UC Berkeley
UC Berkeley Previously Published Works

Title
Too Many Materials and Too Many Applications: An Experimental Problem Waiting for a Computational Solution

Permalink
https://escholarship.org/uc/item/7522j5c5

Journal
ACS Central Science, 6(11)

ISSN
2374-7943

Authors
Ongari, Daniele
Talirz, Leopold
Smit, Berend

Publication Date
2020-11-25

DOI
10.1021/acscentsci.0c00988

Peer reviewed

Too Many Materials and Too Many Applications: An Experimental Problem Waiting for a Computational Solution

Daniele Ongari, Leopold Talirz, and Berend Smit*

Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: Finding the best material for a specific application is the ultimate goal of materials discovery. However, there is also the reverse problem: when experimental groups discover a new material, they would like to know all the possible applications this material would be promising for. Computational modeling can aim to fulfill this expectation, thanks to the sustained growth of computing power and the collective engagement of the scientific community in developing more efficient and accurate workflows for predicting materials' performances. We discuss the impact that reproducibility and automation of the modeling protocols have on the field of gas adsorption in nanoporous crystals. We envision a platform that combines these tools and enables effective matching between promising materials and industrial applications.
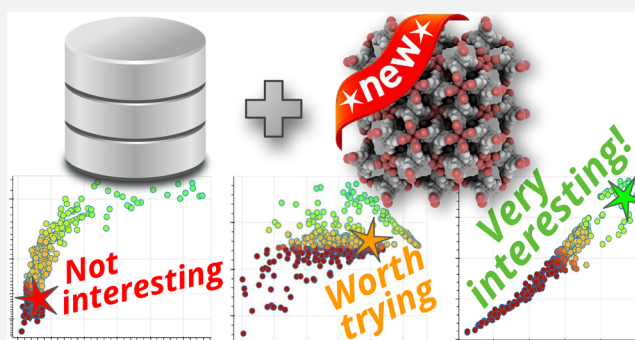
INTRODUCTION

In this Outlook, we are indulging in a luxury problem: too many materials with too many possible applications. In the field of nanoporous materials, more than 100 000 metal−organic frameworks (MOFs) have been reported (Figure 1), with large design spaces also guaranteed for covalent organic frameworks (COFs), zeolites, porous organic cages, etc. At the same time, the range of applications of these materials is expanding from gas separation[1] and gas storage[2] to fields such as catalysis[3,4] and sensing.[5,6] A new material is often designed and tested for one specific application; testing it for a wide range of applications may exceed the expertise, time resources, and/or infrastructure of the research group synthesizing the material. Conversely, research groups focused on the application side can no longer afford to test all materials of possible interest.

Let us illustrate this point with a few examples. Al-PMOF was first synthesized for its photocatalytic activity but was later discovered to be promising for the separation of CO_2 from wet flue gases.[13] MOF SBMOF-1 was synthesized to capture CO_2 but turned out to be an excellent material for separating Xe from Kr.[14] UMCM-152 was first reported in 2010 and tested for H_2 adsorption[15] but was rediscovered as a record-breaking material for oxygen storage eight years later.[16] In these and further examples,[17] computational screening studies discovered the potential of existing materials for new applications. Since the number of available materials is so large (and becomes even larger when including materials generated *in silico*[18]), computational modeling is at present the only feasible screening method.

A typical computational screening study aims to rank a set of materials for a given application: the first step is to determine key performance indicators (KPIs) and the ranking criteria for the comparison of different materials. KPIs are typically related to material properties, such as the electronic band structure (for optical or electronic KPIs), or adsorption isotherms (for KPIs related to gas storage or separations). While some properties, such as stability, remain difficult to predict from first-principles, density functional theory (DFT) calculations provide access to a wide range of properties of MOFs, including the band gap and band structure,[19] mechanical properties,[20,21] and catalytic properties.[22,23] DFT also allows us to make accurate predictions of the interactions of guest molecules inside the pores.[24,25] In addition, classical molecular simulations enable the computation of thermodynamic and transport properties of these guest molecules.[26,27]

When screening studies involve large numbers of materials and multiple steps, inspecting each simulation step manually is no longer feasible. It becomes necessary to develop robust workflows that provide automatic fallback routes for common errors, signal potential issues to the user (e.g., related to the atomic structures provided as input), manage calculations with
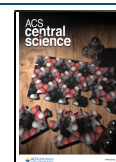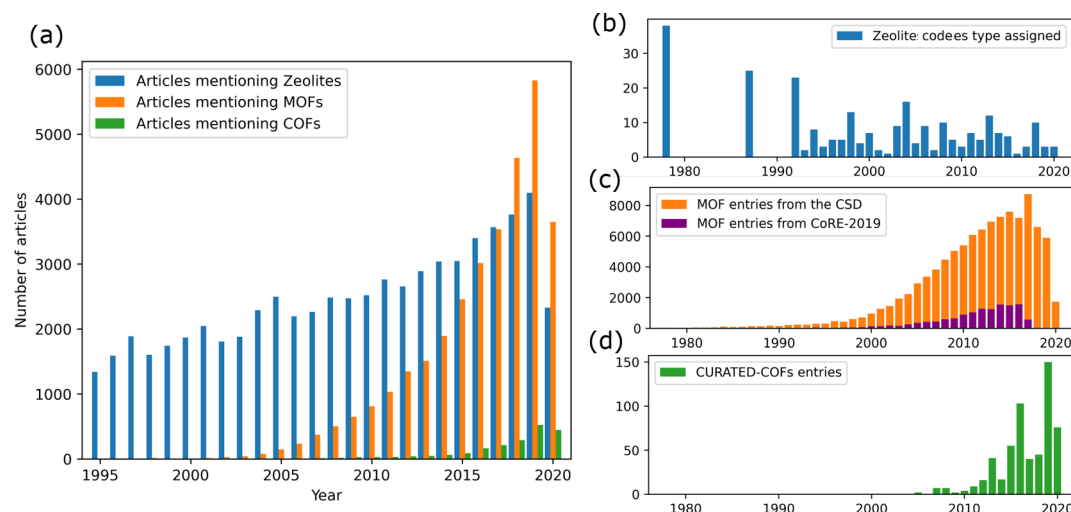
Figure 1. (a) Papers mentioning "Zeolite", "Metal Organic Framework", and "Covalent Organic Framework" in the title or the abstract, as parsed from Scopus in July 2020.[7] The right column collects histograms for the deposition of materials in publicly available databases. (b) Zeolite code types by year of assignment, from the database of the International Zeolite Association (IZA).[8] (c) MOF-subset of the Cambridge Structural Database (CSD, May 20 update) by year of publication (orange).[9] MOFs in the CoRE-2019 "All solvent Removed" (ASR) subset (purple) are selected from the CSD release of November 2017 with criteria such as three-dimensionality of the framework and permeability to small molecules.[10] (d) COFs in the CURATED-COFs database (June 20 update), by year of publication.[11,12]
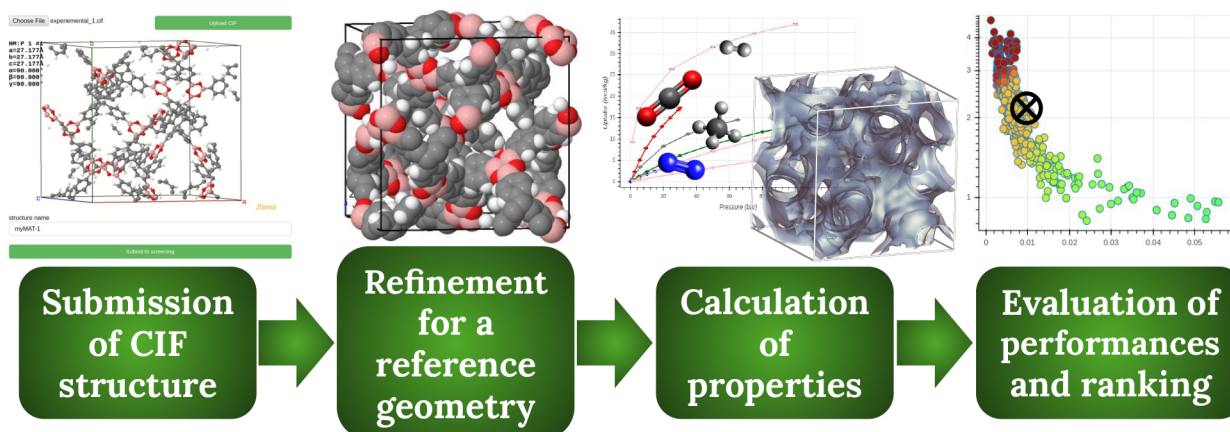

Figure 2. Scheme of exemplary workflow. The user starts by uploading the atomic structure of a crystalline materials in the CIF format, which triggers the refinement of the atomic positions, the computation of pore geometry, and thermodynamic and transport properties. Finally, its performance for specific applications is evaluated, and the material is ranked versus other candidates.

a variety of software packages, and allow for easy retrieval of the resulting data. The development of automated workflows is a major effort, and the community would benefit enormously from open-source projects that allow the sharing of such tools.

> The development of automated workflows is a major effort, and the community would benefit enormously from open-source projects that allow the sharing of such tools.

We can envision taking this approach even one step further: What if the workflow infrastructure was made available for experimental groups to use? They could upload the crystal structure of a new nanoporous material (MOF, COF, or zeolite) and—with close to zero effort—obtain calculations of

thermodynamic and transport properties for a range of guest molecules, as well as predictions of how this new material would perform in screening studies previously published in the literature (Figure 2), or if an experimental group develops a novel separation process, having a database of thermodynamic data would allow this group to identify top performing materials for their applications. Finally, if a computational group improves an existing force field or protocol for a specific class of materials, the updated workflow could be available whenever a material of that class is uploaded.

In this Outlook, we introduce the challenge of matching nanoporous materials and applications and argue that an open-science platform is a promising and practical approach to address this need. Looking back at almost two decades of progress in the study and modeling of MOFs and COFs, we observe that it is often difficult to build upon and extend the output of the computational community. While many authors nowadays make the source code underlying their screening studies openly available, the barrier to putting them into action

remains high for anyone who is not a computational expert. We suggest that most of the puzzle pieces are already there and, if put together, they have the potential to dramatically increase the impact of computational studies, both by activating new classes of users and by encouraging collaboration in the design of computational modules.

> Most of the puzzle pieces are already there and, if put together, they have the potential to dramatically increase the impact of computational studies.

■ DATABASES OF NANOPOROUS MATERIALS

Curated Structures from Experimental Syntheses. Computational screening studies rely on large databases of materials. A first step is to collect all the reported structures in a consistent format: today, the Crystallographic Information File (CIF) format is the most common one. The minimal information needed to proceed with a computational screening study is the dimensions of the unit cell and the elements and coordinates of the atoms that compose the framework. This information is obtained ideally from single crystal X-ray diffraction studies. When obtaining single crystals is not possible, one can rely on powder X-ray diffraction or other indirect measurements instead.[28]

In the MOF community, it is standard practice to publish new structures in the Cambridge Structural Database (CSD)[29] and to report the reference code assigned to the entry in the article. With its more than one million entries (102 508 of which were recognized as MOFs until May 2020),[9] the CSD is the oldest and largest data set of organic and metal−organic crystals.[30]

Unfortunately, many of the reported structures in the CSD are not suitable for a computational screening study out of the box. From X-ray diffraction data, it is difficult to locate hydrogen. If the material is charged, locating the counterions can also be challenging. For porous MOFs, the crystal structure is often determined with solvent molecules present inside the framework, while many practical applications require *activating* the material by removing the solvent molecules. In addition, disorder in the material is often indicated via partial occupancies, which need to be resolved to unique structures (in most of the cases manually) before using the structures as input for simulations.

The group around D. Sholl pioneered the extraction of MOFs from the CSD for a computational screening purpose. In two studies from 2012 (seeking materials for kinetic separation of noble gases[31] and CO_2/N_2[32]) they distilled a set of 3432 and 1163 MOFs, respectively, from which they discarded entries with atomic disorder, and they algorithmically removed solvent molecules. One year later, Siegel et al.[33] targeted hydrogen storage and identified ca. 38 800 crystals from the CSD as MOFs but had to exclude ca. 16 000 problematic structures due to missing H, disorder, etc. For these first studies, the final database of filtered and curated CIFs was only accessible upon request to the authors. In 2014, Chung et al. created a set of 5109 "Computation-Ready Experimental" (CoRE) structures, selected to be three-dimensional, porous (i.e., with a pore limiting diameter >2.4

Å), and fully desolvated, and made the database openly available for download.[34] Recently, this database was updated to include 14 142 MOFs, 546 of which were collected from sources other than the CSD.[10] This update also included a version of the database where solvent molecules coordinated to metal sites were not removed (i.e., the MOFs were not "computationally activated"). In two separate projects, Nazarian et al. used DFT to provide partial charges[35] and geometry-optimized structures[36] for the first set of CoRE-MOFs. Only 502 (i.e., ca. 10%) passed both refinements. The database with partial charges was used by other groups in several screening studies,[16,37−39] demonstrating the impact of providing an open-access, curated, and extensive database of structures to the computational community.

Ideally, for every new MOF structure deposited in the CSD, a computation-ready structure would be generated as well. At present, however, there is no standardized protocol for these steps—e.g., removal of solvents, addition of missing hydrogen atoms, resolution of partial occupancies, correction of atomic overlaps, and structural distortion—and as a result, each group may generate slightly different structures that make it difficult to compare predicted properties.[40,41] This cleaning procedure can be seen as a continuous process where more and more checks and fixes are added in a collaborative effort.[42,43] For the platform we envision, it would be highly desirable to eventually arrive at an internally consistent and extensive database of nanoporous structures that are "ready" for molecular simulations or electronic structure calculations, and made available in a way that satisfies the FAIR data principles: Findable, Accessible, Interoperable, and Reusable.[44,45]

Extending this curation to different classes of materials, such as COFs, inevitably results in new considerations and challenges. For MOFs, the CSD imposes quality controls on the accuracy of the crystal structure. COFs typically have short-range crystallinity but long-range disorder, preventing the refinement of the crystal structure directly from X-ray diffraction data[46] and thus inclusion in the CSD. As a consequence, experimental groups develop their own protocols to generate the crystal structures reported with their publications. These difficulties motivated us to create a database that combines the advantages of both the CoRE and CSD protocols and provides a high level of transparency and consistency in the refinements including cell optimization and the calculation of partial charges. Branching off from the second version of the CoRE-COF database by Tong et al.,[47] we extended the database to 574 COFs in the June 2020 update. The relevant literature is monitored by the @COF_Papers Twitter bot, and structures in CIF format are collected in a public repository, where researchers can suggest new additions or report errors.[12] All modifications, corrections, and additions are tracked by the Git version control system. Moreover, an automated routine, orchestrated by the AiiDA workflow manager,[48,49] computes the DFT-optimized structures and partial charges following a published protocol.[11] The results are made available periodically in the CURATED COF database hosted on the Materials Cloud open science platform.[50] A recent independent study compared the gas-separation performances as computed from the structures of this database and the original COF structures, highlighting the importance of the curation process.[51]

Hypothetical Structures. In addition to the databases of experimentally determined structures, there is an even larger number of structures generated *in silico*, which are further

candidates for computational screenings. Replacing the experimental synthesis of new materials with computational algorithms increases the number of atomic structures that can be assembled (zeolites, MOFs, and COFs) by orders of magnitude. To give an idea of how large these databases can get, we just mention two very recent works, which reported 325 000 hypothetical MOFs[13] and 471 990 hypothetical COFs.[52]

As the growth of the number of hypothetical structures even outpaces the continued increase of computational power, brute force screening will become increasingly unfeasible. A promising alternative is to select only a modest subset of most *diverse* structures to perform accurate calculations (comparable to those on experimental structures), train machine learning methods to capture the structure—performance relations, and use them to extend the screening to the remaining materials.[53] Indeed, the key to the success of this approach is to find effective metrics for the "diversity" between structures in the context of a particular application.[54]

■ COMPUTATION OF MATERIALS' PROPERTIES

Gas Adsorption Properties. Once one has a set of computation-ready structures, one can start computing the properties that are relevant for the application(s) of interest— here, gas adsorption in nanoporous materials. If interested in comparing properties and performances among many materials (and updating this ranking over time), one needs to pay particular attention to the consistency of results. Consistency means, for example, applying the same protocol to curate the crystals' atomic structure, to estimate partial charges, to exclude inaccessible pores, etc. in order to enable the comparison of the final results. This includes both choices of the model, such as the DFT functional or the force field (UFF, DREIDING, TraPPE, etc.), as well as secondary parameters, such as the choice of the DFT basis set or whether to use tail-corrections in grand canonical Monte Carlo (GCMC) adsorption calculations.[55]

A first step may involve a relaxation of the atomic positions using force fields, semiempirical methods, or DFT in order to ensure that the atomic structure is consistent with the computational method employed. This step can also help identify mistakes in input structures and take the effect of solvent removal on the framework into account.

Gas—framework interactions are often evaluated using GCMC insertion techniques with classical force fields.[17] In this approach, the Coulomb interaction is modeled by partial charges, which are tabulated for popular gas molecules but need to be computed for the framework. If DFT was used for the geometry relaxation, partial charges can be computed at negligible extra cost from the electrostatic potential, usually preferring protocols that aim at reproducing the electrostatic potential (e.g., REPEAT, DDEC) over the others (e.g., Mulliken, Hirshfeld, Bader).[56,57] Alternatively, cheaper charge equilibration methods can be used, but with extreme care (see ref 37). The partial charges on the framework atoms need to be combined with a force field to describe the interaction between the gas molecule and the framework. Many studies opt for off-the-shelf parameters for the dispersion interaction, such as DREIDING[58] or UFF[59] for the framework and TraPPE[60] for the adsorbate. Steps outside the original design space of existing force fields or modifications of their parameters need to be carefully validated in order to ensure that the behavior of the gas molecules in the pores is reproduced with sufficient accuracy (e.g., adsorption isotherms, heats of adsorption, etc.).

As one moves to larger numbers of structures and more complex workflows, it becomes increasingly challenging to manage the calculations and to provide all information required to reproduce a particular result. This is where workflow managers can come to the rescue, and numerous open-source infrastructures are available for orchestrating computational chemistry codes with advanced logic,[61] such as AiiDA,[48,49] FireWorks,[62] AFLOW,[63] or signac.[64]

Open Challenges. Once we obtain a reliable force field for molecule—framework interactions, we are still left with a number of open challenges. One challenge is the modeling of defects: as high-throughput computations typically assume perfect crystals, they will not capture properties that are dominated by crystal defects present in the experimental material. Another challenge is the modeling of framework mechanics upon adsorption: most screening studies assume a rigid framework. For many structures, this approximation is reasonable, but some materials are known to display structural changes upon gas adsorption, which can affect performance in relevant applications.[65] Assuming the structure to be rigid may lead to incorrect identification of pore accessibility for gas molecules. Algorithms based on geometrical assessment of channel diameters can easily recognize nonaccessible pores and exclude them from the adsorption calculation.[66] However, it is less trivial to routinely identify those cases where a small rotation of the ligands can allow the gas molecule to permeate (such as in the well-studied case of ZIF-8[67,68]).

There are other material properties relevant to the process modeling of gas-related applications that can be evaluated from the unit cell, such as gas diffusion,[69] heat capacity,[70,71] mechanical stability,[72,73] and chemical stability. The studies cited above propose computational protocols for investigating these properties, which might be extended for high-throughput screenings. Combining all these properties in the same screening platform would allow the filtering out of structures that are unstable or show poor thermal or molecular diffusion and provide more information to the process model.

Moving beyond the field of gas adsorption brings yet more properties into focus, such as more accurate electronic properties for applications in sensing, semiconductors, and photocatalysis,[19,74] which put increased emphasis on the choice of the electronic structure method and are beyond the scope of this Outlook.

■ RANKING MATERIALS

Accurately predicting material properties is the aim of molecular simulations, but it represents only half of the story: our ultimate goal is to rank materials for a given application, based on key performance indicators (KPIs). In the following, we discuss the KPIs for two important applications of nanoporous materials: hydrogen storage[33,75−79] and CO_2 separation from nitrogen.[11,40,80−83]

> Our ultimate goal is to rank materials for a given application, based on key performance indicators

For H_2 storage, the main KPI is the deliverable (or "working") capacity, i.e., the difference in gas uptake between the *loading conditions* at higher pressure and/or lower temperature, and the *discharging conditions* at lower pressure and/or higher temperature. Therefore, the evaluation of hydrogen storage performance requires calculations at only these two conditions of temperature and pressure, and molecular simulations have been shown to be feasible for screening more than half a million structures.[78,79] Similar considerations are also valid for the evaluation of natural gas deliverable capacity.[34,47,84,85] Other important KPIs may focus on the diffusion of gas and heat inside the framework, in order to enable fast loading/discharge and heat dissipation.

For CO_2 capture, finding KPIs for the ranking is more complex. In 2012, our group developed a simplified thermodynamic model for carbon capture and sequestration (CCS) involving a temperature−pressure-swing adsorption process, considering inlet gases from a coal-fired power plant (14:86 ratio of CO_2:N_2), a natural-gas-fired power plant (4% CO_2), or air (400 ppm of CO_2). This model was used to evaluate different classes of nanoporous materials,[82,86] and we recently expanded the study to COFs.[11] Two KPIs were identified: The "parasitic energy" is defined as the energy needed to separate 1 kg of CO_2 and compress the purified gas to 150 bar for underground storage. The parasitic energy can be taken as a measure of the operating cost of the separation (OPEX). The working capacity, on the other hand, determines the amount of adsorbent material needed and thus the capital cost (CAPEX). The study showed that the minimal parasitic energy can be related to an optimal value for the Henry coefficient of CO_2 around 10^{-3} mol/(kg Pa) (at 300 K) for power plants—stronger affinity between CO_2 and the framework would result in higher energy needed for the regeneration of the adsorbent. For direct-air capture, the optimal value lies above 10^{-1} mol/(kg Pa), and chemisorption appears to be a more promising solution. Recently, simulations have been coupled with more advanced models of the pressure-swing process,[87−89] indicating subtle relations between the properties of the material and their performance in the process. In particular, the often overlooked nitrogen isotherm was identified as a key indicator.

The case of CO_2 separation highlights the importance of connecting materials' properties to process modeling: on one hand, the process modelers need to be aware of the uncertainties in material property predictions and how they propagate through their model. If small perturbations in the inputs alter the outcome significantly, this "butterfly effect" will compromise the reliability of the final ranking. On the other hand, the molecular simulation community should focus its efforts on improving the predictive accuracy for those properties that are shown to have the largest influence on the process models. In this context, modular workflows and automated provenance tracking can simplify investigations of individual workflow components and help trace their impact on the final rankings.

■ **TOWARD BEST PRACTICES**

While the vision of a common platform with easily extendable and fully interoperable workflows may appear somewhat utopian today, there are a number of concrete practices that researchers can adopt to move closer toward this goal.

Reproducibility. In order for a reader of a scientific publication to be able to reproduce its results, the study should report—among the analysis and discussion of its scientific results—also all the information needed to reproduce them. However, for screening studies that involve large numbers of materials and/or multistep workflows, achieving this "radical transparency" is easier said than done: manually collecting all necessary input files, postprocessing scripts, software versions, etc. can be time-consuming, and completeness is difficult to ensure. Here, workflow managers can help by tracking the provenance automatically and providing ways to export and share this information with peers.

For example, our recent work on parsing COFs from the literature and assessing their performance for carbon capture tries to follow this approach, publishing both the full provenance graph of the study and the source code of the workflow used to orchestrate the calculations.[11,50,90] The provenance graph gives any interested researcher the ability to click on a data point and to inspect every step of the workflow that was used to compute it, try to reproduce an individual calculation themselves, or report mistakes they encounter. Sharing the source code of the corresponding workflows on collaborative platforms like GitHub further enables direct suggestions of bug fixes or improvements to the protocols, both in code and in narrative form.

> The provenance graph gives any interested researcher the ability to click on a data point and to inspect every step of the workflow that was used to compute it.

Automation. Moving from the study of a few materials to hundreds or thousands of them puts an emphasis on automation. One needs an effective way not only to manage the sequence of calculations but also to handle common errors and perform preliminary data analysis. We illustrate this using simple, practical examples from gas adsorption in nanoporous materials: Before submitting a crystal structure to GCMC simulations one has to detect and block the inaccessible pores and expand the simulation cell to include twice the cutoff used for the potential. These operations may need some external packages or *ad hoc* scripts and can be fully automated using a workflow manager. Another notable case is the handling of DFT calculations in which the self-consistent field cycle fails to converge. Depending on the system under study, remedies can include automatic resubmission with slower, more conservative minimization schemes (e.g., switching from orbital transformation to diagonalization methods) or turning on electronic smearing.[11]

When considering one's own use cases, and realizing the many intermediate steps that would need to be automated to go from an input structure to the final result, one inevitably arrives at the question whether the effort of full automation is worth the time investment. While this determination needs to be made case by case, it is easy to forget that each manual operation makes results harder to reproduce and entails substantial time investments when others (or even our future self) go on to extend and build upon the work. At the same time, modern workflow managers provide time-saving convenience features, such as automatic translations of job parameters to the language of various queuing systems,

automatic file transfers between the local workstation and the cluster, and automatic record keeping.

The development of a robust workflow can be challenging, but is a valuable outcome from a computational study on its own: it ensures that when new sets of materials are released, they can be included in the screening with minimal additional effort.

Open Source. The idea of open science does not stop at access to papers and data but extends naturally to the software used to obtain the data: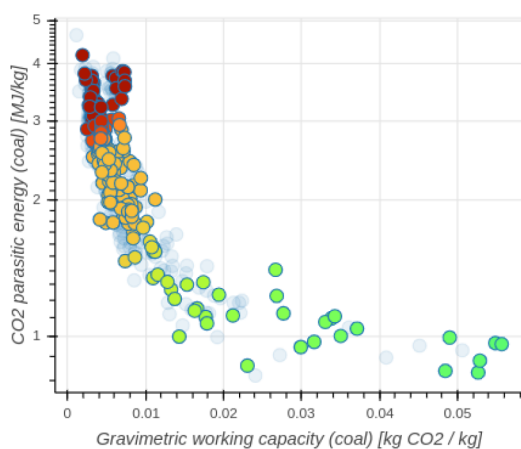 the use of free and open-source software (FOSS) lowers the barriers to reusing, reproducing, and building upon prior work. This is particularly true for materials screening studies, where readers may want to compare a new material to the set of the already screened ones. When data, software, and workflows are made openly accessible, the barrier for such checks reduces to the marginal computational cost of screening just one more material.

Despite these obvious benefits, the demand for FOSS has been notably missing from declarations in the open science context.[91] One of the reasons may be that developing and maintaining high-quality scientific software takes years of teamwork, and commercial licenses have proven to be a successful model for funding such efforts in the past.[92] Today, however, FOSS alternatives exist for most applications in computational materials science, and we do seem to observe a trend of increasing adoption of these codes vs their commercial competitors over the course of the past decade:[92,93] CA-STEP[94,95] (restricted to academic use) and openMolcas[96] being two recent examples of codes that have decided to switch to a more open licensing model.

We believe that the question of sustainable software development for open science needs to be on the table and discussed by all stakeholders.

Setting the Stage for Machine Learning. In recent years, machine learning (ML) has been rapidly mixing with molecular simulations,[97] and we expect the advent of automated workflows in the field of nanoporous materials modeling to amplify this trend.[53] Among the first applications of ML methods is the prediction of the Henry coefficient or a full isotherm in a fraction of a second, from conventional geometric properties of the crystal structure (such as pore volume and atoms' connectivity) and/or more advanced descriptors.[98−101] This massive speed up enables the screening
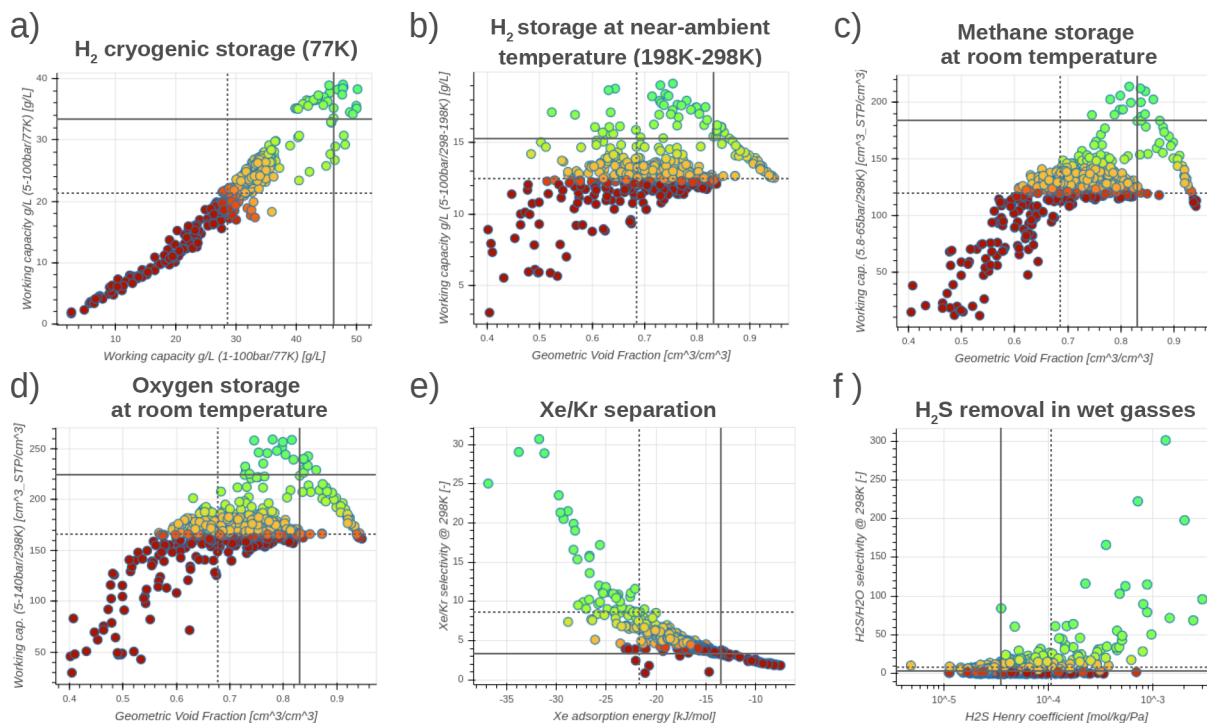

Figure 3. Performance of COF structures for CO_2 capture: parasitic energy required for the process versus gravimetric working capacity. Markers of the 250 new COFs are color-coded based on their ranking from high performance (low parasitic energy and high working capacity, green) to low performance (red). Markers of materials already included in ref 11 are shown in light gray.

Figure 4. Performance of CURATED-COFs for H_2 storage at (a) cryogenic and (b) near-ambient conditions, (c) methane storage, (d) oxygen storage, (e) Xe/Kr separation, and (f) (H_2S)/water separation. The ranking is color-coded from high performance (green) to low performance (red). Selectivities are computed as the ratio of the Henry coefficients of the two gases at 300 K. The coordinates of the markers for T-COF-2 and JUC-509 are highlighted by dashed and solid lines, respectively.

of even millions of materials at affordable computational cost, shifting the role of molecular simulation to providing sufficient training data for the ML. The main question for a new material then becomes is this structure different enough from the others already included in the training set to justify the use of expensive molecular simulations over ML predictions?[54]

However, at present, ML studies trained on data published by other groups are rare in the field. In many cases, data are recomputed specifically for the training, even when similar (but not identical and consistent!) data are available. In this context, moving as a community from delivering just a final set of results to including also the infrastructure needed to obtain them could allow ML experts to easily extend the training set with new consistent data. Reusing the same data sets in multiple ML studies would also enable effective assessments of the models themselves, which is less trivial when the training data differ. Finally, the training routines should be automated and made reproducible as well.[53]

■ TOWARD A PROTOTYPE OF A MATERIALS MATCHING PLATFORM

As a first step toward realizing the ideas put forward in this Outlook, we have extended our work of curating COFs and screening these materials for carbon capture[11] to six new applications and 250 new COFs.[50] The new structures include mostly COFs published after the original work (as tracked by the @COF_PAPERS Twitter bot), and the applications are based on previous screening studies focused on gas storage (methane,[85] hydrogen,[78,79,102] and oxygen[16]) and gas separations (Xe/Kr,[14] H_2S removal in wet gases).

Provided that the input structure is chemically sound (e.g., no missing hydrogens or overlapping atoms) and is charge-neutral (no counterbalancing ions), the workflow CURATED 99% of the structures without human intervention. For all CURATED structures, we automatically computed the adsorption isotherms and/or Henry coefficients of CO_2, N_2, H_2, CH_4, O_2, H_2S, H_2O, Xe, and Kr. From these isotherms, the KPIs were computed automatically and used to rank the materials as shown in Figure 3 for the extension of our previous study on CO_2 capture[11] and in Figure 4 for the other new applications included. The full workflow typically takes 2–5 days from start to finish, using ≈1000 core hours. In other words, it costs about the price of three cups of coffee: two cups for the curation of the structure and one for all KPIs.[103] The full provenance graph of each workflow, shown in Figure 5, is tracked automatically by the AiiDA workflow manager.

It is interesting to discuss two examples of recently reported structures that were included in the update. T-COF-2 (Figure 6a) was synthesized in 2020 and tested for a photocatalysis application.[108] The simulations do not predict this material to be among the top performers for any of the gas adsorption applications implemented so far. While this is the most likely outcome, as the screening studies are extended to more applications, the probability of discovering unexpected hits should increase.

The other COF, JUC-509 (Figure 6b), seems more promising. This material was synthesized in 2019 for catalysis.[109] Based on the atomic structure of the material, our screening predicts it to be among the top performing materials for the storage of H_2, CH_4, and O_2 (Figure 4). We acknowledge that there are other factors that may play a role in deciding whether it is worth testing JUC-509 for these
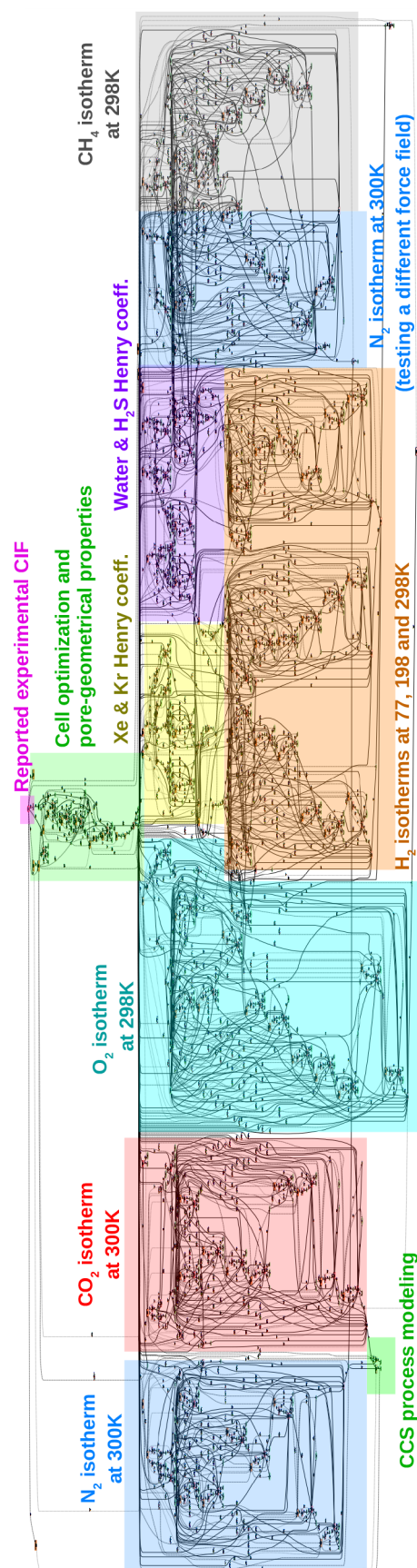

Figure 5. AiiDA provenance graph of the workflow tracing the entire path from the initial CIF file to the properties and performance computed for it. The graph shows process and data as nodes, and their connection[49] in an interactive visualization, each node can be browsed to explore the input parameters of the calculation, its output results, and the details of the processes.[104] Colors distinguish different modules of the workflow, whose source code is available online.[90] The modules make use of other popular open-source tools, such as CP2K,[105] Raspa,[106] Zeo++,[66] and chargemol.[107]
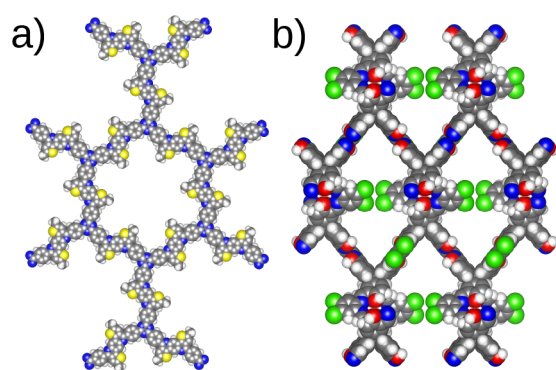
Figure 6. Crystal structures of (a) T-COF-2 and (b) JUC-509. Elements: H (white), C (gray), N (blue), oxygen (red), S (yellow), Cl (green).

applications, but it is our hope that cases like this one will lead to interesting, unexpected discoveries going forward.

Many things remain to be done in order to transform this prototype into a platform of broad impact. We have limited ourselves to COFs, and it is essential to extend it to MOFs and other porous materials. We have used relatively elementary KPIs to illustrate the concept, which we hope to replace with more advanced ones, and so far, the predictions rely on generic force fields which may not always be the optimal choice. Finally, we would like to extend the range of applications beyond the current scope of gas storage and separation. The results of this screening are updated periodically and can be accessed from materialscloud.org/discover/curated-cofs. Both the data and the source code of the underlying workflows are made available online.[110]

Over time, we hope to inspire other research groups to build upon the existing open infrastructure and develop their own modules for new applications, resulting in "living" screening studies that are regularly updated with new materials. Infrastructure projects like these require long-term commitments, which are notoriously difficult to make in today's research funding landscape. Thanks to support from the MARVEL National Centre of Competence for Research, we feel ready to accept the challenge and, given the enormous potential impact for the field, hope to be able to convince other funding agencies and possibly commercial partners to join.

In summary, the idea of this Outlook is to illustrate a dormant potential in the computational materials science community that can be unlocked by moving toward a more open, collaborative way of doing science—not necessarily inventing something spectacularly new, but simply putting together the pieces of a large puzzle. While we have made the case for the field of nanoporous materials for gas adsorption applications, the basic concept would seem extensible to further classes of materials and applications.

AUTHOR INFORMATION

Corresponding Author

Berend Smit − *Laboratory of Molecular Simulation (LSMO), Institut des Sciences et Ingénierie Chimiques, École Polytechnique Fédérale de Lausanne (EPFL), Sion CH-1951, Valais, Switzerland;* orcid.org/0000-0003-4653-8562; Email: berend.smit@epfl.ch

Authors

Daniele Ongari − *Laboratory of Molecular Simulation (LSMO), Institut des Sciences et Ingénierie Chimiques, École Polytechnique Fédérale de Lausanne (EPFL), Sion CH-1951, Valais, Switzerland;* orcid.org/0000-0001-6197-2901

Leopold Talirz − *Laboratory of Molecular Simulation (LSMO), Institut des Sciences et Ingénierie Chimiques, École Polytechnique Fédérale de Lausanne (EPFL), 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;* orcid.org/0000-0002-1524-5903

Complete contact information is available at:
https://pubs.acs.org/10.1021/acscentsci.0c00988

Notes

The authors declare no competing financial interest.

ACRONYMS

AiiDA	Automated Interactive Infrastructure and DAtabase for computational science
CIF	crystallographic information file
COF	covalent organic framework
CoRE	computation-ready experimental
CSD	Cambridge Structural Database
CURATED	Clean, Uniform, Refined with Automated Tracking from Experimental Database
DFT	density functional theory
GCMC	grand canonical Monte Carlo
KPI	key performance indicator
ML	machine learning
MOF	metal−organic framework

REFERENCES

(1) 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.

(2) Mason, J. A.; Veenstra, M.; Long, J. R. Evaluating metal−organic frameworks for natural gas storage. *Chem. Sci.* **2014**, *5*, 32−51.

(3) Zhu, L.; Liu, X.-Q.; Jiang, H.-L.; Sun, L.-B. Metal−Organic Frameworks for Heterogeneous Basic Catalysis. *Chem. Rev.* **2017**, *117*, 8129.

(4) Bavykin, A.; Kolobov, N.; Khan, I. S.; Bau, J. A.; Ramirez, A.; Gascon, J. Metal−Organic Frameworks in Heterogeneous Catalysis: Recent Progress, New Trends, and Future Perspectives. *Chem. Rev.* **2020**, *120*, 8468.

(5) Liu, X.; Huang, D.; Lai, C.; Zeng, G.; Qin, L.; Wang, H.; Yi, H.; Li, B.; Liu, S.; Zhang, M.; Deng, R.; Fu, Y.; Li, L.; Xue, W.; Chen, S. Recent advances in covalent organic frameworks (COFs) as a smart sensing material. *Chem. Soc. Rev.* **2019**, *48*, 5266−5302.

(6) Kreno, L. E.; Leong, K.; Farha, O. K.; Allendorf, M.; Van Duyne, R. P.; Hupp, J. T. Metal−Organic Framework Materials as Chemical Sensors. *Chem. Rev.* **2012**, *112*, 1105−1125.

(7) Burnham, J. F. Scopus database: a review. *Biomed. Digit. Libr.* **2006**, *3*, 1.

(8) *Database of Zeolite Structures.* www.iza-structure.org/databases/ (accessed on June 2020).

(9) 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 subset: a collection of metal-organic frameworks for past, present, and future. *Chem. Mater.* **2017**, *29*, 2618−2625.

(10) Chung, Y. G.; Haldoupis, E.; Bucior, B. J.; Haranczyk, M.; Lee, S.; Zhang, H.; Vogiatzis, K. D.; Milisavljevic, M.; Ling, S.; Camp, J. S.; Slater, B.; Siepmann, J. I.; Sholl, D. S.; Snurr, R. Q. Advances, Updates, and Analytics for the Computation-Ready, Experimental Metal−Organic Framework Database: CoRE MOF 2019. *J. Chem. Eng. Data* **2019**, *64*, 5985−5998.

(11) Ongari, D.; Yakutovich, A. V.; Talirz, L.; Smit, B. Building a Consistent and Reproducible Database for Adsorption Evaluation in Covalent−Organic Frameworks. *ACS Cent. Sci.* **2019**, *5*, 1663−1675.

(12) *CURATED-COFs.* github.com/danieleongari/egulp (accessed on June 2020).

(13) Boyd, P. G.; et al. Data-driven design of metal−organic frameworks for wet flue gas CO2 capture. *Nature* **2019**, *576*, 253−256.

(14) Banerjee, D.; Simon, C. M.; Plonka, A. M.; Motkuri, R. K.; Liu, J.; Chen, X.; Smit, B.; Parise, J. B.; Haranczyk, M.; Thallapally, P. K. Metal−organic framework with optimally selective xenon adsorption and separation. *Nat. Commun.* **2016**, *7*, 11831.

(15) Schnobrich, J. K.; Lebel, O.; Cychosz, K. A.; Dailly, A.; Wong-Foy, A. G.; Matzger, A. J. Linker-Directed Vertex Desymmetrization for the Production of Coordination Polymers with High Porosity. *J. Am. Chem. Soc.* **2010**, *132*, 13941−13948.

(16) Moghadam, P. Z.; Islamoglu, T.; Goswami, S.; Exley, J.; Fantham, M.; Kaminski, C. F.; Snurr, R. Q.; Farha, O. K.; Fairen-Jimenez, D. Computer-aided discovery of a metal−organic framework with superior oxygen uptake. *Nat. Commun.* **2018**, *9*, 1378.

(17) Sturluson, A.; Huynh, M. T.; Kaija, A. R.; Laird, C.; Yoon, S.; Hou, F.; Feng, Z.; Wilmer, C. E.; Colón, Y. J.; Chung, Y. G.; Siderius, D. W.; Simon, C. M. The role of molecular modelling and simulation in the discovery and deployment of metal-organic frameworks for gas storage and separation. *Mol. Simul.* **2019**, *45*, 1082−1121.

(18) Boyd, P. G.; Lee, Y.; Smit, B. Computational development of the nanoporous materials genome. *Nat. Rev. Mater.* **2017**, *2*, 17037.

(19) Mancuso, J. L.; Mroz, A. M.; Le, K. N.; Hendon, C. H. Electronic Structure Modeling of Metal−Organic Frameworks. *Chem. Rev.* **2020**, *120*, 8641.

(20) Tan, J.-C.; Civalleri, B.; Erba, A.; Albanese, E. Quantum mechanical predictions to elucidate the anisotropic elastic properties of zeolitic imidazolate frameworks: ZIF-4 vs. ZIF-zni. *CrystEngComm* **2015**, *17*, 375−382.

(21) Ryder, M. R.; Maul, J.; Civalleri, B.; Erba, A. Quasi-Harmonic Lattice Dynamics of a Prototypical Metal−Organic Framework. *Adv. Theory Simulations* **2019**, *2*, 1900093.

(22) Bernales, V.; Ortuño, M. A.; Truhlar, D. G.; Cramer, C. J.; Gagliardi, L. Computational Design of Functionalized Metal−Organic Framework Nodes for Catalysis. *ACS Cent. Sci.* **2018**, *4*, 5−19.

(23) Rosen, A. S.; Notestein, J. M.; Snurr, R. Q. Identifying promising metal−organic frameworks for heterogeneous catalysis via high-throughput periodic density functional theory. *J. Comput. Chem.* **2019**, *40*, 1305.

(24) Lee, K.; Howe, J. D.; Lin, L. C.; Smit, B.; Neaton, J. B. Small-Molecule Adsorption in Open-Site Metal-Organic Frameworks: A Systematic Density Functional Theory Study for Rational Design. *Chem. Mater.* **2015**, *27*, 668−678.

(25) Odoh, S. O.; Cramer, C. J.; Truhlar, D. G.; Gagliardi, L. Quantum-Chemical Characterization of the Properties and Reactivities of Metal−Organic Frameworks. *Chem. Rev.* **2015**, *115*, 6051−6111.

(26) Frenkel, D.; Smit, B. *Understanding Molecular Simulations: from Algorithms to Applications*, 2nd ed.; Academic Press: San Diego, 2002.

(27) Smit, B.; Maesen, T. L. M. Molecular Simulations of Zeolites: Adsorption, Diffusion, and Shape Selectivity. *Chem. Rev.* **2008**, *108*, 4125−4184.

(28) Gropp, C.; Canossa, S.; Wuttke, S.; Gándara, F.; Li, Q.; Gagliardi, L.; Yaghi, O. M. Standard Practices of Reticular Chemistry. *ACS Cent. Sci.* **2020**, *6*, 1255.

(29) *The Cambridge Structural Database.* ccdc.cam.ac.uk/solutions/csd-system/components/csd (accessed on June 2020).

(30) Shevchenko, A. P.; Eremin, R. A.; Blatov, V. A. The CSD and knowledge databases: from answers to questions. *CrystEngComm* **2020**, in press. DOI: 10.1039/D0CE00265H

(31) Van Heest, T.; Teich-McGoldrick, S. L.; Greathouse, J. A.; Allendorf, M. D.; Sholl, D. S. Identification of metal-organic framework materials for adsorption separation of rare gases: Applicability of ideal adsorbed solution theory (IAST) and effects of inaccessible framework regions. *J. Phys. Chem. C* **2012**, *116*, 13183−13195.

(32) Watanabe, T.; Sholl, D. S. Accelerating Applications of Metal−Organic Frameworks for Gas Adsorption and Separation by Computational Screening of Materials. *Langmuir* **2012**, *28*, 14114−14128.

(33) Goldsmith, J.; Wong-Foy, A. G.; Cafarella, M. J.; Siegel, D. J. Theoretical Limits of Hydrogen Storage in Metal−Organic Frameworks: Opportunities and Trade-Offs. *Chem. Mater.* **2013**, *25*, 3373−3382.

(34) 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.

(35) Nazarian, D.; Camp, J. S.; Sholl, D. S. A Comprehensive Set of High-Quality Point Charges for Simulations of Metal−Organic Frameworks. *Chem. Mater.* **2016**, *28*, 785−793.

(36) 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.

(37) 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.

(38) Qiao, Z.; Zhang, K.; Jiang, J. In silico screening of 4764 computation-ready, experimental metal−organic frameworks for CO 2 separation. *J. Mater. Chem. A* **2016**, *4*, 2105−2114.

(39) Erdös, M.; de Lange, M. F.; Kapteijn, F.; Moultos, O. A.; Vlugt, T. J. H. In Silico Screening of Metal−Organic Frameworks for Adsorption-Driven Heat Pumps and Chillers. *ACS Appl. Mater. Interfaces* **2018**, *10*, 27074−27087.

(40) Altintas, C.; Avci, G.; Daglar, H.; Nemati Vesali Azar, A.; Erucar, I.; Velioglu, S.; Keskin, S. An extensive comparative analysis of two MOF databases: high-throughput screening of computation-ready MOFs for CH 4 and H 2 adsorption. *J. Mater. Chem. A* **2019**, *7*, 9593−9608.

(41) Li, A.; Bueno-Perez, R.; Wiggin, S.; Fairen-Jimenez, D. Enabling efficient exploration of metal−organic frameworks in the Cambridge Structural Database. *CrystEngComm* **2020**, in press. DOI: 10.1039/D0CE00299B

(42) Barthel, S.; Alexandrov, E. V.; Proserpio, D. M.; Smit, B. Distinguishing Metal−Organic Frameworks. *Cryst. Growth Des.* **2018**, *18*, 1738−1747.

(43) Chen, T.; Manz, T. A. Identifying misbonded atoms in the 2019 CoRE metal−organic framework database. *RSC Adv.* **2020**, *10*, 26944−26951.

(44) Wilkinson, M. D.; et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **2016**, *3*, 160018.

(45) Coudert, F. Materials Databases: The Need for Open, Interoperable Databases with Standardized Data and Rich Metadata. *Adv. Theory Simulations* **2019**, *2*, 1900131.

(46) 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.

(47) 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.

(48) 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.

(49) Huber, S. P.; et al. AiiDA 1.0, a scalable computational infrastructure for automated reproducible workflows and data provenance. *Sci. Data* **2020**, *7*, 300.

(50) *CURATED covalent organic frameworks database.* materialscloud.org/discover/curated-cofs (accessed on June 2020).

(51) Altundal, O. F.; Altintas, C.; Keskin, S. Can COFs replace MOFs in flue gas separation? high-throughput computational screening of COFs for CO 2 /N 2 separation. *J. Mater. Chem. A* **2020**, *8*, 14609.

(52) 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.

(53) Jablonka, K. M.; Ongari, D.; Moosavi, S. M.; Smit, B. Big-Data Science in Porous Materials: Materials Genomics and Machine Learning. *Chem. Rev.* **2020**, *120*, 8066.

(54) Moosavi, S. M.; Nandy, A.; Jablonka, K. M.; Ongari, D.; Janet, J. P.; Boyd, P. G.; Lee, Y.; Smit, B.; Kulik, H. J. Understanding the diversity of the metal-organic framework ecosystem. *Nat. Commun.* **2020**, *11*, 4068.

(55) Jablonka, K. M.; Ongari, D.; Smit, B. Applicability of tail corrections in the molecular simulations of porous materials. *J. Chem. Theory Comput.* **2019**, *15*, 5635−5641.

(56) Yang, Q.; Liu, D.; Zhong, C.; Li, J.-R. Development of Computational Methodologies for Metal−Organic Frameworks and Their Application in Gas Separations. *Chem. Rev.* **2013**, *113*, 8261−8323.

(57) Hamad, S.; Balestra, S. R.; Bueno-Perez, R.; Calero, S.; Ruiz-Salvador, A. R. Atomic charges for modeling metal-organic frameworks: Why and how. *J. Solid State Chem.* **2015**, *223*, 144−151.

(58) Mayo, S. L.; Olafson, B. D.; Goddard, W. a. DREIDING: a generic force field for molecular simulations. *J. Phys. Chem.* **1990**, *94*, 8897−8909.

(59) 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.

(60) Potoff, J. J.; Siepmann, J. I. Vapor-liquid equilibria of mixtures containing alkanes, carbon dioxide, and nitrogen. *AIChE J.* **2001**, *47*, 1676−1682.

(61) *Awesome Materials Informatics.* github.com/tilde-lab/awesome-materials-informatics (accessed on July 2020).

(62) Jain, A.; Ong, S. P.; Chen, W.; Medasani, B.; Qu, X.; Kocher, M.; Brafman, M.; Petretto, G.; Rignanese, G.-M.; Hautier, G.; Gunter, D.; Persson, K. A. FireWorks: a dynamic workflow system designed for high-throughput applications. *Concurr. Comput. Pract. Exp.* **2015**, *27*, 5037−5059.

(63) Curtarolo, S.; Setyawan, W.; Hart, G. L.; Jahnatek, M.; Chepulskii, R. V.; Taylor, R. H.; Wang, S.; Xue, J.; Yang, K.; Levy, O.; Mehl, M. J.; Stokes, H. T.; Demchenko, D. O.; Morgan, D. AFLOW: An automatic framework for high-throughput materials discovery. *Comput. Mater. Sci.* **2012**, *58*, 218−226.

(64) Adorf, C. S.; Dodd, P. M.; Ramasubramani, V.; Glotzer, S. C. Simple data and workflow management with the signac framework. *Comput. Mater. Sci.* **2018**, *146*, 220−229.

(65) Simon, C. M.; Braun, E.; Carraro, C.; Smit, B. Statistical mechanical model of gas adsorption in porous crystals with dynamic moieties. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, E287−E296.

(66) Haranczyk, M.; Sethian, J. A. Automatic structure analysis in high-throughput characterization of porous materials. *J. Chem. Theory Comput.* **2010**, *6*, 3472−3480.

(67) Fairen-Jimenez, D.; Moggach, S. A.; Wharmby, M. T.; Wright, P. A.; Parsons, S.; Düren, T. Opening the gate: Framework flexibility in ZIF-8 explored by experiments and simulations. *J. Am. Chem. Soc.* **2011**, *133*, 8900−8902.

(68) Kolokolov, D. I.; Stepanov, A. G.; Jobic, H. Mobility of the 2-Methylimidazolate Linkers in ZIF-8 Probed by 2H NMR: Saloon Doors for the Guests. *J. Phys. Chem. C* **2015**, *119*, 27512−27520.

(69) Mace, A.; Barthel, S.; Smit, B. Automated Multiscale Approach to Predict Self-Diffusion from a Potential Energy Field. *J. Chem. Theory Comput.* **2019**, *15*, 2127−2141.

(70) Kloutse, F. A.; Zacharia, R.; Cossement, D.; Chahine, R. Specific heat capacities of MOF-5, Cu-BTC, Fe-BTC, MOF-177 and MIL-53 (Al) over wide temperature ranges: Measurements and application of empirical group contribution method. *Microporous Mesoporous Mater.* **2015**, *217*, 1−5.

(71) Kapil, V.; Wieme, J.; Vandenbrande, S.; Lamaire, A.; Van Speybroeck, V.; Ceriotti, M. Modeling the Structural and Thermal Properties of Loaded Metal−Organic Frameworks. An Interplay of Quantum and Anharmonic Fluctuations. *J. Chem. Theory Comput.* **2019**, *15*, 3237−3249.

(72) Moosavi, S. M.; Boyd, P. G.; Sarkisov, L.; Smit, B. Improving the Mechanical Stability of Metal-Organic Frameworks Using Chemical Caryatids. *ACS Cent. Sci.* **2018**, *4*, 832−839.

(73) Moghadam, P. Z.; Rogge, S. M.; Li, A.; Chow, C.-M.; Wieme, J.; Moharrami, N.; Aragones-Anglada, M.; Conduit, G.; Gomez-Gualdron, D. A.; Van Speybroeck, V.; Fairen-Jimenez, D. Structure-Mechanical Stability Relations of Metal-Organic Frameworks via Machine Learning. *Matter* **2019**, *1*, 219−234.

(74) Fumanal, M.; Capano, G.; Barthel, S.; Smit, B.; Tavernelli, I. Energy-based descriptors for photo-catalytically active metal−organic framework discovery. *J. Mater. Chem. A* **2020**, *8*, 4473−4482.

(75) Gomez, D. A.; Toda, J.; Sastre, G. Screening of hypothetical metal−organic frameworks for H 2 storage. *Phys. Chem. Chem. Phys.* **2014**, *16*, 19001−19010.

(76) Colón, Y. J.; Fairen-Jimenez, D.; Wilmer, C. E.; Snurr, R. Q. High-throughput screening of porous crystalline materials for hydrogen storage capacity near room temperature. *J. Phys. Chem. C* **2014**, *118*, 5383−5389.

(77) Bobbitt, N. S.; Chen, J.; Snurr, R. Q. High-Throughput Screening of Metal-Organic Frameworks for Hydrogen Storage at Cryogenic Temperature. *J. Phys. Chem. C* **2016**, *120*, 27328−27341.

(78) Thornton, A. W.; Simon, C. M.; Kim, J.; Kwon, O.; Deeg, K. S.; Konstas, K.; Pas, S. J.; Hill, M. R.; Winkler, D. A.; Haranczyk, M.; Smit, B. Materials Genome in Action: Identifying the Performance Limits of Physical Hydrogen Storage. *Chem. Mater.* **2017**, *29*, 2844−2854.

(79) Ahmed, A.; Seth, S.; Purewal, J.; Wong-Foy, A. G.; Veenstra, M.; Matzger, A. J.; Siegel, D. J. Exceptional hydrogen storage achieved by screening nearly half a million metal-organic frameworks. *Nat. Commun.* **2019**, *10*, 1568.

(80) Yazaydın, A. O.; Snurr, R. Q.; Park, T.-H.; Koh, K.; Liu, J.; LeVan, M. D.; Benin, A. I.; Jakubczak, P.; Lanuza, M.; Galloway, D. B.; Low, J. J.; Willis, R. R. Screening of Metal-Organic Frameworks for

Carbon Dioxide Capture from Flue Gas Using a Combined Experimental and Modeling Approach. *J. Am. Chem. Soc.* **2009**, *131*, 18198−18199.

(81) Haldoupis, E.; Nair, S.; Sholl, D. S. Finding MOFs for highly selective CO 2/N 2 adsorption using materials screening based on efficient assignment of atomic point charges. *J. Am. Chem. Soc.* **2012**, *134*, 4313−4323.

(82) 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.

(83) Qiao, Z.; Zhang, K.; Jiang, J. In silico screening of 4764 computation-ready, experimental metal-organic frameworks for CO2 separation. *J. Mater. Chem. A* **2016**, *4*, 2105−2114.

(84) Wilmer, C. E.; Leaf, M.; Lee, C. Y.; Farha, O. K.; Hauser, B. G.; Hupp, J. T.; Snurr, R. Q. Large-scale screening of hypothetical metal-organic frameworks. *Nat. Chem.* **2012**, *4*, 83−89.

(85) 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.

(86) 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.

(87) Burns, T. D.; Pai, K. N.; Subraveti, S. G.; Collins, S. P.; Krykunov, M.; Rajendran, A.; Woo, T. K. Prediction of MOF Performance in Vacuum Swing Adsorption Systems for Postcombustion CO 2 Capture Based on Integrated Molecular Simulations, Process Optimizations, and Machine Learning Models. *Environ. Sci. Technol.* **2020**, *54*, 4536.

(88) Farmahini, A. H.; Krishnamurthy, S.; Friedrich, D.; Brandani, S.; Sarkisov, L. From Crystal to Adsorption Column: Challenges in Multiscale Computational Screening of Materials for Adsorption Separation Processes. *Ind. Eng. Chem. Res.* **2018**, *57*, 15491−15511.

(89) Farmahini, H.; Friedrich, A.; Brandani, D.; Sarkisov, S.; Exploring, L. new sources of efficiency in process-driven materials screening for post-combustion carbon capture. *Energy Environ. Sci.* **2020**, *13*, 1018−1037.

(90) *AiiDA-LSMO plugin.* github.com/lsmo-epfl/aiida-lsmo (accessed on June 2020).

(91) *Open Science Means Open Source—Or, at Least, It Should.* linuxjournal.com/content/open-science-means-open-source-or-least-it-should (accessed on July 2020).

(92) *Codes—NOMAD CoE.* www.nomad-coe.eu/index.php?page=codes (accessed on July 2020).

(93) Ghiringhelli, L. M.; Carbogno, C.; Levchenko, S.; Mohamed, F.; Huhs, G.; Lüders, M.; Oliveira, M.; Scheffler, M. Towards efficient data exchange and sharing for big-data driven materials science: Metadata and data formats. *npj Comput. Mater.* **2017**, *3*, 1−9.

(94) Clark, S. J.; Segall, M. D.; Pickard, C. J.; Hasnip, P. J.; Probert, M. I. J.; Refson, K.; Payne, M. C. First principles methods using CASTEP. *Z. Kristallogr. - Cryst. Mater.* **2005**, *220*, 567−570.

(95) *CASTEP website.* www.castep.org/CASTEP/GettingCASTEP (accessed on June 2020).

(96) Fdez. Galván, I.; et al. OpenMolcas: From Source Code to Insight. *J. Chem. Theory Comput.* **2019**, *15*, 5925−5964.

(97) Butler, K. T.; Davies, D. W.; Cartwright, H.; Isayev, O.; Walsh, A. Machine learning for molecular and materials science. *Nature* **2018**, *559*, 547−555.

(98) Anderson, R.; Biong, A.; Gómez-Gualdrón, D. A. Adsorption Isotherm Predictions for Multiple Molecules in MOFs Using the Same Deep Learning Model. *J. Chem. Theory Comput.* **2020**, *16*, 1271−1283.

(99) Borboudakis, G.; Stergiannakos, T.; Frysali, M.; Klontzas, E.; Tsamardinos, I.; Froudakis, G. E. Chemically intuited, large-scale screening of MOFs by machine learning techniques. *npj Comput. Mater.* **2017**, *3*, 40.

(100) Fanourgakis, G. S.; Gkagkas, K.; Tylianakis, E.; Klontzas, E.; Froudakis, G. A Robust Machine Learning Algorithm for the Prediction of Methane Adsorption in Nanoporous Materials. *J. Phys. Chem. A* **2019**, *123*, 6080−6087.

(101) Korolev, V.; Mitrofanov, A.; Korotcov, A.; Tkachenko, V. Graph Convolutional Neural Networks as "General-Purpose" Property Predictors: The Universality and Limits of Applicability. *J. Chem. Inf. Model.* **2020**, *60*, 22−28.

(102) Kapelewski, M. T.; Runčevski, T.; Tarver, J. D.; Jiang, H. Z.; Hurst, K. E.; Parilla, P. A.; Ayala, A.; Gennett, T.; Fitzgerald, S. A.; Brown, C. M.; Long, J. R. Record High Hydrogen Storage Capacity in the Metal-Organic Framework Ni 2 (m-dobdc) at Near-Ambient Temperatures. *Chem. Mater.* **2018**, *30*, 8179−8189.

(103) We calculated an average of ca. 1000 core hours, with parallel calculations, for the evaluation of a COF for the seven applications reported as proof of concept. Of these, about two-thirds are dedicated to the cell optimization via DFT. The pricing for the coffee analogy has been estimated from the cost of commercial pay-per-use computation time.

(104) Talirz, L.; et al. Materials Cloud, a platform for open computational science. *Sci. Data* **2020**, *7*, 299.

(105) Kühne, T. D.; et al. CP2K: An electronic structure and molecular dynamics software package - Quickstep: Efficient and accurate electronic structure calculations. *J. Chem. Phys.* **2020**, *152*, 194103.

(106) 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.

(107) Manz, T. A.; Limas, N. G. Introducing DDEC6 atomic population analysis: part 1. charge partitioning theory and methodology. *RSC Adv.* **2016**, *6*, 47771−47801.

(108) Wang, Y.; Liu, H.; Pan, Q.; Wu, C.; Hao, W.; Xu, J.; Chen, R.; Liu, J.; Li, Z.; Zhao, Y. Construction of Fully Conjugated Covalent Organic Frameworks via Facile Linkage Conversion for Efficient Photoenzymatic Catalysis. *J. Am. Chem. Soc.* **2020**, *142*, 5958−5963.

(109) Yan, S.; Guan, X.; Li, H.; Li, D.; Xue, M.; Yan, Y.; Valtchev, V.; Qiu, S.; Fang, Q. Three-dimensional Salphen-based Covalent-Organic Frameworks as Catalytic Antioxidants. *J. Am. Chem. Soc.* **2019**, *141*, 2920−2924.

(110) Ongari, D.; Yakutovich, A. V.; Talirz, L.; Smit, B.*Building a consistent and reproducible database for adsorption evaluation in Covalent-Organic Frameworks*; Materials Cloud Archive 2020.107. DOI: 10.24435/materialscloud:kd-wj.