASPA: molecular simulation software for adsorption and diffusion in flexible nanoporous materials. *Mol. Simul.* **2016**, *42*, 81−101.
(56) Potoff, J. J.; Siepmann, J. I. Vapor-liquid equilibria of mixtures containing alkanes, carbon dioxide, and nitrogen. *AIChE J.* **2001**, *47*, 1676−1682.
(57) Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. a.; Skiff, W. M. UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *J. Am. Chem. Soc.* **1992**, *114*, 10024−10035.
(58) Dubbeldam, D.; Torres-Knoop, A.; Walton, K. S. On the inner workings of monte carlo codes. *Mol. Simul.* **2013**, *39*, 1253−1292.
(59) danieleongari/calc_pe. http://doi.org/10.5281/zenodo.3352154.

DOI: 10.1021/acscentsci.9b00619
ACS Cent. Sci. 2019, 5, 1663−1675