Computers and Chemical Engineering 167 (2022) 108022

Contents lists available at ScienceDirect

Computers and Chemical Engineering

journal homepage: www.elsevier.com/locate/cace

Computational discovery of Metal–Organic Frameworks for sustainable energy systems: Open challenges

Xiangyu Yin, Chrysanthos E. Gounaris *

Department of Chemical Engineering, Carnegie Mellon University, Pittsburgh, PA 15213, United States of America

ARTICLE INFO

Keywords:
Metal–organic frameworks
Microporous materials
Gas separations
Structure–function relationships

ABSTRACT

Metal–Organic Frameworks (MOFs) are promising functional microporous materials for a variety of next-generation sustainable energy systems. Their large design space makes it impossible to synthesize, test, and screen them all to identify best candidates. The computational discovery of MOFs has thus become a popular research topic, with methodological advances in computational chemistry and data science heavily contributing to this. Structure databases, materials representation, property evaluation methodologies, performance metrics, and search algorithms all pose open challenges for the community to solve. These challenges are summarized and briefly discussed in this study, with a focus on the engineering aspects required for computational MOF discovery to become a reliable tool for industry. As computational discovery workflows are complicated and necessitate skills from a variety of disciplines, bridging the knowledge gap and enhancing collaboration are critical. Despite the challenges, we remain optimistic about the great potential of computational MOF discovery technology.

1. Introduction

The increasing fossil-fuel based energy consumption is causing energy crises, global warming, climate change, and other severe issues. The necessity to move to sustainable energy has become widely recognized, as reflected in international agreements (Kyoto, Paris, etc.). However, in order to achieve the aspired global energy transition, it is critical to develop new sustainable energy technologies and systems. Transitioning present energy systems and discovering new sustainable energy sources have sparked a lot of research. For traditional energy systems such as power plants, refineries, and transportation vehicles, advances have been made in carbon capture and conversion (Bui et al., 2018), hydrogen production and utilization (Ishaq et al., 2021), and methane reduction and storage (Collins et al., 2018; He et al., 2018a), to name but a few major areas. Meanwhile, research into the harvesting, conversion, storage, and utilization of novel energy sources such as wind, hydro, photovoltaic energy, solar energy, and biomass energy is rising (Chu and Majumdar, 2012; Qazi et al., 2019). Those revolutionary technologies and advanced applications rely on the development of novel high-performing materials. Metal–organic frameworks (MOFs) have lately emerged as one of the most active research topics in the field of sustainable energy materials.

MOFs have been used in a variety of sustainable energy applications with great success. For example, MOFs have been intensively investigated for adsorption-based applications due to their intrinsic high porosity, leading to outstanding gas storage and separation capabilities (Li et al., 2018, 2019). MOFs with photo- and electro-active ligands, as well as metal ions, can be employed as energy acceptors or catalytic sites in alternative energy systems (Reddy et al., 2020; Liao et al., 2018; Bavykina et al., 2020). Supercapacitors can be made from MOFs with electrical conductivity (Zhang et al., 2019a). Furthermore, because of the synergistic effects among the functional units, a careful combination of MOFs with other functional materials (semiconductors, graphene, etc.) can result in advanced composites with superior performance than their individual components (Stock and Biswas, 2012). These MOF composites serve as novel platforms for investigating sustainable energy applications.

MOFs are customizable because of their simple coordination chemistry, which allows for rational design. MOFs are often synthesized modularly by connecting organic molecules to metal ions, clusters, or chains, to build pre-determined extended-network architectures. Researchers can modify the properties of these materials with great control, and can develop materials for specific applications thanks to the modular synthesis approach. It is also feasible to control the properties of these materials further by carrying out linker functionalization (Henke et al., 2013; Lyu et al., 2019), metal/linker exchange (Lalonde et al., 2013; Karagiaridi et al., 2014), guest installation (Suh et al., 2019; Talin et al., 2014), defect engineering (Fang et al., 2015; Chong et al., 2017), among other techniques.

* Corresponding author.
E-mail address: gounaris@cmu.edu (C.E. Gounaris).

https://doi.org/10.1016/j.compchemeng.2022.108022
Received 24 April 2022; Received in revised form 25 September 2022; Accepted 29 September 2022
Available online 14 October 2022

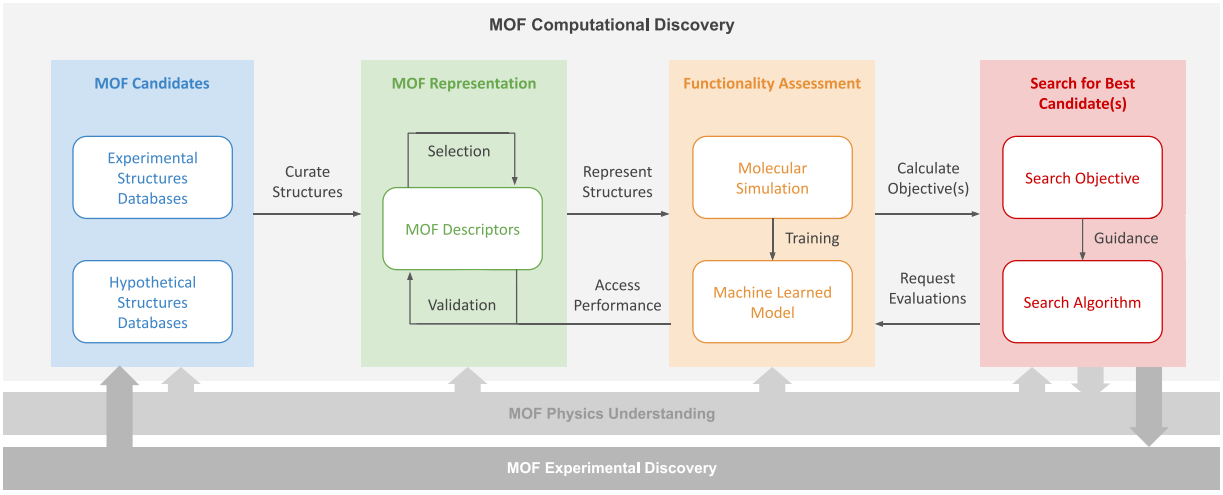

Fig. 1. Schematic representation of MOF computational discovery stages and how they may relate to experimental discovery and physics understanding.

The number of synthesized MOFs has seen an exponential increase, with the number of possible MOF structures being essentially infinite, given the enormous set of possible linkers and metal nodes as well as the different ways to combine them. The presence of a significant number of MOFs is both a challenge and an opportunity. Because the experimental synthesis, characterization, and testing of a novel material typically takes significant amount of time and resources, it is impossible to synthesize and test millions of MOFs for each application of interest. With the rapid development in improved computational chemistry techniques and access to lots of computational power, recent studies have focused on computational discovery of MOFs, where the vast materials design space is explored to identify promising candidates in a time-effective manner. The computational discovery of MOFs begins with a database of candidates. The materials design space is then described using a materials representation scheme. The candidates' functionalities are then computationally assessed using empirical correlations, machine-learned models, or molecular simulations. The search objective/performance metrics are then determined. Finally, a search (e.g., screening, heuristic, or optimization) is conducted to explore the design space. Following the identification of the best candidates, experimental efforts can be focused on these materials and physical insights could be inferred. To aid readers in understanding the big picture of MOF computational discovery, we provide Fig. 1, where we schematically represent the development stages and also coarsely illustrate how computational discovery interacts with experimental discovery and physical understanding. We further refer readers to recent reviews (Ludwig, 2019; Lyu et al., 2020; Montoya et al., 2022) for details on how computational, experimental and/or physical discovery can synergize towards accelerating materials discovery.

In the remainder of this paper, we will highlight some of the current open challenges in terms of each of the above-mentioned stages in the computational discovery process for MOFs, focusing on the engineering aspects. In Section 2, we discuss challenges in experimentally generated and hypothetical MOF databases, limitations of current materials representation techniques, and engineering challenges. Then, in Section 3, we summarize open challenges in properties evaluation methods, including molecular simulation and machine learning methods, as well as present common engineering concerns. Finally, in Section 4, we discuss open challenges related to both search objectives/performance metrics and search algorithms. We conclude with some final remarks in Section 5. Targeting for a short review paper, we keep our exposition of ideas at a high level, providing references to detailed studies for readers to further review.

2. Databases and representation

Computational MOF discovery starts from structures of candidates. A *crystallographic information file* (CIF) is a typical digital representation of a MOF structure that contains information about atomic positions, bonds, symmetry groups, and lattice factors. The Cambridge Structural Database (CSD) (Groom and Allen, 2014) and the Computation-Ready, Experimental (CoRE) MOFs database (Chung et al., 2014, 2019) are two popular databases of MOF structures sourced from experiments. Whereas the entries in the CoRE database are pre-curated so they can be readily used in computational workflows, we highlight that many experimentally obtained structures need to first be curated (e.g., by fixing missing atoms and incorrect bonds) before becoming input. In recent years, approaches that *in silico* generate hypothetical crystal structure databases (e.g., hMOF Wilmer et al., 2012; Bobbitt et al., 2016) have been developed, taking advantage of the modular nature of MOF structures and constructing structures from building blocks. We refer the readers to recent publications (Daglar and Keskin, 2020; Ongari et al., 2020) for overviews of MOF database development.

CIF files are not directly used in computational workflows, however, despite the fact that they represent complete MOF structures. To build a more comprehensive description of a material, scientists normally need to condense the information in a CIF and also dig up additional hidden information about geometries, topologies, chemistries, and energy. There is no one-size-fits-all formalism, and there are numerous representations in the literature. Common descriptors for representing a MOF include *geometrical* descriptors that describe the pore environment, *chemical* descriptors that effectively account for differences in the chemical environments, *topological* descriptors that capture details of the pore structures, and *energy-based* descriptors that consider the electronic structure and energy surfaces. In the literature, there are several ways to categorize all descriptors. We refer the readers to a recent review of MOF representation and selection (Mukherjee and Colón, 2021). Normally, when researchers need a representation of MOFs, they need to gather descriptor information from a variety of sources and choose amongst them using experiments or chemical intuition. Recent research has focused on automated generation of a comprehensive representation of MOFs (i.e., learn latent space with machine-learned models), paving the path for CIF files to be directly incorporated into computational discovery workflows. Here, we summarize several open challenges related to MOF databases and representations:

2.1. Hypothetical databases

Diversity One issue with current hypothetical MOF databases is that they lack the chemical and structural diversity of experimentally synthesized MOF collections. Furthermore, different hypothetical MOF

databases can cover different regions of the MOF design space (Moosavi et al., 2020). Most hypothetical MOF databases usually only consider a limited number of building blocks and coordinating topologies. Such an approach overlooks the huge design space potentially accessible when other metals, inorganic linkers and topologies are considered, which will make the computational discovery incomplete and inefficient. To address the issues related to database diversity, a diverse selection of building blocks must be used when constructing those databases. Recently, more studies have discussed the diversity issue of hypothetical MOF databases while newer and more diverse MOF databases have been reported (Gómez-Gualdrón et al., 2016; Nicholas et al., 2021; Majumdar et al., 2021). However, current studies focus mainly on a few aspects of diversity (e.g., chemical species, building blocks, typologies, distance on a certain projected latent space) without generally validating those diversity metrics. In our opinion, open challenges that the community still has to address include how to: (1) properly define (and, preferably, quantify) the diversity of MOF databases; (2) evaluate the diversity metrics of current MOF databases; (3) create diverse databases of building blocks for the on-demand construction of hypothetical MOFs; and (4) develop methodologies to diversify given structure databases (e.g., filter down to a diverse subset of structures or generate a superset using artificial structure components).

Synthesizability Another significant issue associated with hypothetical MOF databases is the synthesizability of such structures. When the MOFs considered in a discovery workflow to identify a promising candidate are sourced from experiments, we could potentially refer to the available experimental protocols to synthesize the material, but with hypothetical MOFs that are generated *in silico*, we have little or no knowledge of whether they are practically synthesizable, let alone how to do this. Some recent studies (Witman et al., 2016; Polat et al., 2020) have integrated computational screening with experimental synthesis, which is the ideal research paradigm to address this issue when resources are plentiful. It could be more practical if we had the ability to predict whether or not a hypothetical MOF is synthesizable, and even recommend a synthesis route. In fact, since synthesizability arises from a combination of factors and can be difficult to define, it may be better to use multiple layers of criteria (e.g., thermodynamic, mechanical stability), as inspired from directed evolution (Wang et al., 2021). Understanding, quantifying, and predicting the synthesizability of MOFs is currently a research hotspot (Ding et al., 2019; Anderson and Gómez-Gualdrón, 2020; Park et al., 2021; Nandy et al., 2021a,b) and various challenges exist throughout the process, including obtaining experimental metadata, improving molecular simulation methods, defining synthesizability metrics, and developing synthesizability prediction methods.

2.2. Experimental databases

From Experiments to Database MOF structures derived experimentally are still a safe and reliable choice for many computational studies. However, there are numerous challenges to overcome between synthesizing a MOF structure and the latter becoming an entry in a database. Accelerating experimental synthesis is one of the first that come to mind, and to that end, we note that studies towards automated high-throughput synthesis, characterization, and testing are recently on the rise (Clayson et al., 2020). Another challenge is to produce computable CIF files from characterizing experimental results. In some circumstances, current characterization analyses can be ineffective or inaccurate. This may result in CIFs not accurately reflecting the experimentally synthesized MOF structure. Furthermore, because experimental setups vary, structures generated from several sources may not be consistent. It has been reported that there are 52 different experimentally observed lattice parameters for the same HKUST-1 structure (Nazarian et al., 2017). Finally, experimental metadata (particularly "failed" trials, which are critical for understanding synthesizability) are rarely made publicly available. Challenges remain in establishing standardized experimental protocols and also recording experimental metadata (including failed ones) in machine-readable formats along with the structure files.

From Database to Computational Workflow From entries in structure databases to actual inputs of computational workflows, challenges still exist. To begin with, not all structures in experimental structure databases are labeled as MOFs accurately. Furthermore, those unprocessed MOF structure files may contain undesirable characteristics, such as solvent molecules, missing hydrogen atoms, and overlapping/disordered atoms, which can cause properties evaluation methods such as molecular simulation to be erroneous. With such a vast number of MOF structures determined, manually resolving these issues becomes impossible. To produce computation-ready structure databases, automated approaches have been developed to sift through experimental databases to detect MOF structures, remove solvent molecules, address disorder, and so on Chung et al. (2014, 2019). Because automated methods are not perfect and different methods are used in different databases, some structures in those automatically-curated databases are inaccurate, and same MOFs in two databases can have different structures (Altintas et al., 2019). Challenges remain in further improving such automatic pre-processing methods and standardizing a set of methodologies for the community.

2.3. MOFs representation

Selection Selecting a MOF representation can be fairly difficult due to the enormous number and diversity of possible methods. The best representation will be determined by a number of factors, including the dataset used, the computational model built, the information retrieved, and the problem under investigation. There is currently no universal representation of MOFs. In fact, depending on the problem settings, practically every MOFs computational discovery study in the literature has a different representation. Traditionally, scientists with specific domain knowledge determine what should be included in a MOF representation. This approach usually requires a significant amount of manual operations which reduces reproducibility. Furthermore, the performance of these representations does not always translate across workflows. In order to compose a representation from all possible descriptors, feature selection and feature extraction approaches have recently been applied (Altintas et al., 2021). They are usually based on dataset information and do not have a feedback loop from overall performance. Building an automated systematic representation selection pipeline is currently an open challenge.

Evaluation The computational workflow outcome (e.g., prediction accuracy) and computational cost are commonly utilized to evaluate a MOF representation. Recently, there are some extra concerns for MOF representation that create new challenges for the community. In particular, a representation should be: (1) *invariant* in terms of translation, rotation, and periodicity, reducing the number of symmetries and increasing the efficiency of the search (for example, the EGNN representation has recently been investigated as a way to circumvent some of the drawbacks of standard graph-based representation Satorras et al., 2021); (2) *invertible*, which refers to the existence of an inverse transform from representation to crystal structure, which is critical for the development of generative models and inverse material design methods (ideally, a one-to-one mapping of representations and structures should exist); and (3) *interpretable* and *understandable* inasmuch as, when combined with computational investigations, the representation should be able to disclose underlying physical principles and inform experimental studies. Challenges certainly remain in addressing each above mentioned requirement for MOF representation.

2.4. Engineering concerns

Comparability When combining and comparing research data from various sources, it is necessary to integrate, as well as be in position differentiate, information across databases. Generally, this can be done with data engineering methods. One current challenge, however, is how to identify and locate a specific MOF (or set of MOFs) within databases. This is partly due to the existence of various naming conventions. Dubiously, the material HKUST-1 is also known as CuBTC and MOF-199 (Sturluson et al., 2019). This creates a number of issues, namely: (1) experimental data cannot be linked to database structure files automatically; (2) structure and property databases may not be joined easily; (3) redundant information exists in multiple databases, creating potential data inconsistencies. A universal MOF identification scheme that is consistent across literature and databases may help tackle this challenge. Such schemes already exist for simple organic compounds. For example, the Simplified Molecular-Input Line-Entry System (SMILES) (Weininger, 1988) and the more recent SELF-referencIng Embedded Strings (SELFIES) (Krenn et al., 2020) are two representation systems that are well-known and widely used, but they are not perfectly compatible with MOFs because of the extra complexity and unit cell periodicity. Thus, there is a need to develop a scheme specific to MOFs. Recently, MOF-compatible identifiers have been developed such as MOFid/MOFkey (Bucior et al., 2019) and RFcode (Yao et al., 2021). Challenges remain in semantics design and ontology engineering to improve the MOF identification methodology. There also exist challenges to dynamically adapt universal identifiers to specific applications. We refer the readers to Scheffler et al. (2022) for a more in-depth discussion on achieving FAIR (findable, accessible, interoperable and reusable) materials data infrastructures.

Reproducibility Transparency and standardization are keys for reproducibility of MOFs research data. This goal should be aided by recent trends towards open-access publications and numerous data repositories, such as Zenodo (Sicilia et al., 2017), the Computational Material Repository (Landis et al., 2012), and the Open Science Framework (Foster and Deardorff, 2017). In addition to transparency, the community needs standardized data generation, storage, and processing methods. To that end, recent advances in databases with built-in standardized toolkits for computational studies, such as the Materials Project (Jain et al., 2013) and AFLOWLIB (Curtarolo et al., 2012), will facilitate the realization of this goal. As for experimental data, digitizing experimental records, results and metadata in a standardized and machine-readable format remains a great challenge. It is also crucial to disclose the findings of unsuccessful experiments and the performances of poor-performing materials in order to better understand essential properties like synthesizability. Unfortunately, this practice is far from the norm when publishing scholarly results.

3. Properties evaluation

Efficient and accurate properties evaluation methods that could predict/calculate target functions of MOFs from the representation/descriptors are the key to computational discovery of MOFs. A variety of computational methods have been utilized to calculate MOF properties, among which the molecular simulation approach is the most investigated one. Molecular simulation via *ab initio* quantum chemistry computations gives relatively accurate predictions. For example, Barona et al. (2019) carried out DFT calculations to predict MOF catalysts and the results agreed well with experimental validation. But *ab initio* calculations require significant amount of computational resources. On the contrary, Monte Carlo or molecular dynamics simulations with classical force-fields are less accurate but much more efficient, and are thus used more widely in high-throughput studies. The readers are directed to a recent review (Sturluson et al., 2019) for detailed discussion on molecular simulation approaches. Recently, data-driven (e.g., machine learning) approaches have been widely investigated, given the increasing availability of MOF molecular simulation tools and the development of MOF property databases. Machine learning (ML) is very efficient in terms of inference (i.e., prediction) and the community has focused on improving its accuracy and training cost, among other aspects. Although in this section we consider data-driven approaches as property evaluators (mostly via regression-type models), we note that there exist other ways of integrating data-driven technologies. For example, using classification and regression models to predict synthesizability (Anderson and Gómez-Gualdrón, 2020; Jang et al., 2020; Park et al., 2022), using generative models to generate hypothetical structures (Yao et al., 2021; Altintas et al., 2021), and extracting literature information using natural language models (Tshitoyan et al., 2019; Park et al., 2021) are some examples. The readers are directed to recent publications (Gu et al., 2019; Chong et al., 2020; Altintas et al., 2021; Batra et al., 2021; Rosen et al., 2022) for more details of machine-learned models in computational MOF discovery. Below, we articulate open challenges related to the computational evaluation of MOF properties, focusing on molecular simulation and machine learning methods.

3.1. Molecular simulation

Accuracy Molecular simulations are commonly used for the *in silico* evaluation of MOF properties, and most machine learning models use datasets generated from such simulations. As a result, improving the accuracy of the molecular simulation methods is critical. Handling the flexibility of MOFs is currently one of the challenges in improving simulation accuracy. MOFs are assumed to be rigid in practically all computational studies in the literature by neglecting bound intramolecular interactions. However, MOFs are indeed flexible, and studies have demonstrated that this flexibility has a substantial impact on a MOF's properties (Sarkisov et al., 2014; Aljammal et al., 2019). For the vast majority of MOFs in the databases, the effect of removing rigidity assumptions has not been examined. Besides flexibility, particular instances, such as open metal sites, special functional groups and defects, are also rarely investigated in the literature. Considering all of those cases would require significant algorithmic advancements and computational capacity, and that those cases may only be relevant for a limited subset of MOFs, it may not be practical to design universal molecular simulation methodologies suitable for all MOFs. While scientists are surely making improvements in understanding those limitations and developing new technologies, it remains a challenge for engineers to design automated and systematic strategies for validating various assumptions and methodologies (e.g., force fields, charge assignment methods) and selecting the combination with the lowest computing cost that meets the accuracy requirement.

Efficiency Another key challenge in molecular simulation is to improve its efficiency, as researchers frequently have to trade-off between accuracy and computational cost in computational discovery studies. Many studies in the field focus on using machine learning methods to replace expensive simulations. However, as molecular simulation is the source of training data in many data-driven method studies, and data-driven methods generally require a large amount of data, we would argue that the need to accelerate molecular simulation is even greater with the advent of machine learning methods. To improve the efficiency of molecular simulation, the first challenge is to reduce the time to develop a new force field. Transfer learning approaches have recently been utilized to construct computationally cheap force fields that gain knowledge from one system and apply it to another (Smith et al., 2019). The second challenge is to use the latest technological advances in parallel computing hardware and software to speed up the simulations. Recently, progress has been made in developing Grand-Canonical Monte Carlo algorithms and density functional theory algorithms on GPUs (Nejahi et al., 2019; Zhou and Wu, 2020) and gained orders of magnitudes of speedups.

3.2. Data-driven machine learning

Generalizability and Transferability One commonplace goal of data-driven approaches is that the learned models perform well with unknown data (i.e., *generalizability*) and/or other problem settings (i.e., *transferability*). To realize generalizability and transferability are especially of interest to MOF discovery because the vast design space (i.e., unknown candidates) and many potential applications (e.g., unseen problem settings) of MOFs. For the generalizability of MOF property prediction models, there are additional factors that need to be considered beyond testing on validation datasets: (1) databases, which should be diverse and representative; (2) MOF representation, which should be effective and capture essential information; (3) MOF properties, which should fall within expected value ranges and exhibit enough spread; and (4) machine learning models, which should be efficiently "overparameterized" (i.e., large enough) (Brutzkus and Globerson, 2019). Overall, we could argue that a major challenge remains in systematically diagnosing and fixing the ML model for better generalization. In terms of transferability, one aspect is to learn a transferable representation (e.g., embedding) that may be used for a variety of tasks. Meta-learning has also been used to predict the properties of MOFs under various situations (Sun et al., 2021). Inspired by the rapid development of pre-trained models in other domains of interest to the ML community (Wolf et al., 2019), it remains a challenge to construct a generic and transferable pre-trained MOFs model that can be applied to various applications via re-training with a small number of new data points.

Working with Limited Data ML models often require a large number of data points. Creating large datasets of MOF properties and/or representation can be computationally expensive, and the number of data points available for challenging properties is sometimes very limited. This imposes a great challenge for pursuing machine learning models, as smaller datasets result in worse predictability and poor generalization of those. As a result, novel machine learning techniques and methodologies for small datasets have been investigated in the literature, such as transfer learning (He et al., 2018b; Ma et al., 2020) (using pre-trained parameters before training the model on the limited database) and active learning (Xue et al., 2016) (essentially sampling the training set from the entire database). Another challenge with limited datasets is the prevalence of missing values, which could come from failed computations or unreported experimental data. Recently, the literature has suggested methods such as using a recommendation system to estimate missing values and address this issue (Sturluson et al., 2021).

3.3. Engineering concerns

Data Availability The majority of computational discovery studies in the literature rely on in-house molecular simulation capabilities to evaluate material properties, and the simulated data is almost never publicly available. Even when similar data is available, data is frequently re-computed. It is an open challenge to create open MOF properties databases, which would save computing resources, enable cross-disciplinary collaboration, and allow different studies to be compared. Another challenge with data availability is that research reported in the literature has mostly concentrated on a few number of MOF properties predominantly related to small molecule adsorption, while other relevant properties have received less attention. In particular, additional thermodynamic properties can be important when evaluating industrial-level gas separation/storage processes, while electronic and catalytic properties are crucial for energy-related applications, yet those properties are rarely explored in the literature. We have recently seen the creation of databases like QMOF (Rosen et al., 2021) that attempt to address this issue. While various property databases are being created, it would be a challenge for engineers to create a centralized interface to link and query them. Last but not least, validating computed properties remains a challenge. The ideal way to validate computational studies is via experiments. But experimental validations of all computational results is impractical. Instead, databases of experimental properties could be used as benchmarks to validate and evaluate computational workflows in this context. Selecting and building such robust benchmark property databases remains a challenge.

Workflow Automation Manual operation makes it hard to reproduce results and requires extra time when there is need to extend the work. Therefore, automation is critical in the reproducibility of molecular simulation and machine learning studies. Recently, building software infrastructures to automate various aspects of molecular simulation and machine learning has become a hot research topic. The Open Force Field Initiative (Shirts et al., 2019), for example, includes a number of tools for automating molecular simulation, while AutoML tools have been used to build machine learning models (Borboudakis et al., 2017). Beyond software infrastructure, there exist challenges in automating the whole computational workflow. Here, comprehensive workflow manager infrastructures, such as AiiDA (Pizzi et al., 2016) and FireWorks (Jain et al., 2015), have recently been built to automate, manage, persist, distribute, and recreate complicated computational workflows. We also remark that computing environment management, which includes the computing hardware, operating system, and software employed, is crucial for the reproducibility of computational operations and information reported should generally include versions and configurations. The development of such environment management infrastructure specialized for computational investigations is still a challenge. Quantum Mobile virtual machines (Talirz et al., 2020) for simulation workflows and cloud-based tools like Codeocean (Staubitz et al., 2016) for machine learning activities are some of the current available options.

4. Metrics and search

Conceptually, computational materials discovery is an optimization problem aimed at searching the materials design space for the best candidate with the optimal target functionality. We note that, due to uncertainties, the candidate can be a cluster/group of structures instead of a single structure. With the MOF databases, representation, and property evaluation methods in place, a search might be as simple as an enumerative process, where a finite subset of candidates in the design space is evaluated and the best gets selected. In fact, much of the field's literature is built on this concept, with high-throughput screening using a variety of datasets, representations, and property evaluation methods. Some high-throughput screening studies have achieved success in discovering new MOFs – or new usages for MOFs – that already existed but had never been applied in the desired context (Moghadam et al., 2018; Boyd et al., 2019). However, the screening strategy generally suffers from high computation cost and limited search space.

To that end, systematic search/optimization approaches have been investigated recently. We note that nearly all current Structure–function relationships (e.g., property evaluators) are considered as black-box systems, in which we can see the inputs and outputs but do not have access to a mathematical description of the relationship. Accordingly, the search methodologies developed to accommodate such relationships are simulation-based black-box search/sampling algorithms, including genetic algorithms (Chung et al., 2016), Monte Carlo tree search (Zhang et al., 2019b), and Bayesian optimization (Deshwal et al., 2021). Such black-box algorithms usually necessitate the specification of structure transformation rules or surrogate model forms, and if not carefully designed, they will require a very high number of data points and evaluation. Moreover, without explicitly understanding and exploring the underlying physics (i.e., Structure–function relationship), they may identify top-performing MOFs but fail to identify the best candidates (i.e., global optimality). Another existing strategy for searching the design space is to use generative machine learning models to produce

new MOF structures with good performance metrics directly, allowing for the so-called *inverse design*.

Before applying any search method for computational discovery, it is critical to choose a proper performance metric/search objective. For many straightforward applications, search metrics could simply be material attributes. However, simplistic measurements would not adequately reflect our goal in many other situations, particularly when considering MOFs in multi-scale systems and taking more realistic factors into account. As a result, various functional and hybrid metrics have been proposed in the literature to address this issue. We refer the readers to recent publications (Erucar and Keskin, 2018; Leperi et al., 2019; Farmahini et al., 2021) for a detailed treatment of MOF performance metrics. In the remainder, we will discuss challenges related to metrics and search methods.

4.1. Search objectives

Process-Level Information To find the optimum material for a specific process, one requires process-level information and should ideally aim to optimize both the process and the material design at the same time. It has been demonstrated that existing performance metrics, despite the fact that they frequently allow for the elimination of inferior materials and may be adequate proxies in idealized cases, do not provide an accurate ranking of the MOFs (Farmahini et al., 2018). The reason is that overall process-materials system performance is the outcome of a complex system with highly coupled materials design and process design decisions. There is rarely a direct link between process performance and material properties. Furthermore, measuring the entire process system with a single performance metric is usually inadequate. Process analog/simplified/surrogate models have recently been used to evaluate process-level performance in materials search (Subramanian Balashankar and Rajendran, 2019; Arora et al., 2020). Combining technical and economic evaluation of the entire process with the computational discovery workflow is still a challenge.

Cost Another issue with present performance metrics is the lack of cost objective, which is critical for process development and technology commercialization, as cost is frequently the deciding factor. Currently, the cost of uncommercialized MOFs is frequently thought to be comparable to that of existing materials when, in fact, MOFs have completely different raw ingredients and manufacturing processes than typical materials. To further complicate the matter, new technologies are constantly being developed and need to be reflected on the cost. As a result, developing costing models for MOFs is an extremely difficult open challenge. Aside from material costs, determining additional operational costs, such as regeneration costs, is another issue that requires a thorough understanding of the entire process. Readers should refer to recent publications on techno-economic analysis of MOFs-based adsorption processes (Danaci et al., 2020; DeSantis et al., 2017; Shi et al., 2021; Severino et al., 2021).

Sustainability Whereas the majority of existing MOF computational discovery studies focus on MOF functionality and/or performance, sustainability studies are largely conducted in parallel and focus more on assessment and analysis rather than discovery. To that end, there exist opportunities to incorporate sustainability considerations into the MOF discovery process itself. However, relying on a computational discovery workflow to identify MOFs that are both high-performing and sustainable constitutes a great challenge for the community. We note that sustainability as a high-level concept cannot be readily described with a simple metric, as multiple factors from different perspectives need to be taken into consideration. In our view, some factors that should be considered for evaluating sustainability of MOFs are: (1) safety and "greenness" of MOF synthesis and production; (2) energy requirements of and/or savings from the applications enabled by MOFs; and (3) environmental, economic and societal impacts associated with the complete MOF life cycle. We refer the readers to reviews discussing sustainability aspects of MOFs (Julien et al., 2017; Chen et al., 2017; Kumar et al., 2019; Woodliffe et al., 2021; Faust, 2016; Grande et al., 2017). Besides defining and quantifying sustainability metrics, it will also be challenging to use those – often convoluted ones – to guide the materials discovery, which calls for advanced optimization algorithms.

4.2. Search methods

Optimization In our opinion, a MOF computational discovery methodology should be evaluated, at a minimum, along the following dimensions: (1) search efficiency in terms of computational time, number of evaluations, CPU resource usage, and other related metrics; (2) search effectiveness in terms of exploration–exploitation trade-off and optimality gap; (3) scalability with respect to data size and system size; and (4) flexibility in terms of dependence on specific system and domain knowledge. Currently, systematizing search methodologies for MOF discovery is still a relatively unexplored research area with studies focusing mainly on the first two of the aforementioned dimensions. Challenges exist for advancing along each one of those dimensions as well as for developing all-rounder methodologies. Another group of challenges arise from more complex and realistic search objectives, such as those discussed previously. To accommodate such objectives requires advanced search techniques, such as multi-objective constrained optimization. Furthermore, to deal with uncertainties (e.g., data inaccuracy) in the discovery process, we need to resort to more involved – and generally less tractable – optimization under uncertainty techniques. Last, but not least, there exist opportunities and challenges to go beyond current simulation-based methodologies and to develop physics-informed technologies that can help us obtain physical understanding during the materials discovery process. We refer the readers to recent reviews (Karniadakis et al., 2021; Peng et al., 2022) discussing the background and impact of physics-informed computational materials in general. For MOF discovery specifically, with understanding of the underlying physics, we can formulate and exploit the Structure–function relationships in glass-box forms and break the inherent limitations of black-box search. This is intriguing from an optimization perspective since it allows for the use of rigorous mathematical optimization procedures. Besides its rigor, mathematical optimization also gives formal methods for dealing with multi-objective optimization problems, restrictions, and various sources of uncertainty. Thus, it is an open challenge both to learn physics from and to utilize the physics back into the materials discovery process. To that end, technologies such as physics-informed feature engineering, pattern recognition, interpretable ML, causality inference, symbolic regression and reinforcement learning could be further investigated and integrated with MOF discovery.

Bridging the Knowledge Gap The computational discovery of MOFs is a complex task that incorporates knowledge and technologies from a wide range of disciplines. It requires the participation and collaboration of materials scientists, computational chemists, data scientists, machine learning engineers, process engineers, and often also software engineers. The knowledge gap that exists between different domains presents a great challenge for productive collaboration on computational discovery/search. For example, despite all improvements in the automation and user interface of molecular simulation toolkits, process engineers may still find them difficult to access and utilize. Process modeling and optimization tools, on the other hand, are not always readily available to the materials science community. What is more, the latest machine learning advancements may not be phrased with notations that is familiar to chemical engineers. In order to close the knowledge gap between different fields, it is critical to make technologies more accessible, user-friendly and automated, while also providing enough documentation for each technology. The development of an all-purpose software program/platform with a user-friendly interface dedicated to one or more applications is possible

in the future. This aspired tool would require MOF structures and process settings as input, and it will output an integrated model for the material-process-environment system. It could then be linked with an optimization engine to enable an optimization workflow to recommend MOF candidates, process design and operating condition, as well as estimated environmental impacts.

5. Conclusions

Open challenges in computational MOF discovery were summarized and briefly explored in this study, covering database and representation, properties evaluation, performance metrics, and search algorithms, among other topics. We should highlight that this review is written from the perspective of a process engineer, and that it is primarily concerned with engineering challenges. Making the computational discovery more realistic and reliable, as well as applying the results in real-world industrial production, is the ultimate challenge for all participants in the field. Unfortunately, the majority of current research utilizes highly idealized structures and assumptions, representations without systematic validations, simplified properties evaluation methods, performance metrics that are not aligned with process goals, and heuristic search methods, among other weaknesses. As a result, computationally discovered MOFs can rarely be synthesized or confirmed to perform well in the laboratory, let alone make their way to adoption in real-world applications. Furthermore, many experimentally-proven top-performing MOFs have never been identified in computational studies (Taddei and Petit, 2021). Although we are confronted with significant challenges, it is remarkable to see how quickly this field is evolving and progressing, with new articles and innovations being introduced on a daily basis. Considering the potential and importance of computational discovery of MOFs in the future, we are quite optimistic.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgments

We graciously acknowledge funding from the U.S. Department of Energy, Office of Fossil Energy's Crosscutting Research Program through the Institute for the Design of Advanced Energy Systems.

Disclaimer

This paper was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

References

Aljammal, Noor, Jabbour, Christia, Chaemchuen, Somboon, Juzsakova, Tatjána, Verpoort, Francis, 2019. Flexibility in metal–organic frameworks: A basic understanding. Catalysts 9 (6), 512.

Altintas, Cigdem, Altundal, Omer Faruk, Keskin, Seda, Yildirim, Ramazan, 2021. Machine learning meets with metal organic frameworks for gas storage and separation. J. Chem. Inf. Model. 61 (5), 2131–2146.

Altintas, Cigdem, Avci, Gokay, Daglar, Hilal, Azar, Ayda Nemati Vesali, Erucar, Ilknur, Velioglu, Sadiye, Keskin, Seda, 2019. An extensive comparative analysis of two MOF databases: High-throughput screening of computation-ready MOFs for CH4 and H2 adsorption. J. Mater. Chem. A 7 (16), 9593–9608.

Anderson, Ryther, Gómez-Gualdrón, Diego A., 2020. Large-scale free energy calculations on a computational metal–organic frameworks database: Toward synthetic likelihood predictions. Chem. Mater. 32 (19), 8106–8119.

Arora, Akhil, Iyer, Shachit S., Hasan, M.M. Faruque, 2020. Computational material screening using artificial neural networks for adsorption gas separation. J. Phys. Chem. C (ISSN: 19327455) 124 (39), 21446–21460. http://dx.doi.org/10.1021/acs.jpcc.0c05900.

Barona, Melissa, Ahn, Sol, Morris, William, Hoover, William, Notestein, Justin M., Farha, Omar K., Snurr, Randall Q., 2019. Computational predictions and experimental validation of alkane oxidative dehydrogenation by fe2 m mof nodes. ACS Catal. 10 (2), 1460–1469.

Batra, Rohit, Song, Le, Ramprasad, Rampi, 2021. Emerging materials intelligence ecosystems propelled by machine learning. Nature Rev. Mater. 6 (8), 655–678.

Bavykina, Anastasiya, Kolobov, Nikita, Khan, Il Son, Bau, Jeremy A., Ramirez, Adrian, Gascon, Jorge, 2020. Metal–organic frameworks in heterogeneous catalysis: Recent progress, new trends, and future perspectives. Chem. Rev. 120 (16), 8468–8535.

Bobbitt, N. Scott, Chen, Jiayi, Snurr, Randall Q., 2016. High-throughput screening of metal–organic frameworks for hydrogen storage at cryogenic temperature. J. Phys. Chem. C 120 (48), 27328–27341.

Borboudakis, Giorgos, Stergiannakos, Taxiarchis, Frysali, Maria, Klontzas, Emmanuel, Tsamardinos, Ioannis, Froudakis, George E., 2017. Chemically intuited, large-scale screening of MOFs by machine learning techniques. NPJ Comput. Mater. 3 (1), 1–7.

Boyd, Peter G., Chidambaram, Arunraj, García-Díez, Enrique, Ireland, Christopher P., Daff, Thomas D., Bounds, Richard, Gładysiak, Andrzej, Schouwink, Pascal, Moosavi, Seyed Mohamad, Maroto-Valer, M. Mercedes, et al., 2019. Data-driven design of metal–organic frameworks for wet flue gas CO2 capture. Nature 576 (7786), 253–256.

Brutzkus, Alon, Globerson, Amir, 2019. Why do larger models generalize better? A theoretical perspective via the XOR problem. In: International Conference on Machine Learning. PMLR, pp. 822–830.

Bucior, Benjamin J., Rosen, Andrew S., Haranczyk, Maciej, Yao, Zhenpeng, Ziebel, Michael E., Farha, Omar K., Hupp, Joseph T., Siepmann, J. Ilja, Aspuru-Guzik, Alán, Snurr, Randall Q., 2019. Identification schemes for metal–organic frameworks to enable rapid search and cheminformatics analysis. Cryst. Growth Des. 19 (11), 6682–6697.

Bui, Mai, Adjiman, Claire S., Bardow, André, Anthony, Edward J., Boston, Andy, Brown, Solomon, Fennell, Paul S., Fuss, Sabine, Galindo, Amparo, Hackett, Leigh A., et al., 2018. Carbon capture and storage (CCS): The way forward. Energy Environ. Sci. 11 (5), 1062–1176.

Chen, Junying, Shen, Kui, Li, Yingwei, 2017. Greening the processes of metal–organic framework synthesis and their use in sustainable catalysis. ChemSusChem 10 (16), 3165–3187.

Chong, Sanggyu, Lee, Sangwon, Kim, Baekjun, Kim, Jihan, 2020. Applications of machine learning in metal-organic frameworks. Coord. Chem. Rev. 423, 213487.

Chong, Sanggyu, Thiele, Günther, Kim, Jihan, 2017. Excavating hidden adsorption sites in metal-organic frameworks using rational defect engineering. Nature Commun. 8 (1), 1–10.

Chu, Steven, Majumdar, Arun, 2012. Opportunities and challenges for a sustainable energy future. Nature 488 (7411), 294–303.

Chung, Yongchul G., Camp, Jeffrey, Haranczyk, Maciej, Sikora, Benjamin J., Bury, Wojciech, Krungleviciute, Vaiva, Yildirim, Taner, Farha, Omar K., Sholl, David S., Snurr, Randall Q., 2014. Computation-ready, experimental metal–organic frameworks: A tool to enable high-throughput screening of nanoporous crystals. Chem. Mater. 26 (21), 6185–6192.

Chung, Yongchul G., Gómez-Gualdrón, Diego A., Li, Peng, Leperi, Karson T., Deria, Pravas, Zhang, Hongda, Vermeulen, Nicolaas A., Stoddart, J. Fraser, You, Fengqi, Hupp, Joseph T., et al., 2016. In silico discovery of metal-organic frameworks for precombustion CO2 capture using a genetic algorithm. Sci. Adv. 2 (10), e1600909.

Chung, Yongchul G., Haldoupis, Emmanuel, Bucior, Benjamin J., Haranczyk, Maciej, Lee, Seulchan, Zhang, Hongda, Vogiatzis, Konstantinos D., Milisavljevic, Marija, Ling, Sanliang, Camp, Jeffrey S., et al., 2019. Advances, updates, and analytics for the computation-ready, experimental metal–organic framework database: Core MOF 2019. J. Chem. Eng. Data 64 (12), 5985–5998.

Clayson, Ivan G., Hewitt, Daniel, Hutereau, Martin, Pope, Tom, Slater, Ben, 2020. High throughput methods in the synthesis, characterization, and optimization of porous materials. Adv. Mater. 32 (44), 2002780.

Collins, William J., Webber, Christopher P., Cox, Peter M., Huntingford, Chris, Lowe, Jason, Sitch, Stephen, Chadburn, Sarah E., Comyn-Platt, Edward, Harper, Anna B., Hayman, Garry, et al., 2018. Increased importance of methane reduction for a 1.5 degree target. Environ. Res. Lett. 13 (5), 054003.

Curtarolo, Stefano, Setyawan, Wahyu, Wang, Shidong, Xue, Junkai, Yang, Kesong, Taylor, Richard H., Nelson, Lance J., Hart, Gus L.W., Sanvito, Stefano, Buongiorno-Nardelli, Marco, et al., 2012. AFLOWLIB. ORG: A distributed materials properties repository from high-throughput ab initio calculations. Comput. Mater. Sci. 58, 227–235.

Daglar, Hilal, Keskin, Seda, 2020. Recent advances, opportunities, and challenges in high-throughput computational screening of MOFs for gas separations. Coord. Chem. Rev. 422, 213470.

Danaci, David, Bui, Mai, Mac Dowell, Niall, Petit, Camille, 2020. Exploring the limits of adsorption-based CO2 capture using MOFs with PVSA–from molecular design to process economics. Mol. Syst. Des. Eng. 5 (1), 212–231.

DeSantis, Daniel, Mason, Jarad A., James, Brian D., Houchins, Cassidy, Long, Jeffrey R., Veenstra, Mike, 2017. Techno-economic analysis of metal–organic frameworks for hydrogen and natural gas storage. Energy Fuels 31 (2), 2024–2032.

Deshwal, Aryan, Simon, Cory M., Doppa, Janardhan Rao, 2021. Bayesian optimization of nanoporous materials. Mol. Syst. Des. Eng. 6 (12), 1066–1086.

Ding, Meili, Cai, Xuechao, Jiang, Hai-Long, 2019. Improving MOF stability: Approaches and applications. Chem. Sci. 10 (44), 10209–10230.

Erucar, Ilknur, Keskin, Seda, 2018. High-throughput molecular simulations of metal organic frameworks for CO2 separation: Opportunities and challenges. Front. Mater. 5, 4.

Fang, Zhenlan, Bueken, Bart, De Vos, Dirk E., Fischer, Roland A., 2015. Defect-engineered metal–organic frameworks. Angew. Chem. Int. Ed. 54 (25), 7234–7254.

Farmahini, Amir H., Krishnamurthy, Shreenath, Friedrich, Daniel, Brandani, Stefano, Sarkisov, Lev, 2018. From crystal to adsorption column: Challenges in multiscale computational screening of materials for adsorption separation processes. Ind. Eng. Chem. Res. 57 (45), 15491–15511.

Farmahini, Amir H., Krishnamurthy, Shreenath, Friedrich, Daniel, Brandani, Stefano, Sarkisov, Lev, 2021. Performance-based screening of porous materials for carbon capture. Chem. Rev. 121 (17), 10666–10741.

Faust, Thomas, 2016. MOFs move to market. Nature Chem. 8 (11), 990–991.

Foster, Erin D., Deardorff, Ariel, 2017. Open science framework (OSF). J. Med. Libr. Assoc.: JMLA 105 (2), 203.

Gómez-Gualdrón, Diego A., Colón, Yamil J., Zhang, Xu, Wang, Timothy C., Chen, Yu-Sheng, Hupp, Joseph T., Yildirim, Taner, Farha, Omar K., Zhang, Jian, Snurr, Randall Q., 2016. Evaluating topologically diverse metal–organic frameworks for cryo-adsorbed hydrogen storage. Energy Environ. Sci. 9 (10), 3279–3289.

Grande, Carlos A., Blom, Richard, Spjelkavik, Aud, Moreau, Valentine, Payet, Jérôme, 2017. Life-cycle assessment as a tool for eco-design of metal-organic frameworks (MOFs). Sustain. Mater. Technol. 14, 11–18.

Groom, Colin R., Allen, Frank H., 2014. The cambridge structural database in retrospect and prospect. Angew. Chem. Int. Ed. 53 (3), 662–671.

Gu, Geun Ho, Noh, Juhwan, Kim, Inkyung, Jung, Yousung, 2019. Machine learning for renewable energy materials. J. Mater. Chem. A 7 (29), 17096–17117.

He, Yabing, Chen, Fengli, Li, Bin, Qian, Guodong, Zhou, Wei, Chen, Banglin, 2018a. Porous metal–organic frameworks for fuel storage. Coord. Chem. Rev. 373, 167–198.

He, Yuping, Cubuk, Ekin D., Allendorf, Mark D., Reed, Evan J., 2018b. Metallic metal–organic frameworks predicted by the combination of machine learning methods and ab initio calculations. J. Phys. Chem. Lett. 9 (16), 4562–4569.

Henke, Sebastian, Schneemann, Andreas, Fischer, Roland A., 2013. Massive anisotropic thermal expansion and thermo-responsive breathing in metal–organic frameworks modulated by linker functionalization. Adv. Funct. Mater. 23 (48), 5990–5996.

Ishaq, Haris, Dincer, Ibrahim, Crawford, Curran, 2021. A review on hydrogen production and utilization: Challenges and opportunities. Int. J. Hydrogen Energy.

Jain, Anubhav, Ong, Shyue Ping, Chen, Wei, Medasani, Bharat, Qu, Xiaohui, Kocher, Michael, Brafman, Miriam, Petretto, Guido, Rignanese, Gian-Marco, Hautier, Geoffroy, et al., 2015. FireWorks: A dynamic workflow system designed for high-throughput applications. Concurr. Comput.: Pract. Exp. 27 (17), 5037–5059.

Jain, Anubhav, Ong, Shyue Ping, Hautier, Geoffroy, Chen, Wei, Richards, William Davidson, Dacek, Stephen, Cholia, Shreyas, Gunter, Dan, Skinner, David, Ceder, Gerbrand, et al., 2013. Commentary: The materials project: A materials genome approach to accelerating materials innovation. APL Mater. 1 (1), 011002.

Jang, Jidon, Gu, Geun Ho, Noh, Juhwan, Kim, Juhwan, Jung, Yousung, 2020. Structure-based synthesizability prediction of crystals using partially supervised learning. J. Am. Chem. Soc. 142 (44), 18836–18843.

Julien, Patrick A., Mottillo, Cristina, Friščić, Tomislav, 2017. Metal–organic frameworks meet scalable and sustainable synthesis. Green Chem. 19 (12), 2729–2747.

Karagiaridi, Olga, Bury, Wojciech, Mondloch, Joseph E., Hupp, Joseph T., Farha, Omar K., 2014. Solvent-assisted linker exchange: An alternative to the de novo synthesis of unattainable metal–organic frameworks. Angew. Chem. Int. Ed. 53 (18), 4530–4540.

Karniadakis, George Em, Kevrekidis, Ioannis G., Lu, Lu, Perdikaris, Paris, Wang, Sifan, Yang, Liu, 2021. Physics-informed machine learning. Nature Rev. Phys. 3 (6), 422–440.

Krenn, Mario, Häse, Florian, Nigam, AkshatKumar, Friederich, Pascal, Aspuru-Guzik, Alan, 2020. Self-referencing embedded strings (SELFIES): A 100% robust molecular string representation. Mach. Learn.: Sci. Technol. 1 (4), 045024.

Kumar, Pawan, Anand, Bhaskar, Tsang, Yiu Fai, Kim, Ki-Hyun, Khullar, Sadhika, Wang, Bo, 2019. Regeneration, degradation, and toxicity effect of MOFs: Opportunities and challenges. Environ. Res. 176, 108488.

Lalonde, Marianne, Bury, Wojciech, Karagiaridi, Olga, Brown, Zachary, Hupp, Joseph T., Farha, Omar K., 2013. Transmetalation: Routes to metal exchange within metal–organic frameworks. J. Mater. Chem. A 1 (18), 5453–5468.

Landis, David D., Hummelshøj, Jens S., Nestorov, Svetlozar, Greeley, Jeff, Dułak, Marcin, Bligaard, Thomas, Nørskov, Jens K., Jacobsen, Karsten W., 2012. The computational materials repository. Comput. Sci. Eng. 14 (6), 51–57.

Leperi, Karson T., Chung, Yongchul G., You, Fengqi, Snurr, Randall Q., 2019. Development of a general evaluation metric for rapid screening of adsorbent materials for postcombustion CO2 capture. ACS Sustain. Chem. Eng. 7 (13), 11529–11539.

Li, Hao, Li, Libo, Lin, Rui-Biao, Zhou, Wei, Zhang, Zhangjing, Xiang, Shengchang, Chen, Banglin, 2019. Porous metal-organic frameworks for gas storage and separation: Status and challenges. EnergyChem 1 (1), 100006.

Li, Hao, Wang, Kecheng, Sun, Yujia, Lollar, Christina T., Li, Jialuo, Zhou, Hong-Cai, 2018. Recent advances in gas storage and separation using metal–organic frameworks. Mater. Today 21 (2), 108–121.

Liao, Pei-Qin, Shen, Jian-Qiang, Zhang, Jie-Peng, 2018. Metal–organic frameworks for electrocatalysis. Coord. Chem. Rev. 373, 22–48.

Ludwig, Alfred, 2019. Discovery of new materials using combinatorial synthesis and high-throughput characterization of thin-film materials libraries combined with computational methods. NPJ Comput. Mater. 5 (1), 1–7.

Lyu, Hao, Ji, Zhe, Wuttke, Stefan, Yaghi, Omar M., 2020. Digital reticular chemistry. Chemistry 6 (9), 2219–2241.

Lyu, Jiafei, Zhang, Xuan, Otake, Ken-ichi, Wang, Xingjie, Li, Peng, Li, Zhanyong, Chen, Zhijie, Zhang, Yuanyuan, Wasson, Megan C., Yang, Ying, et al., 2019. Topology and porosity control of metal–organic frameworks through linker functionalization. Chem. Sci. 10 (4), 1186–1192.

Ma, Ruimin, Colon, Yamil J., Luo, Tengfei, 2020. Transfer learning study of gas adsorption in metal–organic frameworks. ACS Appl. Mater. Interfaces 12 (30), 34041–34048.

Majumdar, Sauradeep, Moosavi, Seyed Mohamad, Jablonka, Kevin Maik, Ongari, Daniele, Smit, Berend, 2021. Diversifying databases of metal organic frameworks for high-throughput computational screening. ACS Appl. Mater. Interfaces.

Moghadam, Peyman Z., Islamoglu, Timur, Goswami, Subhadip, Exley, Jason, Fantham, Marcus, Kaminski, Clemens F., Snurr, Randall Q., Farha, Omar K., Fairen-Jimenez, David, 2018. Computer-aided discovery of a metal–organic framework with superior oxygen uptake. Nature Commun. 9 (1), 1–8.

Montoya, Joseph H., Aykol, Muratahan, Anapolsky, Abraham, Gopal, Chirranjeevi B., Herring, Patrick K., Hummelshøj, Jens S., Hung, Linda, Kwon, Ha-Kyung, Schweigert, Daniel, Sun, Shijing, et al., 2022. Toward autonomous materials research: Recent progress and future challenges. Appl. Phys. Rev. 9 (1), 011405.

Moosavi, Seyed Mohamad, Nandy, Aditya, Jablonka, Kevin Maik, Ongari, Daniele, Janet, Jon Paul, Boyd, Peter G., Lee, Yongjin, Smit, Berend, Kulik, Heather J., 2020. Understanding the diversity of the metal-organic framework ecosystem. Nature Commun. 11 (1), 1–10.

Mukherjee, Krishnendu, Colón, Yamil J., 2021. Machine learning and descriptor selection for the computational discovery of metal-organic frameworks. Mol. Simul. 1–21.

Nandy, Aditya, Duan, Chenru, Taylor, Michael G., Liu, Fang, Steeves, Adam H., Kulik, Heather J., 2021a. Computational discovery of transition-metal complexes: From high-throughput screening to machine learning. Chem. Rev. 121 (16), 9927–10000.

Nandy, Aditya, Terrones, Gianmarco, Arunachalam, Naveen, Duan, Chenru, Kastner, David W., Kulik, Heather J., 2021b. MOFSimplify: Machine learning models with extracted stability data of three thousand metal-organic frameworks. arXiv preprint arXiv:2109.08098.

Nazarian, Dalar, Camp, Jeffrey S., Chung, Yongchul G., Snurr, Randall Q., Sholl, David S., 2017. Large-scale refinement of metal- organic framework structures using density functional theory. Chem. Mater. 29 (6), 2521–2528.

Nejahi, Younes, Barhaghi, Mohammad Soroush, Mick, Jason, Jackman, Brock, Rushaidat, Kamel, Li, Yuanzhe, Schwiebert, Loren, Potoff, Jeffrey, 2019. GOMC: GPU optimized Monte Carlo for the simulation of phase equilibria and physical properties of complex fluids. SoftwareX 9, 20–27.

Nicholas, Thomas C., Alexandrov, Eugeny V., Blatov, Vladislav A., Shevchenko, Alexander P., Proserpio, Davide M., Goodwin, Andrew L., Deringer, Volker L., 2021. Visualization and quantification of geometric diversity in metal–organic frameworks. Chem. Mater. 33 (21), 8289–8300.

Ongari, Daniele, Talirz, Leopold, Smit, Berend, 2020. Too many materials and too many applications: An experimental problem waiting for a computational solution. ACS Cent. Sci. 6 (11), 1890–1900.

Park, Hyunsoo, Kang, Yeonghun, Choe, Wonyoung, Kim, Jihan, 2021. Mining insights on metal-organic framework synthesis from scientific literature texts. arXiv preprint arXiv:2108.13590.

Park, Hyunsoo, Kang, Yeonghun, Choe, Wonyoung, Kim, Jihan, 2022. Mining insights on metal–organic framework synthesis from scientific literature texts. J. Chem. Inf. Model. 62 (5), 1190–1198.

Peng, Jiayu, Schwalbe-Koda, Daniel, Akkiraju, Karthik, Xie, Tian, Giordano, Livia, Yu, Yang, Eom, C. John, Lunger, Jaclyn R., Zheng, Daniel J., Rao, Reshma R., et al., 2022. Human-and machine-centred designs of molecules and materials for sustainability and decarbonization. Nature Rev. Mater. 1–19.

Pizzi, Giovanni, Cepellotti, Andrea, Sabatini, Riccardo, Marzari, Nicola, Kozinsky, Boris, 2016. AiiDA: Automated interactive infrastructure and database for computational science. Comput. Mater. Sci. 111, 218–230.

Polat, H. Mert, Kavak, Safiyye, Kulak, Harun, Uzun, Alper, Keskin, Seda, 2020. CO2 separation from flue gas mixture using [BMIM][BF4]/MOF composites: Linking high-throughput computational screening with experiments. Chem. Eng. J. 394, 124916.

Qazi, Atika, Hussain, Fayaz, Rahim, Nasrudin A.B.D., Hardaker, Glenn, Alghazzawi, Daniyal, Shaban, Khaled, Haruna, Khalid, 2019. Towards sustainable energy: A systematic review of renewable energy sources, technologies, and public opinions. IEEE Access 7, 63837–63851.

Reddy, Ch Venkata, Reddy, Kakarla Raghava, Harish, V.V.N. al, Shim, Jaesool, Shankar, M.V., Shetti, Nagaraj P., Aminabhavi, Tejraj M., 2020. Metal-organic frameworks (MOFs)-based efficient heterogeneous photocatalysts: Synthesis, properties and its applications in photocatalytic hydrogen generation, CO2 reduction and photodegradation of organic dyes. Int. J. Hydrogen Energy 45 (13), 7656–7679.

Rosen, Andrew S., Iyer, Shaelyn M., Ray, Debmalya, Yao, Zhenpeng, Aspuru-Guzik, Alán, Gagliardi, Laura, Notestein, Justin M., Snurr, Randall Q., 2021. Machine learning the quantum-chemical properties of metal–organic frameworks for accelerated materials discovery. Matter 4 (5), 1578–1597.

Rosen, Andrew S., Notestein, Justin M., Snurr, Randall Q., 2022. Realizing the data-driven, computational discovery of metal-organic framework catalysts. Curr. Opin. Chem. Eng. 35, 100760.

Sarkisov, Lev, Martin, Richard L., Haranczyk, Maciej, Smit, Berend, 2014. On the flexibility of metal–organic frameworks. J. Am. Chem. Soc. 136 (6), 2228–2231.

Satorras, Victor Garcia, Hoogeboom, Emiel, Welling, Max, 2021. E (n) equivariant graph neural networks. arXiv preprint arXiv:2102.09844.

Scheffler, Matthias, Aeschlimann, Martin, Albrecht, Martin, Bereau, Tristan, Bungartz, Hans-Joachim, Felser, Claudia, Greiner, Mark, Groß, Axel, Koch, Christoph T., Kremer, Kurt, et al., 2022. FAIR data enabling new horizons for materials research. Nature 604 (7907), 635–642.

Severino, Maria Inês, Gkaniatsou, Effrosyni, Nouar, Farid, Pinto, Moisés L, Serre, Christian, 2021. MOFs industrialization: A complete assessment of production costs. Faraday Discuss. 231, 326–341.

Shi, Zenan, Yuan, Xueying, Yan, Yaling, Tang, Yuanlin, Li, Junjie, Liang, Hong, Tong, Lianpeng, Qiao, Zhiwei, 2021. Techno-economic analysis of metal–organic frameworks for adsorption heat pumps/chillers: From directional computational screening, machine learning to experiment. J. Mater. Chem. A 9 (12), 7656–7666.

Shirts, Michael R, Mobley, David L, Chodera, John, Wang, Lee-Ping, Gilson, Michael K, 2019. The open force field initiative: Better force fields through open, data-driven science. In: 2019 AIChE Annual Meeting. AIChE.

Sicilia, Miguel-Angel, García-Barriocanal, Elena, Sánchez-Alonso, Salvador, 2017. Community curation in open dataset repositories: Insights from zenodo. Procedia Comput. Sci. 106, 54–60.

Smith, Justin S., Nebgen, Benjamin T., Zubatyuk, Roman, Lubbers, Nicholas, Devereux, Christian, Barros, Kipton, Tretiak, Sergei, Isayev, Olexandr, Roitberg, Adrian E., 2019. Approaching coupled cluster accuracy with a general-purpose neural network potential through transfer learning. Nature Commun. 10 (1), 1–8.

Staubitz, Thomas, Klement, Hauke, Teusner, Ralf, Renz, Jan, Meinel, Christoph, 2016. CodeOcean-a versatile platform for practical programming excercises in online environments. In: 2016 IEEE Global Engineering Education Conference (EDUCON). IEEE, pp. 314–323.

Stock, Norbert, Biswas, Shyam, 2012. Synthesis of metal-organic frameworks (MOFs): Routes to various MOF topologies, morphologies, and composites. Chem. Rev. 112 (2), 933–969.

Sturluson, Arni, Huynh, Melanie T., Kaija, Alec R., Laird, Caleb, Yoon, Sunghyun, Hou, Feier, Feng, Zhenxing, Wilmer, Christopher E., Colón, Yamil J., Chung, Yongchul G., et al., 2019. The role of molecular modelling and simulation in the discovery and deployment of metal-organic frameworks for gas storage and separation. Mol. Simul. 45 (14–15), 1082–1121.

Sturluson, Arni, Raza, Ali, McConachie, Grant D., Siderius, Daniel, Fern, Xiaoli, Simon, Cory, 2021. A recommendation system to predict missing adsorption properties of nanoporous materials.

Subramanian Balashankar, Vishal, Rajendran, Arvind, 2019. Process optimization-based screening of zeolites for post-combustion CO2 capture by vacuum swing adsorption. ACS Sustain. Chem. Eng. (ISSN: 21680485) 7 (21), 17747–17755. http://dx.doi.org/10.1021/acssuschemeng.9b04124, URL www.hypotheticalzeolites.net/database/deem/.

Suh, Bong Lim, Chong, Sanggyu, Kim, Jihan, 2019. Photochemically induced water harvesting in metal–organic framework. ACS Sustain. Chem. Eng. 7 (19), 15854–15859.

Sun, Yangzesheng, DeJaco, Robert F., Li, Zhao, Tang, Dai, Glante, Stephan, Sholl, David S., Colina, Coray M., Snurr, Randall Q., Thommes, Matthias, Hartmann, Martin, et al., 2021. Fingerprinting diverse nanoporous materials for optimal hydrogen storage conditions using meta-learning. Sci. Adv. 7 (30), eabg3983.

Taddei, Marco, Petit, Camille, 2021. Engineering metal–organic frameworks for adsorption-based gas separations: From process to atomic scale. Mol. Syst. Des. Eng. 6 (11), 841–875.

Talin, A. Alec, Centrone, Andrea, Ford, Alexandra C., Foster, Michael E., Stavila, Vitalie, Haney, Paul, Kinney, R. Adam, Szalai, Veronika, El Gabaly, Farid, Yoon, Heayoung P., et al., 2014. Tunable electrical conductivity in metal-organic framework thin-film devices. Science 343 (6166), 66–69.

Talirz, Leopold, Kumbhar, Snehal, Passaro, Elsa, Yakutovich, Aliaksandr V., Granata, Valeria, Gargiulo, Fernando, Borelli, Marco, Uhrin, Martin, Huber, Sebastiaan P., Zoupanos, Spyros, et al., 2020. Materials cloud, a platform for open computational science. Sci. Data 7 (1), 1–12.

Tshitoyan, Vahe, Dagdelen, John, Weston, Leigh, Dunn, Alexander, Rong, Ziqin, Kononova, Olga, Persson, Kristin A., Ceder, Gerbrand, Jain, Anubhav, 2019. Unsupervised word embeddings capture latent knowledge from materials science literature. Nature 571 (7763), 95–98.

Wang, Yajie, Xue, Pu, Cao, Mingfeng, Yu, Tianhao, Lane, Stephan T., Zhao, Huimin, 2021. Directed evolution: Methodologies and applications. Chem. Rev. 121 (20), 12384–12444.

Weininger, David, 1988. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. J. Chem. Inf. Comput. Sci. 28 (1), 31–36.

Wilmer, Christopher E., Leaf, Michael, Lee, Chang Yeon, Farha, Omar K., Hauser, Brad G., Hupp, Joseph T., Snurr, Randall Q., 2012. Large-scale screening of hypothetical metal–organic frameworks. Nature Chem. 4 (2), 83–89.

Witman, Matthew, Ling, Sanliang, Anderson, Samantha, Tong, Lianheng, Stylianou, Kyriakos C., Slater, Ben, Smit, Berend, Haranczyk, Maciej, 2016. In silico design and screening of hypothetical MOF-74 analogs and their experimental synthesis. Chem. Sci. 7 (9), 6263–6272.

Wolf, Thomas, Debut, Lysandre, Sanh, Victor, Chaumond, Julien, Delangue, Clement, Moi, Anthony, Cistac, Pierric, Rault, Tim, Louf, Rémi, Funtowicz, Morgan, et al., 2019. Huggingface's transformers: State-of-the-art natural language processing. arXiv preprint arXiv:1910.03771.

Woodliffe, John Luke, Ferrari, Rebecca S., Ahmed, Ifty, Laybourn, Andrea, 2021. Evaluating the purification and activation of metal-organic frameworks from a technical and circular economy perspective. Coord. Chem. Rev. 428, 213578.

Xue, Dezhen, Balachandran, Prasanna V., Hogden, John, Theiler, James, Xue, Deqing, Lookman, Turab, 2016. Accelerated search for materials with targeted properties by adaptive design. Nature Commun. 7 (1), 1–9.

Yao, Zhenpeng, Sánchez-Lengeling, Benjamín, Bobbitt, N. Scott, Bucior, Benjamin J., Kumar, Sai Govind Hari, Collins, Sean P., Burns, Thomas, Woo, Tom K., Farha, Omar K., Snurr, Randall Q., et al., 2021. Inverse design of nanoporous crystalline reticular materials with deep generative models. Nat. Mach. Intell. 3 (1), 76–86.

Zhang, Xu, Chen, An, Zhong, Ming, Zhang, Zihe, Zhang, Xin, Zhou, Zhen, Bu, Xian-He, 2019a. Metal–organic frameworks (MOFs) and MOF-derived materials for energy storage and conversion. Electrochem. Energy Rev. 2 (1), 29–104.

Zhang, Xiangyu, Zhang, Kexin, Lee, Yongjin, 2019b. Machine learning enabled tailor-made design of application-specific metal–organic frameworks. ACS Appl. Mater. Interfaces 12 (1), 734–743.

Zhou, Musen, Wu, Jianzhong, 2020. A GPU implementation of classical density functional theory for rapid prediction of gas adsorption in nanoporous materials. J. Chem. Phys. 153 (7), 074101.