The role of molecular modelling and simulation in the discovery and deployment of metal-organic frameworks for gas storage and separation

Arni Sturluson[1], **Melanie T. Huynh**[1], **Alec R. Kaija**[2], **Caleb Laird**[1], **Sunghyun Yoon**[3], **Feier Hou**[4], **Zhenxing Feng**[1], **Christopher E. Wilmer**[2,*], **Yamil J. Colón**[5,†], **Yongchul G. Chung**[3,‡], **Daniel W. Siderius**[6,§], **Cory M. Simon**[1,¶]

[1]School of Chemical, Biological, and Environmental Engineering, Oregon State University. Corvallis, OR, USA

[2]Department of Chemical and Petroleum Engineering, University of Pittsburgh, Pittsburgh, PA, USA

[3]School of Chemical and Biomolecular Engineering, Pusan National University, Busan, Korea (South)

[4]Western Oregon University. Department of Chemistry, Monmouth, OR, USA

[5]Department of Chemical and Biomolecular Engineering, University of Notre Dame, Notre Dame, IN, USA

[6]Chemical Sciences Division, National Institute of Standards and Technology. Gaithersburg, MD, USA

Abstract

Metal-organic frameworks (MOFs) are highly tuneable, extended-network, crystalline, nanoporous materials with applications in gas storage, separations, and sensing. We review how molecular models and simulations of gas adsorption in MOFs have informed the discovery of performant MOFs for methane, hydrogen, and oxygen storage, xenon, carbon dioxide, and chemical warfare agent capture, and xylene enrichment. Particularly, we highlight how large, open databases of MOF crystal structures, post-processed to enable molecular simulations, are a platform for computational materials discovery. We discuss how to orient research efforts to routinise the computational discovery of MOFs for adsorption-based engineering applications.

[*]Corresponding author, wilmer@pitt.edu. [†]Corresponding author, ycolon@nd.edu. [‡]Corresponding author, drygchung@gmail.com. [§]Corresponding author, daniel.siderius@nist.gov. [¶]Corresponding author, Cory.Simon@oregonstate.edu.

[8]https://adsorption.nist.gov/isodb

I. INTRODUCTION

> I can't see exactly what would happen, but I can hardly doubt that when we have
> some control of the arrangement of things on a small scale, we will get an
> enormously greater range of possible properties that substances can have and of
> different things that we can do.
>
> Richard Feynman, American Physical Society Meeting,
> 1959

Metal-organic frameworks (MOFs) [1] are solid-state materials that commonly harbour
nano-sized pores and enormous internal surface areas (> 7 000 m^2/g) [2]. Their consequent
gas adsorption properties lend them applications in storing [3], separating [4], and sensing
[5] gases. Since the first MOF was reported by Omar Yaghi in 1999 [6], a few MOF-based
products have appeared on the market [7], including for safe sub-atmospheric storage of
toxic gases (NuMat Technologies) and carbon dioxide capture in submarines (Mosaic
Materials).

MOFs, heralded as "designer materials" [8], are synthesised modularly by linking organic
molecules, serving as struts, to metals or metal clusters, serving as nodes, to form pre-
determined, extended-network structures [1]. See Figure 1. Owing to their synthetic
adjustability, over 80 000 MOFs displaying diverse pore geometries and surface chemistries
and thus, adsorption properties, have been experimentally reported [9, 10]. MOFs have
garnered much attention because of this ability to exert control over the self-assembly of
linkers and metal nodes/clusters at the nano-scale; judiciously choosing the molecular
building blocks and (sometimes arduously [11]) finding the synthetic conditions to yield a
pre-determined, extended network structure is coined as *reticular chemistry* [12, 13].

For adsorption-based engineering applications, a coveted aim beyond reticular chemistry is
to specify a desired (optimal) adsorption property, then synthesise the MOF that exhibits it.
For example, consider the search for a MOF to densify and store natural gas onboard a
passenger vehicle and deliver it to the engine for fuel [18]. The driving distance of a vehicle
equipped with an adsorbed natural gas fuel tank of a given volume is determined primarily
by the usable capacity of the material [19]; operating at room temperature via a pressure
swing between 65 bar and 5 bar, the usable capacity is the equilibrium methane adsorption
after filling up at the fuel station (65 bar) minus that retained when the tank sustains
insufficient pressure to drive flow to the engine (5 bar)[1]. For the near-term deployment of a
MOF in an adsorbed natural gas fuel tank, a lofty goal is to reliably identify, among the 80
000 experimentally synthesised, which MOF exhibits the maximal methane usable capacity.

---

[1] In addition to thermal and chemical stability and the cost of the MOF, several other material properties influence the performance of
a MOF in adsorbed natural gas storage, such as heat of adsorption, specific heat, thermal conductivity, diffusion coefficients, and
adsorption of impurities that can "poison" adsorption sites (note we approximated natural gas as methane).

Clearly, it is impractically time- and resource-intensive to synthesise, characterise, and measure the high-pressure methane adsorption isotherm in every MOF, despite underway efforts for high-throughput robotic synthesis [20–22] driven by machine learning algorithms [23].

The topic of this review is the role of molecular modelling and simulation in enabling, to an increasing extent, the identification of an existing MOF[2] to exhibit a desired (optimal) adsorption property (e.g., maximal methane usable capacity) through high-throughput computational screenings. The simple idea, encapsulated in the Materials Project [24], is to computationally predict adsorption performance of each existing MOF structure to shortlist materials for experimental synthesis and testing. We take a unique angle from previous reviews [25–27] by highlighting how large, open data sets of MOF crystal structures provide a platform for the computational identification of existing MOFs for gas storage, separation, and sensing.

We begin in Sec. II by briefly describing methodologies to computationally screen MOF structures for gas adsorption applications. We then discuss in Sec. III how MOF crystal structures are determined experimentally from X-ray diffraction (XRD) studies. In Sec. IV, we describe how, owing to artefacts of XRD studies, many MOF structures deposited in the Cambridge Structural Database (CSD) typically contain solvent and/or are chemically invalid, precluding their direct use for computational screenings. In Sec. V, we sketch how these structures are processed to curate an open database of *computation-ready*, experimental (CoRE) MOF crystal structures that resemble the structure used in gas adsorption studies. Following, in Sec. VI, we survey high-throughput computational screenings of MOF crystal structures for gas adsorption and separation applications that have directly motivated the experimental synthesis and characterisation of a performant MOF. In line with the theme of how the open CoRE MOF database spurred the computational identification of performant materials, in Sec. VII we review efforts by NIST to curate open databases of adsorption isotherms in MOFs. Interspersed, we draw analogies with the impacts of open data in molecular biology and machine learning.

II.   MOLECULAR MODELS AND METHODS

Accurately and efficiently predicting the adsorption properties of a given MOF structure via computation remains a formidable challenge and is still an active, progressing area of research.

A.   Energetic modelling

As a mathematical description of the potential energy of the many-body system consisting of the MOF structure and its adsorbed gas molecules, *ab initio* quantum chemical calculations are accurate but require significant computational resources. On the other hand, classical force fields, whose interatomic potential parameters are fit to experimental data or *ab initio*

---

[2]We specify the MOF to be existing, as opposed to hypothetical or conceived, according to whether synthesis protocols have already been reported in the literature. Knowledge of these protocols, as well as activation procedures and information about stability, can expedite deployment.

calculations, are more computationally efficient than *ab initio* calculations but often less accurate when transferred to systems different to which they were fit. For the former reason, classical force fields such as the Universal Force Field (UFF) [28] and DREIDING [29] are typically used to describe interactions of gas molecules with the MOF in high-throughput screenings. The interatomic potentials in DREIDING, for example, were tuned to reproduce crystal structures and sublimation energies of a large set of compounds. For gas–gas interactions, the Transferable Potentials for Phase Equilibria (TraPPE) [30] cover many adsorbate molecules and were tuned to reproduce vapour–liquid equilibria of the fluids. Gas–MOF interactions are typically obtained by mixing rules that determine interatomic potential parameters for cross-species interactions from pure-species parameters. Electrostatic interactions are usually modelled (with Coulomb's law) by assigning point charges to the atoms of the MOFs and, if appropriate, the adsorbate molecule (e.g., TraPPE assigns charges to the atoms of CO_2 [31], which has polar bonds, but not to CH_4, which lacks polar bonds [30]). There exists a hierarchy of methods for assigning point charges to MOFs to model their electrostatic interaction with an adsorbate molecule that possesses polar bonds [32] (see Sec. IX B 2), and a database of experimental MOF crystal structures with high-quality point charges assigned can be downloaded (see Sec. V A 1).

Notably, off-the-shelf force fields such as UFF often inaccurately model the interaction of electron-donating adsorbates with coordinatively unsaturated/open metal sites in MOFs [33]; as they were designed/tuned to describe van der Waals interactions, complicated electronic interactions such as π-complexes are outside their scope of applicability. An example interaction between an adsorbate and unsaturated metal site is ethylene forming a π-complex with unsaturated Ag^+ sites in MOFs [34, 35]. A typical approach to accurately model the electronic interaction between an adsorbate and unsaturated metal site is to retrofit the forcefield with a specially designed and tuned pair potential between, e.g., an unsaturated Cu in a Cu_2 paddlewheel and the centre of the carbon-carbon double bond of ethylene [36]. The pair potential is chosen and tuned using first principles calculations of the potential energy of the adsorbate near the unsaturated metal atom (within the MOF or within a representative cluster) as training data. Such an approach (with variations) has been used to retrofit off-the-shelf force fields such as UFF with a pair potential to describe C_2H_4 [36, 37], H_2 [38], C_2H_2 [39, 40], CO_2 [39, 41, 42], CH_4 [41, 43], and CO [44] interactions with un-saturated Cu within Cu_2 paddlewheels as well as for CO_2, CH_4, and H_2O interactions with unsaturated metals in the MOF-74 series [41, 45, 46]. Retrofitting an off-the-shelf force field to properly describe interactions of adsorbates with unsaturated metals in MOFs involves significant effort; Campbell et al. showed that such a retrofitted force field can transfer to different adsorbates (e.g., ethylene to propylene) and to different MOFs (e.g., CuBTC to PCN-16) [47].

The many-body potential energy description is then used as input to Monte Carlo (MC) or molecular dynamics (MD) simulations [48] to predict properties of the MOF crystal, such as adsorption isotherms and diffusion coefficients, respectively. See the recent review by Cho et al. for more on force field development [49].

B. Structural modelling

To simulate adsorption of gas in a given MOF, its crystal structure must be known. Typically in high-throughput computational screenings, the MOF structure is, as an approximation of varying severity (see Sec. IX A), taken as rigid for two reasons. First, the computational expense to sample and compute the intrahost energy of the possible configurations of the host in addition to the gas (under the osmotic ensemble [50]) is prohibitive for a large number of structures. Second, an (accurate) intrahost force field to model the potential energy of different MOF conformations may not be available for certain coordination environments encountered in MOFs. However, progress is underway in the development of (i) transferable, accurate intrahost force fields for MOFs (e.g., UFF4MOF[51], MOF-FF[52], BTW-FF[53]) and tools to automatically generate intrahost force field input files for simulation software [54], (ii) automated procedures to generate intrahost force fields from *ab initio* calculations (QuickFF [55, 56]), and (iii) efficient algorithms to simulate their flexibility [57].

Clearly, databases of crystal structures of MOFs are a prerequisite for the high-throughput computational screening of MOFs for adsorption-based applications. The accuracy of the crystal structure is imperative, as simulated gas adsorption can be sensitive to the crystal structure assumed [58]. Generally, the experimentally determined crystal structure of a MOF is regarded as the gold standard, with the following caveats. First, if the crystal structure was experimentally determined directly after synthesis, with solvent in its pores, it could change upon the evacuation of solvent (activation) [59]. Second, there can be variation in the experimentally determined lattice constants of a given MOF among different research groups (e.g., see Figure S1 in Ref. 60 for the spread of 52 experimentally reported lattice parameters of HKUST-1). In the absence of an experimentally determined crystal structure, we note that one could computationally place the appropriate linker molecules and metal nodes into the appropriate network topology to build a predicted MOF structural model [61], then refine this crude geometrical approximation by minimising the potential energy of the assembled structure over the atomic positions using a classical force field or electronic structure calculations.

C. Molecular simulation of adsorption in MOFs

For general background on molecular simulation techniques, we suggest that readers consult the seminal textbook by Frenkel and Smit [48]. More recent reviews from the group of Coudert [50, 62, 63] outline the application of molecular simulation techniques specifically to the prediction of many properties of nanoporous materials.

1. Monte Carlo—The Markov chain Monte Carlo (MC) technique enables efficient sampling of a probability distribution defined on a high-dimensional sample space when a function *proportional to* the probability density is known (e.g., without exact knowledge of the normalisation factor) [64]. As such, it is useful for sampling statistical mechanical ensembles [65] that, at thermodynamic equilibrium, govern the probability of each possible microstate of a molecular system composed of a MOF and gas molecules and computing ensemble averages of random variables of interest.

In the most widely invoked grand-canonical (μ V T) ensemble, a MOF crystal of fixed volume V is immersed in a bath of (mixed) gas at temperature T and chemical potential μ. The probability of a given microstate, denoted by ν, under the grand-canonical ensemble is:

$$p_\nu \propto e^{-\beta E_\nu + \beta \mu \cdot \mathbf{N}_\nu},$$

(1)

with $\beta = 1/(k_B T)$, k_B the Boltzmann constant, E_ν the energy of microstate ν, and N_ν the number of particles in microstate ν. The contribution of the potential energy to E_ν is computed via a molecular model/force field. To explore microstates, Monte Carlo proposals typically include particle insertions, deletions, and translations. Equipped with a force field and a MOF crystal structure, therefore, we can conduct Markov Chain Monte Carlo simulations of the (μ V T) ensemble to simulate equilibrium gas adsorption in MOFs assumed to be rigid. Properties such as the expected number of adsorbed particles of each species $\langle \mathbf{N} \rangle$ and spatial probability density of gas molecules in the MOF are computed from the simulation. The latter is of interest to determine the most favourable adsorption sites in the MOF. Lastly, the chemical potential of the bulk gas μ is related to experimentally relevant variables (e.g., pressure and composition) through an empirical or theoretical equation of state or separate simulation(s) of the bulk, unconfined gas.

The Gibbs ensemble [66] was specifically devised for molecular simulations of phase coexistence of a fluid within a single simulation, without introducing spurious interfacial effects from a finite-sized system. The Gibbs ensemble can be applied to simulate coexistence of a bulk gas with an adsorbed phase in a MOF [67]. The Gibbs ensemble imposes the Canonical (N V T) ensemble on a system comprised of two distinct subsystems: a volume containing the gas phase and a volume containing the adsorbed phase with the MOF. The two subsystems are in thermodynamic equilibrium internally and with each other, but periodic boundary conditions are applied to each subsystem separately to obtain bulk properties and eliminate interfacial effects. In a Monte Carlo simulation of the Gibbs ensemble, Monte Carlo proposals include (i) particle displacements within each subsystem that ensure internal equilibrium within each subsystem and (ii) particle transfers between the two subsystems to ensure the equality of chemical potential μ between the two subsystems. To treat a flexible MOF, we include a Monte Carlo proposal to (iii) equally and oppositely change the volume of the two subsystems [68]. The simulator sets the *total* system composition $\mathbf{N}$, temperature T, and volumes of the two subsystems; the pressure P and composition of the gas phase can be computed in the equilibrated system to link back to the experimental conditions. The Gibbs ensemble reduces to the grand-canonical ensemble in the limit of infinite bulk gas volume [67]. The main advantage of the Gibbs ensemble over the grand-canonical ensemble is that we do not need to specify μ of the bulk phase in advance since the bulk phase is explicitly simulated [67, 69]. The main disadvantage of the Gibbs ensemble over the grand-canonical ensemble is that we can only directly specify the

total system composition, not the bulk gas composition, which is the more commonly encountered situation in experiments/industrial settings.

When the unit cell of the MOF is flexible, the appropriate ensemble for Monte Carlo simulation is the osmotic ensemble [70], in which the system volume (commensurate with the MOF crystal) is allowed to fluctuate against fixed mechanical stress, but is otherwise similar to a grand-canonical ensemble. (Hence, the osmotic ensemble is sometimes called the isothermal-isobaric-semigrand ensemble.) An important consequence of the set of constraints in the osmotic ensemble is that the chemical potential of the gas is decoupled from the mechanical stress imposed on the crystal. Simulations of the osmotic ensemble come at a large computational cost because, compared to the grand-canonical ensemble, (i) the number of accessible microstates to sample increases dramatically and (ii) during Monte Carlo moves that involve changes in the MOF structure, one must additionally calculate the intrahost potential energy (so, an intrahost force field for the MOF is required). Interestingly, the osmotic ensemble allows us to predict gas adsorption when the MOF is under mechanical stress *in addition* to the pressure of the gas, such the mechanical stress experiments on flexible MOF Co(bdp) by Mason et al. [71].

Notably, the Henry coefficient of an adsorbate in a MOF, pertinent to dilute conditions, can be calculated from an ordinary Monte Carlo integration [48]. Such a calculation is accomplished by a series of single-particle "ghost" insertions [72]; thus, estimation of a Henry coefficient is less computationally demanding than full grand-canonical Monte Carlo simulations.

Readers can additionally consult Ref. 69 for more details on Monte Carlo algorithms in the context of simulations of adsorption in MOFs.

2. Molecular dynamics—The other molecular simulation technique in widespread use is molecular dynamics (MD) simulation, which simulates a system comprised of atoms and/or molecules by propagating Newton's equations of motion forward in time. As for MC simulation, a user may obtain time averages of random variables (which, according to the ergodic hypothesis, are equivalent to ensemble averages given the same constraints) and, hence, thermophysical properties of the model system. Unlike MC, however, MD also enables the measurement of transport properties, such as the diffusivity of gas inside a MOF [73], as it samples a sequence of transient microstates in a time series[3].

As is done in MC, classical MD simulations of a gas in a MOF utilise a force field and a MOF crystal structure, which provide the quantities necessary to define the system's *Hamiltonian* and, therefrom, its governing equations of motions. In the simplest applications, the combined gas–MOF system is constrained to a fixed total energy (the sum of all kinetic and potential energies), fixed volume (and system shape), and fixed number of atoms/molecules, i.e., a microcanonical ensemble. Other ensembles, such as the canonical ensemble (fixed temperature) or the isothermal-isobaric ensemble (fixed temperature and stress [e.g., thermodynamic pressure and/or external mechanical stress]), are simulated by

---

[3]That said, under application of transition state theory, kinetic MC algorithms can simulate the diffusion of gases in MOFs [74].

extending the system Hamiltonian [48] to impose the desired constraints on the system and then propagating the resultant equations of motion for that constrained ensemble.

Among the challenges encountered in using MD simulations to model gas in a MOF is the issue of how to set or determine the gas pressure when the MOF structure is rigid; since the simulation cell cannot change shape, the isothermal-isobaric ensemble is not applicable. This challenge led to the development of various "grand canonical" molecular dynamics simulation techniques, in which a portion of the simulation cell acts as a reservoir of guest molecules at a fixed chemical potential. Gas molecules are added to and removed from the reservoir in events that resemble GCMC insertion and deletion moves, which fixes the chemical potential (and, hence, the pressure) of the remainder of the simulation cell once the system has adequately equilibrated. Readers can consult work by the group of Parrinello [75–77] for details of the general approach or others [78–80] for specific applications to adsorption of gases in porous materials.

Additional challenges arise in the implementation of MD for flexible MOFs. In principle, one can directly simulate adsorption-induced deformation of a MOF using isothermalisobaric ensemble MD, provided a model for the intrahost potential energy. However, we reiterate that intrahost force fields have not been extensively developed and the use of stock force fields for deformable materials is questionable[54, 81]. An approach that has been used to address these challenges is that of hybrid MD-MC, in which the total simulation cycles between periods of isothermal-isobaric ensemble MD (fixed-N, but changing volume) and GCMC (fixed volume, but fluctuating N). The MD portions relax the MOF structure while allowing the thermodynamic pressure to change whereas the MC portions restore the chemical potential to the desired value while the MOF is rigid. Readers can consult Refs. 82–84 for more details and particular implementations of this type of approach.

D. High-throughput computational screening

Equipped with a force field, a set of MOF crystal structures, and software to conduct e.g. a grand-canonical Monte Carlo (GCMC) simulation [68, 69, 85], the *brute-force computational screening strategy* is to loop over all material candidates and simulate gas adsorption in each material:

```
for material in materials
simulate_adsorption(material)
end
```

After all simulations have finished, we sort the materials by the desired property obtained by the simulation and, voilá, shortlist the top few for experimental testing[4]. This is the obvious value of high-throughput computational screenings, and it is predicated on sufficiently

---

[4]Typically, as illustrated in our survey in Sec. VI, some human judgment on e.g., ease of synthesis is also exercised in further shortlisting materials.

accurate molecular models and simulations (i.e., sufficient sampling [86, 87]) to rank the materials by their desired adsorption property with high statistical confidence.

A less obvious value of high-throughput computational screenings, which negates their trivialisation as a for loop, is to extract relationships between the structure of the MOF and the (simulated) adsorption property to reveal insights for rational design. For example, Wilmer et al. [88] found that MOFs exhibiting the highest CO_2/N_2 selectivities for flue gas separations harbour pore sizes no larger than $\approx$ 5 Å to 6 Å. Often, experimental data are too sparse to recognise such structure-property relationships.

Finally, another value of simulating adsorption in thousands of MOFs is to set expectations of performance limitations. For example, by simulating methane adsorption in 650 000 nanoporous materials, some of us suggested that the usable capacity targets set for vehicular methane storage and delivery were likely too high because all materials fell short of the target [89]. Again, the validity of conclusions of this nature is predicated on the accuracy of the force field and assumptions built into the molecular models. For example, that work [89] held MOFs rigid during the simulation, neglecting the effect of structural flexibility on the usable capacity. Currently, MOF Co(bdp) boasts the highest methane usable capacity [71]; Co(bdp) possesses a flexible backbone that collapses and expels residual gas at the discharge pressure to obtain a large usable capacity.

A brute-force screening strategy (the for loop above) for thousands of structures may entail an infeasible computational expense, especially for complicated adsorbates (e.g., inserting chain molecules such as hexane during an MC simulation via a configurational bias algorithm [90], modelling polarisability of carbon dioxide by open-metal sites in MOFs [91], modelling water adsorption in MOFs, which requires many MC samples [92]), high pressures (where many molecules are typically present in the system), and treatment of structural flexibility [70]. Two methods have emerged to circumvent conducting simulations in *all* material candidates in a brute-force screening, thereby saving computational expense: statistical machine learning and genetic algorithms.

a.   Statistical machine learning.—A statistical machine learning [93] (regression or classification) model can be trained to predict adsorption properties using geometric/ structural [94] and/or chemical [95] descriptors representing the MOF as input. The key idea is that the dependent variable in the model (the adsorption property) is expensive to compute, whereas the independent variables (the MOF descriptors) are cheap to compute. So, a trained statistical model can be used to *cheaply* predict adsorption properties on the basis of MOF descriptors. In a machine learning-accelerated high-throughput computational screening, simulated properties in only a (diverse) *subset* of material candidates are used to train the model (i.e., to identify the parameters of the statistical model). For the remaining materials where simulations were forgone, the trained statistical model is then used to predict their adsorption properties on the basis of their (cheaply computed) descriptors. For example, chemical, geometrical (surface area, void fraction, largest cavity diameter, etc.), and/or potential energy-based descriptors were used to train statistical models to screen MOFs for CO_2 adsorption [96], Xe/Kr separations [97], hydrogen adsorption [98, 99], and methane adsorption [100], conducting simulations on only a subset of training MOFs. See

Ref. 101 for a review on Quantitative Structure Property Relationship (QSPR) modelling in materials science in general.

Fernandez et al. [102] trained a variety of machine learning models to predict simulated methane adsorption in MOFs from geometric features such as void fraction, pore size, surface area, and density; perhaps most interestingly, the decision trees learned and, owing to their interpretability, informed useful MOF design rules for methane storage. Fernandez et al. [103] later invented a more information-rich descriptor of MOF structures based on the radial distribution function, weighted by atomic properties. Principal component analysis of the new (multi-dimensional) descriptor revealed its correlation with pore size, surface area, and simulated CH_4 and CO_2 uptakes. Motivated by the latter correlation, they trained support vector machines and multilinear regression models to predict simulated CH_4, CO_2, and N_2 adsorption in hypothetical MOFs from the atomic property-weighted radial distribution function. Fernandez et al. [96] later used the atomic property weighted radial distribution descriptor to train a support vector machine to discriminate between MOFs with high and low simulated CO_2 adsorption, viewing computational screening as a classification problem. Next, Fernandez et al. [104] similarly binarised simulated low-pressure CO_2 and N_2 adsorption in hypothetical MOFs as high or low, then trained a variety of machine learning classifiers to discriminate between MOFs with high/low CO_2 and N_2 adsorption on the basis of geometric descriptors.

Simon et al. [97] trained a random forest to predict simulated Xe/Kr selectivity in nanoporous materials using cheaply computed structural and energetic descriptors as input. While a structurally diverse set of 15 000 materials was used for training (where GCMC simulations of Xe/Kr adsorption were required to label each structure for the random forest), the trained random forest was used to discard materials unlikely to exhibit a high Xe/Kr selectivity from a pool of 655 000 materials without conducting GCMC simulations of Xe/Kr adsorption. This allowed more computationally expensive GCMC simulations to be focused on the most promising materials as determined by the random forest, which predicts Xe/Kr selectivity from the material descriptors very quickly. Feature importance methods showed that the energy-based descriptor was the most predictive feature for the random forest.

Thornton et al. [99] trained a neural network to predict simulated hydrogen adsorption in MOFs using structural descriptors (void fraction, pore size, surface area) and simulated adsorption energy as input, then used it accelerate a high-throughput computational screening of hundreds of thousands of materials.

Pardakhti et al. [100] combined the commonly used structural descriptors with chemical descriptors (such as the number of H, C, O, etc. atoms in a unit cell) to describe a MOF, then trained a variety of machine learning models to predict methane adsorption in hypothetical MOFs. They found a random forest to be most predictive, using the combination of structural and chemical features as input.

Borboudakis et al. [105] gathered H_2 and CO_2 adsorption measurements (target variables) in 100 MOFs from the literature. The authors then encoded the identity of the linker, metal

centre, and functional groups comprising each MOF (independent variables) into a data matrix. They then used a black box machine learning pipeline, Just Add Data, for training and testing, which led to a modest predictive capacity (Pearson correlation of 0.68 and 0.61 between predicted and true CO_2 and H_2 adsorption, respectively).

Anderson et al. [106] constructed over 400 hypothetical MOFs, then used simulated adsorption data and a variety of structural and chemical descriptors to train an (interpretable) decision tree to predict if the CO_2/N_2 selectivity of a given parent MOF will increase (or not) when functionalised. The authors also trained several machine learning regressors (decision tree, random forest, support vector machine, neural network, gradient boosted machine) to predict the simulated CO_2/N_2 selectivity of a given MOF; the gradient boosted machine ranked the materials according to CO_2/N_2 selectivity most accurately.

Motivated by the promise of metalated catechols for hydrogen storage [107], Anderson et al. [108] constructed 105 MOF crystals with "alchemical" metalated catechols on their linkers. The sites are "alchemical" because the interaction of H_2 with the metalated catechol was allowed to artificially vary to account for variations in chemistry without the explicit specification of it. The authors trained a neural network to predict simulated hydrogen adsorption of a given MOF from structural descriptors, the density of catechol sites, and the strength of the alchemical H_2 interaction with the catechol site. A single neural network was trained to predict hydrogen adsorption at different temperatures and pressures by including them as inputs. The authors then used the neural network to predict the maximum attainable H_2 deliverable capacity at different storage/discharge conditions.

Wu et al. [109] trained a random forest and a gradient boosting regression tree to predict simulated methane adsorption in hypothetical MOFs using structural descriptors and the simulated methane Henry coefficient (cheaper to compute than adsorption from a GCMC simulation) as inputs.

Bucior et al. [98] engineered an energy-based feature vector for H_2 adsorption in a MOF by computing the potential energy of a H_2 adsorbate at each point on a grid overlaying the unit cell of the MOF, binning the energies into a histogram, and stacking the bin heights into a vector. They then trained an L1-regularised multilinear regression model to predict the H_2 usable capacity at 77 K from the descriptor, using molecular simulations of H_2 adsorption in a training set of 1 000 hypothetical MOFs. Inspection of the coefficients in the regression model revealed the energetics of the adsorption sites that tend to equip a MOF with a high/low H_2 usable capacity. Given the predictiveness of the model, the authors proceeded to screen 54 776 MOF structures for H_2 storage and delivery with the trained model, eventually pinpointing MFU-4*l* as a promising candidate (see Sec. VI A 2 c).

We anticipate machine learning to (i) more commonly play a role in high-throughput computational screenings and (ii) to be further transferred to and developed for the domain of MOFs, particular deep learning [110, 111].

b. Genetic algorithms.—The second method to avert a brute-force screening is to employ a genetic algorithm to search for and sample regions of chemical space where the

most performant MOFs lie. A genetic algorithm [112] begins with an initial *population* of MOFs, then iteratively evolves the population of MOFs towards MOFs displaying higher *fitness*, defined to be e.g., the CO_2 uptake, which we can compute with a simulation [113]. The population of MOFs present at iteration i is *generation i*. The fitness and characteristics of MOFs in generation i are used to stochastically evolve generation i, thereby determining generation $i + 1$ of MOFs, with the aim of selecting/generating MOFs that exhibit maximal fitness. First, in *elitism*, a fraction of the MOFs with the highest fitness simply proceed to the next generation (without modification) [113, 114]. Second, in *selection*, a subset of the population is stochastically selected to be *parents* that produce *children*, which inherit features from their two parents, that belong to the next generation. MOFs displaying higher fitness are more likely to be selected to be parents. MOFs with the lowest fitness are unlikely to (i) proceed to the next generation or (ii) pass off their genes (features) to the next generation. Each possible MOF must have an [ideally, unique] *genetic representation*, usually a categorical array; one entry of this array, a *gene*, might represent, e.g., the type of functional group that decorates the linker [114]. *Crossover* and *mutation* determine the child of two selected parents from the genetic representations of the two parents. In the simplest variant, uniform crossover, each gene for the child is chosen from a randomly chosen parent. Mutations are random changes in randomly selected genes to maintain a diverse population and prevent finding a local as opposed to the global maximum of fitness. The idea is to generate children for the next generation by mixing the features (genes) of the MOFs exhibiting the highest fitness, but also by exploring small genetic variations. A genetic algorithm terminates after a specified number of generations or when a satisfactory fitness is reached.

Bao et al. [115] used a genetic algorithm to mutate the chemistry of MOF linkers to arrive at MOFs with high methane usable capacity. Chung et al. [114] used a genetic algorithm to search for MOFs for CO_2 capture. Collins et al.[113] began with 141 experimentally reported MOF structures as "parents", to focus on a more synthetically viable subspace of MOFs, and used a genetic algorithm to optimise the functionalisation of their linkers for postcombustion CO_2 capture.

III. DETERMINING CRYSTAL STRUCTURES OF MOFS

Single crystal X-ray diffraction (SC-XRD) is the most utilised and powerful technique to quantitatively determine the detailed crystal structure of a MOF, although sometimes neutron diffraction is used [124]. Structural information can be extracted from the XRD pattern by careful analysis. Take a one-dimensional XRD pattern (e.g., X-ray intensity vs. the diffraction angle) as an example. The position of the peak, according to Braggs law, is related to the d-spacing of the unit cell of the material, which is defined by the lattice parameters and symmetry. The area of the peak, i.e., the integral of the intensity, is a result of the types and relative positions of atoms in the unit cell; the peak width and shape are affected by defects and the size of the the crystallite. [125, 126]

The crystal structure of a MOF can be determined from XRD data through a process called Rietveld refinement[127]. First, we start by building a crystal unit cell structure model (or use an existing one) as an initial guess. Then, we refine the unit cell structural model, i.e.,

adjust the atomic positions, unit cell parameters, and atomic occupancy (but not symmetry), to minimise the difference between the experimental and the simulated (in the structural model) SC-XRD pattern [126]. Clearly, building a crystal unit cell requires prior information (i.e, atom types, possible chemical formula, crystal symmetry) about the MOF. The crystallographic *R*-factor describes the difference between the refined crystallographic structure model and the experimental X-ray diffraction pattern:

$$R := \frac{\sum \left| I_{\exp} - I_{\sim} \right|}{\sum \left| I_{\exp} \right|},$$

(2)

where $I_{\exp}$ is the experimentally measured X-ray intensity and $I_{\sim}$ is the simulated intensity; the sum is over all diffraction angles in the data. A trustable structure is achieved when the crystallographic *R*-factor is less than 10 % [128]; a high quality structure usually has $R <$ 5 %.

Several difficulties are encountered in SC-XRD analysis that complicate or prevent definitive structural solutions. (1) As X-rays interact with electrons of matter, light elements such as carbon, oxygen, and hydrogen weakly scatter X-rays compared to heavy metals. Consequently, the refinement of such light elements is extremely difficult or sometimes impossible, resulting in e.g., commonly missing hydrogen atoms in Rietveld-refined crystal structure files. (2) As extended-network structures, MOFs can be deconstructed into their underlying topological nets by treating the points of extension on the secondary building units and linkers, respectively, as nodes and edges of a network [129]. During MOF synthesis, high temperature and pressure could introduce network disorder such as twinning [130] and pseudo-symmetry [131]. Relatedly, interpenetration of nets in MOFs [132] introduce complexity in the XRD pattern. (3) If the building blocks do not assemble into the anticipated net, Rietveld refinement will be unsuccessful, and the crystallographer must brainstorm other possible nets with which to compare with the XRD pattern; a degree of experience, intuition, and prior knowledge is required; (4) Since most MOFs are synthesised using solvent-based methods, solvent molecules are present in the pores. Strongly coordinating solvents (e.g., water, nitrobenzene) may assume a structure or pseudo structure and introduce foreign diffraction peaks that complicate XRD analysis [133]. Even if the solvent can be evacuated, the MOF may still adsorb moisture rapidly from the air; keeping the MOF in an inert (dry) atmosphere during experiments may be beneficial. Solvent masking (typically by software such as SQUEEZE[134]) is an effective way to remove the contribution by the solvent from the XRD pattern. (5) Because XRD is an ensemble measurement, the specific tilt of certain ligands in the MOF structure could be averaged out and thus undetermined by refinement [135, 136].

In addition to SC-XRD, powder X-ray diffraction (PXRD) is used to obtain MOF structural information. This is because it is sometimes difficult to synthesise large single crystals ca.

100 nm, as required for SC-XRD. The relationship between the crystal structure and XRD peak positions and intensities is the same as in SC-XRD, however, unlike SC-XRD, PXRD is the average of the diffraction of all small MOF crystals in different orientations. Thus, unlike SC-XRD, PXRD is often used to check the phase, crystallinity, and purity of the sample instead of determining the detailed atomic positions. However, it has become possible to obtain crystal structures from powder diffraction data using various refinement and simulation methods [125, 137, 138]. See Figure 2b for an example experimental PXRD pattern, of IRMOF-74-VII, whose crystal structure is in Figure 2a.

In some cases, computational methods can assist the experimental determination of MOF crystal structures. Poor crystallinity, large unit cells, high void fraction, and low symmetry sometimes impede MOF structure determination from XRD data with conventional methods such as Rietveld refinement [139, 140]. On the computer, we can explore the possible ways in which the (known) building blocks can be arranged under constraints (unit cell dimensions, maybe space group) imposed by the experimental XRD pattern. After building a set of candidate structural models, we can simulate the XRD pattern in the candidates and compare to the experimental pattern to identify the most likely structure [139, 141]. For example, Li et al. [140] computationally constructed a model of the suspected, complex structure of NU-1301, whose 173.3 Å cubic unit cell is comprised of 816 organic linkers and 816 uranium nodes; the correspondence of the simulated PXRD pattern of the crystal model with the experimental PXRD pattern was evidence that the crystal model was an accurate representation of NU-1301. Deria et al. [142] used computer-generated structural models to assist elucidation of the closed pore and open pore structures of breathing MOF NU-1105. Deng et al. [143] synthesised a series of MOFs expected to possess the same topology as Mg-MOF-74, but with longer linkers, up to 50 Å long. They verified the newly synthesised MOFs share the topology of Mg-MOF-74 by constructing structural models, simulating their PXRD patterns, and comparing to the corresponding experimental PXRD patterns (see Figure 2).

Finally, *in situ* XRD has become popular during the synthesis process [144, 145] to study the growth kinetics and mechanism [146–148] and during gas adsorption experiments to determine which adsorption sites are populated at different pressures [149, 150] and investigate gas-induced structural changes [151, 152].

IV. THE CAMBRIDGE STRUCTURAL DATABASE

The Cambridge Structural Database (CSD) [153] is a widely-used repository of crystal structures of organic, metal-organic, and organometallic compounds that are mostly determined by X-ray diffraction. Anyone can initiate the deposition of a structure in the Cambridge Crystallographic Data Center (CCDC), a nonprofit organisation. Before entering the database, these structures are processed both computationally and by expert structural chemistry editors [154]. Each crystal structure in the database is assigned a unique six-letter identifier (a "refcode", sometimes also including two digits appended at the end corresponding to different structure determinations). Anyone can access and download CSD data freely via the online search engine of the CSD, the ConQuest programme [155], and a Python API[154]. For each crystal structure entry, the CCDC website has a chemical

diagram, 3D viewer, and link to the associated publication. The data from CCDC, downloadable as a .cif file, contains basic structure parameters (e.g., space group, lattice constants, unit cell angles, fractional coordinates of atoms, etc.). More than 50 000 new structures are entered and updated each year [156], and 900 000 entries were recorded in 2018, according to the CSD website (https://www.ccdc.cam.ac.uk/). The number of MOFs estimated to be in the CSD is ca. 70 000 [10].

There are a few issues with MOF structures deposited in the CSD that preclude computational scientists from directly using them to conduct molecular simulations of gas adsorption. The summation of these issues, described in the following subsections, means that most MOF structures deposited in the CSD are not *computation-ready*.

A. Problem 0: Identifying which crystal structures within the CSD are MOFs

Because each CSD entry is not labelled according to the class of material (e.g., MOF versus covalent organic framework), goal zero is to compile a list of entries in the CSD that can be classified as MOFs via an automatic routine. The IUPAC provisional recommendation for the definition of a MOF is "… is a Coordination Polymer (or alternatively Coordination Network) with an open framework containing potential voids." [157], but this definition is not universally accepted [157, 158]. The most common method to search for MOFs in the CSD is to use a chemical bond criteria [10, 159, 160]; MOFs usually have a metal atom connected to specific atoms and/or ligands, which can be searched for in the database. For example, carboxylate-based linkers coordinated to metals can be found using the CSD Python API or ConQuest [155]. Beginning 2016, the Cambridge Crystallographic Data Centre (CCDC) maintains a subset of all structures in their database that they classify as a MOF [10] using seven different chemical bond criteria (see Figure 4), narrowing 850 000 structures in the CSD down to 69 699 MOFs [10]; 1D, 2D, and 3D network structures were included. On the other hand, the CoRE MOF dataset (see Sec. V A) contains only structures with 3-D connected frameworks [9]. The CSD subset reports 8 388 non-disordered structures with a pore limiting diameter greater than 3.7 Å.

B. Problem 1: solvent in the pores

The first problem with MOF structures in the CSD is that solvent molecules are often included in their pores. This is an artefact of XRD studies conducted after solvent-based synthesis [11]. However, before MOFs are deployed for use as adsorbents, heat and/or vacuum is applied to drive off residual solvent in the pores, a process known as *activation* [161], thereby allocating space for gas molecules to adsorb. Thus, the solvent molecules must be computationally removed from each structure, mimicking the experimental activation process, before simulating gas adsorption in it. An underlying assumption here is that removing the solvent does not change the MOF structure or cause it to collapse, which sometimes occurs [59, 162]; also, the structure of the MOF could differ depending on the solvent in its pores [163]. Almost 90 % of MOFs in the CSD contain solvent in their pores; water is most common [10].

C. Problem 2: structural disorder and missing hydrogen atoms

The second problem with many MOF structures in the CSD is an artefact of XRD; many structures are incomplete (e.g., missing hydrogen atoms) and chemically invalid (e.g., exhibit disorder). As Sec. III describes, (1) it is difficult to refine from XRD patterns the atoms, such as hydrogen, that only weakly scatter X-rays. Therefore, often, hydrogen atoms are omitted entirely from the .cif file. (2) Because XRD is an ensemble measurement, ligands that adopt multiple e.g., rotational conformations will appear disordered, resulting in a chemically invalid crystal structure. See Figure 3 for examples. The disorder must be repaired and missing atoms must be added in appropriate orientations to render a MOF structure computation-ready.

V. COMPUTATION-READY CRYSTAL STRUCTURES

We now review efforts to compile a database of computation-ready MOF structures, i.e., by addressing the problems noted above, to facilitate virtual screenings. We declare a crystal structure model (unit cell information, list of atoms and their coordinates) to be *computation-ready* if and only if the crystal comprises chemically valid building blocks and resembles the experimentally activated crystal structure used for gas adsorption measurements. For example, a MOF is not computation-ready if hydrogen atoms are missing from its linker, residual solvent (which ideally is removed during activation) remains in its pores, or linkers appear in multiple rotational conformations owing to disorder. To address the problems discussed in Sec. IV, the curation of a computation-ready database of experimental MOFs requires (a) sifting through the CSD to pick out the MOFs, (b) removing solvent molecules to mimic the experimental activation procedure, and (c) correcting artefacts of XRD that result in chemically invalid structures, by, e.g., adding missing hydrogen atoms and choosing one conformation of a disordered ligand.

Manually inspecting each structure in the CSD, removing solvent, adding missing hydrogen atoms, and repairing disorder, i.e., rendering it computation-ready, would be extremely time-consuming. Therefore, several authors have developed automatic routines using computer programmes to curate sets of computation-ready MOF structures. However, early databases prior to Chung et al. [9] remained private and thus could not serve as a platform for materials discovery for the community as a whole. We briefly review them here regardless.

In 2005, Ockwig et al. [159] identified a set of 1 127 MOFs from the CSD using structural queries that searched for crystals with metals coordinated to organic linkers to form 3D structures. The motivation of Ockwig et al. to compile these MOFs was to analyse and rationalise the distribution of net topologies among MOFs synthesised to date and shed light on how to design structures and predict the topology in which building blocks will assemble. While these MOFs were not rendered computation-ready, the authors made available as supplementary material a list of the CSD refcodes of these 3D MOFs along with their net. Haldoupis et al. [164] leveraged this early list of experimentally synthesised MOFs in the CSD to demonstrate a high-throughput screening of MOFs for kinetic-based separations (estimating permeability) of small, approximately spherical adsorbates. The authors manually repaired disorder in several of the structures listed by Ockwig et al. to render them computation-ready. In 2012, Van Heest et al. [165] extended the database of Haldoupis et al.

[164] to 3 432 MOFs and computationally screened them for kinetic-based noble gas separations. The authors mentioned that they excluded materials with a "significant degree of disorder" and removed solvent from the pores, but did not provide details or publicly release the database of MOFs.

In 2013, Goldsmith et al. [160] compiled a database of 22 700 computation-ready MOFs using an automated routine and divulged details of their procedure. First, they specified criteria to classify a structure as a MOF ("structures that contain carbon, a metal, a ligand, and a metal-ligand bond; and structures labeled as an extended structure" [160]) and scanned the CSD for MOFs. They then detected symmetry-related disorder, ionic species, and missing hydrogen atoms in these structures and excluded them. Finally, they developed an algorithm to remove residual, free-floating (as opposed to bound) solvents from the pores to mimic the experimental activation process. The authors removed solvent by first constructing a periodic graph model of a MOF, where the nodes represent atoms and edges represent bonds. The atoms that were not a member of the largest connected graph that included the metal were assumed solvent and removed. Goldsmith et al. used the Chahine rule for carbonaceous materials [166], a linear relationship between the gravimetric surface area and excess hydrogen uptake[5], to estimate the hydrogen uptake capacity in these MOFs. In addition to shortlisting top candidates for experimental testing, the authors revealed a trade-off between gravimetric and volumetric capacity and indicated that targeting MOFs with the highest surface areas is not coincident with targeting MOFs with the highest hydrogen capacity.

The curation of these early databases of computation-ready MOFs enabled high-throughput computation screenings of experimental MOFs for adsorption-based engineering applications and the generation of considerable insights [160, 164, 165, 169]. However, the fruits of these *private* computation-ready crystal structures were available only to the authors that were in possession of them. We now highlight the development of *open* computation-ready, experimental MOF databases [9, 10] and review their lucid impact on the computation-informed discovery of MOFs for adsorption-based engineering applications.

A. The (open) CoRE MOF database

In 2014, Chung et al. released a free and open, computation-ready, experimental (CoRE) MOF database [9][6]. The workflow to construct the CoRE MOF database is shown in Figure 3a. The authors first searched the CSD for potential MOF crystal structures using the CCDC Conquest programme; the search was for structures with more than one bond between metals and the elements O, N, B, P, S, and C. Additionally, the structures were required to form any kind of bond from these six elements to C, N, P, or S atoms. Then, the framework

---

[5]Consider a MOF immersed in a bath of gas at chemical potential μ and temperature T. Absolute adsorption is the number of adsorbate molecules in the MOF. Excess adsorption is the absolute adsorption minus the number of adsorbate molecules present in a volume V_p of the bulk gas phase at chemical potential μ and temperature T, where V_p is the accessible pore volume offered by the MOF. While absolute adsorption is directly obtained in molecular simulation, excess adsorption is more directly obtained in experimental gas adsorption measurements. [167, 168]
[6]The term "CoRE MOF" was coined by Prof. David Sholl, who wrote a number of different combinations of words on a napkin during the Nanoporous Materials Genome Center Meeting (2013, Berkeley, CA).

structures were analysed using the routine implemented in Zeo++ [170, 171] to identify 3-dimensional MOFs. The procedure led to 20 000 3D MOF structures.

Several automated in-house scripts were developed to further modify the framework structures to make them computation-ready crystals. This includes the removal of solvents (see Figure 3b), of which, by the definition provided by Chung et al., there are two types: free and bound. Free solvents (i.e., unbound solvents) are molecules that are not part of the framework, and bound solvents (i.e., coordinated solvents) are molecules that are part of the framework atoms as determined by the van der Waals radii of atoms plus a skin distance of 0.4 Å. Bound solvents are usually coordinated to the open metal sites in MOFs, such as the copper site in HKUST-1. Figure 3b illustrates the removal of free and bound solvents from a MOF. To remove solvents bound to the open metal sites of a given structure, the algorithm first constructed an adjacency matrix that describes the bonding network of atoms within it, viewed as an undirected graph (nodes: atoms, edges: bonds). Entry (i, j) of the adjacency matrix is one if atom i and j are bonded and zero otherwise. Second, for each oxygen atom–the typical bridging atom belonging to a solvent molecule in the pores–that is connected to a metal atom in the structure, the oxygen-metal bond was temporarily removed from the adjacency matrix by modifying the two corresponding elements to be zero. Following the modification, the adjacency matrix was passed to the SciPy connected components module to check if removal of the bond changed the number of edge-disjoint graph clusters comprising the structure. If the number of edge-disjoint clusters changed (indicating a solvent-MOF bond was eliminated), then the matrix element was left modified and the disjoint graph with the lowest molecular weight (the solvent) was deleted. If the number of clusters did not change, then the matrix component was changed back to one to reintroduce the bond. Also, the symmetry-related copies of atoms arising from disorder (see Figure 3c) were deleted by removing the lines from the .cif files that contained coordinates with asterisks or question marks. Lastly, the charge-balancing ions in the structures were kept on the basis of the chemical formula provided by the CSD and the adjacency matrix. For instance, if the chemical formula of an ion, which can be distinguished on the basis of a (+) or (−) mark, provided by the CSD, matches the chemical formula derived using the connected component algorithm applied to the adjacency matrix, the ion is kept as part of the framework. Some chemically invalid structures were manually edited using Materials Studio to add missing hydrogen atoms (that were not resolved from XRD), repair other related disorders, such as overlapping atoms, and remove solvent molecules (which are often critical in maintaining the structural integrity of the MOF) to mimic experimental activation. A set of 4 764 computation-ready structures were made available to the public via Github [172].

As evidence that the CoRE MOF database has had a significant impact on high-throughput computational MOF screening and discovery, it has collected over 220 citations on Google Scholar (as of 12 July 2019) since its publication in 2014 and has enabled several computation-inspired MOF discoveries, as outlined in our survey in Sec. VI. Its impact was predicated on making the structures freely and easily accessible via Github.

1. CoRE MOF charge assignment—Electrostatic forces are an important component of the interaction between a MOF and an adsorbate possessing polar bonds, such as carbon

dioxide [173, 174]. The electrostatic potential inside the MOF is typically modelled by assigning a point charge to each MOF atom; point charges are assigned on the molecular model for the adsorbate as well (as is the case for CO_2 [31]), thereby completing the description of the electrostatic interaction of the adsorbate with the MOF. Quick/cheap charge equilibration [Qeq [175]] methods and its variants [EEM[176], , PQeq[177], SQE[178], SCQeq[179], EQeq[180], MEPO-Qeq[181], FCQeq[182], I-Qeq[182], , EQeq +C[183], and SQE-MEPO[184]] are commonly used to assign point charges to atoms of a MOF. However, point charges derived from the electronic density obtained from a first principles calculation on the particular MOF are generally considered more reliable [32]. The Repeating Electrostatic Potential Extracted ATomic (REPEAT) method [185] assigns charges to MOF atoms by fitting to the electrostatic potential (outside the van der Waals radii of the MOF atoms) obtained from a periodic, first principles calculation. The density derived electrostatic and chemical (DDEC) method [186] assigns charges to MOF atoms by both fitting to the electrostatic potential (outside the van der Waals radii of the MOF atoms) and partitioning the electron density obtained from a periodic, first principles calculation. Both REPEAT and DDEC methods address issues caused by buried atoms and result in chemically meaningful charges. Nazarian et al. computed DDEC charges for more than 2 900 CoRE MOFs from the electronic density calculated with the PBE Density Functional Theory (DFT) functional and made these charges openly available [187]. DDEC charges were not assigned to a fraction of the CoRE MOF structures owing to failures in the electronic density calculations arising from computer memory limitations (affecting MOFs whose primitive unit cells have many atoms) and unresolved errors in VASP [187]. As the electrostatic potential inside a MOF is assumed to be independent of the adsorbate studied, with the view of the MOF as 'hosting' an electrostatic potential field in which the 'guest' adsorbate sits, these adsorbate-agnostic DDEC charges will facilitate high-throughput computational screenings of the CoRE MOFs when electrostatic interactions are involved.

2.   DFT-energy-minimised CoRE MOFs—The crystal structure of a newly synthesised MOF is often determined from XRD preceding activation, therefore with its pores still filled with solvent molecules [60]. Upon activation (evacuation of solvent) for gas adsorption, the structure of the MOF could change significantly [59]. Some MOFs may be destabilised by removing solvent, causing the framework to collapse [188]. Therefore, simply deleting solvent molecules from crystal structure files obtained via XRD studies, as in the generation of the CoRE MOF database, may not represent the structure upon activation. In addition, as Nazarian et al. [60] demonstrate for HKUST-1, MOF structures determined by XRD are subject to variation among research groups.

To address these issues, Nazarian et al. [60] used DFT to minimise the potential energy (by changing the atomic coordinates) of the structures of 879 CoRE MOF structures. Indeed, several DFT-optimised CoRE MOFs showed significant changes in the structure and simulated adsorption compared to their cognate unoptimised structures [60]. These DFT-optimised structures are likely more reliable then the vanilla CoRE MOF structures [9] because (a) they maintained their structural integrity (i.e., still have a void) after DFT-optimisation, (b) the DFT-optimisation accounted for any change in structure that would result from the solvent removal, and (c) the DFT-optimised structures serve as a standard for

a given MOF as opposed to (perhaps) arbitrarily choosing from the multiple structures of the same MOF deposited into the CSD by different research groups. Notably, the DFT calculations of only 879 of the 2 612 optimisation-attempted, lanthanide- and actinide-free CoRE MOF structures converged within the 30 000 CPU hours ceiling dedicated to each structure [60]; reasons for lack of convergence in the allotted time include a large number of atoms and a poor initial geometry.

B. CSD-maintained structures

In 2017, Moghadam et al. [10] reported a CCDC-maintained MOF subset of the CSD integrated into the CSD to allow for substructure searches using CSD tools and automatically update every quarter to account for newly deposited MOFs. They used seven chemical bonding criteria (see Figure 4) to sift through the CSD and identify which structures are MOFs, resulting in 69 666 MOF structures. Then, to mimic the experimental activation process, the authors wrote and released a Python script to remove bound and unbound solvent from the structures by searching the structure for 74 common solvent molecules in the MOF (or for lone oxygen atoms representing a partially resolved water molecule); 88 % of the MOFs were found to have solvent (water most common), of which 52 % is unbound and 48 % bound. A filtering process then flagged structures with disorder present, resulting in a non-disordered subset, within the MOF subset, containing 54 808 structures. According to computed geometric properties of the non-disordered subset, such as surface area and pore size, a large portion of the structures in the non-disordered subset (85 %) exhibit negligible internal surface areas, while the remaining 8 388 structures had notable pores, with a pore limiting diameter of at least 3.7 Å and a gravimetric surface area between 500 m^2/g and 2000 m^2/g.

The CSD-maintained MOF database by Moghadam et al. [10] offers several advantages over the CoRE MOF database: it is more comprehensive (including 1D, 2D, and 3D structures), automatically updates quarterly to account for newly deposited structures, allows for bond-type or cluster-type searches using the CSD tools, makes the Python scripts used for solvent removal open, and grants the user more granularity during solvent removal by allowing the removal of bound and unbound solvent separately. Two major shortcomings of the CSD MOF subset [10] compared to the CoRE MOF database [9], however, are (a) many structures are still invalid and not computation-ready because crystallographic disorder was not repaired and missing hydrogen atoms were not added and (b) DFT-optimised structures and DFT-assigned point charges are not available for the CSD MOF subset (unlike for the CoRE MOF database [60, 187]).

C. Shortcomings of automatically curated MOF databases

There are several criticisms/shortcomings of the computation-ready, experimental MOF databases. (1) The method used for the removal of solvent could be too aggressive, in that the structural integrity of the framework without solvent may become questionable for some structures [188]. The newly updated CoRE MOF database will contain structures with and without bound solvents, and the CSD-maintained subset [10] grants the user granularity in removing free or bound solvents via the Python scripts. (2) The MOF structure could change upon activation, relaxing to a different state when the solvent is removed [10, 50]; this is

particularly a concern for soft porous crystals [189]. (3) Redundant (duplicate) structures are present [190]. (4) As the solvent-removal is automatic and disorder-flagging are imperfect, several structures in both databases are not chemically accurate [191–193]. A recommended way to report such structures for the CoRE MOF database is to report an issue via Github. [7](5) When repairing disorder, one must choose a certain conformation of the ligand. The choice of ligand conformation could significantly influence the adsorption properties. This has been demonstrated in the case of Xe adsorption in SIFSIX-3-Ni structure, where the tilts of the pyrazine ligands are disordered in the XRD-determined structure [136]. (6) Several MOFs are missing from the CoRE MOF database, and it is not automatically updated when a new MOF is deposited in the CSD [10]; this issue is addressed with the CSD-maintained subset [10] which is updated every quarter.

Altintas et al. [191] comprehensively compared the CoRE [9] and CSD-maintained, non-disordered [10] MOF databases and found several discrepancies and shortcomings in both. The authors assembled 3 490 MOFs (characterised by CSD codes) in the intersection of the two databases and found that simulated gas adsorption (for both methane and hydrogen) differed significantly depending on from which database it was pulled for 387 MOFs. The differences emanated from the different and often erroneous methods among the two database curators to remove solvent, repair disorder, and address charge balancing ions. For example, errors in the automatic solvent removal routine sometimes erroneously removed a metal, ligand, or functional group that is part of the MOF structure. Several MOFs (total of 54) lacked a structure in either database that properly represents the experimentally reported structure. Altintas et al. [191] remark "it is not completely possible to fully automate establishment of a MOF [database] because treatment of some MOFs requires personal, hands-on manipulation and detailed chemistry knowledge". They released a spreadsheet with a list of problematic MOFs and manually corrected set of 54 MOFs with improper structures in both databases.

D. Computation-ready crystal structures for other classes of materials

Despite our focus on MOFs, open databases of computation-ready covalent organic frameworks (COFs) [194] and porous organic cage molecules [195] have emerged as well. Tong et al. [196, 197] prepared a database of 280 disorder- and solvent-free, experimentally synthesised COF structures (both 2D and 3D) ready for molecular simulations. Miklitz et al. [198] compiled a database of 41 intrinsically porous cage molecules from the CSD. Recent, exploratory work on applying an unsupervised machine learning algorithm to encode the shapes of cavities of porous cage molecules into latent vector representations [199] was enabled by the open porous cage database of Miklitz et al. [198]. We also duly mention the open International Zeolite Association (IZA) database of zeolite structures [200], which is widely used for computational studies of gas adsorption in zeolites.

---

[7]http://dx.doi.org/10.11578/1118280

VI. SURVEY OF HIGH-THROUGHPUT COMPUTATIONAL SCREENINGS WITH EXPERIMENTAL CONFIRMATION

We now survey high-throughput computational screenings of MOFs for gas storage and separation that directly motivated the synthesis and testing of a MOF in the bona fide (as opposed to *in silico*) laboratory. These computation-driven MOF discoveries demonstrate the practical impact of computational materials science. However, we do not discount the many high-throughput computational screenings lacking an experimental component, as (i) these computational predictions could be followed up in the future and (ii) insights into structure-property relationships from computational studies can (albeit perhaps indirectly) prompt the experimental discovery of new, performant MOFs.

A. Gas storage and delivery

For applications of MOFs in storing gases, we exploit the interactions [van der Waals, electrostatic, transition metal complexation (π-backbonding with olefins [201], complexes of H_2 with metalated catechol [107, 202, 203]), etc.] of a gas molecule with the surface of the MOF to achieve a greater density of adsorbed gas than in the corresponding bulk gas phase at the same temperature and pressure. In practice, deploying a MOF for gas storage entails packing a pressure vessel with a MOF adsorbent.

1.   Natural gas storage and delivery—Natural gas, composed of mostly methane, is regarded as a transition fuel from petroleum-based to renewable and clean fuels (i.e., renewably produced hydrogen) [18]. First, natural gas is abundant and cheap. Second, compared to e.g., gasoline, natural gas emits 25 % less carbon dioxide per energy harvested from its combustion [204], as well as less volatile organic compounds, carbon monoxide, particulate matter, and sulphur oxides (but more nitrogen oxides) [205]. However, the greenhouse effects of fugitive emissions (methane is a potent greenhouse gas itself) [206] and groundwater contamination by hydraulic fracturing [207] may diminish these environmental benefits if not controlled. Third, in the United States, the pipeline infrastructure for natural gas delivery is already in place. The transportation sector accounts for 28 % of energy consumption in the United States, and petroleum-based fuels comprise 93 % of transportation fuels [208]. The widespread adoption of natural gas as a transportation fuel could, therefore, reduce transportation costs and emissions.

A technical barrier to the widespread adoption of natural gas as a fuel for passenger vehicles is that, as a gas, compared to (liquid) gasoline, its volumetric energy density is low. Therefore, to obtain a reasonable driving range under the constraint of limited space for an onboard fuel tank, natural gas must be densified [19]. Two incumbent methods to densify natural gas are liquefaction at low temperature (111.7 K, 1 atm) and compression to high pressures ($\approx$ 200 bar, 298 K). These methods require bulky, heavy, expensive fuel tanks and expensive infrastructure at refuelling stations; further, boil-o losses from liquefied natural gas are an environmental concern [209]. Alternatively, MOFs have demonstrated the ability to densify natural gas for onboard vehicular storage at room temperature and significantly lower pressures (35 bar to 65 bar) than compressed natural gas [18, 19, 210]. So far, no MOF has met the most recent usable capacity target of 12.5 MJ methane/L MOF set by

ARPA-E to compete with compressed natural gas [211], using a pressure swing between 65 bar (storage pressure) and 5.8 bar (minimum engine inlet pressure needed) [89].

Note that in the high-throughput screenings below, natural gas is approximated as methane. As a caveat, a computational study by Zhang et al. found that, when considering the influence of larger hydrocarbons contained in natural gas, the ranking of MOFs for natural gas storage could differ from when approximating natural gas as pure methane [212].

a. **NOTT-107 and NU-125:** Wilmer et al. [213] developed a computational approach to generate MOF structural models from a chemical library of building blocks, then screened them to identify candidate materials for methane storage at 35 bar and 298 K. This study considered only the absolute volumetric methane loading at 35 bar as opposed to other studies, which consider a usable capacity, defined as the difference in absolute volumetric methane loadings at 35 bar and 5 bar. To generate hypothetical MOFs, Wilmer et al. curated a library of 102 building blocks. The building blocks varied substantially in their geometry, number of connection sites, and chemical composition. These building blocks could be divided into three categories: inorganic, organic, and functional groups. The algorithm constructed crystals with at most one kind of inorganic building block, two kinds of organic building blocks, and one functional group (see Figure 5). Building blocks could combine if the geometry and chemical composition at the point of connection was the same as in the synthesisable structure. Connections between building blocks were determined solely based on geometric rules; that is, the structures were not energy-minimised. "The approach is very much like snapping Tinkertoys or Lego bricks together," said Wilmer et al. [213]. The combinations of building blocks were exhaustively explored, resulting in 137 953 hypothetical MOF structures.

The generated structures were validated by comparing a subset of them to their energetically relaxed counterparts. Choosing the appropriate building blocks, structures were generated resembling HKUST-1 [16], IRMOF-1 [6], PCN-14 [214], and MIL-47 [215]. Allowing each "pseudo-MOF" to relax energetically using the UFF [28], Wilmer et al. found that every atom within the pseudo-MOF structures was typically < 0.8 Å distance from the crystallographically measured position. Further, simulated methane adsorption isotherms at 298 K in the pseudo-MOFs agreed with both simulated adsorption in the crystallographic structure as well as with experimental adsorption data.

After validation of the structure generation algorithm, each hypothetical MOF was screened for methane storage at 35 bar and 298 K. The screening of the 137 953 structures was conducted in three successive stages of increasing Monte Carlo cycles in GCMC simulations. Of these top performers, a structural analogue to PCN-14 [PCN-14 had a predicted methane storage capacity of 197 L(STP)/L] was predicted to have a record-breaking methane storage capacity of 213 L(STP)/L.[213] This structure, unbeknownst at the time of synthesis as having been the previously-reported MOF NOTT-107 [216], was synthesised and found to have an experimentally measured methane capacity ≈ 8 % lower than the prediction [213].

In addition to identifying promising candidate structures for synthesis, this library of hypothetical structures provided insights into relationships between the structure of the MOF and high-pressure methane storage. For each hypothetical MOF, geometric properties such as surface area, void fraction, and pore size distribution were computed and correlated with the simulated methane adsorption. Though maximising gravimetric surface area had been a common strategy for designing methane adsorbents, exceeding an optimal surface area (≈ 2 500 m^2/g to 3 000 m^2/g) was found to diminish the methane storage capacity. The void fractions of the best adsorbents were found to fall within a narrow range around ≈ 0.8, and the majority of these contained methyl, ethyl, or propyl functional groups, with pore sizes between 4 Å and 8 Å. These insights led to the discovery of MOF NU-125 [217], which was designed to have a void fraction of 0.8 and demonstrated promise as an adsorbent with an exceptionally high methane uptake (see Figure 6) and a usable capacity (58 bar to 5.8 bar) that is 67 % that of the typical compressed natural gas tanks used in American transportation vehicles, which are pressurised to 3600 psi (248 bar).

b.   NU-800: Gomez-Gualdron et al. [218] constructed a set of 204 zirconium-based, hypothetical MOF structures to search for optimal and stable MOFs for storage and delivery of methane. The authors constructed the MOF structural models by computationally arranging a highly stable inorganic secondary building unit, $(Zr_6O_4)(OH)_4(CO_2)_n$, with various building blocks (see Figures 7a and 7b) to form MOFs in four network topologies, **fcu**, **ftw**, **scu** and **csq**. The structures were energy-minimised by using the UFF [28] to describe the intrahost energetics.

In each hypothetical MOF, the authors conducted GCMC simulations of methane adsorption at 65 bar and 5.8 bar to compute the usable capacity. Of the 204 Zr-based hypothetical MOFs, the one based on the ditopic building unit TPT in Figure 7a was predicted to exhibit the highest methane usable capacity (197 L(STP)/L) and was coined NU-800. See Figure 7c. As a consequence, NU-800 was synthesised, and its methane (also nitrogen, carbon dioxide and hydrogen) adsorption isotherms were measured and compared to the simulation (see Figure 7d) with overall good agreement. The measured experimental methane usable capacity of NU-800 is 167 L(STP)/L (10 % lower than the simulated value), which is the best among Zr-based MOFs and better than many previously reported MOFs for methane storage, such as Ni-MOF-74 [121 L(STP)/L] [219] and PCN-14 [149 L(STP)/L] [214], yet 15 % lower than MOF-519 [203 L(STP)/L] [220], the record-holder at the time of publication. Repeated adsorption and desorption cycles indicated that NU-800 was highly stable.

2.   Hydrogen storage and delivery—Hydrogen (H_2) is an ideally clean transportation fuel since it emits only water and heat when it combines with oxygen in a fuel cell. If hydrogen is produced renewably, such as via electrolysis of water with electricity generated from wind turbines [221] as opposed to via the (currently widely-used) steam reforming of natural gas, its adoption as a fuel could significantly reduce the rate of greenhouse gas emissions associated with the transportation sector. Moreover, hydrogen is abundant (though bonded with oxygen in water or with carbon in hydrocarbons). Hydrogen possesses a larger gravimetric energy density than any fossil fuel; however, as a gas at ambient conditions,

hydrogen suffers from a very low volumetric energy density compared to (liquid) gasoline. Therefore, for a passenger vehicle to drive an acceptable distance on a single, reasonably sized tank of hydrogen fuel, the hydrogen must be densified. Incumbent densification schemes include room-temperature storage by compression up to 700 bar and cryogenic storage (liquefaction at 20.4 K at 1 bar). Both require significant energy input, heavy and bulky fuel tanks, and costly infrastructure at refilling stations. Safety is a concern particularly for hydrogen compressed to 700 bar. Another well-researched densification strategy is to react hydrogen with metals to form metal hydrides [222], but these metal hydrides often require high temperatures to release the hydrogen and are very heavy [223]. For an ultimate hydrogen storage goal, the United States Department of Energy (DOE) set 50 g/L and 6.5 weight percent storage targets for an onboard vehicular hydrogen storage system operating at temperatures ranging from −40 °C to 60 °C [224] and pressures below 100 bar [225]. To meet this target and densify hydrogen at 100 bar for onboard vehicle storage (significantly lower than compressed hydrogen storage at 350 bar to 700 bar), much current research is focused on exploiting physical adsorption in MOFs [225]. Thus far, no MOF has met the DOE storage target in the specified temperature range because the van der Waals interactions of hydrogen with a MOF are too weak [18, 223, 225].

a.   she-MOF-1: Gómez-Gualdrón and co-workers[226] constructed a set of 13 000 hypothetical MOF structures falling in 41 different topologies and screened them for cryogenic hydrogen storage. The hydrogen usable capacities in each hypothetical MOF were predicted from the difference in simulated hydrogen adsorption at 77 K and 100 bar and at 160 K and 5 bar (using a combination of a pressure- and temperature-swing for hydrogen storage and delivery). To generate the hypothetical structures, the authors took a "top-down" approach by, first, specifying the topology of the extended network based on the points of connection emanating from the building blocks, then, placing building blocks in the topological net, spatially scaled to accommodate them. The building blocks were selected from those seen in existing MOFs. The authors then synthesised, activated, and measured hydrogen adsorption isotherms in a hypothetical MOF in the rare **she** topology, **she-MOF-1**. While **she-MOF-1** showed moderate thermal stability up to 548 K, its pore volume reduced by 30% (according to nitrogen adsorption isotherms) during shipping for hydrogen adsorption measurements, indicating a lack of long-term stability after activation. To confirm the adsorption prediction, though, simulated and experimental hydrogen adsorption isotherms match very well at 160 K after scaling the experimental data by 1.3 to account for the loss in pore volume after shipment. This study demonstrates the need for the ability to predict the stability of hypothetical MOFs (see Sec. IX D).

b.   IRMOF-20: Ahmed and co-workers[227] sought to find MOFs with both high gravimetric and volumetric usable hydrogen capacity. To do so, they simulated hydrogen adsorption at 77 K in a pressure range from 1 bar to 100 bar in each MOF in the CoRE MOF database [9] and the (privatized) database from Goldsmith et al. [160]. On the basis of the simulated usable capacity of hydrogen at 77 K using a pressure swing between 100 bar and 5 bar, the authors targeted the synthesis of a MOF exhibiting gravimetric and volumetric usable capacities that surpass those measured in MOF-5 (4.5 weight percent and 31.1 g/L), considered a benchmark material for hydrogen storage [228]. Among the 90 MOFs

predicted to surpass the performance of MOF-5, they targeted IRMOF-20 [229], with a 6.1 weight percent and 35.5 g H_2/L predicted usable capacity, for synthesis and measured its hydrogen adsorption isotherm at 77 K and up to 100 bar. The simulated and experimental gravimetric and volumetric hydrogen adsorption isotherms agreed very well. Notably, the authors quantified the degree to which the Chahine rule [166], an empirical correlation that relates gravimetric excess hydrogen uptake to the surface area of a material, can predict simulated hydrogen adsorption at 77 K and 35 bar using two different molecular models for hydrogen; the correlation is reasonable in both cases. IRMOF-20 has a greater surface area than MOF-5 (measured BET areas of 4 073 m^2/g vs. 3 512 m^2/g), rationalising its greater gravimetric usable capacity of hydrogen under the Chahine rule.

c.   MFU-4l: Bucior et al. [98] screened the CSD MOF subset [10] of ca. 55 000 MOFs for hydrogen storage at 77 K and 100 bar using a combination of machine learning and molecular simulations. First, they trained an L1-regularised linear regression model to predict the simulated usable capacity of hydrogen at 77 K using a pressure swing between 100 bar and 2 bar. Engineering a feature vector to represent each MOF, they binned into a histogram the computed van der Waals potential energy of interaction between hydrogen and the MOF at a grid of points overlaid the unit cell. To serve as training data for the regression model, they simulated hydrogen adsorption in a diverse set of hypothetical MOFs [141, 213, 226]. After ensuring their trained regression model was sufficiently accurate on test data, Bucior et al. then, on the basis of computed potential energy histograms, applied the model to predict the hydrogen usable capacity of the 55 776 MOFs in the CCDC subset [10]. To refine the usable capacity predictions by the regression model, they conducted GCMC simulations in the 1 000 MOFs predicted by the regression model to have the highest usable capacity. MFU-4*l* [230] (CSD refcode: UPOZAB) was among the top 25 3D MOF candidates according to these targeted GCMC simulations; its hydrogen adsorption isotherms were measured before but only up to 20 bar [230]. Bucior et al. then experimentally synthesised MFU-4*l* and measured its high-pressure hydrogen adsorption isotherms at 77 K, 160 K, and 296 K, with which the simulated adsorption isotherms agreed very (160 K and 296 K) or reasonably (77 K) well. See Figure 8. MFU-4*l* exhibited a usable capacity of 29 g/L (77 K, between 100 bar and 5 bar), which ranks it among the top reported MOFs for hydrogen storage at these conditions [231].

d.   SNU-70, UMCM-9, and PCN-610/NU-100: Ahmed et al. [232] used the Chahine rule to down-select ca. 44 000 MOFs from a set of ca. 500 000 experimentally reported and hypothetical MOFs for focused GCMC simulations of hydrogen adsorption at 77 K and 100 bar and 5 bar. They targeted the experimental synthesis of materials with predicted volumetric and gravimetric usable hydrogen usable capacities larger than benchmark materials MOF-5 and IRMOF-20, also considering "perceived stability and synthetic accessibility": PCN-610/NU-100, an experimentally reported MOF (CSD refcodes HABQUY/GAGZEV); a hypothetical MOF [233] that was a variant of experimentally reported MOF, SNU-70 (CSD refcode GEBPEK) [234]; and MOF with CSD refcode ZELROZ [235]. The latter could not be fully activated, which motivated the authors to construct a model of and synthesise (motivated by high predicted H_2 capacity) a mixed-linker, non-interpenetrated, MOF, UMCM-9 [236], iso-reticular to ZELROZ. Experimentally

measured H_2 adsorption isotherms at 77 K up to 100 bar demonstrated each UMCM-9, SNU-70, and PCN-610/NU-100 to have higher H_2 volumetric and gravimetric usable capacities than both benchmark materials MOF-5 and IRMOF-20.

3.   Oxygen storage and delivery—Oxygen (O_2) gas is used in healthcare to treat a variety of respiratory illnesses including chronic obstructive pulmonary disease and pulmonary fibrosis. [237, 238] Pure O_2 is also necessary for industrial processes such as Linz-Donawitz-steelmaking, which uses O_2 to reduce the carbon content of molten carbon-rich pig-iron to create steel; this process comprises 60 % of all steel production [239, 240]. Another use of pure oxygen is to increase the efficiency of the regeneration of catalyst in fluid catalytic cracking, which is an essential process in petroleum refinement [241]. To store oxygen gas, it is typically densified via compression to high-pressures of around 200 bar and stored at room temperature. Liquid oxygen is also used, though this storage method requires temperatures be maintained below 90 K in insulated vessels. MOFs are a novel alternative for O_2 storage [242], enabling, compared to standard compressed oxygen storage, (a) an increased oxygen storage density at comparable pressures and/or (b) a comparable stored oxygen density but at a reduced storage pressure, thereby alleviating safety concerns and the need for heavy storage tanks.

a.   UMCM-152: Moghadam et al. [243] conducted a high-throughput computational screening of 2 392 previously synthesised MOFs from a subset of the CoRE database chosen because of the high-quality partial charges assigned to the MOF atoms [9, 187]. The authors performed GCMC simulations of oxygen adsorption at 298 K and at pressures between 1 bar and 200 bar in each MOF. The volumetric oxygen usable capacity, using a pressure swing between 140 bar and 5 bar, was then computed. High volumetric oxygen usable capacities were correlated with the largest cavity diameters above 8 Å, void fractions larger than 0.7, and geometric surface areas larger than 2 600 m^2/g. See Fig 9a. The most promising MOF for oxygen delivery was UMCM-152 (CSD refcode: ANUGIA; see Figure 9b), with a predicted usable capacity of 249 L(STP)/L. Thus, UMCM-152 was targeted for synthesis and oxygen adsorption isotherm measurement. UMCM-152 displayed the highest volumetric O_2 delivery of any material reported, 249 L(STP)/L, 22.5 % higher than the previously best reported material, NU-125.[242]. At room temperature, the density of oxygen in a UMCM-152-packed tank at 140 bar is 96 % higher than in a traditional O_2 gas tank storage at the same pressure; to achieve the same density in a UMCM-152-packed tank at 140 bar, a compressed cylinder would exhibit up to 300 bar of pressure. Notably, the simulated and experimental oxygen adsorption isotherms at 298 K agreed very well. In summary, computational screening was used to identify UMCM-152 as exhibiting a large volumetric oxygen usable capacity to enable safer (low-pressure) and more compact adsorption-based oxygen storage.

B.   Gas separations

For applications of MOFs in gas separations, we exploit differences among gas species in their (i) affinity for the surface, an energetic effect, (ii) packing into the pores, an entropic effect [244], and/or (iii) rate of transport through the material. Chemical separations account for 10 % to 15 % of the world's energy consumption [245]. Therefore, improving the

efficiency of incumbent separation processes, e.g., distillation in the petroleum industry, could reduce pollution and make goods cheaper to produce. Moreover, the highly tuneable pore shapes and surface chemistries of MOFs could enable molecular separations that were once infeasible [246]. In practice, deploying a MOF for gas separations entails (a) packing a column with a MOF adsorbent, then passing the gaseous mixture through the column or (b) embedding the MOF adsorbent within a membrane, both of which allow the MOF to selectively capture certain gas species.

1.   Xenon/krypton separations—Life-cycle analysis indicates that generating electricity by nuclear fission emits less greenhouse gases than by fossil fuels, with emissions on par with solar photovoltaics [247]. Reprocessing used nuclear fuel recovers unused uranium for further electricity generation, thereby maximally utilising our uranium reserves, and reduces the volume of nuclear waste to sequester [248]. During the aqueous reprocessing of used nuclear fuel, volatile, radioactive nuclides of xenon and krypton evolve into the off-gases in parts-per-million concentrations [249]. MOFs could potentially be used in an adsorption-based process at ambient conditions to capture the xenon and krypton from the off-gases to prevent their emission into the environment [250]. Using two adsorption processes in series, one to remove xenon, and the next to remove krypton, is one strategy, where a material with a high Xe/Kr selectivity is desired for the first process [251]. The radioactive krypton (^{85}Kr, half-life ca. 10.7 years) recovered from the second process can be sequestered, while the xenon, which has a much shorter half-life (longest-lived ^{127}Xe, half-life 36.4 days [251]), recovered from the first process could be sold in the market for use in medicine, ion propulsion, lighting, and insulation [252].

a.   SBMOF-1: Searching for a MOF harbouring a high Xe/Kr selectivity, Banerjee et al. [253] calculated the Henry coefficients of xenon and krypton at 298 K in the set of CoRE and hypothetical MOFs, relevant to the dilute conditions encountered in the off-gases of used nuclear fuel reprocessing. The MOF that exhibited the highest Xe/Kr selectivity, SBMOF-1, was a member of the CoRE MOF database and thus has already been synthesised [254], but not characterised for Xe/Kr separations. Motivated by the computational prediction, Banerjee et al. [253] synthesised SBMOF-1, measured its pure-component adsorption isotherms, and conducted column breakthrough experiments using a surrogate gas mixture that mimics the off-gas of used nuclear fuel reprocessing facilities. SBMOF-1 was found to exhibit the highest experimentally reported equilibrium Xe/Kr selectivity at dilute conditions (on the basis of experimental Henry coefficients) and to show good breakthrough performance, even in the presence of humidity. See Figure 10. We duly note that several computational screenings of MOFs for xenon/krypton separations have been carried out at different conditions prior to the release of the CoRE MOF database [97, 255, 256].

2.   Chemical warfare agent capture—Nerve agents, such as sarin and soman, are among the most lethal chemical warfare agents due to their high levels of neurotoxicity. These synthetically produced toxins are readily absorbed through dermal contact, inhalation, and ingestion [257]. The primary mechanism of nerve agent function is by disrupting nerve signals to the organs in the body, resulting in symptoms such as seizures, cardiac arrest, and

potentially death by asphyxiation [258, 259]. Acute exposure to nerve agents can lead to long-term cognitive and behavioural deficits [260]. The adsorption capabilities of MOFs can potentially be exploited to capture chemical warfare agents from the air [261], e.g., as a filter in a gas mask [262]. Using molecular simulations to rank MOFs according to their ability to capture CWAs underlines a classic role of computer simulations: reducing the need to conduct dangerous experiments.

Due to the lethality of chemical warfare agents, surrogate molecules that share key characteristics are used in research to avoid exposure. For example, a commonly used surrogate for mustard gas is diethyl sulphide (DES). See Figure 11b. Sholl et al. [259] simulated adsorption of nerve agents [soman, sarin] at dilute conditions in the CoRE MOFs and compared their heats of adsorption to that of four common surrogates [dimethyl methylphosphonate (DMMP), diethyl chlorophosphate (DCP), diisopropyl fluorophosphate (DFP), and dimethyl *p*-nitrophenyl phosphate (DMNP)]. Judging from the correlation of the simulated heats of adsorption of the authentic nerve agents in the CoRE MOFs with the heats of adsorption of surrogates, e.g., DMMP, DCP, and DFP are poor surrogates for soman adsorption in MOFs, with DMNP its best surrogate.

a. $Ni_3(BTP)_2$: Matito-Martos et al. [263] designed a high-throughput screening strategy to identify MOFs for the capture of chemical warfare agents (CWAs) sarin, soman, mustard in humid environments. The authors screened a subset of 2 932 MOFs from the CoRE database with point charges assigned [187] to account for CWA-MOF electrostatic interactions during the simulations. First, a subset of 1 275 MOFs were excluded because they exhibited pore limiting diameters too narrow (lower than 3.72 Å) to accomodate CWA molecules. Second, the authors simulated adsorption of the CWAs and their surrogates in the 1 647 remaining MOFs at dilute conditions, computing their Henry coefficients and isosteric heats of adsorption via Widom insertions. The heats of adsorption of the authentic CSW and its surrogate were reasonably correlated for mustard gas and its surrogate diethyl sulphide (DES), but less so for sarin and somin and their surrogates. MOFs displaying the highest Henry coefficients of CWAs tended to harbour a largest cavity diameter of around 5 Å. Next, to account for competitive adsorption of water from the environment, Matito-Martos et al. computed the Henry coefficients of water in each MOF, then shortlisted 156 hydrophobic structures displaying Henry coefficients and heats of adsorption lower than hydrophobic ZIF-8 [264]. In the shortlist of hydrophobic MOFs, the authors ran (more expensive) GCMC simulations of mustard gas and nerve agents sarin and soman at 13.8 Pa and 0.6 Pa, respectively, an estimate of the lethal concentrations. Of eight MOFs predicted to exhibit the largest sarin, soman, and mustard gas uptakes (the three were strongly correlated), they selected $Ni_3(BTP)_2$ (CSD refcode: UTEWOG; see Figure 11a) for experimental synthesis and column breakthrough experiments on the basis of its reported thermal and chemical stability [265]. The authors conducted a column breakthrough experiment with 150 mg $Ni_3(BTP)_2$, flowing nitrogen gas with 80 % relative humidity (water) and 1 ppm diethyl sulphide (DES, a mustard gas surrogate) through the column at room temperature and 20 mL/min and measuring the composition of DES at the exit of the column with a gas chromatograph. Figure 11c shows that $Ni_3(BTP)_2$ readily captured DES for more than 7 hours, at which point the MOF became saturated with DES, and DES broke through the

column. The concentration of DES in the eluted gas (before saturation) was only 0.05 ppm, mimicking the mustard gas concentration that would be inhaled if $Ni_3(BTP)_2$ were a filter in a gas mask within an environment of 1 ppm mustard gas and 80 % relative humidity. Further, adsorption in $Ni_3(BTP)_2$ was reversible; thermogravimetric analysis and temperature programmed desorption indicated water was desorbed at a lower temperature than for DES, proving that the framework is selective for DES over water. In summary, Matito-Martos et al. computationally pinpointed $Ni_3(BTP)_2$ as readily adsorbing mustard gas in the presence of humidity and demonstrated capture of its surrogate molecule DES through column breakthrough experiments.

C.   Carbon dioxide capture

Fossil fuels have been one of the main energy sources since the twentieth century in the United States [266] and are the dominant source of CO_2 emissions into the atmosphere worldwide [267]. The increasing CO_2 concentration in the atmosphere is leading to significant changes in the climate, and the average global temperature is projected to rise by 2.6 °C to 4.8 °C by the end of the twenty-first century if CO_2 emissions are not mitigated from their current trajectory [268]. In the United States, roughly 33 % of energy-related CO_2 emissions are directly tied to the burning of fossil fuels to generate electricity [269]. Significant effort is devoted to develop efficient technologies to capture CO_2 from the flue gas of fossil fuel-fired power plants before it is released into the atmosphere (post-combustion capture). The captured CO_2 can then be sequestered in a geological reservoir [270]. The technology currently used to capture CO_2, such as absorbing CO_2 with aqueous alkanolamine absorbents, has not proven to be energy efficient and would reduce the energy output of power plants by ca. 30 % [271]. The energy penalty is mainly due to the high cost regenerating the amine solvent used to capture the CO_2 [271]. A more recent development sees MOFs selectively capturing CO_2 from flue gas, which could reduce the energy penalty significantly [272]. The removal of CO_2 from the flue gas is called post-combustion carbon capture. An alternative is to capture CO_2 prior to fuel combustion, i.e., a pre-combustion strategy. Instead of directly burning natural gas to produce electricity, natural gas and steam are converted, in the presence of a catalyst, to H_2 and CO (steam reforming). The CO is then reacted with water (water–gas shift reaction) to produce CO_2 and H_2. CO_2 can then be separated from H_2 (rather than from the flue gases) and sequestered, and the pure H_2 can be burned to produce electricity (without CO_2 emissions). An advantage to employing MOFs for pre-combustion carbon capture is that the CO_2/H_2 mixture is at high pressure already, and a pressure swing down to atmospheric pressure can readily push CO_2 out and regenerate the MOF.

a.   NOTT-101/OEt—Chung et al. [114] used a genetic algorithm to find a performant MOF for pre-combustion CO_2 capture. To apply genetic algorithms to search for MOFs with good selectivity for CO_2 over H_2 and a high CO_2 usable capacity, Chung et al. explored the hypothetical MOF (hMOF) database by Wilmer et al. [213]. The database contains some duplicate "twin" structures, so the database was reduced from 130 000 to ≈ 55 000 structures. The genetic representation of each hMOF consisted of a set of six integers representing the maximum and actual degree of interpenetration and the species of the inorganic node, organic linkers, and the appended functional group (see Figure 12a). In a

genetic algorithm, the series of integers is called a chromosome, and each integer within the chromosome is called a gene. A genetic algorithm compares fitness between different chromosomes, where the fitness is a property of the chromosomes that we seek to optimise. The chromosomes with higher fitness are allowed to advance to the next generation and are then allowed to undergo genetic changes, such as a random change in one of their genes (mutation) or a gene swap between two chromosomes (crossover). This process repeats, allowing the chromosomes to "evolve". Eventually, the genetic algorithm creates a generation that optimises the fitness function[273]. All hMOFs were described by a chromosome and a subset of 100 diverse hMOF chromosomes were chosen for the initial generation. Three genetic algorithms were initiated, each with a different adsorption property serving as the fitness: CO_2 usable capacity, selectivity for CO_2 over H_2, and an adsorbent performance score (APS), a product of the two aforementioned properties. The genetic algorithms were run for 10 generations, at which point Chung et al. analysed the MOFs in the last generations and found that only a few organic linker genes and inorganic node genes were represented (see Figure 12b), meaning that those genes led to higher fitness throughout the evolutionary stages of the genetic algorithm. These genes were used to obtain a preliminary list of MOFs that exhibited both good selectivity for CO_2 over H_2, a high CO_2 usable capacity, and APS (see Figure 12c). Because of its high fitness, as well as previous experience in MOF synthesis, Chung et al. synthesised NOTT-101/OEt and measured both CO_2 and H_2 adsorption isotherms (see Figure 12d). Good agreement was found with simulated isotherms. NOTT-101/OEt had a usable capacity of 3.8 mmol/g and a CO_2/H_2 selectivity of 60. Comparatively, other notable MOFs studied for CO_2/H_2 separations in pre-combustion are Mg-MOF-74, with a 2.6 mmol/g usable capacity and a 365 CO_2/H_2 selectivity [274], and Cu-BTTri, with a 3.7 mmol/g usable capacity and a 20 CO_2/H_2 selectivity [275].

Selectivity and usable capacity, however, are based on the equilibrium adsorption isotherms and do not consider process objectives, such as required purity and recovery. To check if there exists a certain "threshold" selectivity to achieve the hydrogen purity requirement of 99.999 % for combustion, Chung et al. carried out a series of pressure-swing adsorption (PSA) simulations to find the lower limit of CO_2/H_2 selectivity. They found that, for precombustion CO_2/H_2 separation, the CO_2/H_2 selectivity of the material needs to be greater than 30 to meet the process objective of 99.999 % H_2 purity. On the basis of process modelling, Chung et al. concluded that, while NOTT-101/OEth meets the process requirement, another porous material, Cu-BTTri, cannot be used for precombustion carbon capture application because the CO_2/H_2 selectivity is not high enough to generate the high purity H_2 stream required for subsequent energy generation.

1.   Xylene enrichment: Mixtures of *ortho*-, *para*-, and *meta*-xylene and ethylbenzene (C_8 aromatics) are obtained from the catalytic reforming of crude oil [276, 277]. The *p*-xylene isomer is the most valuable component of the C_8 aromatic mixture. It is oxidised to yield terephthalic acid or dimethyl terephthalate, both feedstocks for the production of polyethylene terephthalate (PET) [276], which is widely used for synthetic fibres (polyester) and bottles [245]. Pure *o*-xylene is also valuable to synthesise phthalic anhydride, a precursor for the production of plasticisers [276]. However, mixtures of C_8 aromatics are

very challenging to separate because of their similar shapes, boiling points, and polarities [278]. The two incumbent industrial processes to separate C_8 aromatic mixtures are crystallisation and, more often, selective adsorption onto a solid-state material [276, 277, 279]. MOFs are promising adsorbent materials for separating C_8 aromatics more effectively than zeolites, which are currently used in a simulated moving bed process to obtain high-purity p-xylene [280].

a. ***MOF-48:*** Gee et al. [281] conducted multi-component GCMC simulations of adsorption in the CoRE MOFs immersed in a 1:3:6:3 ethylbenzene/o-X/m-X/p-X (X=xylene) mixture at 9 bar and 50 °C (conditions for liquid phase). Among the CoRE MOFs with the highest predicted selectivity for and capacity of p-xylene, Gee et al. selected MIL-140B, MOF-48, MIL-47, and MIL-125-NH$_2$ to target for experimental investigation after also considering chemical and thermal stability and the commercial availability of their linkers and metals. Liquid-phase breakthrough adsorption measurements then tested the capability of each MIL-140B, MOF-48, and MIL-125-NH$_2$ to separate p-xylene. Of these, MIL-140B exhibited the highest breakthrough p-xylene selectivities (1.8 over o-xylene, 1.6 over m-xylene, and 2.1 over ethylbenzene). The authors claim that their column breakthrough experiments indicated MIL-140B exhibits a higher p-xylene selectivity than zeolite BaX currently used in industry, but their breakthrough experiment for BaX was conducted at 180 °C compared to 50 °C for MIL-140B. Interestingly, though MOF-48 and MIL-47 differ only by a dimethyl-functionalisation, MOF-48 exhibits p-xylene selectivity while MIL-47 exhibits o-xylene selectivity, emphasising that subtle distinctions in pore features can lead to selectivity switching for these similarly-shaped C_8 aromatic isomers.

VII. NIST RESOURCES FOR ADSORPTION MEASUREMENTS

In 2014, the National Institute of Standards and Technology (NIST) officially launched a programme devoted to adsorption science, with two main aspects: a measurement laboratory named the NIST Facility for Adsorbent Characterization and Testing (FACT)[292] and an adsorption data repository. The purpose of the FACT laboratory is to support programmes related to research, development, and engineering of adsorbent materials by developing testing procedures, disseminating reference measurements, and providing impartial testing and characterisation of adsorbent materials. A notable accomplishment of the FACT is the dissemination of a reference carbon dioxide adsorption isotherm on a NIST Reference Material, NIST RM-8852 (an Ammonium ZSM-5 zeolite), which was developed via an interlaboratory study[293]. With an emphasis on development and dissemination of standard methods and measurements, measurement outputs of the FACT may prove useful as reference points for validation of laboratory measurements or as reference properties for future modelling efforts.

The data component of NIST's efforts was released in 2014 as a free, web-based database of adsorption experiments, including measured adsorption isotherms, entitled the NIST/ARPA-E Database of Novel and Emerging Adsorbent Materials[294] (NIST-ISODB)[8]. The initial iteration of the database included a list of previously-published journal articles that describe adsorption experiments (with a broad definition of "experiments," including molecular simulations, *ab initio* simulations, model-based approaches, etc.) with tagged metadata

describing experimental parameters such as the adsorbent material, adsorptive gas, measurement temperatures, and pressure range among other descriptors. The data contents of NIST-ISODB were to be from two major sources: the open scientific literature and measurements from the FACT laboratory itself. The initial contents of the database targeted materials in the MOF family, though it also included carbon materials, zeolites, and other common porous adsorbents. Adsorption isotherms present in the journal articles that compose database entries were converted from the source graphical or tabular form in the article to a format compatible with NIST-ISODB, which could then be accessed either by displaying the isotherm graphically in the NIST-ISODB web application (cf. Figure 13) or by downloading a structured data file from NIST-ISODB. Furthermore, the NIST-ISODB web application was and is capable of plotting multiple isotherms, from the same or different source articles, simultaneously, allowing for online comparison of isotherms; simultaneous plotting of isotherms for the same adsorbent/adsorptive/temperature combination enables a sort-of "virtual interlaboratory study" of the particular adsorption experiment. The NIST-ISODB has steadily grown to over 3 500 articles and more than 30 000 isotherms as of publication of the present manuscript. Data additions to the database are chosen either from the results of string-based searches of the extant literature or by direct submission of data by outside laboratories. Other additions to NIST-ISODB since its 2014 launch include improved database vocabulary for adsorptive species (via the InChIKey scheme[295]) and adsorbent materials (see following paragraph), an application programming interface (API) for accessing the database contents (in particularly, the library of adsorption isotherms), isotherm fitting tools inside the online isotherm plotting utility, an ideal adsorbed solution theory (IAST) calculator that integrates with the API to estimate multicomponent adsorption equilibrium, and a simple adsorption column simulator that also uses IAST in conjunction with the isotherm API functions to estimate column breakthrough time.

One challenge identified early in the NIST-ISODB project was that of the naming scheme(s) for adsorbent materials and MOFs in particular. In short, there is no standard method for naming MOFs and, perhaps more critically, a specific MOF may go by multiple names, easily leading to confusion for novices and experts alike. For example, the material named HKUST-1 in the disclosure of its initial synthesis[16] is now more commonly known as CuBTC (short for copper benzene-1,3,5-tricarboxylate), but is also known as MOF-199 [296] and is sold by BASF under the name Basolite™ C300[297]. Multiplicity of names, such as for CuBTC, hinders effective searches in NIST-ISODB. To solve this problem, the NIST Registry of Adsorbent Materials[298] (NIST-MATDB)[9] was released in 2017 as a companion to NIST-ISODB. The overarching purpose of NIST-MATDB is to identify adsorbent materials via unique identifiers (based on SHA-256 cryptographic hash digests) that cross-references the names that have been or will be applied to those materials. Additionally, the NIST-MATDB provides for the association of external resources to the unique identifiers, allowing, for example, an association of a CSD entry with an adsorbent material in the NIST-MATDB. As for the NIST-ISODB, the contents of NIST-MATDB are accessible via both a web application and an API. Additionally, the NIST-ISODB was

---

[9]https://adsorption.nist.gov/matdb

reconfigured to rely on NIST-MATDB for resolution of adsorbent material names. Lastly, the NIST-MATDB web application includes a feedback tool by which the user community can provide corrections, metadata improvements, and additions to the registry. The intention is that, via crowd-sourcing and followup auditing by subject matter experts, the registry will be improved using the collective knowledge of the adsorbent materials, synthetic chemistry, and crystallography communities.

The availability of both the NIST-ISODB and NIST-MATDB provide a large quantity of freely-accessible data on adsorbents and adsorption experiments that can be leveraged for computationally-driven approaches to material development and refinement, e.g., through validation of molecular simulations and benchmarking of force fields. Additionally, the APIs of both databases provide platforms for automated exploitation of the open datasets through either straightforward data mining or more opaque machine-learning approaches. For example, Park et al. used the NIST-ISODB API to investigate the reproducibility of experimentally-measured adsorption isotherms and reported the perhaps not surprising, but certainly concerning, conclusion that few experimental adsorption isotherm experiments (e.g., for carbon dioxide adsorptive, perhaps only 15 MOFs out of thousands) can be clearly identified as reproducible based on literature data in the NIST-ISODB[299]. For example, Figure 14 shows the authors' compilation of isotherms of carbon dioxide adsorption in HKUST-1 near 298 K from the NIST-ISODB, which provides a graphical indication of the variability present in reported experimental isotherms. Additionally, Figure 15 graphically summarises their results, relating reproducibility of experimental isotherms to consistency while also indicating the number of independent isotherms available and the outlier types. One can envision other relatively straightforward uses for the NIST-ISODB dataset by identifying specific materials or families of materials that could then be reevaluated or evolved via computational approaches to achieve specific performance objectives. As one example, the NIST-ISODB isotherms could be data mined to search for candidate adsorbents for chemical separations by applying a theory such as IAST to suitable isotherms in the database. (Such a use is already envisioned via example tools in the NIST-ISODB application that integrate its isotherm API functions with the pyIAST software package[300, 301].) Similarly, integration of NIST-ISODB and NIST-MATDB with chemical insight into adsorbents (e.g., via the CSD) could be leveraged to identify families of MOFs that could be the starting point for computational material evolution toward specific performance metrics via genetic algorithm-driven mutation of those MOF coupled with the computational evaluation of the o spring materials for various material properties and adsorption characteristics. Such approaches are similar in principal to the computational screening of the hMOFs set by Snurr and co-workers[213, 302, 303], though with experimental adsorption isotherm data as a starting point.

Another opportunity for computation-driven materials development based on the NISTISODB is in force field tuning and development, an ongoing need that we discuss further in Sec. IX B. The isotherm dataset of NIST-ISODB could serve as a massive training set for the development of force fields specifically for adsorptive fluids confined in MOFs and other adsorbent materials. In fact, one can argue that there is a strong analogy with the machine learning competitions mentioned earlier: the large, freely available datasets can promote the development of standards for benchmarking force fields (e.g., resultant force

fields must satisfy essential performance metrics) and then open competitions can drive progress in molecular modelling. Such competitions could be based on the provision of limited training data from the NIST-ISODB and requirements to predict isotherms for specific adsorptive/adsorbent/temperature combinations. A competition along these lines could be ongoing, with a regularly or continuously updated leaderboard ranking the submitted isotherms. Competitions of this type would be similar to the Industrial Fluid Properties Simulation Challenge (IFPSC)[304] (nine editions to date), in which challenge entrants computationally predict some thermophysical property based on limited experimental measurements on which to tune their simulations or other predictive method. For example, the 2012 and 2014 IFPSC competitions [305–308] involved prediction of adsorption isotherms for perfluorohexane adsorption on zeolite and activated carbon adsorbents, respectively, with only simple isotherms (Nitrogen and/or Argon), pore-size distribution, and other structural characteristics as training data. Lastly, force field development and competitions based on open data resources like those from NIST may also adopt workflow practices similar to "continuous integration" (CI) that is widely used in software engineering. Given some guiding parameters (e.g., training force fields for a particular material class against a specified set of adsorbates), a CI workflow could monitor a database of adsorption isotherms, retrain a force field whenever new data that fits the training specification is available, and then re-run simulations to predict material and adsorption characteristics for cases outside the training set.

The introduction of open data resources for adsorption has also revealed opportunities and challenges that stem from a lack of standardisation among researchers of adsorption and material scientists beyond that of naming adsorbent materials. (We note that one of the goals of the FACT laboratory at NIST is to develop and disseminate best practices for adsorption measurements, which addresses this point in part.) One specific issue is the difficulty encountered in comparing isotherms from different laboratories when the adsorption measurand is presented in different units. NIST-ISODB reports more than 60 unique, non-reducible types of adsorption units, including mmol/g (millimoles of adsorbate per gram of adsorbent), the volumetric units cm^3(STP)/g (cubic centimetres of adsorbate gas *at standard temperature and pressure* (STP) conditions per gram of adsorbent), fractional units such as weight-percent, units normalised by surface area, or mass or mole units per unit cell, to name only a few[294]. Each type of units has its own advantages and particular uses, but conversion to a different unit type often requires extra information that may not be present in a manuscript describing an adsorption experiment or simulation (e.g., unit cell dimensions, bulk density, etc.). A broader issue in the area of standardisation involves the description of adsorption experiments, the quantities that are actually measured, and the clear presentation of both. For example, experimental isotherms are typically presented as *excess* adsorption isotherms[167, 309], but rarely described explicitly as such. Conversely, isotherms obtained from molecular simulations are usually *absolute* adsorption isotherms [167]. Other examples of lack of description of adsorption experiments include poor identification of the adsorbent material and, in the case of multicomponent adsorption, poor or incomplete description of the adsorptive gas composition. To address some of these concerns, NIST developed an isotherm data file based on the JavaScript Object Notation (JSON) standard to contain both the isotherm data and experimental metadata, and the NIST-ISODB API serves isotherm

data in this format (cf. Ref. 310). It is similar in principle to the Crystallographic Information File (CIF) developed and promoted by the International Crystallography Union[311, 312] for representing periodic crystal structures. For the sake of experimental/ simulation reproducibility and ease of data re-use, we strongly encourage the MOF and adsorption modelling communities to firstly describe their experimental/simulation setup with sufficient detail to enable reproduction by other groups, to adopt conventions that provide clarity in data representation and interpretation, and perhaps utilise common data formats like the JSON isotherm file.

VIII. DISCUSSION

In Sec. VI, we surveyed instances of the computational identification of performant MOFs for hydrogen, methane, and oxygen storage, carbon dioxide, xenon, and nerve agent capture, and xylene enrichment. The computational identifications of many of these near-term MOFs for adsorption-based engineering applications were often predicated on a database of *open*, computation-ready, experimental MOF crystal structures[9, 10]. There are striking parallels here with molecular biology and machine learning (the "hypothesis-free science" that enabled the discovery of CRISPR [119]) and large, open datasets that spurred developments in recommendation algorithms [286] and computer vision [289]. We anticipate the open NIST databases of adsorption isotherm measurements will spur further advances in force field development, by taking advantage of the large sets of isotherms available as training data and through the development of standard benchmarking targets and common data formats that facilitate comparison between groups worldwide.

High-throughput computational screening techniques are not unique to the domain of nanoporous materials [313, 314]. High-throughput, first principles calculations have propelled the discovery of organic light-emitting diodes [315], Li battery materials [316], organic solar cell materials [317], and catalysts [318–320]. The field of drug discovery has long adopted principles of high-throughput computational screening and cheminformatics [321, 322].

In addition to open databases of crystal structures, open data on (even failed) MOF synthesis experiments could also help the MOF community. Finding the optimal synthesis conditions (solvent, temperature, reaction time) to yield a high-quality (e.g., high surface area) MOF crystal can be difficult and time-consuming [11]. Moosavi et al. [20] showed that machines can learn from failed attempts to synthesise a MOF, automatically altering the synthesis conditions towards those yielding a higher quality crystal.

While we focus on *computation-driven* MOF discovery, we do not discount the chemical intuition of experimental MOF chemists that often results in successful, rational design of a MOF selective for a particular gas. First, one can graft functional groups onto the surface or use metals that are known to attract (e.g., amine functionalisation to target CO_2 [323], open transition metal sites to target ethylene [324]) or exclude (e.g., functionalisation with hydrophobic alkyl chains to exclude water [325]) certain adsorbates. Second, one can gauge the length of the linker required to manifest in a pore size accommodating of and commensurate with the target adsorbate [12]. However, MOFs often exhibit complicated

pore geometries, which, from the perspective of a configurational integral in a partition function [65], dictate the ensemble of configurations of the adsorbate(s) within the pores and the energetics of its interactions with the pore walls; subtle differences in pore geometry can have large effects on adsorption when the pore size is commensurate with the molecule [281]. Also, interactions with e.g., coordinatively unsaturated metal sites in MOFs are difficult to predict [326]. Moreover, some MOFs harbour flexible backbones or rotating/wobbling constituents [189] and undergo structural changes upon adsorption of gas [327], sometimes involving delicate competition between entropy and energy [328]. Thus, the rational design of a MOF to exhibit a desired adsorption property is very challenging, warranting the use of molecular models and simulations to carry out high-throughput computational screenings to account for these subtle factors through molecular models.

Finally, in addition to open databases of crystal structures, we reinforce that releasing computer codes and/or input files is imperative for reproducibility and efficient progress. See the review of Coudert [329], who predicts that the phrase "data available upon request" will become obsolete.

IX. ORIENTING THE FIELD

We now opine the most important future research directions to *routinely* identify a set of MOFs with optimal adsorption properties for an engineering application via high-throughput computational screening; the list of computationally-identified MOFs in Sec. VI could be longer. Rigorously, the survey in Sec. VI only indicates a degree of statistical signal in the rankings predicted in high-throughput computational screenings; there could be many false negatives in the studies in Sec. VI. Moreover, there could be under-reporting of experimental followups to computational predictions that failed to agree.

In a computational MOF utopia, all MOFs are rigid, perfect crystals lacking defects, and the molecular models accurately and cheaply describe potential energies. Future research directions are based on leaving behind the presumption of a computational MOF utopia.

A. Treating MOF flexibility

MOFs are typically treated as rigid in high-throughput computational screenings owing to (a) the exorbitant cost to sample configurations of the MOF and compute intrahost energies and (b) the lack of an accurate force field that covers all coordination environments found in MOFs. Often, treating the MOF as a rigid "host" is an adequate approximation [330]. Some MOFs however, are known to have flexible backbones [331] or constituents (e.g., rotating ligands [332, 333]) that adopt different ensembles of configurations depending on temperature, the amount of adsorbed gas, and the mechanical stress imposed on the framework; the flexible modes of a MOF can dramatically influence adsorption [189, 327]. Even small pore size fluctuations can be important when the (average) pore size is commensurate with the size of the adsorbate [334]. Therefore, we opine that accounting for MOF flexibility in high-throughput computational screenings is a significant next step to more accurately predict gas adsorption properties. This requires the development of (i) accurate intrahost force fields for MOFs, which is underway [52, 53, 55, 56, 188], (ii) efficient algorithms to sample MOF configurations, also underway [57], and (iii) increasing

computing power. Recently, to search for MOFs that exhibit negative gas adsorption, Krause et al. [335] used DFT to compute stress-strain relationships for linker molecules that buckle under stress. After assembling structural models with these linkers, the authors used classical force fields to compute the intrahost free energy as a function of the unit cell volume. Bistability of the intrahost free energy suggested the potential for negative gas adsorption [336] and helped inspire the discovery of DUT-50, a new material (though isoreticular to DUT-49 [337]) that exhibits negative gas adsorption.

B. Developing more accurate force fields

Arguably, the most important future research direction lies in the development of accurate classical force fields (molecular models). Given an accurate many-body potential energy description of the MOF and adsorbed gas system, the equilibrium adsorption properties follow from Monte Carlo simulations of a statistical mechanical ensemble [48, 65]. Owing to the exorbitant computation required for first principles calculations of the potential energy of a single MOF-gas configuration, it is often impractical to conduct a GCMC simulation of gas adsorption in a MOF using a first principles description of the potential energy[10]. As a consequence, simulations typically employ classical interatomic potentials to model the potential energy of an atomic system, whose parameters are tuned to reproduce experimental data or *ab initio* calculations. These cheaply-computed potentials enable the simulation of gas in a large number of MOFs but usually incur a loss in accuracy compared to *ab initio* methods, particularly when transferred to a system different from which they were tuned.

1.   Van der Waals interactions—Off-the-shelf, generic force fields such as the UFF [28], DREIDING [29], AMBER [341], OPLS [342], etc. are typically employed to describe van der Waals interactions for high-throughput screening McDaniel et al. [343] systematically compared predictions of CO_2 and CH_4 adsorption in 424 MOFs among UFF and an *ab initio* force field [based on symmetry-adapted perturbation theory (SAPT) [344]]; predicted adsorption often differed significantly between the generic and *ab initio* force field, but the statistics of the ranking of the MOFs according to adsorption was good (Spearman's rank correlation coefficient squared for CH_4 and CO_2 ranges from 0.81 to 0.98, considering uptake at both 1 bar and 30 bar and 298 K). The later finding, though specific to CO_2 and CH_4, emphasises that high-throughput computational screenings using generic force fields may rank materials with sufficient statistics but not quantitatively predict adsorption in each MOF satisfactorily. To an extent reducing the importance of generic force field choice, Dokur et al. [345] showed that simulated CO_2, H_2, N_2, and CH_4 uptakes in 100 MOFs with UFF are well-correlated with those using DREIDING (binary gas mixtures at 1 bar, 298 K).

Still, predictions of adsorption using generic force fields are often not satisfactorily accurate, and significant research efforts should be devoted to the development of accurate and

---

[10]Though, notably, DFT calculations were used as an energetic description of small gas molecules in MOF-74 to compute Henry coefficients via Widom insertions [338] by biasing the samples towards low-energy regions. Similarly, Fetisov et al. [339] conducted first principle Monte Carlo simulations of CO_2, N_2, and H_2O adsorption in Mg-MOF-74. To avoid wastefully devoting DFT calculations to high-energy trial configurations, the authors employed (i) a configurational-bias Monte Carlo algorithm and (ii) a cheaper, approximate potential for pre-sampling configurations. Chen et al. [340] simulated CH_4 adsorption in CuBTC using *ab initio* calculations on a grid to characterise the CuBTC-CH_4 interaction.

transferable force fields for predicting adsorption in MOFs. A flagship failure of generic force fields is in accurately modelling the interaction of some adsorbates with coordinatively unsaturated/open metal sites in MOFs [33, 45] because they were not tuned to represent interactions in these environments involving complicated electronic interactions such as π-complexes [34, 35]. For example, Mercado et al. [46] found that simulated adsorption of CO_2 and CH_4 in M-MOF-74 (M = Mg, Fe, Co, Ni, Zn) using UFF exhibits very poor agreement with experiment owing to the presence of open metal sites. On the basis of periodic DFT calculations of the energy of the adsorbate at judiciously placed positions in the pores, they derived a force field for CO_2, CH_4, and H_2O in the M-MOF-74 family, which yielded a more accurate prediction of the adsorption isotherm. Fetisov et al. [339] conducted first principles Monte Carlo simulations of CO_2, N_2, and H_2O adsorption in Mg-MOF-74 and achieved good agreement with the experimentally reported adsorption isotherms without the manual effort of tuning a force field, but at a large computational expense. Because unsaturated metal sites sometimes provide the strongest adsorption sites in a MOF [201], a lofty goal is to develop a generic (across e.g., metal paddlewheels in MOFs) force field for interactions of adsorbates with unsaturated metals [47]. Aside from unsaturated metals, still predictions of adsorption via generic force fields are often unsatisfactory. For example, though SBMOF-1 was correctly predicted by the simulations to be a highly-ranked material for Xe/Kr separations (see Figure 10), the predicted Xe and Kr adsorption isotherms deviate significantly from the experiment (see Supplementary Figure 22 in Ref. 253).

A more radical idea is to abandon efforts to develop a generic, transferable force field and, instead, fine-tune/tailor a force field to each MOF in an automated manner. Considerable effort is currently spent on force field development for a single adsorbate in a single MOF [45]. However, automatic routines that judiciously sample positions of adsorbates in the MOF for *ab initio* calculations could then be used to correct a prior assumption about the force field, e.g., a generic force field description. Development of such automatic force field tuners would result in more accurate high-throughput computational screenings of MOFs with minimal human intervention.

Another important and related direction in force field development is to leverage machine learning models as force fields, trained on data obtained from *ab initio* calculations, e.g., neural network force fields [346–349] and Gaussian Approximation Potentials[350]. Typically in force field development, a strict functional form for the interatomic potential is chosen, at least partially motivated by physics (e.g., Lennard Jones, Buckingham potentials). If the imposed functional form does not adequately allow the important physics to be captured, the tuned force field will be inaccurate. In contrast, deep neural networks are highly adaptable and capable of representing highly non-linear potential energy surfaces [351] with minimal human intervention. Therefore, in principle, neural networks can accurately reproduce the potential energy surface governed by *ab initio* theory but at a drastically lower computational cost. The disadvantage of a neural network force field is that much more data (*ab initio* calculations) is needed to train it than for a traditional force field where a physically motivated function form is imposed [348]. The reason is that the neural network must learn the shape of the potential energy surface in addition to the quantitative details[11].

Notably, assessing the accuracy of a force field via comparison with an experimental adsorption measurement is complicated because attributing deviations between simulation and experiment to specific causal factors is extremely difficult. First, there is significant variation in the experimental adsorption isotherm measurements [299], perhaps owing to varying synthesis and activation protocols. Second, under the assumption of quality experimental data, several factors (e.g., neglect of flexibility, poor guest-host interatomic potential [functional form and parameters], neglect of polarisability of the adsorbate, etc.) could contribute to poor agreement with the simulation, and it is often difficult to definitively attribute error to any particular one of these factors. Ideally, one could compare force field predictions to several measured properties (adsorption isotherms at different temperatures, heat of adsorption, vapour–liquid equilibria of the bulk gas phase, compressibility and thermal expansion coefficient of the MOF, etc.) when assessing force field predictiveness and tuning a new force field. A sensitivity analysis of how force field parameters affect the predictions could shed light on the most important effects to describe in a particular MOF-adsorbate system. A standard and comprehensive methodology to benchmark both interatomic potentials for van der Waals interactions and electrostatic potential modelling could propel force field development. Such a methodology should be collaboratively developed by the relevant user communities and, consequently, we do not yet suggest specific criteria for tuning and defining both dispersion and electrostatics. However, we can envision a methodology in which a force field is tuned against a well-defined set of isotherms (e.g., certain adsorbates and temperatures associated with conventional material characterisation techniques and conditions[352]) and, perhaps, other important material characteristics such as pore volume, BET area, and (if appropriate) pore-size distributions. Furthermore, this type of methodology can and should leverage existing data resources (e.g., NIST-ISODB) and encourage the collection and distribution (again, via open data resources) of specific experimental measurements to facilitate force field development and consequential acceleration of materials discovery and refinement.

2. Electrostatic interactions—The electrostatic potential field within a MOF is typically described by assigning point charges to the atoms of the framework. See Ongari et al. [32] for a summary and comparative assessment of the hierarchy of methods to assign these point charges to MOF atoms. There is signficant variance among the charges predicted by the different methods, and predicted adsorption of adsorbates with polar bonds can be sensitive to these variations [173].

Within the high-throughput screening paradigm of generating a large library of hypothetical MOF structures and screening them for their adsorption properties via molecular models and simulations, developing accurate and computationally efficient methods to assign point charges to MOF structures is imperative. One future direction is to train a machine learning model to assign reliable charges on MOF atoms. Towards this, two data-based methods, the connectivity-based atom contribution (CBAC) method [353] and the molecular building

---

[11]The following thought experimental clarifies why more data is needed to fit a neural network to the PES surface than to fit a traditional force field with an interatomic potential imposed. Assume that the Lennard-Jones potential is the ground truth for an interaction between two atoms of A. Then, two independent data points, i.e., the potential energy at two distances, is enough to determine the 12–6 Lennard-Jones σ and ϵ. In contrast, the neural network would need many more data points to learn the 12–6 scaling with interatomic distance.

block-based (MBBB) method [354] assign a charge to a given MOF atom based on its local bonding environment and building block to which it belongs, respectively. In the CBAC method, an atom in a MOF with an unknown charge with a given bonding environment is assigned a charge equal to the average charge of all atoms with that bonding environment found in a training set of MOFs for which charges were assigned by first principles calculations in conjunction with the CHELP method [355]. In the MBBB method, charges are first assigned to molecular building blocks constituting the library of MOFs using first principles calculations in conjunction with the CHELP method. Then, when the building blocks are combined to construct a library of MOF models, the building blocks in the assembled MOF inherit the charges from the isolated building blocks. Both methods require post-point charge assignment adjustments to enforce charge neutrality.

An alternative to assigning point charges to MOF atoms to describe the electrostatic potential in the MOF is to directly use an electrostatic potential grid obtained from electronic structure calculations. Such an approach requires more computer memory and disk space to store the grid but (i) reduces computational cost to compute electrostatic potential during a simulation (Ewald sum vs. grid interpolation) and (ii) is more accurate than translating the electrostatic potential into a set of point charges.

A daunting challenge is to account for the change in the electrostatic potential within the MOF as it flexes or as gas molecules adsorb on it, transferring charge between the adsorbate and the MOF. One could envision a protocol where the point charges on the MOF are updated every time the microstate of the system is updated during the simulation. This however, would result in more computational resources needed for the calculation.

C.   Treating MOF defects

Molecular simulations of adsorption in MOFs apply periodic boundary conditions to mimic a perfect, defect-free, infinite crystal. In practice, MOFs can exhibit a significant degree of defects and disorder (i.e., non-crystallinity) that affect their adsorption properties [356, 357]. A flagship example is UiO-66 [358], whose inorganic Zr-based node is coordinated to 12 benzene-1,4- dicarboxylate (BDC) ligands to afford a highly stable structure. UiO-66 can possess a significant amount of linker vacancies that can be systematically tuned by varying the synthesis conditions [359]. Both experimental [359] and computational [360, 361] studies have elucidated how the linker defects significantly influence the adsorption of CO_2 and H_2O in UiO-66. Through molecular modelling, Bristow et al. [362] investigated the mechanism by which linker vacancies form in UiO-66.

Zhang et al. [363] used DFT calculations to assess the thermodynamic stability and kinetic accessibility of zinc metal node vacancies, linker vacancies, and dangling linker defects in ZIF-8. They found defects to be relatively low in energy compared to the crystalline structure but prohibitive yet surmountable kinetic barriers exist to their formation from the crystalline state. Molecular dynamic simulations showed that defective 6-member rings in ZIF-8 can enhance the hopping rate of adsorbates through them [364].

A means to predict which MOFs in a high-throughput screening are most susceptible to forming defects that significantly influence adsorption would be useful for flagging

computational predictions of adsorption predicated on the perfect crystal assumption. That said, arguably, synthetic conditions could be tuned to eliminate defects, placing this problem in the hands of experimental MOF chemists.

D. Predicting MOF synthesisability and stability

Stability is a prerequisite for deploying a MOF for most practical applications [14]. An important research goal is to predict *a priori* the thermal, chemical, and mechanical stability of a given hypothetical MOF. Most directly stated by Zunger, within the context of topological materials, but also applicable to hypothetical MOFs, "theorists who design materials must add extra filters to avoid sending their chemistry colleagues off to the lab on a pointless quest" to synthesise "fantasy materials" [365]. Therefore, it is important to develop computational methods to predict the synthetic feasibility of hypothetical MOFs, as it will help the community to focus on the synthesis of high-performing hypothetical MOFs with high fidelity for practical applications. Efforts are underway to computationally discern the feasibility of synthesising a hypothetical MOF and identify synthetic conditions that provide favourable thermodynamics and kinetics for its formation, mainly by modelling MOF self-assembly. Cantu et al.[366] used DFT calculations and *ab initio* molecular dynamics to uncover the formation pathway (and energies of the intermediates) of the secondary building block of MIL-101. They found that the highest energy barrier is the formation of the metal centre, where the organic linkers play a key role; linker addition reactions have lower energy barriers that vary depending on the arrangement of water and incumbent linkers coordinated to the metal centre. Classical simulations have also been used to study MOF self-assembly. Yoneya and co-workers[367] demonstrated how tuning the interaction parameters between the building blocks can lead to the spontaneous formation of MOFs in molecular dynamics simulations. Biswal and Kusalik[368, 369] also simulated MOF self-assembly and observed moieties that are important in MOF formation. Recent work complements these efforts by employing enhanced-sampling techniques. Colón and co-workers[370] used enhanced-sampling techniques to simulate the self-assembly of MOF-5, finding the process is entropically favourable for the ordered MOF structure. They also found free energy barriers that are associated with structural rearrangements and solvent interactions that are disrupted in an amorphous-to-crystalline transition. Similarly, Kollias and co-workers[371] calculate free energies to explore the role that solvent and counterions play in the early-stage nucleation of MIL-101. They found that the stability of the early structures and the presence of amorphous vs. crystalline moieties can be affected by the identity of the solvent and ions present in solution.

Regarding chemical stability, in an early study of MOF stability in the presence of water, Greathouse and Allendorf [372] conducted molecular dynamics simulations of water in MOF-5, treating the flexibility of MOF-5 and modelling the ZnO coordination with a nonbonded potential. They showed distortion of the MOF structure at low water concentrations and the collapse of the framework at higher water concentrations as the water molecules attack the ZnO_4 polyhedra. In subsequent work, Bellarosa and co-workers[373] used Born–Oppenheimer molecular dynamics simulations and showed the degradation of MOF-5 instead occurs through the replacement of the organic linkers by water molecules. Han et al. [374] used density functional theory calculations to elucidate the mechanism by

which acid gases degrade zeolitic imidazolate frameworks (ZIFs) both inside the pores and at the external surface of a crystallite.

Mechanical stability of MOFs under high pressures is important when densifying MOF powders for deployment in practical applications. Rogge et al. [375] provide an account of advances in computationally predicting the mechanical stability of MOFs. Moghadam et al. [81] recently trained an artificial neural network to predict mechanical properties of MOFs and found that topological features of the MOF are strongly correlated to mechanical stability.

E. Multi-scale modelling

So far, the majority of high-throughput computational screenings relied on simple performance metrics based on equilibrium adsorption properties of the material. Although these simple properties are important indicators of the performance of these materials for adsorption-based engineering applications, these metrics do not completely determine process-level objectives [376–378]. For instance, the usable capacity and selectivity of CO_2 are commonly used performance metrics for CO_2 capture applications, while the overall process objectives are the cost of capturing and recovering the CO_2 ($/g CO_2) and the productivity of the material [378], both under CO_2 purity and recovery constraints. Since the improvements in equilibrium adsorptive selectivity and usable capacity do not necessarily translate into better process performance, we envisage the integration of process-level simulations and molecular simulations, feeding innate material properties obtained from molecular simulations into process-level (larger length scale) simulations to account for heat and mass transfer kinetics, pressure drops in columns, etc. to properly rank materials [378, 379]. However, (i) process-level modelling requires many physical properties of the material to be known, and (ii) often, process-level detriments to material performance can be corrected via engineering, e.g., poor thermal conductivity can be addressed by incorporating heat exchangers into the process.

Krishna [380] discussed and contrasted the material properties that determine performance in two distinct separation processes using adsorbent materials: pressure-swing adsorption units (a bed packed with material) and membrane permeation units (a thin layer of material). The performance of a pressure-swing adsorption unit is determined primarily by both the equilibrium gas uptake and selectivity of the material, and influences of intracrystalline diffusion are typically considered undesirable. In contrast, both equilibrium adsorption properties of the material and intracrystalline diffusion rates are important determinants of membrane permeation unit performance. Therefore, the criteria on intrinsic properties for evaluating a material for a separation process critically depend on the mode of its intended use.

First et al. [377] combined molecular simulations (to predict adsorption isotherms) and pressure swing adsorption process modelling and optimisation of e.g., the column length and pressure of adsorption and desorption, under methane recovery and purity constraints, to rank zeolites for CO_2 capture from natural gas. Interestingly, the authors found "no clear correlation between the overall cost and material-centric metrics, such as adsorption selectivity" [377].

Recently, Leperi et al.[381] developed a general evaluation metric for post-combustion CO_2 capture materials used in pressure-swing adsorption units. Molecular simulations and full process simulation and optimisations were carried out to screen $\approx 2\,900$ CoRE MOF structures. Commonly used screening metrics, such as the thermodynamic selectivity at dilute conditions, adsorbent figure of merit, separation factor, etc., show poor correlations with the cost of CO_2 capture at the unit operation level. Using the generated data, the authors developed a general evaluation metric, based on the intrinsic materials properties such as those obtained from adsorption isotherms. The authors showed the metric was, compared to other screening metrics, more predictive of the cost of CO_2 capture using a pressure-swing adsorption unit packed with that material.

F NIS

We reviewed the demonstrated impacts of molecular modelling and simulation on the discovery of performant MOFs for adsorption-based engineering applications. Our outlook is that the highly reliable computational identification of MOFs for deployment in gas storage, separations, and sensing will be routinised with well-directed research efforts.

ACKNOWLEDGEMENTS

References

[1]. Furukawa H, Cordova KE, O'Keeffe M, and Yaghi OM, Science 341, 1230444 (2013). [PubMed: 23990564]

[2]. Farha OK, Eryazici I, Jeong NC, Hauser BG, Wilmer CE, Sarjeant AA, Snurr RQ, Nguyen ST, Yazaydın AO, and Hupp JT, Journal of the American Chemical Society 134, 15016 (2012). [PubMed: 22906112]

[3]. Morris RE and Wheatley PS, Angewandte Chemie International Edition 47, 4966 (2008). [PubMed: 18459091]

[4]. Li J-R, Kuppler RJ, and Zhou H-C, Chemical Society Reviews 38, 1477 (2009). [PubMed: 19384449]

[5]. Kreno LE, Leong K, Farha OK, Allendorf M, Van Duyne RP, and Hupp JT, Chemical Reviews 112, 1105 (2012). [PubMed: 22070233]

[6]. Li H, Eddaoudi M, O'Keeffe M, and Yaghi OM, Nature 402, 276 (1999).

[7]. Notman N, Chemistry World (2017), https://www.chemistryworld.com/features/mofs-find-a-use/2500508.article, Accessed 1 April 2019, URL https://www.chemistryworld.com/features/mofs-find-a-use/2500508.article.

[8]. Hendon CH, Rieth AJ, Korzy ski MD, and Dinca M, ACS Central Science 3, 554 (2017). [PubMed: 28691066]

[9]. Chung YG, Camp J, Haranczyk M, Sikora BJ, Bury W, Krungleviciute V, Yildirim T, Farha OK, Sholl DS, and Snurr RQ, Chemistry of Materials 26, 6185 (2014).

[10]. Moghadam PZ, Li A, Wiggin SB, Tao A, Maloney AG, Wood PA, Ward SC, and Fairen-Jimenez D, Chemistry of Materials 29, 2618 (2017).

[11]. Stock N and Biswas S, Chemical Reviews 112, 933 (2011). [PubMed: 22098087]

[12]. Eddaoudi M, Kim J, Rosi N, Vodak D, Wachter J, O'Keeffe M, and Yaghi OM, Science 295, 469 (2002). [PubMed: 11799235]

[13]. Cordova KE and Yaghi OM, Materials Chemistry Frontiers 1, 1304 (2017).

[14]. Howarth AJ, Liu Y, Li P, Li Z, Wang TC, Hupp JT, and Farha OK, Nature Reviews Materials 1, 15018 (2016).

[15]. Rosi NL, Kim J, Eddaoudi M, Chen B, O'Keeffe M, and Yaghi OM, Journal of the American Chemical Society 127, 1504 (2005). [PubMed: 15686384]

[16]. Chui SS-Y, Lo SM-F, Charmant JP, Orpen AG, and Williams ID, Science 283, 1148 (1999). [PubMed: 10024237]

[17]. Simon CM, Kim J, Gómez-Gualdrón DA, Chung YG, Martin RL, Mercado R,Deem MW, Gunter D, Haranczyk M, Snurr RQ, et al., Frontiers for Young Minds 3, 11 (2015).

[18]. Schoedel A, Ji Z, and Yaghi OM, Nature Energy 1, 16034 (2016).

[19]. Mason JA, Veenstra M, and Long JR, Chemical Science 5, 32 (2014).

[20]. Moosavi SM, Chidambaram A, Talirz L, Haranczyk M, Stylianou KC, and Smit B, Nature Communications 10, 539 (2019).

[21]. Greenaway RL, Santolini V, Bennison MJ, Alston BM, Pugh CJ, Little MA, Miklitz M, Eden-Rump EGB, Clowes R, Shakil A, et al., Nature Communications 9, 2849 (2018).

[22]. Steiner S, Wolf J, Glatzel S, Andreou A, Granda JM, Keenan G, Hinkley T, Aragon-Camarasa G, Kitson PJ, Angelone D, et al., Science 363, eaav2211 (2019). [PubMed: 30498165]

[23]. Häse F, Roch LM, and Aspuru-Guzik A, Trends in Chemistry 1, 282 (2019).

[24]. Jain A, Ong SP, Hautier G, Chen W, Richards WD, Dacek S, Cholia S, Gunter D,Skinner D, Ceder G, et al., APL Materials 1, 011002 (2013).

[25]. Boyd PG, Lee Y, and Smit B, Nature Reviews Materials 2, 17037 (2017).

[26]. Fraux G, Chibani S, and Coudert F-X, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 377, 20180220 (2019).

[27]. Colón YJ and Snurr RQ, Chemical Society Reviews 43, 5735 (2014). [PubMed: 24777001]

[28]. Rappé AK, Casewit CJ, Colwell KS, Goddard III WA, and Skiff WM, Journal of the American Chemical Society 114, 10024 (1992).

[29]. Mayo SL, Olafson BD, and Goddard WA, The Journal of Physical Chemistry 94, 8897 (1990).

[30]. Martin MG and Siepmann JI, The Journal of Physical Chemistry B 102, 2569 (1998).

[31]. Potoff JJ and Siepmann JI, AIChE Journal 47, 1676 (2001).

[32]. Ongari D, Boyd PG, Kadioglu O, Mace AK, Keskin S, and Smit B, Journal of Chemical Theory and Computation 15, 382 (2019). [PubMed: 30419163]

[33]. Fischer M, Gomes JR, and Jorge M, Molecular Simulation 40, 537 (2014).

[34]. Zhang Y, Li B, Krishna R, Wu Z, Ma D, Shi Z, Pham T, Forrest K, Space B, and Ma S, Chemical Communications 51, 2714 (2015). [PubMed: 25575193]

[35]. Uchida S, Kawamoto R, Tagami H, Nakagawa Y, and Mizuno N, Journal of the American Chemical Society 130, 12370 (2008). [PubMed: 18722424]

[36]. Fischer M, Gomes JR, Froba M, and Jorge M, Langmuir 28, 8537 (2012). [PubMed: 22574969]

[37]. Kulkarni AR and Sholl DS, The Journal of Physical Chemistry C 120, 23044 (2016).

[38]. Fischer M, Kuchta B, Firlej L, Hoffmann F, and Froba M, The Journal of Physical Chemistry C 114, 19116 (2010).

[39]. Fischer M, Hoffmann F, and Fröba M, ChemPhysChem 11, 2220 (2010). [PubMed: 20540140]

[40]. Zhang C, Lan Y, Guo X, Yang Q, and Zhong C, AIChE Journal 64, 1389 (2018).

[41]. Chen L, Morrison CA, and Duren T, The Journal of Physical Chemistry C 116, 18899 (2012).

[42]. Ongari D, Tiana D, Stoneburner SJ, Gagliardi L, and Smit B, The Journal of Physical Chemistry C 121, 15135 (2017).

[43]. Koh HS, Rana MK, Wong-Foy AG, and Siegel DJ, The Journal of Physical Chemistry C 119, 13451 (2015).

[44]. Campbell C, Gomes JR, Fischer M, and Jorge M, The Journal of Physical Chemistry Letters 9, 3544 (2018). [PubMed: 29886744]

[45]. Dzubak AL, Lin L-C, Kim J, Swisher JA, Poloni R, Maximoff SN, Smit B, and Gagliardi L, Nature Chemistry 4, 810 (2012).

[46]. Mercado R, Vlaisavljevich B, Lin L-C, Lee K, Lee Y, Mason JA, Xiao DJ, Gonzalez MI, Kapelewski MT, Neaton JB, et al., The Journal of Physical Chemistry C 120, 12590 (2016).

[47]. Campbell C, Ferreiro-Rangel CA, Fischer M, Gomes JR, and Jorge M, The Journal of Physical Chemistry C 121, 441 (2016).

[48]. Frenkel D and Smit B, Understanding molecular simulation: from algorithms to applications (Academic Press, San Diego, 2001).

[49]. Cho EH, Lyu Q, and Lin L-C, Molecular Simulation 45, 1 (2019).

[50]. Coudert F-X and Fuchs AH, Coordination Chemistry Reviews 307, 211 (2016).

[51]. Addicoat MA, Vankova N, Akter IF, and Heine T, Journal of Chemical Theory and Computation 10, 880 (2014). [PubMed: 26580059]

[52]. Bureekaew S, Amirjalayer S, Tafipolsky M, Spickermann C, Roy TK, and Schmid R, Physica Status Solidi (b) 250, 1128 (2013).

[53]. Bristow JK, Tiana D, and Walsh A, Journal of Chemical Theory and Computation 10, 4644 (2014). [PubMed: 25574157]

[54]. Boyd PG, Moosavi SM, Witman M, and Smit B, The Journal of Physical Chemistry Letters 8, 357 (2017). [PubMed: 28008758]

[55]. Vanduyfhuys L, Vandenbrande S, Verstraelen T, Schmid R, Waroquier M, and Van Speybroeck V, Journal of Computational Chemistry 36, 1015 (2015). [PubMed: 25740170]

[56]. Vanduyfhuys L, Vandenbrande S, Wieme J, Waroquier M, Verstraelen T, and Van Speybroeck V, Journal of Computational Chemistry 39, 999 (2018). [PubMed: 29396847]

[57]. Heinen J and Dubbeldam D, Wiley Interdisciplinary Reviews: Computational Molecular Science 8, e1363 (2018). [PubMed: 30008812]

[58]. Lawler KV, Hulvey Z, and Forster PM, Physical Chemistry Chemical Physics 17, 18904 (2015). [PubMed: 26133672]

[59]. Li C-P and Du M, Chemical Communications 47, 5958 (2011). [PubMed: 21475759]

[60]. Nazarian D, Camp JS, Chung YG, Snurr RQ, and Sholl DS, Chemistry of Materials 29, 2521 (2017).

[61]. Boyd PG and Woo TK, CrystEngComm 18, 3777 (2016).

[62]. Evans JD, Fraux G, Gaillac R, Kohen D, Trousselet F, Vanson J-M, and Coudert F-X, Chemistry of Materials 29, 199 (2017).

[63]. Fraux G and Coudert F-X, Chemical Communications 53, 7211 (2017). [PubMed: 28634602]

[64]. Tijms H, Probability: A Lively Introduction (Cambridge University Press, Cambridge, 2017).

[65]. Shell MS, Thermodynamics and Statistical Mechanics: An Integrated Approach (Cambridge University Press, Cambridge, 2015).

[66]. Panagiotopoulos AZ, Molecular Physics 61, 813 (1987).

[67]. Panagiotopoulos AZ, Molecular Physics 62, 701 (1987).

[68]. Dubbeldam D, Calero S, Ellis DE, and Snurr RQ, Molecular Simulation 42, 81 (2016).

[69]. Dubbeldam D, Torres-Knoop A, and Walton KS, Molecular Simulation 39, 1253 (2013).

[70]. Jeffroy M, Fuchs AH, and Boutin A, Chemical Communications 28, 3275 (2008).

[71]. Mason JA, Oktawiec J, Taylor MK, Hudson MR, Rodriguez J, Bachman JE, Gonzalez MI, Cervellino A, Guagliardi A, Brown CM, et al., Nature 527, 357 (2015). [PubMed: 26503057]

[72]. Sarkisov L, The Journal of Physical Chemistry C 116, 3025 (2012).

[73]. Sholl DS, Accounts of Chemical Research 39, 403 (2006). [PubMed: 16784218]

[74]. Dubbeldam D and Snurr RQ, Molecular Simulation 33, 305 (2007).

[75]. Karmakar T, Piaggi PM, Perego C, and Parrinello M, Journal of Chemical Theory and Computation 14, 2678 (2018). [PubMed: 29589924]

[76]. Ozcan A, Perego C, Salvalaglio M, Parrinello M, and Yazaydin O, Chemical Science 8, 3858 (2017). [PubMed: 28966778]

[77]. Perego C, Salvalaglio M, and Parrinello M, The Journal of Chemical Physics 142, 144113 (2015). [PubMed: 25877568]

[78]. Sarkisov L and Monson PA, Langmuir 16, 9857 (2000).

[79]. Düren T, Jakobtorweihen S, Keil FJ, and Seaton NA, Physical Chemistry Chemical Physics 5, 369 (2003).

[80]. Delle Site L, Krekeler C, Whittaker J, Agarwal A, Klein R, and Höfling F, Advanced Theory and Simulations 2, 1900014 (2019).

[81]. Moghadam PZ, Rogge SM, Li A, Chow C-M, Wieme J, Moharrami N, Aragones-Anglada M, Conduit G, Gomez-Gualdron DA, Van Speybroeck V, et al., Matter 1, 219 (2019).

[82]. Ghoufi A and Maurin G, The Journal of Physical Chemistry C 114, 6496 (2010).

[83]. Palmer JC and Debenedetti PG, The Journal of Physical Chemistry Letters 3, 2713 (2012). [PubMed: 26295896]

[84]. Rogge SMJ, Goeminne R, Demuynck R, Gutiérrez-Sevillano JJ, Vandenbrande S,Vanduyfhuys L, Waroquier M, Verstraelen T, and Van Speybroeck V, Advanced Theory and Simulations 2, 1800177 (2019).

[85]. Simon C, York AH, Sturluson A, Huynh M, Rosen AS, and Liard C, Porousmaterials.jl (2018), URL https://github.com/SimonEnsemble/PorousMaterials.jl.

[86]. Gowers RJ, Farmahini AH, Friedrich D, and Sarkisov L, Molecular Simulation 44, 309 (2018).

[87]. Grossfield A, Patrone PN, Roe DR, Schultz AJ, Siderius DW, and Zuckerman DM, Living Journal of Computational Molecular Science 1, 5067 (2019).

[88]. Wilmer CE, Farha OK, Bae Y-S, Hupp JT, and Snurr RQ, Energy & Environmental Science 5, 9849 (2012).

[89]. Simon CM, Kim J, Gómez-Gualdrón DA, Camp JS, Chung YG, Martin RL,Mercado R, Deem MW, Gunter D, Haranczyk M, et al., Energy & Environmental Science 8, 1190 (2015).

[90]. Frenkel D, Mooij GCAM, and Smit B, Journal of Physics: Condensed Matter 4, 3053 (1992).

[91]. Becker TM, Heinen J, Dubbeldam D, Lin L-C, and Vlugt TJH, The Journal of Physical Chemistry C 121, 4659 (2017).

[92]. Zhang H and Snurr RQ, The Journal of Physical Chemistry C 121, 24000 (2017).

[93]. Friedman J, Hastie T, and Tibshirani R, The elements of statistical learning, vol. 1 (Springer Series in Statistics, New York, 2001).

[94]. Willems TF, Rycroft CH, Kazi M, Meza JC, and Haranczyk M, Microporous and Mesoporous Materials 149, 134 (2012).

[95]. Collins CR, Gordon GJ, von Lilienfeld OA, and Yaron DJ, The Journal of Chemical Physics 148, 241718 (2018). [PubMed: 29960361]

[96]. Fernandez M, Boyd PG, Daff TD, Aghaji MZ, and Woo TK, The Journal of Physical Chemistry Letters 5, 3056 (2014). [PubMed: 26278259]

[97]. Simon CM, Mercado R, Schnell SK, Smit B, and Haranczyk M, Chemistry of Materials 27, 4459 (2015).

[98]. Bucior BJ, Bobbitt NS, Islamoglu T, Goswami S, Gopalan A, Yildirim T, Farha OK,Bagheri N, and Snurr RQ, Molecular Systems Design & Engineering 4, 162 (2019).

[99]. Thornton AW, Simon CM, Kim J, Kwon O, Deeg KS, Konstas K, Pas SJ, Hill MR, Winkler DA, Haranczyk M, et al., Chemistry of Materials 29, 2844 (2017). [PubMed: 28413259]

[100]. Pardakhti M, Moharreri E, Wanik D, Suib SL, and Srivastava R, ACS Combinatorial Science 19, 640 (2017). [PubMed: 28800219]

[101]. Le T, Epa VC, Burden FR, and Winkler DA, Chemical Reviews 112, 2889 (2012). [PubMed: 22251444]

[102]. Fernandez M, Woo TK, Wilmer CE, and Snurr RQ, The Journal of Physical Chemistry C 117, 7681 (2013).

[103]. Fernandez M, Trefiak NR, and Woo TK, The Journal of Physical Chemistry C 117, 14095 (2013).

[104]. Fernandez M and Barnard AS, ACS Combinatorial Science 18, 243 (2016). [PubMed: 27022760]

[105]. Borboudakis G, Stergiannakos T, Frysali M, Klontzas E, Tsamardinos I, and Froudakis GE, npj Computational Materials 3, 40 (2017).

[106]. Anderson R, Rodgers J, Argueta E, Biong A, and Gómez-Gualdrón DA, Chemistry of Materials 30, 6325 (2018).

[107]. Tsivion E, Long JR, and Head-Gordon M, Journal of the American Chemical Society136, 17827 (2014). [PubMed: 25415078]

[108]. Anderson G, Schweitzer B, Anderson R, and Gomez-Gualdron DA, The Journal of Physical Chemistry C 123, 120 (2019).

[109]. Wu X, Xiang S, Su J, and Cai W, The Journal of Physical Chemistry C 123, 8550 (2019).

[110]. LeCun Y, Bengio Y, and Hinton G, Nature 521, 436 (2015). [PubMed: 26017442]

[111]. Goh GB, Hodas NO, and Vishnu A, Journal of Computational Chemistry 38, 1291 (2017). [PubMed: 28272810]

[112]. Gen M and Lin L, Wiley Encyclopedia of Computer Science and Engineering pp. 1–15 (2007).

[113]. Collins SP, Daff TD, Piotrkowski SS, and Woo TK, Science Advances 2, e1600954 (2016). [PubMed: 28138523]

[114]. Chung YG, Gómez-Gualdrón DA, Li P, Leperi KT, Deria P, Zhang H, Vermeulen NA, Stoddart JF, You F, Hupp JT, et al., Science Advances 2, e1600909 (2016). [PubMed: 27757420]

[115]. Bao Y, Martin RL, Simon CM, Haranczyk M, Smit B, and Deem MW, The Journal of Physical Chemistry C 119, 186 (2015).

[116]. Doudna JA and Charpentier E, Science 346, 1258096 (2014). [PubMed: 25430774]

[117]. Flavahan WA, Gaskell E, and Bernstein BE, Science 357, eaal2380 (2017). [PubMed: 28729483]

[118]. Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, and Horvath P, Science 315, 1709 (2007). [PubMed: 17379808]

[119]. Lander ES, Cell 164, 18 (2016). [PubMed: 26771483]

[120]. Mojica FJ, Díez-Villaseñor C, Soria E, and Juez G, Molecular Microbiology 36, 244 (2000). [PubMed: 10760181]

[121]. Jansen R, Embden J. D. A. v., Gaastra W, and Schouls LM, Molecular Microbiology 43, 1565 (2002). [PubMed: 11952905]

[122]. Mojica FJ, Díez-Villaseñor C, García-Martínez J, and Soria E, Journal of Molecular Evolution 60, 174 (2005). [PubMed: 15791728]

[123]. Goodman L, Genome Research 9, 673 (1999). [PubMed: 10447502]

[124]. Gándara F and Bennett TD, IUCrJ 1, 563 (2014).

[125]. Harris KDM and Tremayne M, Chemistry of Materials 8, 2554 (1996).

[126]. McCusker LB, Von Dreele RB, Cox DE, Louër D, and Scardi P, Journal of Applied Crystallography 32, 36 (1999).

[127]. Rietveld HM, Journal of Applied Crystallography 2, 65 (1969).

[128]. Øien-Ødegaard S, Shearer GC, Wragg DS, and Lillerud KP, Chemical Society Reviews46, 4867 (2017). [PubMed: 28653703]

[129]. O'Keeffe M and Yaghi OM, Chemical Reviews 112, 675 (2012). [PubMed: 21916513]

[130]. Parsons S, Acta Crystallographica Section D 59, 1995 (2003).

[131]. Ma J, Tran LD, and Matzger AJ, Crystal Growth & Design 16, 4148 (2016).

[132]. Jiang H-L, Makal TA, and Zhou H-C, Coordination Chemistry Reviews 257, 2232 (2013).

[133]. Hoshino M, Khutia A, Xing H, Inokuma Y, and Fujita M, IUCrJ 3, 139 (2016).

[134]. Spek AL, Acta Crystallographica Section C: Structural Chemistry 71, 9 (2015). [PubMed: 25567569]

[135]. Winston EB, Lowell PJ, Vacek J, Chocholoušová J, Michl J, and Price JC, Physical Chemistry Chemical Physics 10, 5188 (2008). [PubMed: 18728858]

[136]. Elsaidi SK, Mohamed MH, Simon CM, Braun E, Pham T, Forrest KA, Xu W, Banerjee D, Space B, Zaworotko MJ, et al., Chemical Science 8, 2373 (2017). [PubMed: 28451342]

[137]. Harris KDM, Tremayne M, Lightfoot P, and Bruce PG, Journal of the American Chemical Society 116, 3543 (1994).

[138]. David WI, Shankland K, van de Streek J, Pidcock E, Motherwell WDS, and Cole JC, Journal of Applied Crystallography 39, 910 (2006).

[139]. Wahiduzzaman M, Wang S, Sikora BJ, Serre C, and Maurin G, Chemical Communications 54, 10812 (2018). [PubMed: 30090892]

[140]. Li P, Vermeulen NA, Malliakas CD, Gómez-Gualdrón DA, Howarth AJ, Mehdi BL, Dohnalkova A, Browning ND, Okeeffe M, and Farha OK, Science 356, 624 (2017). [PubMed: 28428318]

[141]. Colón YJ, Gómez-Gualdrón DA, and Snurr RQ, Crystal Growth & Design 17, 5801 (2017).

[142]. Deria P, Gomez-Gualdron DA, Bury W, Schaef HT, Wang TC, Thallapally PK, Sarjeant AA, Snurr RQ, Hupp JT, and Farha OK, Journal of the American Chemical Society 137, 13183 (2015). [PubMed: 26387968]

[143]. Deng H, Grunder S, Cordova KE, Valente C, Furukawa H, Hmadeh M, Gándara F, Whalley AC, Liu Z, Asahina S, et al., Science 336, 1018 (2012). [PubMed: 22628651]

[144]. Taddei M, Casati N, Steitz DA, Dümbgen KC, van Bokhoven JA, and Ranocchiari M, CrystEngComm 19, 3206 (2017).

[145]. Park JH, Min Choi K, Joon Jeon H, Jung Choi Y, and Ku Kang J, Scientific Reports5, 12045 (2015). [PubMed: 26155988]

[146]. Wu Y, Henke S, Kieslich G, Schwedler I, Yang M, Fraser DAX, and O'Hare D, Angewandte Chemie International Edition 55, 14081 (2016). [PubMed: 27726277]

[147]. Dill ED, Josey AA, Folmer JCW, Hou F, and Martin JD, Chemistry of Materials 25, 3932 (2013).

[148]. Julien PA, Užarević K, Katsenis AD, Kimber SAJ, Wang T, Farha OK, Zhang Y, Casaban J, Germann LS, Etter M, et al., Journal of the American Chemical Society 138, 2929 (2016). [PubMed: 26894258]

[149]. Peterson VK, Southon PD, Halder GJ, Price DJ, Bevitt JJ, and Kepert CJ, Chemistry of Materials 26, 4712 (2014).

[150]. Gonzalez MI, Mason JA, Bloch ED, Teat SJ, Gagnon KJ, Morrison GY, Queen WL, and Long JR, Chemical Science 8, 4387 (2017). [PubMed: 28966783]

[151]. Walton RI, Munn AS, Guillou N, and Millange F, Chemistry - A European Journal 17, 7069 (2011).

[152]. Scherb C, Koehn R, and Bein T, Journal of Materials Chemistry 20, 3046 (2010).

[153]. Allen FH, Acta Crystallographica Section B: Structural Science 58, 380 (2002). [PubMed: 12037359]

[154]. Groom CR, Bruno IJ, Lightfoot MP, and Ward SC, Acta Crystallographica Section B: Structural Science, Crystal Engineering and Materials 72, 171 (2016).

[155]. Bruno IJ, Cole JC, Edgington PR, Kessler M, Macrae CF, McCabe P, Pearson J, and Taylor R, Acta Crystallographica Section B: Structural Science 58, 389 (2002). [PubMed: 12037360]

[156]. Bruno IJ and Groom CR, Journal of Computer-Aided Molecular Design 28, 1015 (2014). [PubMed: 25091065]

[157]. Batten SR, Champness NR, Chen X-M, Garcia-Martinez J, Kitagawa S, hrstrm L, O'Keeffe M, Suh MP, and Reedijk J, CrystEngComm 14, 3001 (2012).

[158]. Seth S and Matzger AJ, Crystal Growth & Design 17, 4043 (2017).

[159]. Ockwig NW, Delgado-Friedrichs O, O'Keeffe M, and Yaghi OM, Accounts of Chemical Research 38, 176 (2005). [PubMed: 15766236]

[160]. Goldsmith J, Wong-Foy AG, Cafarella MJ, and Siegel DJ, Chemistry of Materials 25, 3373 (2013).

[161]. Farha OK and Hupp JT, Accounts of Chemical Research 43, 1166 (2010). [PubMed: 20608672]

[162]. Howarth AJ, Peters AW, Vermeulen NA, Wang TC, Hupp JT, and Farha OK, Chemistry of Materials 29, 26 (2017).

[163]. Murdock CR, McNutt NW, Keffer DJ, and Jenkins DM, Journal of the American Chemical Society 136, 671 (2014). [PubMed: 24351165]

[164]. Haldoupis E, Nair S, and Sholl DS, Journal of the American Chemical Society 132, 7528 (2010). [PubMed: 20450176]

[165]. Van Heest T, Teich-McGoldrick SL, Greathouse JA, Allendorf MD, and Sholl DS, The Journal of Physical Chemistry C 116, 13183 (2012).

[166]. Panella B, Hirscher M, and Roth S, Carbon 43, 2209 (2005), ISSN 0008–6223.

[167]. Brandani S, Mangano E, and Sarkisov L, Adsorption 22, 261 (2016).

[168]. Myers AL and Monson PA, Adsorption 20, 591 (2014).

[169]. Watanabe T and Sholl DS, Langmuir 28, 14114 (2012). [PubMed: 22783907]

[170]. Pinheiro M, Martin RL, Rycroft CH, and Haranczyk M, CrystEngComm 15, 7531 (2013).

[171]. Pinheiro M, Martin RL, Rycroft CH, Jones A, Iglesia E, and Haranczyk M, Journal of Molecular Graphics and Modelling 44, 208 (2013), ISSN 1093–3263. [PubMed: 23876827]

[172]. Computation-ready, experimental (core) mofs download link, http://gregchung.github.io/CoRE-MOFs/.

[173]. Watanabe T, Manz TA, and Sholl DS, The Journal of Physical Chemistry C 115, 4824 (2011).

[174]. Walton KS, Millward AR, Dubbeldam D, Frost H, Low JJ, Yaghi OM, and Snurr RQ, Journal of the American Chemical Society 130, 406 (2008). [PubMed: 18154291]

[175]. Rappé AK and Goddard III WA, The Journal of Physical Chemistry 95, 3358 (1991).

[176]. Mortier WJ, Ghosh SK, and Shankar S, Journal of the American Chemical Society 108, 4315 (1986).

[177]. Ramachandran S, Lenz T, Skiff W, and Rappe A, The Journal of Physical Chemistry 100, 5898 (1996).

[178]. Nistor RA, Polihronov JG, Müser MH, and Mosey NJ, The Journal of Chemical Physics 125, 094108 (2006). [PubMed: 16965073]

[179]. Oda A and Takahashi O, Chemical Physics Letters 495, 155 (2010).

[180]. Wilmer CE, Kim KC, and Snurr RQ, The Journal of Physical Chemistry Letters 3, 2506 (2012). [PubMed: 26292141]

[181]. Kadantsev ES, Boyd PG, Daff TD, and Woo TK, The Journal of Physical Chemistry Letters 4, 3056 (2013).

[182]. Wells BA, De Bruin-Dickason C, and Chaffee AL, The Journal of Physical Chemistry C 119, 456 (2015).

[183]. Martin-Noble GC, Reilley D, Rivas LM, Smith MD, and Schrier J, Journal of Chemical Theory and Computation 11, 3364 (2015). [PubMed: 26575770]

[184]. Collins SP and Woo TK, The Journal of Physical Chemistry C 121, 903 (2017).

[185]. Campañá C, Mussard B, and Woo TK, Journal of Chemical Theory and Computation 5, 2866 (2009). [PubMed: 26631798]

[186]. Manz TA and Sholl DS, Journal of Chemical Theory and Computation 6, 2455 (2010). [PubMed: 26613499]

[187]. Nazarian D, Camp JS, and Sholl DS, Chemistry of Materials 28, 785 (2016).

[188]. Coupry DE, Addicoat MA, and Heine T, Journal of Chemical Theory and Computation 12, 5215 (2016). [PubMed: 27580382]

[189]. Horike S, Shimomura S, and Kitagawa S, Nature Chemistry 1, 695 (2009).

[190]. Barthel S, Alexandrov EV, Proserpio DM, and Smit B, Crystal Growth & Design 18, 1738 (2018). [PubMed: 29541002]

[191]. Altintas C, Avci G, Daglar H, Nemati Vesali Azar A, Erucar , Velioglu S, and Keskin S, Journal of Materials Chemistry A 7, 9593 (2019).

[192]. Zarabadi-Poor P and Marek R, ACS Applied Materials & Interfaces 11, 16261 (2019). [PubMed: 30920196]

[193]. Altintas C, Velioglu S, and Keskin S, ACS Applied Materials & Interfaces 11, 16266 (2019). [PubMed: 30920197]

[194]. Diercks CS and Yaghi OM, Science 355, eaal1585 (2017). [PubMed: 28254887]

[195]. Hasell T and Cooper AI, Nature Reviews Materials 1, 16053 (2016).

[196]. Tong M, Lan Y, Yang Q, and Zhong C, Chemical Engineering Science 168, 456 (2017).

[197]. Tong M, Lan Y, Qin Z, and Zhong C, The Journal of Physical Chemistry C 122, 13009 (2018).

[198]. Miklitz M, Jiang S, Clowes R, Briggs ME, Cooper AI, and Jelfs KE, The Journal of Physical Chemistry C 121, 15211 (2017).

[199]. Sturluson A, Huynh MT, York AHP, and Simon CM, ACS Central Science 4, 1663 (2018). [PubMed: 30648150]

[200]. Baerlocher C and McCusker LB, Database of zeolite structures, http://www.iza-structure.org/databases/, Accessed 1 April 2019.

[201]. Barnett BR, Gonzalez MI, and Long JR, Trends in Chemistry 1, 159 (2019).

[202]. Getman RB, Miller JH, Wang K, and Snurr RQ, The Journal of Physical Chemistry C 115, 2066 (2010).

[203]. Weston MH, Farha OK, Hauser BG, Hupp JT, and Nguyen ST, Chemistry of Materials 24, 1292 (2012).

[204]. U. Energy Information Administration, How much carbon dioxide is produced when different fuels are burned?, https://www.eia.gov/tools/faqs/faq.php?id=73&t=11, Accessed 15 May 2019.

[205]. Wang MQ and Huang HS, A full fuel-cycle analysis of energy and emissions impacts of transportation fuels produced from natural gas, Argonne National Laboratory Technical Report, https://publications.anl.gov/anlpubs/2000/01/34988.pdf, Accessed 15 May 2019 (2000).

[206]. Alvarez RA, Pacala SW, Winebrake JJ, Chameides WL, and Hamburg SP, Proceedings of the National Academy of Sciences 109, 6435 (2012).

[207]. Osborn SG, Vengosh A, Warner NR, and Jackson RB, Proceedings of the National Academy of Sciences 108, 8172 (2011).

[208]. Davis SC, Diegel SW, and Boundy RG, Transportation energy data book, Oak Ridge National Laboratory, https://info.ornl.gov/sites/publications/files/pub31202.pdf, Accessed 15 May 2019 (2009).

[209]. Hasan MMF, Zheng AM, and Karimi IA, Industrial & Engineering Chemistry Research 48, 9571 (2009).

[210]. Makal TA, Li J-R, Lu W, and Zhou H-C, Chemical Society Reviews 41, 7761 (2012). [PubMed: 22990753]

[211]. Advanced Research Projects Agency- Energy (ARPA-E), Move (methane opportunities for vehicular energy) program overview, https://arpa-e.energy.gov/sites/default/files/documents/files/MOVE_ProgramOverview.pdf (2012), accessed 4 June 2019.

[212]. Zhang H, Deria P, Farha OK, Hupp JT, and Snurr RQ, Energy & Environmental Science 8, 1501 (2015).

[213]. Wilmer CE, Leaf M, Lee CY, Farha OK, Hauser BG, Hupp JT, and Snurr RQ, Nature Chemistry 4, 83 (2012).

[214]. Ma S, Sun D, Simmons JM, Collier CD, Yuan D, and Zhou H-C, Journal of the American Chemical Society 130, 1012 (2008). [PubMed: 18163628]

[215]. Barthelet K, Marrot J, Riou D, and Férey G, Angewandte Chemie International Edition41, 281 (2002). [PubMed: 12491409]

[216]. Lin X, Telepeni I, Blake AJ, Dailly A, Brown CM, Simmons JM, Zoppi M, Walker GS, Thomas KM, Mays TJ, et al., Journal of the American Chemical Society 131, 2159 (2009). [PubMed: 19159298]

[217]. Wilmer CE, Farha OK, Yildirim T, Eryazici I, Krungleviciute V, Sarjeant AA, Snurr RQ, and Hupp JT, Energy & Environmental Science 6, 1158 (2013).

[218]. Gómez-Gualdrón DA, Gutov OV, Krungleviciute V, Borah B, Mondloch JE, Hupp JT, Yildirim T, Farha OK, and Snurr RQ, Chemistry of Materials 26, 5632 (2014).

[219]. Wu H, Zhou W, and Yildirim T, Journal of the American Chemical Society 131, 4995 (2009). [PubMed: 19275154]

[220]. Gándara F, Furukawa H, Lee S, and Yaghi OM, Journal of the American Chemical Society 136, 5271 (2014). [PubMed: 24661065]

[221]. Koroneos C, Dompros A, Roumbas G, and Moussiopoulos N, International Journal of Hydrogen Energy 29, 1443 (2004).

[222]. Sakintuna B, Lamari-Darkrim F, and Hirscher M, International Journal of Hydrogen Energy 32, 1121 (2007).

[223]. Yang J, Sudik A, Wolverton C, and Siegel DJ, Chemical Society Reviews 39, 656 (2010). [PubMed: 20111786]

[224]. U.S. Department of Energy Office of Energy Efficiency & Renewable Energy, DOE technical targets for onboard hydrogen storage for light-duty vehicles, https://www.energy.gov/eere/fuelcells/doe-technical-targets-onboard-hydrogen-storage-light-duty-vehicles, accessed 12 December 2018.

[225]. Murray LJ, Dinc M, and Long JR, Chemical Society Reviews 38, 1294 (2009). [PubMed: 19384439]

[226]. Gómez-Gualdrón DA, Colón YJ, Zhang X, Wang TC, Chen Y-S, Hupp JT, Yildirim T, Farha OK, Zhang J, and Snurr RQ, Energy & Environmental Science 9, 3279 (2016).

[227]. Ahmed A, Liu Y, Purewal J, Tran LD, Wong-Foy AG, Veenstra M, Matzger AJ, and Siegel DJ, Energy & Environmental Science 10, 2459 (2017).

[228]. Kaye SS, Dailly A, Yaghi OM, and Long JR, Journal of the American Chemical Society 129, 14176 (2007). [PubMed: 17967030]

[229]. Rowsell JLC and Yaghi OM, Journal of the American Chemical Society 128, 1304 (2006). [PubMed: 16433549]

[230]. Denysenko D, Grzywa M, Tonigold M, Streppel B, Krkljus I, Hirscher M, Mugnaioli E,Kolb U, Hanss J, and Volkmer D, Chemistry - A European Journal 17, 1837 (2011).

[231]. García-Holley P, Schweitzer B, Islamoglu T, Liu Y, Lin L, Rodriguez S, Weston MH,Hupp JT, Gómez-Gualdrón DA, Yildirim T, et al., ACS Energy Letters 3, 748 (2018).

[232]. Ahmed A, Seth S, Purewal J, Wong-Foy AG, Veenstra M, Matzger AJ, and Siegel DJ, Nature Communications 10, 1568 (2019).

[233]. Martin RL, Lin L-C, Jariwala K, Smit B, and Haranczyk M, The Journal of Physical Chemistry C 117, 12159 (2013).

[234]. Prasad TK and Suh MP, Chemistry - A European Journal 18, 8673 (2012).

[235]. Rankine D, Avellaneda A, Hill MR, Doonan CJ, and Sumby CJ, Chemical Communications 48, 10328 (2012). [PubMed: 22971691]

[236]. Koh K, Van Oosterhout JD, Roy S, Wong-Foy AG, and Matzger AJ, Chemical Science 3, 2429 (2012).

[237]. Stoller JK, Panos RJ, Krachman S, Doherty DE, Make B, and L. term Oxygen Treatment Trial Research Group, Chest 138, 179 (2010).

[238]. Raghu G, Collard HR, Egan JJ, Martinez FJ, Behr J, Brown KK, Colby TV,Cordier J-F, Flaherty KR, and Lasky JA, American Journal of Respiratory and Critical Care Medicine 183, 788 (2011). [PubMed: 21471066]

[239]. West EJ, Measurement and Control 4, T96 (1971).

[240]. Widlund D, Medvedev A, and Gyllenram R, IFAC Proceedings Volumes 31, 69 (1998).

[241]. McGovern SJ and Yeigh JH Jr, Fcc regeneration (1983), uS Patent 4,370,222.

[242]. DeCoste JB, Weston MH, Fuller PE, Tovar TM, Peterson GW, LeVan MD, and Farha OK, Angewandte Chemie International Edition 53, 14092 (2014). [PubMed: 25319881]

[243]. Moghadam PZ, Islamoglu T, Goswami S, Exley J, Fantham M, Kaminski CF, Snurr RQ, Farha OK, and Fairen-Jimenez D, Nature Communications 9, 1378 (2018).

[244]. Dávila M, Riccardo JL, and Ramirez-Pastor AJ, Chemical Physics Letters 477, 402 (2009).

[245]. Sholl DS and Lively RP, Nature 532, 435 (2016). [PubMed: 27121824]

[246]. Herm ZR, Bloch ED, and Long JR, Chemistry of Materials 26, 323 (2013).

[247]. Lenzen M, Energy Conversion and Management 49, 2178 (2008).

[248]. Sood DD and Patil SK, Journal of Radioanalytical and Nuclear Chemistry 203, 547 (1996).

[249]. Soelberg NR, Garn TG, Greenhalgh MR, Law JD, Jubin R, Strachan DM, and Thallapally PK, Science and Technology of Nuclear Installations 2013, 1 (2013).

[250]. Banerjee D, Cairns AJ, Liu J, Motkuri RK, Nune SK, Fernandez CA, Krishna R, Strachan DM, and Thallapally PK, Accounts of Chemical Research 48, 211 (2015). [PubMed: 25479165]

[251]. Liu J, Fernandez CA, Martin PF, Thallapally PK, and Strachan DM, Industrial & Engineering Chemistry Research 53, 12893 (2014).

[252]. Banerjee D, Simon CM, Elsaidi SK, Haranczyk M, and Thallapally PK, Chem 4, 466 (2018).

[253]. Banerjee D, Simon CM, Plonka AM, Motkuri RK, Liu J, Chen X, Smit B, Parise JB, Haranczyk M, and Thallapally PK, Nature Communications 7, 11831 (2016).

[254]. Banerjee D, Zhang Z, Plonka AM, Li J, and Parise JB, Crystal Growth & Design 12, 2162 (2012).

[255]. Sikora BJ, Wilmer CE, Greenfield ML, and Snurr RQ, Chemical Science 3, 2217 (2012).

[256]. Ryan P, Farha OK, Broadbelt LJ, and Snurr RQ, AIChE Journal 57, 1759 (2011).

[257]. Aroniadou-Anderjaska V, Figueiredo TH, Apland JP, Qashu F, and Braga MF, NeuroToxicology 30, 772 (2009). [PubMed: 19591865]

[258]. Milatovi D and Jokanovi M, in Handbook of Toxicology of Chemical Warfare Agents, edited by Gupta RC(Academic Press, San Diego, 2009), chap. 65, pp. 985–996.

[259]. Agrawal M, Gallis DFS, Greathouse JA, and Sholl DS, The Journal of Physical Chemistry C 122, 26061 (2018).

[260]. Figueiredo TH, Apland JP, Braga MFM, and Marini AM, Epilepsia 59, 92 (2018).

[261]. Barea E, Montoro C, and Navarro JAR, Chemical Society Reviews 43, 5419 (2014). [PubMed: 24705539]

[262]. DeCoste JB and Peterson GW, Chemical Reviews 114, 5695 (2014). [PubMed: 24750116]

[263]. Matito-Martos I, Moghadam PZ, Li A, Colombo V, Navarro JA, Calero S, and Fairen-Jimenez D, Chemistry of Materials 30, 4571 (2018).

[264]. Küsgens P, Rose M, Senkovska I, Fröde H, Henschel A, Siegle S, and Kaskel S, Microporous and Mesoporous Materials 120, 325 (2009).

[265]. Colombo V, Galli S, Choi HJ, Han GD, Maspero A, Palmisano G, Masciocchi N, and Long JR, Chemical Science 2, 1311 (2011).

[266]. Smil V, Annual Review of Energy and the Environment 25, 21 (2000).

[267]. Ciais P, Sabine C, Bala G, Bopp L, Brovkin V, Canadell J, Chhabra A, DeFries R, Galloway J, Heimann M, et al., Carbon and Other Biogeochemical Cycles (Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013), book section 6, pp. 465–570, ISBN ISBN 978-1-107-66182-0, URL www.climatechange2013.org.

[268]. Collins M, Knutti R, Arblaster J, Dufresne J-L, Fichefet T, Friedlingstein P, Gao X, Gutowski W, Johns T, Krinner G, et al., Long-term Climate Change: Projections, Commitments and Irreversibility (Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013), book section 12, pp. 1029–1136, ISBN ISBN 978-1-107-66182-0, URL www.climatechange2013.org.

[269]. US Energy Information Administration, How much of u.s. carbon dioxide emissions are associated with electricity generation? (2019), URL https://www.eia.gov/tools/faqs/faq.php?id=77&t=11.

[270]. Lal R, Philosophical Transactions of the Royal Society B: Biological Sciences 363, 815 (2007).

[271]. Haszeldine RS, Science 325, 1647 (2009). [PubMed: 19779187]

[272]. Sumida K, Rogow DL, Mason JA, McDonald TM, Bloch ED, Herm ZR, Bae T-H, and Long JR, Chemical Reviews 112, 724 (2012). [PubMed: 22204561]

[273]. Whitley D, Statistics and Computing 4, 65 (1994).

[274]. Herm ZR, Swisher JA, Smit B, Krishna R, and Long JR, Journal of the American Chemical Society 133, 5664 (2011). [PubMed: 21438585]

[275]. McDonald TM, D'Alessandro DM, Krishna R, and Long JR, Chemical Science 2, 2022 (2011).

[276]. Cannella WJ, Xylenes and Ethylbenzene (Wiley Online Library, New York, 2000).

[277]. Yang Y, Bai P, and Guo X, Industrial & Engineering Chemistry Research 56, 14725 (2017).

[278]. Gonzalez MI, Kapelewski MT, Bloch ED, Milner PJ, Reed DA, Hudson MR, Mason JA, Barin G, Brown CM, and Long JR, Journal of the American Chemical Society 140, 3412 (2018). [PubMed: 29446932]

[279]. Minceva M and Rodrigues AE, AIChE Journal 53, 138 (2007).

[280]. Moïse J-C and Bellat J-P, The Journal of Physical Chemistry B 109, 17239 (2005). [PubMed: 16853200]

[281]. Gee JA, Zhang K, Bhattacharyya S, Bentley J, Rungta M, Abichandani JS, Sholl DS, and Nair S, The Journal of Physical Chemistry C 120, 12075 (2016).

[282]. LeCun Y, Cortes C, and Burges CJC, The MNIST database of handwritten digits, http://yann.lecun.com/exdb/mnist/, Accessed 1 April 2019 (1998).

[283]. Wan L, Zeiler M, Zhang S, Le Cun Y, and Fergus R, Proceedings of the 30th International Conference on Machine Learning 28, 1058 (2013).

[284]. Krizhevsky A and Hinton G, Learning multiple layers of features from tiny images, University of Toronto Technical Report, https://www.cs.toronto.edu/~kriz/learning-features-2009-TR.pdf, Accessed 15 May 2019 (2009).

[285]. Bennett J and Lanning S, in Proceedings of the KDD Cup Workshop 2007 (ACM, New York, 2007), pp. 3–6, URL http://www.cs.uic.edu/~liub/KDD-cup-2007/NetflixPrize-description.pdf.

[286]. Bell RM and Koren Y, ACM SIGKDD Explorations Newsletter 9, 75 (2007).

[287]. Koren Y, Bell R, and Volinsky C, Computer 42, 30 (2009).

[288]. Koren Y, Netflix prize documentation 81, 1 (2009).

[289]. Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, Huang Z, Karpathy A, Khosla A, Bernstein M, et al., International Journal of Computer Vision 115, 211 (2015).

[290]. Price SL, Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences 474, 20180351 (2018).

[291]. Reilly AM, Cooper RI, Adjiman CS, Bhattacharya S, Boese AD, Brandenburg JG,Bygrave PJ, Bylsma R, Campbell JE, Car R, et al., Acta Crystallographica Section B: Structural Science, Crystal Engineering and Materials 72, 439 (2016).

[292]. Accessed 5 February 2019, URL https://adsorption.nist.gov/factlab.

[293]. Nguyen HGT, Espinal L, van Zee RD, Thomas M, Toman B, Hudson MSL, Mangano E, Brandani S, Broom DP, Benham MJ, et al., Adsorption 24, 531 (2018). [PubMed: 30956405]

[294]. Siderius DW, Shen VK, Johnson RD III, and van Zee RD, eds., NIST/ARPA-E Database of Novel and Emerging Adsorbent Materials (National Institute of Standards and Technology, Gaithersburg, MD, 20899, 2014), 10.18434/T43882, Accessed 14 December 2018, URL 10.18434/T43882.

[295]. Heller SR, McNaught A, Pletnev I, Stein S, and Tchekhovskoi D, Journal of Cheminformatics 7, 23 (2015). [PubMed: 26136848]

[296]. Britt D, Tranchemontagne D, and Yaghi OM, Proceedings of the National Academy of Sciences 105, 11623 (2008).

[297]. Accessed 5 February 2019, URL https://www.sigmaaldrich.com/catalog/product/aldrich/688614?lang=en&region=US.

[298]. Siderius D and Shen V, eds., NIST Registry of Adsorbent Materials (National Institute of Standards and Technology, Gaithersburg, MD, 20899, 2017), 10.18434/t4/1502644, Accessed 14 December 2018, URL 10.18434/t4/1502644.

[299]. Park J, Howe JD, and Sholl DS, Chemistry of Materials 29, 10487 (2017).

[300]. Simon CM, Smit B, and Haranczyk M, Computer Physics Communications 200, 364 (2016).

[301]. Simon CM, Smit B, and Haranczyk M, Documentation for pyiast, accessed 1 February 2019, URL https://pyiast.readthedocs.io.

[302]. Colón YJ, Fairen-Jimenez D, Wilmer CE, and Snurr RQ, The Journal of Physical Chemistry C 118, 5383 (2014).

[303]. Gómez-Gualdrón DA, Wilmer CE, Farha OK, Hupp JT, and Snurr RQ, The Journal of Physical Chemistry C 118, 6941 (2014).

[304]. Accessed 25 February 2019, URL http://www.fluidproperties.org.

[305]. Ross RB, Ahmad R, Brennan JK, Frankel KA, Moore JD, Moore JD, Mountain RD, Shen VK, Schultz NE, Siderius DW, et al., Fluid Phase Equil 366, 136 (2014).

[306]. Ross RB, Brennan JK, Frankel KA, Moore JD, Moore JD, Mountain RD, Ahmad R, Thommes M, Shen VK, Schultz NE, et al., Fluid Phase Equil 366, 141 (2014).

[307]. Schultz NE, Riaz A, Brennan JK, Frankel KA, Moore JD, Moore JD, Mountain RD, Ross RB, Thommes M, Shen VK, et al., Adsorption Sci. Tech 34, 3 (2016).

[308]. Ross RB, Aeschliman DB, Ahmad R, Brennan JK, Brostrom ML, Frankel KA, Moore JD, Moore JD, Mountain RD, Poirier DM, et al., Adsorption Sci. Tech 34, 13 (2016).

[309]. Rouquerol F, Rouquerol J, and Sing K, Adsorption by Powders and Porous Solids (Academic Press, London, 1999).

[310]. Accessed 22 February 2019, URL https://adsorption.nist.gov/isodb/index.php#apis.

[311]. Hall SR, Allen FH, and Brown ID, Acta Crystallographica Section A: Foundations and Devices A47, 655 (1991).

[312]. Brown ID and McMahon B, Acta Crystallographica Section B: Structural Science, Crystal Engineering and Materials 58, 317 (2002).

[313]. Curtarolo S, Hart GL, Nardelli MB, Mingo N, Sanvito S, and Levy O, Nature Materials 12, 191 (2013). [PubMed: 23422720]

[314]. Pyzer-Knapp EO, Suh C, Gómez-Bombarelli R, Aguilera-Iparraguirre J, and Aspuru-Guzik A, Annual Review of Materials Research 45, 195 (2015).

[315]. Gómez-Bombarelli R, Aguilera-Iparraguirre J, Hirzel TD, Duvenaud D, Maclaurin D, Blood-Forsythe MA, Chae HS, Einzinger M, Ha D-G, Wu T, et al., Nature Materials 15, 1120 (2016). [PubMed: 27500805]

[316]. Ceder G, MRS Bulletin 35, 693 (2010).

[317]. Hachmann J, Olivares-Amaya R, Atahan-Evrenk S, Amador-Bedolla C, Sánchez-Carrera RS, Gold-Parker A, Vogt L, Brockway AM, and Aspuru-Guzik A, The Journal of Physical Chemistry Letters 2, 2241 (2011).

[318]. Nørskov JK, Bligaard T, Rossmeisl J, and Christensen CH, Nature Chemistry 1, 37 (2009).

[319]. Greeley J, Jaramillo TF, Bonde J, Chorkendorff I, and Nørskov JK, Nature Materials 5, 909 (2006). [PubMed: 17041585]

[320]. Strasser P, Fan Q, Devenney M, Weinberg WH, Liu P, and Nørskov JK, The Journal of Physical Chemistry B 107, 11013 (2003).

[321]. Kitchen DB, Decornez H, Furr JR, and Bajorath J, Nature Reviews Drug Discovery 3, 935 (2004). [PubMed: 15520816]

[322]. Xu J and Hagler A, Molecules 7, 566 (2002).

[323]. Flaig RW, Osborn Popp TM, Fracaroli AM, Kapustin EA, Kalmutzki MJ, Altamimi RM, Fathieh F, Reimer JA, and Yaghi OM, Journal of the American Chemical Society 139, 12125 (2017). [PubMed: 28817269]

[324]. Bachman J, Reed DA, Kapelewski MT, Chachra G, Jonnavittula D, Radaelli G, and Long JR, Energy & Environmental Science 11, 2423 (2018).

[325]. Nguyen JG and Cohen SM, Journal of the American Chemical Society 132, 4560 (2010). [PubMed: 20232871]

[326]. Poloni R, Lee K, Berger RF, Smit B, and Neaton JB, The Journal of Physical Chemistry Letters 5, 861 (2014). [PubMed: 26274079]

[327]. Coudert F-X, Chemistry of Materials 27, 1905 (2015).

[328]. Wieme J, Lejaeghere K, Kresse G, and Van Speybroeck V, Nature Communications 9, 4899 (2018).

[329]. Coudert F-X, Chemistry of Materials 29, 2615 (2017).

[330]. Greathouse JA, Kinnibrugh TL, and Allendorf MD, Industrial & Engineering Chemistry Research 48, 3425 (2009).

[331]. Boutin A, Springuel-Huet M-A, Nossov A, Gédéon A, Loiseau T, Volkringer C, Férey G, Coudert F-X, and Fuchs AH, Angewandte Chemie International Edition 48, 8464 (2009).

[332]. Gonzalez-Nelson A, Coudert F-X, and van der Veen M, Nanomaterials 9, 330 (2019).

[333]. Catalano L and Naumov P, CrystEngComm 20, 5872 (2018).

[334]. Witman M, Ling S, Jawahery S, Boyd PG, Haranczyk M, Slater B, and Smit B, Journal of the American Chemical Society 139, 5547 (2017). [PubMed: 28357850]

[335]. Krause S, Evans JD, Bon V, Senkovska I, Iacomi P, Kolbe F, Ehrling S, Troschke E, Getzschmann J, Többens DD, et al., ChemRxiv (2019), 10.26434/chemrxiv.7796543.

[336]. Evans JD, Bocquet L, and Coudert F-X, Chem 1, 873 (2016).

[337]. Krause S, Bon V, Senkovska I, Stoeck U, Wallacher D, Többens DM, Zander S, Pillai RS, Maurin G, Coudert F-X, et al., Nature 532, 348 (2016). [PubMed: 27049950]

[338]. Lee Y, Poloni R, and Kim J, Journal of Computational Chemistry 37, 2808 (2016). [PubMed: 27718253]

[339]. Fetisov EO, Shah MS, Long JR, Tsapatsis M, and Siepmann JI, Chemical Communications 54, 10816 (2018). [PubMed: 30152487]

[340]. Chen L, Grajciar L, Nachtigall P, and Duren T, The Journal of Physical Chemistry C 115, 23074 (2011).

[341]. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, and Kollman PA, Journal of the American Chemical Society 117, 5179 (1995).

[342]. Jorgensen WL and Tirado-Rives J, Journal of the American Chemical Society 110, 1657 (1988). [PubMed: 27557051]

[343]. McDaniel JG, Li S, Tylianakis E, Snurr RQ, and Schmidt JR, The Journal of Physical Chemistry C 119, 3143 (2015).

[344]. Jeziorski B, Moszynski R, and Szalewicz K, Chemical Reviews 94, 1887 (1994).

[345]. Dokur D and Keskin S, Industrial & Engineering Chemistry Research 57, 2298 (2018). [PubMed: 29503503]

[346]. Behler J and Parrinello M, Physical Review Letters 98, 146401 (2007). [PubMed: 17501293]

[347]. Smith JS, Isayev O, and Roitberg AE, Chemical Science 8, 3192 (2017). [PubMed: 28507695]

[348]. Behler J, Physical Chemistry Chemical Physics 13, 17930 (2011). [PubMed: 21915403]

[349]. Schütt KT, Arbabzadah F, Chmiela S, Müller KR, and Tkatchenko A, Nature Communications 8, 13890 (2017).

[350]. Bartók AP, Payne MC, Kondor R, and Csányi G, Physical Review Letters 104, 136403 (2010). [PubMed: 20481899]

[351]. Hornik K, Stinchcombe M, and White H, Neural Networks 2, 359 (1989).

[352]. Lowell S, Shields JE, Thomas MA, and Thommes M, Characterization of Porous Solids and Powders: Surface Area, Pore Size and Density (Kluwer, Boston, 2004).

[353]. Xu Q and Zhong C, The Journal of Physical Chemistry C 114, 5035 (2010).

[354]. Argueta E, Shaji J, Gopalan A, Liao P, Snurr RQ, and Gomez-Gualdron DA, Journal of Chemical Theory and Computation 14, 365 (2017). [PubMed: 29227644]

[355]. Breneman CM and Wiberg KB, Journal of Computational Chemistry 11, 361 (1990).

[356]. Bennett TD, Cheetham AK, Fuchs AH, and Coudert F-X, Nature Chemistry 9, 11 (2017).

[357]. Sholl DS and Lively RP, The Journal of Physical Chemistry Letters 6, 3437 (2015). [PubMed: 26268796]

[358]. Cavka JH, Jakobsen S, Olsbye U, Guillou N, Lamberti C, Bordiga S, and Lillerud KP, Journal of the American Chemical Society 130, 13850 (2008). [PubMed: 18817383]

[359]. Wu H, Chua YS, Krungleviciute V, Tyagi M, Chen P, Yildirim T, and Zhou W, Journal of the American Chemical Society 135, 10525 (2013). [PubMed: 23808838]

[360]. Ghosh P, Colón YJ, and Snurr RQ, Chemical Communications 50, 11329 (2014). [PubMed: 25116653]

[361]. Thornton AW, Babarao R, Jain A, Trousselet F, and Coudert F-X, Dalton Transactions 45, 4352 (2016). [PubMed: 26733113]

[362]. Bristow JK, Svane KL, Tiana D, Skelton JM, Gale JD, and Walsh A, The Journal of Physical Chemistry C 120, 9276 (2016).

[363]. Zhang C, Han C, Sholl DS, and Schmidt JR, The Journal of Physical Chemistry Letters 7, 459 (2016). [PubMed: 26771275]

[364]. Han C, Verploegh RJ, and Sholl DS, The Journal of Physical Chemistry Letters 9, 4037 (2018). [PubMed: 29936838]

[365]. Zunger A, Nature 566, 447 (2019). [PubMed: 30814720]

[366]. Cantu DC, McGrail BP, and Glezakou V-A, Chemistry of Materials 26, 6401 (2014).

[367]. Yoneya M, Tsuzuki S, and Aoyagi M, Physical Chemistry Chemical Physics 17, 8649 (2015). [PubMed: 25766115]

[368]. Biswal D and Kusalik PG, ACS Nano 11, 258 (2017). [PubMed: 27997790]

[369]. Biswal D and Kusalik PG, The Journal of Chemical Physics 147, 044702 (2017). [PubMed: 28764378]

[370]. Colón YJ, Guo AZ, Antony LW, Hoffmann KQ, and de Pablo JJ, The Journal of Chemical Physics 150, 104502 (2019). [PubMed: 30876355]

[371]. Kollias L, Cantu DC, Tubbs MA, Rousseau R, Glezakou V-A, and Salvalaglio M, Journal of the American Chemical Society 141, 6073 (2019). [PubMed: 30887804]

[372]. Greathouse JA and Allendorf MD, Journal of the American Chemical Society 128, 10678 (2006). [PubMed: 16910652]

[373]. Bellarosa L, Castillo JM, Vlugt T, Calero S, and López N, Chemistry - A European Journal 18, 12260 (2012).

[374]. Han C, Zhang C, Tymi ska N, Schmidt JR, and Sholl DS, The Journal of Physical Chemistry C 122, 4339 (2018).

[375]. Rogge SMJ, Waroquier M, and Van Speybroeck V, Accounts of Chemical Research 51, 138 (2018). [PubMed: 29155552]

[376]. Rajagopalan AK, Avila AM, and Rajendran A, International Journal of Greenhouse Gas Control 46, 76 (2016).

[377]. First EL, Hasan MMF, and Floudas CA, AIChE Journal 60, 1767 (2014).

[378]. Farmahini AH, Krishnamurthy S, Friedrich D, Brandani S, and Sarkisov L, Industrial & Engineering Chemistry Research 57, 15491 (2018).

[379]. Pullumbi P, Brandani F, and Brandani S, Current Opinion in Chemical Engineering 24, 131 (2019).

[380]. Krishna R, Separation and Purification Technology 194, 281 (2018).

[381]. Leperi K, Chung YG, You F, and Snurr RQ, ACS Sustainable Chemistry & Engineering 7, 11529 (2019), https://doi.org/10.1021/acssuschemeng.9b01418 , URL https://doi.org/10.1021/acssuschemeng.9b01418https://doi.org/10.1021/acssuschemeng.9b01418, URL https://doi.org/10.1021/acssuschemeng.9b01418 .

A lesson from molecular biology

CRISPR-Cas9 technology [116] is revolutionising molecular biology by enabling the facile, precise, and cost-effective editing of genomes [116]. Impacts include accelerating and enabling more systematic experiments to probe gene function and regulation and, potentially, genetic-engineering disease- and stress-resistant crops and correcting genetic and epigenetic human disease, such as cancer [117]. The CRISPR-Cas9 technology originates from an adaptive immune system discovered in bacteria [118]. The fascinating story by which CRISPR was recognised as a bacterial immune system, outlined by Lander [119], bestows useful lessons on materials science, including the importance of "hypothesis-free science".

By 2000, Mojica et al. [120] cataloged the presence of peculiar sequence patterns in the genomes of 20 different microbes using a computer programme to analyse published genomes. Particularly, they found clusters of multiple copies of roughly palindromic sequence base pairs (bp), $\approx$ 24 bp to 40 bp in length, flanking both sides of a unique spacer sequence of roughly consistent length (20 bp to 58 bp). These were descriptively coined clustered regularly interspaced short palindromic repeats (CRISPR) [121]. Mysteriously, the biological role (or lack thereof) of evolutionarily conserved (within species [121]) CRISPRs was unknown.

Five years later, Mojica et al. [122] published evidence that CRISPR is related to a microbial immune system, conferring resistance to e.g., bacteriophages. The link was made by searching databases of DNA molecules for matches of $\approx$ 4 500 known spacer sequences between CRISPRs; 47 of the spacer sequences matched bacteriophage DNA sequences. As further evidence, a microbe strain carrying the CRISPR spacer sequence of a particular virus was found to be immune to infection by that virus, whereas other strains lacking that spacer sequence were susceptible.

Lander [119] credits the role of "hypothesis-free" research in the discovery of CRISPR: "The discovery of the CRISPR loci and their biological function … all emerged not from wet-bench experiments but from open-ended bioinformatic exploration of large-scale, often public, genomic datasets." Similarly, we claim here that hypothesis-free science, i.e., the curation of databases of computation-ready nanoporous crystal structures and adsorption data, can accelerate the pace of nanoporous materials discovery and deployment in un-conceived ways. Of course, mindlessly gathering data is unlikely to be the best allocation of resources [123]; there must be an implicit hypothesis that the data will enable further developments, albeit in an ill-defined or un-conceived context. Our survey of computation-inspired MOF discoveries in Sec. VI will demonstrate the impact of open, computation-ready databases of MOF structures on the discovery of performant MOFs for adsorption-based engineering applications.

A lesson from machine learning

The field of machine learning aims to leverage data to train mathematical models or algorithms to perform a task. For example, instead of explicitly programming a computer to translate speech, identify and classify traffic signs, detect fraudulent financial

transactions, and recommend music, a machine learning algorithm/statistical model is tuned to perform these tasks after taking in many examples (data) as input. As more data is used to train the machine learning model, it generally performs the task more effectively, though with diminishing returns.

The development of the field of machine learning has benefited dramatically from large, open data sets that serve as arenas for machine learning models and methods. These open data sets (i) standardise benchmarking and comparison of models, (ii) reveal insights into the inner-workings and deficiencies of different learning algorithms based on their mistakes while performing intuitive tasks, (iii) foster competition, and (iv) stimulate the development of better-performing methods.

As an example, the open MNIST data set of handwritten digits [282] contains ca. 70 000 labelled, binary images of handwritten digits 0, 1,…, 9. The MNIST data set is partitioned into training and testing sets to provide a widely-used, standard benchmark for classification algorithms. Developments in machine learning algorithms have led to highly accurate handwritten digit classifiers (0.21 % error [283]). As a more challenging arena, CIFAR-10 and -100 [284] data sets each consist of 60 000 colored, labelled images and contain more complicated classes such as frogs, dogs, airplanes, etc.

To directly generate interest and develop new supervised machine learning models and methods, open challenges are held (sometimes prized), where a labelled data set is publicly released for training a model. To rank competitors, a testing data set, where labels are withheld (i.e., the data set has only independent variables), is also released. Teams submit the test set labels predicted by their trained model, and a leaderboard ranks teams according to an evaluation metric e.g., accuracy.

For example, Netflix in 2006 released ratings by ca. 480 000 subscribers on ca. 18 000 movies, comprising ca. 100 million movie ratings from 1 to 5 [285]. Three million ratings (by the same set of subscribers, on the same set of movies) were withheld as test data. In 2009, a $1 million prize was awarded to the team that improved upon the incumbent algorithm of Netflix, Cinematch, by decreasing the root mean square error between predicted and actual ratings on the test set by 10 %. Both the release of the ratings data and the competition generated interest in recommendation systems, spurred the sharing of ideas between groups, and led to advances in recommendation algorithms (which were disseminated) [286, 287]. Koren [288], a member of the team that won the Netflix Prize, noted "a clear spike in related publications, and the Netflix dataset is the direct catalyst to developing some of the better algorithms known in the field". He noted that the teams exhibited a collaborative spirit: "the feeling was of a big community progressing together".

As another example, the ImageNet Large Scale Visual Recognition Challenge is an annual, ongoing challenge since 2010, and it has spurred innovation in object recognition and detection in images [289]. ImageNET is a crowdsource-annotated database of millions of images with hundreds of object categories. An annual workshop is held at the end of the year to disseminate and discuss the most innovative and successful approaches. [289]

Within the realm of materials science, to spur developments and track progress in crystal structure prediction [290], the CCDC holds a challenge to predict crystal structures of molecules [291]. To spur force field development, we envisage holding an open challenge to predict the adsorption isotherms of different gases in MOF structures (holding the experimental MOF adsorption data secret).
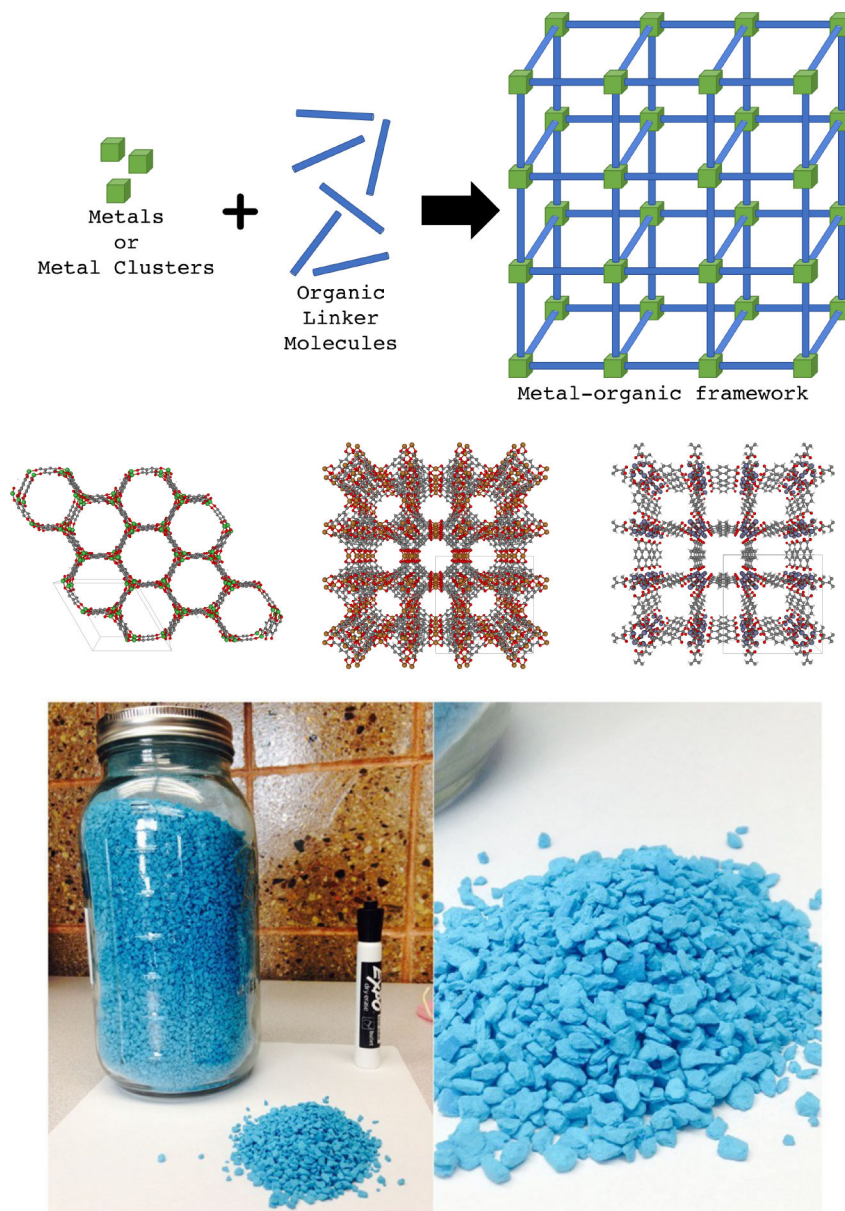
FIG. 1.

The chemistry of metal-organic frameworks (MOFs) is modular and highly tuneable, affording a vast chemical space of crystal structures in which to search for materials exhibiting an optimal adsorption property. (top) MOFs are composed of metal nodes or clusters coordinated to organic linker molecules to form a crystalline, porous framework. By changing the linkers and metal clusters, we can obtain millions of possible materials. Inspired by Ref. 14. (middle) The crystal structures of Ni-MOF-74 [15], HKUST-1 [16], IRMOF-1 [12] are shown as examples. (bottom) A sample of HKUST-1 [16] from Ref. 17. is show to illustrate how a MOF appears from the naked eye.

(a)

(b)

(c)
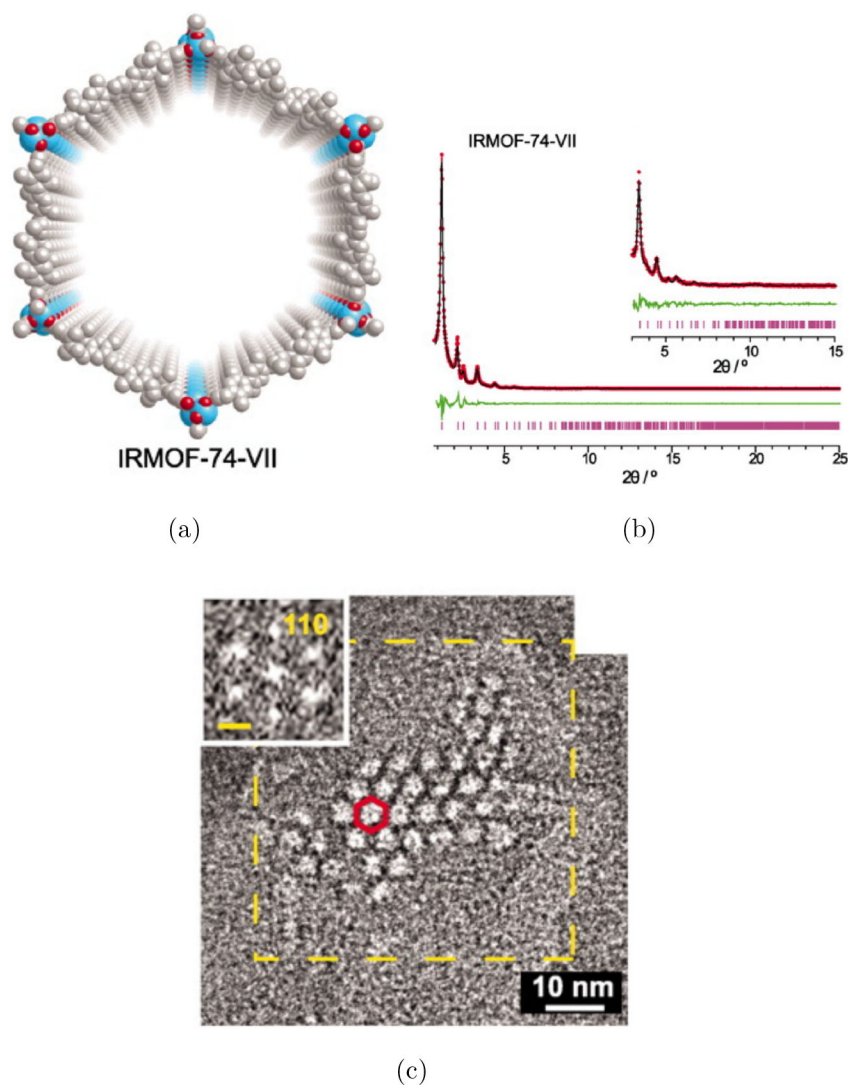
FIG. 2.

The crystal structure of IRMOF-74-VII was determined by PXRD assisted by computational modelling. (a) The suspected and confirmed crystal structure of IRMOF-74-VII. Isoreticular to Mg-MOF-74, organic linkers with seven phenylene rings are connected to magnesium oxide clusters, forming one-dimensional hexagonal channels (blue: Mg, red: O, gray: C). (b) Experimental PXRD pattern (red) and simulated PXRD pattern (black) of a computationally-assembled structure of IRMOF-74-VII. The difference between the experimental and simulated pattern is shown in green. (c) High-resolution transmission electron microscopy images of IRMOF-74-VII show the ordered hexagonal pores as in (a). A fast Fourier transform analysis (inset in upper left; scale bar 2 nm^{-1}) was conducted on the area in the dashed square. The d-spacing from the Fourier transformation, measured from the six reflection spots corresponding to the 110 reflections resolved from the FFT patterns, agreed with those from the PXRD patterns. From Science 2012, 336, 6084, 1018–1023. Reprinted with permission from AAAS.
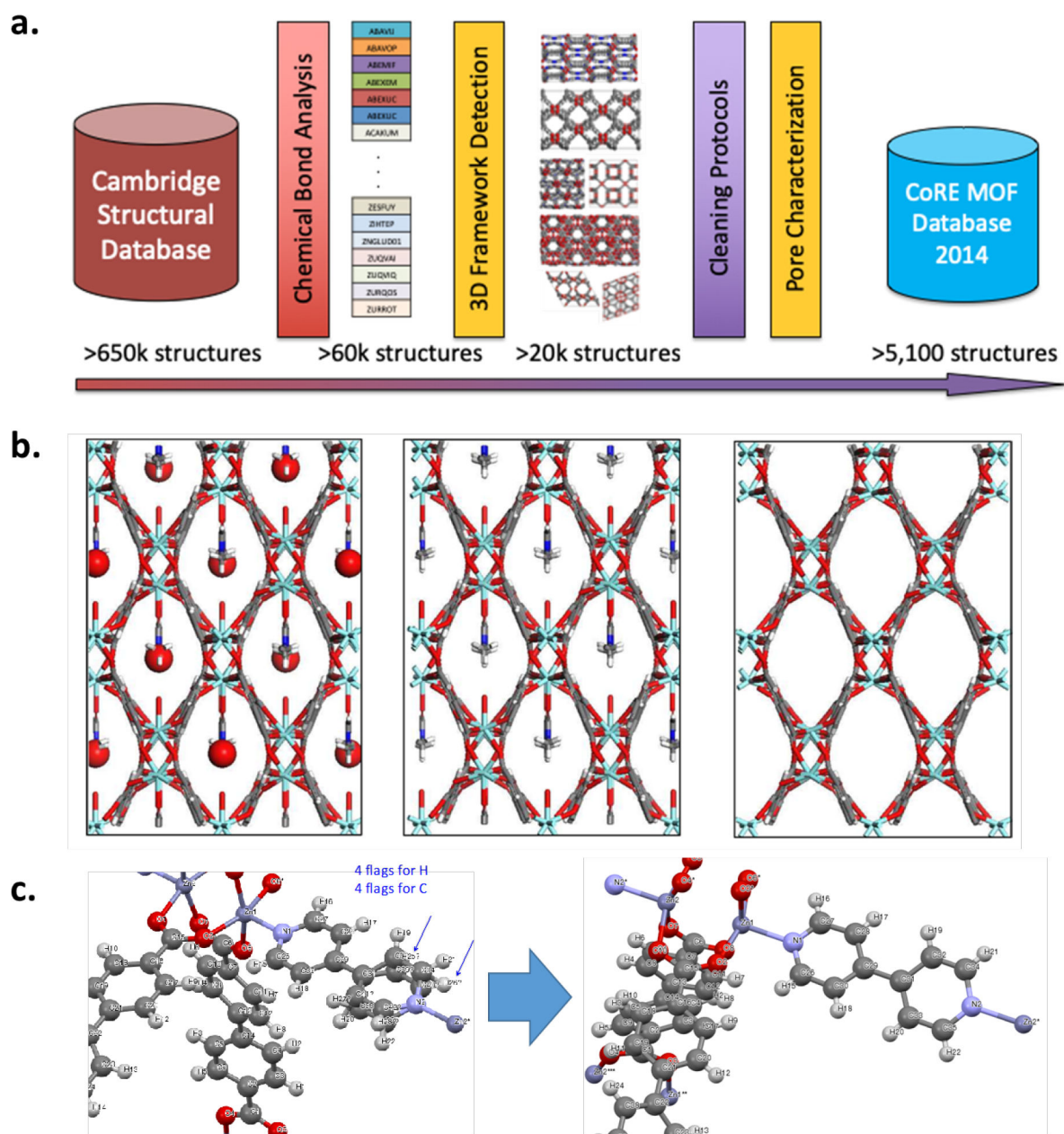
FIG. 3.

Curating the computation-ready, experimental (CoRE) database of MOFs. (a) The workflow of the development of CoRE MOF Database 2014. (b) Automatically removing solvent from MOFs deposited in the CSD. Left to right, as-is structure, free solvent removed, free and bound solvent removed (CSD refcode: VICDOC). (c) Before and after disorder removed (CSD refcode: PIDNEX). Reprinted with permission from Chem. Mater. 2014, 26, 21, 6185–6192. Copyright 2019 American Chemical Society.
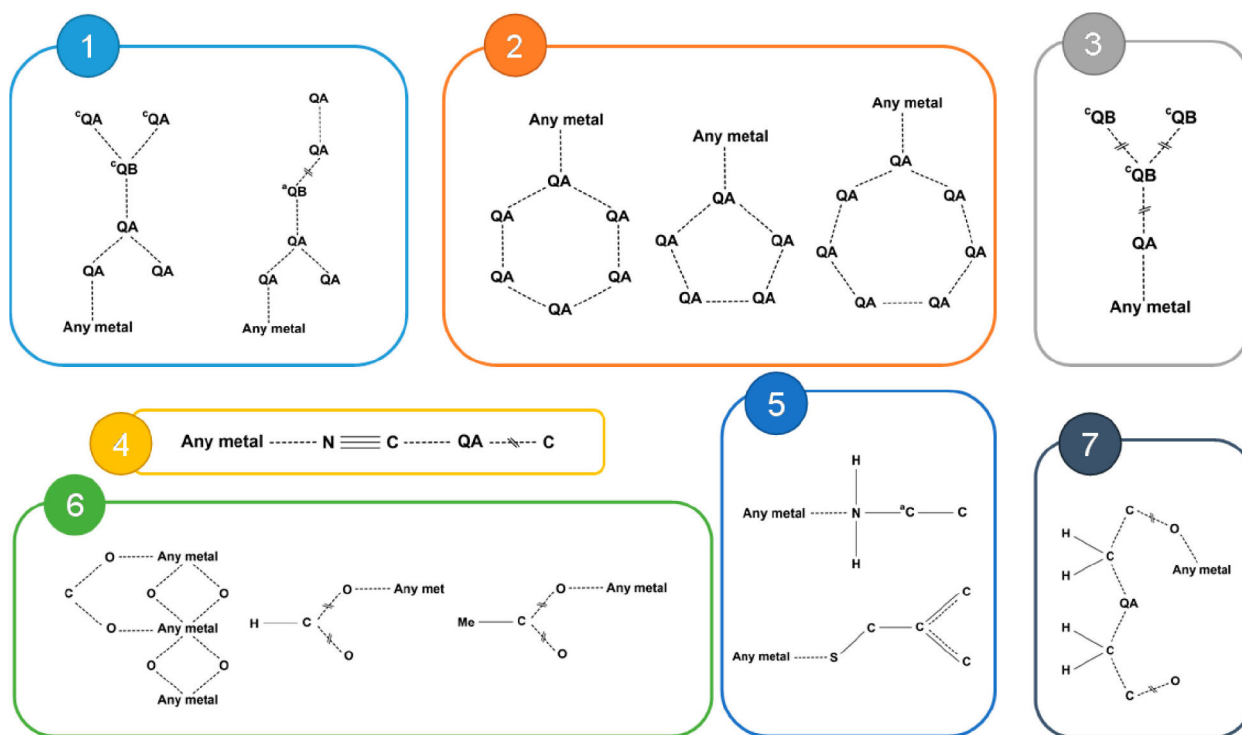
FIG. 4.
The seven chemical bond criteria used by Moghadam and co-workers [10] to search for MOFs in the CSD. Here QA = O, N, P, C, B, S, QB = N, P, B, S, C, ME = methyl group. The superscripts *c* and *a* denote cyclic and acyclic, respectively. Reprinted with permission from Chem. Mater. 2017, 29, 7, 2618–2625. Copyright 2019 American Chemical Society.
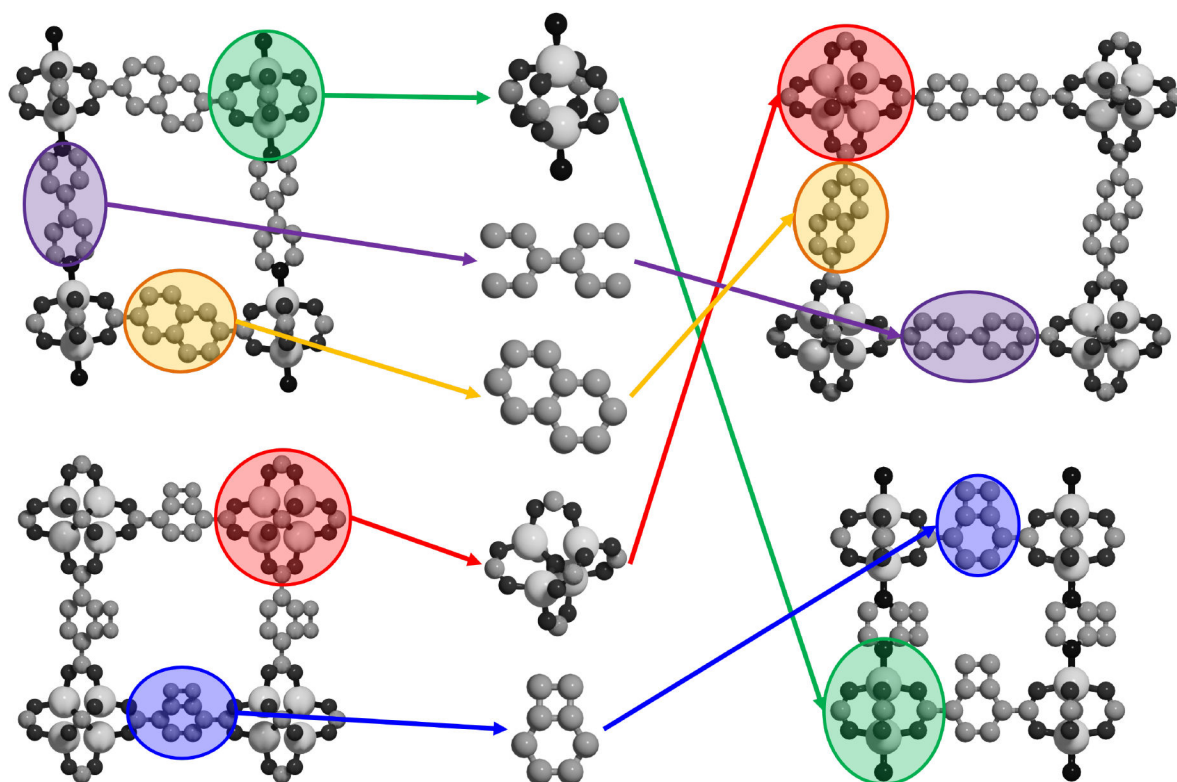
FIG. 5.
The algorithm used by Wilmer et al. [213] to construct hypothetical MOF structures first extracts building blocks from previously-known MOF structures. These building blocks are then combined according to their geometry to create hypothetical MOF structures.
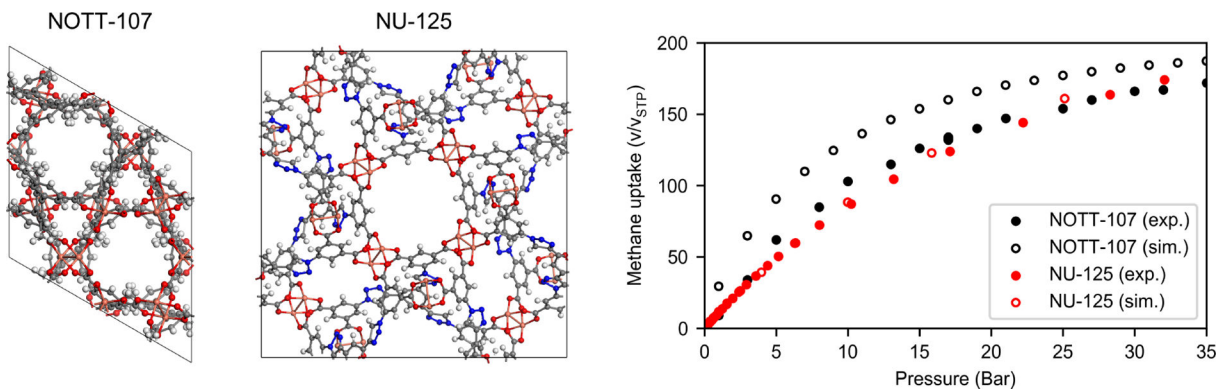
FIG. 6.

The MOFs NOTT-107 and NU-125 were both identified via high throuput computational screening studies [213, 217] as potentially promising materials for methane storage. NU-125, in particular, was pinpointed based on its void fraction of 0.8, which was found as optimal for methane storage from prior large-scale screening studies [213]. Crystal structures of NOTT-107 and NU-125 (left) with methane adsorption isotherms (right) demonstrating comparable performance between the two sorbents.

(a)

(b)

(c)

(d)
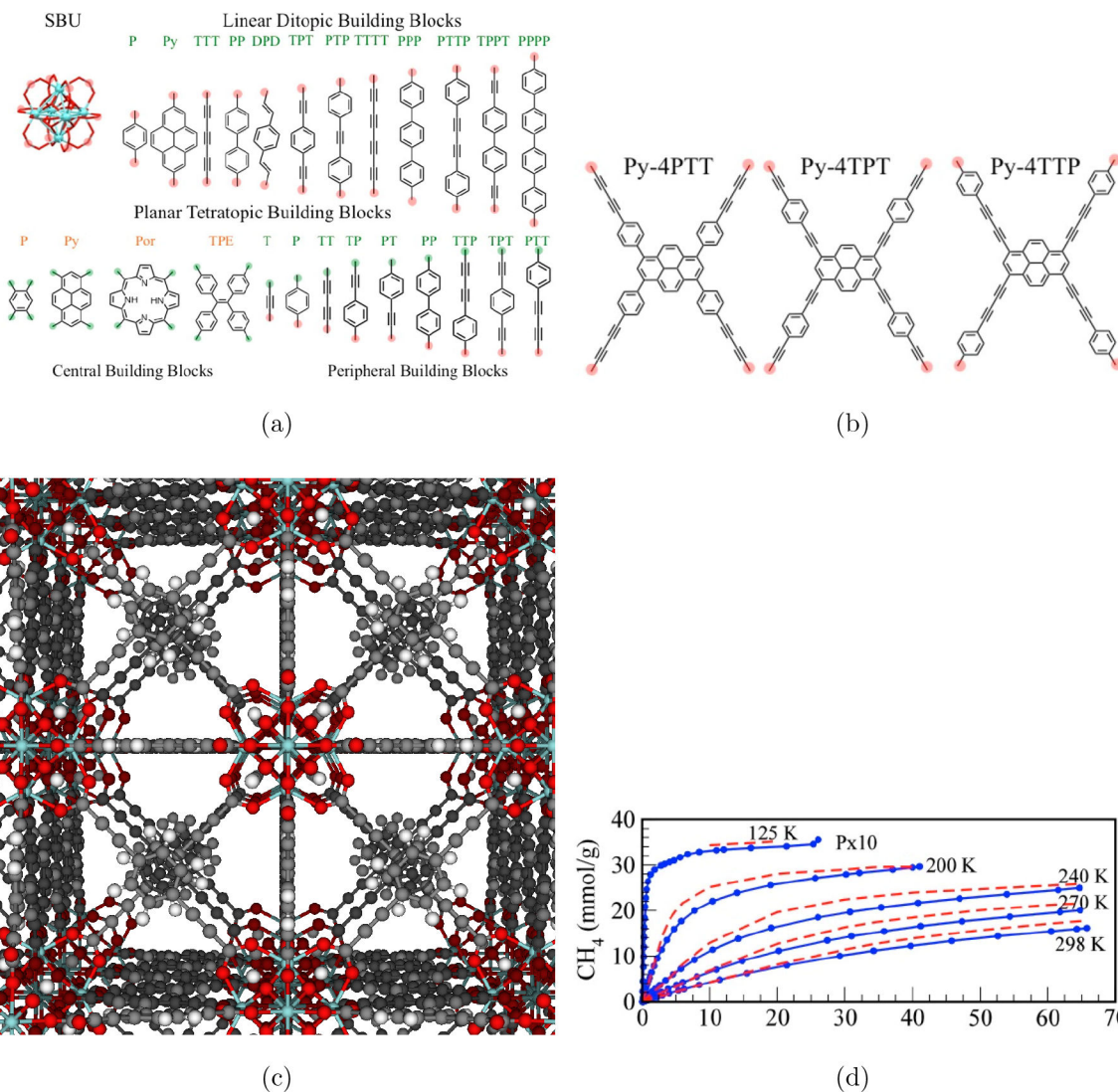
FIG. 7.

Gomez-Gualdron et al. [218] constructed a set of 204 zirconium-based hypothetical MOFs and screened them for storing and delivering methane, leading to the synthesis of NU-800. (a) The basic building blocks used to construct the hypothetical MOFs (connection points highlighted). The ditopic linkers and Zr-SBU are combined in an fcu net. The central and peripheral building blocks are combined to form planar, tetratopic building units (e.g., see (b)), which are then assembled with the Zr-SBU in the ftw, csq, and scu nets. (b) Examples of planar, tetratopic building units obtained by combining one central and four peripheral building blocks (see (a)). (c) The crystal structure of NU-800 has fcu topology and is constructed from the ditopic TPT linker in (a) (cyan: Zr, red: O, grey: C, white: H). (d) Comparison of experimental (points, solid lines) and simulated (dashed lines) isotherms for methane adsorption in NU-800 at various temperatures. Reprinted with permission from Chem. Mater. 2014, 26, 19, 5632–5639. Copyright 2019 American Chemical Society.

(a)
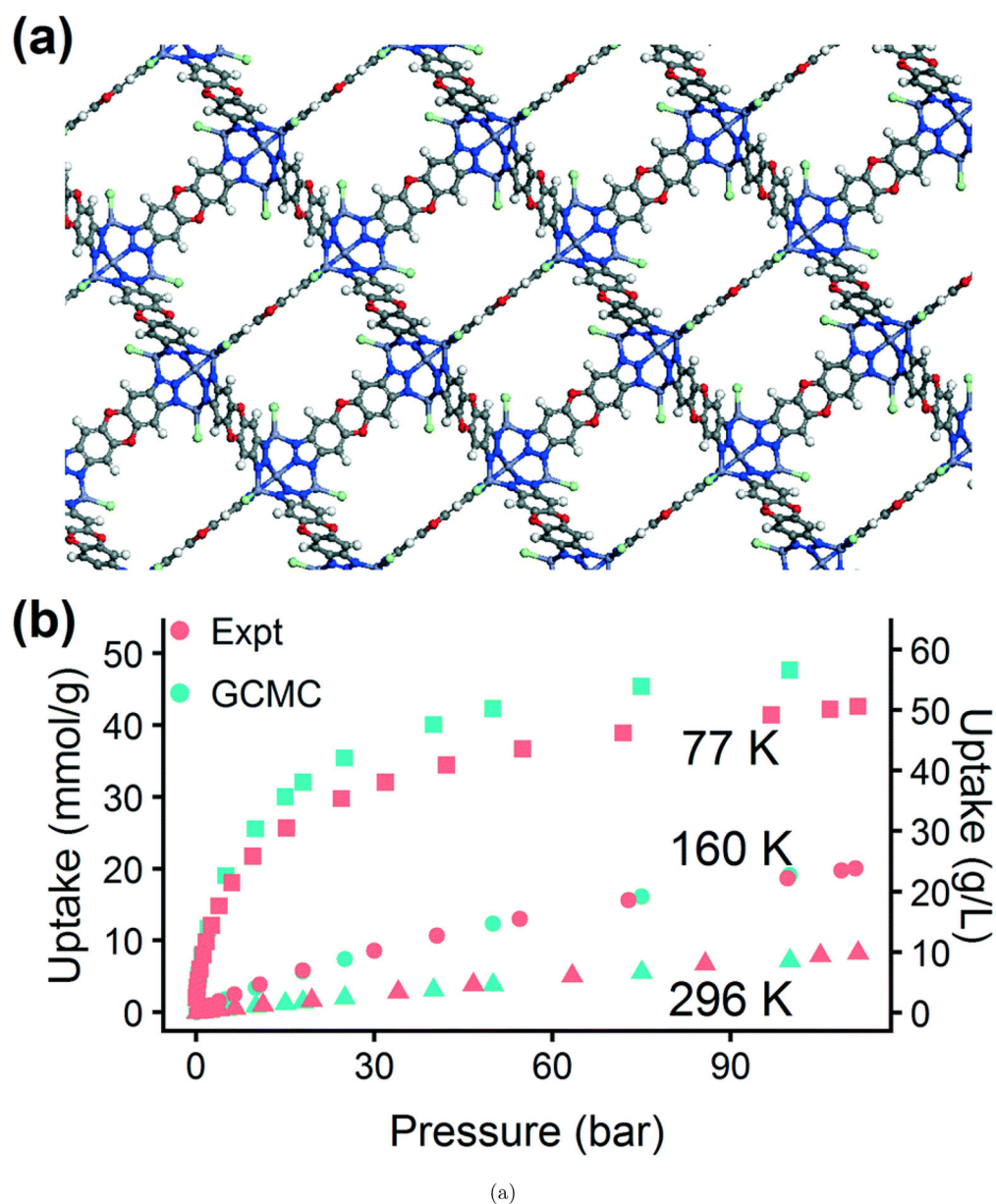
FIG. 8.

A machine-learning accelerated high-throughput screening [98] led to the identification of MFU-4*l* for hydrogen storage. (a) Crystal structure of MFU-4*l*. (b) Simulated and experimental H_2 adsorption isotherms of MFU-4*l* at different temperatures. Reproduced from Mol. Syst. Des. Eng. 2019, 4, 1, 162–174, with permission from The Royal Society of Chemistry.
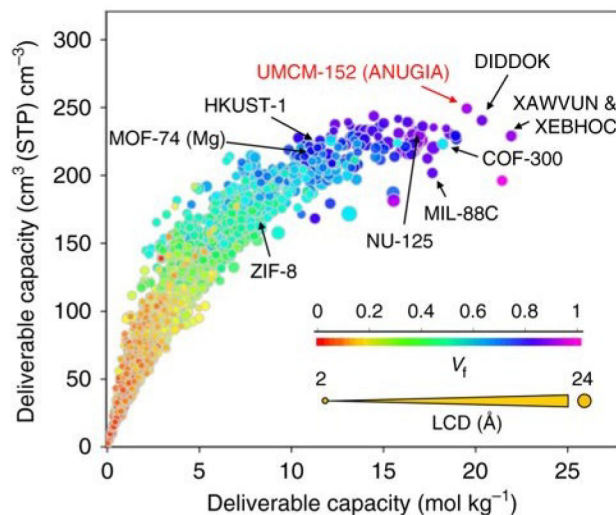
(a)
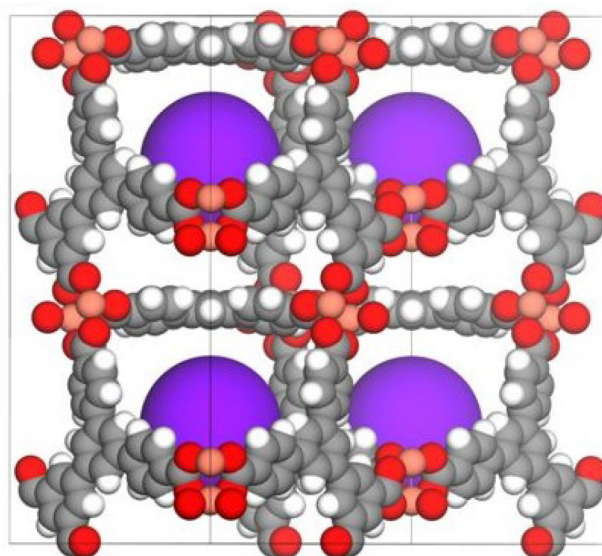
(b)

FIG. 9.

Computational identification of UMCM-125 (CSD refcode: ANUGIA) for oxygen storage and delivery.[243] (a) Computational screening data; each point represents a MOF. The volumetric oxygen deliverable (usable) capacity is largest for MOFs with a void fraction (V_f) above 0.7 and a largest cavity diameter above 7.5 Å to 8 Å. Common MOFs are highlighted, including the MOF predicted to have the largest O_2 usable capacity, UMCM-152. (b) The crystal structure of UMCM-125, with the purple sphere highlighting the main pore. Figure from Nat. Commun., 2018, 9, 1, 1378, under Creative Commons Attribution 4.0 International License https://creativecommons.org/licenses/by/4.0/

(a)

(b)

(c)

(d)
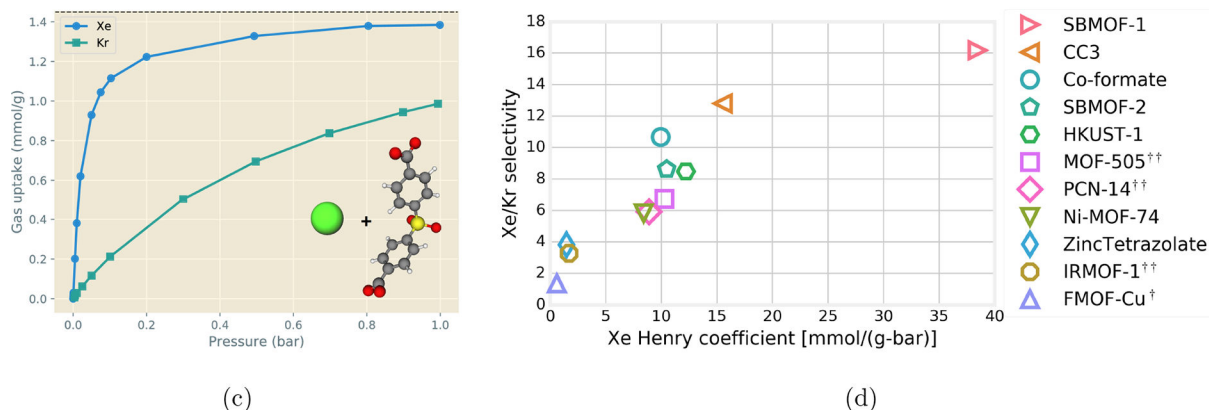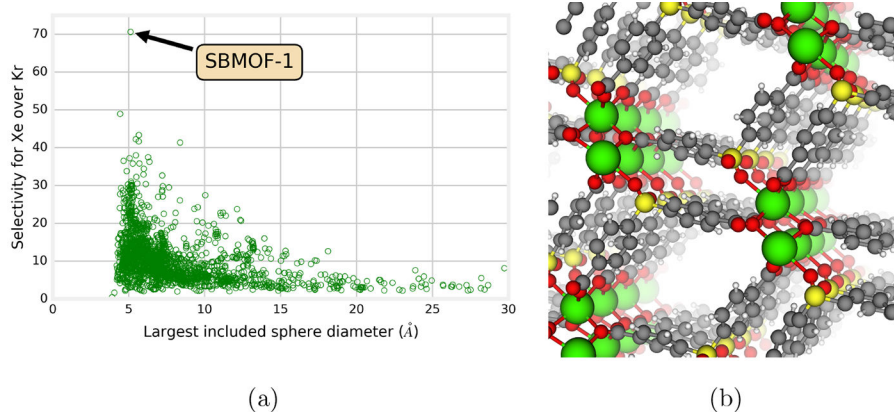
FIG. 10.

SBMOF-1 for Xe/Kr separations [253]. (a) Simulated Xe/Kr selectivity against pore size; each point represents a CoRE MOF; SBMOF-1 is marked. (b) The crystal structure of SBMOF-1 (CSD refcode: KAXQIL) exhibits 1D channels that form well-defined pockets in which xenon can adsorb commensurately with the structure. (green: Ca, yellow: S, red: O, gray: C, white: H) (c) Experimentally measured pure-component Xe and Kr adsorption isotherms in SBMOF-1 at 298 K. Horizontal, dashed line shows one adsorbate per unit cell, indicating commensurate xenon adsorption. Henry coefficients fit to the low-pressure data imply SBMOF-1 harbours an equilibrium Xe/Kr selectivity of 16. Inset shows metal (Ca) and V-shaped organic ligand used to synthesise SBMOF-1. (d) Comparison of experimental, equilibrium Xe/Kr separation capability among MOFs at dilute conditions and at ≈ 298 K; Henry coefficients were extracted from experimental pure-component adsorption isotherms in the literature. Reproduced and adapted from Nat. Commun. 2016, 7, 11831, under Creative Commons Attribution 4.0 International License https://creativecommons.org/licenses/by/4.0/.

(a)

mustard gas
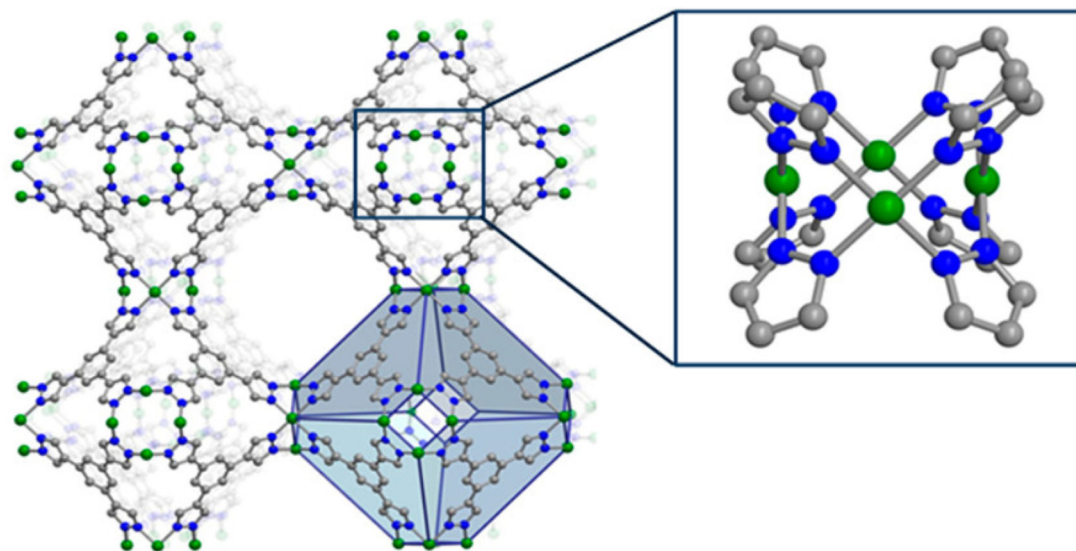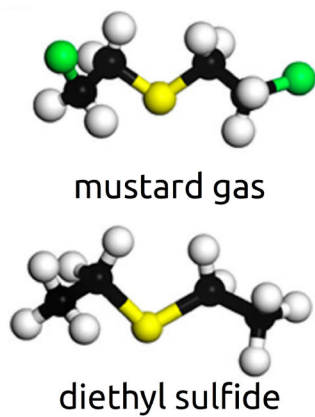
diethyl sulfide

(b)
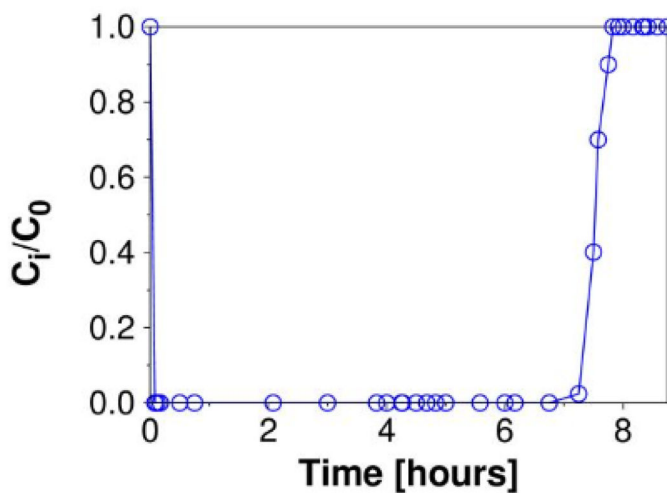
(c)

FIG. 11.

Computational identification of $Ni_3(BTP)_2$ for chemical warfare agent capture [263]. (a) The structure of $Ni_3(BTP)_2$ (CSD refcode: UTEWOG). (b) Mustard gas and its surrogate diethyl sulphide (DES). (c) DES breakthrough curve. A N_2 gas stream with 80 % relative humidity and 1 ppm DES at 298 K is passed, at 20 mL/min, through a column packed with 150 mg of $Ni_3(BTP)_2$. Shown is a measurement, via gas chromatography, of normalised DES concentration at the outlet of the column. Reprinted with permission from Chem. Mater. 2018, 30, 14, 4571–4579. Copyright 2019 American Chemical Society

(a)

(b)

(c)

(d)
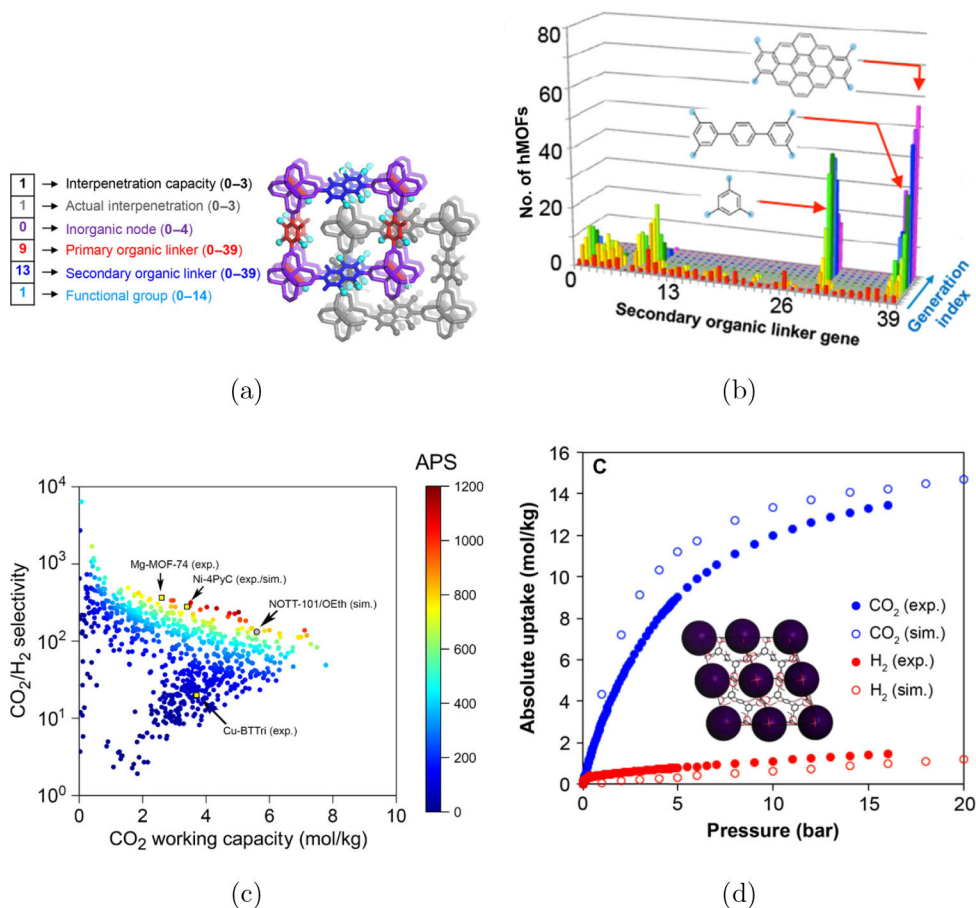
FIG. 12.

The discovery of NOTT-101/OEt was guided using a genetic algorithm in conjunction with molecular simulations to evaluate fitness [114]. (a) A chromosome describes a MOF structure with 6 integers (called genes), each representing a different property of the MOF. The chromosomes define a chemical space that will be explored in the genetic algorithm. (b) A visualisation of the evolution of the primary organic linker gene throughout the genetic algorithm using the adsorption performance score as the fitness. After 10 generations, only 3 genes are predominantly contained within the chromosomes, hinting that these linkers are optimal for CO_2 capture. (c) The simulated adsorption properties of the MOFs from the genetic algorithm search. Each point represents a hypothetical MOF that appeared/evolved in the genetic algorithm. NOTT-101/OEt is marked. (d) A comparison between simulated and experimental isotherms for NOTT-101/OEt. The experimental isotherms were scaled to take into account the 92 % pore activation of the synthesised NOTT-101/OEt sample. The inset shows the structure of NOTT-101/OEt with pores represented by dark spheres. Figure from Sci. Adv., 2016, 2, 10, e1600909, under Creative Commons Attribution-NonCommercial 4.0 International License https://creativecommons.org/licenses/by-nc/4.0/.

FIG. 13.

Screenshot of an article entry[293] available in the NIST-ISODB, selected from results of a search using "NH4-ZSM-5" as the adsorbent material and "carbon dioxide" as the adsorbate gas. Metadata about the experiments described in the article are shown in the upper half of the screenshot. The isotherm visualisation widget shows four separate isotherms from the source article plotted together.
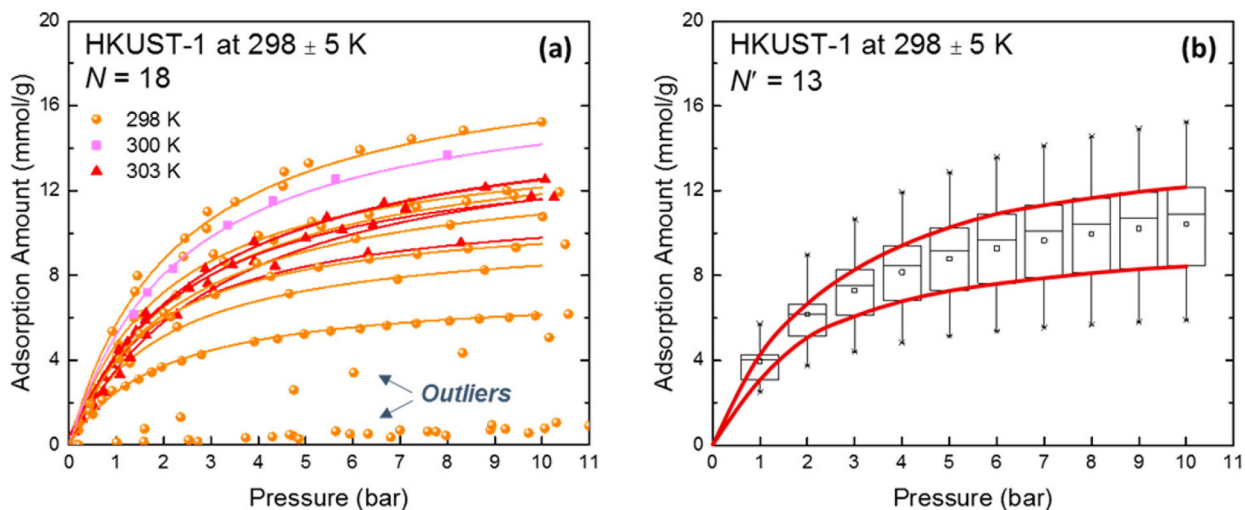
FIG. 14.

On the reproducibility of CO_2 adsorption isotherms in HKUST-1 [299]. (a) Experimental isotherms of carbon dioxide adsorption in HKUST-1 at temperatures of 298 K ± 5 K, as collected and organised by Park et al. [299] from the NIST-ISODB, to show the variability of laboratory measurements for that particular adsorption experiment. (b) Box-and-whisker plot created from 13 (out of 18) non-outlier isotherms in panel (a). Boxes represent the upper and lower quartile, the median is indicated by the straight line, and the small square is the mean. Whiskers represent 1.5 times the interquartile range. Reprinted with permission from Chem. Mater. 2017, 29, 24, 10487–10495. Copyright 2017 American Chemical Society.
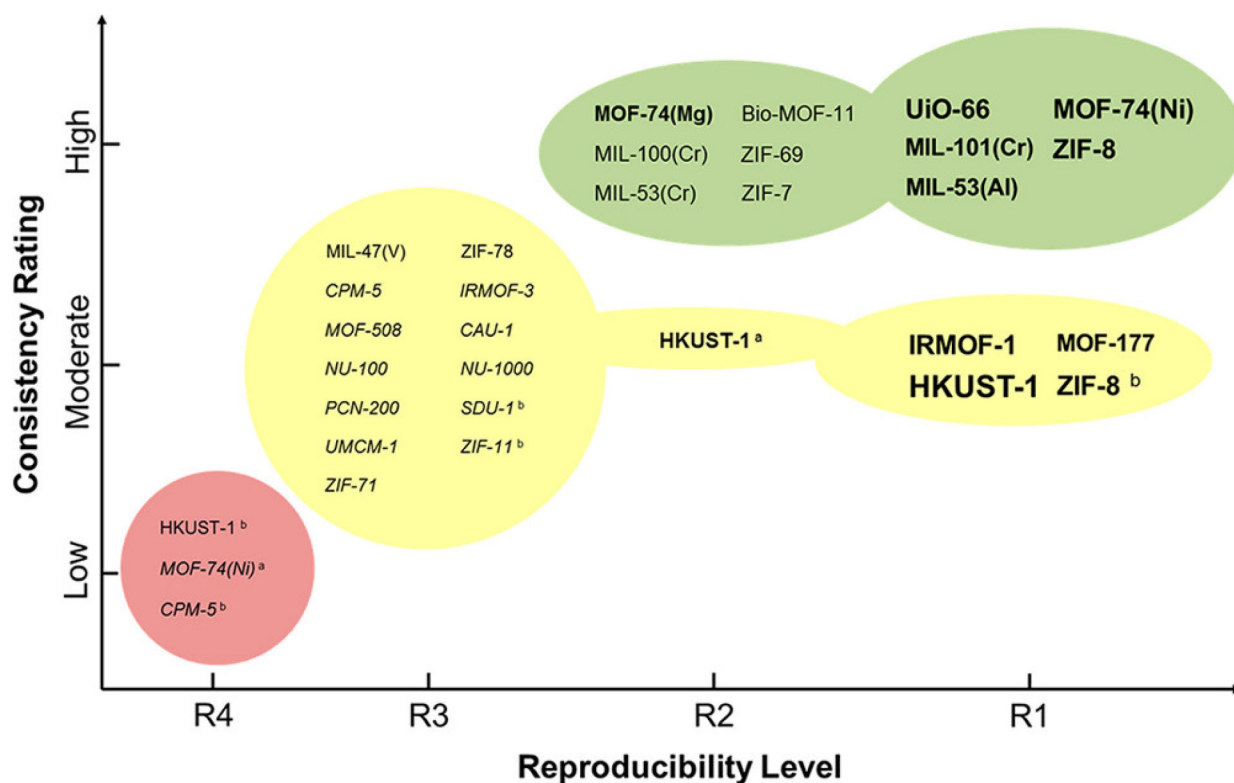
FIG. 15.

Graphical summary of reproducibility and consistency results from Park et al. for adsorption isotherms of carbon dioxide on various MOFs[299]. On the x-axis, reproducibility increases (according to Park et al.'s metrics) from left to right. Font size is indicative of the number of independent isotherms available for the noted material and bolding/italicisation identifies outlier levels. The reader may consult Ref. 299 for full description of the reproducibility level, consistency rating, and outlier levels. Reprinted with permission from Chem. Mater. 2017, 29, 24, 10487–10495. Copyright 2017 American Chemical Society.