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

Arni Surluson,[†] Melanie Huyhn,[†] Alec Kaija,[‡] Caleb Laird,[†] Sunghyun Yoon,[¶] Feier Hou,[§] Zhenxing Feng,[†] Chris 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*

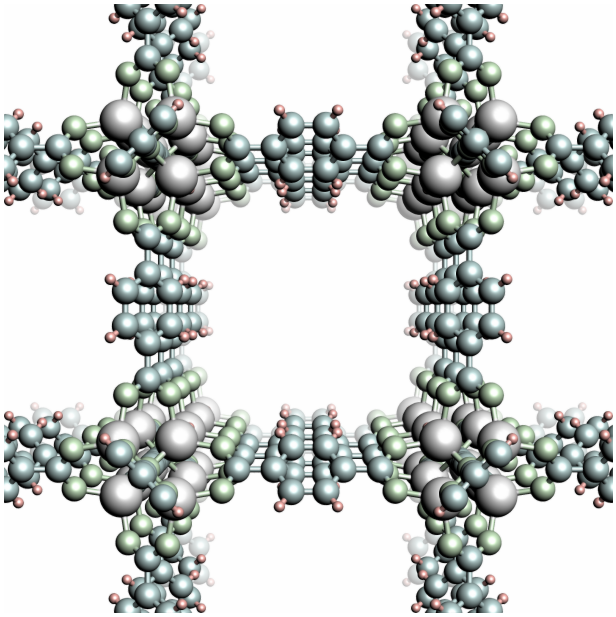
[¶]*Pusan National University. School of Chemical and Biomolecular Engineering, Pusan National University, Busan, Korea (South)*

[§]*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 lucidly impacted 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 for molecular simulations, are a platform for computational materials discovery. We pontificate how to orient research efforts to routinize the computational discovery of MOFs for adsorption-based engineering applications.

everywhere we say a MOF name, we shld make it bold. **MOF-5** not MOF-5

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 m^2/g).[2] Their consequent gas adsorption properties lend them applications in storing,[3] separating,[4] and sensing[5] gases. Since the first MOF was reported by Omar Yaghi in 1999,[6] a few MOF-based products have appeared on the market,[7] including for safe sub-atmospheric storage of toxic gases (NuMat Technologies) and carbon dioxide capture in submarines (Mosaic Materials).

MOFs, heralded as designer materials,[8] are 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 arduously[11]) finding the synthetic conditions to yield a pre-determined, extended network structure is coined as *reticular chemistry*.[12,13]

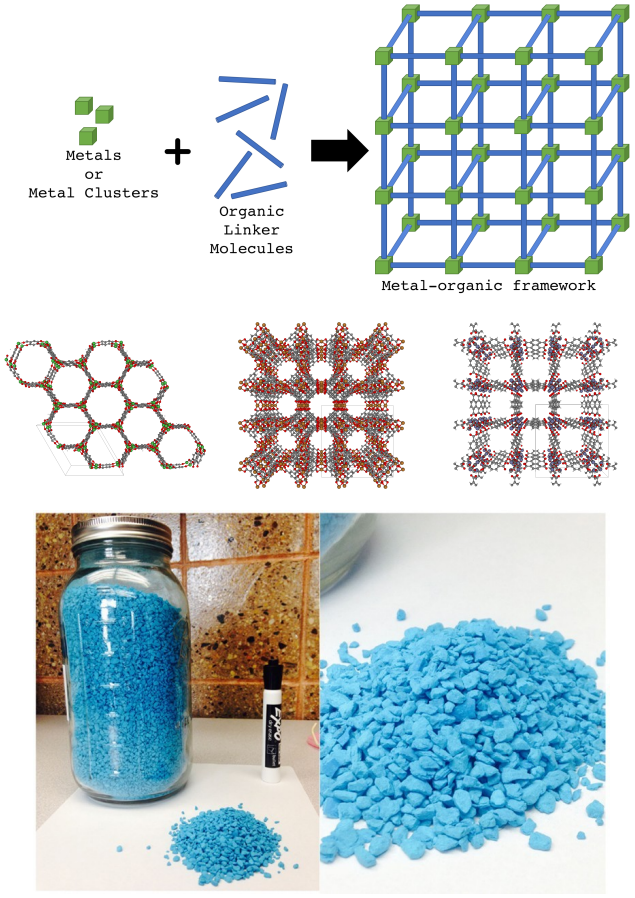


Figure 1: The chemistry of metal-organic frameworks (MOFs) is modular and highly tunable, affording a vast chemical space of crystal structures in which to search for materials exhibiting an optimal adsorption property. (top) MOFs are composed of metal nodes or clusters coordinated to organic linker molecules to form a crystalline, porous framework. By changing the linkers and metal clusters, we can obtain millions of possible materials. (middle) As examples, shown are crystal structures of Ni-MOF-74,[14] HKUST-1,[15] IRMOF-1.[12] (bottom) As an illustration of what a MOF looks like in practice, shown is HKUST-1[15] from Ref. 16.

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.[17] 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;[18] operating at room temperature via a pressure swing between 65 bar and 5 bar, the usable capacity is the equilibrium methane adsorption after filling up at the fuel station (65 bar) minus that retained when the tank sustains insufficient pressure to drive flow to the engine (5 bar)[1]. For the near-term deployment of a MOF in an adsorbed natural gas fuel tank, a coveted aim is to reliably identify, among the 80 000 experimentally synthesized, which MOF exhibits the maximal methane usable capacity. Notable, 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.[19–21]

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[2] to exhibit a desired (optimal) adsorption property (e.g. maximal methane usable capacity) through high-throughput computational screenings. The simple idea, encapsulated in the Materials Project,[22] 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[23–25] by highlighting how large, open data sets of MOF crystal structures provide a platform for the computational identification of existing MOFs for gas storage, separation, and sensing.

We begin in Sec. 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. Finally, in Sec. 6, we survey high-throughput computational screenings of MOF crystal structures for gas adsorption and separation applications that have directly and lucidly 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, classical force fields, whose interatomic potential parameters are fit to experimental data or *ab*

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

[2] We specify the MOF to be existing as opposed to hypothetical or conceived for the sake of near-term as opposed to long-term deployment; the synthesis protocols– and, perhaps, activation procedures and information about stability– of existing MOFs already reported in the literature can expedite deployment.

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[26] and DREIDING[27] 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)[28] 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 polar CO_2[29] but not nonpolar CH_4[28]). There exists a hierarchy of methods for assigning point charges to MOFs to model their electrostatic interaction with a polar adsorbate molecule[30] (see Sec. 9.3), and a database of experimental MOF crystal structures with high-quality point charges assigned can be downloaded (see Sec. 5.1.1).

The many-body potential energy description is then used as input to Monte Carlo (MC) or Molecular Dynamics (MD) simulations[31] to predict properties of the MOF crystal, such as adsorption isotherms and diffusion coefficients, respectively.

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 ensemble[32]) 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 coordination environments encountered in MOFs. However, progress is underway in the development of accurate intrahost force fields for MOFs[33–35] and of efficient algorithms to simulate their flexibility.[36]

Therefore, 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.[37] 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).[38] Second, there can be variation in the experimentally determined lattice constants of a given MOF among different research groups, perhaps due to different synthesis conditions.[39] 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,[40] 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

Consider a MOF crystal immersed in a bath of (mixed) gas at temperature T and chemical potential $\boldsymbol{\mu}$ (linked to pressure $\mathbf{P}$ via an equation of state for the mixed gas). At thermodynamic equilibrium, the grand-canonical statistical mechanical ensemble[41] governs the probability of every possible configuration of gas inside a MOF crystal of fixed volume; the probability of each microstate depends on its potential en-

ergy and particle number of each species. The osmotic ensemble[42] governs the case where the unit cell is flexible and allows decoupling of the chemical potential of the gas from the mechanical stress imposed on the crystal. Equipped with a force field and a MOF crystal structure, therefore, we can conduct Markov Chain Monte Carlo simulations[43] of a statistical ensemble to simulate equilibrium gas adsorption in MOFs. Properties such as the expected number of adsorbed particles of each species (at the given condition) are easily computed from the simulation. On the other hand, Molecular dynamics (MD) simulations calculate dynamic properties such as diffusion coefficients of gases inside MOFs. In an MD simulation, Newton's equations of motion govern the propagation of the system forward in time. See reviews from the group of Coudert[32,44,45] on computational methods for predicting many different properties of nanoporous materials and Smit and Frenkel for a textbook on molecular simulations.[31]

2.4 High-throughput computational screening

Equipped with a force field, a set of MOF crystal structures, and software to conduct a grand-canonical Monte Carlo simulation,[46–48] the *brute-force computational screening strategy* is to loop over all material candidates and simulate gas adsorption in each material:

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

After all simulations have finished, we sort the materials by the desired property obtained by the simulation and, voilá, shortlist the top few for experimental testing[3]. 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[49,50]) 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 trivialization 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.[51] found that MOFs exhibiting the highest CO_2/N_2 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, we 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.[52] 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, our work[52] held MOFs rigid during the simulation and neglected the case where a flexible MOF can collapse and expel residual gas at the discharge pressure, as MOF Co(bdp),[53] which currently boasts the highest methane 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,[54] modeling polarizability of carbon dioxide by open-metal sites in MOFs,[55] modeling water adsorption in MOFs, which requires many MC samples[56]), high pressures (where many molecules are typically present in the system), and treatment of structural flexibility []. Two methods have emerged to circumvent conducting simulations in *all* material candidates in a brute-force screening, thereby saving computational expense. The first method is to use statistical machine learning[57] to build a regression or classification model of an adsorption property with structural[58] and/or chem-

[3] 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.

ical[59] descriptors 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 structural descriptors– are cheap to compute; so a trained model can be used to cheaply predict adsorption on the basis of structural descriptors. Simulated properties in only a (diverse) *sub*set of material candidates can be 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) structural/chemical descriptors. For example, chemical, structural (surface area, void fraction, largest cavity diameter, etc.), and/or potential energy-based descriptors were used to train statistical models to screen MOFs for CO_2 adsorption,[60] Xe/Kr selectivity,[61] hydrogen adsorption,[62] and methane adsorption,[63] conducting simulations on only a subset of training MOFs. See Ref. 64 for a review on Quantitative Structure Property Relationship (QSPR) modeling in materials science in general. The second method to avert a brute-force screening is to employ an evolutionary algorithm to search for and explore regions of chemical space where the most performant MOFs lie. Such an evolutionary algorithm evolves in stages (generations). The simulated performance of MOFs in the first generation is used to select (evolve) the next generation of MOFs in which to simulate gas adsorption, with the aim of reaching a summit of adsorption performance. For example, Bao et al.[65] used an evolutionary algorithm to mutate the chemistry of MOF linkers to arrive at MOFs with high methane usable capacity. Chung et al.[66] used a genetic algorithm to search for MOFs for CO_2 capture.

A lesson from molecular biology

CRISPR-Cas9 technology[67] is revolutionizing molecular biology by enabling the facile, precise, and cost-effective editing of genomes.[67] 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.[68] The CRISPR-Cas9 technology originates from an adaptive immune system discovered in bacteria.[69] The fascinating story by which CRISPR was recognized as a bacterial immune system, outlined by Lander,[70] bestows useful lessons on materials science, including the importance of "hypothesis-free science".

By 2000, Mojica et al.[71] 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 pairs, $\approx$ 24 bp to 40 bp in length, flanking both sides of a unique spacer sequence of roughly consistent length (20 bp to 58 bp). These were descriptively coined clustered regularly interspaced short palindromic repeats (CRISPR).[72] Mysteriously, the biological role (or lack thereof) of evolutionary conserved (within species[72]) CRISPRs was unknown.

Five years later, Mojica et al.[73] published evidence that CRISPR is related to a microbial immune system, conferring resistance to e.g. bacteriophages. The link was made by searching databases of DNA molecules for matches of $\approx$ 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.

Lander[70] 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– the curation of databases of computation-ready nanoporous crystal structures and ad-

sorption 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;[74] 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, however, lucidly demonstrates the impact of open, computation-ready databases of MOF structures.

3 Determining crystal structures of MOFs

We now discuss how the crystal structure of a newly synthesized metal-organic framework is experimentally determined.

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.[75] 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– 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 the crystallite size and defects.[76,77]

Can we say sth abt how wave-length of X-rays appropriate for spatial scales of MOFs? or too basic?

The crystal structure of a MOF can be determined from XRD data through a process called Rietveld refinement. 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.[77] 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:

$$R := \frac{\sum |I_{\text{exp}} - I_{\text{sim}}|}{\sum |I_{\text{exp}}|}, \tag{1}$$

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 %;[78] 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.[79] During MOF synthesis, high temperature and pressure could introduce network disorder such as twinning[80] and pseudo-symmetry.[81] Relatedly, interpenetration of nets in MOFs[82] 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.[83] Even if the solvent can

be evacuated, the MOF may still adsorb moisture rapidly from the air; keeping the MOF in an inert (dry) atmosphere during experiments may be beneficial. Solvent masking (typically by software such as SQUEEZE[84]) is an effective way to remove the contribution by the solvent from the XRD pattern. (5) Because XRD is an ensemble measurement, the specific tilt of certain ligands in the MOF structure could be averaged out and thus undetermined by refinement.[85,86]

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 c.a. 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.[76,87,88]

Finally, *in situ* XRD has become popular during the synthesis process[89,90] to study the growth kinetics and mechanism[91–93] and during gas adsorption experiments to determine which adsorption sites are populated at different pressures[94,95] and investigate gas-induced structural changes.[96,97]

4 The Cambridge Structural Database

The Cambridge Structural Database (CSD)[98] 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.[99] 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,[100] and a Python API.[99] 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,[101] 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 70 000.[10]

There are a few issues with MOF structures in the CSD that preclude computational scientists from simply downloading all MOF crystal structures from the CSD, then conducting molecular simulations in a high-throughput computational screening. That is, not all MOF structures in the CSD are *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 most common method to search for MOFs in the CSD is to use a chemical bond criteria;[10,102–104] 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.[100] Beginning 2016, the Cambridge Crystallographic Data Centre (CCDC) maintains a subset of all structures in

their database that they classify as a MOF[10] using seven different chemical bond criteria (see Fig. 3), narrowing 850 000 structures in the CSD down to 69 699 MOFs;[10] 1D, 2D, and 3D network structures were included. The CSD subset reports 8,388 structures (out of approximately 70 000 structures) with a pore limiting diameter greater than 3.7 Å. On the other hand, the CoRE MOF dataset contains only structures with 3-D connected frameworks.[9] 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.",[105] but this definition is not universally accepted.[105,106]

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*,[107] 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;[38,108] also, the structure of the MOF could differ depending on the solvent in its pores.[109] 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. 2 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– by correcting the problems 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 terephthalic acid linker, residual dimethylformamide (DMF) solvent (which ideally is removed during activation) remains in its pores, or aromatic rings appear in two different rotational conformations owing to disorder. 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, such as by 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 a computation-ready set of MOFs. 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.[102] identified a set of 1127 MOFs from the CSD using structural queries– searching 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.[110] 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.[111] extended the database of Haldoupis et al.[110] to 3432 MOFs and computationally screened them for kinetic-based noble gas separations. The authors mentioned that they excluded materials with a "significant degree of disorder" and removed solvent from the pores, but do not provide details, nor publicly release the database of MOFs.

In 2013, Goldsmith et al.[103] 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"[103]) 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,[112] 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 of computation-ready MOFs enabled high-throughput computation screenings of experimental MOFs for adsorption-based engineering applications and the generation of considerable insights.[103,110,111,113] 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[94]. The workflow to construct the CoRE MOF database is shown in Fig. 2(a). The authors first searched the CSD for potential MOF crystal structures using the CCDC

[4]The term "CoRE MOF" was coined by Prof. David Sholl, who wrote down a number of different combination of words on a piece of napkin during the Nanoporous Materials Genome Center Meeting in 2013 at Berkeley, CA.

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, we required the structures 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++[114,115] 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 bound and unbound solvents (see Fig. 2(b)) by constructing an adjacency matrix of the crystal structure atoms. There are two types of solvents: 1) free solvents, and 2) bound solvents. Free solvents (i.e., unbound solvents) are molecules that are "touching" the framework atoms, and bound solvents (i.e., coordinated solvents) are molecules that are part of the framework atoms as determined by the van der Waals radii of atoms plus a skin distance of 0.4 Å. Bound solvents are usually coordinated to the open metal sites in MOFs, such as the copper site in HKUST-1. Fig. 2(b)) illustrates the removal of free and bound solvents from a MOF. To remove bound solvents from the open metal sites, the algorithm first constructs an adjacency matrix of an input structure. The adjacency matrix represents the connectivity of atoms (bonds) 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. Then, for each atom connected to a metal atom in the structure, the atom-metal bond is temporarily removed from the adjacency matrix by modifying the corresponding elements to be zero. Following the modification, the adjacency matrix was passed to the SciPy connected components module to check the changes in the bonding of the structure with respect to the modification. If the number of the clusters changes, then the matrix element is left modified. If the number of clusters did not change, then the matrix component is changed back to one (i.e., the bond is reintroduced). This process is repeated for all atoms connected to a metal atoms in the structure. Also, the symmetry-related copies of atoms arising from disorder (see Fig. 2(c)) were deleted by removing the lines from the files that contains the 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 (+) 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) and repair other related disorders, such as overlapping atoms. Following these fixes, solvent molecules (which are sometimes critical in maintaining the structural integrity of the MOF) were removed to mimic the experimental activation condition. A set of 4,764 computation-ready structures were made available to the public via Github.[116]

As evidence that the CoRE MOF database has had a significant impact on high-throughput computational MOF screening and discovery, it has collected over 187 citations on Google Scholar (as of March 5, 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.

Still, there are several criticisms related to the methods employed in the construction of computation-ready MOF databases. (1) The method used for the removal of solvent is too aggressive, in that the structural integrity of the framework without solvent may become questionable for some structures.[117] The newly updated CoRE MOF database[104] contains structures with and without bound solvents. (2) Redundant duplicate structures are present.[118] (3) The MOF structure could change upon activation, relaxing to a different state when solvent is removed;[32] this is particularly a concern for soft porous crystals.[119] (4) As the solvent-removal is automatic and disorder-flagging is imperfect,

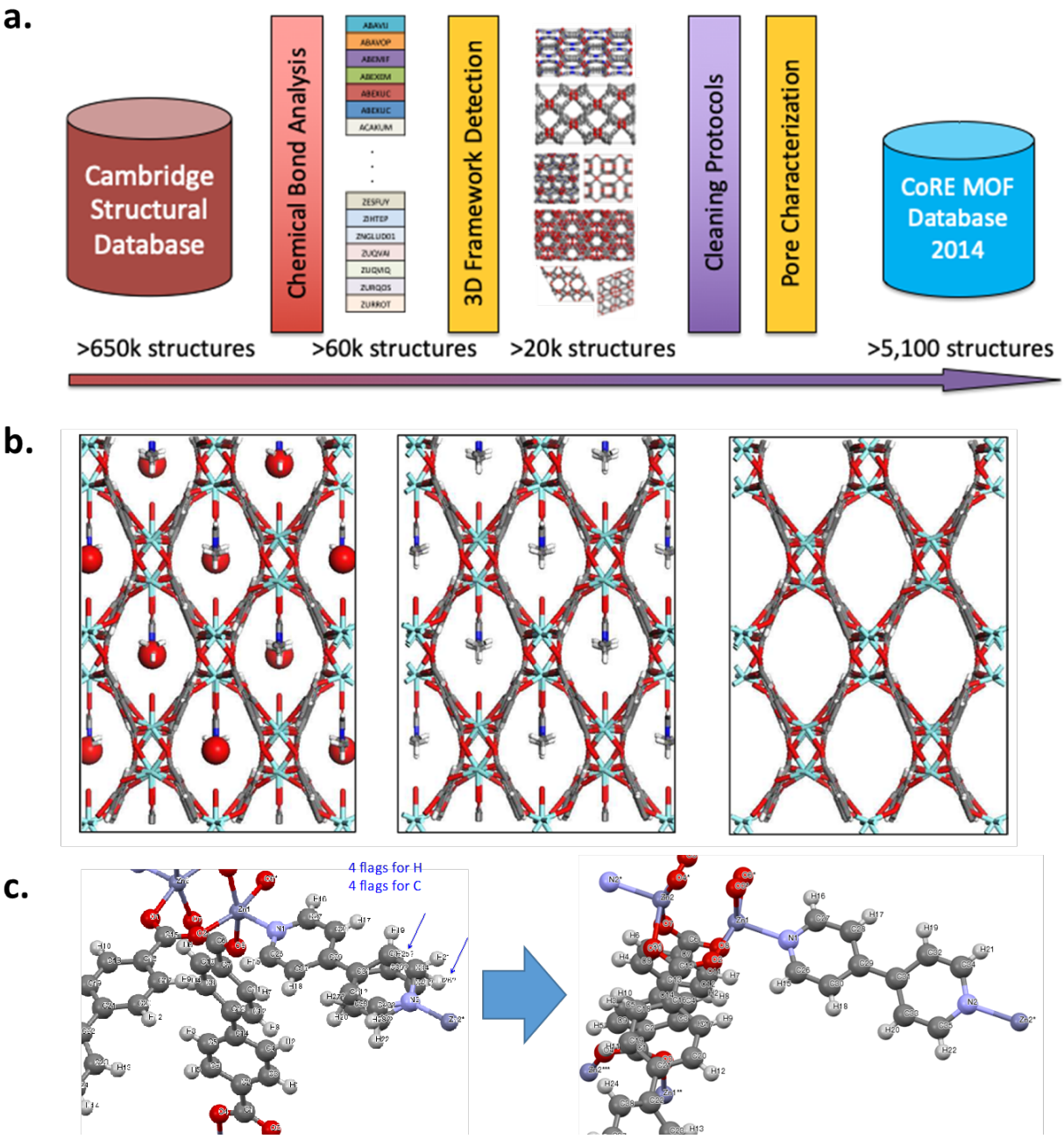


Figure 2: Curating the computation-ready, experimental (CoRE) database of MOFs. (a) The workflow of the development of CoRE MOF Database 2014. (b) Automatically removing solvent from MOFs deposited in the CSD. Left to right, as-is structure, free solvent removed, free and bound solvent removed (CSD refcode: 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.

several structures in the CoRE MOF database are not chemically accurate. A recommended way to report such structures is to send a pull request on the CoRE MOF database website.[5] (5) When repairing disorder, one must choose a certain conformation of the ligand. The choice of ligand comformation 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-

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

determined structure.[86] (6) Several MOFs are missing from the CoRE MOF database, and it is not automatically updated when a new MOF is deposited in the CSD.[10]

5.1.1 CoRE MOF charge assignment

Electrostatic interactions are an important component of the potential energy of adsorption of polar molecules such as carbon dioxide.[120,121] 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 CO_2[29]), thereby completing the description of the electrostatic interaction of the adsorbate with the MOF. While quick/cheap methods such as charge equilibration[122] are commonly used to assign point charges to MOF atoms, point charges derived from electronic structure calculations are generally considered more reliable.[30] Density derived electrostatic and chemical (DDEC) charges[123] is one such method to associate chemically meaningful charges with atoms in MOFs from a calculated electronic density. Nazarian et al. computed DDEC charges for $>$ 2 000 CoRE MOFs from the electronic density calculated with the PBE DFT functional and made these charges openly available.[124] 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.[39] Upon activation (evacuation of solvent) for gas adsorption, the structure of the MOF could change significantly.[38] Some MOFs may be destabilized by removing solvent, causing the framework to collapse.[117] 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.[39] demonstrate for HKUST-1, MOF structures determined by XRD are subject to variation among research groups. To address these issues, Nazarian et al.[39] used DFT to relax the structures of 879 CoRE MOF structures to minimize their potential energy. Indeed, several DFT-optimized CoRE MOFs showed significant changes in the structure and simulated adsorption compared to cognate unoptimized structure.[39] These DFT-optimized structures are likely more reliable then the vanilla CoRE MOF structures[9] because (a) they maintained their structure 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 2 612 optimization-attempted, lanthanide- and actinide-free CoRE MOF structures converged within the 30 000 CPU hours ceiling dedicated to each structure;[39] 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 searches and automatic updates when new MOF structures are deposited. They used seven chemical bonding criteria (see Fig. 3) 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 struc-

tures; 88 % of the MOFs were found to have solvent, of which 52 % is unbound and 48 % bound. The authors employed a complicated filtering process to find disorder present in the MOF structures, resulting in a non-disordered subset, within the MOF subset, containing 54 808 structures. Next, the authors computed geometric properties of the non-disordered subset, such as surface area and pore size. The authors found that a large portion of the structures in the non-disordered subset (85 %) exhibit negligible internal surface areas, while the remaining 8388 structures had notable pores, with a pore limiting diameter of at least 3.7 Å and a gravimetric surface area between 500 m^2/g and 2000 m^2/g.

The CSD-maintained MOF database by Moghadam et al.[10] offers several advantages over the CoRE MOF database: it is more comprehensive (including 1D, 2D, and 3D structures), automatically updated when new MOFs are deposited in the CSD, allows for bond-type or cluster-type searches using the CSD tools, releases the Python scripts used for cleaning, and, during solvent removal, gives the user more granularity by allowing the removal of bound and unbound solvent separately. Two major shortcomings of the CSD MOF subset[10] compared to the CoRE MOF database,[9] however, are (a) many structures are still invalid and not computation-ready because crystallographic disorder was not repaired and missing hydrogen atoms were not added and (b) DFT-optimized structures and DFT-assigned point charges are not available for the CSD MOF subset (unlike for the CoRE MOF database[39,124]).

5.3 Computation-ready crystal structures for other classes of materials

Despite our focus on MOFs, open databases of computation-ready covalent organic frameworks (COFs)[125] and porous organic cage molecules[126] have emerged as well. Tong et al.[127,128] prepared a database of 280 disorder- and solvent-free, experimentally synthesized COF structures (both 2D and 3D) ready for molecular simulations. Miklitz et al.[129] compiled a database of 41 intrinsically porous cage molecules from the CSD. Our recent exploratory work on applying an unsupervised machine learning algorithm to encode the shapes of cavities of porous cage molecules into latent vector representations[130] was built upon the porous cage database of Miklitz et al.[129] We duly mention the open International Zeolite Association (IZA) database of zeolite structures,[131] 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 lucidly 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 van der Waals and electrostatic interactions 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.

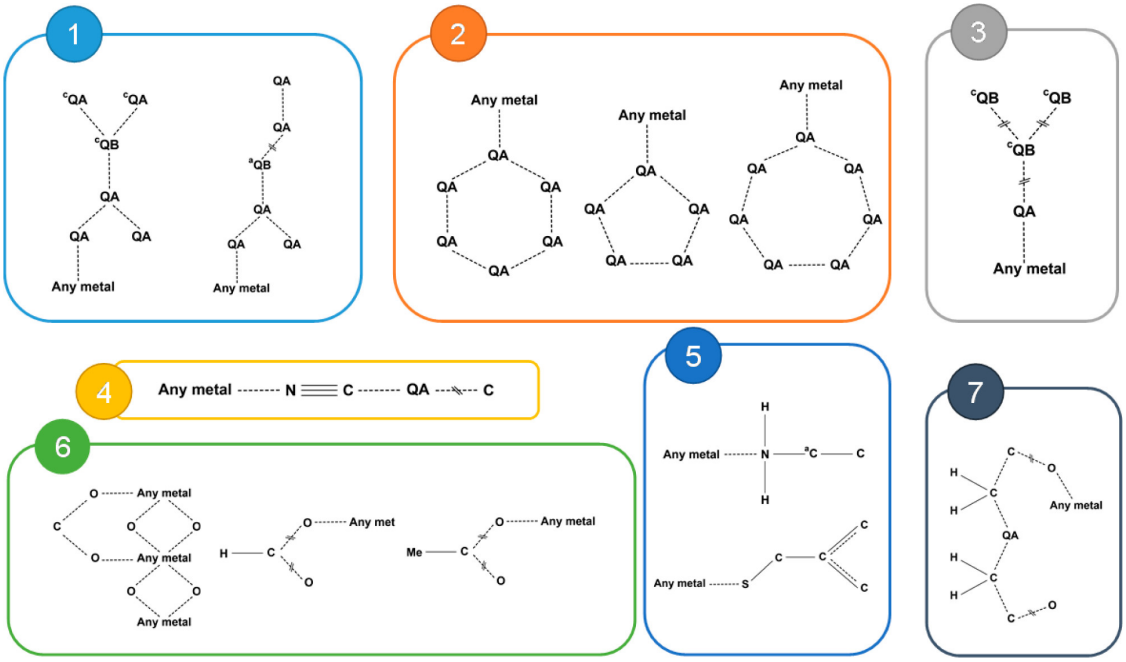


Figure 3: The seven chemical bond criteria used by Moghadam and co-workers[10] to search for MOFs in the CSD. Here A = O, N, P, C, B, S, QB = N, P, B, S, C, ME = methyl group. The superscripts *c* and *a* denote cyclic and acyclic, respectively. Reprinted with permission from Chem. Mater. 2017, 29, 7, 2618–2625. Copyright 2019 American Chemical Society.

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).[17] 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,[132] as well as less VOCs, CO, PM10, and SO_x (but more NO_x).[133] However, the greenhouse effects of fugitive emissions (methane is a potent greenhouse gas itself)[134] and groundwater contamination by hydraulic fracturing[135] may diminish these environmental benefits if not controlled. Third, the pipeline infrastructure for natural gas delivery is already in place in the US. In the US, the transportation sector accounts for 28 % of US energy consumption, and petroleum-based fuels comprise 93 % of transportation fuels.[136] 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.[18] Two incumbent methods to densify natural gas are liquefaction at low temperature (111.7 K, 1 atm) and compression to high pressures (≈ 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.[137] 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.[17,18,138] 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,[139] using a pressure swing between 65 bar (storage pressure) and 5.8 bar (minimum engine inlet pressure needed).[52]

Note that in the high-throughput screenings below, natural gas is approximated as methane.

As a caveat, a computational study by Zhang et al. noted 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.[140]

NOTT-107 and NU-125 Wilmer et al.[141] 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 derived from crystallographic data of already synthesized MOFs. 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. 4). 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.[141] The combinations of building blocks were exhaustively explored, resulting in 137,953 hypothetical MOF structures.

The generated structures were validated by comparing a set of generated structures to their energetically relaxed counterparts. Choosing the appropriate building blocks, structures were generated resembling HKUST-1,[15] IRMOF-1,[6] PCN-14,[142] and MIL-47.[143] Allowing each "pseudo-MOF" to relax energetically using the Universal Force Field,[26] Wilmer et al. found that every atom within the pseudo-

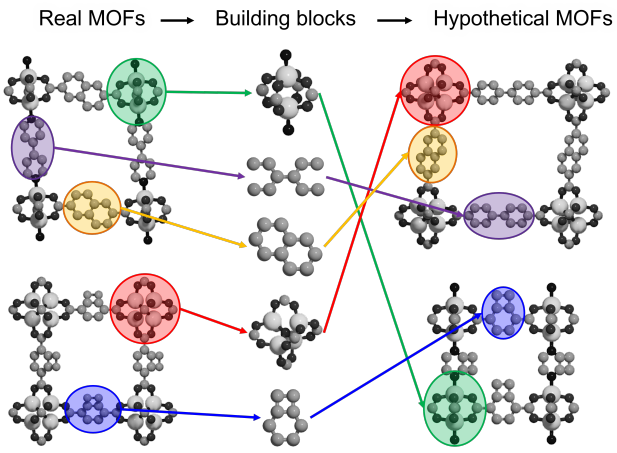


Figure 4: The algorithm used by Wilmer et al.[141] to construct hypothetical MOF structures first extracts building blocks from previously-known MOF structures. These building blocks are then combined according to their geometry to create hypothetical MOF structures.

MOF structures was typically < 0.8 Å distance from the crystallographically measured position. Further, simulated methane adsorption isotherms at 298 K in the pseudo-MOFs agreed with both simulated adsorption in the crystallographic structure as well as with experimental adsorption data.

After validation of the structure generation algorithm, each hypothetical MOF was screened for methane storage at 35 bar and 298 K. The screening of the 137,953 structures was conducted in three successive stages of increasing Monte Carlo cycles in the 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.[141] This structure, unbeknownst at the time of synthesis as having been the previously-reported MOF NOTT-107,[144] was synthesized and found to have an experimentally measured methane capacity $\approx 8\,\%$ lower than the prediction.[141]

In addition to identifying promising candidate structures for synthesis, this library of hypothetical structures provided insights into relationships between the structure of the MOF and high-pressure methane storage. In addition to simulating methane adsorption in each hypo-

thetical MOF, geometric properties such as surface area, void fraction, and pore size distribution were computed. Though maximizing gravimetric surface area had been a common strategy for designing methane adsorbents, exceeding the optimal surface area ($\approx$ 2 500 m^2/g to 3 000 m^2/g) was found to diminish the methane storage capacity. The best adsorbents were found within a narrow range of void fractions $\approx$ 0.8, and the majority of these contained methyl, ethyl, or propyl functional groups, with pore sizes between 4 and 8 Å. These insights led to the discovery of MOF NU-125,[145] 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. 5) and a usable capacity (58 bar to 5.8 bar) that is 67 % that of a compressed natural gas tank at 248 bar.

248 bar? why that comparison?

NU-800 Gomez-Gualdron et al.[146] constructed a set of 204 zirconium-based, hypothetical MOF structures to search for optimal and stable MOFs for storage and delivery of methane. The authors constructed the MOF structural models by computationally arranging a highly stable inorganic secondary building unit, $(Zr_6O_4)(OH)_4(CO_2)_n$, with various building blocks (see Fig 6a) to form MOFs in four network topologies, **fcu**, **ftw**, **scu** and **csq**. Several of the MOFs featured isomeric building blocks (see Fig. 6b) to shed light on possible differences in the adsorption properties of MOFs with the same chemical formula and similar pore volume. The structures were optimized by using the UFF[26] to describe the intrahost energetics.

In each hypothetical MOF, the authors conducted grand canonical Monte Carlo 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 tetratopic building unit TTP in Fig. 6b was predicted to exhibit the highest methane usable capacity (197 L(STP)/L) and was coined NU-8000. See Fig. 6d. As a consequence, NU-800 was synthesized, and its methane (also nitrogen, carbon dioxide and hydrogen) adsorption isotherms were measured and compared to the simulation (see Fig 6d) with overall good agreement. Repeated adsorption and desorption cycles indicated that NU-800 was highly stable. 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)[147] and PCN-14 (149 L(STP)/L),[148] yet 15 % lower than MOF-519 (203 L(STP)/L),[149] the record-holder at the time of publication.

6.1.2 Hydrogen storage and delivery

Hydrogen (H_2) is an ideally clean transportation fuel since it emits only water when it combines with oxygen in a fuel cell. If hydrogen is produced renewably, such as via electrolysis of water with electricity generated from wind turbines[150] 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). Though hydrogen possesses a larger gravimetric energy density than any fossil fuel, 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; hydrogen compressed to 700 bar is a safety concern. Another well-researched densification strategy is to react hydrogen with metals to form metal hydrides,[151] but these metal hydrides often require high temperatures to release the hydrogen and are very heavy.[152] For an ultimate hydrogen storage goal, the US DOE set 50 g/L and 6.5 weight percent storage targets for an onboard vehicular hydrogen

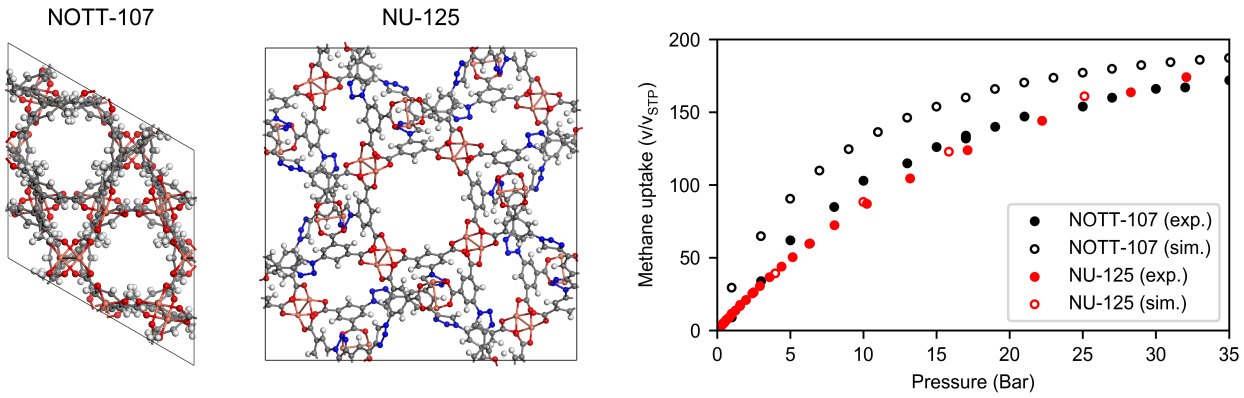


Figure 5: say something about how Wilmer et al. motivated the synthesis and characterization of these in caption so it is self-contained (not clear reading this caption why NU-125 and NOTT-107 are being shown. Crystal structures of NOTT-107 and NU-125 (left) with methane adsorption isotherms (right) demonstrating comparable performance between the two sorbents. []

storage system operating at temperatures ranging from -40 °C to 60 °C[153] and pressures below 100 bar.[154] 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 on MOFs.[154] Thus far, no MOF has met the DOE storage target in the specified temperature range because the van der Waals interactions of hydrogen with a MOF are too weak.[17,152,154]

IRMOF-20 Ahmed and co-workers[155] sought to find MOFs with both high gravimetric and volumetric usable hydrogen capacity. To do so, they simulated hydrogen adsorption at 77 K in a pressure range from 1 bar to 100 bar in each MOF in the CoRE MOF database[9] and the (privatized) database from Goldsmith et al.[103] 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 wt% and 31.1 g H_2/L), considered a benchmark material for hydrogen storage.[156] Among the 90 MOFs predicted to surpass the performance of MOF-5, they targeted IRMOF-20,[157] with a 6.1 wt% and 35.5 g H_2 L^{-1} 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,[112] 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 surface areas of 4073 m^2/g vs. 3512 m^2/g), rationalizing its greater gravimetric hydrogen usable capacity under the Chahine rule.

MFU-4*l* Bucior et al.[62] screened the CSD MOF subset[10] of c.a. 8 000 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 between 100 and 2 bar at 77 K. To engineer a feature vector representing a 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

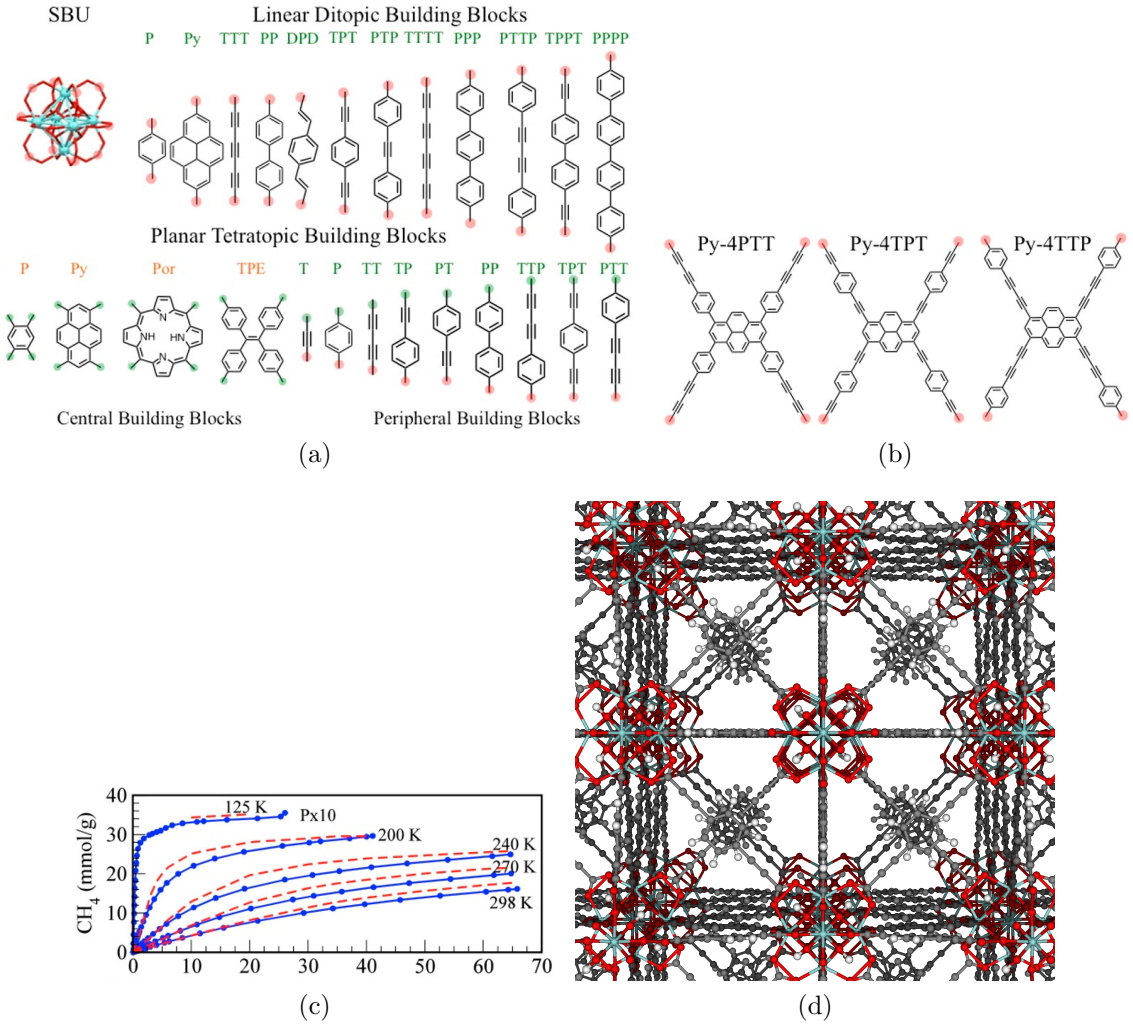
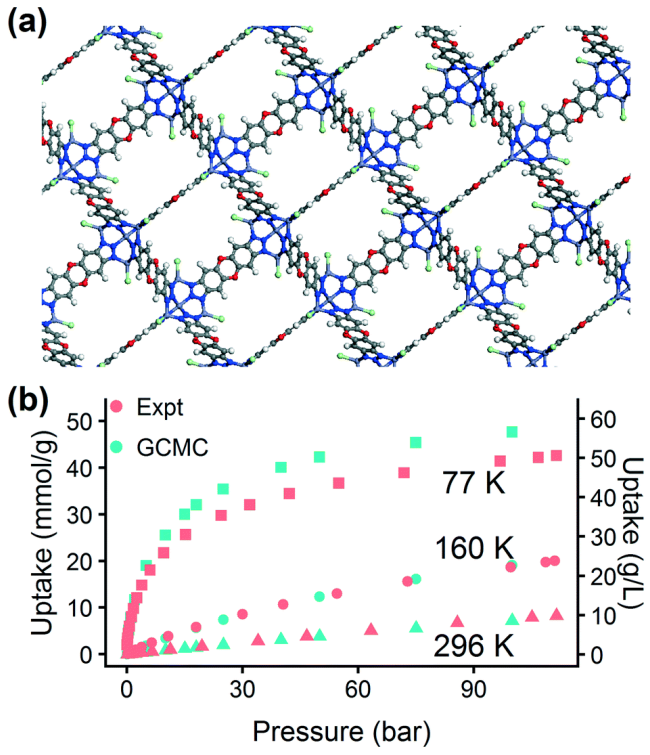


Figure 6: Gomez-Gualdron et al.[146] constructed a set of 204 zirconium-based hypothetical MOFs and screened them for storing and delivering methane, leading to the synthesis of NOTT-800. (a) The basic building blocks used to construct the hypothetical MOFs (connection points highlighted). The ditopic linkers and Zr-SBU are combined in an **fcu** net. The central and peripheral building blocks are combined to form 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 tetratopic building units obtained by combining one central and four peripheral building blocks (see (a)). These particular building units are isomers. (c) Comparison of experimental (solid lines) and simulated (dashed lines) isotherms for methane adsorption in NU-800 at various temperatures. (d) The crystal structure of NU-800 has **fcu** topology and is constructed from the TPT linker in (b) (cyan: Zr, red: O, grey: C, white: H). Reprinted with permission from Chem. Mater. 2014, 26, 19, 5632–5639. Copyright 2019 American Chemical Society.

simulated hydrogen adsorption in a diverse set of hypothetical MOFs. 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 55000 MOFs in the CCDC subset.[10] To refine the usable capacity predictions by the regression model, they conducted GCMC simulations in the 1000 MOFs predicted by the regression model to have the highest usable capacity. MFU-4ℓ[158] (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.[158] Bucior

et al. then experimentally synthesized MFU-4*l*(Zn) 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. 7. MFU-4*l*(Zn) 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.[159]



(a)

Figure 7: A machine-learning accelerated high-throughput screening[62] led to the identification of MFU-4*l* for hydrogen storage. (a) Crystal structure of MFU-4*l*. (b) Simulated and experimental H_2 adsorption isotherms of MFU-4*l* at different temperatures. Reproduced from Ref. 62 with permission from The Royal Society of Chemistry.

she-MOF-1 Gómez-Gualdrón and coworkers[160] constructed a set of 13 000 hypothetical MOF structures falling in 41 different topologies and screened them for cryogenic hydrogen storage. The hydrogen usable capacities in each hypothetical MOF were predicted from the difference in simulated hydrogen adsorption at 77 K and 100 bar and at 160 K and 5 bar (using a combination of a pressure- and temperature-swing for hydrogen storage and delivery). To generate the hypothetical structures, the authors took a "top-down" approach that first specifies the topology of the extended network based on the points of connection emanating from the building blocks, then places 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 of 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) after shipping it off 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.

6.1.3 Oxygen storage and delivery

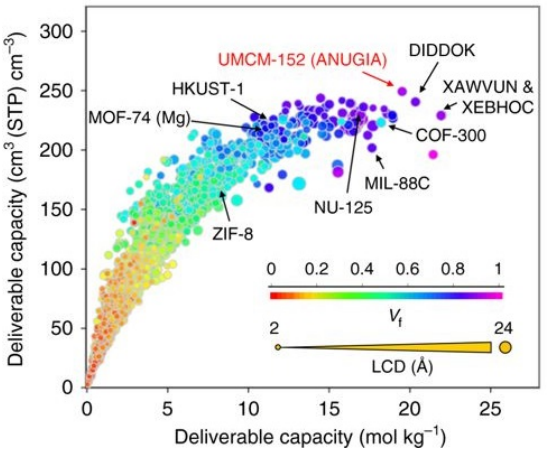
Oxygen (O_2) gas is used in healthcare to treat a variety of respiratory illnesses including chronic obstructive pulmonary disease and pulmonary fibrosis.[161,162] Pure O_2 is also necessary for industrial processes such as Linz-Donawitz-steelmaking which uses O_2 to reduce the carbon content of molten carbon-rich pig-iron to create steel; this process comprises 60 % of all steel production.[163,164] Another use of pure oxygen is to increase the efficiency of the regeneration of catalyst in fluid catalytic cracking – an essential process in petroleum refinement.[165] To store oxygen gas, it is typically densified via compression to high-pressures of around 200 bar and stored at room temperature. MOFs are a novel alternative for O_2 storage,[166] enabling, compared to standard compressed oxygen storage, (a) an increased oxygen storage density

is this a significant use?

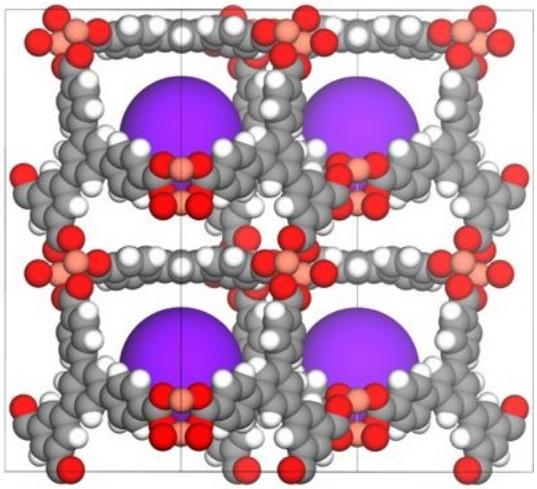
at comparable pressures and/or (b) a comparable stored oxygen density but at a reduced storage pressure, thereby alleviating safety concerns and the need for heavy storage tanks.

this is missing mention of liquid oxygen storage

UMCM-152 Moghadam et al.[167] conducted a high-throughput computational screening of 2392 previously synthesized MOFs from a subset of the CoRE database chosen because of the high-quality partial charges assigned to the MOF atoms.[9,124] The authors performed grand-canonical Monte Carlo simulations of oxygen adsorption at 298 K and at pressures between 1 bar and 200 bar in each MOF. The volumetric oxygen usable capacity at 140 bar storage 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 m^2/g. See Fig 8. The most promising MOF for oxygen delivery was UMCM-152 (CSD refcode: ANUGIA), with a predicted usable capacity of 249 L(STP)/L, and thus UMCM-152 was targeted for synthesis and oxygen adsorption isotherm measurement. UMCM-152 displayed the highest volumetric O_2 delivery for any material reported, 249 L(STP)/L, 22.5 % higher than the previously best reported material, NU-125.[166] At room temperature, the density of oxygen in a UMCM-152-packed tank at 140 bar is 96 % higher than in a traditional O_2 gas tank storage at the same pressure; to achieve the same density in a UMCM-152-packed tank at 140 bar, a compressed cyclinder would have to be up to 300 bar. 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 compacter adsorption-based oxygen storage.



(a)



(b)

Figure 8: Computational identification of UMCM-125 (CSD refcode ANUGIA) for oxygen storage and delivery.[167] (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 (LCD) above 7.5 Å to 8 Å. Common MOFs are highlighted, including the MOF predicted to have the largest O_2 usable capacity, UMCM-152. (b) The crystal structure of UMCM-125, with pore space for oxygen adsorption shown in purple. Figure from Ref. 167 under Creative Commons Attribution 4.0 International License `https://creativecommons.org/licenses/by/4.0/`.

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,[168] and/or (iii) rate of transport through the material. Chemical separations account for 10 % to 15 % of the world's energy consumption.[169] 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 impossible.[170] In practice, deploying a MOF for gas separations entails (a) packing a column with a MOF adsorbent, then passing the gaseous mixture through the column or (b) embedding the MOF adsorbent within a membrane, both allowing the MOF to selective 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.[171] 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.[172] 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.[173] 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 uncontrolled release into the environment.[174] 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.[175] The radioactive krypton from the second process (^{85}Kr, half-life ca. 10.7 years) can be sequestered, while the xenon from the first process, which has a much shorter half-life (longest-lived ^{127}Xe, half-life 36.4 days[175]) could be sold in the market for use in medicine, ion propulsion, lighting, and insulation.[176]

SBMOF-1 Searching for a MOF harboring a high Xe/Kr selectivity, Banerjee et al.[177] 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,[178] but not characterized for Xe/Kr separations. Motivated by the computational prediction, Banerjee et al.[177] synthesized SBMOF-1, measured its pure-component adsorption isotherms, and conducted column breakthrough experiments using a surrogate used nuclear fuel reprocessing off-gas mixture as the inlet. 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. 9. 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 MOFs.[61,179,180]

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.[181] 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.[182,183] Acute exposure to nerve agents can lead to long-term cognitive and behavioral deficits.[184] The adsorption capabilities of MOFs can potentially be exploited to capture chemical warfare agents from the air,[185] e.g., as a filter in a gas mask.[186] 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.

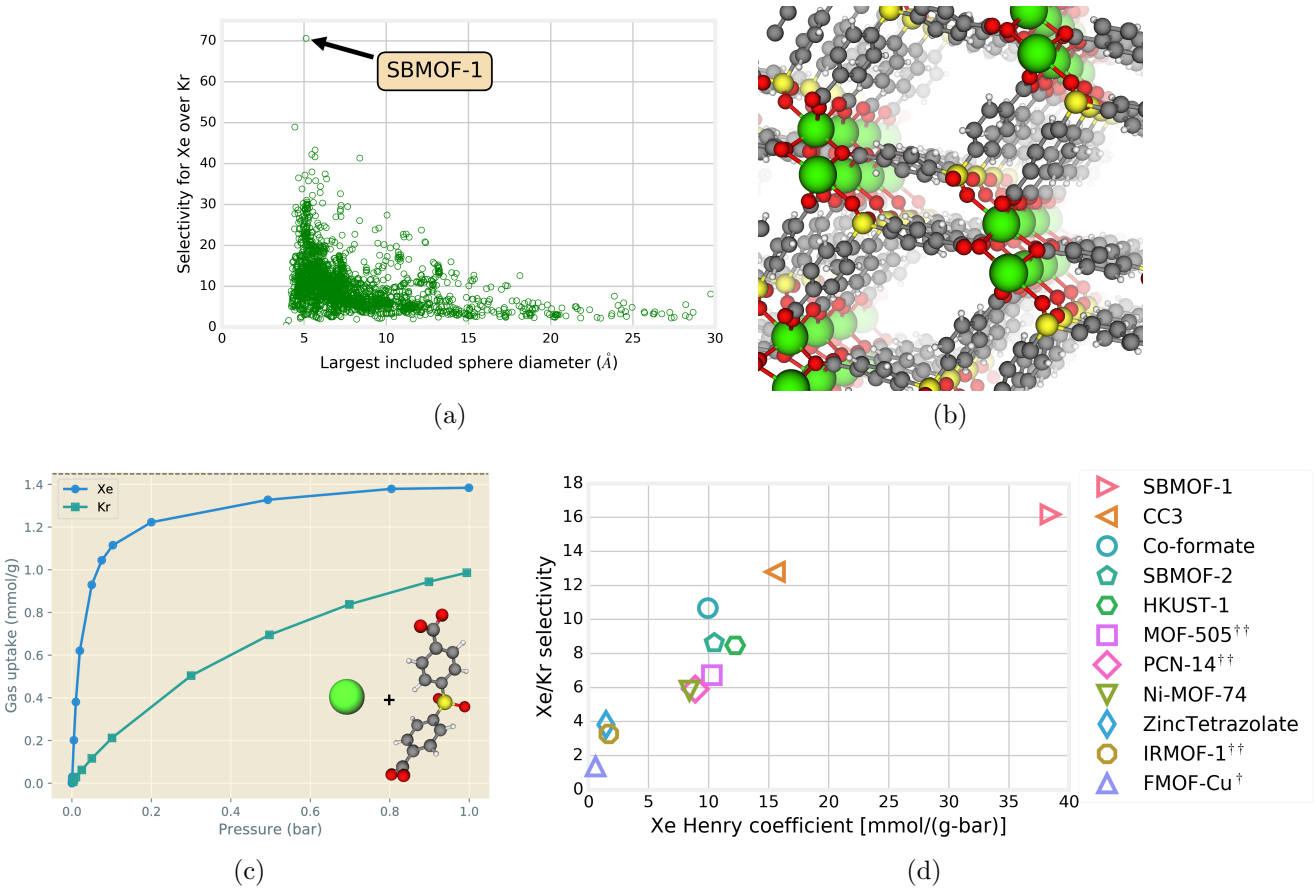


Figure 9: SBMOF-1 for Xe/Kr separations.[177] (a) Simulated Xe/Kr selectivity against pore size; each point represents a CoRE MOF; SBMOF-1 is marked. (b) The crystal structure of SBMOF-1 (CSD refcode KAXQIL) exhibits 1D channels that form well-defined pockets in which xenon can adsorb commensurately 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 and organic ligand used to synthesize SBMOF-1. (d) Comparison of Xe/Kr separation performance among MOFs, at dilute conditions; Henry coefficients were extracted from experimental pure-component adsorption isotherms in the literature. Reproduced and adapted from Ref. 177 under Creative Commons Attribution 4.0 International License `https://creativecommons.org/licenses/by/4.0/`.

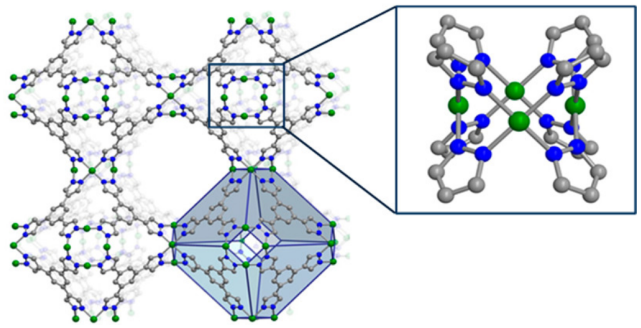
Due to the lethality of chemical warfare agents, surrogate molecules that share key characteristics are used in research to avoid exposure. For example, a commonly used surrogate for mustard gas is diethyl sulfide (DES). See Fig. 10b. Sholl et al.[183] simulated adsorption of nerve agents [soman, sarin] at dilute conditions in the CoRE MOFs and compared their heats of adsorption to that of four common surrogates [dimethyl methylphosphonate (DMMP), diethyl chlorophosphate (DCP), diisopropyl fluorophosphate (DFP), and dimethyl *p*-nitrophenyl phosphate (DMNP)]. Judging from the correlation of the simulated heats of adsorption of the authentic nerve agents in the CoRE MOFs with the heats of adsorption of surrogates, e.g. DMMP, DCP, and DFP are poor surrogates for soman adsorption in MOFs, with DMNP its best surrogate.

$Ni_3(BTP)_2$ Matito-Martos et al.[187] designed a high-throughput screening strategy to iden-

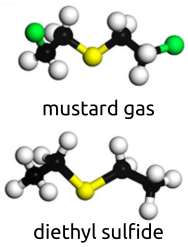
tify MOFs for the capture of chemical warfare agents (CWAs) sarin, soman, mustard in humid environments. The authors screened a subset of 2 932 MOFs from the CoRE database with point charges assigned[124] to account for CWA-MOF electrostatic interactions during the simulations. First, a subset of 1 275 MOFs were excluded because they exhibited pore limiting diameters too narrow (lower than 3.72 Å) to accomodate CWA molecules. Second, the authors simulated adsorption of the CWAs and their surrogates in the 1647 remaining MOFs at dilute conditions, computing the Henry coefficient and isosteric heat 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 MOF ZIF-8.[188] In the shortlist of hydrophobic MOFs, the authors ran more expensive GCMC simulations of mustard gas and nerve agents sarin and soman at 13.8 Pa and 0.6 Pa, respectively, an estimate of the lethal concentrations. Of eight MOFs predicted to exhibit the largest sarin, soman, and mustard gas uptakes (the three were strongly correlated), they selected $Ni_3(BTP)_2$ (CSD refcode UTEWOG) for experimental synthesis and column breakthrough experiments on the basis of its reported thermal and chemical stability.[189] The authors conducted a column breakthrough experiment with 150 mg $Ni_3(BTP)_2$, flowing nitrogen gas with 80% relative humidity (water) and 1 ppm diethyl 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 10c shows that $Ni_3(BTP)_2$ readily captured DES for more than 7 hours, at which point the MOF became saturated with DES, and DES broke through the column. The concentration of DES in the eluted gas (before saturation) was only 0.05 ppm, mimicking the mustard gas concentration that would be inhaled if $Ni_3(BTP)_2$ were a filter in a gas mask within an environment of 1 ppm mustard gas and 80% relative humidity. Further, adsorption in $Ni_3(BTP)_2$ was reversible; thermogravimetric analysis and temperature programmed desorption indicated water was desorbed at a lower temperature than for DES, proving that the framework is selective for DES over water. In summary, Matito-Martos et al. computationally pinpointed $Ni_3(BTP)_2$ as readily adsorbing mustard gas in the presence of humidity and demonstrated capture of its surrogate molecule DES through column breakthrough experiments.

6.3 Carbon dioxide capture

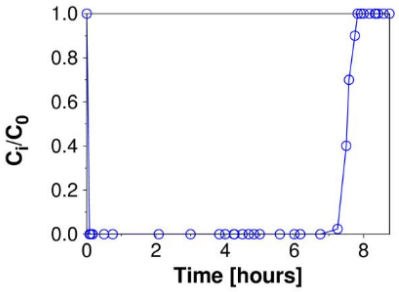
Fossil fuels have been one of the main energy sources since the 20th century in the US[190] and are the main source of carbon dioxide emissions worldwide.[191] The increasing carbon dioxide concentration in the atmosphere is leading to significant changes in the climate, and the average global temperature is projected to rise by 2.6 °C to 4.8 °C by the end of the 21st century if carbon dioxide emissions are not mitigated from their current trajectory.[192] In the United States, roughly 40 % of CO_2 emissions are directly tied to the burning of fossil fuels in power plants,[193] and significant effort is devoted to develop efficient technologies to capture CO_2 from the flue gas of fossil fuel-fired power plants before it is released into the atmosphere. The captured CO_2 can then be sequestered in a geological reservoir.[194] The methods currently used to capture carbon dioxide, such as absorbing CO_2 with aqueous alkanolamine absorbents have not proven to be energy efficient and reduce energy output by approximately 30 % in most power plants.[195] The energy penalty is mainly due to the high cost of separating CO_2 from the flue gas after fuel combustion and regenerating the the amine solvent used to capture the CO_2.[195] A more recent development sees MOFs selec-



(a)



(b)



(c)

Figure 10: Computational identification of $Ni_3[BTP]_2$ for chemical warfare agent capture.[187] (a) The structure of $Ni_3[BTP]_2$ (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 through a column packed with 150 mg of $Ni_3[BTP]_2$ at 20 mL/min. Shown is a measurement of normalized DES concentration at the outlet of the column via gas chromatography. Reprinted with permission from Chem. Mater. 2018, 30, 14, 4571–4579. Copyright 2019 American Chemical Society

tively capturing the CO_2 gas, which could reduce the energy penalty significantly.[193] The removal of CO_2 from the flue gas is called post-combustion carbon capture. An alternative is to capture CO_2 prior to the fuel combustion– a pre-combustion strategy. Instead of directly burning natural gas to produce electricity, natural gas and steam are converted, in the presence of a catalyst, to H_2 and CO. The CO is then reacted with water, in a water-gas shift reaction, to produce CO_2 and H_2. Carbon dioxide can then be separated from H_2 rather than from exhaust gases, and the pure H_2 can be burned to produce electricity. An advantage to employing MOFs for pre-combustion carbon capture is that the CO_2/H_2 mixture is at high pressure already, and a pressure swing down to atmospheric pressure can readily push CO_2 out and regenerate the MOF.

NOTT-101/OEt Chung et al.[66] used a genetic algorithm to find a MOF for pre-combustion carbon dioxide capture. To apply genetic algorithms to search for MOFs with good selectivity for CO_2 over H_2 and a high CO_2 usable capacity, Chung et al. explored the hypothetical MOF (hMOF) database by Wilmer et al.[141] The database contains some duplicate "twin" structures, so the database was reduced from 130k to 55k structures. Every hMOF was described with six digit integers, representing the maximum and actual degree of interpenetration and the species of the inorganic nodes, functional group, primary organic linker, and secondary organic linker (see Fig 11a). In a genetic algorithm, this representation and each digit integer is called a chromosome and a gene, respectively. 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 better 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 is allowed to repeat itself, allowing the chromosomes to "evolve" and eventually the fitness converges to a local minima.[196] All hMOFs were described by a chromosome and a subset of 100 diverse hMOF chromosomes were chosen for the initial generation. Three genetic algorithms were initiated, each with a different adsorption property serving as the fitness: CO_2 usable capacity, selectivity for CO_2 over H_2, and an adsorbent performance score (APS), which is a product of the two previously mentioned 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 11b), meaning that those genes led to higher fitness throughout the evolutionary stages of the genetic algorithm. These genes were used to obtain a preliminary list of MOFs that exhibited both good selectivity for CO_2 over H_2, a high CO_2 usable capacity, and APS (see Fig 11c). Governed by previous experiences in MOF synthesis and its fitness in the genetic algorithms, Chung et al. synthesized and tested NOTT-101/OEt experimentally. Both CO_2 and H_2 adsorption isotherms were measured (see Fig 11d) and showed a good agreement with simulated isotherms. NOTT-101/OEt had a usable capacity of 3.8 mmol/g and a CO_2/H_2 selectivity of 60. Other notable MOFs studied for CO_2/H_2 separations in pre-combustion are Mg-MOF-74, with a 2.6 mmol/g usable capacity and a 365 CO_2/H_2 selectivity,[197] and Cu-BTTri, with a 3.7 mmol/g usable capacity and a 20 CO_2/H_2 selectivity.[198]

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 reaction, Chung et al. carried out a series of pressure-swing adsorption (PSA) simulations to find the lower limit of CO_2/H_2 selectivity. They found that the materials to be used for the precombustion CO_2/H_2 separation, the CO_2/H_2 selectivity needs to be greater than 30 to meet the process objective of 99.999% H_2 purity. On the basis of the process modeling, Chung et al concluded that while NOTT-101/OEth meets the process requirement, another porous material, Cu-BTTri, cannot be used for precombustion carbon capture application because the CO_2/H_2 selectivity is not high enough to generate the high purity H_2 stream required for subsequent energy generation.

doesn't purity follow from selectivity? I am naive.

6.3.1 Xylene enrichment

Mixtures of C_8 aromatics– *ortho*-, *para*-, and *meta*-xylene and ethylbenzene– are obtained from the catalytic reforming of crude oil.[199,200] The *p*-xylene isomer is the most valuable component of the C_8 aromatic mixture. It is oxidized to yield terephthalic acid or dimethyl terephthalate, both feedstocks for the production of polyethylene terephthalate (PET),[199] which is widely used for synthetic fibers (polyester) and bottles.[169] Pure *o*-xylene is also valuable to synthesize phthalic anhydride, a precursor to produce plasticizers.[199] However, mixtures of C_8 aromatics are very challenging to separate because of their similar shapes, boiling points, and polarities.[201] The two incumbent industrial processes to separate C_8 aromatic mixtures are crystallization and, more often, selective adsorption onto a solid-state material.[199,200,202] MOFs are promising adsorbent materials for separating C_8 aromatics more effectively than zeolites, which are currently used in a simulated moving bed process to obtain high-purity *p*-xylene.[203]

MOF-48 Gee et al.[204] conducted multicomponent grand-canonical Monte Carlo simulations of adsorption in the CoRE MOFs immersed in a 0.33:1:2:1 ethylbenzene/*o*-X/*m*-X/*p*-X (X=xylene) mixture at 9 bar and 50 °C (conditions for liquid phase). Among the CoRE MOFs with the highest predicted selectivity for and capacity of *p*-xylene, Gee et al. selected MIL-140B, MOF-48, MIL-47, and MIL-125-NH_2 to target for experimental investigation after also considering chemical and thermal stability and the commercial availability of their linkers and metals. Liquid-phase breakthrough adsorption measurements then tested the capability of each MIL-140B, MOF-48, and MIL-125-NH_2 to separate *p*-xylene. Of these, MIL-140B exhibited the highest breakthrough *p*-xylene selectivities (1.8 over *o*-xylene, 1.6 over *m*-xylene, and 2.1 over ethylbenzene). The authors claim that their column breakthrough experiments indicated MIL-140B exhibits a higher *p*-xylene selectivity than zeolite BaX currently used in industry, but their breakthrough experiment for BaX was conducted at 180 °C compared to 50 °C for MIL-140B. Interestingly, though MOF-48 and MIL-47 differ only by a dimethyl-functionalization, MOF-48 exhibits *p*-xylene selectivity while MIL-47 exhibits *o*-xylene selectivity, emphasizing that

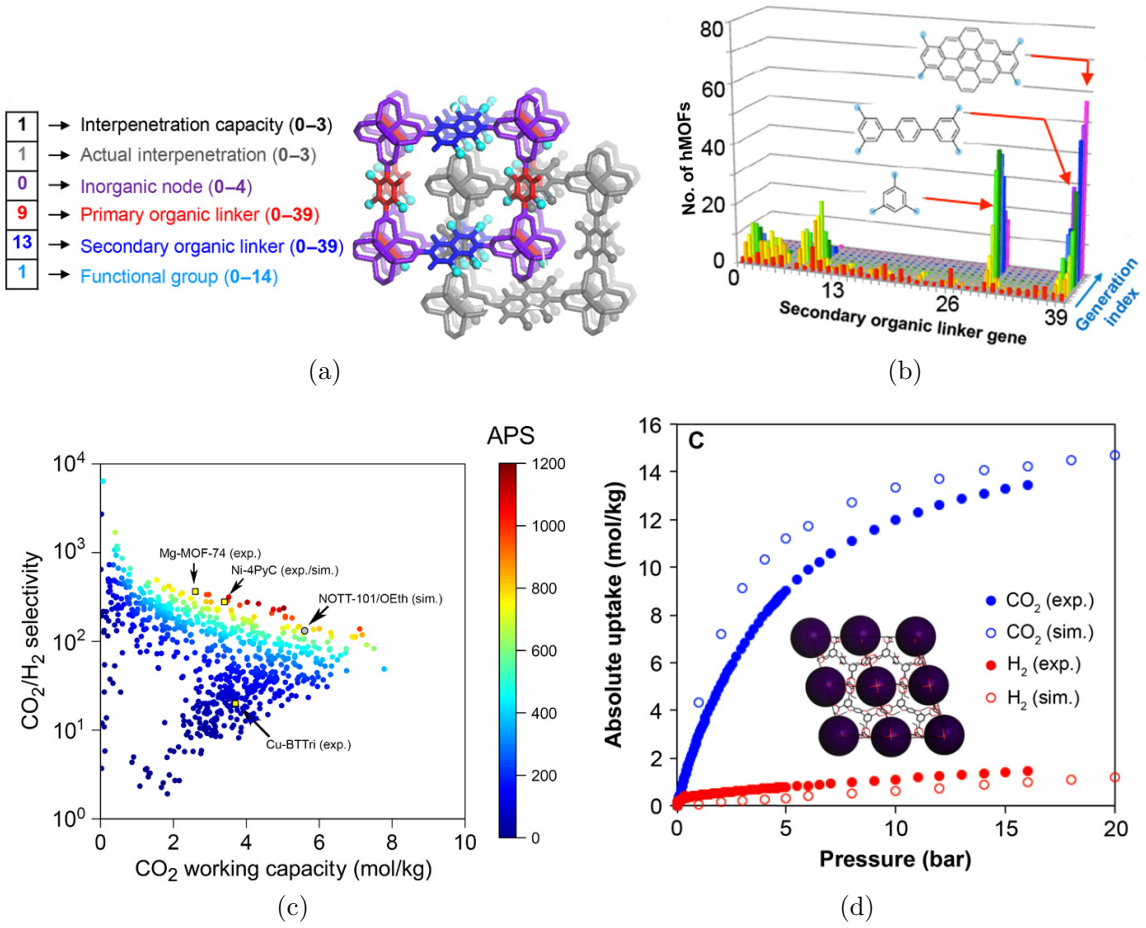


Figure 11: The discovery of NOTT-101/OEt was guided using a genetic algorithm, leading to a good comparison with experimental data.[66] (a) A chromosome describes a MOF structure with 6 integers (called genes), each representing a different property of the MOF. The genes form a chemical space that will 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 described by the chromosomes, hinting that these linkers are optimal for CO_2 capture. (c) The adsorption properties of the MOFs be more specific than vague "results" from the genetic algorithm search. Each point represents a hypothetical MOF generated in the genetic algorithm. (d) A comparison between simulated and experimental isotherms for NOTT-101/OEt. The experimental isotherms were multiplied with a correction factor for this figure, to take into account the 92 % pore activation of the synthesized NOTT-101/OEt. The structure of NOTT-101/OEt can be seen in the inset, where pores are represented by dark spheres.

subtle distinctions in pore features can lead to selectivity switching for these similarly-shaped C_8 aromatic isomers.

A lesson from machine learning

The field of machine learning aims to leverage data to train mathematical models or algorithms to perform a task. For example, instead of explicitly programming a computer to translate speech, identify and classify traffic signs, detect fraudulent financial transactions, and recommend music, a machine learning algorithm/statistical model is tuned to perform these tasks after taking in many examples (data) as input. As more data is used to train the ma-

chine 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 digits[205] contains c.a. 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 % error[206]). As a more challenging arena, CIFAR-10 and -100[207] 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), whereby publicly released are (1) a labeled data set for training a model and (ii) a data set where labels are withheld for testing/evaluating the model (i.e. only independent variables are 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 c.a. 480 000 subscribers on c.a. 18 000 movies, comprising c.a. 100 million movie ratings from 1 to 5.[208] Three million ratings (by the same set of subscribers, on the same set of movies) were withheld as test data. In 2009, a $1 million prize was awarded to the team that improved upon the incumbent algorithm of Netflix, Cinematch, by decreasing the root mean square error between predicted and actual ratings on the test set by 10 %. Both the release of the ratings data and the competition generated interest in recommendation systems, spurred the sharing of ideas between groups, and led to advances in recommendation algorithms (which were disseminated).[209,210] Koren,[211] 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.[212] ImageNET is a crowdsource-annotated database of millions of images with hundreds of object categories. An annual workshop is held at the end of the year to disseminate and discuss the most innovative and successful approaches.[212]

Within the realm of materials science, to spur developments and track progress in crystal structure prediction,[213] the CCDC holds a challenge to predict crystal structures of molecules.[214] To spur force field development, an open challenge could be held 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)[215] 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 mea-

surements, 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 the NIST Reference Material, NIST RM-8852 (an Ammonium ZSM-5 zeolite), which was developed via an interlaboratory study.[216] 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 Materials[217] (NIST-ISODB)[6]. 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. 12) or by downloading a structured data file from NIST-ISODB. Furthermore, the NIST-ISODB web application was and is capable of plotting multiple isotherms, from the same or different source articles, simultaneously, allowing for online comparison of isotherms; simultaneous plotting of isotherms for the same adsorbent/adsorptive/temperature combination enables a sort-of "virtual interlaboratory study" of the particular adsorption experiment. The NIST-ISODB has steadily grown to over 3 500 articles and more than 30 000 isotherms as of publication of the present manuscript. Data additions to the database are chosen either from the results of string-based searches of the extant literature or by direct submission of data by outside laboratories. Other additions to NIST-ISODB since its 2014 launch include improved database vocabulary for adsorptive species (via the InChIKey scheme[218]) and adsorbent materials (see following paragraph), an application programming interface (API) for accessing the database contents (in particularly, the library of adsorption isotherms), isotherm fitting tools inside the online isotherm plotting utility, an 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.

One challenge identified early in the NIST-ISODB project was that of the naming scheme(s) for adsorbent materials and MOFs in particular. In short, there is no standard method for naming MOFs and, perhaps more critically, a specific MOF may go by multiple names, easily leading to confusion for novices and experts alike. For example, the material named HKUST-1 in the disclosure of its initial synthesis[15] is now more commonly known as CuBTC (short for Copper Benzene-1,3,5-tricarbolyxate), but is also known as MOF-199[219] and is sold by BASF under the name Basolite™ C300.[220] Multiplicity of names, such as for CuBTC, hinders effective searches in NIST-ISODB. To solve this problem, the NIST Registry of Adsorbent Materials[221] (NIST-MATDB)[7] was released in 2017 as a companion to NIST-ISODB. The overarching purpose of NIST-MATDB is to identify adsorbent materials via unique iden-

[6] `https://adsorption.nist.gov/isodb`

[7] `https://adsorption.nist.gov/matdb`

Figure 12: Screenshot of an article entry[216] 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.

tifiers (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.[222] For example, Fig. 13 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. 14 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.[223,224]) 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,[141,225,226] 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 then open competitions can drive

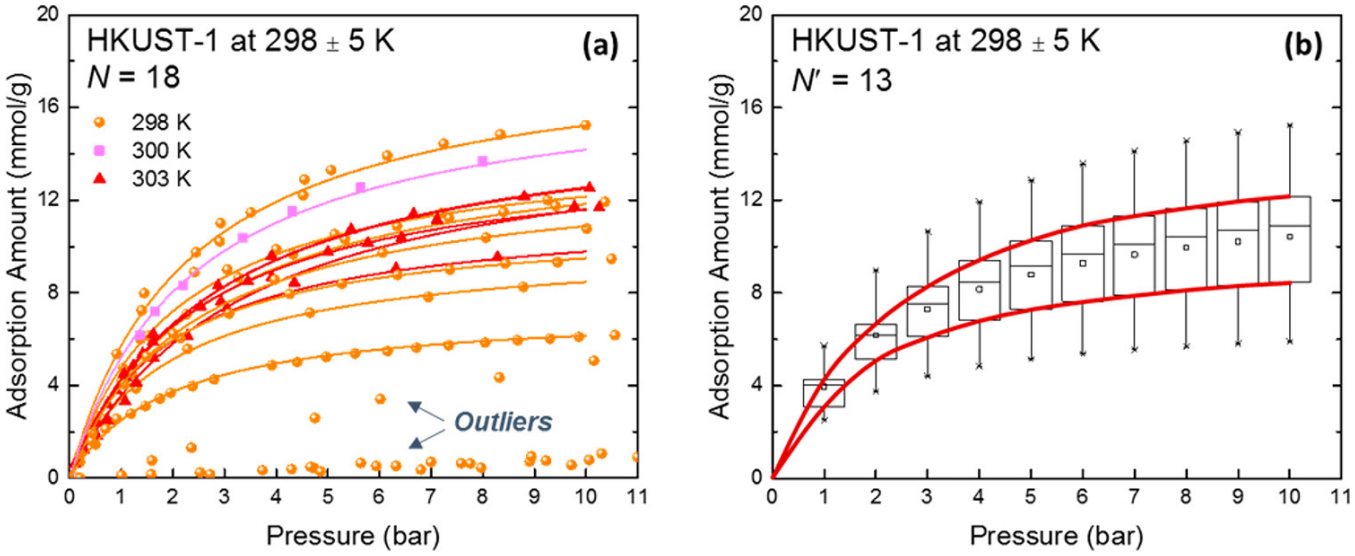


Figure 13: On the reproducibility of CO_2 adsorption isotherms in HKUST-1.[222] (a) Experimental isotherms of carbon dioxide adsorption in HKUST-1 at temperatures of 298 K ± 5 K, as collected and organized by Park et al.[222] from the NIST-ISODB, to show the variability of laboratory measurements for 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.

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. In fact, 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)[227] (nine editions to date), in which challenge entrants are 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[228–231] 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 FACT laboratory at NIST is to develop and disseminate best practices for adsorption measurements, which addresses this point in part.) One specific issue is the difficulty encountered in comparing isotherms from different laboratories when the adsorption measurand is pre-

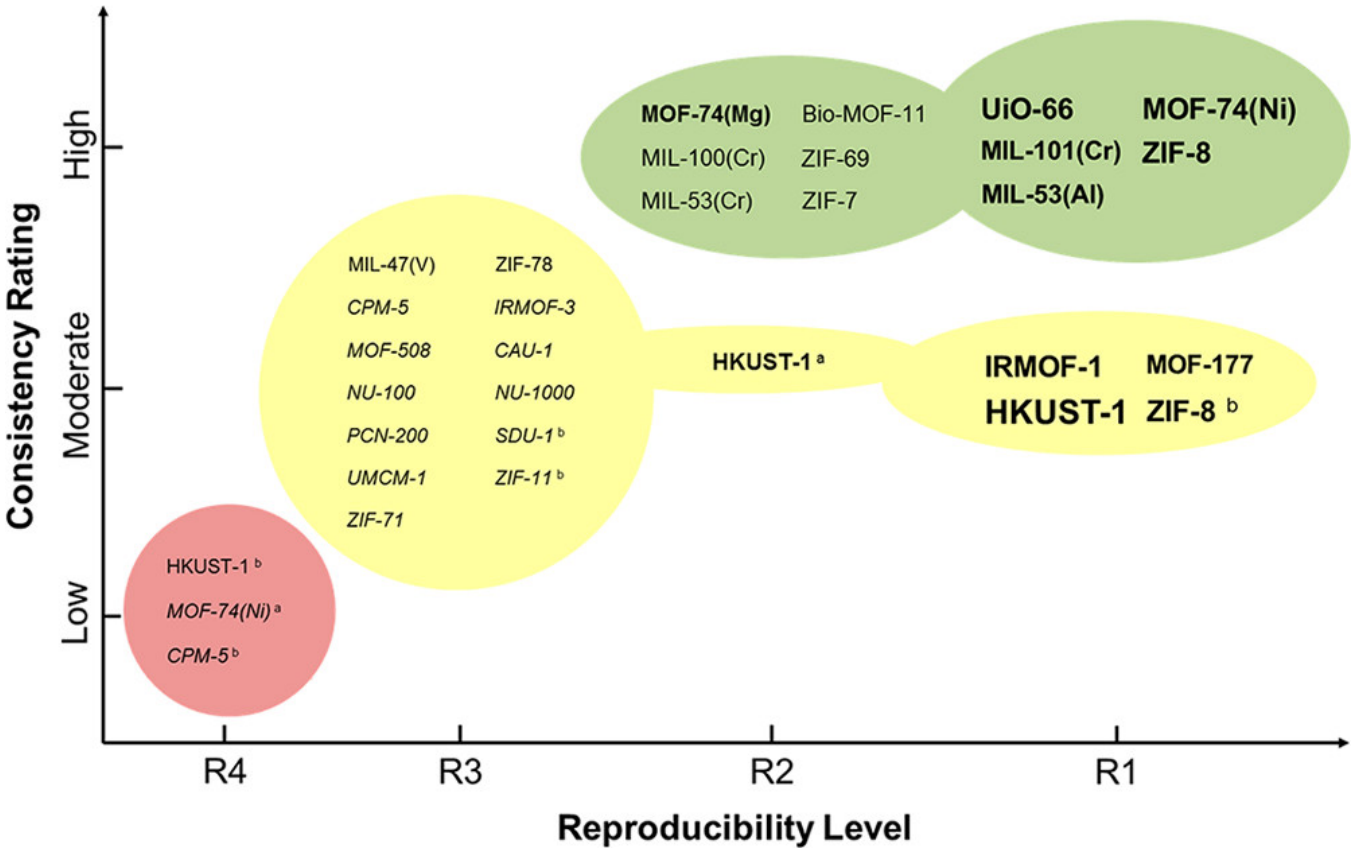


Figure 14: Graphical summary of reproducibility and consistency results from Park et al. for adsorption isotherms of carbon dioxide on various MOFs.[222] On the x-axis, reproducibility increases (according to Park et al.'s metrics) from left to right. Font size is indicative of the number of independent isotherms available for the noted material and bolding/italicization identifies outlier levels. The reader may consult Ref. 222 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.

sented 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) to the volumetric units cm^3(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.[217]) 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 description of adsorption experiments, the quantities that are actually measured, and the clear presentation of both. For example, experimental isotherms are typically presented as *excess* isotherms,[232] but rarely described explicitly as such. Conversely, isotherms obtained from molecular simulations are usually of the *absolute* type. 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 a 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. 233). It is similar in principle to the Crys-

tallographic Information File (CIF) developed and promoted by the International Crystallography Union[234,235] 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 CRISPR[70] and large, open datasets that spurred developments in recommendation algorithms[209] and computer vision.[212] 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 also applied in other domains to discover e.g. organic light-emitting diodes[236] and Li battery materials.[237] The field of drug discovery has long adopted principles of high-throughput computational screening and cheminformatics.[238,239]

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. showed that machines can learn from failed attempts to synthesize a MOF to identify better synthesis conditions that yield a higher-quality crystal.[19]

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 CO_2,[240] open transition metal sites to target ethylene[241]) or exclude (e.g. functionalization with hydrophobic alkyl chains to exclude water[242]) 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,[41] 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 when the pore size is commensurate with the molecule []. Also, interactions with e.g. coordinatively unsaturated metal sites in MOFs are difficult to predict.[243] Moreover, some MOFs harbor flexible backbones or rotating/wobbling constituents[119] and undergo structural changes upon adsorption of gas,[244] sometimes involving delicate competition between entropy and energy.[245] 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 in the computational MOF community is imperative for repro-

ducibility and efficient progress. See the review of Coudert,[246] 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.[247] Some MOFs however, are known to have flexible backbones[248] or constituents (e.g. rotating ligands[249,250]) 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.[119,244] Even small pore size fluctuations can be important to account for when the (average) pore size is commensurate with the size of the adsorbate.[251] 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,[34,35,117] (ii) efficient algorithms to sample MOF configurations, also underway,[36] and (iii) increasing computing power.

9.2 Developing more accurate force fields

Arguably, the most important future research direction lies in classical force field (molecular model) development. 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.[31,41] Owing to the exorbitant computation required for *ab initio* methods to compute the potential energy of a single MOF-gas configuration, it is impractical to e.g. conduct a grand-canonical Monte Carlo simulation of gas adsorption in a MOF using an *ab initio* description of the potential energy[8]. As a consequence, classical interatomic potentials are used for an energetic description, 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 incur a loss in accuracy compared to *ab initio* methods. Off-the-shelf, generic force fields such as the Universal Force Field,[26] DREIDING,[27] AMBER,[253] OPLS,[254] etc. are typically employed to describe van der Waals interactions for high-throughput screening, but especially often suffer from low accuracy because these force fields were not tuned specifically for chemical environments found in MOFs, e.g. coordinatively unsaturated metal sites.[255]

right?

[8] Though, notably, density functional theory (DFT) calculations were used as an energetic description of small gas molecules in MOF-74 to compute Henry coefficients via Widom insertions[252] by biasing the samples towards low-energy regions.

McDaniel et al.[256] systematically compared predictions of CO_2 and CH_4 adsorption in 424 MOFs among UFF and an *ab initio* force field (based on symmetry-adapted perturbation theory (SAPT)[257]); predicted adsorption often differed significantly between the generic and *ab initio* force field, but the statistics of the ranking of the MOFs according to adsorption was good (Spearman's rank correlation coefficient squared for CH_4 and CO_2 ranges from 0.81 to 0.98, considering uptake at both 1 bar and 30 bar and 298 K). The later finding, though specific to CO_2 and CH_4, emphasizes that high-throughput computational screenings (using generic force fields) may rank materials with sufficient statistics but not qualitatively predict adsorption in each MOF satisfactorily. To an extent reducing the importance of which generic force field to employ in a screening, Dokur et al.[258] showed that simulated CO_2, H_2, N_2, and CH_4 uptakes in 100 MOFs with UFF are well-correlated with those using Dreiding (binary gas mixtures at 1 bar, 298 K).

Still, the adsorption predictions using these off-the-shelf force fields are not quantitatively accurate. Using our own work as an example, though SBMOF-1 was correctly predicted by the simulations to be a highly-ranked material for Xe/Kr separations, the predicted Xe and Kr adsorption isotherms deviate significantly from the experiment (Supplementary Figure 22 in Ref. 177). Furthermore, accurately modeling interactions of adsorbates with coordinatively unsaturated metal sites in MOFs is notoriously difficult with classical force fields.[255]

Notably, assessing the accuracy of a force field by comparing to an experimental adsorption measurement is complicated because attributing deviations between simulation and experiment to causal factors is extremely difficult. First, there is significant variation in the experimental adsorption isotherm measurements,[222] 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 van der Waals interatomic potential parameters, inappropriate functional forms of the guest-host van der Waals interatomic potential polarizability of the adsorbate, etc.) could contribute to poor agreement with the simulation, and it is often difficult to definitively attribute error to any 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 particul 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. Dan, anything to say here?

One idea is to cede 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.[255] 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 computational expense.

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.[259–262] Typically in force field development, a strict functional form for the interatomic potential is chosen, at least partially motivated by physics (e.g. Lennard Jones, Buckingham potentials). If the imposed functional form does not adequately allow the important physics to be captured, the tuned force field will be inaccurate. In contrast, deep neural networks are highly adaptable and capable of representing highly non-linear potential energy

surfaces[263] 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. The reason is that the neural network must learn the shape of the potential energy surface in addition to the quantitative details[2619].

this section needs help

9.3 Developing more accurate models of electrostatics

The electrostatic potential field within a MOF is typically described by assigning point charges to the MOFs. See Ongari et al.[30] for a summary and comparative assessment of the hierarchy of methods to assign these point charges to MOF atoms. There is signficant variance among the charges predicted by the different methods, and predicted adsorption can be sensitive to the charges assigned to the MOF.[120]

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.

Ideally, one could use raw the electrostatic potential grid as output from electronic structure calculations.

this section needs help

[9]The following thought experimental clarifies why more data is needed to fit a neural network to the PES surface than to fit a traditional force field with an interatomic potential imposed. Consider if 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.

9.4 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.[264,265] A flagship example is UiO-66,[266] 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.[267] Both experimental[267] and computational[268] studies have elucidated how the linker defects significantly influence adsorption of CO_2 in UiO-66. Through molecular modeling, Bristow et al.[269] investigated the mechanism by which linker vacancies form in UiO-66.

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.

9.5 Predicting MOF stability

Stability is a prerequisite for deploying a MOF for most practical applications. An important area of research is predicting *a priori* which hypothetical MOFs can be experimentally synthesized and their thermal, chemical, and mechanical stability.

For example, Gomez-Gualdron et al.[160] constructed a set of hypothetical MOFs by computationally assembling molecular building blocks into topological nets and computationally identified **she-MOF-1** for hydrogen storage; however, upon synthesis and after shipment for adsorption measurements, **she-MOF-1** lost 30% of its pore volume (i.e. it was not stable). In addition to predicting their adsorption properties, a means to predict the stability of hypothetical MOFs would further accelerate the deployment

of MOFs for adsorption-based engineering applications.

In a rare study of MOF stability in the presence of water, Greathouse and Allendorf[270] 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 ZnO_4 polyhedra.

know of other studies of MOF stability?

9.6 Multi-scale modeling

So far, the majority of computational high-throughput screening relied on the computation of performance metrics based on equilibrium properties, such as usable capacity and selectivity. Although these properties are important indicator of the performance of these materials for adsorption applications, improving these metrics do not guarantee improvements in the process objectives. For instance, the usable capacity and selectivity of CO_2 are commonly used performance metrics for CO_2 capture applications, while the overall "process" objectives are the cost of CO_2 and the product purity and recovery. Since the improvements in adsorption selectivity and usable capacity do not necessarily translate into process performance, we envisage the development of more accurate performance metrics that ties the properties obtained from molecular simulations and the process objectives will be important in the future.

nice. can u give concrete example of multi-scale modeling? like modeling the breakthrough curve?

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−2016R1D1A1B3934484).

Supporting Information Available: Supporting Info here. This material is available free of charge via the Internet at `http://pubs.acs.org/`.

References

(1) Furukawa, H.; Cordova, K. E.; OKeeffe, 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* **2011**, *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.

(7) Notman, N. MOFs find a use. *Chemistry World* **2017**,

(8) Hendon, C. H.; Rieth, A. J.; Korzynski, 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.; Wiggin, 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 folkloreand reality of reticular chemistry. *Materials Chemistry Frontiers* **2017**, *1*, 1304–1309.

(14) 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.

(15) Chui, S. S.-Y.; Lo, S. M.-F.; Charmant, J. P. et al. A chemically functionalizable nanoporous material $[Cu_3(TMA)_2(H_2O)_3]_n$. *Science* **1999**, *283*, 1148–1150.

(16) Simon, C. M.; Kim, J.; Gomez-Gualdron, D. A. et al. Computer-aided search for materials to store natural gas for vehicles. *New Discovery, Front. Young Minds-Understanding the Earth and its Resources* **2015**,

(17) Schoedel, A.; Ji, Z.; Yaghi, O. M. The role of metal–organic frameworks in a carbon-neutral energy cycle. *Nature Energy* **2016**, *1*, 16034.

(18) Mason, J. A.; Veenstra, M.; Long, J. R. Evaluating metal–organic frameworks for natural gas storage. *Chemical Science* **2014**, *5*, 32–51.

(19) Moosavi, S. M.; Chidambaram, A.; Talirz, L. et al. Capturing chemical intuition in synthesis of metal-organic frameworks. *Nature Communications* **2019**, *10*, 539.

(20) 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.

(21) 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.

(22) 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.

(23) Boyd, P. G.; Lee, Y.; Smit, B. Computational development of the nanoporous materials genome. *Nature Reviews Materials* **2017**, *2*, 17037.

(24) Fraux, G.; Chibani, S.; Coudert, F.-X. Modelling of framework materials at multiple scales: current practices and open questions. **2018**,

(25) Coln, Y. J.; Snurr, R. Q. High-throughput computational screening of metalorganic frameworks. *Chemical Society Reviews* **2014**, *43*, 5735–5749.

(26) Rappé, A. K.; Casewit, C. J.; Colwell, K. 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.

(27) Mayo, S. L.; Olafson, B. D.; Goddard, W. A. DREIDING: a generic force field for molecular simulations. *Journal of Physical chemistry* **1990**, *94*, 8897–8909.

(28) 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.

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

(30) 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* **2018**,

(31) Frenkel, D.; Smit, B. *Understanding molecular simulation: from algorithms to applications*; Elsevier, 2001; Vol. 1.

(32) Coudert, F.-X.; Fuchs, A. H. Computational characterization and prediction of metal–organic framework properties. *Coordination Chemistry Reviews* **2016**, *307*, 211–236.

(33) 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.

(34) 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.

(35) 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.

(36) Heinen, J.; Dubbeldam, D. On flexible force fields for metal–organic frameworks: Recent developments and future prospects. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2018**, e1363.

(37) 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.

(38) Li, C.-P.; Du, M. Role of solvents in coordination supramolecular systems. *Chemical Communications* **2011**, *47*, 5958–5972.

(39) 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* **2016**, *29*, 2521–2528.

(40) 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.

(41) Shell, M. S. *Thermodynamics and statistical mechanics: an integrated approach*; Cambridge University Press, 2015.

(42) 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.

(43) Tijms, H. *Probability: A Lively Introduction*; Cambridge University Press, 2017.

(44) Evans, J. D.; Fraux, G.; Gaillac, R. et al. Computational chemistry methods for nanoporous materials. *Chemistry of Materials* **2016**, *29*, 199–212.

(45) Fraux, G.; Coudert, F.-X. Recent advances in the computational chemistry of soft porous crystals. *Chemical Communications* **2017**, *53*, 7211–7221.

(46) Dubbeldam, D.; Torres-Knoop, A.; Walton, K. S. On the inner workings of Monte Carlo codes. *Molecular Simulation* **2013**, *39*, 1253–1292.

(47) 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.

(48) Simon, C.; York, A. H.; Sturluson, A. et al. PorousMaterials.jl. 2018; `https://github.com/SimonEnsemble/PorousMaterials.jl`.

(49) 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.

(50) 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* **2018**, *1*.

(51) 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.

(52) Simon, C. M.; Kim, J.; Gomez-Gualdron, D. A. et al. The materials genome in action: identifying the performance limits for methane storage. *Energy & Environmental Science* **2015**, *8*, 1190–1199.

(53) 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.

(54) 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.

(55) Becker, T. M.; Heinen, J.; Dubbeldam, 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.

(56) Zhang, H.; Snurr, R. Q. Computational Study of Water Adsorption in the Hydrophobic MetalOrganic Framework ZIF-8: Adsorption Mechanism and Acceleration of the Simulations. *The Journal of Physical Chemistry C* **2017**, *121*, 24000–24010.

(57) Friedman, J.; Hastie, T.; Tibshirani, R. *The elements of statistical learning*; Springer series in statistics New York, NY, USA:, 2001; Vol. 1.

(58) 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.

(59) 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.

(60) 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.

(61) 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.

(62) 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**,

(63) Pardakhti, M.; Moharreri, E.; Wanik, D. et al. Machine learning using combined structural and chemical descriptors for prediction of methane adsorp-

tion performance of metal organic frameworks (MOFs). *ACS combinatorial science* **2017**, *19*, 640–645.

(64) 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.

(65) 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* **2014**, *119*, 186–195.

(66) Chung, Y. G.; Gómez-Gualdrón, D. A.; Li, P. et al. In silico discovery of metal-organic frameworks for precombustion CO_2 capture using a genetic algorithm. *Science Advances* **2016**, *2*, e1600909.

(67) Doudna, J. A.; Charpentier, E. The new frontier of genome engineering with CRISPR-Cas9. *Science* **2014**, *346*, 1258096.

(68) Flavahan, W. A.; Gaskell, E.; Bernstein, B. E. Epigenetic plasticity and the hallmarks of cancer. *Science* **2017**, *357*, eaal2380.

(69) Barrangou, R.; Fremaux, C.; Deveau, H. et al. CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **2007**, *315*, 1709–1712.

(70) Lander, E. S. The heroes of CRISPR. *Cell* **2016**, *164*, 18–28.

(71) 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.

(72) Jansen, R.; Embden, J. D. v.; Gaastra, W. et al. Identification of genes that are associated with DNA repeats in prokaryotes. *Molecular microbiology* **2002**, *43*, 1565–1575.

(73) Mojica, F. J.; García-Martínez, J.; Soria, E. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *Journal of Molecular Evolution* **2005**, *60*, 174–182.

(74) Goodman, L. Hypothesis-limited research. *Genome research* **1999**, *9*, 673–676.

(75) Gándara, F.; Bennett, T. D. Crystallography of metal–organic frameworks. *IUCrJ* **2014**, *1*, 563–570.

(76) Tremayne, M.; Harris, K. D. M. Crystal structure determination from powder diffraction data. *Chemistry of Materials* **1996**, *8*, 2554–2570.

(77) McCusker, L. B.; Von Dreele, R. B.; Cox, D. E. et al. Rietveld refinement guidelines. *Journal of Applied Crystallography* **1999**, *32*, 36–50.

(78) 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.

(79) OKeeffe, M.; Yaghi, O. M. Deconstructing the crystal structures of metal–organic frameworks and related materials into their underlying nets. *Chemical Reviews* **2011**, *112*, 675–702.

(80) Parsons, S. Introduction to twinning. *Acta Crystallographica Section D* **2003**, *59*, 1995–2003.

(81) 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.

(82) 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.

(83) Hoshino, M.; Khutia, A.; Xing, H. et al. The crystalline sponge method updated. *IUCrJ* **2016**, *3*, 139–151.

(84) Spek, A. L. PLATON SQUEEZE: a tool for the calculation of the disordered solvent contribution to the calculated structure factors. *Acta Crystallographica Section C: Structural Chemistry* **2015**, *71*, 9–18.

(85) 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.

(86) 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.

(87) 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.

(88) 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.

(89) 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.

(90) Park, J. H.; Choi, K. M.; Jeon, H. J. 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.

(91) 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.

(92) 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.

(93) Julien, P. A.; Užarević, K.; Katsenis, 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.

(94) Peterson, V. K.; Southon, P. D.; Halder, G. J. et al. Guest Adsorption in the Nanoporous Metal–Organic Framework Cu3 (1, 3, 5-Benzenetricarboxylate) 2: Combined In Situ X-Ray Diffraction and Vapor Sorption. *Chemistry of Materials* **2014**, *26*, 4712–4723.

(95) Gonzalez, M. I.; Mason, J. A.; Bloch, E. D. et al. Structural characterization of framework–gas interactions in the metal–organic framework Co 2 (dobdc) by in situ single-crystal X-ray diffraction. *Chemical Science* **2017**, *8*, 4387–4398.

(96) Walton, R. I.; Munn, A. S.; Guillou, N. et al. Uptake of Liquid Alcohols by the Flexible FeIII Metal–Organic Framework MIL-53 Observed by Time-Resolved In Situ X-ray Diffraction. *Chemistry–A European Journal* **2011**, *17*, 7069–7079.

(97) 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.

(98) Allen, F. H. The Cambridge Structural Database: a quarter of a million crys-

tal structures and rising. *Acta Crystallographica Section B: Structural Science* **2002**, *58*, 380–388.

(99) 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.

(100) 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* **2002**, *58*, 389–397.

(101) Bruno, I. J.; Groom, C. R. A crystallographic perspective on sharing data and knowledge. *Journal of Computer-aided Molecular Design* **2014**, *28*, 1015–1022.

(102) 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.

(103) Goldsmith, J.; Wong-Foy, A. G.; Cafarella, M. J. et al. Theoretical limits of hydrogen storage in metal–organic frameworks: Opportunities and tradeoffs. *Chemistry of Materials* **2013**, *25*, 3373–3382.

(104) Chung, Y. G.; Haldoupis, E.; Bucior, B. J. et al. Advances, Updates, and Analytics for the Computation-Ready, Experimental Metal-Organic Framework Database: CoRE MOF 2019. *???* **???**, *??*, ???

(105) Batten, S. R.; Champness, N. R.; Chen, X.-M. et al. Coordination polymers, metalorganic frameworks and the need for terminology guidelines. *CrystEngComm* **2012**, *14*, 3001–3004.

(106) Seth, S.; Matzger, A. J. MetalOrganic Frameworks: Examples, Counterexamples, and an Actionable Definition. *Crystal Growth & Design* **2017**, *17*, 4043–4048.

(107) 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.

(108) 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* **2016**, *29*, 26–39.

(109) Murdock, C. R.; McNutt, N. W.; Keffer, 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.

(110) 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.

(111) Van Heest, T.; Teich-McGoldrick, S. L.; Greathouse, J. A. et al. Identification of MetalOrganic 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.

(112) Panella, B.; Hirscher, M.; Roth, S. Hydrogen adsorption in different carbon nanostructures. *Carbon* **2005**, *43*, 2209 – 2214.

(113) Watanabe, T.; Sholl, D. S. Accelerating Applications of MetalOrganic Frameworks for Gas Adsorption and Separation

by Computational Screening of Materials. *Langmuir* **2012**, *28*, 14114–14128.

(114) Pinheiro, M.; Martin, R. L.; Rycroft, C. H. et al. High accuracy geometric analysis of crystalline porous materials. *CrystEngComm* **2013**, *15*, 7531–7538.

(115) 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.

(116) Computation-ready, Experimental (CoRE) MOFs download link. `http://gregchung.github.io/CoRE-MOFs/`.

(117) 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.

(118) Barthel, S.; Alexandrov, E. V.; Proserpio, D. M. et al. Distinguishing MetalOrganic Frameworks. *Crystal Growth & Design* **2018**, *18*, 1738–1747.

(119) Horike, S.; Shimomura, S.; Kitagawa, S. Soft porous crystals. *Nature Chemistry* **2009**, *1*, 695.

(120) 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.

(121) Walton, K. S.; Millward, A. R.; Dubbeldam, 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.

(122) Rappe, A. K.; Goddard III, W. A. Charge equilibration for molecular dynamics simulations. *The Journal of Physical Chemistry* **1991**, *95*, 3358–3363.

(123) 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.

(124) 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.

(125) Diercks, C. S.; Yaghi, O. M. The atom, the molecule, and the covalent organic framework. *Science* **2017**, *355*, eaal1585.

(126) Hasell, T.; Cooper, A. I. Porous organic cages: soluble, modular and molecular pores. *Nature Reviews Materials* **2016**, *1*, 16053.

(127) 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.

(128) 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**,

(129) Miklitz, M.; Jiang, S.; Clowes, R. et al. Computational screening of porous organic molecules for xenon/krypton separation. *The Journal of Physical Chemistry C* **2017**, *121*, 15211–15222.

(130) Sturluson, A.; Huynh, M. T.; York, A. H. P. et al. Eigencages: Learning a Latent Space of Porous Cage Molecules. *ACS Central Science* **2018**,

(131) Baerlocher, C.; McCusker, L. B. Database of Zeolite Structures. http://www.iza-structure.org/databases/.

(132) Energy Information Administration, U. How much carbon dioxide is produced when different fuels are burned? 2018; https://www.eia.gov/tools/faqs/faq.php?id=73&t=11.

(133) Wang, M. Q.; Huang, H. S. *A full fuel-cycle analysis of energy and emissions impacts of transportation fuels produced from natural gas*; 2000.

(134) 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**,

(135) 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.

(136) Davis, S. C.; Diegel, S. W.; Boundy, R. G. *Transportation energy data book*; 2009.

(137) 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.

(138) Makal, T. A.; Li, J.-R.; Lu, W. et al. Methane storage in advanced porous materials. *Chemical Society Reviews* **2012**, *41*, 7761–7779.

(139) (ARPA-E), A. R. P. A.-E. MOVE (Methane Opportunities for Vehicular Energy) Program Overview. https://arpa-e.energy.gov/sites/default/files/documents/files/MOVE_ProgramOverview.pdf, 2012.

(140) 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.

(141) Wilmer, C. E.; Leaf, M.; Lee, C. Y. et al. Large-scale screening of hypothetical metal–organic frameworks. *Nature Chemistry* **2012**, *4*, 83.

(142) 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.

(143) 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.

(144) Lin, X.; Blake, A. J.; Dailly, A. 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.

(145) Wilmer, C. E.; Farha, O. K.; Yildirim, T. et al. Gram-scale, high-yield synthesis of a robust metalorganic framework for storing methane and other gases. *Energy & Environmental Science* **2013**, *6*, 1158–1163.

(146) Gomez-Gualdron, 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.

(147) Wu, H.; Zhou, W.; Yildirim, T. High-capacity methane storage in metal- organic frameworks M2 (dhtp): the important role of open metal sites. *Journal*

of the American Chemical Society **2009**, *131*, 4995–5000.

(148) 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.

(149) Gandara, 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.

(150) 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.

(151) 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.

(152) 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.

(153) DOE Technical Targets for Onboard Hydrogen Storage for Light-Duty Vehicles. `https://www.energy.gov/eere/fuelcells/doe-technical-targets-onboard-hydrogen-storage-light-duty-vehicles`. Accessed: 2018-12-12.

(154) Murray, L. J.; Dincă, M.; Long, J. R. Hydrogen storage in metal–organic frameworks. *Chemical Society Reviews* **2009**, *38*, 1294–1314.

(155) Ahmed, A.; Liu, Y.; Purewal, J. et al. Balancing gravimetric and volumetric hydrogen density in MOFs. *Energy & Environmental Science* **2017**, *10*, 2459–2471.

(156) Kaye, S. S.; Dailly, A.; Yaghi, O. M. et al. Impact of preparation and handling on the hydrogen storage properties of Zn4O (1, 4-benzenedicarboxylate) 3 (MOF-5). *Journal of the American Chemical Society* **2007**, *129*, 14176–14177.

(157) Rowsell, J. L.; 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.

(158) 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.

(159) 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.

(160) 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.

(161) 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.

(162) 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.

(163) West, E. 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.

(164) 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.

(165) McGovern, S. J.; Yeigh Jr, J. H. FCC Regeneration. 1983; US Patent 4,370,222.

(166) 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.

(167) 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.

(168) 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.

(169) 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*, 435–438.

(170) Herm, Z. R.; Bloch, E. D.; Long, J. R. Hydrocarbon separations in metal–organic frameworks. *Chemistry of Materials* **2013**, *26*, 323–338.

(171) Lenzen, M. Life cycle energy and greenhouse gas emissions of nuclear energy: A review. *Energy conversion and management* **2008**, *49*, 2178–2199.

(172) Sood, D. D.; Patil, S. K. Chemistry of nuclear fuel reprocessing: Current status. *Journal of radioanalytical and nuclear chemistry* **1996**, *203*, 547–573.

(173) 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*.

(174) Banerjee, D.; Cairns, A. J.; Liu, J. et al. Potential of metal–organic frameworks for separation of xenon and krypton. *Accounts of chemical research* **2014**, *48*, 211–219.

(175) 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.

(176) Banerjee, D.; Simon, C. M.; Elsaidi, S. K. et al. Xenon Gas Separation and Storage Using Metal-Organic Frameworks. *Chem* **2018**, *4*, 466–494.

(177) Banerjee, D.; Simon, C. M.; Plonka, A. M. et al. Metal–organic framework with optimally selective xenon adsorption and separation. *Nature Communications* **2016**, *7*, ncomms11831.

(178) Banerjee, D.; Zhang, Z.; Plonka, A. M. et al. A calcium coordination framework having permanent porosity and high CO_2/N_2 selectivity. *Crystal Growth & Design* **2012**, *12*, 2162–2165.

(179) Sikora, B. J.; Wilmer, C. E.; Greenfield, M. L. et al. Thermodynamic analysis of Xe/Kr selectivity in over 137000 hypothetical metal–organic frameworks. *Chemical Science* **2012**, *3*, 2217–2223.

(180) 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.

(181) Vassiliki Aroniadou-Anderjaska, J. P. A. F. Q., Taiza H. Figueiredo; Braga, M. F. Primary brain targets of nerve agents. *National Institute of Health* **2009**, *30*, 772–776.

(182) Milatović, D.; Jokanović, M. In *Handbook of Toxicology of Chemical Warfare Agents*; Gupta, R. C., Ed.; Academic Press: San Diego, 2009; pp 985–996.

(183) Mayank Agrawal, J. A. G., Dorina F. Sava Gallis; Sholl, D. S. How Useful Are Common Simulants of Chemical Warfare Agents at Predicting Adsorption Behavior? *The Royal Society of Chemistry* **2018**, *122*, 26061–26069.

(184) Taiza H. Figueiredo, M. F. M. B. A. M. M., James P. Apland Acute and long-term consequences of exposure to organophosphate nerve agents in humans. *Epilepsia* **2018**, *59*, 92–99.

(185) Barea, E.; Montoro, C.; Navarro, J. A. R. Toxic gas removal metalorganic frameworks for the capture and degradation of toxic gases and vapours. *The Royal Society of Chemistry* **2014**, *43*, 5419–5430.

(186) DeCoste, J. B.; Peterson, G. W. Metal–organic frameworks for air purification of toxic chemicals. *Chemical Reviews* **2014**, *114*, 5695–5727.

(187) 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**,

(188) Küsgens, P.; Rose, M.; Senkovska, I. et al. Characterization of metal-organic frameworks by water adsorption. *Microporous and Mesoporous Materials* **2009**, *120*, 325–330.

(189) 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.

(190) Smil, V. Energy in the twentieth century: resources, conversions, costs, uses, and consequences. *Annual Review of Energy and the Environment* **2000**, *25*, 21–51.

(191) 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, p 465570.

(192) 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, p 10291136.

(193) Sumida, K.; Rogow, D. L.; Mason, J. A. et al. Carbon dioxide capture in metal–organic frameworks. *Chemical Reviews* **2011**, *112*, 724–781.

(194) Lal, R. Carbon sequestration. *Philosophical Transactions of the Royal Society B: Biological Sciences* **2007**, *363*, 815–830.

(195) Haszeldine, R. S. Carbon capture and storage: how green can black be? *Science* **2009**, *325*, 1647–1652.

(196) Whitley, D. A genetic algorithm tutorial. *Statistics and computing* **1994**, *4*, 65–85.

(197) 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.

(198) 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 CuBTTri. *Chemical Science* **2011**, *2*, 2022–2028.

(199) Cannella, W. J. Xylenes and ethylbenzene. *Kirk-Othmer Encyclopedia of Chemical Technology* **2000**,

(200) 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.

(201) 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.

(202) Minceva, M.; Rodrigues, A. E. Understanding and revamping of industrial scale SMB units for p-xylene separation. *AIChE journal* **2007**, *53*, 138–149.

(203) 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.

(204) 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.

(205) LeCun, Y.; Cortes, C.; Burges, C. J. C. The MNIST database of handwritten digits. *http://yann. lecun. com/exdb/mnist/* **1998**,

(206) Wan, L.; Zeiler, M.; Zhang, S. et al. Regularization of neural networks using dropconnect. International Conference on Machine Learning. 2013; pp 1058–1066.

(207) Krizhevsky, A.; Hinton, G. *Learning multiple layers of features from tiny images*; 2009.

(208) Bennett, J.; Lanning, S. The Netflix Prize. Proceedings of the KDD Cup Workshop 2007. New York, 2007; pp 3–6.

(209) Bell, R. M.; Koren, Y. Lessons from the Netflix prize challenge. *ACM Sigkdd Explorations Newsletter* **2007**, *9*, 75–79.

(210) Koren, Y.; Bell, R.; Volinsky, C. Matrix factorization techniques for recommender systems. *Computer* **2009**, 30–37.

(211) Koren, Y. The bellkor solution to the netflix grand prize. *Netflix prize documentation* **2009**, *81*, 1–10.

(212) Russakovsky, O.; Deng, J.; Su, H. et al. Imagenet large scale visual recognition challenge. *International Journal of Computer Vision* **2015**, *115*, 211–252.

(213) 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.

(214) 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.

(215) `https://adsorption.nist.gov/factlab`, Accessed 5 February 2019.

(216) 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.

(217) 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.

(218) Heller, S. R.; McNaught, A.; Pletnev, I. et al. InChI, the IUPAC International Chemical Identifier. *Journal of Cheminformatics* **2015**, *7*, 23.

(219) 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.

(220) `https://www.sigmaaldrich.com/catalog/product/aldrich/688614?lang=en&region=US`, Accessed 5 February 2019.

(221) 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.

(222) Park, J.; Howe, D. S., Joshua D. Sholl How Reproducible Are Isotherm Measurements in Metal-Organic Frameworks? *Chemistry of Materials* **2017**, *29*, 10487–10495.

(223) Simon, C. M.; Smith, B.; Haranczyk, M. pyIAST: Ideal adsorbed solution theory (IAST) Python package. *Computer Physics Communications* **2016**, *200*.

(224) Simon, C. M.; Smit, B.; Haranczyk, M. Documentation for pyIAST. `https://pyiast.readthedocs.io`, Accessed 1 February 2019.

(225) 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.

(226) 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.

(227) `http://www.fluidproperties.org`, Accessed 25 February 2019.

(228) Ross, R. B.; Ahmad, R.; Brennan, J. K. et al. The Seventh Industrial Fluid Properties Simulation Challenge. *Fluid Phase Equil.* **2014**, *366*, 136–140.

(229) 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.

(230) Schultz, N. E.; Riaz, A.; Brennan, J. K. et al. The Eighth Industrial Fluid Properties Simulation Challenge. *Adsorption Sci. Tech.* **2016**, *34*, 3–12.

(231) 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.

(232) Rouquerol, F.; Rouquerol, J.; Sing, K. *Adsorption by Powders and Porous Solids*; Academic Press: London, 1999.

(233) `https://adsorption.nist.gov/isodb/index.php#apis`, Accessed 22 February 2019.

(234) 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* **1991**, *A47*, 688–685.

(235) 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.

(236) 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.

(237) Ceder, G. Opportunities and challenges for first-principles materials design and applications to Li battery materials. *MRS bulletin* **2010**, *35*, 693–701.

(238) 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.

(239) Xu, J.; Hagler, A. Chemoinformatics and drug discovery. *Molecules* **2002**, *7*, 566–600.

(240) 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.

(241) Bachman, J.; Reed, D. A.; Kapelewski, M. T. et al. Enabling Alternative Ethylene Production through Its Selective Adsorption in the Metal–Organic Framework Mn_2(*m*-dobdc). *Energy & Environmental Science* **2018**,

(242) 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.

(243) 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.

(244) 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.

(245) 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.

(246) Coudert, F.-X. Reproducible Research in Computational Chemistry of Materials. *Chemistry of Materials* **2017**, *29*, 2615–2617.

(247) Greathouse, J. A.; Kinnibrugh, 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.

(248) Boutin, A.; Springuel-Huet, M.-A.; Nossov, A. et al. Breathing Transitions in MIL-53 (Al) Metal–Organic Framework Upon Xenon Adsorption. *Angewandte Chemie* **2009**, *121*, 8464–8467.

(249) Gonzalez-Nelson, A.; Coudert, F.; van der Veen, M. Rotational Dynamics of Linkers in MetalOrganic Frameworks. *Nanomaterials* **2019**, *9*, 330.

(250) Catalano, L.; Naumov, P. Exploiting rotational motion in molecular crystals. *CrystEngComm* **2018**, *20*, 5872–5883.

(251) 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.

(252) 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.

(253) 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.

(254) 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.

(255) 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.

(256) 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.

(257) 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.

(258) 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.

(259) Behler, J.; Parrinello, M. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Physical Review Letters* **2007**, *98*, 146401.

(260) 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.

(261) Behler, J. Neural network potential-energy surfaces in chemistry: a tool for large-scale simulations. *Physical Chemistry Chemical Physics* **2011**, *13*, 17930–17955.

(262) Schütt, K. T.; Arbabzadah, F.; Chmiela, S. et al. Quantum-chemical insights from deep tensor neural networks. *Nature Communications* **2017**, *8*, 13890.

(263) Hornik, K.; Stinchcombe, M.; White, H. Multilayer feedforward networks are universal approximators. *Neural Networks* **1989**, *2*, 359–366.

(264) 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.

(265) Sholl, D. S.; Lively, R. P. Defects in metal–organic frameworks: challenge or opportunity? *The Journal of Physical Chemistry Letters* **2015**, *6*, 3437–3444.

(266) 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.

(267) 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.

(268) Thornton, A.; Babarao, R.; Jain, A. et al. Defects in metal–organic frameworks: a compromise between adsorption and stability? *Dalton Transactions* **2016**, *45*, 4352–4359.

(269) 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.

(270) 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.

Graphical TOC Entry