

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

Arni Sturluson,[†] Melanie T. Huynh,[†] Alec R. Kaija,[‡] Caleb Laird,[†] Sunghyun Yoon,[¶] Feier Hou,[§] Zhenxing Feng,[†] Christopher E. Wilmer,^{*,‡} Yamil J. Colón,^{*,||} Yongchul G. Chung,^{*,¶} Daniel W. Siderius,^{*,⊥} and Cory M. Simon^{*,†}

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

[‡]*University of Pittsburgh. Department of Chemical and Petroleum Engineering, Pittsburgh, PA, USA*

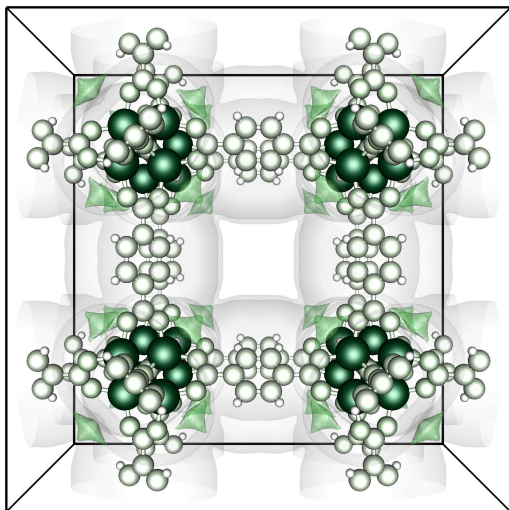
[¶]*Pusan National University. School of Chemical and Biomolecular Engineering, Pusan National University, Busan, Korea (South)*

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

^{||}*University of Notre Dame. Department of Chemical and Biomolecular Engineering, University of Notre Dame, Notre Dame, IN, USA*

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

E-mail: wilmer@pitt.edu; ycolon@nd.edu; drygchung@gmail.com; daniel.siderius@nist.gov; Cory.Simon@oregonstate.edu



Abstract

Metal-organic frameworks (MOFs) are highly tunable, 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 routinize the computational discovery of MOFs for adsorption-based engineering applications.

1 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 harbor nano-sized pores and enormous internal surface areas ($> 7000 \text{ m}^2/\text{g}$).² Their consequent gas adsorption properties lend them applications in storing,³ separating,⁴ and sensing⁵ 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 (Nu-Mat Technologies) and carbon dioxide capture in submarines (Mosaic Materials).

MOFs, heralded as "designer materials",⁸ are synthesized 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 Fig. 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 arduously¹¹) finding the synthetic conditions to yield a pre-determined, extended

network structure is coined as *reticular chemistry*.^{12,13}

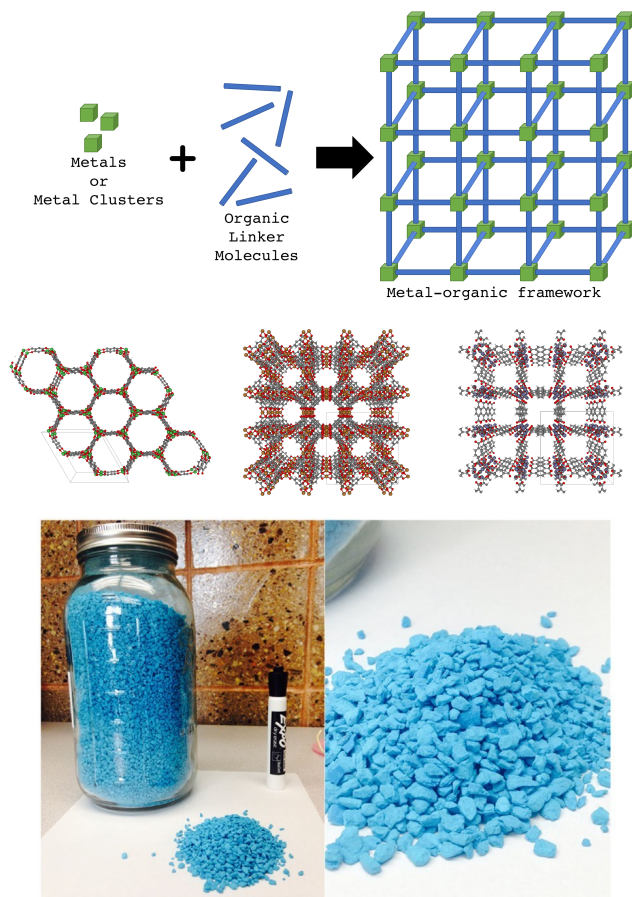


Figure 1: The chemistry of metal-organic frameworks (MOFs) is modular and highly tunable, affording a vast chemical space of crystal structures in which to search for materials exhibiting an optimal adsorption property. (top) MOFs are composed of metal nodes or clusters coordinated to organic linker 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-1¹² are shown as examples. (bottom) A sample of HKUST-1¹⁶ from Ref. 17. is shown to illustrate how a MOF appears from the naked eye.

For adsorption-based engineering applications, a coveted aim beyond reticular chemistry is to specify a desired (optimal) adsorption property, then synthesize 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)*. 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 synthesized, which MOF exhibits the maximal methane usable capacity. Clearly, it is impractically time- and resource-intensive to synthesize, characterize, 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 modeling and simulation in enabling, to an increasing extent, the identification of an existing MOF[†] 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

*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).

[†]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.

for gas storage, separation, and sensing.

We begin in Sec. 2 by briefly describing methodologies to computationally screen MOF structures for gas adsorption applications. We then discuss in Sec. 3 how MOF crystal structures are determined experimentally from X-ray diffraction (XRD) studies. In Sec. 4, we describe how, owing to artifacts 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. 5, 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. 6, we survey high-throughput computational screenings of MOF crystal structures for gas adsorption and separation applications that have directly motivated the experimental synthesis and characterization 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. 7 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.

2 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.

2.1 Energetic modeling

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, clas-

sical force fields, whose interatomic potential parameters are fit to experimental data or *ab initio* 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 DREIDING²⁹ 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 vapor-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 modeled (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 CO₂,³¹ which has polar bonds, but not to CH₄, which lacks polar bonds³⁰). There exists a hierarchy of methods for assigning point charges to MOFs to model their electrostatic interaction with an adsorbate molecule that possesses polar bonds³² (see Sec. 9.2.2), and a database of experimental MOF crystal structures with high-quality point charges assigned can be downloaded (see Sec. 5.1.1).

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 Cu₂ paddlewheel and the center 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 C₂H₄,^{36,37} H₂,³⁸ C₂H₂,^{39,40} CO₂,^{39,41,42} CH₄,^{41,43} and CO⁴⁴ interactions with unsaturated Cu within Cu₂ paddlewheels as well as for CO₂, CH₄, and H₂O 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) simulations⁴⁸ 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

2.2 Structural modeling

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. 9.1), 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 ensemble⁵⁰) 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 intra-

host force fields for MOFs (e.g., UFF4MOF,⁵¹ MOF-FF,⁵² BTW-FF⁵³) 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 Fig. 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 minimizing the potential energy of the assembled structure over the atomic positions using a classical force field or electronic structure calculations.

2.3 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.

2.3.1 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 normalization factor).⁶⁴ As such, it is useful for sampling statistical mechanical ensembles⁶⁵ 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 (μVT) 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:

$$p_\nu \propto e^{-\beta E_\nu + \beta \mu \cdot \mathbf{N}_\nu}, \quad (1)$$

with $\beta = 1/(k_B T)$, k_B the Boltzmann constant, E_ν the energy of microstate ν , and \mathbf{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 (μVT) 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 favorable 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 ensemble⁶⁶ was specifically devised for molecular simulations of phase coexistence of a fluid within a single simulation, with-

out 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 (NVT) 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.

2.3.2 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[‡].

As is done in MC, classical MD simulations of a gas in a MOF utilize 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 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 extending the system Hamiltonian⁴⁸ 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 a reservoir of guest molecules at 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 isothermal-

isobaric 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.

2.4 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, voilà, shortlist the top few for experimental testing[§]. This is the obvious value of high-throughput computational screenings, and it is predicated on sufficiently 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 triv-

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

[§]Typically, as illustrated in our survey in Sec. 6, some human judgment on e.g., ease of synthesis is also exercised in further shortlisting materials.

ialization 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 CO₂/N₂ selectivities for flue gas separations harbor pore sizes no larger than ≈ 5 Å to 6 Å. Often, experimental data is too sparse to recognize 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, validity of conclusions of this nature are predicated on the accuracy of the force field and assumptions built into the molecular models. For example, that work⁸⁹ 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 a MC simulation via a configurational bias algorithm,⁹⁰ modeling polarizability of carbon dioxide by open-metal sites in MOFs,⁹¹ modeling water adsorption in MOFs, which requires many MC samples⁹²), 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.

Statistical machine learning. A statistical machine learning⁹³ (regression or classification) model can be trained to predict adsorption

properties using geometric/structural⁹⁴ and/or chemical⁹⁵ 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 CO₂ 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) modeling 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 CH₄ and CO₂ uptakes. Motivated by the latter correlation, they trained support vector machines and multilinear regression models to predict simulated CH₄, CO₂, and N₂ 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 CO₂ adsorption, viewing computational screening as a classification problem. Next, Fernandez et al.¹⁰⁴ similarly binarized simulated low-pressure CO₂ and N₂ adsorption in hypothetical MOFs as high or low, then trained a variety of machine learning classifiers to discriminate between MOFs with high/low CO₂ and N₂ 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 to 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 H₂ and CO₂

adsorption measurements (target variables) in 100 MOFs from the literature. The authors then encoded the identity of the linker, metal center, 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 CO₂ and H₂ 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 CO₂/N₂ selectivity of a given parent MOF will increase (or not) when functionalized. The authors also trained several machine learning regressors (decision tree, random forest, support vector machine, neural network, gradient boosted machine) to predict the simulated CO₂/N₂ selectivity of a given MOF; the gradient boosted machine ranked the materials according to CO₂/N₂ 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 H₂ with the metalated catechol was allowed to artificially vary to account for variations in chemistry without 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 H₂ 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 H₂ 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 H₂ adsorption in a MOF by computing the potential energy of a H₂ 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-regularized multilinear regression model to predict the H₂ usable capacity at 77 K from the descriptor, using molecular simulations of H₂ 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 H₂ usable capacity. Given the predictiveness of the model, the authors proceeded to screen 54 776 MOF structures for H₂ storage and delivery with the trained model, eventually pinpointing MFU-4l as a promising candidate (see Sec. 6.1.2).

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}

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 algorithm¹¹² 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 CO₂ 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 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 CO₂ 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 optimize the functionalization of their linkers for postcombustion CO₂ capture.

A lesson from molecular biology

CRISPR-Cas9 technology¹¹⁶ is revolutionizing 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 recognized 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 program to analyze published genomes. Particularly, they found clusters of multiple copies of roughly palindromic sequence base pairs (bp), ≈ 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 species¹²¹) 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 ≈ 4500 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.

Lander¹¹⁹ 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. 6 will demonstrate the impact of open, computation-ready databases of MOF structures on the discovery of performant MOFs for adsorption-based engineering applications.

3 Determining crystal structures of MOFs

Single crystal X-ray diffraction (SC-XRD) is the most utilized 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 minimize 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:

$$R := \frac{\sum |I_{\text{exp}} - I_{\text{sim}}|}{\sum |I_{\text{exp}}|}, \quad (2)$$

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 twinning¹³⁰ and pseudo-symmetry.¹³¹ Relatedly, interpenetration of nets in MOFs¹³² 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 synthesized 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 SQUEEZE¹³⁴) is an effec-

tive 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 synthesize 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 Fig. 2b for an example experimental PXRD pattern, of IRMOF-74-VII, whose crystal structure is in Fig. 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 rep-

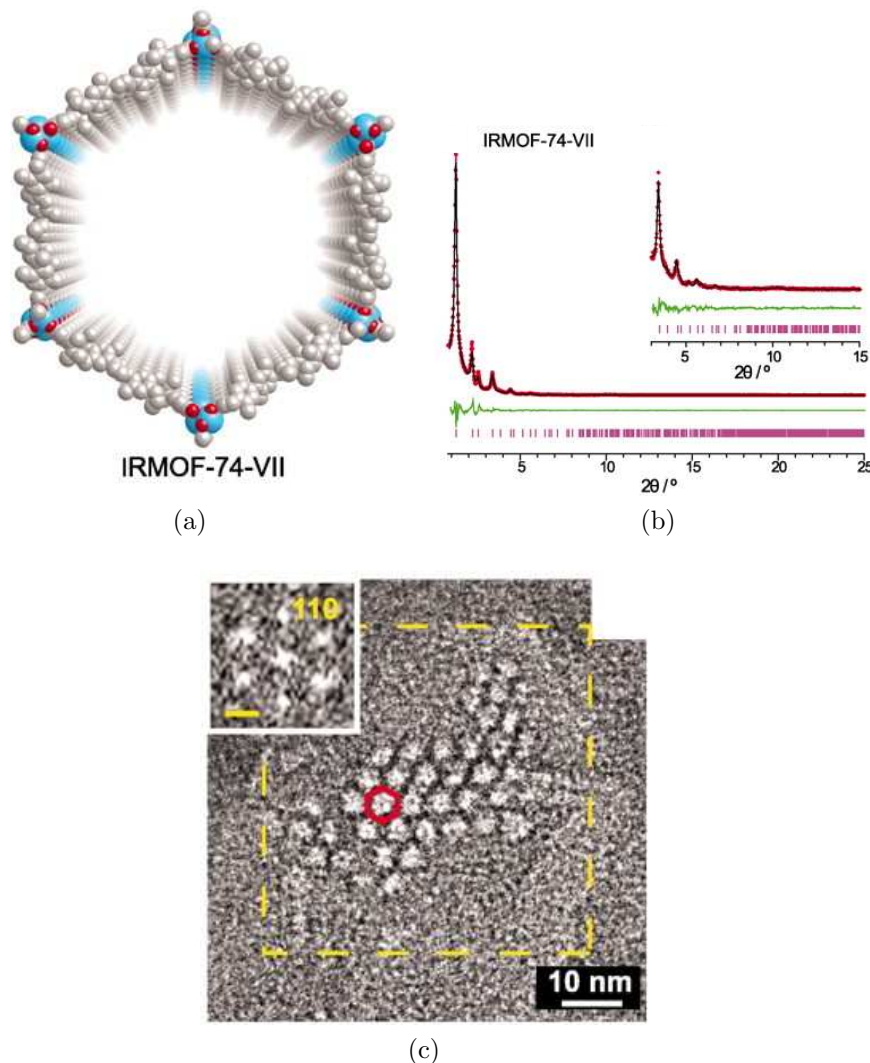


Figure 2: The crystal structure of IRMOF-74-VII was determined by PXRD assisted by computational modeling. (a) The suspected and confirmed crystal structure of IRMOF-74-VII. Isorecticular to Mg-MOF-74, organic linkers with seven phenylene rings are connected to magnesium oxide clusters, forming one-dimensional 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.

resentation 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 synthesized 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 synthesized 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 Fig. 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}

4 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 organization. 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 program,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*.

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

Because each CSD entry is not labeled 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 MOF¹⁰ using seven different chemical bond criteria (see Fig. 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. 5.1) 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.

4.2 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 artifact 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

4.3 Problem 2: structural disorder and missing hydrogen atoms

The second problem with many MOF structures in the CSD is an artifact of XRD; many structures are incomplete (e.g., missing hydrogen atoms) and chemically invalid (e.g., exhibit disorder). As Sec. 3 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 Fig. 3 for examples. The disorder must be repaired and missing atoms must be added in appropriate orientations to render a MOF structure computation-ready.

5 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 sol-

vent (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. 4, 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 artifacts 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 programs 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 analyze and rationalize the distribution of net topologies among MOFs synthesized 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 synthesized 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 com-

putationally 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[¶], 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

[¶]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}

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.

5.1 The (open) CoRE MOF database

In 2014, Chung et al. released a free and open, computation-ready, experimental (CoRE) MOF database^{9||}. The workflow to construct the CoRE MOF database is shown in Fig. 3a. The authors first searched the CSD for potential MOF crystal structures using the CCDC Conquest program; 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 structures were analyzed 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 Fig. 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

^{||}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).

of 0.4 Å. Bound solvents are usually coordinated to the open metal sites in MOFs, such as the copper site in HKUST-1. Fig. 3b illustrates the removal of free and bound solvents from a MOF. To remove solvents bound to the open metal sites, the algorithm first constructed an adjacency matrix of an input structure. The adjacency matrix represents the bonding network of atoms in the structure 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 lone oxygen atom 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 MOF. If the number of edge-disjoint clusters changed (indicating a solvent-MOF bond was eliminated), then the matrix element was left modified and the subgraph 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 (i.e., the bond is reintroduced). This process was repeated for all oxygen atoms connected to a metal atom in the structure. Also, the symmetry-related copies of atoms arising from disorder (see Fig. 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 re-

move solvent molecules (which are often critical in maintaining the structural integrity of the MOF) to mimic experimental activation. A set of 4764 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 July 12, 2019) since its publication in 2014 and has enabled several computation-inspired MOF discoveries, as outlined in our survey in Sec. 6. Its impact was predicated on making the structures freely and easily accessible via Github.

5.1.1 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 modeled 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 CO₂³¹), thereby completing the description of the electrostatic interaction of the adsorbate with the MOF. Quick/cheap charge equilibration [Qeq¹⁷⁵] methods and its variants [EEM,¹⁷⁶ PQeq,¹⁷⁷ SQE,¹⁷⁸ SCQeq,¹⁷⁹ EQeq,¹⁸⁰ MEPO-Qeq,¹⁸¹ FC-Qeq,¹⁸² I-Qeq,¹⁸² EQeq+C,¹⁸³ and SQE-MEPO¹⁸⁴] 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) method¹⁸⁵ 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) method¹⁸⁶ 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

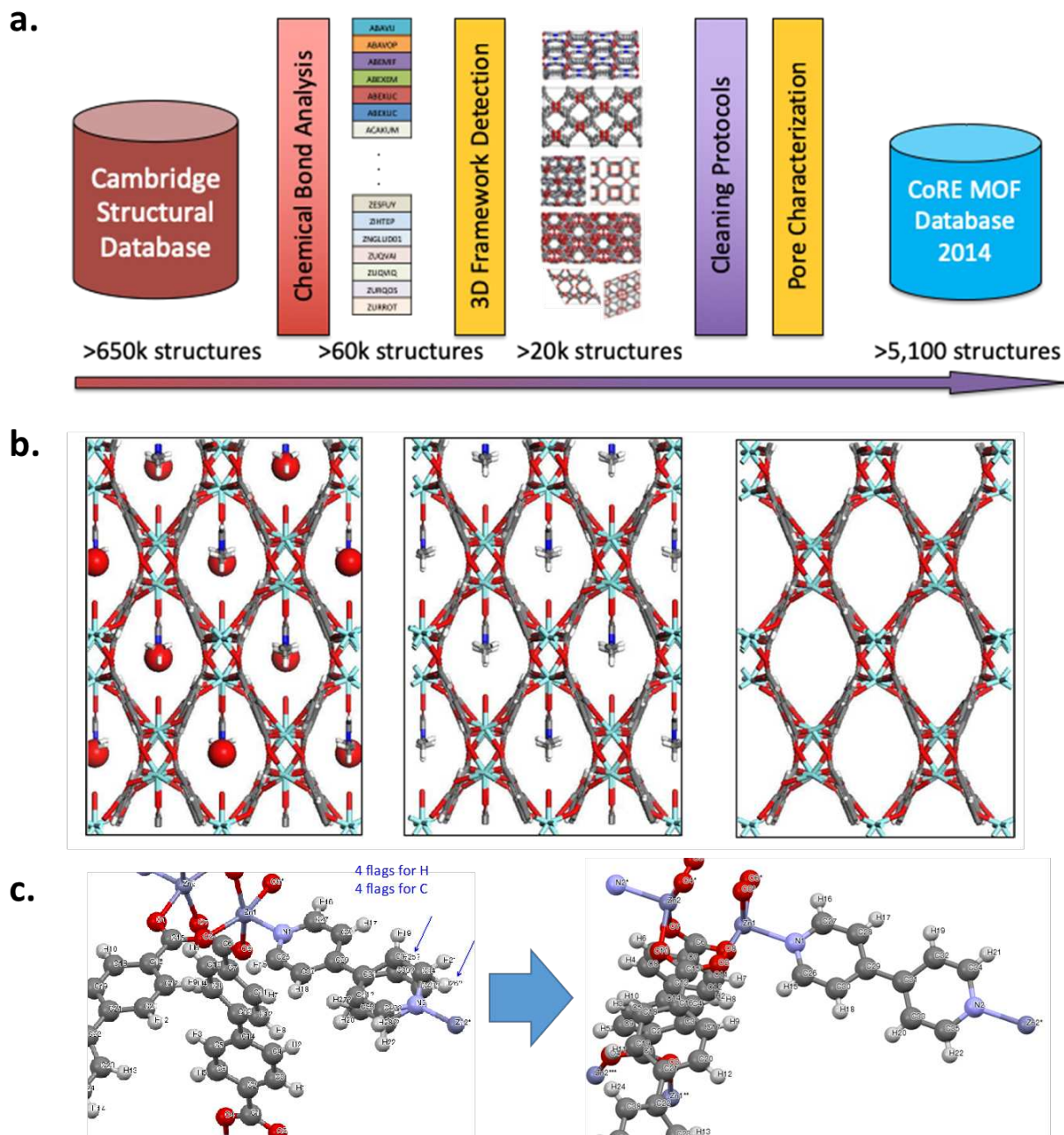


Figure 3: Curating the computation-ready, experimental (CoRE) database of MOFs. (a) The workflow of the development of CoRE MOF Database 2014. (b) Automatically removing solvent from MOFs deposited in the CSD. Left to right, as-is structure, free solvent removed, free and bound solvent removed (CSD refcode: 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.

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 cal-

culated 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 mem-

ory 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.

5.1.2 DFT-energy-minimized CoRE MOFs

The crystal structure of a newly synthesized 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 destabilized 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 minimize the potential energy (by changing the atomic coordinates) of the structures of 879 CoRE MOF structures. Indeed, several DFT-optimized CoRE MOFs showed significant changes in the structure and simulated adsorption compared to their cognate unoptimized structures.⁶⁰ These DFT-optimized structures are likely more reliable than the vanilla CoRE MOF structures⁹ because (a) they maintained their structural integrity (i.e., still have a void) after DFT-optimization, (b) the DFT-optimization accounted for any change in structure that would result from the solvent removal, and (c) the DFT-optimized 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 2612 optimization-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.

5.2 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 Fig. 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 m²/g and 2000 m²/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

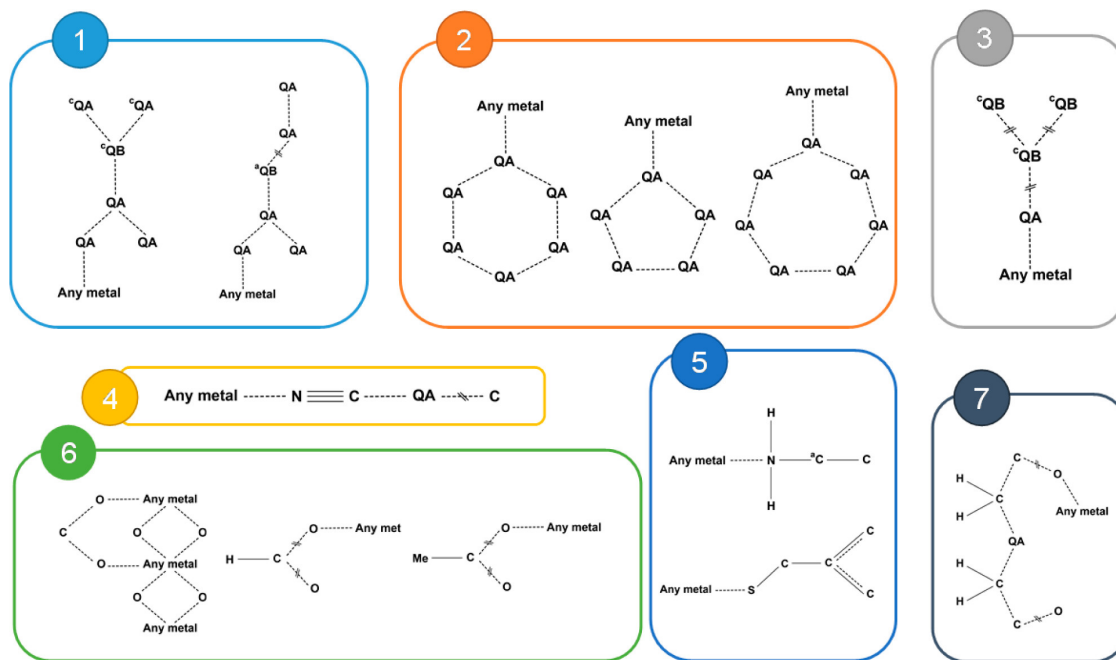


Figure 4: The seven chemical bond criteria used by Moghadam and co-workers¹⁰ to search for MOFs in the CSD. Here QA = O, N, P, C, B, S, QB = N, P, B, S, C, ME = methyl group. The superscripts *c* and *a* denote cyclic and acyclic, respectively. Reprinted with permission from Chem. Mater. 2017, 29, 7, 2618–2625. Copyright 2019 American Chemical Society.

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 subset¹⁰ 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-optimized structures and DFT-assigned point charges are not available for the CSD MOF subset (unlike for the CoRE MOF database^{60,187}).

5.3 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 with-

out bound solvents, and the CSD-maintained subset¹⁰ 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 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.** (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

**<http://dx.doi.org/10.11578/1118280>

MOF is deposited in the CSD;¹⁰ this issue is addressed with the CSD-maintained subset¹⁰ which is updated every quarter.

Altintas et al.¹⁹¹ comprehensively compared the CoRE⁹ and CSD-maintained, non-disordered¹⁰ MOF databases and found several discrepancies and shortcomings in both. The authors assembled 3 490 MOFs (characterized 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.

5.4 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 molecules¹⁹⁵ have emerged as well. Tong et al.^{196,197} prepared a database of 280 disorder- and solvent-free, experimentally synthesized 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 cav-

ities of porous cage molecules into latent vector representations¹⁹⁹ 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.

6 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.

6.1 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 H₂ 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.

6.1.1 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 sulfur 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 fracturing²⁰⁷ 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 refueling stations; further, boil-off 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

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 Fig. 5). Building blocks could combine if the geometry and chemical composition at the point of connection was the same as in the synthesizable structure. Connections between building blocks were determined solely based on geometric rules; that is, the structures were not energy-minimized. “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-

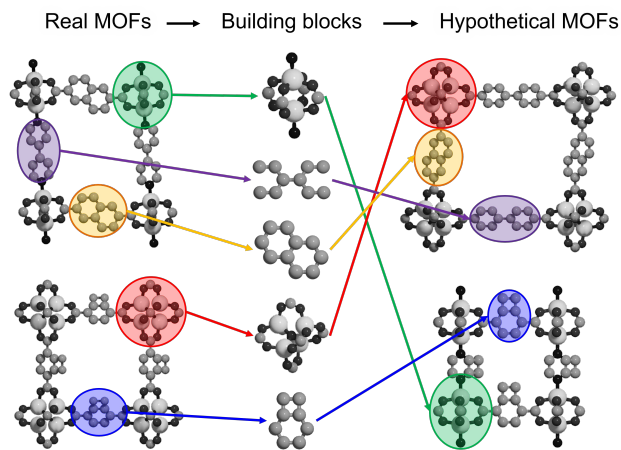


Figure 5: The algorithm used by Wilmer et al.²¹³ to construct hypothetical MOF structures first extracts building blocks from previously-known MOF structures. These building blocks are then combined according to their geometry to create hypothetical MOF structures.

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 $< 0.8 \text{ \AA}$ 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 synthesized and found to have an experimentally measured methane capacity $\approx 8 \%$ lower than the prediction.²¹³

In addition to identifying promising candidate structures for synthesis, this library of hypothetical structures provided insights into rela-

tionships 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 maximizing gravimetric surface area had been a common strategy for designing methane adsorbents, exceeding an optimal surface area ($\approx 2500 \text{ m}^2/\text{g}$ to $3000 \text{ m}^2/\text{g}$) 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 \AA and 8 \AA . 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 Fig. 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 pressurized to 3600 psi (248 bar).

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, $(\text{Zr}_6\text{O}_4)(\text{OH})_4(\text{CO}_2)_n$, with various building blocks (see Figs. 7a and 7b) to form MOFs in four network topologies, **fcu**, **ftw**, **scu** and **csq**. The structures were energy-minimized by using the UFF²⁸ 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 Fig. 7a was predicted to exhibit the highest methane usable capacity (197 L(STP)/L) and was coined NU-800. See Fig. 7c. As a consequence, NU-800 was synthesized, and its methane (also nitrogen, carbon dioxide and hydrogen) adsorption isotherms

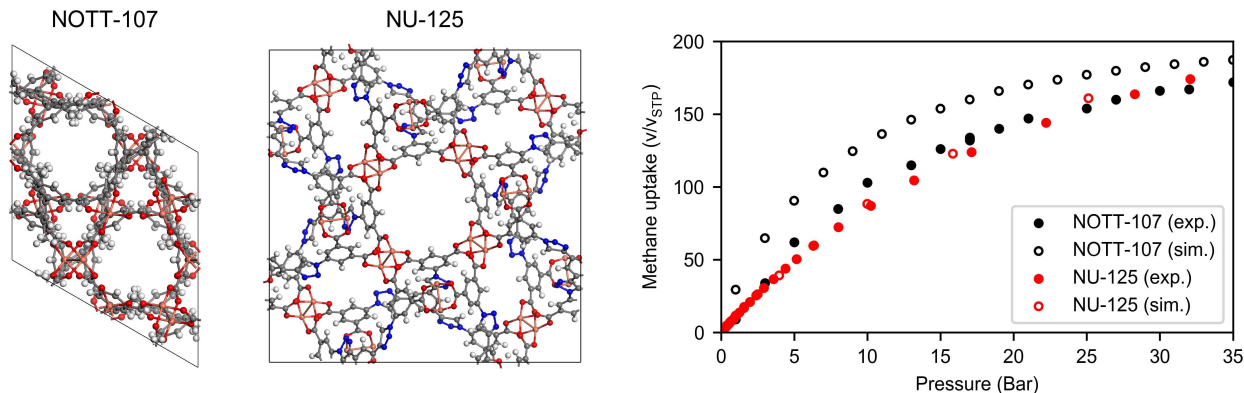


Figure 6: The MOFs NOTT-107 and NU-125 were both identified via high throughput computational screening studies^{213,217} as potentially promising materials for methane storage. NU-125, in particular, was pinpointed based on its void fraction of 0.8, which was found as optimal for methane storage from 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.

were measured and compared to the simulation (see Fig. 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.

6.1.2 Hydrogen storage and delivery

Hydrogen (H₂) 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 turbines²²¹ 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 °C²²⁴ 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

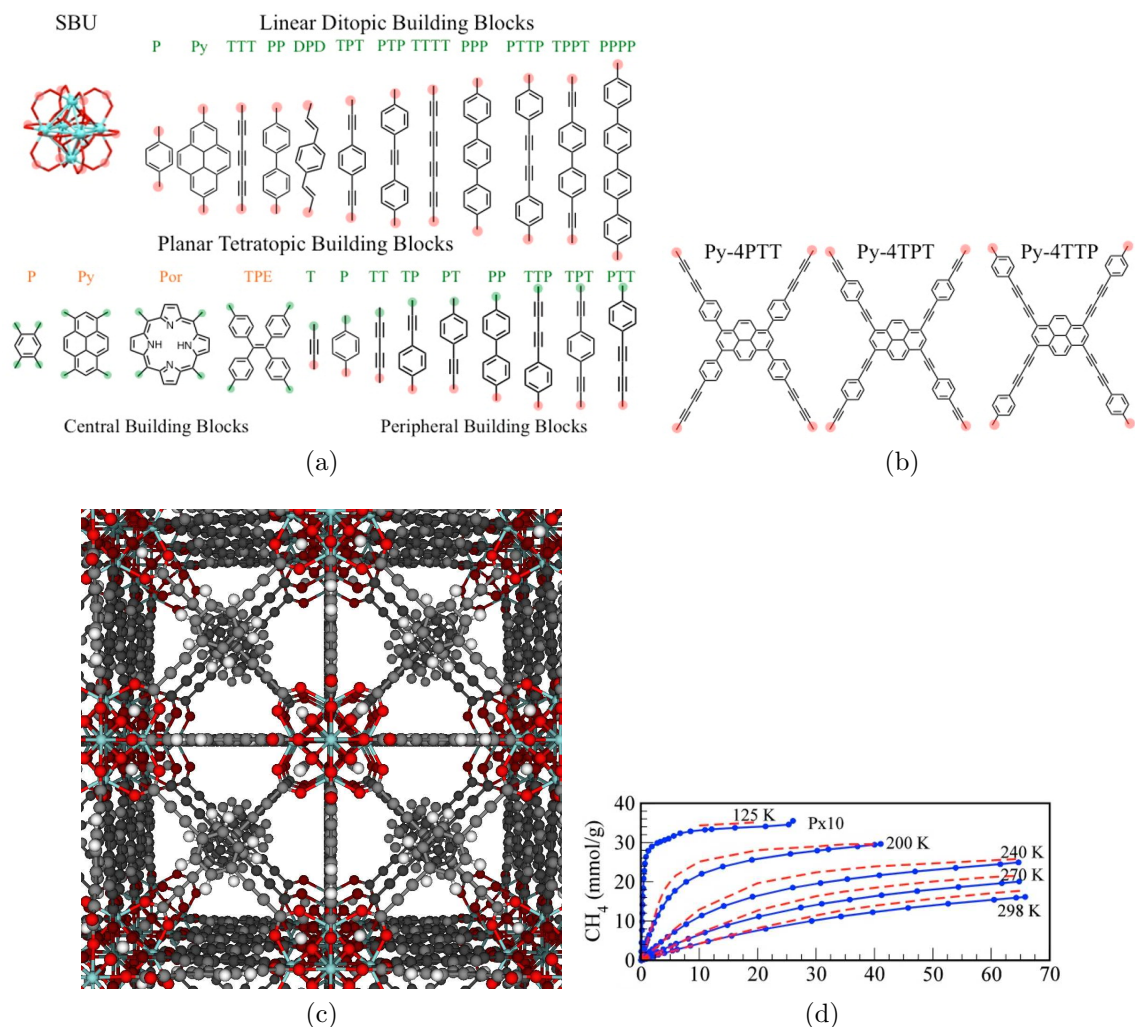


Figure 7: Gomez-Gualdron et al.²¹⁸ constructed a set of 204 zirconium-based hypothetical MOFs and screened them for storing and delivering methane, leading to the synthesis of NU-800. (a) The basic building blocks used to construct the hypothetical MOFs (connection points highlighted). The ditopic linkers 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.

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}

she-MOF-1 Gómez-Gualdrón and co-workers²²⁶ 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 struc-

tures, 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 synthesized, 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. 9.4).

IRMOF-20 Ahmed and co-workers²²⁷ 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 database⁹ 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 H₂/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 4073 m²/g vs. 3512 m²/g), rationalizing its greater gravimetric usable capacity of hydrogen under the Chahine rule.

MFU-4l Bucior et al.⁹⁸ screened the CSD MOF subset¹⁰ 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-regularized linear regression model to predict 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-4l²³⁰ (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 synthesized MFU-4l 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 Fig. 8. MFU-4l 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

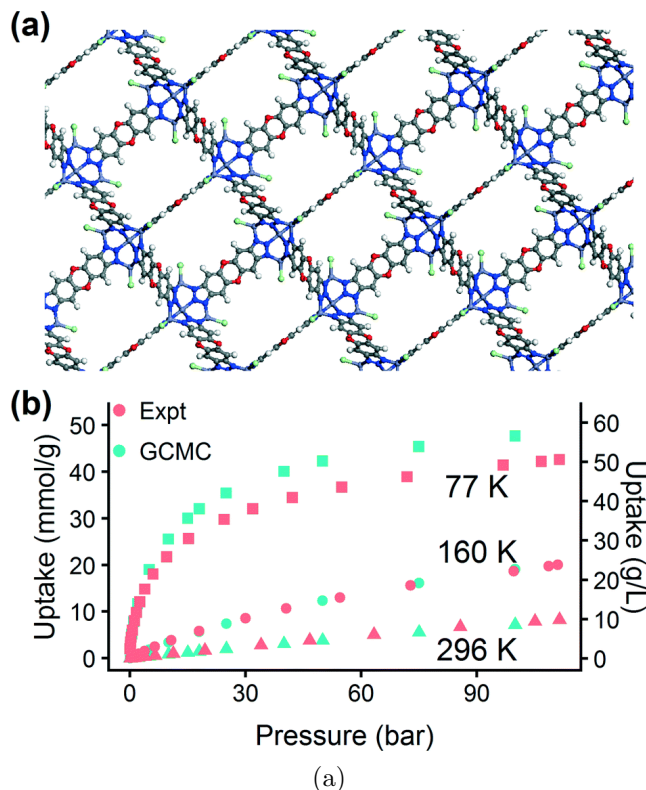


Figure 8: A machine-learning accelerated high-throughput screening⁹⁸ led to the identification of MFU-4l for hydrogen storage. (a) Crystal structure of MFU-4l. (b) Simulated and experimental H_2 adsorption isotherms of MFU-4l at different temperatures. Reproduced from Mol. Syst. Des. Eng. 2019, 4, 1, 162–174, with permission from The Royal Society of Chemistry.

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 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 MOF²³³ that was a variant of exper-

imentally 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 synthesize (motivated by high predicted H_2 capacity) a mixed-linker, non-interpenetrated, MOF, UMCM-9,²³⁶ isoreticular 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.

6.1.3 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.

UMCM-152 Moghadam et al.²⁴³ conducted a high-throughput computational screening of 2 392 previously synthesized 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 largest cavity diameters above 8 Å, void fractions larger than 0.7, and geometric surface areas larger than 2600 m²/g. See Fig 9a. The most promising MOF for oxygen delivery was UMCM-152 (CSD refcode: ANUGIA; see Fig. 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 O₂ 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 O₂ 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.

6.2 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 tunable pore shapes and surface chemistries of MOFs could enable molecular separations that were once infeasible.²⁴⁶ In

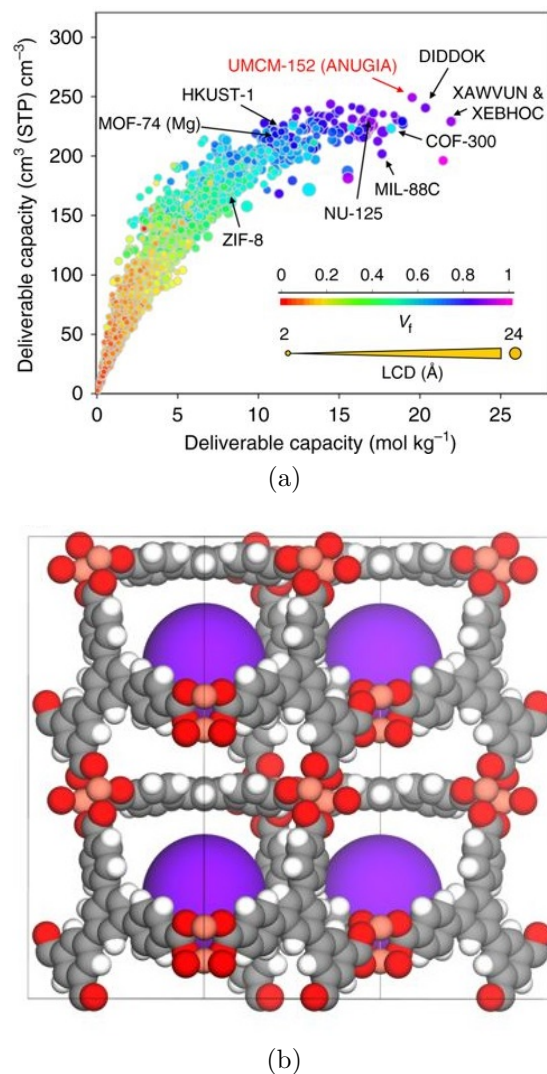


Figure 9: Computational identification of UMCM-125 (CSD refcode: ANUGIA) for oxygen storage and delivery.²⁴³ (a) Computational screening data; each point represents a MOF. The volumetric oxygen deliverable (usable) capacity is largest for MOFs with a void fraction (V_f) above 0.7 and a largest cavity diameter above 7.5 Å to 8 Å. Common MOFs are highlighted, including the MOF predicted to have the largest O₂ 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/>

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.

6.2.1 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 utilizing 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 days²⁵¹), recovered from the first process could be sold in the market for use in medicine, ion propulsion, lighting, and insulation.²⁵²

SBMOF-1 Searching for a MOF harboring 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 synthesized,²⁵⁴ but

not characterized for Xe/Kr separations. Motivated by the computational prediction, Banerjee et al.²⁵³ synthesized 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 Fig. 10. We duly note that several computational screenings of MOFs for xenon/krypton separations have been carried out at different conditions prior to release of the CoRE MOF database.^{97,255,256}

6.2.2 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 behavioral 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 sulfide (DES). See Fig. 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

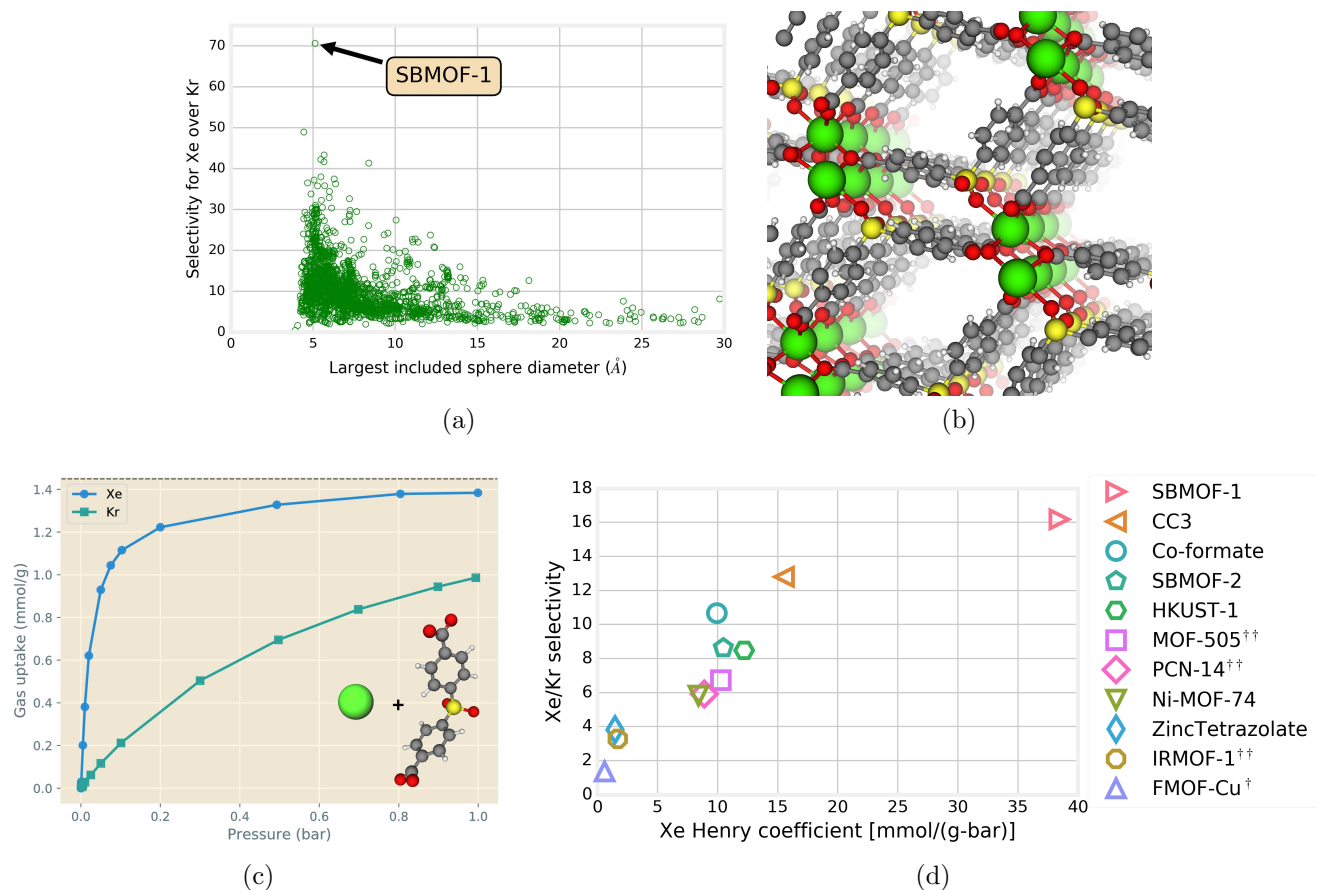


Figure 10: SBMOF-1 for Xe/Kr separations.²⁵³ (a) Simulated Xe/Kr selectivity against pore size; each point represents a CoRE MOF; SBMOF-1 is marked. (b) The crystal structure of SBMOF-1 (CSD refcode: KAXQIL) exhibits 1D channels that form well-defined pockets in which xenon can adsorb commensurately 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 harbors an equilibrium Xe/Kr selectivity of 16. Inset shows metal (Ca) and V-shaped organic ligand used to synthesize 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/>.

(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.

Ni₃(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 2932 MOFs from the CoRE database with point charges assigned¹⁸⁷ to account for CWA-MOF electrostatic interactions during the simulations. First, a subset of 1275 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 1647 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 sulfide (DES), but less so for sarin and somin and their surrogates. MOFs displaying the highest Henry coefficients of CWAs tended to harbor 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 short-list 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 $\text{Ni}_3(\text{BTP})_2$ (CSD refcode: UTEWOG; see Fig. 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 $\text{Ni}_3(\text{BTP})_2$, flowing nitrogen gas with 80 % relative humidity (water) and 1 ppm diethyl sulfide (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. Fig. 11c shows that $\text{Ni}_3(\text{BTP})_2$ 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 $\text{Ni}_3(\text{BTP})_2$ were a filter in a gas mask within an environment of 1 ppm mustard gas and 80 % relative humidity. Further, adsorp-

tion in $\text{Ni}_3(\text{BTP})_2$ 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 $\text{Ni}_3(\text{BTP})_2$ as readily adsorbing mustard gas in the presence of humidity and demonstrated capture of its surrogate molecule DES through column breakthrough experiments.

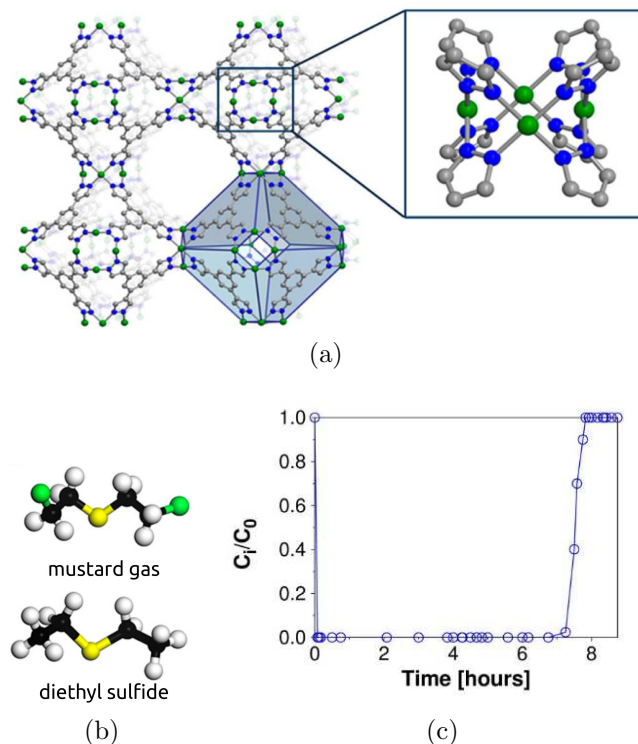


Figure 11: Computational identification of $\text{Ni}_3(\text{BTP})_2$ for chemical warfare agent capture.²⁶³ (a) The structure of $\text{Ni}_3(\text{BTP})_2$ (CSD refcode: UTEWOG). (b) Mustard gas and its surrogate diethyl sulfide (DES). (c) DES breakthrough curve. A N_2 gas stream with 80 % relative humidity and 1 ppm DES at 298 K is passed, at 20 mL/min, through a column packed with 150 mg of $\text{Ni}_3(\text{BTP})_2$. Shown is a measurement, via gas chromatography, of normalized DES concentration at the outlet of the column. Reprinted with permission from Chem. Mater. 2018, 30, 14, 4571–4579. Copyright 2019 American Chemical Society

6.3 Carbon Dioxide capture

Fossil fuels have been one of the main energy sources since the 20th century in the United States²⁶⁶ and are the dominant source of CO₂ emissions into the atmosphere worldwide.²⁶⁷ The increasing CO₂ 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 21st century if CO₂ emissions are not mitigated from their current trajectory.²⁶⁸ In the United States, roughly 33 % of energy-related CO₂ emissions are directly tied to the burning of fossil fuels to generate electricity.²⁶⁹ Significant effort is devoted to develop efficient technologies to capture CO₂ from the flue gas of fossil fuel-fired power plants before it is released into the atmosphere (post-combustion capture). The captured CO₂ can then be sequestered in a geological reservoir.²⁷⁰ The technology currently used to capture CO₂, such as absorbing CO₂ with aqueous alkanolamine absorbents, has not proven to be energy efficient and would reduce 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 CO₂.²⁷¹ A more recent development sees MOFs selectively capturing CO₂ from flue gas, which could reduce the energy penalty significantly.²⁷² The removal of CO₂ from the flue gas is called post-combustion carbon capture. An alternative is to capture CO₂ 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 H₂ and CO (steam reforming). The CO is then reacted with water (water-gas shift reaction) to produce CO₂ and H₂. CO₂ can then be separated from H₂ (rather than from the flue gases) and sequestered, and the pure H₂ can be burned to produce electricity (without CO₂ emissions). An advantage to employing MOFs for pre-combustion carbon capture is that the CO₂/H₂ mixture is at high pressure already, and a pressure swing down to atmospheric pressure can readily push CO₂ out and regenerate the MOF.

NOTT-101/OEt Chung et al.¹¹⁴ used a genetic algorithm to find a performant MOF for pre-combustion CO₂ capture. To apply genetic algorithms to search for MOFs with good selectivity for CO₂ over H₂ and a high CO₂ 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 \approx 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 Fig. 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 optimize. 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 optimizes 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: CO₂ usable capacity, selectivity for CO₂ over H₂, 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. analyzed the MOFs in the last generations and found that only a few organic linker genes and inorganic node genes were represented (see Fig. 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 CO₂ over H₂, a high CO₂ usable capacity, and APS

(see Fig. 12c). Because of its high fitness, as well as previous experience in MOF synthesis, Chung et al. synthesized NOTT-101/OEt and measured both CO₂ and H₂ adsorption isotherms (see Fig. 12d). Good agreement was found with simulated isotherms. NOTT-101/OEt had a usable capacity of 3.8 mmol/g and a CO₂/H₂ selectivity of 60. Comparatively, other notable MOFs studied for CO₂/H₂ separations in pre-combustion are Mg-MOF-74, with a 2.6 mmol/g usable capacity and a 365 CO₂/H₂ selectivity,²⁷⁴ and Cu-BTtri, with a 3.7 mmol/g usable capacity and a 20 CO₂/H₂ 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 CO₂/H₂ selectivity. They found that, for precombustion CO₂/H₂ separation, the CO₂/H₂ selectivity of the material needs to be greater than 30 to meet the process objective of 99.999 % H₂ purity. On the basis of process modeling, Chung et al. concluded that, while NOTT-101/OEt meets the process requirement, another porous material, Cu-BTtri, cannot be used for pre-combustion carbon capture application because the CO₂/H₂ selectivity is not high enough to generate the high purity H₂ stream required for subsequent energy generation.

6.3.1 Xylene enrichment

Mixtures of *ortho*-, *para*-, and *meta*-xylene and ethylbenzene (C₈ aromatics) are obtained from the catalytic reforming of crude oil.^{276,277} The *p*-xylene isomer is the most valuable component of the C₈ aromatic mixture. It is oxidized to yield terephthalic acid or dimethyl terephthalate, both feedstocks for the production of polyethylene terephthalate (PET),²⁷⁶ which is widely used for synthetic fibers (polyester) and bottles.²⁴⁵ Pure *o*-xylene is also valuable to synthesize phthalic anhydride, a precursor for pro-

duction of plasticizers.²⁷⁶ However, mixtures of C₈ aromatics are very challenging to separate because of their similar shapes, boiling points, and polarities.²⁷⁸ The two incumbent industrial processes to separate C₈ aromatic mixtures are crystallization and, more often, selective adsorption onto a solid-state material.^{276,277,279} MOFs are promising adsorbent materials for separating C₈ aromatics more effectively than zeolites, which are currently used in a simulated moving bed process to obtain high-purity *p*-xylene.²⁸⁰

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-NH₂ 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-NH₂ 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-functionalization, MOF-48 exhibits *p*-xylene selectivity while MIL-47 exhibits *o*-xylene selectivity, emphasizing that subtle distinctions in pore features can lead to selectivity switching for these similarly-shaped C₈ aromatic isomers.

A lesson from machine learning

The field of machine learning aims to leverage data to train mathematical models

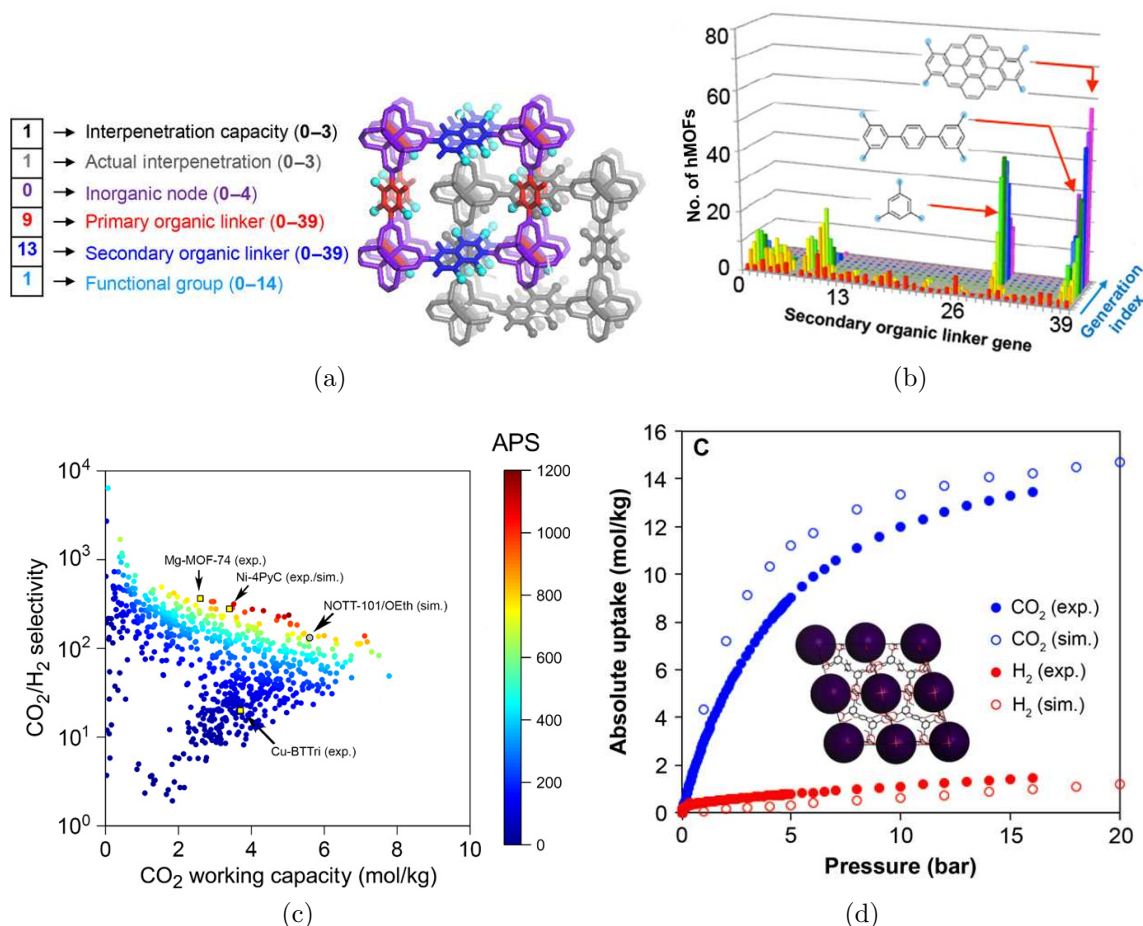


Figure 12: The discovery of NOTT-101/OEt was guided using a genetic algorithm in conjunction with molecular simulations to evaluate fitness.¹¹⁴ (a) A chromosome describes a MOF structure with 6 integers (called genes), each representing a different property of the MOF. The chromosomes define a chemical space that will be explored in the genetic algorithm. (b) A visualization 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 synthesized 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/>.

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) standardize 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 digits²⁸² contains ca. 70 000 labeled, 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 % error²⁸³). As a more challenging arena, CIFAR-10 and -100²⁸⁴ data sets each consist of 60 000 colored, labeled 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 labeled 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 in-

terest 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 crowdsourced 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).

7 NIST Resources for Adsorption Measurements

In 2014, the National Institute of Standards and Technology (NIST) officially launched a program 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 programs related to research, development, and engineering of adsorbent materials by developing testing procedures, disseminating reference measurements, and providing impartial testing and

characterization 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 modeling 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 Materials²⁹⁴ (NIST-ISODB)^{††}. 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. Fig. 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 scheme²⁹⁵) and adsorbent materials (see following paragraph), an application programming interface (API) for accessing the database contents (in particular, 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 synthesis¹⁶ is now more commonly known as CuBTC (short for copper benzene-1,3,5-tricarboxylate), but is also known as MOF-199²⁹⁶ and is sold by BASF under the name BasoliteTM C300.²⁹⁷ Multiplicity of names, such as for CuBTC, hinders effective searches in NIST-ISODB. To solve this problem, the NIST Registry of Adsorbent Materials²⁹⁸ (NIST-MATDB)^{‡‡} was released in 2017 as a companion to NIST-ISODB. The overarching purpose of NIST-MATDB is to

^{††}<https://adsorption.nist.gov/isodb>

^{‡‡}<https://adsorption.nist.gov/matdb>

DOI: 10.1007/s10450-018-9958-x

Title

A Reference High-Pressure CO₂ Adsorption Isotherm for Ammonium ZSM-5 Zeolite: Results of an Interlaboratory Study

Authors

Huong Giang T Nguyen, [Laura Espinal](#), Roger D van Zee, Matthias Thommes, Blaza Toman, [M Sterlin Leo Hudson](#), Enzo Mangano, Stefano Brandani, [Darren P Broom](#), Michael J Benham, Katie A. Cychosz, [Pieter Bertier](#), F Yang, Bernhard M. Krooss, [Rebecca L Siegelman](#), Masako Hakuman, Kazuyuki Nakai, Armin Ebner, Lutfi Erden, James A. Ritter, Aaron Moran, Orhan Talu, Yi Wang, Krista S. Walton, Pierre Billefont, Guy De Weireld

Journal

Adsorption

Year

2018

Adsorbents

[Ammonium ZSM-5](#)

Adsorbates

[Carbon Dioxide](#)

Category

Experiment, Interlaboratory Study

Temperature (K)

293

Pressure range (bar)

[0.00,45.00]

17 Views



[Leave feedback](#)

Isotherm Visualization

The Isotherm Visualization Tool can plot multiple isotherms and fits of isotherms. Options are described in the [User Guide](#).

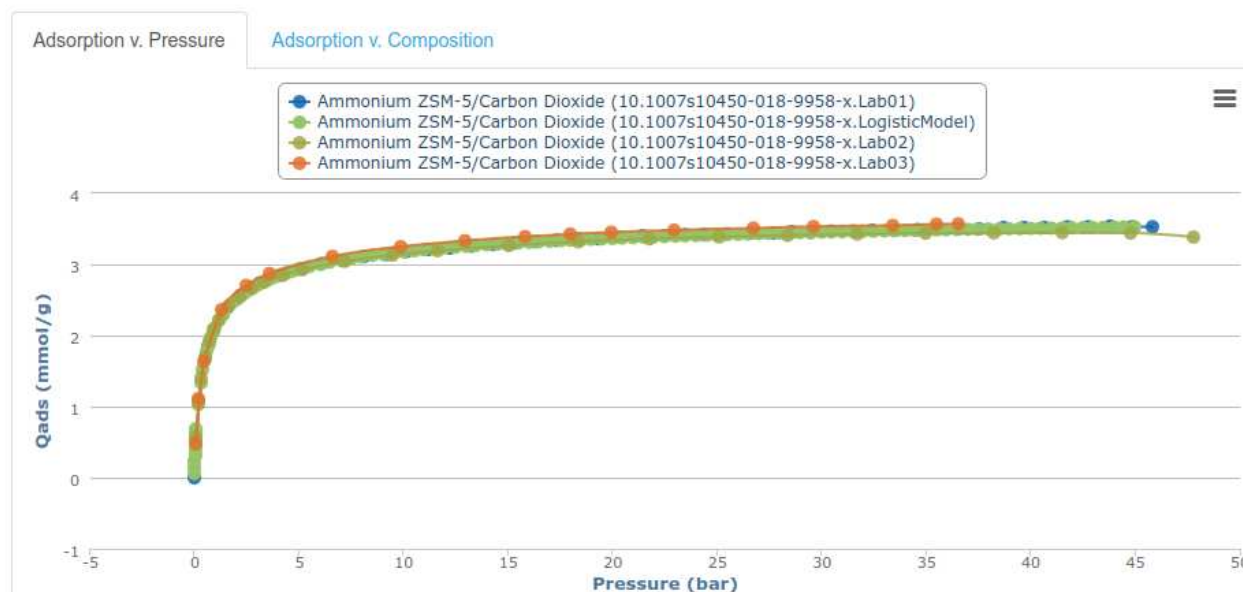


Figure 13: Screenshot of an article entry²⁹³ available in the NIST-ISODB, selected from results of a search using “NH₄-ZSM-5” as the adsorbent material and “carbon dioxide” as the adsorbate gas. Metadata about the experiments described in the article are shown in the upper half of the screenshot. The isotherm visualization widget shows four separate isotherms from the source article plotted together.

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 association of external resources to the unique identifiers, allowing, for example, 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 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, Fig. 14 shows the authors’ compilation of isotherms of carbon dioxide adsorption in HKUST-1 near 298 K from the NIST-ISODB, which provides graphical indication of the variability present

in reported experimental isotherms. Additionally, Fig. 15 graphically summarizes 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 computational evaluation of the offspring 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 NIST-ISODB is in force field tuning and development, an ongoing need that we discuss further in Sec. 9.2. 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

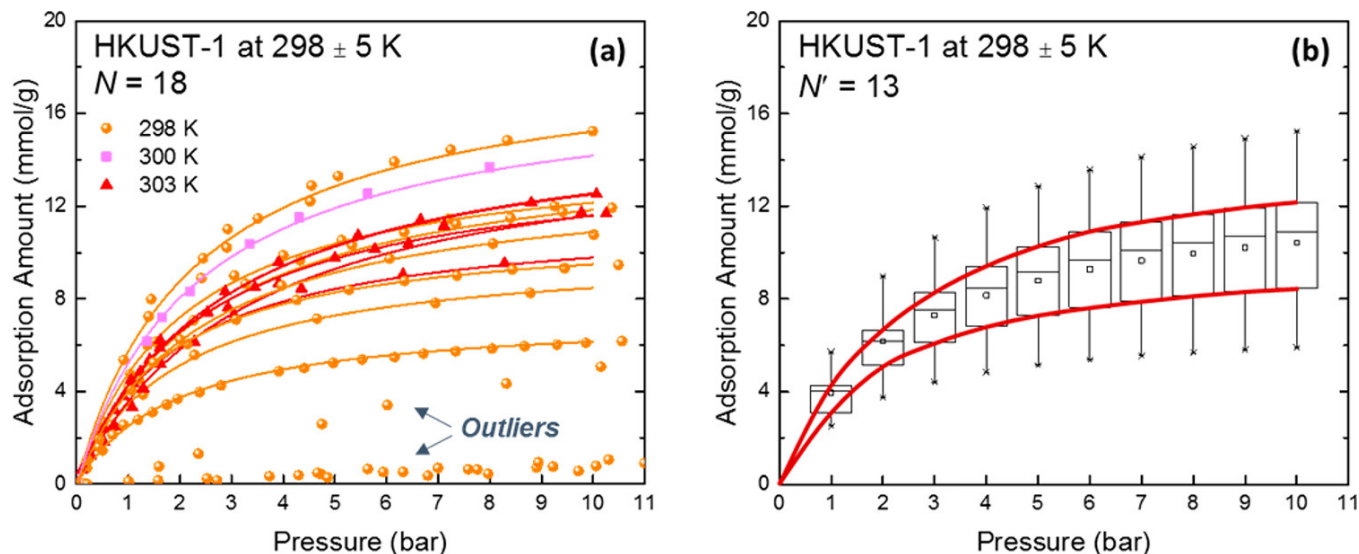


Figure 14: On the reproducibility of CO_2 adsorption isotherms in HKUST-1.²⁹⁹ (a) Experimental isotherms of carbon dioxide adsorption in HKUST-1 at temperatures of $298 \text{ K} \pm 5 \text{ K}$, as collected and organized by Park et al.²⁹⁹ from the NIST-ISODB, to show the variability of laboratory 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.

then open competitions can drive progress in molecular modeling. 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 standardization 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

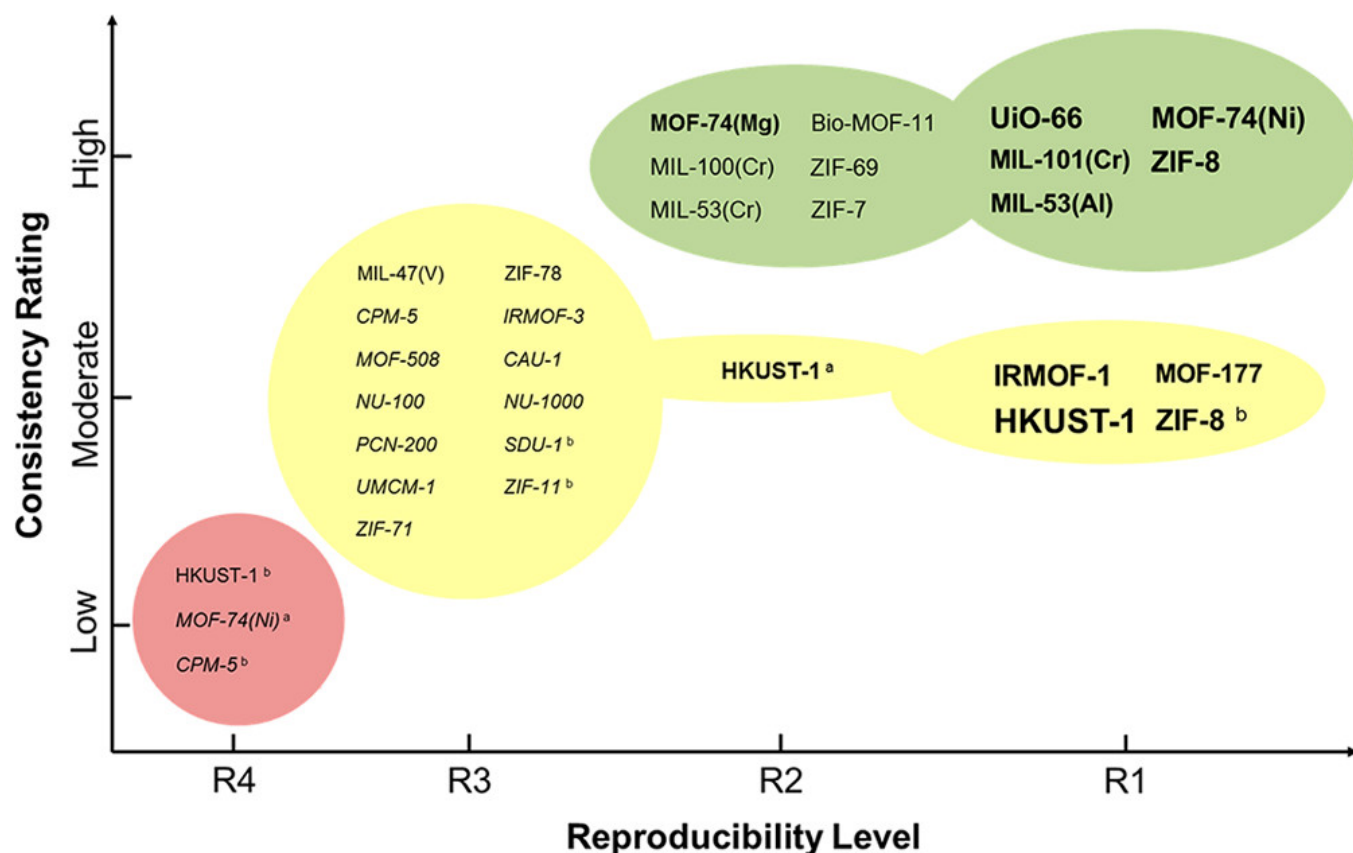


Figure 15: Graphical summary of reproducibility and consistency results from Park et al. for adsorption isotherms of carbon dioxide on various MOFs.²⁹⁹ On the x-axis, reproducibility increases (according to Park et al.’s metrics) from left to right. Font size is indicative of the number of independent isotherms available for the noted material and bolding/italicization 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.

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 cm³(STP)/g (cubic centimeters of adsorbate gas *at standard temperature and pressure* (STP) conditions per gram of adsorbent), fractional units such as weight-percent, units normalized 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 standardization involves descrip-

tion 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 modeling 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 utilize common data formats like the JSON isotherm file.

8 Discussion

In Sec. 6, 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 CRISPR¹¹⁹) and large, open datasets that spurred developments in recommendation algorithms²⁸⁶ 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 synthesize 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 functionalization to target CO₂,³²³ open transition metal sites to target ethylene³²⁴) or exclude (e.g., functionalization with hydrophobic alkyl chains to exclude water³²⁵) 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 harbor flexible backbones or rotating/wobbling constituents¹⁸⁹ 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.

9 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. 6 could be longer. Rigorously, the survey in Sec. 6 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. 6. 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.

9.1 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 backbones³³¹ 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 adsorption³³⁶ and helped inspire the discovery of DUT-50, a new material (though isoreticular to DUT-49³³⁷) that exhibits negative gas adsorption.

9.2 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^{§§}. As a conse-

^{§§}Though, notably, DFT calculations were used as an energetic description of small gas molecules in MOF-74 to compute Henry coefficients via Widom insertions³³⁸ by biasing the samples towards low-energy regions. Similarly, Fetisov et al.³³⁹ conducted first principle Monte

quence, 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.

9.2.1 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 CO₂ and CH₄ 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 CH₄ and CO₂ 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 CO₂ and CH₄, emphasizes 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 CO₂, H₂, N₂, and CH₄ 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 accu-

Carlo simulations of CO₂, N₂, and H₂O 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 CH₄ adsorption in CuBTC using *ab initio* calculations on a grid to characterize the CuBTC-CH₄ interaction.

rate, and significant research efforts should be devoted to the development of accurate and transferable force fields for predicting adsorption in MOFs. A flagship failure of generic force fields is in accurately modeling 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 CO₂ and CH₄ 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 CO₂, CH₄, and H₂O 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 CO₂, N₂, and H₂O 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 Fig. 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 surfaces³⁵¹ 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^{¶¶}.

Notably, assessing the accuracy of a force field via comparison with an experimental adsorption measurement is complicated because attributing deviations between simulation and ex-

periment 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 polarizability 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, vapor-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 modeling 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 characterization techniques and conditions³⁵²) 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.

¶¶The following thought experiment 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.

9.2.2 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 significant 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) method³⁵³ and the molecular building block-based (MBBB) method³⁵⁴ 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.

9.3 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 experimental³⁵⁹ and computational^{360,361} studies have elucidated how the linker defects significantly influence adsorption of CO₂ and H₂O in UiO-66. Through molecular modeling, 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.

9.4 Predicting MOF synthesizability 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 synthesize “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 synthesizing a hypothetical MOF and identify synthetic conditions that provide favorable thermodynamics and kinetics for its formation, mainly by modeling 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 center, 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 center. Classical simulations have also been used

to study MOF self-assembly. Yoneya and co-workers³⁶⁷ 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-workers³⁷⁰ used enhanced-sampling techniques to simulate the self-assembly of MOF-5, finding the process is entropically favorable 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-workers³⁷¹ 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 Allendorf³⁷² conducted molecular dynamics simulations of water in MOF-5, treating the flexibility of MOF-5 and modeling 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 ZnO₄ polyhedra. In subsequent work, Bellarosa and co-workers³⁷³ 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.

9.5 Multi-scale modeling

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 CO₂ are commonly used performance metrics for CO₂ capture applications, while the overall process objectives are the cost of capturing and recovering the CO₂ (\$/g CO₂) and the productivity of the material,³⁷⁸ both under CO₂ 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 modeling 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.

Krishna³⁸⁰ 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 modeling and optimization of e.g., the column length and pressure of adsorption and desorption, under methane recovery and purity constraints, to rank zeolites for CO₂ 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 CO₂ capture materials used in pressure-swing adsorption units. Molecular simulations and full process simulation and optimizations were carried out to screen ≈ 2900 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 CO₂ 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 CO₂ capture using a pressure-swing adsorption unit packed with that material.

Finis

We reviewed the demonstrated impacts of molecular modeling 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 routinized with well-directed research efforts.

Acknowledgement C. M. S. thanks the School of Chemical, Biological, and Environment Engineering (CBEE) at Oregon State University for start-up funds. Y. G. C. thanks the financial support from the National Research Foundation of Korea (NRF) grant funded by the Ministry of Education (NRF-2016R1D1A1B03934484).

References

- (1) Furukawa, H.; Cordova, K. E.; O’Keeffe, M. et al. The chemistry and applications of metal–organic frameworks. *Science* **2013**, *341*, 1230444.
- (2) Farha, O. K.; Eryazici, I.; Jeong, N. C. et al. Metal–organic framework materials with ultrahigh surface areas: is the sky the limit? *Journal of the American Chemical Society* **2012**, *134*, 15016–15021.
- (3) Morris, R. E.; Wheatley, P. S. Gas storage in nanoporous materials. *Angewandte Chemie International Edition* **2008**, *47*, 4966–4981.
- (4) Li, J.-R.; Kuppler, R. J.; Zhou, H.-C. Selective gas adsorption and separation in metal–organic frameworks. *Chemical Society Reviews* **2009**, *38*, 1477–1504.
- (5) Kreno, L. E.; Leong, K.; Farha, O. K. et al. Metal–organic framework materials as chemical sensors. *Chemical Reviews* **2012**, *112*, 1105–1125.
- (6) Li, H.; Eddaoudi, M.; O’Keeffe, M. et al. Design and synthesis of an exceptionally stable and highly porous metal–organic framework. *Nature* **1999**, *402*, 276–279.
- (7) Notman, N. MOFs find a use. *Chemistry World* **2017**, <https://www.chemistryworld.com/features/mofs-find-a-use/2500508.article>, Accessed 1 April 2019.
- (8) Hendon, C. H.; Rieth, A. J.; Kozłyński, M. D. et al. Grand challenges and future opportunities for metal–organic frameworks. *ACS Central Science* **2017**, *3*, 554–563.
- (9) Chung, Y. G.; Camp, J.; Haranczyk, M. et al. Computation-ready, experimental metal–organic frameworks: A tool to enable high-throughput screening of nanoporous crystals. *Chemistry of Materials* **2014**, *26*, 6185–6192.
- (10) Moghadam, P. Z.; Li, A.; Wiggins, S. B. et al. Development of a Cambridge Structural Database subset: a collection of metal–organic frameworks for past, present, and future. *Chemistry of Materials* **2017**, *29*, 2618–2625.
- (11) Stock, N.; Biswas, S. Synthesis of metal–organic frameworks (MOFs): routes to various MOF topologies, morphologies, and composites. *Chemical Reviews* **2011**, *112*, 933–969.
- (12) Eddaoudi, M.; Kim, J.; Rosi, N. et al. Systematic design of pore size and functionality in isoreticular MOFs and their application in methane storage. *Science* **2002**, *295*, 469–472.
- (13) Cordova, K. E.; Yaghi, O. M. The ‘folklore’ and reality of reticular chemistry. *Materials Chemistry Frontiers* **2017**, *1*, 1304–1309.
- (14) Howarth, A. J.; Liu, Y.; Li, P. et al. Chemical, thermal and mechanical stabilities of metal–organic frameworks. *Nature Reviews Materials* **2016**, *1*, 15018.
- (15) Rosi, N. L.; Kim, J.; Eddaoudi, M. et al. Rod packings and metal–organic frameworks constructed from rod-shaped secondary building units. *Journal of the American Chemical Society* **2005**, *127*, 1504–1518.
- (16) Chui, S. S.-Y.; Lo, S. M.-F.; Charman, J. P. et al. A chemically

- p>functionalizable nanoporous material
- $[\text{Cu}_3(\text{TMA})_2(\text{H}_2\text{O})_3]_n$
- .
- Science*
- 1999**
- ,
- 283*
- , 1148–1150.
- (17) Simon, C. M.; Kim, J.; Gómez-Gualdrón, D. A. et al. Computer-aided search for materials to store natural gas for vehicles. *Frontiers for Young Minds* **2015**, *3*, 11.
 - (18) Schoedel, A.; Ji, Z.; Yaghi, O. M. The role of metal–organic frameworks in a carbon-neutral energy cycle. *Nature Energy* **2016**, *1*, 16034.
 - (19) Mason, J. A.; Veenstra, M.; Long, J. R. Evaluating metal–organic frameworks for natural gas storage. *Chemical Science* **2014**, *5*, 32–51.
 - (20) Moosavi, S. M.; Chidambaram, A.; Talirz, L. et al. Capturing chemical intuition in synthesis of metal–organic frameworks. *Nature Communications* **2019**, *10*, 539.
 - (21) Greenaway, R. L.; Santolini, V.; Bennison, M. J. et al. High-throughput discovery of organic cages and catenanes using computational screening fused with robotic synthesis. *Nature Communications* **2018**, *9*, 2849.
 - (22) Steiner, S.; Wolf, J.; Glatzel, S. et al. Organic synthesis in a modular robotic system driven by a chemical programming language. *Science* **2019**, *363*, eaav2211.
 - (23) Häse, F.; Roch, L. M.; Aspuru-Guzik, A. Next-Generation Experimentation with Self-Driving Laboratories. *Trends in Chemistry* **2019**, *1*, 282–291.
 - (24) Jain, A.; Ong, S. P.; Hautier, G. et al. Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL Materials* **2013**, *1*, 011002.
 - (25) Boyd, P. G.; Lee, Y.; Smit, B. Computational development of the nanoporous materials genome. *Nature Reviews Materials* **2017**, *2*, 17037.
 - (26) Fraux, G.; Chibani, S.; Coudert, F.-X. Modelling of framework materials at multiple scales: current practices and open questions. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **2019**, *377*, 20180220.
 - (27) Colón, Y. J.; Snurr, R. Q. High-throughput computational screening of metal–organic frameworks. *Chemical Society Reviews* **2014**, *43*, 5735–5749.
 - (28) Rappé, A. K.; Casewit, C. J.; Colwell, K. S. et al. UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *Journal of the American Chemical Society* **1992**, *114*, 10024–10035.
 - (29) Mayo, S. L.; Olafson, B. D.; Goddard, W. A. DREIDING: a generic force field for molecular simulations. *The Journal of Physical Chemistry* **1990**, *94*, 8897–8909.
 - (30) Martin, M. G.; Siepmann, J. I. Transferable potentials for phase equilibria. 1. United-atom description of n-alkanes. *The Journal of Physical Chemistry B* **1998**, *102*, 2569–2577.
 - (31) Potoff, J. J.; Siepmann, J. I. Vapor–liquid equilibria of mixtures containing alkanes, carbon dioxide, and nitrogen. *AIChE Journal* **2001**, *47*, 1676–1682.
 - (32) Ongari, D.; Boyd, P. G.; Kadioglu, O. et al. Evaluating charge equilibration methods to generate electrostatic fields in nanoporous materials. *Journal of Chemical Theory and Computation* **2019**, *15*, 382–401.
 - (33) Fischer, M.; Gomes, J. R.; Jorge, M. Computational approaches to study adsorption in MOFs with unsaturated metal sites. *Molecular Simulation* **2014**, *40*, 537–556.
 - (34) Zhang, Y.; Li, B.; Krishna, R. et al. Highly selective adsorption of ethylene

- over ethane in a MOF featuring the combination of open metal site and π -complexation. *Chemical Communications* **2015**, 51, 2714–2717.
- (35) Uchida, S.; Kawamoto, R.; Tagami, H. et al. Highly selective sorption of small unsaturated hydrocarbons by nonporous flexible framework with silver ion. *Journal of the American Chemical Society* **2008**, 130, 12370–12376.
 - (36) Fischer, M.; Gomes, J. R.; Froba, M. et al. Modeling adsorption in metal–organic frameworks with open metal sites: propane/propylene separations. *Langmuir* **2012**, 28, 8537–8549.
 - (37) Kulkarni, A. R.; Sholl, D. S. Screening of copper open metal site MOFs for olefin/paraffin separations using DFT-derived force fields. *The Journal of Physical Chemistry C* **2016**, 120, 23044–23054.
 - (38) Fischer, M.; Kuchta, B.; Firlej, L. et al. Accurate prediction of hydrogen adsorption in metal–organic frameworks with unsaturated metal sites via a combined density-functional theory and molecular mechanics approach. *The Journal of Physical Chemistry C* **2010**, 114, 19116–19126.
 - (39) Fischer, M.; Hoffmann, F.; Fröba, M. New microporous materials for acetylene storage and C_2H_2/CO_2 separation: insights from molecular simulations. *ChemPhysChem* **2010**, 11, 2220–2229.
 - (40) Zhang, C.; Lan, Y.; Guo, X. et al. Materials genomics-guided ab initio screening of MOFs with open copper sites for acetylene storage. *AIChE Journal* **2018**, 64, 1389–1398.
 - (41) Chen, L.; Morrison, C. A.; Duren, T. Improving predictions of gas adsorption in metal–organic frameworks with coordinatively unsaturated metal sites: model potentials, ab initio parameterization, and GCMC simulations. *The Journal of Physical Chemistry C* **2012**, 116, 18899–18909.
 - (42) Ongari, D.; Tiana, D.; Stoneburner, S. J. et al. Origin of the Strong Interaction between Polar Molecules and Copper (II) Paddle-Wheels in Metal Organic Frameworks. *The Journal of Physical Chemistry C* **2017**, 121, 15135–15144.
 - (43) Koh, H. S.; Rana, M. K.; Wong-Foy, A. G. et al. Predicting methane storage in open-metal-site metal–organic frameworks. *The Journal of Physical Chemistry C* **2015**, 119, 13451–13458.
 - (44) Campbell, C.; Gomes, J. R.; Fischer, M. et al. New Model for Predicting Adsorption of Polar Molecules in Metal–Organic Frameworks with Unsaturated Metal Sites. *The Journal of Physical Chemistry Letters* **2018**, 9, 3544–3553.
 - (45) Dzubak, A. L.; Lin, L.-C.; Kim, J. et al. Ab initio carbon capture in open-site metal–organic frameworks. *Nature Chemistry* **2012**, 4, 810–816.
 - (46) Mercado, R.; Vlaisavljevich, B.; Lin, L.-C. et al. Force field development from periodic density functional theory calculations for gas separation applications using metal–organic frameworks. *The Journal of Physical Chemistry C* **2016**, 120, 12590–12604.
 - (47) Campbell, C.; Ferreira-Rangel, C. A.; Fischer, M. et al. A transferable model for adsorption in MOFs with unsaturated metal sites. *The Journal of Physical Chemistry C* **2016**, 121, 441–458.
 - (48) Frenkel, D.; Smit, B. *Understanding molecular simulation: from algorithms to applications*; Elsevier, 2001; Vol. 1.
 - (49) Cho, E. H.; Lyu, Q.; Lin, L.-C. Computational discovery of nanoporous materials

- for energy-and environment-related applications. *Molecular Simulation* **2019**, 1–26.
- (50) Coudert, F.-X.; Fuchs, A. H. Computational characterization and prediction of metal–organic framework properties. *Coordination Chemistry Reviews* **2016**, 307, 211–236.
 - (51) Addicoat, M. A.; Vankova, N.; Akter, I. F. et al. Extension of the universal force field to metal–organic frameworks. *Journal of Chemical Theory and Computation* **2014**, 10, 880–891.
 - (52) Bureekaew, S.; Amirjalayer, S.; Tafipolsky, M. et al. MOF-FF-A flexible first-principles derived force field for metal–organic frameworks. *Physica Status Solidi (b)* **2013**, 250, 1128–1141.
 - (53) Bristow, J. K.; Tiana, D.; Walsh, A. Transferable force field for metal–organic frameworks from first-principles: BTW-FF. *Journal of Chemical Theory and Computation* **2014**, 10, 4644–4652.
 - (54) Boyd, P. G.; Moosavi, S. M.; Witman, M. et al. Force-field prediction of materials properties in metal-organic frameworks. *The Journal of Physical Chemistry Letters* **2017**, 8, 357–363.
 - (55) Vanduyfhuys, L.; Vandenbrande, S.; Verstraelen, T. et al. QuickFF: A program for a quick and easy derivation of force fields for metal–organic frameworks from ab initio input. *Journal of Computational Chemistry* **2015**, 36, 1015–1027.
 - (56) Vanduyfhuys, L.; Vandenbrande, S.; Wieme, J. et al. Extension of the QuickFF force field protocol for an improved accuracy of structural, vibrational, mechanical and thermal properties of metal–organic frameworks. *Journal of Computational Chemistry* **2018**, 39, 999–1011.
 - (57) Heinen, J.; Dubbeldam, D. On flexible force fields for metal–organic frameworks: Recent developments and future prospects. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2018**, 8, e1363.
 - (58) Lawler, K. V.; Hulvey, Z.; Forster, P. M. On the importance of a precise crystal structure for simulating gas adsorption in nanoporous materials. *Physical Chemistry Chemical Physics* **2015**, 17, 18904–18907.
 - (59) Li, C.-P.; Du, M. Role of solvents in coordination supramolecular systems. *Chemical Communications* **2011**, 47, 5958–5972.
 - (60) Nazarian, D.; Camp, J. S.; Chung, Y. G. et al. Large-Scale Refinement of Metal–Organic Framework Structures Using Density Functional Theory. *Chemistry of Materials* **2017**, 29, 2521–2528.
 - (61) Boyd, P. G.; Woo, T. K. A generalized method for constructing hypothetical nanoporous materials of any net topology from graph theory. *CrystEngComm* **2016**, 18, 3777–3792.
 - (62) Evans, J. D.; Fraux, G.; Gaillac, R. et al. Computational chemistry methods for nanoporous materials. *Chemistry of Materials* **2017**, 29, 199–212.
 - (63) Fraux, G.; Coudert, F.-X. Recent advances in the computational chemistry of soft porous crystals. *Chemical Communications* **2017**, 53, 7211–7221.
 - (64) Tijms, H. *Probability: A Lively Introduction*; Cambridge University Press, 2017.
 - (65) Shell, M. S. *Thermodynamics and Statistical Mechanics: An Integrated Approach*; Cambridge University Press, 2015.
 - (66) Panagiotopoulos, A. Z. Direct determination of phase coexistence properties of fluids by Monte Carlo simulation in a new

- p>ensemble.
- Molecular Physics*
- 1987**
- ,
- 61*
- , 813–826.
- (67) Panagiotopoulos, A. Z. Adsorption and capillary condensation of fluids in cylindrical pores by Monte Carlo simulation in the Gibbs ensemble. *Molecular Physics* **1987**, *62*, 701–719.
- (68) Dubbeldam, D.; Calero, S.; Ellis, D. E. et al. RASPA: molecular simulation software for adsorption and diffusion in flexible nanoporous materials. *Molecular Simulation* **2016**, *42*, 81–101.
- (69) Dubbeldam, D.; Torres-Knoop, A.; Walton, K. S. On the inner workings of Monte Carlo codes. *Molecular Simulation* **2013**, *39*, 1253–1292.
- (70) Jeffroy, M.; Fuchs, A. H.; Boutin, A. Structural changes in nanoporous solids due to fluid adsorption: thermodynamic analysis and Monte Carlo simulations. *Chemical Communications* **2008**, 3275–3277.
- (71) Mason, J. A.; Oktawiec, J.; Taylor, M. K. et al. Methane storage in flexible metal–organic frameworks with intrinsic thermal management. *Nature* **2015**, *527*, 357–361.
- (72) Sarkisov, L. Toward Rational Design of Metal–Organic Frameworks for Sensing Applications: Efficient Calculation of Adsorption Characteristics in Zero Loading Regime. *The Journal of Physical Chemistry C* **2012**, *116*, 3025–3033.
- (73) Sholl, D. S. Understanding macroscopic diffusion of adsorbed molecules in crystalline nanoporous materials via atomistic simulations. *Accounts of Chemical Research* **2006**, *39*, 403–411.
- (74) Dubbeldam, D.; Snurr, R. Q. Recent developments in the molecular modeling of diffusion in nanoporous materials. *Molecular Simulation* **2007**, *33*, 305–325.
- (75) Karmakar, T.; Piaggi, P. M.; Perego, C. et al. A Cannibalistic Approach to Grand Canonical Crystal Growth. *Journal of Chemical Theory and Computation* **2018**, *14*, 2678–2683.
- (76) Ozcan, A.; Perego, C.; Salvalaglio, M. et al. Concentration gradient driven molecular dynamics: a new method for simulations of membrane permeation and separation. *Chemical Science* **2017**, *8*, 3858–3865.
- (77) Perego, C.; Salvalaglio, M.; Parrinello, M. Molecular dynamics simulations of solutions at constant chemical potential. *The Journal of Chemical Physics* **2015**, *142*, 144113.
- (78) Sarkisov, L.; Monson, P. A. Hysteresis in Monte Carlo and Molecular Dynamics Simulations of Adsorption in Porous Materials. *Langmuir* **2000**, *16*, 9857–9860.
- (79) Düren, T.; Jakobtorweihen, S.; Keil, F. J. et al. Grand canonical molecular dynamics simulations of transport diffusion in geometrically heterogeneous pores. *Physical Chemistry Chemical Physics* **2003**, *5*, 369–375.
- (80) Delle Site, L.; Krekeler, C.; Whittaker, J. et al. Molecular Dynamics of Open Systems: Construction of a Mean-Field Particle Reservoir. *Advanced Theory and Simulations* **2019**, *2*, 1900014.
- (81) Moghadam, P. Z.; Rogge, S. M.; Li, A. et al. Structure-Mechanical Stability Relations of Metal-Organic Frameworks via Machine Learning. *Matter* **2019**,
- (82) Ghoufi, A.; Maurin, G. Hybrid Monte Carlo Simulations Combined with a Phase Mixture Model to Predict the Structural Transitions of a Porous Metal–Organic Framework Material upon Adsorption of Guest Molecules. *The Journal of Physical Chemistry C* **2010**, *114*, 6496–6502.

- (83) Palmer, J. C.; Debenedetti, P. G. Computer Simulation of Water Sorption on Flexible Protein Crystals. *The Journal of Physical Chemistry Letters* **2012**, *3*, 2713–2718.
- (84) Rogge, S. M. J.; Goeminne, R.; Demuyne, R. et al. Modeling Gas Adsorption in Flexible Metal–Organic Frameworks via Hybrid Monte Carlo/Molecular Dynamics Schemes. *Advanced Theory and Simulations* **2019**, *2*, 1800177.
- (85) Simon, C.; York, A. H.; Sturluson, A. et al. PorousMaterials.jl. 2018; <https://github.com/SimonEnsemble/PorousMaterials.jl>.
- (86) Gowers, R. J.; Farmahini, A. H.; Friedrich, D. et al. Automated analysis and benchmarking of GCMC simulation programs in application to gas adsorption. *Molecular Simulation* **2018**, *44*, 309–321.
- (87) Grossfield, A.; Patrone, P. N.; Roe, D. R. et al. Best Practices for Quantification of Uncertainty and Sampling Quality in Molecular Simulations [Article v1.0]. *Living Journal of Computational Molecular Science* **2019**, *1*, 5067.
- (88) Wilmer, C. E.; Farha, O. K.; Bae, Y.-S. et al. Structure–property relationships of porous materials for carbon dioxide separation and capture. *Energy & Environmental Science* **2012**, *5*, 9849–9856.
- (89) Simon, C. M.; Kim, J.; Gómez-Gualdrón, D. A. et al. The materials genome in action: identifying the performance limits for methane storage. *Energy & Environmental Science* **2015**, *8*, 1190–1199.
- (90) Frenkel, D.; Mooij, G. C. A. M.; Smit, B. Novel scheme to study structural and thermal properties of continuously deformable molecules. *Journal of Physics: Condensed Matter* **1992**, *4*, 3053–3076.
- (91) Becker, T. M.; Heinen, J.; Dubbel-dam, D. et al. Polarizable force fields for CO₂ and CH₄ adsorption in M-MOF-74. *The Journal of Physical Chemistry C* **2017**, *121*, 4659–4673.
- (92) Zhang, H.; Snurr, R. Q. Computational Study of Water Adsorption in the Hydrophobic Metal–Organic Framework ZIF-8: Adsorption Mechanism and Acceleration of the Simulations. *The Journal of Physical Chemistry C* **2017**, *121*, 24000–24010.
- (93) Friedman, J.; Hastie, T.; Tibshirani, R. *The elements of statistical learning*; Springer series in statistics New York, NY, USA.; 2001; Vol. 1.
- (94) Willems, T. F.; Rycroft, C. H.; Kazi, M. et al. Algorithms and tools for high-throughput geometry-based analysis of crystalline porous materials. *Microporous and Mesoporous Materials* **2012**, *149*, 134–141.
- (95) Collins, C. R.; Gordon, G. J.; von Lilienfeld, O. A. et al. Constant size descriptors for accurate machine learning models of molecular properties. *The Journal of Chemical Physics* **2018**, *148*, 241718.
- (96) Fernandez, M.; Boyd, P. G.; Daff, T. D. et al. Rapid and accurate machine learning recognition of high performing metal organic frameworks for CO₂ capture. *The Journal of Physical Chemistry Letters* **2014**, *5*, 3056–3060.
- (97) Simon, C. M.; Mercado, R.; Schnell, S. K. et al. What are the best materials to separate a xenon/krypton mixture? *Chemistry of Materials* **2015**, *27*, 4459–4475.
- (98) Bucior, B. J.; Bobbitt, N. S.; Islamoglu, T. et al. Energy-based descriptors to rapidly predict hydrogen storage in metal–organic frameworks. *Molecular Systems Design & Engineering* **2019**, *4*, 162–174.

- (99) Thornton, A. W.; Simon, C. M.; Kim, J. et al. Materials genome in action: identifying the performance limits of physical hydrogen storage. *Chemistry of Materials* **2017**, *29*, 2844–2854.
- (100) Pardakhti, M.; Moharreri, E.; Wanik, D. et al. Machine learning using combined structural and chemical descriptors for prediction of methane adsorption performance of metal organic frameworks (MOFs). *ACS Combinatorial Science* **2017**, *19*, 640–645.
- (101) Le, T.; Epa, V. C.; Burden, F. R. et al. Quantitative structure–property relationship modeling of diverse materials properties. *Chemical Reviews* **2012**, *112*, 2889–2919.
- (102) Fernandez, M.; Woo, T. K.; Wilmer, C. E. et al. Large-scale quantitative structure–property relationship (QSPR) analysis of methane storage in metal-organic frameworks. *The Journal of Physical Chemistry C* **2013**, *117*, 7681–7689.
- (103) Fernandez, M.; Trefiak, N. R.; Woo, T. K. Atomic property weighted radial distribution functions descriptors of metal–organic frameworks for the prediction of gas uptake capacity. *The Journal of Physical Chemistry C* **2013**, *117*, 14095–14105.
- (104) Fernandez, M.; Barnard, A. S. Geometrical properties can predict CO₂ and N₂ adsorption performance of metal-organic frameworks (MOFs) at low pressure. *ACS Combinatorial Science* **2016**, *18*, 243–252.
- (105) Borboudakis, G.; Stergiannakos, T.; Frysali, M. et al. Chemically intuited, large-scale screening of MOFs by machine learning techniques. *npj Computational Materials* **2017**, *3*, 40.
- (106) Anderson, R.; Rodgers, J.; Argueta, E. et al. Role of pore chemistry and topology in the CO₂ capture capabilities of MOFs: from molecular simulation to machine learning. *Chemistry of Materials* **2018**, *30*, 6325–6337.
- (107) Tsivion, E.; Long, J. R.; Head-Gordon, M. Hydrogen physisorption on metal–organic framework linkers and metalated linkers: a computational study of the factors that control binding strength. *Journal of the American Chemical Society* **2014**, *136*, 17827–17835.
- (108) Anderson, G.; Schweitzer, B.; Anderson, R. et al. Attainable Volumetric Targets for Adsorption-Based Hydrogen Storage in Porous Crystals: Molecular Simulation and Machine Learning. *The Journal of Physical Chemistry C* **2019**, *123*, 120–130.
- (109) Wu, X.; Xiang, S.; Su, J. et al. Understanding quantitative relationship between methane storage capacities and characteristic properties of metal organic frameworks based on machine learning. *The Journal of Physical Chemistry C* **2019**,
- (110) LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436.
- (111) Goh, G. B.; Hodas, N. O.; Vishnu, A. Deep learning for computational chemistry. *Journal of Computational Chemistry* **2017**, *38*, 1291–1307.
- (112) Gen, M.; Lin, L. Genetic Algorithms. *Wiley Encyclopedia of Computer Science and Engineering* **2007**, 1–15.
- (113) Collins, S. P.; Daff, T. D.; Piotrkowski, S. S. et al. Materials design by evolutionary optimization of functional groups in metal-organic frameworks. *Science Advances* **2016**, *2*, e1600954.
- (114) Chung, Y. G.; Gómez-Gualdrón, D. A.; Li, P. et al. In silico discovery of metal–organic frameworks for precombustion CO₂ capture using a genetic algorithm. *Science Advances* **2016**, *2*, e1600909.

- (115) Bao, Y.; Martin, R. L.; Simon, C. M. et al. In silico discovery of high deliverable capacity metal–organic frameworks. *The Journal of Physical Chemistry C* **2015**, *119*, 186–195.
- (116) Doudna, J. A.; Charpentier, E. The new frontier of genome engineering with CRISPR-Cas9. *Science* **2014**, *346*, 1258096.
- (117) Flavahan, W. A.; Gaskell, E.; Bernstein, B. E. Epigenetic plasticity and the hallmarks of cancer. *Science* **2017**, *357*, eaal2380.
- (118) Barrangou, R.; Fremaux, C.; Deveau, H. et al. CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **2007**, *315*, 1709–1712.
- (119) Lander, E. S. The heroes of CRISPR. *Cell* **2016**, *164*, 18–28.
- (120) Mojica, F. J.; Díez-Villaseñor, C.; Soria, E. et al. Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. *Molecular Microbiology* **2000**, *36*, 244–246.
- (121) Jansen, R.; Embden, J. D. A. v.; Gaastera, W. et al. Identification of genes that are associated with DNA repeats in prokaryotes. *Molecular Microbiology* **2002**, *43*, 1565–1575.
- (122) Mojica, F. J.; Díez-Villaseñor, C.; García-Martínez, J. et al. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *Journal of Molecular Evolution* **2005**, *60*, 174–182.
- (123) Goodman, L. Hypothesis-limited research. *Genome Research* **1999**, *9*, 673–676.
- (124) Gándara, F.; Bennett, T. D. Crystallography of metal–organic frameworks. *IUCrJ* **2014**, *1*, 563–570.
- (125) Harris, K. D. M.; Tremayne, M. Crystal structure determination from powder diffraction data. *Chemistry of Materials* **1996**, *8*, 2554–2570.
- (126) McCusker, L. B.; Von Dreele, R. B.; Cox, D. E. et al. Rietveld refinement guidelines. *Journal of Applied Crystallography* **1999**, *32*, 36–50.
- (127) Rietveld, H. M. A profile refinement method for nuclear and magnetic structures. *Journal of Applied Crystallography* **1969**, *2*, 65–71.
- (128) Øien-Ødegaard, S.; Shearer, G. C.; Wragg, D. S. et al. Pitfalls in metal–organic framework crystallography: towards more accurate crystal structures. *Chemical Society Reviews* **2017**, *46*, 4867–4876.
- (129) O’Keeffe, M.; Yaghi, O. M. Deconstructing the crystal structures of metal–organic frameworks and related materials into their underlying nets. *Chemical Reviews* **2012**, *112*, 675–702.
- (130) Parsons, S. Introduction to twinning. *Acta Crystallographica Section D* **2003**, *59*, 1995–2003.
- (131) Ma, J.; Tran, L. D.; Matzger, A. J. Toward topology prediction in Zr-based microporous coordination polymers: the role of linker geometry and flexibility. *Crystal Growth & Design* **2016**, *16*, 4148–4153.
- (132) Jiang, H.-L.; Makal, T. A.; Zhou, H.-C. Interpenetration control in metal–organic frameworks for functional applications. *Coordination Chemistry Reviews* **2013**, *257*, 2232–2249.
- (133) Hoshino, M.; Khutia, A.; Xing, H. et al. The crystalline sponge method updated. *IUCrJ* **2016**, *3*, 139–151.
- (134) Spek, A. L. PLATON SQUEEZE: a tool for the calculation of the disordered solvent contribution to the calculated struc-

- ture factors. *Acta Crystallographica Section C: Structural Chemistry* **2015**, *71*, 9–18.
- (135) Winston, E. B.; Lowell, P. J.; Vacek, J. et al. Dipolar molecular rotors in the metal–organic framework crystal IRMOF-2. *Physical Chemistry Chemical Physics* **2008**, *10*, 5188–5191.
- (136) Elsaidi, S. K.; Mohamed, M. H.; Simon, C. M. et al. Effect of ring rotation upon gas adsorption in SIFSIX-3-M (M= Fe, Ni) pillared square grid networks. *Chemical Science* **2017**, *8*, 2373–2380.
- (137) Harris, K. D. M.; Tremayne, M.; Lightfoot, P. et al. Crystal structure determination from powder diffraction data by Monte Carlo methods. *Journal of the American Chemical Society* **1994**, *116*, 3543–3547.
- (138) David, W. I.; Shankland, K.; van de Streek, J. et al. DASH: a program for crystal structure determination from powder diffraction data. *Journal of Applied Crystallography* **2006**, *39*, 910–915.
- (139) Wahiduzzaman, M.; Wang, S.; Sikora, B. J. et al. Computational structure determination of novel metal–organic frameworks. *Chemical Communications* **2018**, *54*, 10812–10815.
- (140) Li, P.; Vermeulen, N. A.; Malliakas, C. D. et al. Bottom-up construction of a superstructure in a porous uranium-organic crystal. *Science* **2017**, *356*, 624–627.
- (141) Colón, Y. J.; Gómez-Gualdrón, D. A.; Snurr, R. Q. Topologically Guided, Automated Construction of Metal–Organic Frameworks and Their Evaluation for Energy-Related Applications. *Crystal Growth & Design* **2017**, *17*, 5801–5810.
- (142) Deria, P.; Gomez-Gualdrón, D. A.; Bury, W. et al. Ultraporous, water stable, and breathing zirconium-based metal–organic frameworks with ftw topology. *Journal of the American Chemical Society* **2015**, *137*, 13183–13190.
- (143) Deng, H.; Grunder, S.; Cordova, K. E. et al. Large-pore apertures in a series of metal-organic frameworks. *Science* **2012**, *336*, 1018–1023.
- (144) Taddei, M.; Casati, N.; Steitz, D. A. et al. In situ high-resolution powder X-ray diffraction study of UiO-66 under synthesis conditions in a continuous-flow microwave reactor. *CrystEngComm* **2017**, *19*, 3206–3214.
- (145) Park, J. H.; Min Choi, K.; Joon Jeon, H. et al. In-situ observation for growth of hierarchical metal–organic frameworks and their self-sequestering mechanism for gas storage. *Scientific Reports* **2015**, *5*, 12045.
- (146) Wu, Y.; Henke, S.; Kieslich, G. et al. Time-Resolved In Situ X-ray Diffraction Reveals Metal-Dependent Metal–Organic Framework Formation. *Angewandte Chemie International Edition* **2016**, *55*, 14081–14084.
- (147) Dill, E. D.; Josey, A. A.; Folmer, J. C. W. et al. Experimental determination of the crystallization phase-boundary velocity in the halozeotype CZX-1. *Chemistry of Materials* **2013**, *25*, 3932–3940.
- (148) Julien, P. A.; Užarević, K.; Katse- nis, A. D. et al. In situ monitoring and mechanism of the mechanochemical formation of a microporous MOF-74 framework. *Journal of the American Chemical Society* **2016**, *138*, 2929–2932.
- (149) Peterson, V. K.; Southon, P. D.; Halder, G. J. et al. Guest Adsorption in the Nanoporous Metal–Organic Framework $\text{Cu}_3(1,3,5\text{-Benzenetricarboxylate})_2$: Combined In Situ X-Ray Diffraction and Vapor Sorption. *Chemistry of Materials* **2014**, *26*, 4712–4723.

- (150) Gonzalez, M. I.; Mason, J. A.; Bloch, E. D. et al. Structural characterization of framework–gas interactions in the metal–organic framework Co₂(dobdc) by in situ single-crystal X-ray diffraction. *Chemical Science* **2017**, *8*, 4387–4398.
- (151) Walton, R. I.; Munn, A. S.; Guillou, N. et al. Uptake of Liquid Alcohols by the Flexible Fe^{III} Metal–Organic Framework MIL-53 Observed by Time-Resolved In Situ X-ray Diffraction. *Chemistry - A European Journal* **2011**, *17*, 7069–7079.
- (152) Scherb, C.; Koehn, R.; Bein, T. Sorption behavior of an oriented surface-grown MOF-film studied by in situ X-ray diffraction. *Journal of Materials Chemistry* **2010**, *20*, 3046–3051.
- (153) Allen, F. H. The Cambridge Structural Database: a quarter of a million crystal structures and rising. *Acta Crystallographica Section B: Structural Science* **2002**, *58*, 380–388.
- (154) Groom, C. R.; Bruno, I. J.; Lightfoot, M. P. et al. The Cambridge structural database. *Acta Crystallographica Section B: Structural Science, Crystal Engineering and Materials* **2016**, *72*, 171–179.
- (155) Bruno, I. J.; Cole, J. C.; Edgington, P. R. et al. New software for searching the Cambridge Structural Database and visualizing crystal structures. *Acta Crystallographica Section B: Structural Science* **2002**, *58*, 389–397.
- (156) Bruno, I. J.; Groom, C. R. A crystallographic perspective on sharing data and knowledge. *Journal of Computer-Aided Molecular Design* **2014**, *28*, 1015–1022.
- (157) Batten, S. R.; Champness, N. R.; Chen, X.-M. et al. Coordination polymers, metal–organic frameworks and the need for terminology guidelines. *CrystEngComm* **2012**, *14*, 3001–3004.
- (158) Seth, S.; Matzger, A. J. Metal–Organic Frameworks: Examples, Counterexamples, and an Actionable Definition. *Crystal Growth & Design* **2017**, *17*, 4043–4048.
- (159) Ockwig, N. W.; Delgado-Friedrichs, O.; O’Keeffe, M. et al. Reticular chemistry: occurrence and taxonomy of nets and grammar for the design of frameworks. *Accounts of Chemical Research* **2005**, *38*, 176–182.
- (160) Goldsmith, J.; Wong-Foy, A. G.; Cafarella, M. J. et al. Theoretical limits of hydrogen storage in metal–organic frameworks: Opportunities and trade-offs. *Chemistry of Materials* **2013**, *25*, 3373–3382.
- (161) Farha, O. K.; Hupp, J. T. Rational design, synthesis, purification, and activation of metal–organic framework materials. *Accounts of Chemical Research* **2010**, *43*, 1166–1175.
- (162) Howarth, A. J.; Peters, A. W.; Vermeulen, N. A. et al. Best practices for the synthesis, activation, and characterization of metal–organic frameworks. *Chemistry of Materials* **2017**, *29*, 26–39.
- (163) Murdock, C. R.; McNutt, N. W.; Keffeler, D. J. et al. Rotating phenyl rings as a guest-dependent switch in two-dimensional metal–organic frameworks. *Journal of the American Chemical Society* **2014**, *136*, 671–678.
- (164) Haldoupis, E.; Nair, S.; Sholl, D. S. Efficient Calculation of Diffusion Limitations in Metal Organic Framework Materials: A Tool for Identifying Materials for Kinetic Separations. *Journal of the American Chemical Society* **2010**, *132*, 7528–7539.
- (165) Van Heest, T.; Teich-McGoldrick, S. L.; Greathouse, J. A. et al. Identification of Metal–Organic Framework Materials for Adsorption Separation of Rare Gases:

- Applicability of Ideal Adsorbed Solution Theory (IAST) and Effects of Inaccessible Framework Regions. *The Journal of Physical Chemistry C* **2012**, *116*, 13183–13195.
- (166) Panella, B.; Hirscher, M.; Roth, S. Hydrogen adsorption in different carbon nanostructures. *Carbon* **2005**, *43*, 2209–2214.
- (167) Brandani, S.; Mangano, E.; Sarkisov, L. Net, excess and absolute adsorption and adsorption of helium. *Adsorption* **2016**, *22*, 261–276.
- (168) Myers, A. L.; Monson, P. A. Physical adsorption of gases: the case for absolute adsorption as the basis for thermodynamic analysis. *Adsorption* **2014**, *20*, 591–622.
- (169) Watanabe, T.; Sholl, D. S. Accelerating Applications of Metal–Organic Frameworks for Gas Adsorption and Separation by Computational Screening of Materials. *Langmuir* **2012**, *28*, 14114–14128.
- (170) Pinheiro, M.; Martin, R. L.; Rycroft, C. H. et al. High accuracy geometric analysis of crystalline porous materials. *CrystEngComm* **2013**, *15*, 7531–7538.
- (171) Pinheiro, M.; Martin, R. L.; Rycroft, C. H. et al. Characterization and comparison of pore landscapes in crystalline porous materials. *Journal of Molecular Graphics and Modelling* **2013**, *44*, 208–219.
- (172) Computation-ready, Experimental (CoRE) MOFs download link. <http://gregchung.github.io/CoRE-MOFs/>.
- (173) Watanabe, T.; Manz, T. A.; Sholl, D. S. Accurate treatment of electrostatics during molecular adsorption in nanoporous crystals without assigning point charges to framework atoms. *The Journal of Physical Chemistry C* **2011**, *115*, 4824–4836.
- (174) Walton, K. S.; Millward, A. R.; Dubbel-dam, D. et al. Understanding inflections and steps in carbon dioxide adsorption isotherms in metal–organic frameworks. *Journal of the American Chemical Society* **2008**, *130*, 406–407.
- (175) Rappé, A. K.; Goddard III, W. A. Charge equilibration for molecular dynamics simulations. *The Journal of Physical Chemistry* **1991**, *95*, 3358–3363.
- (176) Mortier, W. J.; Ghosh, S. K.; Shankar, S. Electronegativity-equalization method for the calculation of atomic charges in molecules. *Journal of the American Chemical Society* **1986**, *108*, 4315–4320.
- (177) Ramachandran, S.; Lenz, T.; Skiff, W. et al. Toward an understanding of zeolite Y as a cracking catalyst with the use of periodic charge equilibration. *The Journal of Physical Chemistry* **1996**, *100*, 5898–5907.
- (178) Nistor, R. A.; Polihronov, J. G.; Müser, M. H. et al. A generalization of the charge equilibration method for non-metallic materials. *The Journal of Chemical Physics* **2006**, *125*, 094108.
- (179) Oda, A.; Takahashi, O. Parameter determination for the charge equilibration method including third-and fourth-order terms applied to non-metallic compounds. *Chemical Physics Letters* **2010**, *495*, 155–159.
- (180) Wilmer, C. E.; Kim, K. C.; Snurr, R. Q. An extended charge equilibration method. *The Journal of Physical Chemistry Letters* **2012**, *3*, 2506–2511.
- (181) Kadantsev, E. S.; Boyd, P. G.; Daff, T. D. et al. Fast and accurate electrostatics in metal organic frameworks with a robust charge equilibration parameterization for high-throughput virtual screening of gas adsorption. *The Journal of Physical Chemistry Letters* **2013**, *4*, 3056–3061.

- (182) Wells, B. A.; De Bruin-Dickason, C.; Chaffee, A. L. Charge equilibration based on atomic ionization in metal–organic frameworks. *The Journal of Physical Chemistry C* **2015**, *119*, 456–466.
- (183) Martin-Noble, G. C.; Reilley, D.; Rivas, L. M. et al. EQeq+ C: an empirical bond-order-corrected extended charge equilibration method. *Journal of Chemical Theory and Computation* **2015**, *11*, 3364–3374.
- (184) Collins, S. P.; Woo, T. K. Split-Charge Equilibration Parameters for Generating Rapid Partial Atomic Charges in Metal–Organic Frameworks and Porous Polymer Networks for High-Throughput Screening. *The Journal of Physical Chemistry C* **2017**, *121*, 903–910.
- (185) Campaña, C.; Mussard, B.; Woo, T. K. Electrostatic potential derived atomic charges for periodic systems using a modified error functional. *Journal of Chemical Theory and Computation* **2009**, *5*, 2866–2878.
- (186) Manz, T. A.; Sholl, D. S. Chemically meaningful atomic charges that reproduce the electrostatic potential in periodic and nonperiodic materials. *Journal of Chemical Theory and Computation* **2010**, *6*, 2455–2468.
- (187) Nazarian, D.; Camp, J. S.; Sholl, D. S. A comprehensive set of high-quality point charges for simulations of metal–organic frameworks. *Chemistry of Materials* **2016**, *28*, 785–793.
- (188) Coupry, D. E.; Addicoat, M. A.; Heine, T. Extension of the universal force field for metal–organic frameworks. *Journal of Chemical Theory and Computation* **2016**, *12*, 5215–5225.
- (189) Horike, S.; Shimomura, S.; Kitagawa, S. Soft porous crystals. *Nature Chemistry* **2009**, *1*, 695–704.
- (190) Barthel, S.; Alexandrov, E. V.; Proserpio, D. M. et al. Distinguishing Metal–Organic Frameworks. *Crystal Growth & Design* **2018**, *18*, 1738–1747.
- (191) Altintas, C.; Avci, G.; Daglar, H. et al. An Extensive Comparative Analysis of Two MOF Databases: High-Throughput Screening of Computation-Ready MOFs for CH₄ and H₂ Adsorption. *Journal of Materials Chemistry A* **2019**, *7*, 9593–9608.
- (192) Zarabadi-Poor, P.; Marek, R. Comment on “Database for CO₂ Separation Performances of MOFs Based on Computational Materials Screening”. *ACS Applied Materials & Interfaces* **2019**, *11*, 16261–16265.
- (193) Altintas, C.; Velioglu, S.; Keskin, S. Reply to Comment on “Database for CO₂ Separation Performances of MOFs Based on Computational Materials Screening”. *ACS Applied Materials & Interfaces* **2019**, *11*, 16266–16271.
- (194) Diercks, C. S.; Yaghi, O. M. The atom, the molecule, and the covalent organic framework. *Science* **2017**, *355*, eaal1585.
- (195) Hasell, T.; Cooper, A. I. Porous organic cages: soluble, modular and molecular pores. *Nature Reviews Materials* **2016**, *1*, 16053.
- (196) Tong, M.; Lan, Y.; Yang, Q. et al. Exploring the structure-property relationships of covalent organic frameworks for noble gas separations. *Chemical Engineering Science* **2017**, *168*, 456–464.
- (197) Tong, M.; Lan, Y.; Qin, Z. et al. Computation-Ready, Experimental Covalent Organic Framework for Methane Delivery: Screening and Material Design. *The Journal of Physical Chemistry C* **2018**, *122*, 13009–13016.
- (198) Miklitz, M.; Jiang, S.; Clowes, R. et al. Computational screening of porous organic molecules for xenon/krypton sep-

- aration. *The Journal of Physical Chemistry C* **2017**, *121*, 15211–15222.
- (199) Sturluson, A.; Huynh, M. T.; York, A. H. P. et al. Eigencages: Learning a Latent Space of Porous Cage Molecules. *ACS Central Science* **2018**, *4*, 1663–1676.
- (200) Baerlocher, C.; McCusker, L. B. Database of Zeolite Structures. <http://www.iza-structure.org/databases/>, Accessed 1 April 2019.
- (201) Barnett, B. R.; Gonzalez, M. I.; Long, J. R. Recent Progress Towards Light Hydrocarbon Separations Using Metal–Organic Frameworks. *Trends in Chemistry* **2019**,
- (202) Getman, R. B.; Miller, J. H.; Wang, K. et al. Metal alkoxide functionalization in metal-organic frameworks for enhanced ambient-temperature hydrogen storage. *The Journal of Physical Chemistry C* **2010**, *115*, 2066–2075.
- (203) Weston, M. H.; Farha, O. K.; Hauser, B. G. et al. Synthesis and metalation of catechol-functionalized porous organic polymers. *Chemistry of Materials* **2012**, *24*, 1292–1296.
- (204) Energy Information Administration, U. 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, M. Q.; Huang, H. S. 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, R. A.; Pacala, S. W.; Winebrake, J. J. et al. Greater focus needed on methane leakage from natural gas infrastructure. *Proceedings of the National Academy of Sciences* **2012**, *109*, 6435–6440.
- (207) Osborn, S. G.; Vengosh, A.; Warner, N. R. et al. Methane contamination of drinking water accompanying gas-well drilling and hydraulic fracturing. *Proceedings of the National Academy of Sciences* **2011**, *108*, 8172–8176.
- (208) Davis, S. C.; Diegel, S. W.; Boundy, R. G. Transportation Energy Data Book. Oak Ridge National Laboratory, <https://info.ornl.gov/sites/publications/files/pub31202.pdf>, Accessed 15 May 2019, 2009.
- (209) Hasan, M. M. F.; Zheng, A. M.; Karimi, I. A. Minimizing boil-off losses in liquefied natural gas transportation. *Industrial & Engineering Chemistry Research* **2009**, *48*, 9571–9580.
- (210) Makal, T. A.; Li, J.-R.; Lu, W. et al. Methane storage in advanced porous materials. *Chemical Society Reviews* **2012**, *41*, 7761–7779.
- (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, O. K. et al. A thermodynamic tank model for studying the effect of higher hydrocarbons on natural gas storage in metal–organic frameworks. *Energy & Environmental Science* **2015**, *8*, 1501–1510.
- (213) Wilmer, C. E.; Leaf, M.; Lee, C. Y. et al. Large-scale screening of hypothetical metal–organic frameworks. *Nature Chemistry* **2012**, *4*, 83–89.

- (214) Ma, S.; Sun, D.; Simmons, J. M. et al. Metal-organic framework from an anthracene derivative containing nanoscopic cages exhibiting high methane uptake. *Journal of the American Chemical Society* **2008**, *130*, 1012–1016.
- (215) Barthelet, K.; Marrot, J.; Riou, D. et al. A breathing hybrid organic-inorganic solid with very large pores and high magnetic characteristics. *Angewandte Chemie International Edition* **2002**, *41*, 281–284.
- (216) Lin, X.; Telepeni, I.; Blake, A. J. et al. High capacity hydrogen adsorption in Cu(II) tetracarboxylate framework materials: the role of pore size, ligand functionalization, and exposed metal sites. *Journal of the American Chemical Society* **2009**, *131*, 2159–2171.
- (217) Wilmer, C. E.; Farha, O. K.; Yildirim, T. et al. Gram-scale, high-yield synthesis of a robust metal-organic framework for storing methane and other gases. *Energy & Environmental Science* **2013**, *6*, 1158–1163.
- (218) Gómez-Gualdrón, D. A.; Gutov, O. V.; Krungleviciute, V. et al. Computational design of metal-organic frameworks based on stable zirconium building units for storage and delivery of methane. *Chemistry of Materials* **2014**, *26*, 5632–5639.
- (219) Wu, H.; Zhou, W.; Yildirim, T. High-capacity methane storage in metal-organic frameworks $M_2(dhtp)$: the important role of open metal sites. *Journal of the American Chemical Society* **2009**, *131*, 4995–5000.
- (220) Gándara, F.; Furukawa, H.; Lee, S. et al. High methane storage capacity in aluminum metal-organic frameworks. *Journal of the American Chemical Society* **2014**, *136*, 5271–5274.
- (221) Koroneos, C.; Dompros, A.; Roumbas, G. et al. Life cycle assessment of hydrogen fuel production processes. *International Journal of Hydrogen Energy* **2004**, *29*, 1443–1450.
- (222) Sakintuna, B.; Lamari-Darkrim, F.; Hirscher, M. Metal hydride materials for solid hydrogen storage: a review. *International Journal of Hydrogen Energy* **2007**, *32*, 1121–1140.
- (223) Yang, J.; Sudik, A.; Wolverton, C. et al. High capacity hydrogen storage materials: attributes for automotive applications and techniques for materials discovery. *Chemical Society Reviews* **2010**, *39*, 656–675.
- (224) of Energy Office of Energy Efficiency & Renewable Energy, U. D. DOE Technical Targets for Onboard Hydrogen Storage for Light-Duty Vehicles.
- (225) Murray, L. J.; Dincă, M.; Long, J. R. Hydrogen storage in metal-organic frameworks. *Chemical Society Reviews* **2009**, *38*, 1294–1314.
- (226) Gómez-Gualdrón, D. A.; Colón, Y. J.; Zhang, X. et al. Evaluating topologically diverse metal-organic frameworks for cryo-adsorbed hydrogen storage. *Energy & Environmental Science* **2016**, *9*, 3279–3289.
- (227) Ahmed, A.; Liu, Y.; Purewal, J. et al. Balancing gravimetric and volumetric hydrogen density in MOFs. *Energy & Environmental Science* **2017**, *10*, 2459–2471.
- (228) Kaye, S. S.; Dailly, A.; Yaghi, O. M. et al. Impact of preparation and handling on the hydrogen storage properties of $Zn_4O(1,4\text{-benzenedicarboxylate})_3$ (MOF-5). *Journal of the American Chemical Society* **2007**, *129*, 14176–14177.

- (229) Rowsell, J. L. C.; Yaghi, O. M. Effects of functionalization, catenation, and variation of the metal oxide and organic linking units on the low-pressure hydrogen adsorption properties of metal–organic frameworks. *Journal of the American Chemical Society* **2006**, *128*, 1304–1315.
- (230) Denysenko, D.; Grzywa, M.; Tonigold, M. et al. Elucidating Gating Effects for Hydrogen Sorption in MFU-4-Type Triazolate-Based Metal–Organic Frameworks Featuring Different Pore Sizes. *Chemistry - A European Journal* **2011**, *17*, 1837–1848.
- (231) García-Holley, P.; Schweitzer, B.; Islamoglu, T. et al. Benchmark Study of Hydrogen Storage in Metal–Organic Frameworks under Temperature and Pressure Swing Conditions. *ACS Energy Letters* **2018**, *3*, 748–754.
- (232) Ahmed, A.; Seth, S.; Purewal, J. et al. Exceptional hydrogen storage achieved by screening nearly half a million metal–organic frameworks. *Nature Communications* **2019**, *10*, 1568.
- (233) Martin, R. L.; Lin, L.-C.; Jariwala, K. et al. Mail-order metal–organic frameworks (MOFs): designing isorecticular MOF-5 analogues comprising commercially available organic molecules. *The Journal of Physical Chemistry C* **2013**, *117*, 12159–12167.
- (234) Prasad, T. K.; Suh, M. P. Control of Interpenetration and Gas-Sorption Properties of Metal–Organic Frameworks by a Simple Change in Ligand Design. *Chemistry - A European Journal* **2012**, *18*, 8673–8680.
- (235) Rankine, D.; Avellaneda, A.; Hill, M. R. et al. Control of framework interpenetration for in situ modified hydroxyl functionalised IRMOFs. *Chemical Communications* **2012**, *48*, 10328–10330.
- (236) Koh, K.; Van Oosterhout, J. D.; Roy, S. et al. Exceptional surface area from coordination copolymers derived from two linear linkers of differing lengths. *Chemical Science* **2012**, *3*, 2429–2432.
- (237) Stoller, J. K.; Panos, R. J.; Krachman, S. et al. Oxygen therapy for patients with COPD: current evidence and the long-term oxygen treatment trial. *Chest* **2010**, *138*, 179–187.
- (238) Raghu, G.; Collard, H. R.; Egan, J. J. et al. An official ATS/ERS/JRS/ALAT statement: idiopathic pulmonary fibrosis: evidence-based guidelines for diagnosis and management. *American Journal of Respiratory and Critical Care Medicine* **2011**, *183*, 788–824.
- (239) West, E. J. Error Analysis and Calibration Techniques to Determine the Measurement System Requirements for an LD Steelmaking Dynamic Control Application. *Measurement and Control* **1971**, *4*, T96–T108.
- (240) Widlund, D.; Medvedev, A.; Gyllenram, R. Towards model-based closed-loop control of the basic oxygen steelmaking process. *IFAC Proceedings Volumes* **1998**, *31*, 69–74.
- (241) McGovern, S. J.; Yeigh Jr, J. H. FCC Regeneration. 1983; US Patent 4,370,222.
- (242) DeCoste, J. B.; Weston, M. H.; Fuller, P. E. et al. Metal–organic frameworks for oxygen storage. *Angewandte Chemie International Edition* **2014**, *53*, 14092–14095.
- (243) Moghadam, P. Z.; Islamoglu, T.; Goswami, S. et al. Computer-aided discovery of a metal–organic framework with superior oxygen uptake. *Nature Communications* **2018**, *9*, 1378.
- (244) Dávila, M.; Riccardo, J. L.; Ramirez-Pastor, A. J. Exact statistical thermodynamics of alkane binary mixtures in

- zeolites: New interpretation of the adsorption preference reversal phenomenon from multisite-occupancy theory. *Chemical Physics Letters* **2009**, *477*, 402–405.
- (245) Sholl, D. S.; Lively, R. P. Seven chemical separations: to change the world: purifying mixtures without using heat would lower global energy use, emissions and pollution—and open up new routes to resources. *Nature* **2016**, *532*, 435–437.
- (246) Herm, Z. R.; Bloch, E. D.; Long, J. R. Hydrocarbon separations in metal–organic frameworks. *Chemistry of Materials* **2013**, *26*, 323–338.
- (247) Lenzen, M. Life cycle energy and greenhouse gas emissions of nuclear energy: A review. *Energy Conversion and Management* **2008**, *49*, 2178–2199.
- (248) Sood, D. D.; Patil, S. K. Chemistry of nuclear fuel reprocessing: Current status. *Journal of Radioanalytical and Nuclear Chemistry* **1996**, *203*, 547–573.
- (249) Soelberg, N. R.; Garn, T. G.; Greenhalgh, M. R. et al. Radioactive iodine and krypton control for nuclear fuel reprocessing facilities. *Science and Technology of Nuclear Installations* **2013**, *2013*, 1–12.
- (250) Banerjee, D.; Cairns, A. J.; Liu, J. et al. Potential of metal–organic frameworks for separation of xenon and krypton. *Accounts of Chemical Research* **2015**, *48*, 211–219.
- (251) Liu, J.; Fernandez, C. A.; Martin, P. F. et al. A two-column method for the separation of Kr and Xe from process off-gases. *Industrial & Engineering Chemistry Research* **2014**, *53*, 12893–12899.
- (252) Banerjee, D.; Simon, C. M.; Elsaidi, S. K. et al. Xenon Gas Separation and Storage Using Metal–Organic Frameworks. *Chem* **2018**, *4*, 466–494.
- (253) Banerjee, D.; Simon, C. M.; Plonka, A. M. et al. Metal–organic framework with optimally selective xenon adsorption and separation. *Nature Communications* **2016**, *7*, 11831.
- (254) Banerjee, D.; Zhang, Z.; Plonka, A. M. et al. A calcium coordination framework having permanent porosity and high CO₂/N₂ selectivity. *Crystal Growth & Design* **2012**, *12*, 2162–2165.
- (255) Sikora, B. J.; Wilmer, C. E.; Greenfield, M. L. et al. Thermodynamic analysis of Xe/Kr selectivity in over 13 7000 hypothetical metal–organic frameworks. *Chemical Science* **2012**, *3*, 2217–2223.
- (256) Ryan, P.; Farha, O. K.; Broadbelt, L. J. et al. Computational screening of metal–organic frameworks for xenon/krypton separation. *AIChE Journal* **2011**, *57*, 1759–1766.
- (257) Aroniadou-Anderjaska, V.; Figueiredo, T. H.; Aplan, J. P. et al. Primary brain targets of nerve agents: the role of the amygdala in comparison to the hippocampus. *NeuroToxicology* **2009**, *30*, 772–776.
- (258) Milatović, D.; Jokanović, M. In *Handbook of Toxicology of Chemical Warfare Agents*; Gupta, R. C., Ed.; Academic Press: San Diego, 2009; pp 985–996.
- (259) Agrawal, M.; Gallis, D. F. S.; Greathouse, J. A. et al. How Useful Are Common Simulants of Chemical Warfare Agents at Predicting Adsorption Behavior? *The Journal of Physical Chemistry C* **2018**, *122*, 26061–26069.
- (260) Figueiredo, T. H.; Aplan, J. P.; Braga, M. F. M. et al. Acute and long-term consequences of exposure to organophosphate nerve agents in humans. *Epilepsia* **2018**, *59*, 92–99.
- (261) Barea, E.; Montoro, C.; Navarro, J. A. R. Toxic gas removal – metal–organic frameworks for the capture and degradation of

- toxic gases and vapours. *Chemical Society Reviews* **2014**, *43*, 5419–5430.
- (262) DeCoste, J. B.; Peterson, G. W. Metal–organic frameworks for air purification of toxic chemicals. *Chemical Reviews* **2014**, *114*, 5695–5727.
- (263) Matito-Martos, I.; Moghadam, P. Z.; Li, A. et al. Discovery of an Optimal Porous Crystalline Material for the Capture of Chemical Warfare Agents. *Chemistry of Materials* **2018**, *30*, 4571–4579.
- (264) Küssgens, P.; Rose, M.; Senkovska, I. et al. Characterization of metal–organic frameworks by water adsorption. *Microporous and Mesoporous Materials* **2009**, *120*, 325–330.
- (265) Colombo, V.; Galli, S.; Choi, H. J. et al. High thermal and chemical stability in pyrazolate-bridged metal–organic frameworks with exposed metal sites. *Chemical Science* **2011**, *2*, 1311–1319.
- (266) Smil, V. Energy in the twentieth century: resources, conversions, costs, uses, and consequences. *Annual Review of Energy and the Environment* **2000**, *25*, 21–51.
- (267) Ciais, P.; Sabine, C.; Bala, G. et al. In *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*; Stocker, T., Qin, D., Plattner, G.-K. et al., Eds.; Cambridge University Press: Cambridge, United Kingdom and New York, NY, USA, 2013; Chapter 6, pp 465–570.
- (268) Collins, M.; Knutti, R.; Arblaster, J. et al. In *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*; Stocker, T., Qin, D., Plattner, G.-K. et al., Eds.; Cambridge University Press: Cambridge, United Kingdom and New York, NY, USA, 2013; Chapter 12, pp 1029–1136.
- (269) US Energy Information Administration, How much of U.S. carbon dioxide emissions are associated with electricity generation? 2019; <https://www.eia.gov/tools/faqs/faq.php?id=77&t=11>.
- (270) Lal, R. Carbon sequestration. *Philosophical Transactions of the Royal Society B: Biological Sciences* **2007**, *363*, 815–830.
- (271) Haszeldine, R. S. Carbon capture and storage: how green can black be? *Science* **2009**, *325*, 1647–1652.
- (272) Sumida, K.; Rogow, D. L.; Mason, J. A. et al. Carbon dioxide capture in metal–organic frameworks. *Chemical Reviews* **2012**, *112*, 724–781.
- (273) Whitley, D. A genetic algorithm tutorial. *Statistics and Computing* **1994**, *4*, 65–85.
- (274) Herm, Z. R.; Swisher, J. A.; Smit, B. et al. Metal–organic frameworks as adsorbents for hydrogen purification and precombustion carbon dioxide capture. *Journal of the American Chemical Society* **2011**, *133*, 5664–5667.
- (275) McDonald, T. M.; D’Alessandro, D. M.; Krishna, R. et al. Enhanced carbon dioxide capture upon incorporation of N, N-dimethylethylenediamine in the metal–organic framework CuBTri. *Chemical Science* **2011**, *2*, 2022–2028.
- (276) Cannella, W. J. *Kirk-Othmer Encyclopedia of Chemical Technology*; Wiley Online Library, 2000.
- (277) Yang, Y.; Bai, P.; Guo, X. Separation of xylene isomers: A review of recent advances in materials. *Industrial & Engineering Chemistry Research* **2017**, *56*, 14725–14753.
- (278) Gonzalez, M. I.; Kapelewski, M. T.; Bloch, E. D. et al. Separation of Xylene Isomers through Multiple Metal Site Interactions in Metal–Organic Frameworks. *Journal of the American Chemical Society* **2018**, *140*, 3412–3422.

- (279) Minceva, M.; Rodrigues, A. E. Understanding and revamping of industrial scale SMB units for p-xylene separation. *AIChE Journal* **2007**, *53*, 138–149.
- (280) Moïse, J.-C.; Bellat, J.-P. Effect of preadsorbed water on the adsorption of p-xylene and m-xylene mixtures on BaX and BaY zeolites. *The Journal of Physical Chemistry B* **2005**, *109*, 17239–17244.
- (281) Gee, J. A.; Zhang, K.; Bhattacharyya, S. et al. Computational identification and experimental evaluation of metal–organic frameworks for xylene enrichment. *The Journal of Physical Chemistry C* **2016**, *120*, 12075–12082.
- (282) LeCun, Y.; Cortes, C.; Burges, C. J. C. The MNIST database of handwritten digits. <http://yann.lecun.com/exdb/mnist/>, Accessed 1 April 2019, 1998.
- (283) Wan, L.; Zeiler, M.; Zhang, S. et al. Regularization of neural networks using dropconnect. International Conference on Machine Learning. 2013; pp 1058–1066.
- (284) Krizhevsky, A.; 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.; Lanning, S. The Netflix Prize. Proceedings of the KDD Cup Workshop 2007. New York, 2007; pp 3–6.
- (286) Bell, R. M.; Koren, Y. Lessons from the Netflix prize challenge. *ACM SIGKDD Explorations Newsletter* **2007**, *9*, 75–79.
- (287) Koren, Y.; Bell, R.; Volinsky, C. Matrix factorization techniques for recommender systems. *Computer* **2009**, *42*, 30–37.
- (288) Koren, Y. The bellkor solution to the netflix grand prize. *Netflix prize documentation* **2009**, *81*, 1–10.
- (289) Russakovsky, O.; Deng, J.; Su, H. et al. ImageNet large scale visual recognition challenge. *International Journal of Computer Vision* **2015**, *115*, 211–252.
- (290) Price, S. L. Control and prediction of the organic solid state: a challenge to theory and experiment. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **2018**, *474*, 20180351.
- (291) Reilly, A. M.; Cooper, R. I.; Adjiman, C. S. et al. Report on the sixth blind test of organic crystal structure prediction methods. *Acta Crystallographica Section B: Structural Science, Crystal Engineering and Materials* **2016**, *72*, 439–459.
- (292) <https://adsorption.nist.gov/factlab>, Accessed 5 February 2019.
- (293) Nguyen, H. G. T.; Espinal, L.; van Zee, R. D. et al. A reference high-pressure CO₂ adsorption isotherm for ammonium ZSM-5 zeolite: Results of an interlaboratory study. *Adsorption* **2018**, *24*, 531–539.
- (294) Siderius, D. W., Shen, V. K., Johnson III, R. D. et al., Eds. *NIST/ARPA-E Database of Novel and Emerging Adsorbent Materials*; National Institute of Standards and Technology: Gaithersburg, MD, 20899, 2014; <https://dx.doi.org/10.18434/T43882>, Accessed 14 December 2018.
- (295) Heller, S. R.; McNaught, A.; Pletnev, I. et al. InChI, the IUPAC International Chemical Identifier. *Journal of Cheminformatics* **2015**, *7*, 23.
- (296) Britt, D.; Tranchemontagne, D.; Yaghi, O. M. Metal–organic frameworks with high capacity and selectivity for harmful gases. *Proceedings of the National Academy of Sciences* **2008**, *105*, 11623–11627.

- (297) <https://www.sigmaaldrich.com/catalog/product/aldrich/688614?lang=en®ion=US>, Accessed 5 February 2019.
- (298) Siderius, D., Shen, V., Eds. *NIST Registry of Adsorbent Materials*; National Institute of Standards and Technology: Gaithersburg, MD, 20899, 2017; <https://dx.doi.org/10.18434/t4/1502644>, Accessed 14 December 2018.
- (299) Park, J.; Howe, J. D.; Sholl, D. S. How Reproducible Are Isotherm Measurements in Metal–Organic Frameworks? *Chemistry of Materials* **2017**, *29*, 10487–10495.
- (300) Simon, C. M.; Smit, B.; Haranczyk, M. pyIAST: Ideal adsorbed solution theory (IAST) Python package. *Computer Physics Communications* **2016**, *200*, 364–380.
- (301) Simon, C. M.; Smit, B.; Haranczyk, M. Documentation for pyIAST. <https://pyiastr.readthedocs.io>, Accessed 1 February 2019.
- (302) Colón, Y. J.; Fairen-Jimenez, D.; Wilmer, C. E. et al. High-Throughput Screening of Porous Crystalline Materials for Hydrogen Storage Capacity near Room Temperature. *The Journal of Physical Chemistry C* **2014**, *118*, 5383–5389.
- (303) Gómez-Gualdrón, D. A.; Wilmer, C. E.; Farha, O. K. et al. Exploring the Limits of Methane Storage and Delivery in Nanoporous Materials. *The Journal of Physical Chemistry C* **2014**, *118*, 6941–6951.
- (304) <http://www.fluidproperties.org>, Accessed 25 February 2019.
- (305) Ross, R. B.; Ahmad, R.; Brennan, J. K. et al. The Seventh Industrial Fluid Properties Simulation Challenge. *Fluid Phase Equil.* **2014**, *366*, 136–140.
- (306) Ross, R. B.; Brennan, J. K.; Frankel, K. A. et al. Perfluorohexane Adsorption in BCR-704 Faujasite Zeolite Benchmark Studies for the Seventh Industrial Fluid Properties Simulation Challenge. *Fluid Phase Equil.* **2014**, *366*, 141–145.
- (307) Schultz, N. E.; Riaz, A.; Brennan, J. K. et al. The Eighth Industrial Fluid Properties Simulation Challenge. *Adsorption Sci. Tech.* **2016**, *34*, 3–12.
- (308) Ross, R. B.; Aeschliman, D. B.; Ahmad, R. et al. Adsorption, X-ray Diffraction, Photoelectron, and Atomic Emission Spectroscopy Benchmark Studies for the Eighth Industrial Fluid Properties Simulation Challenge. *Adsorption Sci. Tech.* **2016**, *34*, 13–41.
- (309) Rouquerol, F.; Rouquerol, J.; Sing, K. *Adsorption by Powders and Porous Solids*; Academic Press: London, 1999.
- (310) <https://adsorption.nist.gov/isodb/index.php#apis>, Accessed 22 February 2019.
- (311) Hall, S. R.; Allen, F. H.; Brown, I. D. The crystallographic information file (CIF): a new standard archive file for crystallography. *Acta Crystallographica Section A: Foundations and Devices* **1991**, *A47*, 655–685.
- (312) Brown, I. D.; McMahon, B. CIF: the computer language of crystallography. *Acta Crystallographica Section B: Structural Science, Crystal Engineering and Materials* **2002**, *58*, 317–324.
- (313) Curtarolo, S.; Hart, G. L.; Nardelli, M. B. et al. The high-throughput highway to computational materials design. *Nature Materials* **2013**, *12*, 191.
- (314) Pyzer-Knapp, E. O.; Suh, C.; Gómez-Bombarelli, R. et al. What Is High-Throughput Virtual Screening? A Perspective from Organic Materials Discov-

ery. *Annual Review of Materials Research* **2015**, *45*, 195–216.

- (315) Gómez-Bombarelli, R.; Aguilera-Iparraguirre, J.; Hirzel, T. D. et al. Design of efficient molecular organic light-emitting diodes by a high-throughput virtual screening and experimental approach. *Nature Materials* **2016**, *15*, 1120–1127.
- (316) Ceder, G. Opportunities and challenges for first-principles materials design and applications to Li battery materials. *MRS Bulletin* **2010**, *35*, 693–701.
- (317) Hachmann, J.; Olivares-Amaya, R.; Atahan-Evrenk, S. et al. The Harvard clean energy project: large-scale computational screening and design of organic photovoltaics on the world community grid. *The Journal of Physical Chemistry Letters* **2011**, *2*, 2241–2251.
- (318) Nørskov, J. K.; Bligaard, T.; Rossmeisl, J. et al. Towards the computational design of solid catalysts. *Nature Chemistry* **2009**, *1*, 37.
- (319) Greeley, J.; Jaramillo, T. F.; Bonde, J. et al. Computational high-throughput screening of electrocatalytic materials for hydrogen evolution. *Nature Materials* **2006**, *5*, 909–913.
- (320) Strasser, P.; Fan, Q.; Devenney, M. et al. High throughput experimental and theoretical predictive screening of materials—A comparative study of search strategies for new fuel cell anode catalysts. *The Journal of Physical Chemistry B* **2003**, *107*, 11013–11021.
- (321) Kitchen, D. B.; Decornez, H.; Furr, J. R. et al. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nature Reviews Drug Discovery* **2004**, *3*, 935–949.
- (322) Xu, J.; Hagler, A. Chemoinformatics and Drug Discovery. *Molecules* **2002**, *7*, 566–600.
- (323) Flaig, R. W.; Osborn Popp, T. M.; Fracaroli, A. M. et al. The chemistry of CO₂ capture in an amine-functionalized metal–organic framework under dry and humid conditions. *Journal of the American Chemical Society* **2017**, *139*, 12125–12128.
- (324) Bachman, J.; Reed, D. A.; Kapelewski, M. T. et al. Enabling Alternative Ethylene Production through Its Selective Adsorption in the Metal–Organic Framework Mn₂(*m*-dobdc). *Energy & Environmental Science* **2018**, *11*, 2423–2431.
- (325) Nguyen, J. G.; Cohen, S. M. Moisture-resistant and superhydrophobic metal–organic frameworks obtained via post-synthetic modification. *Journal of the American Chemical Society* **2010**, *132*, 4560–4561.
- (326) Poloni, R.; Lee, K.; Berger, R. F. et al. Understanding trends in CO₂ adsorption in metal–organic frameworks with open-metal sites. *The Journal of Physical Chemistry Letters* **2014**, *5*, 861–865.
- (327) Coudert, F.-X. Responsive metal–organic frameworks and framework materials: under pressure, taking the heat, in the spotlight, with friends. *Chemistry of Materials* **2015**, *27*, 1905–1916.
- (328) Wieme, J.; Lejaeghere, K.; Kresse, G. et al. Tuning the balance between dispersion and entropy to design temperature-responsive flexible metal–organic frameworks. *Nature Communications* **2018**, *9*, 4899.
- (329) Coudert, F.-X. Reproducible Research in Computational Chemistry of Materials. *Chemistry of Materials* **2017**, *29*, 2615–2617.
- (330) Greathouse, J. A.; Kinniburgh, T. L.; Allendorf, M. D. Adsorption and separation of noble gases by IRMOF-1: grand

- canonical Monte Carlo simulations. *Industrial & Engineering Chemistry Research* **2009**, *48*, 3425–3431.
- (331) Boutin, A.; Springuel-Huet, M.-A.; Nossov, A. et al. Breathing Transitions in MIL-53 (Al) Metal–Organic Framework Upon Xenon Adsorption. *Angewandte Chemie International Edition* **2009**, *48*, 8464–8467.
- (332) Gonzalez-Nelson, A.; Coudert, F.-X.; van der Veen, M. Rotational Dynamics of Linkers in Metal–Organic Frameworks. *Nanomaterials* **2019**, *9*, 330.
- (333) Catalano, L.; Naumov, P. Exploiting rotational motion in molecular crystals. *CrystEngComm* **2018**, *20*, 5872–5883.
- (334) Witman, M.; Ling, S.; Jawahery, S. et al. The influence of intrinsic framework flexibility on adsorption in nanoporous materials. *Journal of the American Chemical Society* **2017**, *139*, 5547–5557.
- (335) Krause, S.; Evans, J. D.; Bon, V. et al. Towards General Network Architecture Design Criteria for Negative Gas Adsorption Transitions in Ultraporous Frameworks. *ChemRxiv* **2019**, <https://doi.org/10.26434/chemrxiv.7796543>.
- (336) Evans, J. D.; Bocquet, L.; Coudert, F.-X. Origins of negative gas adsorption. *Chem* **2016**, *1*, 873–886.
- (337) Krause, S.; Bon, V.; Senkovska, I. et al. A pressure-amplifying framework material with negative gas adsorption transitions. *Nature* **2016**, *532*, 348–352.
- (338) Lee, Y.; Poloni, R.; Kim, J. Probing gas adsorption in MOFs using an efficient ab initio widom insertion Monte Carlo method. *Journal of Computational Chemistry* **2016**, *37*, 2808–2815.
- (339) Fetisov, E. O.; Shah, M. S.; Long, J. R. et al. First principles Monte Carlo simulations of unary and binary adsorption: CO₂, N₂, and H₂O in Mg-MOF-74. *Chemical Communications* **2018**, *54*, 10816–10819.
- (340) Chen, L.; Grajciar, L.; Nachtigall, P. et al. Accurate prediction of methane adsorption in a metal–organic framework with unsaturated metal sites by direct implementation of an ab initio derived potential energy surface in GCMC simulation. *The Journal of Physical Chemistry C* **2011**, *115*, 23074–23080.
- (341) Cornell, W. D.; Cieplak, P.; Bayly, C. I. et al. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *Journal of the American Chemical Society* **1995**, *117*, 5179–5197.
- (342) Jorgensen, W. L.; Tirado-Rives, J. The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *Journal of the American Chemical Society* **1988**, *110*, 1657–1666.
- (343) McDaniel, J. G.; Li, S.; Tylianakis, E. et al. Evaluation of force field performance for high-throughput screening of gas uptake in metal–organic frameworks. *The Journal of Physical Chemistry C* **2015**, *119*, 3143–3152.
- (344) Jeziorski, B.; Moszynski, R.; Szalewicz, K. Perturbation theory approach to intermolecular potential energy surfaces of van der Waals complexes. *Chemical Reviews* **1994**, *94*, 1887–1930.
- (345) Dokur, D.; Keskin, S. Effects of Force Field Selection on the Computational Ranking of MOFs for CO₂ Separations. *Industrial & Engineering Chemistry Research* **2018**, *57*, 2298–2309.
- (346) Behler, J.; Parrinello, M. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Physical Review Letters* **2007**, *98*, 146401.

- (347) Smith, J. S.; Isayev, O.; Roitberg, A. E. ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost. *Chemical Science* **2017**, *8*, 3192–3203.
- (348) Behler, J. Neural network potential-energy surfaces in chemistry: a tool for large-scale simulations. *Physical Chemistry Chemical Physics* **2011**, *13*, 17930–17955.
- (349) Schütt, K. T.; Arbabzadah, F.; Chmiela, S. et al. Quantum-chemical insights from deep tensor neural networks. *Nature Communications* **2017**, *8*, 13890.
- (350) Bartók, A. P.; Payne, M. C.; Kondor, R. et al. Gaussian approximation potentials: The accuracy of quantum mechanics, without the electrons. *Physical Review Letters* **2010**, *104*, 136403.
- (351) Hornik, K.; Stinchcombe, M.; White, H. Multilayer feedforward networks are universal approximators. *Neural Networks* **1989**, *2*, 359–366.
- (352) Lowell, S.; Shields, J. E.; Thomas, M. A. et al. *Characterization of Porous Solids and Powders: Surface Area, Pore Size and Density*; Kluwer: Boston, 2004.
- (353) Xu, Q.; Zhong, C. A general approach for estimating framework charges in metal-organic frameworks. *The Journal of Physical Chemistry C* **2010**, *114*, 5035–5042.
- (354) Argueta, E.; Shaji, J.; Gopalan, A. et al. Molecular Building Block-Based Electronic Charges for High-Throughput Screening of Metal–Organic Frameworks for Adsorption Applications. *Journal of Chemical Theory and Computation* **2017**, *14*, 365–376.
- (355) Breneman, C. M.; Wiberg, K. B. Determining atom-centered monopoles from molecular electrostatic potentials. The need for high sampling density in formamide conformational analysis. *Journal of Computational Chemistry* **1990**, *11*, 361–373.
- (356) Bennett, T. D.; Cheetham, A. K.; Fuchs, A. H. et al. Interplay between defects, disorder and flexibility in metal–organic frameworks. *Nature Chemistry* **2017**, *9*, 11–16.
- (357) Sholl, D. S.; Lively, R. P. Defects in metal–organic frameworks: challenge or opportunity? *The Journal of Physical Chemistry Letters* **2015**, *6*, 3437–3444.
- (358) Cavka, J. H.; Jakobsen, S.; Olsbye, U. et al. A new zirconium inorganic building brick forming metal organic frameworks with exceptional stability. *Journal of the American Chemical Society* **2008**, *130*, 13850–13851.
- (359) Wu, H.; Chua, Y. S.; Krungleviciute, V. et al. Unusual and highly tunable missing-linker defects in zirconium metal–organic framework UiO-66 and their important effects on gas adsorption. *Journal of the American Chemical Society* **2013**, *135*, 10525–10532.
- (360) Ghosh, P.; Colón, Y. J.; Snurr, R. Q. Water adsorption in UiO-66: the importance of defects. *Chemical Communications* **2014**, *50*, 11329–11331.
- (361) Thornton, A. W.; Babarao, R.; Jain, A. et al. Defects in metal–organic frameworks: a compromise between adsorption and stability? *Dalton Transactions* **2016**, *45*, 4352–4359.
- (362) Bristow, J. K.; Svane, K. L.; Tiana, D. et al. Free Energy of Ligand Removal in the Metal–Organic Framework UiO-66. *The Journal of Physical Chemistry C* **2016**, *120*, 9276–9281.
- (363) Zhang, C.; Han, C.; Sholl, D. S. et al. Computational characterization of defects in metal–organic frameworks: Spontaneous and water-induced point defects in ZIF-8. *The Journal of Physical Chemistry Letters* **2016**, *7*, 459–464.

- (364) Han, C.; Verploegh, R. J.; Sholl, D. S. Assessing the impact of point defects on molecular diffusion in ZIF-8 using molecular simulations. *The Journal of Physical Chemistry Letters* **2018**, *9*, 4037–4044.
- (365) Zunger, A. Beware of plausible predictions of fantasy materials. *Nature* **2019**, *566*, 447–449.
- (366) Cantu, D. C.; McGrail, B. P.; Glezakou, V.-A. Formation Mechanism of the Secondary Building Unit in a Chromium Terephthalate Metal–Organic Framework. *Chemistry of Materials* **2014**, *26*, 6401–6409.
- (367) Yoneya, M.; Tsuzuki, S.; Aoyagi, M. Simulation of metal–organic framework self-assembly. *Physical Chemistry Chemical Physics* **2015**, *17*, 8649–8652.
- (368) Biswal, D.; Kusalik, P. G. Probing Molecular Mechanisms of Self-Assembly in Metal–Organic Frameworks. *ACS Nano* **2017**, *11*, 258–268.
- (369) Biswal, D.; Kusalik, P. G. Molecular simulations of self-assembly processes in metal–organic frameworks: Model dependence. *The Journal of Chemical Physics* **2017**, *147*, 044702.
- (370) Colón, Y. J.; Guo, A. Z.; Antony, L. W. et al. Free energy of metal–organic framework self-assembly. *The Journal of Chemical Physics* **2019**, *150*, 104502.
- (371) Kollias, L.; Cantu, D. C.; Tubbs, M. A. et al. Molecular Level Understanding of the Free Energy Landscape in Early Stages of Metal–Organic Framework Nucleation. *Journal of the American Chemical Society* **2019**, *141*, 6073–6081.
- (372) Greathouse, J. A.; Allendorf, M. D. The interaction of water with MOF-5 simulated by molecular dynamics. *Journal of the American Chemical Society* **2006**, *128*, 10678–10679.
- (373) Bellarosa, L.; Castillo, J. M.; Vlugt, T. et al. On the Mechanism Behind the Instability of Isoreticular Metal–Organic Frameworks (IRMOFs) in Humid Environments. *Chemistry - A European Journal* **2012**, *18*, 12260–12266.
- (374) Han, C.; Zhang, C.; Tyminska, N. et al. Insights into the stability of zeolitic imidazolate frameworks in humid acidic environments from first-principles calculations. *The Journal of Physical Chemistry C* **2018**, *122*, 4339–4348.
- (375) Rogge, S. M. J.; Waroquier, M.; Van Speybroeck, V. Reliably modeling the mechanical stability of rigid and flexible metal–organic frameworks. *Accounts of Chemical Research* **2018**, *51*, 138–148.
- (376) Rajagopalan, A. K.; Avila, A. M.; Rajendran, A. Do adsorbent screening metrics predict process performance? A process optimisation based study for post-combustion capture of CO₂. *International Journal of Greenhouse Gas Control* **2016**, *46*, 76–85.
- (377) First, E. L.; Hasan, M. M. F.; Floudas, C. A. Discovery of novel zeolites for natural gas purification through combined material screening and process optimization. *AIChE Journal* **2014**, *60*, 1767–1785.
- (378) Farmahini, A. H.; Krishnamurthy, S.; Friedrich, D. et al. From crystal to adsorption column: challenges in multiscale computational screening of materials for adsorption separation processes. *Industrial & Engineering Chemistry Research* **2018**, *57*, 15491–15511.
- (379) Pullumbi, P.; Brandani, F.; Brandani, S. Gas separation by adsorption: technological drivers and opportunities for improvement. *Current Opinion in Chemical Engineering* **2019**, *24*, 131–142.
- (380) Krishna, R. Methodologies for screening and selection of crystalline microporous

materials in mixture separations. *Separation and Purification Technology* **2018**, *194*, 281–300.

- (381) Leperi, K.; Chung, Y. G.; You, F. et al. Development of a general evaluation metric for rapid screening of adsorbent materials for post-combustion CO₂ capture. *ACS Sustainable Chemistry & Engineering* **2019**, *7*, 11529–11539.