Coordination Chemistry Reviews 484 (2023) 215112

Contents lists available at ScienceDirect

Coordination Chemistry Reviews

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

Review

Recent advances in computational modeling of MOFs: From molecular simulations to machine learning

Hakan Demir, Hilal Daglar [1], Hasan Can Gulbalkan [1], Gokhan Onder Aksu [1], Seda Keskin *

Department of Chemical and Biological Engineering, Koc University, 34450 Istanbul, Turkey

ARTICLE INFO

Article history:
Received 15 November 2022
Received in revised form 22 February 2023
Accepted 5 March 2023
Available online 21 March 2023

Keywords:
Metal organic framework
Molecular simulation
Computational screening
Big data
Machine learning
Data science

ABSTRACT

The reticular chemistry of metal–organic frameworks (MOFs) allows for the generation of an almost boundless number of materials some of which can be a substitute for the traditionally used porous materials in various fields including gas storage and separation, catalysis, drug storage and delivery. The number of MOFs and their potential applications are growing so quickly that, when novel MOFs are synthesized, testing them for all possible applications is not practical. High-throughput computational screening approaches based on molecular simulations of materials have been widely used to investigate MOFs and identify the optimal MOFs for a specific application. Despite the growing computational resources, given the enormous MOF material space, computational identification of promising MOFs requires more efficient approaches in terms of time and effort. Leveraging data-driven science techniques can offer key benefits such as accelerated MOF design and discovery pathways via the establishment of machine learning (ML) models and interpretation of complex structure-performance relationships that can reach beyond expert intuition. In this review, we present key scientific breakthroughs that propelled computational modeling of MOFs and discuss the state-of-the-art approaches extending from molecular simulations to ML algorithms. Finally, we provide our perspective on the potential opportunities and challenges for the future of big data-driven MOF design and discovery.
© 2023 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Contents

1. Background . . . 2
2. A synopsis of the milestones in MOF research using traditional methods . . . 3
3. Linking scientific paradigms to implement data-driven material research . . . 6
   3.1. Material representation and data collection methods for ML studies . . . 7
   3.2. Recent applications of ML methods in MOF research . . . 8
      3.2.1. Predicting synthesis conditions . . . 8
      3.2.2. Predicting stability . . . 9
      3.2.3. Predicting guest accessibility . . . 10
      3.2.4. Inverse material design . . . 11
      3.2.5. Universal models for different applications . . . 11
      3.2.6. Bridging ML with experiments . . . 12
   3.3. Developing ML workflows: from datasets up to publication . . . 13
4. Outlook . . . 14
Declaration of Competing Interest . . . 15
Acknowledgements . . . 15
References . . . 15

* Corresponding author.
E-mail address: skeskin@ku.edu.tr (S. Keskin).
[1] These authors contributed equally.

https://doi.org/10.1016/j.ccr.2023.215112
0010-8545/© 2023 The Authors. Published by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Background

Porous materials, either in amorphous or crystalline form, comprise extensive families of structures such as metal–organic frameworks (MOFs), covalent-organic frameworks (COFs), zeolites, and activated carbons [1]. Since some of these porous structures possess a broad physical and chemical diversity with the ability to selectively separate molecules based on their pore size and chemistry, they have been widely used in many different applications ranging from catalysis to medicine [2]. The tunability of chemistry, porosity, and surface area of some classes of porous materials can enable the development of new materials bringing about more energy efficient, environmentally friendly, and low-cost technologies. MOFs are a relatively new generation of crystalline porous materials with favorable properties including but not limited to very high surface areas (up to ~10,000 m^2/g) [3], large variety in their porosity and pore sizes, low densities, and chemical tunability. As a result, MOFs have been extensively studied in a wide variety of fields, mostly for gas storage and separation [4,5], catalysis [6,7], biomedicine (e.g., drug storage and delivery) [8–11], chemical sensing [12], electrical conductivity [13], and light harvesting [14].

With the recent achievements in experimental synthesis and modification techniques developed for MOFs, their structural and chemical properties can be facilely altered through introducing functional groups [15], exchanging metal and/or ligand [16], encapsulation of nano-sized guests [17], and interpenetration of frameworks [18]. The number of synthesized MOFs has immensely increased and surpassed 100,000, whose crystallographic information files (CIFs) have been deposited into the Cambridge Structural Database (CSD) [19]. Analogous to the experiments, hypothetical MOFs (hMOFs) have been computationally generated by looping through possible combinations of building units in various topologies. While hMOFs are much larger in number compared to the synthesized MOFs, only a few hMOFs has been synthesized. Fig. 1 **(a)** represents the enormous spectrum of MOF material space that can be experimentally synthesized and computationally generated. Amongst the huge number of MOFs, there can be many materials potentially offering solutions to important societal problems such as climate change, environmental pollution, energy efficiency, and drug therapy. However, considering more than trillions of MOFs that can be constructed [20], the number of materials has become far larger than the "tractable" number for which trial-and-error experiments and brute-force molecular simulations can be carried out in a reasonable amount of time.

Molecular simulation methods (e.g., Grand Canonical Monte Carlo (GCMC), equilibrium molecular dynamics (EMD) and non-equilibrium molecular dynamics (NEMD) simulations) have been generally used to compute gas adsorption and separation in MOFs. While early molecular simulation studies focused on only a few MOFs, more recent studies probed a much larger number of MOFs thanks to the recently developed high-throughput computational screening (HTCS) approaches. The main goal of employing HTCS approaches is to evaluate the potential of a large set of structures for an application in a time-efficient manner and eventually identify the most promising materials to guide the experimental efforts. Thus, it can be anticipated that the HTCS studies will hugely benefit from the deployment of bigger computational resources that can produce even bigger data than done so far. Yet, as the number of MOFs increases exponentially, performing molecular simulations for all of them would be highly inefficient. To conduct material research in a time and cost-effective manner, data-driven approaches cutting down the number of simulations and experiments performed for new material discovery are highly warranted. Even if it was plausible to carry out molecular simulations for every material and application, analyzing the complex "big data" with high volume and dimensionality would still be cumbersome without data science techniques. This is generally described as the problem of "finding the needle in the haystack" since the traditional approaches solely based on the use of empirical, theoretical, or simulation data involve a brute-force approach (or at best some expert intuition) with limited data guidance. Conversely, the integration of data science into material research is deemed as a scientific paradigm shift enabling a new approach where data is systematically retrieved and processed from experimental, theoretical, simulation sources to establish machine learning (ML) based workflows [21].

Data/text mining techniques, supervised, and unsupervised ML models can facilitate understanding of the data since they can help structure the data, establish structure–property-performance relations, and generate deep generative models [22]. ML is a branch of artificial intelligence (AI) that can analyze the relationship between input variables and output to predict target properties of new materials or cluster unlabeled data sets [23]. The ML

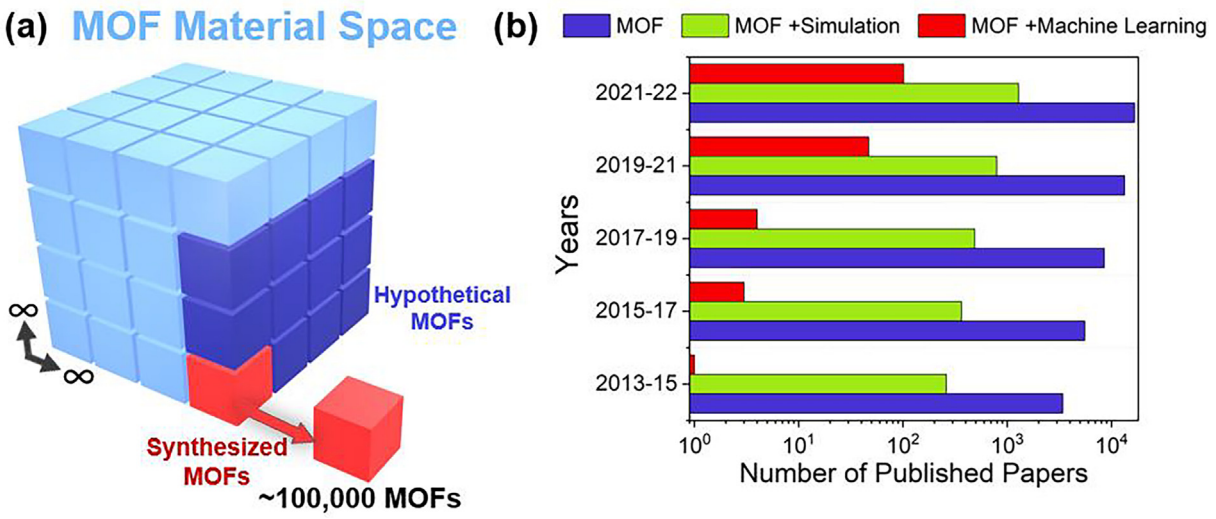


Fig. 1. (a) The large cube represents the MOF space with an almost infinite number of materials, including synthesized and hypothetical MOFs. (b) The number of published papers having keywords (i) only "MOF", (ii) "MOF" and "Simulation", (iii) "MOF" and "Machine Learning" in their titles and abstract. Data were retrieved from the Web of Science on 17th October 2022.

models developed for a target application can offer substantial benefits for elucidating the complex structure–property relationships of MOFs. In addition, the development of ML models can significantly boost the material discovery process, allowing performance predictions to be made within seconds for thousands of materials. The ML algorithms developed for MOFs to date have primarily focused on predicting their gas storage and separation properties [5–10], colors [11], oxidation states [12], heat capacities [13], in addition to assigning partial charges to MOF atoms [24,25], and predicting their performance as heat pumps [15]. The number of published articles analyzing MOFs with ML algorithms dramatically increased from one in 2013–2015 to >50 in 2019–2021 reaching close to 100 in 2021–2022, as shown in Fig. 1**(b)**.

Motivated by the huge potential of incorporating ML techniques into computational modelling of MOFs, in this review, we highlight different approaches combining molecular simulations and ML methods to accurately assess both the properties and potential of MOFs for a variety of processes such as gas storage, separation, catalysis in addition to predicting synthesizability, guest accessibility and stability of MOFs. We first present the chronological framework for the computational modelling of MOFs, starting from molecular simulations of a handful of materials to the implementation of HTCS approach and finally ML-guided modeling of MOFs. Moving forward, we discuss in detail some of the recent innovative studies that contribute to the material design by defining new material representations, generating diverse material databases, applying new learning methods, and constructing the ML workflows. Finally, we present our perspective on the existing/foreseen challenges and opportunities for the AI-aided MOF research to accelerate the discovery of novel and high-performing MOFs.

2. A synopsis of the milestones in MOF research using traditional methods

The last several decades have witnessed significant progress in experimental and computational approaches used for MOF research. Fig. 2 represents the brief history of the major developments in the field starting from the early experiments and molecular simulations and extending to HTCS and ML studies. The MOF research has been heavily stimulated by the breakthroughs made in the MOF synthesis together with proven permanent porosity of MOFs in 1990′s enabling their practical use [26,27]. Between 1999 and 2002, after the synthesis of highly porous MOFs (such as HKUST-1 [28], MOF-2 [29], MOF-3 [29], MOF-4 [29], MOF-5 [26,29], MOF-14 [30], QuartzMOF-1 [31], QuartzMOF-2 [31]), the field gained a lot of attention. Due to their high surface areas and porosities, MOFs achieved high H_2, CH_4, CO_2 uptakes, outperforming traditional porous materials such as zeolites and activated carbons [32]. While environmental concerns have grown over conventional transportation fuel usage, CH_4 has emerged as an alternative fuel [33]. Peng et al. 33] showed that HKUST-1 has an exceptional CH_4 storage capacity of 267 cm^3 (STP) cm^{-3} at 65 bar, 298 K as the first MOF to exceed the initial U.S. Department of Energy (DOE) target (200 cm^3 (STP) cm^{-3}) at the time of publication. Capturing CO_2 from large volumes of gas streams is one of the industrially relevant challenges which can be achieved through adsorption in MOFs [34]. Yaghi and coworkers [34] reported the high CO_2 uptake of MOF-177 as 320 cm^3 (STP) cm^{-3} up to 40 bar, 298 K outperforming other MOFs such as IRMOF-1, -3, -6, -11, MOF-74 and Cu-BTC. Hydrogen is considered as a promising energy carrier for the transportation sector since it has clean combustion characteristics [35]. Wong-Foy et al. [36] reported H_2 adsorption capacity of different MOFs (HKUST-1, MOF-74, MOF-177, IRMOF-1, IRMOF-6, IRMOF-11, and IRMOF-20 where IR stands for isoreticular) under various pressures up to 80 bar, 77 K and showed that MOF-177 achieved a high excess H_2 uptake (~80 mg/g). Yan et al. [37] investigated the H_2 storage capacity of NOTT-112 and reported an excess H_2 uptake of 76 mg/g at 77 K and 70 bar. Furukawa et al. [3] examined a series of MOFs (MOF-200, -205, and -210) for H_2 storage and demonstrated that MOF-210 has an excess H_2 uptake of 86 mg/g at 77 K and 70 bar, surpassing that of MOF-177. Hupp and coworkers [38] reported an exceptional excess H_2 uptake of 164 mg/g for NU-100 at the same conditions. These results show that MOFs are promising candidates for gas storage applications.

Thanks to their tunable pore sizes, which are fundamental for an efficient separation, the design and development of MOFs for gas separation applications has also been another main research field [39,40]. Mason et al. [41] reported the CO_2/N_2 selectivity of Mg-MOF-74 as 148.1 at 1 bar, 323 K, which is higher than that of

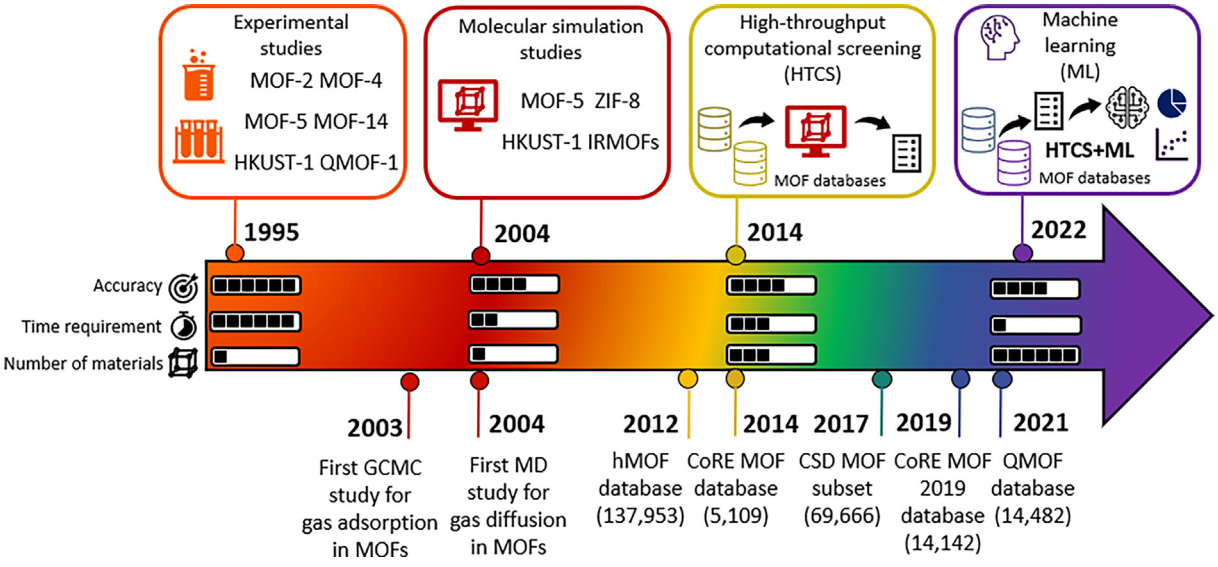


Fig. 2. A timeline of major milestones in computational MOF research. The number of black boxes inside the arrow comparatively represents the accuracy of the corresponding method, time requirement to obtain target properties, and the number of materials that can be studied using that method (experiments, molecular simulations, HTCS and HTCS + ML). The numbers in the parenthesis for the databases represent the number of materials in these databases.

zeolite NaX (87.4). Ma and coworkers [42] synthesized a microporous MOF, MMCF-1, and estimated its CO_2/N_2 selectivity as 114 at 273 K, 1 bar. Bae et al. [43] studied carborane-based MOFs for CO_2/CH_4 separation and showed that they can achieve high CO_2 selectivities over CH_4 (~17). Herm et al. [44] investigated a number of MOFs, such as MOF-177 and Mg-MOF-74, for CO_2/H_2 separation at pressure swing adsorption (PSA) condition (up to 40 bar and 313 K) and demonstrated that these MOFs exhibit high CO_2/H_2 selectivities (~860) outperforming the selectivities of activated carbon JX101 (~100) and zeolite 13X (400) at 1 bar, 313 K. These experimental studies proved that MOFs have huge potential for gas separation applications making them strong alternatives to traditional materials, particularly for CO_2 capture.

Aside from gas adsorption and separation applications, the massive porous space of MOFs can be exploited for storing and delivering drug molecules. Several properties of MOFs, such as the presence of metals with low toxicity (Fe, Zn, Ca, Mg, etc.) and ligands (carboxylic and phosphonic acids), biodegradability in the aqueous medium, ability to respond to external stimuli, and tailorable chemistry make them potential candidates for drug storage and delivery applications [45–49]. One of the first usage of MOFs as drug carriers was demonstrated for MIL-100(Cr) and MIL-101(Cr) which adsorb high amounts (0.35 and 1.38 g/g MOF) of ibuprofen (IBU) thanks to their large pore sizes and high pore volumes [48]. Complete release of the drug was achieved after 3 days for MIL-100(Cr) and 6 days for MIL-101(Cr) due to the higher proportion of the aromatic rings in the latter, which increases the interactions between IBU and the framework. Horcajada et al. [49] examined the impact of flexibility of MOFs (MIL-53(Cr) and MIL-53(Fe)) on IBU delivery and showed that frameworks adapt their pore openings to the drug size to optimize the drug-matrix interactions. Although these early studies mostly focused on MOFs having metals which may have biocompatibility issues due to their toxicity to humans [50], they serve as proof-of-concept studies for the use of MOFs as drug delivery platforms. The potential toxicity problems associated with MOF degradation in the aqueous media fueled the search for alternative MOFs. For example, Bio-MIL-1, consisting of Fe metal and nicotinic acid was synthesized, and the importance of using a therapeutic agent as a framework component in a MOF has been demonstrated [51]. An et al. [52] synthesized Bio-MOF-1 and reported the loading of procainamide, an antiarrhythmic drug, in this MOF as 0.22 g/g. These studies showed that MOFs can adsorb significant amounts of drugs and release them slowly or rapidly depending on the framework.

As Fig. 2 represents, molecular simulations of MOFs mainly started in the early 2000s. Based on theoretical models, molecular simulations can provide atomically detailed information on thermodynamic and kinetic properties of systems that may be hard or even impossible to acquire through experiments. They can also be deemed as a complementary tool to experiments to validate experimental findings and set up theoretical basis for the observed phenomena. Typically, it is less laborious to set up MOF models and carry out molecular simulations than testing MOFs in the wet lab rendering large-scale computational screening easier than experimental screening [53]. The success of molecular simulations relies on the precision of atomic positions as the atomic interactions hinge upon them. The high crystallinity of MOFs facilitates representing their atomic positions precisely while their high symmetry allows reducing MOFs to small primitive cells (typically having ~100 atoms/unit cell or less). These provided great advantages for the computational investigation of MOFs as the former eliminates atomic position-related inaccuracies in computed properties while the latter acts as a factor for cutting down the computational cost of calculations, rendering MOFs as a material class of interest for molecular simulations.

So far, numerous molecular simulation studies have been carried out for MOFs where properties such as guest adsorption amounts, isosteric heat of adsorption of guests, diffusivity of guests, and thermal conductivity of materials were the main focus [54]. Among various fields, small gas molecule adsorption/separation has been one of the most popular topics perhaps due to the successes of initial molecular simulations. The gas adsorption and diffusion in MOFs can be described using two different types of simulations (GCMC and MD simulations, respectively) [55]. Snurr and colleagues [56] performed the first GCMC simulation study of MOFs by computing the adsorption isotherms of CH_4 for IRMOF-1 and IRMOF-6 at 35 bar, 298 K and demonstrated that simulation results were consistent with those obtained from experimental measurements. In the following years, the number of MOFs simulated for gas separation has expanded. For instance, Wang [57] performed GCMC simulations to compute CH_4 storage of 10 MOFs at 35 bar, 298 K, and concluded that MOFs with high isosteric heats of adsorption at infinite dilution (characterizing energetic interaction between sorbate and sorbent), large surface areas, free volumes, and low densities can achieve high CH_4 storage capacities. Babarao et al. [58] utilized GCMC simulations to compute the CO_2 storage capacities of nine MOFs at 30 bar, 298 K and showed that the organic linker has a significant role to achieve high CO_2 adsorption ($\geq$40 mmol/g). Five distinct IRMOFs (IRMOF-1, -8, -10, -14, and -16) were examined by Düren and Snurr as potential adsorbents for the separation of hydrocarbons (C_4H_{10}/CH_4) via GCMC simulations [59]. Based on their findings, they proposed a hypothetical IRMOF structure (IRMOF-993) with high C_4H_{10}/CH_4 selectivities (up to 2720), outperforming selectivity of already-synthesized IRMOF-1 (up to 200). Babarao et al. [62] examined the adsorption and separation of the CO_2/CH_4 mixture in seven MOFs (Cu-BTC, PCN-6, PCN-6′ where PCN is porous coordination network, IRMOF-1, IRMOF-13, IRMOF-14 and soc-MOF) via GCMC simulations and revealed that the CO_2/CH_4 selectivity of soc-MOF can reach 36, which is one order of magnitude greater than the CO_2/CH_4 selectivities of IRMOF and PCN structures.

Fast kinetics of adsorption, governed by diffusion rates, is as important as high adsorption capacities. To computationally determine the transport properties of systems, Newton's law of motion is solved for a set of atoms in EMD simulations [61]. Compared to GCMC, EMD simulations are computationally much more demanding. Sarkisov et al. [62] performed the first EMD simulations for MOFs to compute self-diffusivities of CH_4, *n*-pentane (C_5H_{12}), *n*-hexane (C_6H_{14}), and *n*-heptane (C_7H_{16}) in IRMOF-1 at 300 K and showed that the self-diffusivities of *n*-alkanes in IRMOF-1 is comparable to those in the crystalline bipyridine systems. Skoulidas [63] performed EMD simulations to compute argon (Ar) diffusivities in HKUST-1 at 298 K and found out that the diffusivities are very similar to those in silica zeolites (ITQ-3, and ITQ-7). More detailed discussions on using molecular simulations for gas separation applications of MOFs can be found in a recent review [64].

Molecular simulations also have the potential to provide valuable insights into how drug molecules are stored and released in MOFs. Compared to the studies on gas adsorption and separation, there are fewer number of molecular simulation studies on MOFs for drug storage and delivery due to the immense computational cost of calculating a multitude of interactions of large drug molecules in the frameworks [47]. Early computational research focused on the adsorption and diffusion of IBU within the MIL (-47, -53, -101) [65,66] and bio-MOF family [67–69], UMCM-1 [65,69], HMOF-1 [66], CD-MOF-1 [68], and Mg-MOF-74 [67,68]. For example, Erucar and Keskin studied the effect of MOF flexibility on IBU, caffeine, and urea diffusion and showed that the slow diffusion of these molecules inside Bio-MOF-100 renders this material promising for drug delivery [67]. Snurr and coworkers

demonstrated that Bio-MOF-100 outperforms traditional zeolites and mesoporous silicas in terms of IBU capacity [68]. These studies are exemplary of efficiently identifying a promising material for drug storage and delivery using different computational methodologies. More recently, molecular simulations were performed to study storage and delivery of anticancer drugs in MOFs. The majority of the studies focused on methotrexate (MTX) [70], 5-fluorouracil (5-FU) [70–73], gemcitabine [74,75], doxorubicin [76,77] and busulfan [78]. Three new MOFs were synthesized by Liu et al. [71], which were found to be capable of adsorbing large amounts (0.40–0.45 g/g) of 5-FU, and GCMC simulation results were found to be in good agreement with the experimental measurements. Li et al. [73] synthesized and tested a MOF for 5-FU adsorption and used GCMC simulations to demonstrate that hydrogen bonding was the determining factor for drug loading inside the MOF. Storage of multiple anticancer drugs is important for combination therapy in cancer treatment. Erucar and Keskin studied the potential of 10 different MOF-74 structures for the coadsorption of 5-FU and MTX using molecular simulations [70]. MTX adsorption at low fugacity was favored over 5-FU adsorption since the former has a higher number of atoms and stronger interactions with MOFs, whereas 5-FU adsorption at high fugacity was favored over MTX due to entropic effects. More details about the computational methods (configurational-bias Monte Carlo (CBMC), GCMC, MD simulations and density functional theory (DFT) calculations) employed to compute drug storage and delivery in MOFs can be found in the literature [79].

Computational methods have been helpful to assess the catalytic activities of MOFs for which experimental methodologies are very costly to apply. MOFs were evaluated for thermocatalytic, photocatalytic, and electrocatalytic applications where DFT calculations are used to obtain reaction and activation energies [80–85]. Mao et al. [84] computed the electrocatalytic activities of titanium and copper-based MOFs for CO_2 reduction using DFT, and found that FeN_4-MOF has a very low limiting potential, obtained from the largest free energy change, outperforming several reported catalysts. Grau-Crespo et al. [85] designed a series of zeolitic imidazolate frameworks (ZIFs) and calculated their band gaps using DFT for visible-light water splitting and CO_2 reduction photocatalysis. Speybroeck and co-workers [86] utilized DFT calculations to examine the role of defects in catalysis and showed that defective UiO-66 structures are more stable compared to their non-defective counterparts due to the higher entropy of the former. They revealed that the defective structures have higher number of Lewis acidic sites exhibiting higher activity and selectivity for an isomerization reaction. These studies show the importance of employing computational tools to better understand novel MOF catalysts.

The use of molecular simulations for a large number of MOFs has become easier with the establishment of several MOF databases (DBs), as represented in Fig. 2**.** Cambridge Structural Database (CSD) is one of the main avenues to deposit experimentally synthesized CIFs of MOF structures. However, due to the presence of solvent molecules, disorders, and missing hydrogen (H) atoms in CIFs, a curation step is required before structures can be directly used in molecular simulations. To this end, Computation-Ready and Experimental MOF (CoRE MOF) database was built by Snurr et al. [87] in 2014, which consists of 5,109 experimentally synthesized MOFs refined from CSD by implementing several curation steps. In the second version of CoRE MOF, released in 2019, the number of MOFs available in this DB increased to 14,142 [88]. With the help of automated curation steps, the CSD MOF subset consisting of 69,666 materials, has also been made available in 2017 [89]. A detailed comparison of computation-ready MOF databases used in HTCS studies and the impact of MOF database selection on their performance assessment were also examined in the literature [90,91]. Additionally, thanks to new algorithms developed in the last decade, generating novel hypothetical MOFs (hMOFs) through building block assembly has become easier [92–96]. As shown in Fig. 2, Wilmer et al. [92] introduced the first hMOF database consisting of 137,953 MOFs in 2012. There are other hMOF databases constructed by Gomez-Gualdron et al. [93], and Boyd and Woo [94] by using different building blocks, topologies, and methodologies [95]. With the development of hypothetical databases, researchers are no longer restricted to only experimentally generated MOF structures and can benefit from a large pool of hMOFs with varied structural and chemical properties to identify the best materials for various applications.

As shown in Fig. 2, the research focus in the MOF field has shifted from purely experimental methods to HTCS which is a powerful approach to identify the best MOFs by calculating various performance metrics using the gas uptakes and diffusivities obtained from molecular simulations. Haldoupis et al. [96] computed the Henry's constants (representing gas affinities of frameworks at low pressure) and self-diffusion coefficients of CH_4 and H_2 in 504 MOFs and showed that MOFs can achieve large H_2/CH_4 membrane selectivities (up to 10^{12}) outperforming zeolites and polymers. In another study [97], they utilized the same group of MOFs and performed GCMC simulations to identify the best candidates for CO_2/N_2 and CO_2/CH_4 separations, and demonstrated that MOFs achieve very high CO_2/N_2 (up to 10^4) and CO_2/CH_4 (up to 10^6) adsorption selectivities. Watanabe and Sholl [98] selected 1163 MOFs from the CSD and identified promising MOFs for membrane-based CO_2/N_2 separation using GCMC and EMD simulations and showed that MOF membranes can achieve very high membrane selectivities (>800), outperforming traditional polymer membranes. Altintas et al. [99] screened a large CSD MOF subset consisting of 3816 MOFs to unlock their CO_2/N_2 and CO_2/CH_4 separation performances by performing GCMC simulations. It was demonstrated that the best candidates for selective CO_2 separation exhibit narrow pore sizes (3.8–5 Å), low surface areas (<1000 m^2/g), and high densities (>1 g/cm^3). Avci et al. [100] screened the same MOF DB by using GCMC and EMD simulations to compute both adsorption-based CO_2/H_2 and membrane-based H_2/CO_2 separation performances of MOFs. Based on the computed adsorption and membrane selectivities, 2137 (11) MOFs outperform 10-membered ring (cationic) zeolites at PSA condition and 899 MOFs exceed Robeson's upper bound [101] defined for polymeric membranes. Alternative evaluation metrics and process conditions have also been examined. Leperi et al. [102] screened CoRE MOF 2014 DB by utilizing GCMC simulations combined with a fractionated vacuum pressure swing adsorption (FVPSA) process model and identified 190 MOFs meeting purity (>90 %) and recovery (>90 %) requirements for an effective CO_2/N_2 separation process. It is worth noting that HTCS studies have focused on not only simple non-polar molecules (e.g., H_2, CH_4) but also polar ones (e.g., H_2O). Song et al. [103] conducted a HTCS study and identified 15 MOFs among 5109 CoRE MOFs with high selectivities of CO_2 over H_2O by performing GCMC simulations mimicking CO_2/H_2O and $CO_2/H_2O/N_2$ mixtures at 0.1–1 bar, 298 K.

Wilmer et al. [92] performed one of the largest HTCS to compute CH_4 storage capacities of 137,953 hMOFs, and found that over 300 hMOFs exceed the highest CH_4 storage capacities of experimentally synthesized MOFs. Bobbitt et al. [104] evaluated the same hMOF database for H_2 storage at 2 and 100 bar, 77 K using GCMC simulations, and demonstrated that hMOFs can achieve H_2 volumetric capacities up to 50 g/L, exceeding the DOE 2015 target (40 g/L). The same group also analyzed H_2 storage capacities of 2736 Zr-MOFs together with their functionalized versions using GCMC simulations at 5 and 100 bar, 296 K and showed that MOFs can attain very high deliverable capacities (up to 24 g/L) at room temperature conditions [105]. Siegel and coworkers [106] screened

almost 500,000 MOFs, including both experimentally synthesized MOFs and hMOFs, for H_2 storage using GCMC simulations and identified three novel MOFs (SNU-70, UMCM-9, and PCN-610/NU-100), outperforming IRMOF-20 (33.4 g/L).

HTCS studies assisted with DFT calculations have also proven to be very beneficial for identifying the promising MOFs for catalytic applications. For example, Vogiatzis et al. [107] identified a subset of 7 MOFs, which contain open metal site (Fe atom), among ~5000 CoRE MOFs and demonstrated that Fe-BTT can oxidize ethane to ethanol in the presence of N_2O by using DFT calculations. Investigating CoRE MOF DB, Rosen et al. [108] found promising MOFs for C—H bond activation of CH_4 by oxidation via a completely automated computational approach initializing the locations of small molecules at metal binding sites and performing structural relaxations for MOFs using periodic DFT calculations. In another study, this automated approach was used to screen 60 diverse MOF structures for catalytic activation of CH_4, and it was concluded that the relation between thermodynamic favorability and forming active sites with H-affinity can be used to identify the best MOFs, as this relation gives the active site stabilities and the barrier for C—H activation in CH_4 [109].

The generation of bigger data for MOFs would require more simulations incurring a considerable computational cost [110,111]. The introduction of quantitative structure–property relationships (QSPR) is an alternative and time-efficient approach to correlate a specific performance metric with various chemical and structural properties. Fernandez et al. [112] introduced one of the first QSPR studies by studying 137,953 hMOFs, and investigated the effect of structural properties of MOFs on their CH_4 storage performances. They showed that construction of different regression models based on the data of less than %10 of available hMOFs is very accurate in terms of CH_4 uptake prediction for the remaining material set. Fernandez et al. [113] used atomic property weighted radial distribution functions (AP-RDF), which is the weighted probability of counted number of the atom pairs at a distance from reference particle, and geometric properties to set up QSPR models predicting CO_2, N_2, and CH_4 uptakes of ~83,000 hMOFs refined from Wilmer's hMOF set. They showed that the models constructed from the refined database predict gas uptakes accurately. These studies proved that rather than taking on a brute-force approach, determining material performances through regression models can be more efficient.

3. Linking scientific paradigms to implement data-driven material research

HTCS approaches using molecular simulations and DFT calculations have been groundbreaking for accelerating the design of novel MOFs. However, traditional screening approaches using GCMC, EMD, or DFT-based calculations for each material may not be practical due to the huge and continuously growing number of MOFs and/or computationally demanding systems (i.e., drug loading in MOFs, bulky molecule separation in MOFs, and material properties requiring quantum chemical simulations such as band gap energy and charge density) [114]. This prompts the use of ML methods through which accurate predictive models can be established for a large set of materials [115]. The development of ML model is composed of four main components governing the model accuracy: selection of the material database, representation of materials, selection of the target data, and selection of ML algorithms. While choosing a material database for ML development may appear as an easy task, care should be taken to select a material set consisting of no/minimal structural errors to be able to obtain proper input and output data. Another crucial component for the ML model development is material featurization for which there exists different representations including structural, chemical, energetic features, strings, and images of the structures. As the third component, the target property of MOFs such as gas uptake or bulk properties is employed as the output data in ML models. The input and output data can be collected/generated in different ways such as text-mining the literature, performing molecular simulations, carrying out experiments or combining some/all. The final component for ML model development is the selection of the algorithm having a strong impact on the prediction accuracy.

Fig. 3 shows these four main components of ML model development: databases, features to represent materials, target data and ML algorithms. The relative sizes of the words below reflect their appearance frequency in the collected ML-assisted MOF studies. The most frequently studied MOF database for ML model development in the literature is the CoRE MOF database (used in > 30 studies out of 140) possibly because of having experimentally synthesized materials with diverse physical and chemical properties. Hypothetical MOF databases provide a much larger number of materials which can also be useful for training ML models. However, some of the hMOF databases have been constructed with a limited selection of node, linker and/or topology, hindering the structural and chemical diversity of materials that may result in bias and poor generalization ability of the models. Therefore, several ML studies [116–120] have focused on using different hMOF databases in which MOFs were constructed using different building blocks and topology such as hMOF-Wilmer [92], and hMOF-Boyd [94], ToBaCCo hMOFs [119]. For example, in a recent work, Lee et al. [20] created extremely diverse and large hMOF dataset using 648 node building blocks, 219 edge building blocks and 1775 topologies. While efforts to extend the scale of hMOFs are crucial to have reasonable starting structures for material investigation, databases involving both experimental and hypothetical structures can also be helpful as they can help distinguish structural disparities and performance limits of synthesized and hypothetical MOFs. For example, Quantum MOFs (QMOFs) [120] composed of both experimentally synthesized and computationally generated MOFs is a recently introduced database providing a more balanced structure set in terms of physical and chemical properties. The future studies focusing on databases such as QMOFs can reveal local disparities (e.g., interatomic distance, angle, torsion) in the shared building blocks across experimental and hypothetical MOFs which could help explain differences in structural features and/or performances.

In ML studies focusing on MOFs, the most frequently used descriptors are easily calculable structural properties such as surface area, pore limiting diameter [96] (PLD), density and porosity. Since the structural properties of MOFs might have poor correlation with some of the target properties (e.g., CO_2 uptake, electrical conductivity), different kinds of features (e.g., chemical, atom-centered) have been used to give more detailed representations of MOFs. For example, revised autocorrelation functions (RACs) has been recently introduced as a new kind of graph-based descriptor that defines the chemistry of building blocks of MOFs and used to predict MOFs' low-pressure CO_2 adsorption and deliverable CH_4 capacity [121]. Selected outputs of molecular simulations (such as heat of adsorption and Henry's constants for gas molecules) can also be used as input features of MOFs to train ML models [121]. For example, ML models were developed to predict N_2 uptakes of MOFs and the results showed that incorporating the heat of adsorption values into the list of features increased the prediction power significantly (R^2 going from 0.47 to 0.87) [122]. Adsorbate properties (e.g., kinetic diameter) and operation conditions for an application (e.g., pressure, temperature) can also be used as input features [123–127].

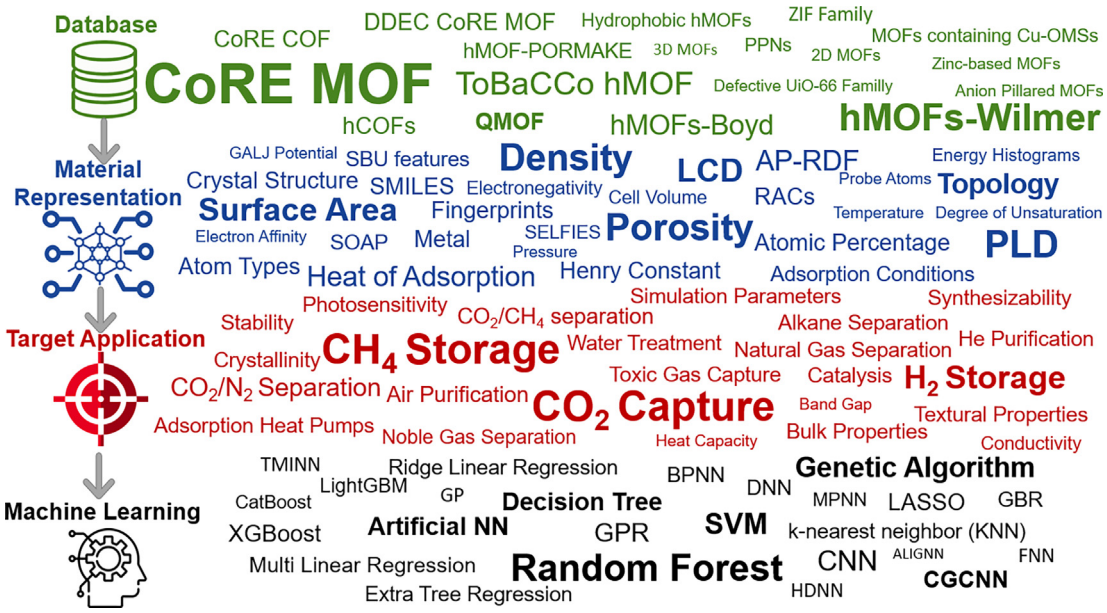


Fig. 3. Word cloud representations for the components of a ML model development process. The four main components are database, material representation, target data, and ML algorithm. The sizes of the words rely on how frequently the relevant words appear in 135 published studies between 2014 and 2022, for example, "random forest" has been used in 37 papers. Data were retrieved from the Web of Science on 17th October 2022 using the keywords of "MOF" and "Machine Learning". We note that words appearing very rarely in the collected studies are omitted.

Another crucial component of ML model development is to select the target application. Since the available data about gas storage and separation are huge in the literature, the majority of ML studies focused on predicting gas uptakes in MOFs exposed to either single-component and/or multicomponent gases (e.g., CH_4, H_2 gases and CO_2 containing mixtures) as shown in Fig. 3 [123,128–132]. In ML-assisted gas separation studies (e.g., flue gas, natural gas separation, air purification) where MOFs are employed as adsorbents and/or membranes, several performance metrics such as gas selectivity, working capacity, permeability are typically predicted by ML models. In recent studies, it has been shown that bulk properties of MOFs such as crystallinity, and stability of structure can also be predicted [131,132]. Among various types of ML approaches, supervised learning methods such as random forest (RF) and decision tree (DT), deep learning methods such as artificial neural network (ANN) have been extensively studied [133,134]. In addition, genetic algorithms can be used to solve optimization problems of ML motivated by genetics and natural selection [135]. They have been used for different purposes such as selection of features, tuning the parameters of ML algorithms, constructing new materials with desired properties [136–138]. The conventional ML models using supervised learning methods can perform on par or better than deep learning algorithms, especially when the amount of data is limited [139]. Therefore, supervised learning-based algorithms, especially RF, are predominant in ML studies as shown in Fig. 3. In the following sections, we will discuss some of the recent groundbreaking ML studies in the field of MOFs and highlight the potential of these studies for inverse material design.

3.1. Material representation and data collection methods for ML studies

There are many techniques to transform materials into machine-readable representations which can be subsequently used for ML-driven research. Chemistry, topology, and pore geometry have been commonly used to define MOFs as a collection of scalar/vectors. In addition to these structure-based representations, string-based representations can be alternatives to define MOFs and direct inverse design of material for a target application [140]. Following the introduction of SMILES (Simplified Molecular Input Line Entry System), string-based representations have been popular [141]. SMILES have been used to describe the molecular structures of various chemical species and their properties such as chirality and stereochemistry [142,143]. However, SMILES strings have weaknesses in terms of their syntactic and semantic robustness. For example, a SMILES string "CCC(OC)CC)SC(ON" has equal number of open and closed parenthesis, but it has a semantic error that the third and fourth parentheses do not match each other. SMILES strings are not unique representations of structures, which might be important to define the exact relationship between the structures and their targeted properties. To address these problems, different string-based representations have been developed such as DeepSMILES [144] (to eliminate paired parentheses using postfix notation) and InChI keys (to create unique representations) [145]. Snurr and coworkers [146] developed two systematic MOF identifiers (MOFid and MOFkey) which can extract SMILES strings and InChI keys of MOF building blocks, respectively, to be subsequently used for ML models. They also developed a web site [147] enabling users to upload the CIFs of MOFs and run the code to obtain MOFid/MOFkey interactively. In 2021, Ramprasad et al. [148] developed Linker SMILES Extraction (LSE), a scalable program extracting the SMILES string of each linker present in a MOF. They then constructed a model using SMILES as inputs and the results of GCMC simulations (CH_4 adsorption amounts at 35 bar) as outputs. The model predicted CH_4 uptake in MOFs with high accuracy (R^2: 0.994).

Recently, SELFIES (SELF-referencing Embedded Strings) was introduced as a robust string representation eliminating semantic and syntactic invalidities of other string-based representation methods. Aspuru-Guzik and coworkers [149] used SELFIES to construct a novel ML model which has the ability of both interpolating and extrapolating the chemical space of molecules, shown to compete well with previous ML models for molecular design. SELFIES representations of MOFs can be obtained as sets of representations derived for MOF building blocks by first breaking down MOFs into

building blocks and then generating SELFIES representations for each building block. However, SELFIES cannot fully represent macromolecules and periodic crystal structures yet implying that MOF deconstruction into molecular clusters is needed. Krenn et al. [150] proposed several future directions for the development of string-based representation and their applications in ML for cheminformatics such as generating crystal SELFIES, testing the human readability of SELFIES and translating between image-based and string-based representations. Feature extraction of crystal structure by image processing has been studied as another option to create ML-readable data of MOFs. Very recently, Sparks and coworkers [151] developed a Python package to represent the crystal structure as PNG files. The package can convert a MOF structure to a PNG file by encoding crystallographic information about unit cells such as atomic numbers, fractional coordinates, lengths and angles of unit cell, space group and cell volume for direct usage with image-based ML models such as convolutional neural network (CNN).

Obtaining data from the literature in the standardized format through various techniques is gaining popularity as it can significantly accelerate creating structured datasets and determining input–output relations [22]. Future efforts as to the data-driven research can significantly benefit from the use of machine-readable file formats facilitating the ML model development. As part of such efforts, recently, adsorption information file (AIF) format was proposed to substantially ease retrieving and standardizing experimental/simulation data for future ML model development and contribute to the adoption of Findability, Accessibility, Interoperability, and Reusability (FAIR) principles [152,153]. One main advantage of reporting adsorption data in AIF format compared to plotting the full adsorption isotherm figures is that it can hinder information/accuracy loss during digitization of data through point picking method [152]. Thus, using AIF format does not only help increase data accuracy and availability but also offers the benefit of linking materials and adsorption data of published studies to incorporate previously published data in the development of ML models easily. Recently, Ongari et al. [154] devised a workflow automatically matching MOFs in the CSD with adsorption data in NIST/ARPA-E Database of Novel and Emerging Adsorbent Materials (NIST-ISODB) [155] using conventional material names, digital object identifier (DOI) of papers, and structure graph. It is known that not all synthesized MOFs were deposited to the CSD [88] thus, future efforts to link MOFs not available in the CSD with their published adsorption data is important for extending the scale of synthesized MOFs that can be used for the development of ML models especially for adsorption purposes.

While one of the main aims of creating structured dataset is to have an accurate set of data, in some cases, there can be inaccuracies in the collected data requiring curation as a further step. With the aim of data curation, several material databases have been made accessible to researchers facilitating the data validation, expediting the material discovery process, and serving as the data source for the ML techniques. QMOF is an exemplary database as it involves a long list of material properties of currently >20,000 MOFs in a ML-friendly format which can be retrieved using an Application Programming Interface (API). Since the diversity of database used in ML models governs the transferability of ML models to unseen materials, several studies focused on introducing new MOF databases. Woo and coworkers [156] introduced a diverse MOF database, named ARC-MOF, along with DFT-based partial atomic charges facilitating their use in molecular simulations where electrostatic interactions play an important role. ARC-MOF database includes 279,610 experimentally synthesized and computer-generated MOFs from 15 different database sources, and it is provided with machine-readable descriptors such as geometric descriptors, RACs and AP-RDF. It was shown that ARC-MOF database have wider ranges of density, surface area, porosity, and pore diameter of MOFs than those in CoRE MOF and QMOF database, rendering ARC-MOF database a good choice for developing transferable ML models. New databases can be constructed using different building blocks to expand the diversity of MOFs for the benefit of ML model development. Conversely, the information learned from ML models can be used to generate databases with specific aims. Nandy et al. [157] used ML models [158] previously trained on the experimental stability data of MOFs to find synthesized and ultrastable MOFs. They separated 474 ultrastable MOFs into their building blocks and mined these blocks comprised of 88 inorganic nodes, 32 organic nodes, and 16 organic edges. The tinker-toy approach in porous materials maker (PORMAKE) [20] was then used to acquire a final set of 54,139 hMOFs. ML model predicted that about half of hMOFs have thermal and mechanical stability upon activation and the computer-generated ultrastable MOFs and their properties were made available online [158]. Expansion of such material databases will be invaluable for the design of optimal materials for practical applications.

3.2. Recent applications of ML methods in MOF research

3.2.1. Predicting synthesis conditions

Tuning synthesis parameters and routes can result in better scalability of MOF production [159]. Based on the CSD MOF subset, Gubsch et al. [160] established a MOF database, DigiMOF database, involving synthesis properties of MOFs (e.g., synthesis method, solvent, metal precursor) extracted from >40,000 published papers using a natural language processing (NLP) tool, ChemDataExtractor [161]. The main challenges encountered during the creation of DigiMOF database were the lack of some of the synthesis parameters in papers, MOF labeling problems, and paper availability issues. While these challenges led to the creation of a database involving about 15% of MOFs in the CSD MOF subset, it presents one of the first examples of identifying production pathways for MOFs. Linking MOFs with their synthesis routes can be highly valuable as this data can subsequently be employed in future HTCS and ML studies to integrate synthesis protocols into the processes of MOF design and scalability for target applications. Similarly, Park et al. [162] extracted synthesis information (e.g., metal/organic precursor type, temperature, solvent) for the CSD MOF subset from $\sim$ 30,000 papers and developed an ANN model for the synthesizability prediction. The model was shown to be successful in differentiating experimental synthesis conditions from computationally generated synthesis conditions with a recall score of 83.1% and predicting the crystallinity outcomes for a list of amorphous and crystalline MOFs, showing that these models can serve as a complementary tool for identifying synthesizable, highly crystalline MOFs.

To predict the synthesis conditions of MOFs, Luo et al. [163] employed a combined data mining and ML approach and compared its success with that of the traditional synthesis approach as shown in Fig. 4. The traditional approach involves searching for probable synthesis conditions for the target MOF in the literature and refining them through trial-and-error and expert experience, which is highly time-consuming. Conversely, the data-driven approach involves the development of ML models based on automatic literature data retrieval whose synthesis conditions recommendations can be improved through incorporating more experimental data, rendering a much more efficient MOF synthesis cycle. To demonstrate the practical benefits of the data-driven approach, they generated the first MOF synthesis database (SynMOF having 983 MOFs) by capturing synthesis parameters via

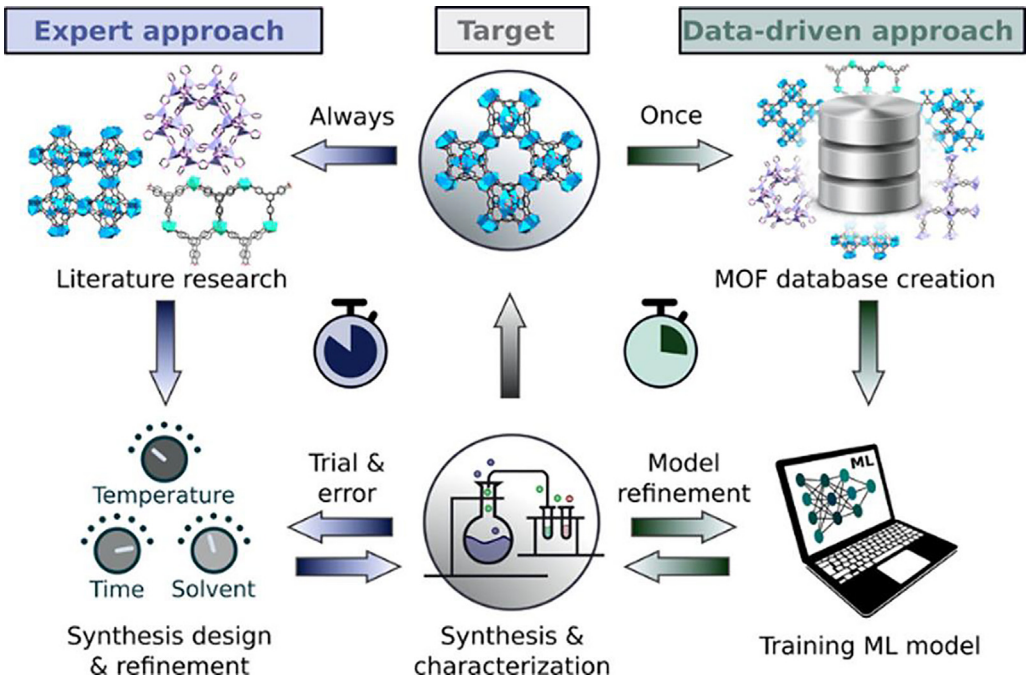


Fig. 4. Comparison of traditional (trial-and-error based) and data-driven (ML-based) approach for MOF synthesis. Reprinted with permission from Ref. [163].

NLP from the literature and structural information from CoRE MOF database which were used together to train a ML model. Their model involving information about metal source, linker, solvent, additive, synthesis time and temperature can serve as an inverse synthesis design tool which can be accessed online [164] where the user feeds in a CIF of a MOF and obtains its synthesis condition estimates. Since the MOF synthesis literature expands rapidly, it was recommended that ML models should be trained by considering new synthesis and crystallographic data for better prediction performance. It is envisioned that as the number of data points on which ML training is carried out increases, graph neural networks can provide more accurate ML models [163]. However, even in its current form, the ML model can estimate synthesis conditions better than experts who were provided MOF structure, metal/linker information. This superiority suggests that rather than a trial-and-error synthesis procedure, initial guesses for synthesis parameters based on ML model can be more beneficial.

ML models were also developed for the prediction of crystallization propensity of metal–organic nanocapsules (MONCs) where failed and successful experimental data were used as input [165]. Specifically, using synthesis parameters as input and reaction outcomes as output, one of the developed ML models predicted the crystallization tendency of MONCs with a high accuracy (>90%). Using reaction outcomes of 20 validation experiments (not involved in training or test set) as the benchmark, it was shown that the developed ML model could outperform expert predictions on the synthesis outcomes based on provided synthesis parameters.

ML models can also be used to optimize synthesis conditions of already synthesized crystalline structures. For instance, Domingues et al. [166] demonstrated that the synthesis conditions of MOFs could be optimized by combining (failed and successful) experimental data and genetic algorithms. Specifically, their approach could provide synthesis conditions which can lead to higher yield of crystalline Al-PMOF in shorter time, compared to the conditions reported in the literature [167]. Besides showing the advances that can be made through ML techniques, these studies underline the importance of publishing not only successful synthesis experiments but also failed ones as they can be beneficial for theoretical models.

3.2.2. Predicting stability

So far, many experimental and computational works identified the top performing MOFs for various applications, however, among them, only those with high structural stabilities (i.e., no/minimal structural degradation) can be employed in industrial applications [131]. While stabilities of many (especially early) MOFs are lower than those of industrially employed zeolites, over time, chemically and thermally stable MOFs have also been constructed [168]. This implies that techniques unraveling stable MOFs among a long list of candidates are highly warranted. One of the most crucial types of stability is water stability as the MOFs may be exposed to humid air or aqueous media during process operation [169]. Batra et al. [169] devised ML models based on experimental water stability data of ~200 MOFs which was, in general, successful in classifying the stability of unseen MOFs (accuracy > 70%). The models, whose only inputs were activated formula units (i.e., no structural information needed), also revealed MOF water stability trends and chemical guidelines based on metal/ligand-based chemical descriptor values. While inputting no structural information to ML models provides some convenience for predictions, incorporation of structural information could improve the predictions. Although MOFs can be synthesized in powder form, they typically need to be manufactured in compact form (e.g., pellet) due to practical operational needs involving reduced pressure drop [170,171]. Many MOFs can degrade/lose crystallinity upon exerting mechanical stress preventing their use in industry for various applications including adsorption/separation where stable pores are required [131]. In the quest of accelerating the identification of mechanically stable MOFs, Moghadam et al. [131] carried out simulations to acquire mechanical properties of thousands of MOFs and established ANN models to predict MOFs' mechanical stability and relate them with topology. Including topological features into ANN model improved the prediction accuracy significantly. Their model can expedite the design and realization of mechanically stable MOFs for which topology, and size, coordination and chemistry of building blocks can play significant role.

One of the most recent works on MOFs revealed that ML models with atomic (e.g., metal electronegativity, ionization energy) and periodic (e.g., MOF density, molecular weight) descriptors can

accurately categorize (with an accuracy > 80%) the MOFs based on their thermal stability points extracted from thermogravimetry analysis (TGA) data [172]. Since thermal stability of MOFs may also depend on other parameters such as solvent type, synthesis time etc. that were not included in the model fitting, a more detailed ML model involving those parameters can improve the predictions and elucidate the effect of synthesis parameters on thermal stability. Nandy et al. [158,173] gleaned structural stability information from thousands of papers via NLP and image analysis to build ML models that can accurately predict solvent-removal stability and thermal stability (mean absolute error (MAE) of 47 °C). The developed ML model allows making MOF stability predictions which are orders of magnitude faster than modeling or experiments. The model can also provide general guidelines to enhance the structural stability of MOFs. For instance, analyzing the entire CoRE MOF database, it was concluded that oxygen-coordinating linkers can form more stable structures than nitrogen-coordinating linkers. One crucial aspect of the study is that they provided a web interface (MOFSimplify) (https://mofsimplify.mit.edu/) that allows uploading new MOF CIFs for which activation and thermal stability predictions can be made.

3.2.3. Predicting guest accessibility

An intriguing application of ML for MOFs is the prediction of guest accessibility (i.e., pore size) based on only chemical information about metals and linkers. This can be beneficial for synthetic exploration of MOFs with sufficiently large pore sizes for applications where guest accessibility is a necessity [174]. Fig. 5 shows ML model development stages (stages 1–6 denoting calculation of structural properties, structure cleaning and decomposition, constructing MOF dataset, MOF featurization, ML model training, ML predictions, respectively) for the prediction of guest accessibility of MOFs. Here, MOFs built by a metal and a linker were focused on for which PLD classifications were performed using only chemical descriptors extracted from metal and linker. The threshold values of PLDs chosen (2.4, 4.4, and 5.9 Å) for the binary classification of MOFs were picked such that the models can identify whether a MOF comprised of a metal-linker combination would involve small, medium, or large pores. Consequently, by selecting only a metal and a linker, one can have an initial assessment about the accessibility of various sized guests in MOFs that would be formed through the combination of the selections.

Likewise, based on metal/linker/functional group selections, Krokidas et al. [175] demonstrated the development of a two-step ML model to predict the aperture sizes of ZIFs with SOD topology and then predict the diffusivity of multiple penetrants (He, H_2, O_2, CO_2, N_2, CH_4, C_2H_4, C_2H_6, C_3H_6, C_3H_8, i-C_4H_{10} and n-C_4H_{10}) inside ZIFs. It was shown that even a simplified ML model with van der Waals diameter of the penetrant and predicted aperture sizes can estimate the penetrant diffusivities accurately (R^2 = 0.86). ML models with less easily calculable descriptors (e.g., stretched aperture size) were also developed and models with simple inputs were shown to provide predictions of a similar quality to those with more complex inputs.

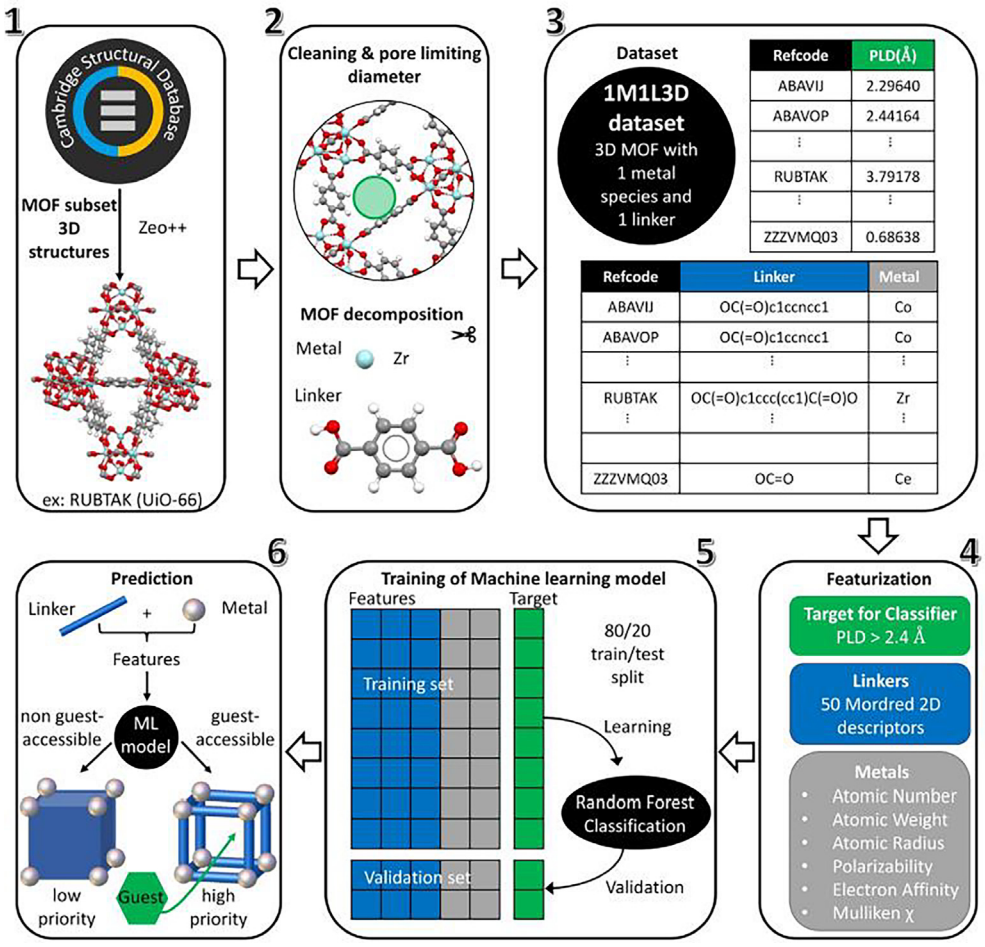


Fig. 5. Schematic summarizing the ML model development stages for the prediction of guest accessibility of MOFs. Reprinted with permission from Ref. [174].

3.2.4. Inverse material design

ML models in which structural features are inputs (and material performances are outputs) can be highly useful for accelerating material discovery with desired performance. However, generalizing the entire MOF space by such ML models is still a daunting task due to the sheer number of MOF structures that can be generated. It was also pointed out earlier that ML models established based on different MOF databases may not be transferable or bias free [115,121]. An alternative is developing methods for the inverse design of MOFs where desired material performance metric (or combination of multiple metrics) is part of the input rather than the output. Lim et al. [176] combined genetic algorithm and deep neural network to inversely design MOFs for Xe/Kr separation. This approach involves fitting a ML model (via MOF representations composed of topology and building blocks) to the simulated Xe/Kr selectivity and generation of candidate structures by genetic algorithm for user-specified selectivity and/or Henry's coefficient value of Xe. A large fraction of the genetically constructed MOFs exhibited desired selectivity compared to the randomly generated MOFs, suggesting that low performing structures can be eliminated through this approach. MOFs can also be inversely designed with threshold performance values so that sufficiently good (in terms of one metric) MOFs can be designed and other factors (stability, cost etc.) can be considered during the final decision on the material selection. In principle, the inverse design of MOFs can also be performed using desired heat of adsorption of sorbates as target value in neural network models which was done earlier for zeolites [177]. It was also shown that the interaction energies between H_2 and MOFs can be used as descriptors in ML models to predict deliverable H_2 amounts between 100 and 2 bar at 77 K [178].

A great example of an inverse design process using SELFIES for MOFs and zeolites was published with the aim of developing deep generative models for CO_2 capture [140]. The CoRE MOF database, including ~14,000 materials, was deconstructed to represent reticular framework identification, including edges, vertices (organic and metal parts), and topology. For reticular framework information, while SELFIES were only used for the edge parts of MOFs to encode/decode in supramolecular variational autoencoder system (SmVAE), vertices and topology information were directly-one-hot encoded [111] (representing presence (absence) of categorical variables with 1 (0) in feature columns) as categories. This reticular framework information was transferred to the latent space of SmVAE for the target property (textural and gas uptake of MOFs) predictions. Then, this autoencoder system predicted all combinations of edge, organic, and metal vertices to create an ideal MOF exhibiting desired value for target properties. Their globally optimized design process was used to discover a MOF (GMOF-9) with improved CO_2 capacity (7.55 mol/kg) and good CO_2/CH_4 selectivity (16). This study was an excellent example of understanding the superior potential of a new global optimization process designing a framework without any constraints unlike the isoreticular optimization process. Fig. 6**(a)** shows the reduced dimension of the properties of MOFs (latent space) constructed via principal component analysis (PCA) based on the similarity of MOF property to represent the material design approaches with two MOF examples: NU-1104 and MOF-5. The isoreticular design for NU-1104 with constrained topology (ftw) and metal node resulted in the maximum CO_2 uptake of 4.33 mol/kg and infinite CO_2/CH_4 selectivity at 5 bar, 313 K as shown in Fig. 6**(b)**. On the other hand, the global optimization design process for MOF-5 without any constraint demonstrated a remarkably high CO_2 uptake (7.55 mol/kg) and high CO_2/CH_4 selectivity (16) at the same conditions as shown in Fig. 6 **(c)**. These results prove that while isoreticular optimization design allows us to construct a framework with one dimensional change, globally optimization process unlocks the material design of a structure with multidimensional change to achieve target property.

3.2.5. Universal models for different applications

Developing a universal ML model which is transferable from one application to another can significantly expedite the material discovery. Kang et al. [179] recently introduced a multi-modal pre-trained transformer architecture in MOFs, MOFTransformer, to develop a universal ML system for different applications. Using local chemical features and global geometric features captured by MOFTransformer, a universal ML model was developed that can be readily transferred from one application to another, such as gas adsorption, gas diffusion, and electronic properties of MOFs. They first designed three pre-training tasks for the features of 1 million hMOFs, such as MOF topology prediction, void fraction prediction, and metal cluster/organic linker classification to capture both local and global features. As a case study, H_2 uptake at 100 bar, H_2 diffusivity at infinite dilution, and band gap of fine-tuned 20,000 QMOFs were predicted using MOFTransformer. The accuracy of H_2 uptake, diffusivity and band gap predictions of MOFTransformer model were compared with those of three other ML models; energy histogram-based ML [178] model in which the bins of an energy histogram were used as features; descriptor-based ML [180] model in which structural and atomic percentage properties of MOFs were used as features, and crystal graph convolution neural network (CGCNN) [120], [181] model where the crystal graph of MOFs was used as features. Fig. 7**(a)** shows that MOFTransformer model has lower MAEs than those of the conventional three ML models using different training sizes, suggesting that MOFTransformer model can outperform the other ML models trained by different features. Depending on the application, the MAEs of the foregoing three ML models can be considerably higher than those of MOFTransformer model. For instance, at a data size of 20,000 MOFs, MAE_{CGCNN}, $MAE_{Energy\ Histogram}$, and $MAE_{Descriptor\text{-}based\ ML}$ were roughly 5, 1.7, and 2.5-fold of those of MOFTransformer for H_2 uptake, H_2 diffusivity, and band gap, respectively.

Another recent example is a self-supervised transformer model developed for property predictions of MOFs, MOFormer [182]. In MOFormer, MOF was represented using SMILES strings of framework subunits in MOFid, not requiring the 3D atomistic structure of MOF, thus, being structure-agnostic. To take advantage of pre-training approach benefits, they introduced a self-supervised learning (SSL) framework combining MOFormer and CGCNN model with > 400,000 MOFs. Then, the graph and string representation of MOF were encoded via CGCNN and MOFormer, respectively. After the pretraining of models, MOFormer and CGCNN models were separately finetuned. The results showed that SSL pretraining framework improved MOF representation accuracy for different datasets such as QMOF and hMOF. The MOFormer pipeline were trained using QMOF dataset to predict the band gap and hMOF dataset to predict CO_2 and CH_4 adsorption. It has been revealed that the accuracy of target property predictions of MOFs using pre-trained CGCNN and MOFormer is generally higher than those using CGCNN and MOFormer from scratch and Smooth Overlap of Atomistic Positions (SOAP) [183]-based ML model. To investigate the efficiency, they compared five different models; two models used in SSL framework from scratch and pretrained ($CGCNN_{scratch}$, $CGCNN_{pretrained}$, $MOFormer_{scratch}$, $MOFormer_{pretrained}$) and SOAP-based ML model based on band gaps of MOFs in QMOF dataset and low-pressure CO_2 adsorption of MOFs in hMOF dataset as shown in Fig. 7(b). Results showed that the accuracy of band gap predictions using $MOFormer_{pretrained}$ model is higher than those of other models when training data size is $\leq$ 1000 MOFs. On the other hand, the accuracy of CO_2 adsorption predictions using CGCNN models is higher than MOFormer models regardless of the training size. Overall, pretrained MOFormer model can give rapid and accurate results especially for properties calculated using quantum chemical simulations even in the case of low training

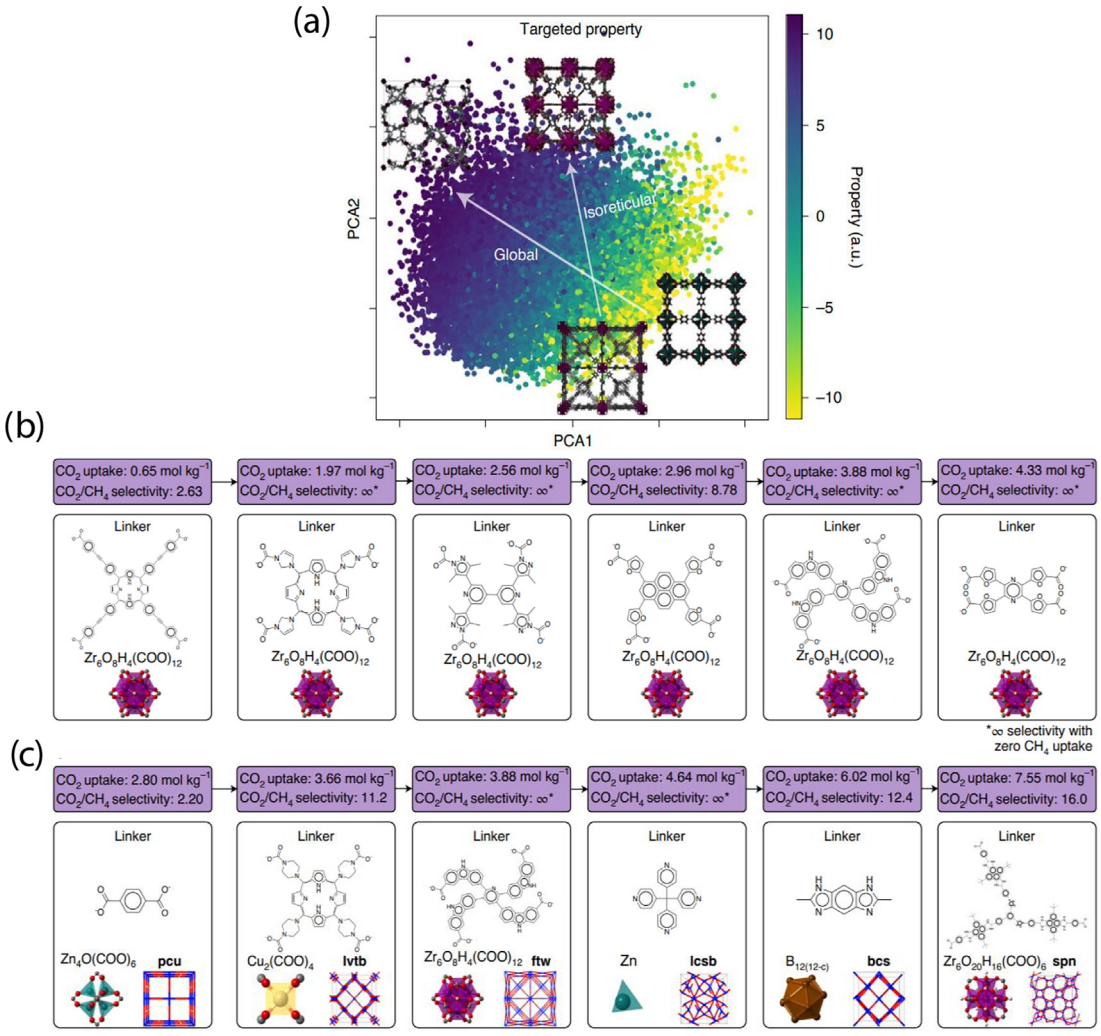


Fig. 6. (a) The comparison of exemplary isoreticular and globally optimized design processes based on principal component analysis (PCA) schematically shown. (b) In isoreticular design process constrained to ftw topology and $Zr_6O_8H_4(COO)_{12}$ node, starting MOF is NU-1104 and discovered MOFs exhibit notable CO_2 uptake and infinite selectivity. (c) In global optimization design process without any constraints, starting MOF is MOF-5 and discovered MOFs exhibit remarkably high CO_2 uptake and high selectivity. Reprinted with permission from Ref. [140].

data size, which is critical to achieve as computationally demanding simulations are required to calculate these properties.

3.2.6. Bridging ML with experiments

As the ultimate goal of developing ML models in the realm of materials science is identifying and realizing high performance materials for a specific application, experimental validation of the ML predictions made for MOFs is crucial yet challenging [184]. To this end, Bucior et al. [178] developed a ML model to predict H_2 storage capabilities of >50,000 MOFs and experimentally validated the H_2 storage predictions for MFU-4*l*(Zn) which was one of the top MOFs according to the ML model. Similar ML model development followed by experimental validation approach was undertaken for H_2 storage in zeolites as well [185]. Thornton et al. [186] developed a neural network-based ML model to find out promising crystalline materials out of ~850,000 materials for deliverable H_2 capacity at cryogenic conditions (77 K). Considering the top materials identified by the ML model, it was concluded that some of the top materials had already been synthesized while there were some hMOFs in the top list as well. The experimental and simulated H_2 isotherms of the synthesized materials (ZIF-8 and MOF-210) validated their predicted high deliverable capacities.

The literature bridging ML models with experimental validation is at its infancy which can be partially attributed to the fact that ML model development is often carried out by purely computational groups. As in all other subfields of MOF research, this underlines the importance of establishing cooperation between experimentalists and theoreticians through which the ML models can be refined further. Such cooperations should involve not only raw data transfer across groups but also addressal of issues that may have deep impact on determining accuracy and generalization ability of ML models such as deficiencies in modelling and unstandardized synthesis, characterization, and testing of materials in the wet lab. For instance, while many computational studies rely on perfect MOF crystals, in reality, structures can be defective which in turn will cause disparities across experimental and simulation results even if interaction parameters were highly accurate. Not surprisingly, the ML models that would be fitted to simulation results of the perfect crystals would not provide accurate representations of properties of defective structures. Therefore, the data-driven approach does not only rely on data collection and ML model development tools but also close cooperation of scientists from multiple disciplines which can help theoreticians to develop better ML models and experimentalists to identify priorly unnoticed properties (e.g., presence of defect) of samples.

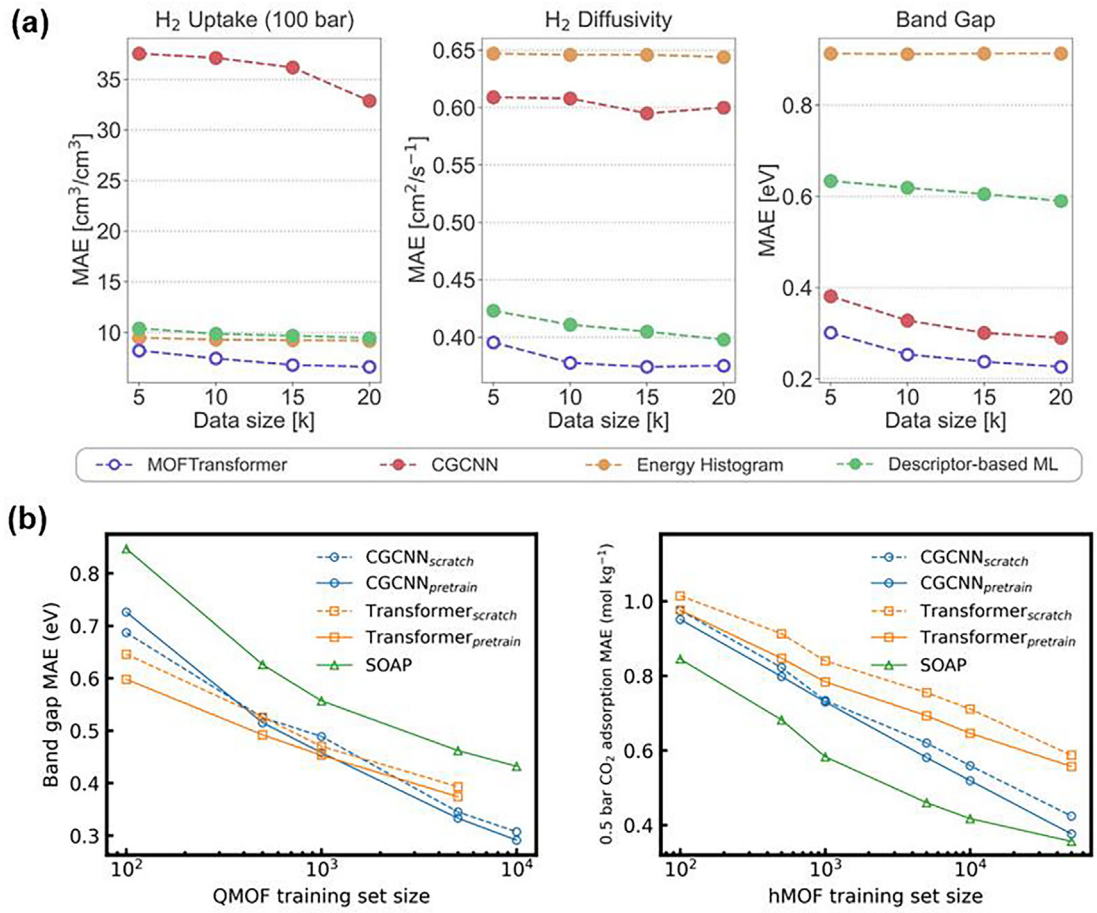


Fig. 7. (a) Comparison of MAE results for H_2 uptake, H_2 diffusivity, and band gap for the finetuning of MOFTransformer model and three baseline models (CGCNN, Energy Histogram and Descriptor-based ML) with respect to data size from 5,000 to 20,000 MOFs. Reprinted with permission from Ref. [179] (b) Data efficiency comparison between CGCNN, Transformer and SOAP-based models on QMOF (left) and hMOF (right) dataset. Reprinted with permission from Ref. [182].

3.3. Developing ML workflows: from datasets up to publication

Each step of the material design process using AI tremendously affects the discovery of novel materials for a target application. The development of new ML algorithms, material databases to optimize the ML models, and descriptors to represent the materials have been studied as mentioned above. In the digital age, automated and user-friendly ML workflows are critical to design new materials quickly and easily. Hachmann et al. [187] recently developed an open ML and informatics program (ChemML) for the analysis, mining, and modelling of chemicals and materials data. ChemML provides a multitude of methods as part of seven core task classes; input acquisition, representing, preparing data, training of models, optimizing the models, visualizing, and obtaining output in addition to different ML techniques such as transfer learning and active learning that improve accuracy and efficiency by closing the loop of data modelling and molecular design. The open-source ChemML package can be modified by adding new methodological contributions and workflows by authors and other researchers to accelerate material design but, to the best of our knowledge, the ChemML package has not been used for the purpose of predicting the properties of MOFs yet.

During the development of a ML workflow, modifications/corrections to the workflow have an important effect on the prediction accuracy of target properties. For this reason, Jablonka et al. [188] recently introduced mofdscribe software providing a ML workflow as well as offering the proper solutions at different stages of ML workflow. The software library consists of collecting a dataset, featurizing materials, splitting the datasets, model evaluation, and comparing the performance with the state-of-the-art models. In the first stage, calculated features and properties of two experimentally synthesized MOF datasets (all-solvent-removed (ASR-CSD) subset of the 2019 MOF subset of the CSD and CoRE MOF), a Boyd-Woo hypothetical dataset (BW) [189], and QMOF dataset were incorporated into mofdscribe. The code allows for identifying and removing duplicate MOFs to curtail the risk of data leakage. To represent the MOF materials in machine-readable format, they integrated packages that enable calculating 40 features of MOFs classified as the atom-centered (RACs, average minimum distance (AMD) [190], building unit-centered (shape, flexibility), and global (pore geometry, topology etc.). They also implemented different data splitting approaches to customize training and test sets based on the problem investigated. For example, if one wants to predict any properties of MOFs discovered after year X by developing ML models for MOFs discovered before year X, the dataset can be split into training and test sets based on time-grouped in mofdscribe. Users can generate a report of the performance of their customized ML pipeline on a public leaderboard, which is highly beneficial to compare the accuracy and results of ML studies with previously reported ones to create the ecosystem for material design. Development of such automated ML workflows is crucial for the field of MOF research as it can facilitate carrying out data-driven research for both non-experts and experts of ML. The increased use of ML techniques can eventually help reach the goal of having accurate predictive models for the design and selection of MOFs for different applications.

4. Outlook

With the steep expansion in the number and diversity of synthesized and computer-generated MOFs, identifying high performing materials for a specific application requires combination of experimental, computational, and ML studies more than ever. As following the traditional method of trial-and-error experimentation for new materials would be highly inefficient and costly, carrying out HTCS studies has become very popular especially in the last decade to direct experiments toward the most promising material candidates. However, the number of MOFs has exceeded even the limit that HTCS approach can handle in terms of computational cost and time. Thus, today's energy and environment-related challenges prompts adoption of a new data-driven approach where the enormous volume of data that can be relatively easily obtained for MOFs should be used to build accurate and fast prediction models. We highlighted our perspective on the opportunities and challenges of recent advancements in the computational MOF field below:

- **Transferable and complete ML workflows for MOFs:** One of the chief aims of the data-driven material research is obtaining transferable ML models as they would reduce the need to train ML models for every material subset and accelerate the material discovery process. The prediction of partial atomic charges is one of the exemplary areas where the transferability of ML models across different porous material databases could be very useful, as these charges have a great impact on the results of classical molecular simulations. Recently, based on DFT data, several ML-based [25,191] models were developed to assign atomic charges in porous materials, which can be useful for the accurate identification of the best materials via molecular simulations.

In a data-driven approach, the ultimate goal should not only be creating those models but also designing the entire workflow such that the computational tools utilized in the workflow are open source, offering reproducible results in an automated fashion [192]. For instance, the retrieved data set for ML model development may involve missing or unrealistic values. If the handling of such cases involves user-specific (e.g., different users setting different thresholds for retention/dismissal of unrealistic values) and undocumented/partially documented decisions as part of data processing stage, the reproducibility of the study can be hampered due to potential data bias [193].

- **Emerging applications of MOFs for ML implementation:** While most studies about the development of ML models have focused on commonly studied applications of MOFs, mostly gas storage and separations, ML models for drug delivery systems are limited due to very scarce data obtained from complex molecular simulations for large drug molecules. He et al. [194] conducted a ML study for drug delivery systems by collecting 40 data points from 100 different published papers, developed different conventional ML models to predict IBU loading as output data and presented a good starting point for ML-based biomaterial design.

Currently, the number of studies implementing ML methods particularly for MOF catalysis is also scarce. Based on the DFT data, Schweitzer et al. [195] established a gradient boosted machine (GBM) based ML model to predict binding energies of several small sorbates (CO, CH_3, OH, H, O, N, C) on surface models of MOF encapsulated metal catalyst (Au, Cu, Pt, Pd, Ni). While their surface models were reduced models representing MOFs only with linkers, the ML model was still instructive as it identified the most important descriptors for binding energy. Li et al. [196] used a different approach and developed a ML model to classify MOFs' potential (good or bad) for CO_2 fixation to cyclic carbonate based on experimental turnover frequency data collected from the literature. Applying the model to $\sim$ 1300 hypothetical MOFs revealed the best metals and linkers that can potentially lead to good performance for fixation. Such insights that can be rapidly obtained through ML models will be of high importance for future MOF-related catalyst design.

- **Accuracy of reference simulation data of MOFs:** So far, there has been a propensity for developing ML models based on simulation data rather than experimental data perhaps due to the wider availability of the former for more materials. As the simulation results rely on the underlying assumptions, ML model predictions can only be as accurate as the simulation results which, in some cases, may not match experimental observations. For instance, while all materials are flexible to some extent, the molecular simulations of MOFs are almost always performed using rigid framework assumption to simplify MOF modeling which may introduce artefacts in the generated data [64]. As the flexible force fields are becoming more widespread, it can be anticipated that future studies will employ them more frequently, eventually creating datasets where the framework flexibility effects are considered. To this end, ML-based force field development studies will be of more importance to accelerate the generation of flexible simulation results. In the future, it would be interesting to see how ML models fitted to flexible simulation results would perform against experimental trends and/or ML models based on rigid simulation results.

Similarly, in most of the computational screening studies, generic force fields which were not specifically derived for MOFs are employed [64]. Depending on the type of interactions involved in the MOF systems, the success of simulations in reproducing experimental trends could be mixed. One of the most renowned examples where the generic force fields can fail to have accurate description is open metal site-adsorbate interactions resulting in significant differences across experimental and simulation data [197]. With the expansion and use of specialized force fields that can provide more accurate descriptions of interactions, the accuracy of molecular simulations (with respect to experimental results) and hence that of ML models (trained on improved simulation results) will improve [197].

- **Potential of ML in experimental field of MOFs:** As the data-driven workflows would benefit from experimental data that are becoming increasingly larger in volume and broader in scope, reporting successful and failed experimental data will be of more importance. The data-driven material research is anticipated to be the mainstream approach in the not too far future, thus, having open access data repositories for both standardized experimental and computational data will be one of the key aspects for material discovery. Collecting both experimental and computational data into databases can reduce the data imbalance or bias in datasets to be subsequently used for the development of ML models which can lead to better generalization abilities.

While it is important to speed up computationally identifying promising materials through various ML techniques, major advancements in the material synthesis, characterization, and testing are also warranted to cut down the experimental time and effort needed. To this end, the use of automated robotic method to perform synthesis experiments can be a game changer in the context of high-throughput experimentation [198]. Incorporating

robots into a diverse set of material discovery processes can dramatically improve the scalability, safety, and reproducibility of the experiments.

There is no doubt that much remains to be done to further improve the material databases, data mining and ML tools etc. Yet, even at the current state of data-driven MOF design and discovery, the implementations of ML techniques are so effective and inspiring. We anticipate that the already fast-paced computational MOF research field will be expedited even further with deeper integration of data science methods into the field in near future.

Data availability

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

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.

Acknowledgements

S.K. acknowledges ERC-2017-Starting Grant. This study has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (ERC-2017-Starting Grant, grant agreement No 756489-COSMOS).

References

[1] G.S. Day, H.F. Drake, H.-C. Zhou, M.R. Ryder, Evolution of porous materials from ancient remedies to modern frameworks, Commun. Chem. 4 (2021) 114, https://doi.org/10.1038/s42004-021-00549-4.
[2] R. Freund, O. Zaremba, G. Arnauts, R. Ameloot, G. Skorupskii, M. Dincă, A. Bavykina, J. Gascon, A. Ejsmont, J. Goscianska, M. Kalmutzki, U. Lächelt, E. Ploetz, C.S. Diercks, S. Wuttke, The current status of MOF and COF applications, Angew. Chemie Int. Ed. 60 (2021) 23975–24001, https://doi.org/10.1002/anie.202106259.
[3] H. Furukawa, N. Ko, Y.B. Go, N. Aratani, S.B. Choi, E. Choi, A.Ö. Yazaydin, R.Q. Snurr, M. O'Keeffe, J. Kim, O.M. Yaghi, Ultrahigh porosity in metal-organic frameworks, Science. 329 (2010) 424–428, https://doi.org/10.1126/science.1192160.
[4] W. Fan, X. Zhang, Z. Kang, X. Liu, D. Sun, Isoreticular chemistry within metal–organic frameworks for gas storage and separation, Coord. Chem. Rev. 443 (2021) 213968, https://doi.org/10.1016/j.ccr.2021.213968.
[5] C.A. Trickett, A. Helal, B.A. Al-Maythalony, Z.H. Yamani, K.E. Cordova, O.M. Yaghi, The chemistry of metal–organic frameworks for CO_2 capture, regeneration and conversion, Nat. Rev. Mater. 2 (2017) 17045, https://doi.org/10.1038/natrevmats.2017.45.
[6] D. Yang, M. Babucci, W.H. Casey, B.C. Gates, The surface chemistry of metal oxide clusters: from metal-organic frameworks to minerals, ACS Cent. Sci. 6 (2020) 1523–1533, https://doi.org/10.1021/acscentsci.0c00803.
[7] Q. Wang, D. Astruc, State of the art and prospects in metal-organic framework (MOF)-based and MOF-derived nanocatalysis, Chem. Rev. 120 (2020) 1438–1511, https://doi.org/10.1021/acs.chemrev.9b00223.
[8] S. Mallakpour, E. Nikkhoo, C.M. Hussain, Application of MOF materials as drug delivery systems for cancer therapy and dermal treatment, Coord. Chem. Rev. 451 (2022) 214262, https://doi.org/10.1016/j.ccr.2021.214262.
[9] K. Suresh, A.J. Matzger, Enhanced drug delivery by dissolution of amorphous drug encapsulated in a water unstable metal-organic framework (MOF), Angew. Chemie Int. Ed. 58 (2019) 16790–16794, https://doi.org/10.1002/anie.201907652.
[10] H.-S. Wang, Metal–organic frameworks for biosensing and bioimaging applications, Coord. Chem. Rev. 349 (2017) 139–155, https://doi.org/10.1016/j.ccr.2017.08.015.
[11] Y.-A. Li, C.-W. Zhao, N.-X. Zhu, Q.-K. Liu, G.-J. Chen, J.-B. Liu, X.-D. Zhao, J.-P. Ma, S. Zhang, Y.-B. Dong, Nanoscale UiO-MOF-based luminescent sensors for highly selective detection of cysteine and glutathione and their application in bioimaging, Chem. Commun. 51 (2015) 17672–17675, https://doi.org/10.1039/C5CC07783D.
[12] L.E. Kreno, K. Leong, O.K. Farha, M. Allendorf, R.P. Van Duyne, J.T. Hupp, Metal-organic framework materials as chemical sensors, Chem. Rev. 112 (2012) 1105–1125, https://doi.org/10.1021/cr200324t.
[13] D. Sheberla, J.C. Bachman, J.S. Elias, C.-J. Sun, Y. Shao-Horn, M. Dincă, Conductive MOF electrodes for stable supercapacitors with high areal capacitance, Nat. Mater. 16 (2017) 220–224, https://doi.org/10.1038/nmat4766.
[14] M.C. So, G.P. Wiederrecht, J.E. Mondloch, J.T. Hupp, O.K. Farha, Metal–organic framework materials for light-harvesting and energy transfer, Chem. Commun. 51 (2015) 3501–3510, https://doi.org/10.1039/C4CC09596K.
[15] A.E. Baumann, X. Han, M.M. Butala, V.S. Thoi, Lithium thiophosphate functionalized zirconium MOFs for Li–S batteries with enhanced rate capabilities, J. Am. Chem. Soc. 141 (2019) 17891–17899, https://doi.org/10.1021/jacs.9b09538.
[16] L. Li, H. Xue, Y. Wang, P. Zhao, D. Zhu, M. Jiang, X. Zhao, Solvothermal metal metathesis on a metal-organic framework with constricted pores and the study of gas separation, ACS Appl. Mater. Interfaces. 7 (2015) 25402–25412, https://doi.org/10.1021/acsami.5b07953.
[17] L. Chen, H. Chen, R. Luque, Y. Li, Metal–organic framework encapsulated Pd nanoparticles: towards advanced heterogeneous catalysts, Chem. Sci. 5 (2014) 3708–3714, https://doi.org/10.1039/C4SC01847H.
[18] A. Ferguson, L. Liu, S.J. Tapperwijn, D. Perl, F.-X. Coudert, S. Van Cleuvenbergen, T. Verbiest, M.A. van der Veen, S.G. Telfer, Controlled partial interpenetration in metal–organic frameworks, Nat. Chem. 8 (2016) 250–257, https://doi.org/10.1038/nchem.2430.
[19] The Cambridge Structural Database (CSD), (2022). https://www.ccdc.cam.ac.uk/CCDCStats/Stats (accessed October 15, 2022).
[20] S. Lee, B. Kim, H. Cho, H. Lee, S.Y. Lee, E.S. Cho, J. Kim, Computational screening of trillions of metal-organic frameworks for high-performance methane storage, ACS Appl. Mater. Interfaces. 13 (2021) 23647–23654, https://doi.org/10.1021/acsami.1c02471.
[21] L. Himanen, A. Geurts, A.S. Foster, P. Rinke, Data-driven materials science: status, challenges, and perspectives, Adv. Sci. 6 (2019) 1900808, https://doi.org/10.1002/advs.201900808.
[22] C.W. Coley, N.S. Eyke, K.F. Jensen, Autonomous discovery in the chemical sciences Part II: outlook, Angew. Chemie Int. Ed. 59 (2020) 23414–23436, https://doi.org/10.1002/anie.201909989.
[23] E. Alpaydin, Introduction to Machine Learning, MIT Press, 2014, https://books.google.com.tr/books?id=NP5bBAAAQBAJ.
[24] V.V. Korolev, A. Mitrofanov, E.I. Marchenko, N.N. Eremin, V. Tkachenko, S.N. Kalmykov, Transferable and extensible machine learning-derived atomic charges for modeling hybrid nanoporous materials, Chem. Mater. 32 (2020) 7822–7831, https://doi.org/10.1021/acs.chemmater.0c02468.
[25] A. Raza, A. Sturluson, C.M. Simon, X. Fern, Message passing neural networks for partial charge assignment to metal-organic frameworks, J. Phys. Chem. C. 124 (2020) 19070–19082, https://doi.org/10.1021/acs.jpcc.0c04903.
[26] H. Li, M. Eddaoudi, M. O'Keeffe, O.M. Yaghi, Design and synthesis of an exceptionally stable and highly porous metal-organic framework, Nature 402 (6759) (1999) 276–279, https://doi.org/10.1038/46248.
[27] O.M. Yaghi, H. Li, Hydrothermal synthesis of a metal-organic framework containing large rectangular channels, J. Am. Chem. Soc. 117 (1995) 10401–10402, https://doi.org/10.1021/ja00146a033.
[28] S.-S.-Y. Chui, S.-M.-F. Lo, J.P.H. Charmant, A.G. Orpen, I.D. Williams, A chemically functionalizable nanoporous material $[Cu_3(TMA)_2(H_2O)_3]_n$, Science. 283 (1999) 1148–1150, https://doi.org/10.1126/science.283.5405.1148.
[29] M. Eddaoudi, H. Li, O.M. Yaghi, Highly porous and stable metal–organic frameworks: structure design and sorption properties, J. Am. Chem. Soc. 122 (2000) 1391–1397, https://doi.org/10.1021/ja9933386.
[30] B. Chen, M. Eddaoudi, S.T. Hyde, M. O'Keeffe, O.M. Yaghi, Interwoven metal-organic framework on a periodic minimal surface with extra-large pores, Science. 291 (2001) 1021–1023, https://doi.org/10.1126/science.1056598.
[31] J. Sun, L. Weng, Y. Zhou, J. Chen, Z. Chen, Z. Liu, D. Zhao, QMOF-1 and QMOF-2: three-dimensional metal-organic open frameworks with a quartzlike topology, Angew. Chemie Int. Ed. 41 (2002) 4471–4473, https://doi.org/10.1002/1521-3773(20021202)41:23<4471::AID-ANIE4471>3.0.CO;2-9.
[32] R.E. Morris, P.S. Wheatley, Gas storage in nanoporous materials, Angew. Chemie Int. Ed. 47 (2008) 4966–4981, https://doi.org/10.1002/anie.200703934.
[33] Y. Peng, V. Krungleviciute, I. Eryazici, J.T. Hupp, O.K. Farha, T. Yildirim, Methane storage in metal-organic frameworks: current records, surprise findings, and challenges, J. Am. Chem. Soc. 135 (2013) 11887–11894, https://doi.org/10.1021/ja4045289.
[34] A.R. Millward, O.M. Yaghi, Metal–organic frameworks with exceptionally high capacity for storage of carbon dioxide at room temperature, J. Am. Chem. Soc. 127 (2005) 17998–17999, https://doi.org/10.1021/ja0570032.
[35] A.M. Seayad, D.M. Antonelli, Recent advances in hydrogen storage in metal-containing inorganic nanostructures and related materials, Adv. Mater. 16 (2004) 765–777, https://doi.org/10.1002/adma.200306557.
[36] A.G. Wong-Foy, A.J. Matzger, O.M. Yaghi, Exceptional H_2 saturation uptake in microporous metal–organic frameworks, J. Am. Chem. Soc. 128 (2006) 3494–3495, https://doi.org/10.1021/ja058213h.
[37] Y. Yan, X. Lin, S. Yang, A.J. Blake, A. Dailly, N.R. Champness, P. Hubberstey, M. Schröder, Exceptionally high H_2 storage by a metal–organic polyhedral framework, Chem. Commun. (2009) 1025–1027, https://doi.org/10.1039/B900013E.
[38] O.K. Farha, A. Özgür Yazaydın, I. Eryazici, C.D. Malliakas, B.G. Hauser, M.G. Kanatzidis, S.T. Nguyen, R.Q. Snurr, J.T. Hupp, De novo synthesis of a metal–organic framework material featuring ultrahigh surface area and gas storage

capacities, Nat. Chem. 2 (11) (2010) 944–948, https://doi.org/10.1038/nchem.834.
[39] J.-R. Li, J. Sculley, H.-C. Zhou, Metal-organic frameworks for separations, Chem. Rev. 112 (2012) 869–932, https://doi.org/10.1021/cr200190s.
[40] L. Pan, D.H. Olson, L.R. Ciemnolonski, R. Heddy, J. Li, Separation of hydrocarbons with a microporous metal-organic framework, Angew. Chem. Int. Ed. 45 (2006) 616–619, https://doi.org/10.1002/anie.200503503.
[41] J.A. Mason, K. Sumida, Z.R. Herm, R. Krishna, J.R. Long, Evaluating metal-organic frameworks for post-combustion carbon dioxide capture via temperature swing adsorption, Energy Environ. Sci. 4 (2011) 3030–3040, https://doi.org/10.1039/C1EE01720A.
[42] W.-Y. Gao, Y. Niu, Y. Chen, L. Wojtas, J. Cai, Y.-S. Chen, S. Ma, Porous metal–organic framework based on a macrocyclic tetracarboxylate ligand exhibiting selective CO_2 uptake, CrystEngComm. 14 (2012) 6115–6117, https://doi.org/10.1039/C2CE25484K.
[43] Y.-S. Bae, O.K. Farha, A.M. Spokoyny, C.A. Mirkin, J.T. Hupp, R.Q. Snurr, Carborane-based metal–organic frameworks as highly selective sorbents for CO2 over methane, Chem. Commun. (2008) 4135–4137, https://doi.org/10.1039/B805785K.
[44] Z.R. Herm, J.A. Swisher, B. Smit, R. Krishna, J.R. Long, Metal−organic frameworks as adsorbents for hydrogen purification and precombustion carbon dioxide capture, J. Am. Chem. Soc. 133 (2011) 5664–5667, https://doi.org/10.1021/ja111411q.
[45] P. Horcajada, R. Gref, T. Baati, P.K. Allan, G. Maurin, P. Couvreur, G. Férey, R.E. Morris, C. Serre, Metal-organic frameworks in biomedicine, Chem. Rev. 112 (2012) 1232–1268, https://doi.org/10.1021/cr200256v.
[46] A.C. McKinlay, R.E. Morris, P. Horcajada, G. Férey, R. Gref, P. Couvreur, C. Serre, BioMOFs: metal-organic frameworks for biological and medical applications, Angew. Chemie Int. Ed. 49 (2010) 6260–6266, https://doi.org/10.1002/anie.201000048.
[47] A. Bieniek, A.P. Terzyk, M. Wiśniewski, K. Roszek, P. Kowalczyk, L. Sarkisov, S. Keskin, K. Kaneko, MOF materials as therapeutic agents, drug carriers, imaging agents and biosensors in cancer biomedicine: recent advances and perspectives, Prog. Mater. Sci. 117 (2021), https://doi.org/10.1016/j.pmatsci.2020.100743.
[48] P. Horcajada, C. Serre, M. Vallet-Regí, M. Sebban, F. Taulelle, G. Férey, Metal-organic frameworks as efficient materials for drug delivery, Angew. Chemie Int. Ed. 45 (2006) 5974–5978, https://doi.org/10.1002/anie.200601878.
[49] P. Horcajada, C. Serre, G. Maurin, N.A. Ramsahye, F. Balas, M. Vallet-Regí, M. Sebban, F. Taulelle, G. Férey, Flexible porous metal-organic frameworks for a controlled drug delivery, J. Am. Chem. Soc. 130 (2008) 6774–6780, https://doi.org/10.1021/ja710973k.
[50] V. Kumar, V. Singh, K.-H. Kim, E.E. Kwon, S.A. Younis, Metal-organic frameworks for photocatalytic detoxification of chromium and uranium in water, Coord. Chem. Rev. 447 (2021) 214148, https://doi.org/10.1016/j.ccr.2021.214148.
[51] S.R. Miller, D. Heurtaux, T. Baati, P. Horcajada, J.-M. Grenèche, C. Serre, Biodegradable therapeutic MOFs for the delivery of bioactive molecules, Chem. Commun. 46 (2010) 4526–4528, https://doi.org/10.1039/C001181A.
[52] J. An, S.J. Geib, N.L. Rosi, Cation-triggered drug release from a porous zinc−adeninate metal−organic framework, J. Am. Chem. Soc. 131 (2009) 8376–8377, https://doi.org/10.1021/ja902972w.
[53] R.B. Getman, Y.-S. Bae, C.E. Wilmer, R.Q. Snurr, Review and analysis of molecular simulations of methane, hydrogen, and acetylene storage in metal-organic frameworks, Chem. Rev. 112 (2012) 703–723, https://doi.org/10.1021/cr200217c.
[54] J. Jiang, R. Babarao, Z. Hu, Molecular simulations for energy, environmental and pharmaceutical applications of nanoporous materials: from zeolites, metal–organic frameworks to protein crystals, Chem. Soc. Rev. 40 (2011) 3599–3612, https://doi.org/10.1039/C0CS00128G.
[55] S. Keskin, J. Liu, R.B. Rankin, J.K. Johnson, D.S. Sholl, Progress, opportunities, and challenges for applying atomically detailed modeling to molecular adsorption and transport in metal−organic framework materials, Ind. Eng. Chem. Res. 48 (2009) 2355–2371, https://doi.org/10.1021/ie800666s.
[56] T. Düren, L. Sarkisov, O.M. Yaghi, R.Q. Snurr, Design of new materials for methane storage, Langmuir. 20 (2004) 2683–2689, https://doi.org/10.1021/la0355500.
[57] S. Wang, Comparative molecular simulation study of methane adsorption in metal−organic frameworks, Energy Fuels 21 (2007) 953–956, https://doi.org/10.1021/ef060578f.
[58] R. Babarao, J. Jiang, Molecular screening of metal−organic frameworks for CO_2 Storage, Langmuir. 24 (2008) 6270–6278, https://doi.org/10.1021/la800369s.
[59] T. Düren, R.Q. Snurr, Assessment of isoreticular metal−organic frameworks for adsorption separations: a molecular simulation study of methane/n-butane mixtures, J. Phys. Chem. B. 108 (2004) 15703–15708, https://doi.org/10.1021/jp0477856.
[60] R. Babarao, J. Jiang, S.I. Sandler, Molecular simulations for adsorptive separation of CO_2/CH_4 mixture in metal-exposed, catenated, and charged metal−organic frameworks, Langmuir. 25 (2009) 5239–5247, https://doi.org/10.1021/la803074g.
[61] A. Sturluson, M.T. Huynh, A.R. Kaija, C. Laird, S. Yoon, F. Hou, Z. Feng, C.E. Wilmer, Y.J. Colón, Y.G. Chung, D.W. Siderius, C.M. Simon, The role of molecular modelling and simulation in the discovery and deployment of metal-organic frameworks for gas storage and separation, Mol. Simul. 45 (2019) 1082–1121, https://doi.org/10.1080/08927022.2019.1648809.
[62] L. Sarkisov, T. Düren, R.Q. Snurr, Molecular modelling of adsorption in novel nanoporous metal–organic materials, Mol. Phys. 102 (2004) 211–221, https://doi.org/10.1080/00268970310001654854.
[63] A.I. Skoulidas, Molecular dynamics simulations of gas diffusion in metal−organic frameworks: argon in CuBTC, J. Am. Chem. Soc. 126 (2004) 1356–1357, https://doi.org/10.1021/ja039215+.
[64] H. Daglar, S. Keskin, Recent advances, opportunities, and challenges in high-throughput computational screening of MOFs for gas separations, Coord. Chem. Rev. 422 (2020) 213470, https://doi.org/10.1016/j.ccr.2020.213470.
[65] R. Babarao, J. Jiang, Unraveling the energetics and dynamics of ibuprofen in mesoporous metal−organic frameworks, J. Phys. Chem. C. 113 (2009) 18287–18291, https://doi.org/10.1021/jp906429s.
[66] R. Bueno-Perez, A. Martin-Calvo, P. Gómez-Álvarez, J.J. Gutiérrez-Sevillano, P. J. Merkling, T.J.H. Vlugt, T.S. van Erp, D. Dubbeldam, S. Calero, Enantioselective adsorption of ibuprofen and lysine in metal−organic frameworks, Chem. Commun. 50 (2014) 10849–10852, https://doi.org/10.1039/C4CC03745F.
[67] I. Erucar, S. Keskin, Efficient storage of drug and cosmetic molecules in biocompatible metal organic frameworks: a molecular simulation study, Ind. Eng. Chem. Res. 55 (2016) 1929–1939, https://doi.org/10.1021/acs.iecr.5b04556.
[68] M.C. Bernini, D. Fairen-Jimenez, M. Pasinetti, A.J. Ramirez-Pastor, R.Q. Snurr, Screening of bio-compatible metal–organic frameworks as potential drug carriers using Monte Carlo simulations, J. Mater. Chem. B. 2 (2014) 766–774, https://doi.org/10.1039/C3TB21328E.
[69] B. Liu, Y. Lian, Z. Li, G. Chen, Molecular simulation of drug adsorption and diffusion in bio-MOFs, Acta Chim. Sin. 72 (8) (2014) 942, https://doi.org/10.6023/A14030221.
[70] I. Erucar, S. Keskin, Computational investigation of metal organic frameworks for storage and delivery of anticancer drugs, J. Mater. Chem. B. 5 (2017) 7342–7351, https://doi.org/10.1039/C7TB01764B.
[71] J.-Q. Liu, X.-F. Li, C.-Y. Gu, J.C.S. da Silva, A.L. Barros, S. Alves-Jr, B.-H. Li, F. Ren, S.R. Batten, T.A. Soares, A combined experimental and computational study of novel nanocage-based metal–organic frameworks for drug delivery, Dalt. Trans. 44 (2015) 19370–19382, https://doi.org/10.1039/C5DT02171E.
[72] J. Wang, J. Jin, F. Li, B. Li, J. Liu, J. Jin, C. Wang, Y. Zeng, Y. Wang, Combined experimental and theoretical insight into the drug delivery of nanoporous metal–organic frameworks, RSC Adv. 5 (2015) 85606–85612, https://doi.org/10.1039/C5RA12543J.
[73] F. Li, B. Li, C. Wang, Y. Zeng, J. Liu, C.-Y. Gu, P. Lu, L. Mei, Encapsulation of pharmaceutical ingredient linker in metal–organic framework: combined experimental and theoretical insight into the drug delivery, RSC Adv. 6 (2016) 47959–47965, https://doi.org/10.1039/C6RA06178H.
[74] M. Kotzabasaki, E. Tylianakis, E. Klontzas, G.E. Froudakis, OH-functionalization strategy in Metal-Organic Frameworks for drug delivery, Chem. Phys. Lett. 685 (2017) 114–118, https://doi.org/10.1016/j.cplett.2017.07.053.
[75] M. Kotzabasaki, I. Galdadas, E. Tylianakis, E. Klontzas, Z. Cournia, G.E. Froudakis, Multiscale simulations reveal IRMOF-74-III as a potent drug carrier for gemcitabine delivery, J. Mater. Chem. B. 5 (2017) 3277–3282, https://doi.org/10.1039/C7TB00220C.
[76] I.B. Vasconcelos, T.G. da Silva, G.C.G. Militão, T.A. Soares, N.M. Rodrigues, M.O. Rodrigues, N.B. da Costa, R.O. Freire, S.A. Junior, Cytotoxicity and slow release of the anti-cancer drug doxorubicin from ZIF-8, RSC Adv. 2 (2012) 9437–9442, https://doi.org/10.1039/C2RA21087H.
[77] G.A.A. AlHazmi, K.S. AbouMelha, M.G. El-Desouky, A.A. El-Bindary, Effective adsorption of doxorubicin hydrochloride on zirconium metal-organic framework: Equilibrium, kinetic and thermodynamic studies, J. Mol. Struct. 1258 (2022), https://doi.org/10.1016/j.molstruc.2022.132679.
[78] D.-Y. Ma, Z. Li, J.-X. Xiao, R. Deng, P.-F. Lin, R.-Q. Chen, Y.-Q. Liang, H.-F. Guo, B. Liu, J.-Q. Liu, Hydrostable and nitryl/methyl-functionalized metal-organic framework for drug delivery and highly selective CO_2 adsorption, Inorg. Chem. 54 (2015) 6719–6726, https://doi.org/10.1021/acs.inorgchem.5b00335.
[79] M. Kotzabasaki, G.E. Froudakis, Review of computer simulations on anti-cancer drug delivery in MOFs, Inorg. Chem. Front. 5 (2018) 1255–1272, https://doi.org/10.1039/C7QI00645D.
[80] Y. Yang, Y. Yang, Y. Liu, S. Zhao, Z. Tang, Metal-organic frameworks for electrocatalysis: beyond their derivatives, Small Sci. 1 (2021) 2100015, https://doi.org/10.1002/smsc.202100015.
[81] X.-P. Wu, I. Choudhuri, D.G. Truhlar, Computational studies of photocatalysis with metal-organic frameworks, Energy Environ. Mater. 2 (2019) 251–263, https://doi.org/10.1002/eem2.12051.
[82] H. Shabbir, S.P. Vicchio, P. Meza-Morales, R.B. Getman, Role of molecular simulations in the design of metal-organic frameworks for gas-phase thermocatalysis: a perspective, J. Phys. Chem. C. 126 (2022) 6111–6118, https://doi.org/10.1021/acs.jpcc.1c10778.
[83] A.S. Rosen, J.M. Notestein, R.Q. Snurr, Realizing the data-driven, computational discovery of metal-organic framework catalysts, Curr. Opin. Chem. Eng. 35 (2022), https://doi.org/10.1016/j.coche.2021.100760.
[84] X. Mao, C. Tang, T. He, D. Wijethunge, C. Yan, Z. Zhu, A. Du, Computational screening of MN_4 (M = Ti–Cu) based metal organic frameworks for CO_2 reduction using the d-band centre as a descriptor, Nanoscale. 12 (2020) 6188–6194, https://doi.org/10.1039/C9NR09529B.
[85] R. Grau-Crespo, A. Aziz, A.W. Collins, R. Crespo-Otero, N.C. Hernández, L.M. Rodriguez-Albelo, A.R. Ruiz-Salvador, S. Calero, S. Hamad, Modelling a linker

mix-and-match approach for controlling the optical excitation gaps and band alignment of zeolitic imidazolate frameworks, Angew. Chemie Int. Ed. 55 (2016) 16012–16016, https://doi.org/10.1002/anie.201609439.
[86] X. Feng, J. Hajek, H.S. Jena, G. Wang, S.K.P. Veerapandian, R. Morent, N. De Geyter, K. Leyssens, A.E.J. Hoffman, V. Meynen, C. Marquez, D.E. De Vos, V. Van Speybroeck, K. Leus, P. Van Der Voort, Engineering a highly defective stable UiO-66 with tunable lewis- brønsted acidity: the role of the hemilabile linker, J. Am. Chem. Soc. 142 (2020) 3174–3183, https://doi.org/10.1021/jacs.9b13070.
[87] Y.G. Chung, J. Camp, M. Haranczyk, B.J. Sikora, W. Bury, V. Krungleviciute, T. Yildirim, O.K. Farha, D.S. Sholl, R.Q. Snurr, Computation-ready, experimental metal-organic frameworks: a tool to enable high-throughput screening of nanoporous crystals, Chem. Mater. 26 (2014) 6185–6192, https://doi.org/10.1021/cm502594j.
[88] Y.G. Chung, E. Haldoupis, B.J. Bucior, M. Haranczyk, S. Lee, H. Zhang, K.D. Vogiatzis, M. Milisavljevic, S. Ling, J.S. Camp, B. Slater, J.I. Siepmann, D.S. Sholl, R.Q. Snurr, Advances, updates, and analytics for the computation-ready, experimental metal-organic framework database: CoRE MOF 2019, J. Chem. Eng. Data. 64 (2019) 5985–5998, https://doi.org/10.1021/acs.jced.9b00835.
[89] P.Z. Moghadam, A. Li, S.B. Wiggin, A. Tao, A.G.P. Maloney, P.A. Wood, S.C. Ward, D. Fairen-Jimenez, Development of a Cambridge structural database subset: a collection of metal-organic frameworks for past, present, and future, Chem. Mater. 29 (2017) 2618–2625, https://doi.org/10.1021/acs.chemmater.7b00441.
[90] H. Daglar, H.C. Gulbalkan, G. Avci, G.O. Aksu, O.F. Altundal, C. Altintas, I. Erucar, S. Keskin, Effect of metal-organic framework (MOF) database selection on the assessment of gas storage and separation potentials of MOFs, Angew. Chemie Int. Ed. 60 (2021) 7828–7837, https://doi.org/10.1002/anie.202015250.
[91] C. Altintas, G. Avci, H. Daglar, A. Nemati Vesali Azar, I. Erucar, S. Velioglu, S. Keskin, 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. 7 (16) (2019) 9593–9608, https://doi.org/10.1039/C9TA01378D.
[92] C.E. Wilmer, M. Leaf, C.Y. Lee, O.K. Farha, B.G. Hauser, J.T. Hupp, R.Q. Snurr, Large-scale screening of hypothetical metal-organic frameworks, Nat. Chem. 4 (2012) 83–89, https://doi.org/10.1038/nchem.1192.
[93] Y.J. Colón, D.A. Gómez-Gualdrón, R.Q. Snurr, Topologically guided, automated construction of metal-organic frameworks and their evaluation for energy-related applications, Cryst. Growth Des. 17 (2017) 5801–5810, https://doi.org/10.1021/acs.cgd.7b00848.
[94] P.G. Boyd, T.K. Woo, A generalized method for constructing hypothetical nanoporous materials of any net topology from graph theory, CrystEngComm. 18 (2016) 3777–3792, https://doi.org/10.1039/C6CE00407E.
[95] P.G. Boyd, Y. Lee, B. Smit, Computational development of the nanoporous materials genome, Nat. Rev. Mater. 2 (2017) 17037, https://doi.org/10.1038/natrevmats.2017.37.
[96] E. Haldoupis, S. Nair, D.S. Sholl, Efficient calculation of diffusion limitations in metal organic framework materials: a tool for identifying materials for kinetic separations, J. Am. Chem. Soc. 132 (2010) 7528–7539, https://doi.org/10.1021/ja1023699.
[97] E. Haldoupis, S. Nair, D.S. Sholl, 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. 134 (2012) 4313–4323, https://doi.org/10.1021/ja2108239.
[98] T. Watanabe, D.S. Sholl, Accelerating applications of metal-organic frameworks for gas adsorption and separation by computational screening of materials, Langmuir. 28 (2012) 14114–14128, https://doi.org/10.1021/la301915s.
[99] C. Altintas, G. Avci, H. Daglar, A. Nemati Vesali Azar, S. Velioglu, I. Erucar, S. Keskin, Database for CO_2 separation performances of MOFs based on computational materials screening, ACS Appl. Mater. Interfaces. 10 (20) (2018) 17257–17268, https://doi.org/10.1021/acsami.8b04600.
[100] G. Avci, S. Velioglu, S. Keskin, High-throughput screening of MOF adsorbents and membranes for H_2 purification and CO_2 capture, ACS Appl. Mater. Interfaces. 10 (2018) 33693–33706, https://doi.org/10.1021/acsami.8b12746.
[101] L.M. Robeson, The upper bound revisited, J. Memb. Sci. 320 (2008) 390–400, https://doi.org/10.1016/j.memsci.2008.04.030.
[102] K.T. Leperi, Y.G. Chung, F. You, R.Q. Snurr, Development of a general evaluation metric for rapid screening of adsorbent materials for postcombustion CO_2 capture, ACS Sustain. Chem. Eng. 7 (2019) 11529–11539, https://doi.org/10.1021/acssuschemeng.9b01418.
[103] S. Li, Y.G. Chung, R.Q. Snurr, High-throughput screening of metal-organic frameworks for CO_2 capture in the presence of water, Langmuir. 32 (2016) 10368–10376, https://doi.org/10.1021/acs.langmuir.6b02803.
[104] N.S. Bobbitt, J. Chen, R.Q. Snurr, High-throughput screening of metal-organic frameworks for hydrogen storage at cryogenic temperature, J. Phys. Chem. C. 120 (2016) 27328–27341, https://doi.org/10.1021/acs.jpcc.6b08729.
[105] H. Chen, R.Q. Snurr, Computational screening of metal-catecholate-functionalized metal-organic frameworks for room-temperature hydrogen storage, J. Phys. Chem. C. 125 (2021) 21701–21708, https://doi.org/10.1021/acs.jpcc.1c06549.
[106] A. Ahmed, S. Seth, J. Purewal, A.G. Wong-Foy, M. Veenstra, A.J. Matzger, D.J. Siegel, Exceptional hydrogen storage achieved by screening nearly half a million metal-organic frameworks, Nat. Commun. 10 (2019) 1568, https://doi.org/10.1038/s41467-019-09365-w.
[107] K.D. Vogiatzis, E. Haldoupis, D.J. Xiao, J.R. Long, J.I. Siepmann, L. Gagliardi, Accelerated computational analysis of metal-organic frameworks for oxidation catalysis, J. Phys. Chem. C. 120 (2016) 18707–18712, https://doi.org/10.1021/acs.jpcc.6b07115.
[108] A.S. Rosen, J.M. Notestein, R.Q. Snurr, Identifying promising metal–organic frameworks for heterogeneous catalysis via high-throughput periodic density functional theory, J. Comput. Chem. 40 (2019) 1305–1318, https://doi.org/10.1002/jcc.25787.
[109] A.S. Rosen, J.M. Notestein, R.Q. Snurr, Structure-activity relationships that identify metal-organic framework catalysts for methane activation, ACS Catal. 9 (2019) 3576–3587, https://doi.org/10.1021/acscatal.8b05178.
[110] H. Lyu, Z. Ji, S. Wuttke, O.M. Yaghi, Digital reticular chemistry, Chem. 6 (2020) 2219–2241, https://doi.org/10.1016/j.chempr.2020.08.008.
[111] K.M. Jablonka, D. Ongari, S.M. Moosavi, B. Smit, Big-data science in porous materials: materials genomics and machine learning, Chem. Rev. 120 (2020) 8066–8129, https://doi.org/10.1021/acs.chemrev.0c00004.
[112] M. Fernandez, T.K. Woo, C.E. Wilmer, R.Q. Snurr, Large-scale quantitative structure-property relationship (QSPR) analysis of methane storage in metal-organic frameworks, J. Phys. Chem. C. 117 (2013) 7681–7689, https://doi.org/10.1021/jp4006422.
[113] M. Fernandez, N.R. Trefiak, T.K. Woo, Atomic property weighted radial distribution functions descriptors of metal-organic frameworks for the prediction of gas uptake capacity, J. Phys. Chem. C. 117 (2013) 14095–14105, https://doi.org/10.1021/jp404287t.
[114] E. Ren, P. Guilbaud, F.-X. Coudert, High-throughput computational screening of nanoporous materials in targeted applications, Digit. Discov. 1 (2022) 355–374, https://doi.org/10.1039/D2DD00018K.
[115] C. Altintas, O.F. Altundal, S. Keskin, R. Yildirim, Machine learning meets with metal organic frameworks for gas storage and separation, J. Chem. Inf. Model. 61 (2021) 2131–2146, https://doi.org/10.1021/acs.jcim.1c00191.
[116] M.Z. Aghaji, M. Fernandez, P.G. Boyd, T.D. Daff, T.K. Woo, Quantitative structure-property relationship models for recognizing metal organic frameworks (MOFs) with high CO_2 working capacity and CO_2/CH_4 selectivity for methane purification, Eur. J. Inorg. Chem. 2016 (2016) 4505–4511, https://doi.org/10.1002/ejic.201600365.
[117] Z. Wang, Y. Zhou, T. Zhou, K. Sundmacher, Identification of optimal metal-organic frameworks by machine learning: structure decomposition, feature integration, and predictive modeling, Comput. Chem. Eng. 160 (2022), https://doi.org/10.1016/j.compchemeng.2022.107739.
[118] Z. Li, B.J. Bucior, H. Chen, M. Haranczyk, J.I. Siepmann, R.Q. Snurr, Machine learning using host/guest energy histograms to predict adsorption in metal–organic frameworks: Application to short alkanes and Xe/Kr mixtures, J. Chem. Phys. 155 (2021) 14701, https://doi.org/10.1063/5.0050823.
[119] R. Anderson, D.A. Gómez-Gualdrón, Increasing topological diversity during computational "synthesis" of porous crystals: how and why, CrystEngComm. 21 (2019) 1653–1665, https://doi.org/10.1039/C8CE01637B.
[120] A.S. Rosen, S.M. Iyer, D. Ray, Z. Yao, A. Aspuru-Guzik, L. Gagliardi, J.M. Notestein, R.Q. Snurr, Machine learning the quantum-chemical properties of metal–organic frameworks for accelerated materials discovery, Matter. 4 (2021) 1578–1597, https://doi.org/10.1016/j.matt.2021.02.015.
[121] S.M. Moosavi, A. Nandy, K.M. Jablonka, D. Ongari, J.P. Janet, P.G. Boyd, Y. Lee, B. Smit, H.J. Kulik, Understanding the diversity of the metal-organic framework ecosystem, Nat. Commun. 11 (2020) 4068, https://doi.org/10.1038/s41467-020-17755-8.
[122] H. Daglar, S. Keskin, Combining machine learning and molecular simulations to unlock gas separation potentials of MOF membranes and MOF/polymer MMMs, ACS Appl. Mater. Interfaces. 14 (2022) 32134–32148, https://doi.org/10.1021/acsami.2c08977.
[123] J. Abdi, F. Hadavimoghaddam, M. Hadipoor, A. Hemmati-Sarapardeh, Modeling of CO_2 adsorption capacity by porous metal organic frameworks using advanced decision tree-based models, Sci. Rep. 11 (2021) 24468, https://doi.org/10.1038/s41598-021-04168-w.
[124] K. Mukherjee, A.W. Dowling, Y.J. Colón, Sequential design of adsorption simulations in metal–organic frameworks, Mol. Syst. Des. Eng. 7 (2022) 248–259, https://doi.org/10.1039/D1ME00138H.
[125] G.S. Fanourgakis, K. Gkagkas, E. Tylianakis, G. Froudakis, A generic machine learning algorithm for the prediction of gas adsorption in nanoporous materials, J. Phys. Chem. C. 124 (2020) 7117–7126, https://doi.org/10.1021/acs.jpcc.9b10766.
[126] M. Pardakhti, P. Nanda, R. Srivastava, Impact of chemical features on methane adsorption by porous materials at varying pressures, J. Phys. Chem. C. 124 (2020) 4534–4544, https://doi.org/10.1021/acs.jpcc.9b09319.
[127] I. Tsamardinos, G.S. Fanourgakis, E. Greasidou, E. Klontzas, K. Gkagkas, G.E. Froudakis, An Automated Machine Learning architecture for the accelerated prediction of Metal-Organic Frameworks performance in energy and environmental applications, Microporous Mesoporous Mater. 300 (2020), https://doi.org/10.1016/j.micromeso.2020.110160.
[128] G.S. Fanourgakis, K. Gkagkas, G. Froudakis, Introducing artificial MOFs for improved machine learning predictions: Identification of top-performing materials for methane storage, J. Chem. Phys. 156 (2022) 54103, https://doi.org/10.1063/5.0075994.
[129] M. Gheytanzadeh, A. Baghban, S. Habibzadeh, A. Esmaeili, O. Abida, A. Mohaddespour, M.T. Munir, Towards estimation of CO_2 adsorption on highly porous MOF-based adsorbents using gaussian process regression approach, Sci. Rep. 11 (2021) 15710, https://doi.org/10.1038/s41598-021-95246-6.

[130] T.D. Burns, K.N. Pai, S.G. Subraveti, S.P. Collins, M. Krykunov, A. Rajendran, T. K. Woo, 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. 54 (2020) 4536–4544, https://doi.org/10.1021/acs.est.9b07407.
[131] P.Z. Moghadam, S.M.J. Rogge, A. Li, C.-M. Chow, J. Wieme, N. Moharrami, M. Aragones-Anglada, G. Conduit, D.A. Gomez-Gualdron, V. Van Speybroeck, D. Fairen-Jimenez, Structure-mechanical stability relations of metal-organic frameworks via machine learning, Matter. 1 (2019) 219–234, https://doi.org/10.1016/j.matt.2019.03.002.
[132] A. Nandy, G. Terrones, N. Arunachalam, et al., MOFSimplify, machine learning models with extracted stability data of three thousand metal–organic frameworks, Sci Data 9 (2022) 74, https://doi.org/10.1038/s41597-022-01181-0.
[133] A. Singh, N. Thakur, A. Sharma, A review of supervised machine learning algorithms, in: 2016 3rd Int. Conf. Comput. Sustain. Glob. Dev., 2016: pp. 1310–1315.
[134] S. Chong, S. Lee, B. Kim, J. Kim, Applications of machine learning in metal-organic frameworks, Coord. Chem. Rev. 423 (2020) 213487, https://doi.org/10.1016/j.ccr.2020.213487.
[135] P. Dhal, C. Azad, A comprehensive survey on feature selection in the various fields of machine learning, Appl. Intell. 52 (2022) 4543–4581, https://doi.org/10.1007/s10489-021-02550-9.
[136] Y. Bao, R.L. Martin, M. Haranczyk, M.W. Deem, In silico prediction of MOFs with high deliverable capacity or internal surface area, Phys. Chem. Chem. Phys. 17 (2015) 11962–11973, https://doi.org/10.1039/C5CP00002E.
[137] Y. Bao, R.L. Martin, C.M. Simon, M. Haranczyk, B. Smit, M.W. Deem, In Silico discovery of high deliverable capacity metal-organic frameworks, J. Phys. Chem. C. 119 (2015) 186–195, https://doi.org/10.1021/jp5123486.
[138] R.S. Olson, J.H.M.B.T.-P. of the W. on A.M. Learning, TPOT: A Tree-based Pipeline Optimization Tool for Automating Machine Learning, 64 (2016) 66–74.
[139] M.-A. Schulz, B.T.T. Yeo, J.T. Vogelstein, J. Mourao-Miranada, J.N. Kather, K. Kording, B. Richards, D. Bzdok, Different scaling of linear models and deep learning in UKBiobank brain images versus machine-learning datasets, Nat. Commun. 11 (2020) 4238, https://doi.org/10.1038/s41467-020-18037-z.
[140] Z. Yao, B. Sánchez-Lengeling, N.S. Bobbitt, B.J. Bucior, S.G.H. Kumar, S.P. Collins, T. Burns, T.K. Woo, O.K. Farha, R.Q. Snurr, A. Aspuru-Guzik, Inverse design of nanoporous crystalline reticular materials with deep generative models, Nat. Mach. Intell. 3 (2021) 76–86, https://doi.org/10.1038/s42256-020-00271-1.
[141] D. Weininger, SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules, J. Chem. Inf. Comput. Sci. 28 (1988) 31–36, https://doi.org/10.1021/ci00057a005.
[142] P. Minkiewicz, A. Iwaniak, M. Darewicz, Annotation of peptide structures using SMILES and other chemical codes-practical solutions, Molecules. 22 (12) (2017) 2075, https://doi.org/10.3390/molecules22122075
[143] R.M. Hanson, Jmol SMILES and Jmol SMARTS: specifications and applications, J. Cheminform. 8 (2016) 50, https://doi.org/10.1186/s13321-016-0160-4.
[144] N. O'Boyle, A. Dalke, DeepSMILES: an adaptation of SMILES for use in machine-learning of chemical structures, ChemRxiv (2018), https://doi.org/10.26434/chemrxiv.7097960.v1.
[145] S. Heller, A. McNaught, S. Stein, D. Tchekhovskoi, I. Pletnev, InChI - the worldwide chemical structure identifier standard, J. Cheminform. 5 (2013) 7, https://doi.org/10.1186/1758-2946-5-7.
[146] B.J. Bucior, A.S. Rosen, M. Haranczyk, Z. Yao, M.E. Ziebel, O.K. Farha, J.T. Hupp, J.I. Siepmann, A. Aspuru-Guzik, R.Q. Snurr, Identification schemes for metal-organic frameworks to enable rapid search and cheminformatics analysis, Cryst. Growth Des. 19 (2019) 6682–6697, https://doi.org/10.1021/acs.cgd.9b01050.
[147] MOFid website, 2019. https://snurr-group.github.io/web-mofid/ (accessed November 1, 2022).
[148] R. Gurnani, Z. Yu, C. Kim, D.S. Sholl, R. Ramprasad, Interpretable machine learning-based predictions of methane uptake isotherms in metal-organic frameworks, Chem. Mater. 33 (2021) 3543–3552, https://doi.org/10.1021/acs.chemmater.0c04729.
[149] A. Nigam, R. Pollice, M. Krenn, G.D.P. Gomes, A. Aspuru-Guzik, Beyond generative models: superfast traversal, optimization, novelty, exploration and discovery (STONED) algorithm for molecules using SELFIES, Chem. Sci. 12 (20) (2021) 7079–7090, https://doi.org/10.1039/D1SC00231G.
[150] M. Krenn, Q. Ai, S. Barthel, N. Carson, A. Frei, N.C. Frey, P. Friederich, T. Gaudin, A.A. Gayle, K.M. Jablonka, R.F. Lameiro, D. Lemm, A. Lo, S.M. Moosavi, J.M. Nápoles-Duarte, A. Nigam, R. Pollice, K. Rajan, U. Schatzschneider, P. Schwaller, M. Skreta, B. Smit, F. Strieth-Kalthoff, C. Sun, G. Tom, G. Falk von Rudorff, A. Wang, A.D. White, A. Young, R. Yu, A. Aspuru-Guzik, SELFIES and the future of molecular string representations, Patterns 3 (10) (2022) 100588, https://doi.org/10.1016/j.patter.2022.100588.
[151] S.G. Baird, K.M. Jablonka, M.D. Alverson, H.M. Sayeed, M.F. Khan, C. Seegmiller, B. Smit, T.D. Sparks, xtal2png: A Python package for representing crystal structure as PNG files, J. Open Source Softw. 7 (2022) 4528, https://doi.org/10.21105/joss.04528.
[152] J.D. Evans, V. Bon, I. Senkovska, S. Kaskel, A universal standard archive file for adsorption data, Langmuir. 37 (2021) 4222–4226, https://doi.org/10.1021/acs.langmuir.1c00122.
[153] M.D. Wilkinson, M. Dumontier, I.J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L.B. da Silva Santos, P.E. Bourne, J. Bouwman, A.J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C.T. Evelo, R. Finkers, A. Gonzalez-Beltran, A.J.G. Gray, P. Groth, C. Goble, J.S. Grethe, J. Heringa, P.A.C. 't Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S.J. Lusher, M.E. Martone, A. Mons, A.L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S.-A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M.A. Swertz, M. Thompson, J. van der Lei, E. van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao, B. Mons, The FAIR Guiding Principles for scientific data management and stewardship, Sci. Data. 3 (1) (2016), https://doi.org/10.1038/sdata.2016.18.
[154] D. Ongari, L. Talirz, K.M. Jablonka, D.W. Siderius, B. Smit, Data-driven matching of experimental crystal structures and gas adsorption isotherms of metal-organic frameworks, J. Chem. Eng. Data. 67 (2022) 1743–1756, https://doi.org/10.1021/acs.jced.1c00958.
[155] D.W. Siderius, V.K. Shen, R.D. Johnson III, R.D. van Zee, NIST/ARPA-E database of novel and emerging adsorbent materials, Natl Inst. Stand. Technol., Gaithersburg, MD, 2014, https://doi.org/10.18434/T43882.
[156] J. Burner, J. Luo, A. White, A. Mirmiran, O. Kwon, P.G. Boyd, S. Maley, M. Gibaldi, S. Simrod, V. Ogden, T.K. Woo, ARC–MOF: a diverse database of metal-organic frameworks with DFT-derived partial atomic charges and descriptors for machine learning, Chem. Mater. 35 (2023) 900–916, https://doi.org/10.1021/acs.chemmater.2c02485.
[157] A. Nandy, S. Yue, C. Oh, C. Duan, G.G. Terrones, Y.G. Chung, H.J. Kulik, A Database of Ultrastable MOFs Reassembled from Stable Fragments with Machine Learning Models, ArXiv (2022), https://doi.org/10.48550/arXiv.2210.14191.
[158] A. Nandy, C. Duan, H.J. Kulik, Using machine learning and data mining to leverage community knowledge for the engineering of stable metal-organic frameworks, J. Am. Chem. Soc. 143 (2021) 17535–17547, https://doi.org/10.1021/jacs.1c07217.
[159] H. Luo, F. Cheng, L. Huelsenbeck, N. Smith, Comparison between conventional solvothermal and aqueous solution-based production of UiO-66-NH_2: Life cycle assessment, techno-economic assessment, and implications for CO_2 capture and storage, J. Environ. Chem. Eng. 9 (2) (2021) 105159, https://doi.org/10.1016/j.jece.2021.105159.
[160] K. Gubsch, R. Bence, L. Glasby, P.Z. Moghadam, DigiMOF: A Database of MOF Synthesis Information Generated via Text Mining, ChemRxiv. (2022). https://doi.org/10.26434/chemrxiv-2022-41t70.
[161] M.C. Swain, J.M. Cole, ChemDataExtractor: a toolkit for automated extraction of chemical information from the scientific literature, J. Chem. Inf. Model. 56 (2016) 1894–1904, https://doi.org/10.1021/acs.jcim.6b00207.
[162] H. Park, Y. Kang, W. Choe, J. Kim, Mining insights on metal-organic framework synthesis from scientific literature texts, J. Chem. Inf. Model. 62 (2022) 1190–1198, https://doi.org/10.1021/acs.jcim.1c01297.
[163] Y. Luo, S. Bag, O. Zaremba, A. Cierpka, J. Andreo, S. Wuttke, P. Friederich, M. Tsotsalas, MOF synthesis prediction enabled by automatic data mining and machine learning, Angew. Chemie Int. Ed. 61 (2022), https://doi.org/10.1002/anie.202200242.
[164] P. Friederich, M. Tsotsalas, MOF Synthesis Prediction Tool, 2022. https://mof-synthesis.aimat.science (accessed November 13, 2022).
[165] Y. Xie, C. Zhang, X. Hu, C. Zhang, S.P. Kelley, J.L. Atwood, J. Lin, Machine learning assisted synthesis of metal-organic nanocapsules, J. Am. Chem. Soc. 142 (2020) 1475–1481, https://doi.org/10.1021/jacs.9b11569.
[166] B. P. Domingues, N.; Moosavi, S. M.; Talirz, L.; P. Ireland, C.; Ebrahim, F. M.; Smit, Using Genetic Algorithms to Systematically Improve the Synthesis Conditions of Al-PMOF, ChemRxiv. (2022). https://doi.org/10.26434/chemrxiv-2022-cgnf5.
[167] A. Fateeva, P.A. Chater, C.P. Ireland, A.A. Tahir, Y.Z. Khimyak, P.V. Wiper, J.R. Darwent, M.J. Rosseinsky, A water-stable porphyrin-based metal-organic framework active for visible-light photocatalysis, Angew. Chemie Int. Ed. 51 (2012) 7440–7444, https://doi.org/10.1002/anie.201202471.
[168] A.J. Howarth, Y. Liu, P. Li, Z. Li, T.C. Wang, J.T. Hupp, O.K. Farha, Chemical, thermal and mechanical stabilities of metal–organic frameworks, Nat. Rev. Mater. 1 (2016) 15018, https://doi.org/10.1038/natrevmats.2015.18.
[169] R. Batra, C. Chen, T.G. Evans, K.S. Walton, R. Ramprasad, Prediction of water stability of metal–organic frameworks using machine learning, Nat. Mach. Intell. 2 (2020) 704–710, https://doi.org/10.1038/s42256-020-00249-z.
[170] H. Wu, T. Yildirim, W. Zhou, Exceptional mechanical stability of highly porous zirconium metal-organic framework UiO-66 and its important implications, J. Phys. Chem. Lett. 4 (2013) 925–930, https://doi.org/10.1021/jz4002345.
[171] M. Rubio-Martinez, C. Avci-Camur, A.W. Thornton, I. Imaz, D. Maspoch, M.R. Hill, New synthetic routes towards MOF production at scale, Chem. Soc. Rev. 46 (2017) 3453–3480, https://doi.org/10.1039/C7CS00109F.
[172] H.U. Escobar-Hernandez, L.M. Pérez, P. Hu, F.A. Soto, M.I. Papadaki, H.-C. Zhou, Q. Wang, Thermal stability of metal-organic frameworks (MOFs): concept, determination, and model prediction using computational chemistry and machine learning, Ind. Eng. Chem. Res. 61 (2022) 5853–5862, https://doi.org/10.1021/acs.iecr.2c00561.
[173] A. Nandy, G. Terrones, N. Arunachalam, C. Duan, D.W. Kastner, H.J. Kulik, MOFSimplify, machine learning models with extracted stability data of three thousand metal–organic frameworks, Sci. Data. 9 (2022) 74, https://doi.org/10.1038/s41597-022-01181-0.
[174] R. Pétuya, S. Durdy, D. Antypov, M.W. Gaultois, N.G. Berry, G.R. Darling, A.P. Katsoulidis, M.S. Dyer, M.J. Rosseinsky, Machine-learning prediction of metal-organic framework guest accessibility from linker and metal chemistry, Angew. Chemie Int. Ed. 61 (2022), https://doi.org/10.1002/anie.202114573.
[175] P. Krokidas, S. Karozis, S. Moncho, G. Giannakopoulos, E.N. Brothers, M.E. Kainourgiakis, I.G. Economou, T.A. Steriotis, Data mining for predicting gas

diffusivity in zeolitic-imidazolate frameworks (ZIFs), J. Mater. Chem. A. 10 (2022) 13697–13703, https://doi.org/10.1039/D2TA02624D.
[176] Y. Lim, J. Park, S. Lee, J. Kim, Finely tuned inverse design of metal–organic frameworks with user-desired Xe/Kr selectivity, J. Mater. Chem. A. 9 (2021) 21175–21183, https://doi.org/10.1039/D1TA03129E.
[177] B. Kim, S. Lee, J. Kim, Inverse design of porous materials using artificial neural networks, Sci. Adv. 6 (2020) eaax9324, https://doi.org/10.1126/sciadv.aax9324.
[178] B.J. Bucior, N.S. Bobbitt, T. Islamoglu, S. Goswami, A. Gopalan, T. Yildirim, O.K. Farha, N. Bagheri, R.Q. Snurr, Energy-based descriptors to rapidly predict hydrogen storage in metal–organic frameworks, Mol. Syst. Des. Eng. 4 (2019) 162–174, https://doi.org/10.1039/C8ME00050F.
[179] Y. Kang, H. Park, B. Smit, et al., A multi-modal pre-training transformer for universal transfer learning in metal–organic frameworks, Nat. Mach. Intell. (2023), https://doi.org/10.1038/s42256-023-00628-2.
[180] I.B. Orhan, H. Daglar, S. Keskin, T.C. Le, R. Babarao, Prediction of O_2/N_2 selectivity in metal-organic frameworks via high-throughput computational screening and machine learning, ACS Appl. Mater. Interfaces. 14 (2022) 736–749, https://doi.org/10.1021/acsami.1c18521.
[181] Tian Xie, C. Grossman Jeffrey, Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties, Physical review letters 120 (14) (2018) 145301, https://doi.org/10.1103/PhysRevLett.120.145301.
[182] Z. Cao, R. Magar, Y. Wang, A. Barati Farimani, MOFormer: self-supervised transformer model for metal-organic framework property prediction, J. Am. Chem. Soc. 145 (2023) 2958–2967, https://doi.org/10.1021/jacs.2c11420.
[183] A.P. Bartók, S. De, C. Poelking, N. Bernstein, J.R. Kermode, G. Csányi, M. Ceriotti, Machine learning unifies the modeling of materials and molecules, Sci. Adv. 3 (2022), https://doi.org/10.1126/sciadv.1701816.
[184] B. Meredig, A. Agrawal, S. Kirklin, J.E. Saal, J.W. Doak, A. Thompson, K. Zhang, A. Choudhary, C. Wolverton, Combinatorial screening for new materials in unconstrained composition space with machine learning, Phys. Rev. B. 89 (2014) 94104, https://doi.org/10.1103/PhysRevB.89.094104.
[185] Y. Sun, R.F. DeJaco, Z. Li, D. Tang, S. Glante, D.S. Sholl, C.M. Colina, R.Q. Snurr, M. Thommes, M. Hartmann, J.I. Siepmann, Fingerprinting diverse nanoporous materials for optimal hydrogen storage conditions using meta-learning, Sci. Adv. 7 (2022) eabg3983, https://doi.org/10.1126/sciadv.abg3983.
[186] A.W. Thornton, C.M. Simon, J. Kim, O. Kwon, K.S. Deeg, K. Konstas, S.J. Pas, M. R. Hill, D.A. Winkler, M. Haranczyk, B. Smit, Materials genome in action: identifying the performance limits of physical hydrogen storage, Chem. Mater. 29 (2017) 2844–2854, https://doi.org/10.1021/acs.chemmater.6b04933.
[187] M. Haghighatlari, G. Vishwakarma, D. Altarawy, R. Subramanian, B.U. Kota, A. Sonpal, S. Setlur, J. Hachmann, ChemML: A machine learning and informatics program package for the analysis, mining, and modeling of chemical and materials data, WIREs Comput. Mol. Sci. 10 (2020) e1458, https://doi.org/10.1002/wcms.1458.
[188] K.M. Jablonka, A.S. Rosen, A.S. Krishnapriyan, B. Smit, An ecosystem for digital reticular chemistry, ChemRxiv. (2022) 1–52, https://doi.org/10.26434/chemrxiv-2022-4g7rx.
[189] P.G. Boyd, A. Chidambaram, E. García-Díez, C.P. Ireland, T.D. Daff, R. Bounds, A. Gładysiak, P. Schouwink, S.M. Moosavi, M.M. Maroto-Valer, J.A. Reimer, J.A. R. Navarro, T.K. Woo, S. Garcia, K.C. Stylianou, B. Smit, Data-driven design of metal–organic frameworks for wet flue gas CO_2 capture, Nature. 576 (2019) 253–256, https://doi.org/10.1038/s41586-019-1798-7.
[190] D. Widdowson, M.M. Mosca, A. Pulido, A.I. Cooper, V. Kurlin, Average Minimum Distances of periodic point sets-foundational invariants for mapping periodic crystals, In: Mathematical and in Computer Chemistry. 87, (2022), pp. 529–559.
[191] S. Kancharlapalli, A. Gopalan, M. Haranczyk, R.Q. Snurr, Fast and accurate machine learning strategy for calculating partial atomic charges in metal-organic frameworks, J. Chem. Theory Comput. 17 (2021) 3052–3064, https://doi.org/10.1021/acs.jctc.0c01229.
[192] D. Ongari, L. Talirz, B. Smit, Too many materials and too many applications: an experimental problem waiting for a computational solution, ACS Cent. Sci. 6 (2020) 1890–1900, https://doi.org/10.1021/acscentsci.0c00988.
[193] A.-Y.-T. Wang, R.J. Murdock, S.K. Kauwe, A.O. Oliynyk, A. Gurlo, J. Brgoch, K.A. Persson, T.D. Sparks, Machine learning for materials scientists: an introductory guide toward best practices, Chem. Mater. 32 (2020) 4954–4965, https://doi.org/10.1021/acs.chemmater.0c01907.
[194] X. Liu, Y. Wang, J. Yuan, X. Li, S. Wu, Y. Bao, Z. Feng, F. Ou, Y. He, Prediction of the ibuprofen loading capacity of MOFs by machine learning, Bioengineering. 9 (10) (2022) 517, https://doi.org/10.3390/bioengineering9100517.
[195] B. Schweitzer, C. Archuleta, B. Seong, R. Anderson, D.A. Gómez-Gualdrón, Electronic effects due to organic linker-metal surface interactions: implications on screening of MOF-encapsulated catalysts, Phys. Chem. Chem. Phys. 22 (2020) 2475–2487, https://doi.org/10.1039/C9CP05380H.
[196] S. Li, Y. Zhang, Y. Hu, B. Wang, S. Sun, X. Yang, H. He, Predicting metal-organic frameworks as catalysts to fix carbon dioxide to cyclic carbonate by machine learning, J. Mater. 7 (2021) 1029–1038, https://doi.org/10.1016/j.jmat.2021.02.005.
[197] A.L. Dzubak, L.-C. Lin, J. Kim, J.A. Swisher, R. Poloni, S.N. Maximoff, B. Smit, L. Gagliardi, Ab initio carbon capture in open-site metal–organic frameworks, Nat Chem. 4 (2012) 810–816, https://doi.org/10.1038/nchem.1432.
[198] B. Burger, P.M. Maffettone, V.V. Gusev, C.M. Aitchison, Y. Bai, X. Wang, X. Li, B. M. Alston, B. Li, R. Clowes, N. Rankin, B. Harris, R.S. Sprick, A.I. Cooper, A mobile robotic chemist, Nature. 583 (2020) 237–241, https://doi.org/10.1038/s41586-020-2442-2.