Advances, Updates, and Analytics for the Computation-Ready, Experimental Metal-Organic Framework Database: CoRE MOF 2019

Yongchul G. Chung[1 ‡*], Emmanuel Haldoupis[2‡], Benjamin J. Bucior[3‡], Maciej Haranczyk[4,5], Seulchan Lee[1], Hongda Zhang[3], Konstantinos D. Vogiatzis[2,6], Marija Milisavljevic[7], Sanliang Ling[8,9], Jeff S. Camp[7], Ben Slater[9], J. Ilja Siepmann[2,10*], David S. Sholl[7*], Randall Q. Snurr[3*]

[1]School of Chemical and Biomolecular Engineering, Pusan National University, Busan 46241, Korea (South)

[2]Department of Chemistry and Chemical Theory Center, University of Minnesota, 207 Pleasant Street SE, Minneapolis, MN 55455, United States

[3]Department of Chemical and Biological Engineering, Northwestern University, Evanston, IL 60208, United States

[4]Computational Research Division, Lawrence Berkeley National Laboratory, One Cyclotron Road, MS 50F-1650, Berkeley, CA 94720, United States

[5]IMDEA Materials Institute, C/Eric Kandel 2, 2890 6 – Getafe, Madrid, Spain

[6]Department of Chemistry, University of Tennessee, Knoxville, TN 37996, United States

[7]School of Chemical & Biomolecular Engineering, Georgia Institute of Technology, Atlanta, GA 30332, United States

[8]Advanced Materials Research Group, Faculty of Engineering, University of Nottingham, University Park, Nottingham NG7 2RD, U.K.

[9]Department of Chemistry, University College London, 20 Gordon Street, London WC1H 0AJ, United Kingdom

[10]Department of Chemical Engineering and Materials Science, University of Minnesota, 421 Washington Avenue SE, Minneapolis, MN 55455, United States

Abstract

Over 14,000 porous, three-dimensional metal-organic framework structures are compiled and analyzed as part of an update to the Computation-Ready, Experimental Metal-Organic Framework Database (CoRE MOF Database). The updated database includes additional structures that were contributed by CoRE MOF users, obtained from updates of the Cambridge Structural Database and a Web of Science search, and derived through semi-automated reconstruction of disordered structures using a topology-based crystal generator. In addition, value is added to the CoRE MOF database through new analyses that can speed up future nanoporous materials discovery activities, including open metal site detection and duplicate searches. Crystal structures (only for the subset that underwent significant changes during curation), pore analytics, and physical property data are included with the publicly available CoRE MOF 2019 database.

INTRODUCTION

Metal-organic frameworks (MOFs) are a class of porous materials assembled from inorganic clusters ("secondary building units") and organic building blocks, where these building blocks are arranged to form an extended 3-dimensional (or 2-dimensional) crystalline solid material.[1-2] This class of materials generally has high porosity, large internal surface area, and often contains active sites that can be used for site-specific adsorption and catalysis. These properties make MOFs attractive platform materials for applications in gas storage,[3] chemical separation[4-5], catalysis[6], drug delivery[7], and chemical sensing.[8] There are many types of organic and inorganic building blocks that can be used for MOF synthesis, and a large number of new MOFs can be rationally designed using different combinations of organic and inorganic building blocks. Because of this, the number of MOFs reported in the literature has seen rapid growth in recent years, and tens of thousands of porous MOFs and MOF-type structures have been synthesized to date.[9] A challenge associated with such a large number of reported structures is to find high-performing materials for an application of interest.

One possible avenue to speed up the discovery of high-performing MOFs is to create a machine-readable database for already synthesized MOFs. As a part of the Nanoporous Materials Genome Center (NMGC), some of us previously created such a database, called the Computation-Ready, Experimental MOF database (CoRE MOF 2014 Database)[10] derived from structures deposited in the Cambridge Structural Database (CSD).[11] Unlike other porous materials databases where the structures have been generated using computer algorithms,[12] MOF structures in the CoRE MOF database are derived from synthesized materials and the synthesis protocols (and sometimes also the activation procedures used to remove solvent molecules trapped in the MOF structure) are usually available; this knowledge can facilitate the synthesis and subsequent testing of high-

performing MOFs emerging from computational high-throughput screening studies. The utility of the CoRE MOF Database has been demonstrated by its use in a growing number of computationally-guided screening studies and materials discovery activities (largely focused on adsorption properties of MOFs) for methane storage,[13] natural gas storage[14] and purification,[15] carbon capture[16] and separation,[17-18] Xe/Kr separation,[19] xylene separations,[20] olefin/paraffin separations,[21-22] conductivity,[23] oxidative catalysis,[24] and hexane isomer separations.[25] Moreover, the CoRE MOF database has been used to obtain adsorption properties in deformed MOFs,[26] to predict gas adsorption properties in MOFs upon introducing defect sites,[27] and to determine the possible crystal structures of synthesized MOFs.[28] These selected examples demonstrate that the CoRE MOF database can help researchers to discover new utility for already existing MOFs.

Several subsets of the CoRE MOF Database have been refined since publication of the database in 2014.[10] Based on the original CoRE MOF structures, Nazarian and co-workers determined partial atomic charges using density functional theory (DFT) calculations for more than 50% of the CoRE MOF structures.[29] This subset is referred to as the CoRE MOF 2014-DDEC database, and it can be used to screen MOFs where first-order electrostatic interactions are important, such as adsorption of molecules with large dipole or quadrupole moments. In another work, the original CoRE MOF structures were optimized using DFT calculations to investigate structural relaxation and, in particular, the effect of solvent removed during the generation of the CoRE MOF structures and also to probe the differences in the adsorption properties (CH_4 and CO_2 guest molecules) in the as-is and optimized structures.[30]

In 2017, a collection of MOF and MOF-type structures found within the CSD was reported independently and incorporated into the CSD itself as a separate MOF database.[9] This was done using a different procedure and a broad definition of what constitutes a MOF and led to the

inclusion of 1D, 2D, and 3D structures even when they have no apparent porosity. Recently, Altintas et al. investigated the adsorption of CH_4 and H_2 in 3,490 structures present in both the CSD and the CoRE MOF 2014 databases and found that 387 MOFs yield significantly different loadings for the structures taken from the two databases for five reasons (missing hydrogen atoms, removal of unbound solvent molecules, removal of bound solvent molecules, removal of (disordered) charge-balancing ions, and removal of MOF fragments).[31] This study illustrates the challenges encountered in the preparation of large databases.

The updated CoRE MOF 2019 Database (as also the CoRE MOF 2014 Database), on the other hand, aims to be a more targeted selection of 3D porous MOFs and is prepared to be directly used in molecular simulations or electronic structure calculations for applications involving storage, separation, and chemical transformations of guest molecules. It should be emphasized that the main purpose of the CoRE MOF database is to allow for high-throughput screening of a large number of MOF structures to find a subset of potentially promising structures for a specific application. Subsequently, this smaller subset of potentially promising structures needs to be analyzed in more detail, e.g., checking for stability of these structures and feasibility of solvent removal, performing geometry optimization, and re-computing properties for optimized structures. Such more detailed analysis would be computationally prohibitive for the entire CoRE MOF database.

In this work, we report a significant update of the CoRE MOF 2014 Database, that nearly triples the number of structures to over 14,000. The additional structures originated from several different sources (contributed by CoRE MOF users, obtained from updates of the Cambridge Structural Database, and a Web of Science search). The updated database is called the CoRE MOF 2019 Database. Following the compilation of the new MOF structures for the database, we analyze the

complete set of structures for the presence of open metal sites (OMS), investigate the effect of bound solvent removal on the calculated geometric properties, and consider how many real MOFs can be found in a set of computer-generated, hypothetical MOFs using a newly introduced molecular fingerprint approach. The atomic structures contained in the CoRE MOF 2019 Database are publicly available at http://dx.doi.org/10.11578/1118280. In cases where entries in the database have been modified from entries in the CSD or are materials not included in the CSD, the full crystal structure of the material is available. In instances where no modifications to the CSD structure was required, our database simply provides the CSD REFCODE.

COMPUTATIONAL METHODS

Datasets in the CoRE MOF 2019 Database. Reported MOF crystal structures often contain coordinates for solvent molecules in the pores. In the CoRE MOF 2014 Database, these solvent molecules were removed, including both molecules that were bound directly to an otherwise under-coordinated metal site and "free" solvent molecules that interact only loosely with the framework. The solvent removal procedure mimics the outcome of the experimental "activation" procedure and creates OMSs in some MOF crystals. Removing the bound solvent can help discover MOFs that would otherwise be labeled as non-porous (defined here as pore size < 2.4 Å; see below) during high-throughput characterization. However, removal of bound (and free) solvent molecules may have a profound impact on the MOF structures because the presence of these solvent molecules may be essential to the framework stability in some MOF crystals or to stabilize charge-balancing ions in their crystallographic locations. In addition, important geometric properties, such as the pore limiting diameter (PLD), largest cavity diameter (LCD), gravimetric and volumetric surface areas (GSA and VSA, respectively)[32] will change upon removal of bound solvent molecules. Since these geometric properties are typically the first descriptor for selecting

promising MOF candidates from a pool of MOF structures, it is important to assess the effect of bound solvent removal on the abovementioned physical properties.

To investigate the effect of removing bound solvents on geometric properties, we prepared the following two sets of the CoRE MOF 2019 Database: (a) Free Solvent Removed (FSR), for which only the free solvent molecules are removed from the structures, and (b) All Solvent Removed (ASR), for which both bound and free solvent molecules are removed from the structures (Figure 1a). The structures in the CoRE MOF Database 2019-ASR set have undergone the same curation procedures as the structures in the CoRE MOF 2014 database. Briefly, the curation process proceeds as follows: 1) removing coordinates with low partial occupancies (as marked by an asterisk next to the atom type in the CSD CIF files), 2) converting the structure to P1 symmetry, 3) removing free solvent molecules, and 4) removing bound solvent molecules. (Removing coordinates for atoms with low occupancy carries the risk of generating partially incomplete structures that cannot be synthesized in the computer-generated form.) The following compounds are the most common solvent molecules removed by the procedure: water, dimethylformamide, N-ethylpropanoamide, dimethyl sulfoxide, 2-imidazolidone, and lone oxygen atom originating, presumably, from a water molecule with crystallographically unresolved hydrogen atoms. On the other hand, when detected, non-neutral hydroxyl groups (-OH) were not removed. A solvent molecule is considered as bound when the metal-oxygen distance is less than the sum of their atomic radii plus 0.4 Å. The atomic radii used in this work are the same as the covalent radii defined by Cordero and coworkers.[33] Following these steps, the structures were passed through the open source code `PorousMaterials.jl`[34] to check for disordered structures. We marked the structure as disordered if the distance between two atoms for a given crystal structure is less than 0.1 Å, as implemented in the `PorousMaterials.jl` function `atomic_overlap`.

Figure 1a illustrates how these different subsets were created. Only porous MOFs were kept in each set, a choice that results in a larger number of MOFs in the ASR set than in the FSR set. A material was considered nonporous if its PLD is less than 2.4 Å. Further details on these methods are given in the CoRE MOF 2014 work.[10]

MOF Structures from the Cambridge Structural Database. The CSD contains the largest number of experimental crystal structures[11, 35,36] of any extant material database. The structures from the CoRE MOF 2014 Database were derived exclusively from the CSD database. However, the CSD continues to update the list of crystal structures and, thus, includes many MOF structures added since 2014. To include these newly reported MOF structures, we obtained the updated crystal structures from the CSD that were included in the November 2017 release. We carried out the same procedure as we did for the CoRE MOF 2014 Database[10] on the newly obtained structures, which includes solvent removal (both bound and free), ion retention, and manual structure adjustments. Based on this procedure, we added more than 9,000 MOF structures to the CoRE MOF 2019-ASR Database.

MOF Structures from the Literature and Duplicate Removal. Not all crystal structures of experimentally synthesized MOFs are available from the CSD. Many MOF crystal structures are deposited to the CSD because some publishers (e.g., Royal Society of Chemistry) require authors to deposit the structures to the CSD before the articles can be published in their journals. Other journals do not have such requirements, which means that some MOF structures may not be deposited in the CSD. In such instances, the authors typically append CIF files as supporting information, or the information is provided as part of the main manuscript. To include some of these experimentally synthesized MOF structures from the literature, we collected missing MOF structures as identified based on their Digital Object Identifier (DOI) strings. Specifically, we

obtained the DOI strings for MOF-related publications by searching the Web of Science (WoS) for the terms "porous organic polymer," and "metal-organic frameworks" (Figure 1c). It should be noted that this search is not exhaustive and some MOF structures described by authors using other terms will be missed, but the effort increases exponentially when including less common search terms to avoid misidentifications (e.g., a significant fraction of "covalent organic frameworks" do not constitute MOFs). Ensuring that every synthesized MOF is included would be an impossible task. To make sure we did not download structures that we already obtained from the CSD, we compared the DOI strings between the structures from the CSD and the Web of Science. We removed DOI strings for structures that were already in the CSD. After this procedure, we manually downloaded the structures, and carried out solvent removal (both bound and free) and manual structure fixing procedures. We note that the ion retention algorithm, which relies on the known stoichiometry information provided by the CSD, was not performed for the structures obtained in this manner because the stoichiometry information is not readily available for these structures. In order to make sure we do not report the same structure multiple times with different filenames, we developed a procedure to compare CIF files to find identical structures. Briefly, the procedure checks the volume, chemical composition, lattice constants, and the position of atoms for a CIF file with the original CIFs from the CSD to determine if the structures are the same or not. A similar method is available from the Pymatgen StructureMatcher algorithm,[37] and we cross-checked both methods as a consistency test. A Python script that implements our own procedure is available as a Jupyter Notebook.[38] Finally, we manually checked the structures that were downloaded from the web against the dataset obtained from CSD and removed the CSD structures if an overlap was found. We added 900 new MOF structures to the CoRE MOF 2019 update using this WoS method.

User Contributions. Since publication of the CoRE MOF 2014 Database, we have received inputs from the user community regarding disorder and errors associated with a number of these structures. Some structures (e.g., the MIL-53 series) were missing hydrogen atoms on the bridging hydroxyl group, others did not contain organic subunits. Nearly 100 MOF structures that were manually fixed by users are updated in the CoRE MOF 2019 Database. We encourage the user community to continue to contribute to the CoRE MOF database by contacting one of the corresponding authors if any problems with structures in the database are found, or by directly submitting the request via the Github repository for the CoRE MOF database. The Github repository can be accessed from the link provided in http://dx.doi.org/10.11578/1118280.

Semi-Automated Corrections to Crystal Structures. Many CIF files obtained from the CSD and the Web of Science search contain a degree of disorder, such as partial occupancies and missing hydrogen atoms, that make them unsuitable for computational screening. Indeed, Moghadam et al. provided a smaller subset of structures identified as non-disordered MOFs.[9] Fixing disorder in MOF structures to generate simulation-ready structures typically requires time-consuming manual work. Instead of relying solely on adjustments by hand, we created a workflow to reconstruct disordered crystal structures by combining a topology-based crystal generator and the Materials Studio Forcite Module (Figure 1b). Briefly, the code constructs a crystal structure based on a given topology, nodes, and edges. Here, nodes and edges are derived from the chemical building blocks that form MOF structures, and topology is a blueprint on which the nodes and edges are placed. Details of the algorithms are reported by Colón and co-workers.[39] Crystal structures were reconstructed based on the topological blueprints obtained from the Reticular Chemistry Structure Resource (RCSR)[40] if not already available in the library of topologies included with the crystal generator. Following the reconstruction of the crystal structure, energy

minimization based on Universal Force Field (UFF) parameters,[41] that are available from the Materials Studio Forcite module, was carried out while holding the experimentally reported lattice constants and cell angles constant (i.e., only positions of atoms were allowed to relax). We restricted ourselves to curate MOF crystals with building blocks that are available within the crystal generator. Using this approach, we generated 28 MOF crystal structures, which are listed in Table S1. We have excluded the disordered structures that we were not able to fix using this method and placed them in a folder named "disorder". The total number of structures that are marked as disordered is 331 for the FSR dataset and 455 for the ASR dataset (including the 331 disordered structures also in the FSR dataset).

Geometric and Topological Analyses. The open-source program Zeo++[42] was used to compute geometric properties of the structures in the CoRE MOF 2019 Database. The PLD, LCD, GSA, and VSA of MOF crystals were calculated using high-accuracy settings (-ha flag), and a hard sphere with a diameter of 3.31 Å (the Lennard-Jones σ parameter of nitrogen in the TraPPE model)[43] was used in computing the surface areas. The underlying topologies of the crystal structures in the CoRE MOF 2019-ASR Database were determined using the ToposPro software program[44-45] with known topologies from the RCSR and EPINET databases.[40, 46]

Linkers in MOFs. Information on the organic building blocks in a subset (~5,000) of the CoRE MOF 2019-ASR Database was identified through a substructure search using the Cambridge Crystallographic Data Centre (CCDC) software's ConQuest program.[47] We restricted the search to the structures with CSD REFCODEs (i.e., no WoS structures) and that were updated from CSD up to March 2016. Common linkers, such as terephthalic acid and trimesic acid, were drawn and identified in the database first. Each linker structure was drawn in the program, which searched through all CoRE MOFs for a given ligand structure to find a match. The hydroxyl hydrogen of

an alcohol or carboxylic acid group was not included in the search because the metal often binds to the oxygen of these functional group via replacement of a proton. Additionally, multiple search queries were performed for a ligand structure with the same atom connectivity to account for delocalized bonds and resonance structures. The search results for a given linker structure included the MOFs that contain the searched subunit structure as a linker as well as MOFs with the structure as a moiety of a larger linker. These were manually distinguished from one another. Many of the MOFs had two types of linkers, and the linker that was not searched for was identified and a substructure search for that linker was performed on all other MOF structures. Based on this search, the ligands from all CoRE MOF structures were grouped by organic linker and common linkers were identified. The list of organic linkers tabulated in this way is not comprehensive given that we only searched for linkers that we considered to be common.

Open Metal Site (OMS) Detection Algorithm. We developed an algorithm to detect the presence of (potential) OMSs in MOFs. The coordinates and unit cell vectors of a MOF structure were read from the CIF files of the CoRE MOF 2019 Database. All metal atoms in the structure were detected together with all atoms in their first coordination spheres, i.e., the framework atoms directly bonded to the metal. A framework atom was considered bonded to a metal atom if the distance between them was smaller than the sum of their covalent radii plus a tolerance that was taken to be 0.2 Å for metal atoms with covalent radius smaller than 1.95 Å and 0.4 Å for larger metal atoms. The covalent radii were taken from Cordero and co-workers,[33] who analyzed crystallographic data for all elements with an atomic number up to 96. The bond detection can, in certain cases, incorrectly detect bonds between atoms that are not truly bonded, and thus incorrectly include these atoms in the first coordination sphere. To eliminate such cases, we checked if any atoms in the coordination sphere are bonded to each other, and if such a case was found we removed the

atom further away from the metal. This rule was not applied to carbon atoms to correctly identify the coordination spheres of metallocene. Metal atoms of the same type were also excluded from this rule to ensure that environments belonging to metal acetate-type ("paddle wheel") building units were described correctly and not considered as tetra-coordinated metals.

Two approaches considering only the geometry of the atoms within the first coordination sphere were used to identify OMSs. The first, plane-based, criterion was tuned to give a binary decision on whether a metal site is open or not. The second metric was based on bending angles and provides a numerical value (τ-factor) that represents the distortion of the first coordination sphere away from a closed configuration and is only applicable to tetra-, penta- and hexa-coordinated metals. The expression for the τ-factors for the tetra-coordinated (τ_4)[48] and penta-coordinated (τ_5)[49] metals were taken from the literature, whereas we developed a new expression for the hexa-coordinated metals (τ_6). Details on both metrics are discussed in **Supporting Information Section S5**. Typically, there are multiple OMSs in a given MOF structure, but some of them could be symmetrically equivalent. For example, M-MOF-74 has only one unique OMS.[50] Knowing which sites are unique in each structure allows for a better representation of the variety of sites found in MOFs and enables a more efficient examination of these OMSs in the computational screening processes. For this reason, all of the unique metal sites in each MOF were identified using their coordination sequence.[51] The OMS code can be accessed from https://github.com/emmhald/open_metal_detector.

Molecular Fingerprint and Overlap between Hypothetical and CoRE MOFs. We developed a molecular fingerprint based on a MOF's elemental composition and geometric properties to determine the overlap between the hypothetical MOFs generated by Wilmer and coworkers[12] and the structures in the CoRE MOF 2019-ASR Database. Our approach begins with the material's

stoichiometry and incrementally considers other geometric properties of MOFs (such as the surface area, density, and pore limiting diameter). As a first step, the chemical formula of each MOF is compared by reducing the stoichiometric coefficients of the elements to their lowest possible integer values. Hydrogen atoms must also be included in the molecular formula to capture the subtle differences between bond order of ligands. For example, PCN-11 and PCN-16 differ by a single bond in their organic ligand. Stoichiometry coefficients of the elements are not enough as a complete fingerprint to identify identical MOF structures from two databases. For example, multiple MOF polymorphs can exist with the same chemical composition. One possible way to recognize the effect of atomic packing inside a crystal structure is to use the crystal density as a descriptor. Depending on the crystal density, the same molecular formula may represent a different structure – for example, interpenetrated MOFs will have the same stoichiometry of elements, but their densities will be different by a factor of two. Since the density is a continuous variable, we must find a threshold tolerance for the property. If the tolerance is too narrow, then the same MOF will be incorrectly classified as distinct, and if the tolerance is too large, then we will not be able to achieve enough resolution to differentiate MOFs. To establish a reasonable density threshold, five common MOFs (HKUST-1, IRMOF-1, MIL-47, MIL-53, and UiO-66) were grouped together with the molecular formula and density filters. Total counts of true duplicates were assigned based on the molecular formula filter and the results were manually checked. Details are provided in the **Supporting Information Section S6**.

Computational Screening. Grand canonical Monte Carlo (GCMC) simulations were carried out to investigate the effect of bound solvent removal on the adsorption properties of a 20:80 mixture of xenon and krypton at 298 K and 1 bar. Lennard-Jones (LJ) potentials were used to model the van der Waals interaction between atoms using a 12.8 Å cut-off without tail-corrections. LJ

parameters for xenon and krypton were taken from Rutkai and coworkers, where the values of the LJ well depth (ε/k_B) and the LJ diameter (σ) are 226.14 K for xenon and 162.58 K for krypton, and 3.949 Å for xenon and 3.627 Å for krypton, respectively.[52] MOF atoms were modeled using the LJ parameters from the UFF force field.[41] The Lorentz-Berthelot[53] mixing rules were used to determine the cross interaction parameters between unlike atoms. The Peng-Robinson equation of state[54] was used to convert the pressure to fugacity necessary to compute the chemical potential of the system. Simulations were carried out on the structures with a PLD value greater than the vdW diameter (as specified in the forcefield) of a xenon atom (3.97 Å). A total of 4,517 structures were examined. GCMC simulations were carried out using the RASPA software package.[55]

Stability Calculations. Calculations of the binding energies of bound solvent molecules, i.e., dimethylformamide (DMF) in FIJDOS[50] and diethyl formamide (DEF) in ja300034j_si_002[56], were performed using the CP2K code (version 6.0), which uses a mixed Gaussian/plane-wave basis set[57-58]. We employed double-ζ polarization quality Gaussian basis sets[59] and a 600 Ry plane-wave cutoff for the auxiliary grid, in conjunction with the Goedecker-Teter-Hutter (GTH) pseudopotentials[60-61]. A convergence threshold of 1.0×10^{-6} Hartree was used for the self-consistent field cycle; structural optimizations were considered as converged when the maximum force on atoms fell below 4.5×10^{-4} Hartree/Bohr. All calculations were performed in the Γ-point approximation for sufficiently large cells, i.e., the total number of atoms (including DMF or DEF solvent molecules) in the FIJDOS and ja300034j_si_002 cells are 252 and 384, respectively. All DFT calculations, including single-point energies and geometry/cell optimizations, were performed using the PBE functional[62], with Grimme's D3 van der Waals correction (PBE+D3)[63]. DMF, DEF, and water molecules were fully optimized in an empty box. The energy of bound solvent removal was computed using the following equation:

$$\Delta E_{\text{solvent removal}} = E_{\text{MOF+solvent}} - (E_{\text{MOF}} + E_{\text{solvent}})$$

Here, $E_{\text{MOF+solvent}}$ is the energy of MOF plus solvent, E_{MOF} is the energy of the optimized MOF without solvent, and E_{solvent} is the energy of the optimized solvent molecules in an empty box.

RESULTS

Structures in the CoRE MOF Database. Table 1 summarizes the different CoRE MOF databases. The CoRE MOF 2014-DDEC and CoRE MOF 2014-DFT-optimized datasets do not cover all the structures found in the CoRE MOF 2014-ASR dataset, because these derived datasets require periodic DFT calculations. The MOFs that are not included in these datasets are those for which DFT calculations were not able to converge in a reasonable amount of time.[29-30]

For the CoRE MOF 2019 Database, we prepared two separate datasets denoted as 2019-FSR and 2019-ASR. There are fewer structures in the CoRE MOF 2019-FSR dataset than in the 2019-ASR dataset since some structures are deemed non-porous (i.e., pore limiting diameter < 2.4 Å) if the bound solvent molecules are not removed. The CoRE MOF 2019-ASR Database is expanded by more than 9,000 additional MOF structures beyond those found in the CoRE MOF 2014-ASR Database. As noted above, all structures that were modified in a significant way from the CIF files in the CSD or were not obtained from the CSD have their atomic coordinates available from http://dx.doi.org/10.11578/1118280. Scripts for detection of OMSs, duplicate structures, and bound/unbound solvent removal are available from the same location.

It must be emphasized that the solvent removal procedure applied to generate the CoRE MOF 2019-ASR dataset (as also for the 2014-ASR dataset) is on purpose inclusive with the aim to yield a maximal dataset for the purpose of high-throughput computational screening. Computational removal of bound solvent molecules may go beyond what can be achieved during experimental activation procedures and may lead to structures that are not mechanically stable. Such unstable

structures would then undergo significant changes either to other crystalline forms potentially with a loss of porosity or convert to disordered structures. Thus, when computational screening of the 2019-ASR dataset yields promising candidate structures not included in the 2019-FSR dataset, then additional steps should be taken by the researcher to assess the stability of the ASR structure. Our prior work using DFT optimization on the 2014-ASR dataset indicated that a fraction of the structures undergo large changes in structure or the optimization did not even converge to a stable structure.[30] Given the computational expense of the DFT optimization, we favor carrying out these optimizations on a case-by-case basis.

The numbers of structures included in the CoRE MOF 2019-ASR and 2019-FSR datasets (14,142 and 9,869 structures, respectively) are about a factor of 5 smaller than the total number of structures included in the MOF dataset reported by Moghadam et al.[9] The primary reasons for this difference are that the CoRE MOF 2019 datasets only include structures that are a) three-dimensional and b) porous, where we define porous as having a PLD greater than 2.4 Å. Note that by applying a PLD filter of >3.7 Å, the CSD yields a subset with ~9,000 porous two- and three-dimensional MOFs.

Solvent Removal Effect on Textural Properties. Following the synthesis of MOFs, the solvent molecules used for the synthesis are generally trapped inside the pores, and in some cases, the residual solvent molecules form a coordination bond with exposed metal sites of the inorganic nodes. These coordinated solvent molecules ("bound" solvents) could be used to stabilize the framework or removed during the activation procedure to leave the metal sites exposed for selective adsorption and catalysis. Removal of such "bound" solvents changes the underlying physical properties, such as the surface area and pore size, of MOFs. **Figure 2** compares the geometric properties of CoRE MOF 2019-ASR and 2019-FSR datasets to analyze the effect of

solvent removal on some key geometric properties of synthesized MOFs. In comparing the effect of bound solvent removal on geometric properties, we did not include user contributed, manually fixed, and structures with ions. We also did not include structures with disorder. In total, we compared 8,815 structures that are in both the ASR and FSR datasets. **Figure 2a** and **Figure 2b** compare the LCD and PLD values from the ASR and FSR sets. We found only 1,426 (or 1,366) structures (~16 % of 8,815 structures) that showed more than a 0.5 Å increase or decrease in LCD (or PLD) value when the bound solvent was removed, with an average increase of ~1.5 Å for both LCD and PLD for the ~ 1,400 structures exceeding the 0.5 Å limit. One of the largest increases in the LCD value with bound solvent removal was observed in FIJDOS (Zn-MOF-74),[50] where the LCD value increased from ~5.0 Å to ~12.0 Å upon removal of dimethylformamide. Similarly, the largest increase in the PLD value was observed in ja300034j_si_002 (Mg_2(dobpdc)),[56] where the PLD value increased from ~10 Å to ~17 Å upon removal of dimethylformamide.

The removal of bound solvents can have a large impact on the accessible volumetric surface area (AVSA), for example, by exposing blocked pores (which will increase the overall surface area) or by flattening the protruding solvents (which may decrease the overall surface area). Figure 2c shows the comparison between the AVSA values from the ASR and FSR sets. In general, we find that the AVSA increases with the removal of bound solvents. 1,738 structures showed at least a 100 m^2/cm^3 increase in their AVSA. The surface area of the structures in this group increased, on average, by 677 m^2/cm^3. The largest increase in surface area was observed for MAGNEQ.[64] We also find that the accessible volumetric surface area decreases more than 100 m^2/cm^3 for 168 structures (~1.9% of 8,815) upon removal of bound solvent. This is because the loss of surface area from the solvent molecules is greater than the MOF surface area exposed upon removing the bound solvent. Of the 168 structures, the average decrease was 167 m^2/cm^3.

Impact of Bound Solvent Removal on Xe/Kr Separation Performance. The removal of bound solvents changes the underlying potential energy surface within the pore, which in turn changes the adsorption properties of the material. The two datasets that we have prepared provide opportunities to test the impact of solvent removal on adsorption properties for many structures with diverse physical properties. As an illustrative example, we have carried out high-throughput computational screening of the ASR and FSR datasets for a 20:80 mixture of xenon and krypton. **Figure 3** compares the Xe uptake, Kr uptake and Xe/Kr selectivity at 1 bar between ASR and FSR datasets. From the screening results, we find that the presence of bound solvent sometimes increases and sometimes decreases the Xe/Kr selectivity. For instance, the presence of coordinated water molecules in AVAGIP (red square) lowers the adsorption of xenon, and because of decreased adsorption of xenon molecules inside the pore, the Xe/Kr selectivity is lower for AVAGIP structure from the FSR dataset (~9) than from the ASR dataset (~21). If the coordinated water molecules are removed (mimicking "activated" condition in experiments), the selectivity of the material increases two-fold from 9 to 21. The increased selectivity is due to increased adsorption of xenon within AVAGIP following the removal of coordinated water molecules (0.44 versus 1.86 mmol/g for FSR and ASR, respectively). On the contrary, the presence of coordinated solvent sometimes increases the selectivity of a material. FEYQIK (blue inverted triangle) is an example where the presence of coordinated water molecules prevents the adsorption of krypton, which leads to higher Xe/Kr selectivity in FEYQIK structure from FSR dataset (~23) than in ASR dataset (~12).

To obtain molecular-level information on the changes of xenon and krypton adsorption, the potential energy surfaces (PES) for AVAGIP and FEYQIK are computed and compared in **Figure S6**. **Figures S6a** and **c** show the PES of ASR and FSR structures of AVAGIP for xenon and

krypton, respectively, which clearly demonstrates that the removal of water molecules creates more favorable adsorption sites for xenon and krypton within the pores. **Figure S6e** shows a comparison between the computed mixture isotherms for the ASR and FSR structures, which demonstrates that the uptakes of both xenon and krypton are increased over the complete range of pressures for the ASR structure; this can be attributed to the creation of more favorable adsorption sites as indicated in **Figure S6a** and **c**. While krypton uptake stays relatively constant before and after the removal of coordinated solvents (0.4 mmol/g for FSR and 0.2 mmol/g for ASR) at the adsorption pressure (1 bar), considerable changes are observed in xenon uptake from 0.44 mmol/g to 1.88 mmol/g at 1 bar. The increased xenon uptake leads to increased Xe/Kr selectivity from 9 to 21.

The removal of coordinated water molecules can also lead to a decrease in the selectivity of the material by altering the underlying PES to make it less favorable for adsorption of xenon. **Figures S6b** and **d** show the PES of FEYQIK FSR (left) and ASR (right) structures for xenon (**Figure S6b**) and krypton (**Figure S6d**). In this case, the presence of water molecules in FSR facilitates the adsorption of xenon, leading to a high Xe/Kr selectivity. Upon the removal of coordinated water molecules, the underlying PES changes and makes it less favorable for xenon adsorption. The mixture isotherms show that the uptake at 1 bar decreases from 2.04 mmol/g to 1.94 mmol/g for xenon and increases from 0.35 mmol/g to 0.59 mmol/g for krypton in FSR and ASR, respectively. The selectivity is decreased from 23 (FSR) to 13 (ASR) because of the increased adsorption of krypton.

Simon and co-workers previously screened the CoRE MOF 2014 dataset to find materials with high Xe/Kr selectivity.[13] The MOF with the highest selectivity found in their work was KAXQIL, also known as SBMOF-1, a calcium based MOF synthesized by Banerjee et al.[65] As shown in

Table 2 and Figure S6, our screening of the CoRE MOF 2019-ASR dataset found a number of materials with higher Xe/Kr selectivity and/or higher Xe uptake than KAXQIL. **Table 2** lists some selected MOFs with high selectivity and high Xe uptake at 1 bar, which are important metrics for screening materials for Xe/Kr separation applications. The results show that several materials with better performance have been synthesized since the computational screening of Simon et al. We find that the presence of solvent molecules sometimes decreases and sometimes increases materials properties due to changes in the PES of the pores and tuning the PES could potentially be achieved by leaving the coordinated solvents intact. We note that blocking of non-accessible pores was not performed for the current study; additional computational and experimental work is needed to validate the high Xe/Kr selectivity for these materials. Data for high-throughput simulation of both the ASR and FSR CoRE MOF datasets for Xe/Kr are available as part of the Supporting Information.

Stability of Selected MOFs Upon Solvent Removal. Removal of both bound and unbound solvents is necessary to utilize the empty space in MOFs for adsorption and catalysis applications, but the removal of solvents within the framework sometimes lead to framework collapse[66-68]. We carried out periodic DFT calculations on three MOFs (FIJDOS, ja300034j_si_002, and AVAGIP) to compute the energy cost to remove solvent molecules. FIJDOs and ja30034j_si_002 were chosen because these MOFs showed largest changes in their LCD values upon bound solvent removal. AVAGIP was chosen on the basis of large increase in Xe/Kr selectivity. The changes in the unit cell volume and physical properties were calculated. Because of the additional computational complexity associated with the Cu/Pr magnetic centers in DFT calculations, we did not attempt to estimate the binding energies of bound solvent molecules in MAGNEQ and FEYQIK. In addition, DFT calculations were not performed for KAXQIL, because it does not

contain bound solvent molecules. The results are summarized in **Table 3**. DFT calculations show that the energy cost of solvent removal is ~113 kJ/mol of dimethylformamide for FIJDOS (with ~0.3% volume decrease) and ~123 kJ/mol of diethyl formamide for ja300034j_si_002 (with 1.4% volume decrease). For the calculation of solvent removal in the FIJDOS, we found that a hydrogen atom was missing from dimethylformamide in the FIJDOS-FSR structure, and we have corrected the chemical structure of dimethylformamide before carrying out the DFT calculations. Based on the minimal volume changes upon solvent removal in these two materials, it is likely that these materials will not collapse upon solvent removal. For AVAGIP, we find the energy of removing the bound solvent (water) is ~45 kJ/mol, and the volume change before/after the bound solvent removal is ~2%. Visualization of DFT-optimized ASR and FSR structures are shown in **Figures S8 – S10**. Experimental validation is necessary to confirm if the calculated $\Delta E_{\text{solvent removal}}$ and the relative volume changes before/after the solvent removal are good descriptors of the framework stability.

Common Topologies. The underlying topologies of the CoRE MOF 2019-ASR dataset were determined using the ToposPro software program.[45] The CoRE MOF 2014 and CoRE MOF 2019 databases include more than 350 topologies. In the CoRE MOF 2019 database, more than 80 of these topologies are represented by 10 or more structures. **Figure 4** shows the 40 most common topologies that together represent more than 50% of the CoRE MOF 2019-ASR dataset. The five most common topologies are **pcu** (26.7 %), **dia** (12.8 %), **rtl** (4.1 %), **ths** (3.9 %), and **bcu** (3.7 %). The 10 most common nets found from the CoRE MOF 2019 Database include all five regular nets reported in the literature (**srs**, **nbo**, **dia**, **pcu**, and **bcu**).[69] From the topological perspective, these nets are the simplest possible nets with minimal transitivity and maximal symmetry.[70] Also, we find the **tbo** net to be quite common in the CoRE MOF databases. This may be due to the

popularity of synthesizing Cu-BTC (HKUST-1)[71] among MOF researchers, as more than 50 Cu-BTC structures were reported in the CoRE MOF 2019-ASR database. We also find that edge-transitive nets (or "default" nets) appear quite frequently. These nets are **tbo**, **pts, stp, sod, lvt, rht, acs, fcu, the, flu, pyr, qtz, sit, pto,** and **ctn**.

Common Linkers. Table 4 summarizes the statistics of some organic linkers contained in the MOFs in the CoRE MOF 2019-ASR dataset. One would expect organic ligands that can be used to construct **pcu**, **dia**, **bcu**, and **tbo** topologies to be popular since organic ligands and topologies are closely related to one another. These topologies require building blocks with two or three "binding points," so we searched for common bidentate or tridentate ligands, such as terephthalic acid, 4,4-bipyridine, and trimesic acid. According to the minimal transitivity principle, the popularity of **pcu**, **dia**, **bcu**, and **tbo** topologies in the CoRE MOF database could be directly linked with the popularity of the linkers with two or three binding points because these are the simplest and most highly symmetric nets that are more likely to be synthesized.

Open Metal Sites. We screened all structures within the CoRE MOF 2019-ASR and 2019-FSR datasets for OMSs. This was done by applying the plane-based criterion and the angle-based metrics (τ-factors) as described in the Methods section and Supporting Information. In total, we found that, out of the 13,687 MOFs in the 2019-ASR dataset, 8,470 (~61.9%) contained at least one OMS. Out of 9,538 MOFs in the 2019-FSR dataset, 3,274 (~34.3%) contained at least one OMS (see Figure S5 for details of number of MOFs). This is a factor of 2.6 smaller than for the 2019-ASR dataset. We would like to emphasize that even 34.2% is a very large fraction and indicates the potential of MOFs for site-specific separation, via dative bonding of guest molecules, and for catalysis. The 2019-ASR dataset contains 8,815 structures that overlap with the FSR dataset and satisfied our porosity constraint only after removal of bound solvent. For the sake of

fair comparison, these 8,815 structures do not include MOFs that have gone through different structure curation procedures, such as ion retention, manual structure modification, and topology-based curation. For this subgroup (8,815 structures), 2,986 (~34%) and 5,219 (~59%) structures were found to include at least one OMS for FSR and ASR datasets, respectively. In other words, 2,233 structures that did not have OMSs in the FSR dataset now have OMSs following the bound solvent removal.

The structural (and chemical) stability of many of the open metal sites in the ASR dataset is questionable since it is unknown if the required activation can take place without framework collapse. Nazarian and coworkers found numerous examples from the CoRE MOF 2014 Database that undergo very significant structural changes during DFT optimization for materials in which the original crystal structure included bound solvents.[30] The OMSs that do not require additional bound solvent molecules to be removed (beyond those removed in experiments before structure determination), found in both the 2019-ASR and 2019-FSR datasets, provide important information about potential stability of the material, since these can more likely be activated under conditions described by the reported experimental procedures. However, the computational study of a diverse range of OMSs, even if some are unstable, can still lead to the identification of motifs with desirable properties.

The percentages of unique metal sites found to have specific τ_4, τ_5 or τ_6 values are presented as histograms in **Figure 5** for the 2019-ASR and 2019-FSR datasets. The colors of the bars indicate whether the sites were identified as open (blue) or not (red) using the plane-based criterion. As discussed in the **Methods section** (and in more detail in the **Supporting Information Section S5**), metal sites with τ-factor values of one, in general, correspond to symmetries that fully surround

the metal site (no open metal site). Deviations from these closed symmetries result in smaller τ-factor values and the presence of OMSs.

It is evident from the figure that indeed almost all coordination spheres with a τ-factor value of zero are identified as open by the plane-based criterion and those with a τ-factor value of one as not open. Since there is a continuous deviation from a closed coordination sphere towards a more exposed environment and the two methods are independent, there are coordination spheres in-between that have the same τ-factor value but a different classification. In addition, the threshold for the plane-based criterion was chosen to be loose to avoid false negatives.

For example, tetra-coordinated metals with $\tau_4 = 0.0$ correspond to a square planar, fully open geometry. Increasing τ_4 values correspond to deviations of the coordination sphere as it approaches a tetrahedral symmetry ($\tau_4 = 1.0$), a closed configuration. This transition can be observed in **Figure 5**, with sites having $\tau_4 > 0.9$ being mostly identified as closed. The most frequent geometry of the penta-coordinated sites, in both datasets, is the square pyramidal (D_{4h} symmetry) which is fully open and has a value of $\tau_5 = 0.0$. Only a small number of sites are found to deviate substantially from this geometry. For the hexa-coordinated metal sites, we observed for the ASR dataset a small broad peak around $\tau_6 = 0.4$, which corresponds to the "see-saw" type coordination sphere (C_{2v} symmetry). Most sites were found to have τ_6 values greater than 0.7 and were mostly identified as closed.

Among the penta-coordinated metal sites, 85% are identified as open within the ASR dataset ($N = 13,687$) and 68% within the FSR dataset ($N = 9,538$). In contrast, only 13% (ASR) and 1.6% (FSR) of the hexa-coordinated metal sites are found to be open. The tetra-coordinated metals are more varied with open percentages of 61% (ASR) and 48% (FSR). This data is presented in **Tables S3**. The data for ASR and FSR overlaps ($N = 8,815$) is presented in **Table S4**. The classification

of the metal coordination spheres used each MOF structure as it is found in each dataset. This means that for the ASR dataset the coordination spheres have undergone bound solvent removal while in the FSR, they have not. This can result in cases where a metal site which, for example, is classified as penta-coordinated in the ASR set may have higher coordination number in the FSR set.

It should be noted that we found a few rare coordination spheres with τ-factors that did not follow the trend that values of zero correspond to open metal sites and values of one to closed sites. This is most clearly seen in **Figure 5** for the tetra-coordinated sites in the ASR set. There is a small percentage of open metal sites with a τ_4 value of one. In these cases, the geometry of the first coordination sphere falls outside the range of deviations that can be captured correctly by the τ-factor formulations. More details on this and examples of such geometries are presented in the **Supporting Information Scheme S4**.

Figure 6 presents an elemental breakdown of the number of unique metal sites found in the ASR and FSR datasets. Although all metal atoms with an atomic number as large as 96 were included in the search, only atoms up to the fifth row of the periodic table, having more than 50 unique sites, are shown in **Figure 6**. The complete information regarding the number of MOFs and sites found for each element are included in the **Supporting Information Tables S6 – S9**.

The most common metals identified in both the ASR and FSR datasets are Zn, Cu, Cd, and Co. Together they account for approximately 50% of the metal sites in both the ASR and FSR datasets. In the FSR dataset, these elements also make up most of the open metal sites found (69%). Na, Li, K and Ag are also to a large extent open in both sets, and this is because they often act as charge-balancing ions (i.e., similar to cation exchanged zeolites).

Estimated Number of Distinct CoRE MOFs. The CoRE MOF 2014 and 2019 databases contain MOFs for which multiple crystallographic structures are reported in the literature, such as IRMOF-1 and Cu-BTC. This observation was already made in the original paper.[10] Recently, Barthel et al.[72] applied a computational methodology to distinguish MOFs in 502 structures from the CoRE MOF 2014-DDEC dataset and found that 78 structures are duplicates, and they also identified 60 structures with problems in network topology and composition. In a similar manner, the molecular fingerprint method outlined in the Methods section was used here to estimate the number of distinct MOFs in the much larger CoRE MOF 2019-ASR dataset. **Table 5** summarizes the results by classifying the structures based on different levels of molecular fingerprints. As more layers are added to the fingerprint, MOFs are increasingly categorized as distinct, until the limit of distinguishing them by their CIF file name. Depending on the fingerprint parameters chosen, there are an estimated 11,000 – 12,000 distinct MOFs in the CoRE MOF 2019 dataset, which corresponds to ~79% of the structures.

Barthel and coworkers[70] deem "duplicate" structures with minor variations in atomic coordinates as redundant. We would like to point out that for high-throughput screening of rigid structures, "duplicates" should not be removed because they provide important information. Similar to the CoRE MOF 2014 and 2019 Databases, the IZA Database[73] of zeolite structures contains multiple variants for many of the topologies resulting from differences in the synthesis conditions and the composition of the tetrahedral atoms and their charge-balancing cations, as well as idealized all-silica structures for zeolites not yet synthesized in all-silica form. Screening is sometimes sensitive to rather small changes in structure (for example, significant differences in water/alcohol adsorption and in performance for sour gas sweetening have been found for different variants of the same zeolite framework topologies[74-75]). That is, when searching for optimal nanoporous

structures, a hit that includes all of the structural variants of the same topology provides much higher degree of certainty that this topology or MOF or zeolite would actually perform well, whereas disagreement between different variants is a good cautionary flag.

Overlap Between Hypothetical MOFs and CoRE MOFs. We also used the molecular fingerprint method outlined in the Methods section to compare the CoRE MOF 2019-ASR dataset and the hypothetical MOF database of Wilmer and co-workers.[12] The algorithm identified 144 families as potentially overlapping between the hypothetical MOF and CoRE MOF 2019-ASR datasets. By visually inspecting these candidates, we confirmed that 19 of these MOFs were true overlaps, and the remaining 125 were false positives (See discussion in the **Supporting Information Section S6**). The results illustrate the challenges of identifying MOFs based on composition and topology. We note that 3 of the 19 structures obtained from the literature are hypothetical MOFs that were reported in the Supporting Information of a paper. This leaves us with 16 synthesized hMOFs. SUMOF-3, HKUST-1 (Zn), HKUST-1 (Cu), IRMOF-1, IRMOF-3, IRMOF-9, IRMOF-14, SUMOF-4, BMOF-1-bpdc, interpenetrated IRMOF-1, MIL-47, $Zn_2(NDC)_2(DPNI)$, and VO(BPDC) are some experimentally reported MOFs that are also found in the hypothetical MOF database based on our molecular fingerprint approach. The summary of confirmed overlap between CoRE MOF 2019-ASR and Wilmer's hypothetical MOF database is provided in the **Supporting Information Table S10**.

CONCLUSIONS

We have reported an updated version of the Computation-Ready, Experimental (CoRE) MOF database with additional analyses and ~ 9,000 additional structures. The detailed analyses of open metal sites presented in this work show that a large fraction of the CoRE MOF structures (34% for

FSR and 62% for ASR) contain at least one open metal site, and researchers are already using this knowledge to screen the database for selective adsorption of small molecules and catalysis applications.[24] As an example application of the database, we screened it to find materials for Xe/Kr separation. We compared our results with previous work of Simon and coworkers, who had computationally screened the CoRE MOF 2014 Database for Xe/Kr separation and found SBMOF-1 (CSD REFCODE: KAXQIL) as the MOF with the highest Xe/Kr selectivity.[19] In our screening of the CoRE MOF 2019 databases, we discovered a number of structures with similar selectivity as that of SBMOF-1 but with higher capacity, which is an important metric in lowering the total operating cost of the adsorption unit for industrial application. The results also show that the Xe/Kr adsorption properties in MOFs are sensitive to the presence of coordinated solvent molecules and the presence (or absence) of coordinated solvents can sometimes increase the selectivity of materials. This paves a new avenue to tune the adsorption selectivity of materials for new separation applications. Finally, based on a molecular fingerprint approach, we compared the CoRE MOF 2019 Database with a database of hypothetical MOFs. We found that there are at least 16 synthesized MOFs present in the hypothetical MOF database of Wilmer et al. Although it is unlikely that any MOF database can ever be comprehensive, the results we have reported here represent the most extensive collection of simulation-ready three-dimensional porous MOF structures available to date. We anticipate that these structures will be of considerable value in computational modeling of MOFs for a diverse range of possible applications and in efforts to compare the structures of putatively new materials in future synthesis studies to the large number of materials that are already known.

Supporting Information. The following files are available free of charge.

Details of Crystal Reconstruction; Potential Energy Surfaces and Mixture Adsorption Isotherms for Xe/Kr from GCMC Simulations; Additional High-throughput Screening Data for CoRE MOF 2019-ASR Dataset for Xe/Kr; Optimized ASR and FSR Structurse; Open Metal Site Detection; Molecular Fingerprints (PDF)

CoRE MOF properties dataset (XLSX)

Overlapping Hypothetical and CoRE MOF Structures (ZIP)

Input Files for ToBaCCo for Structure Restoration (ZIP)

AUTHOR INFORMATION

Corresponding Author

*Email: drygcung@gmail.com

*E-mail:david.sholl@chbe.gatech.edu

*E-mail: snurr@northwestern.edu

*E-mail: siepmann@umn.edu

Author Contributions

Y.G.C., E.H., J.I.S., R.Q.S., and D.S.S. conceived and designed the work. Y.G.C. collected MOFs from CSD and WoS. S. Lee carried out GCMC simulations under the guidance of Y.G.C. E.H. and K.D.V. developed the open metal site detection code and duplicate structure detection code. B.J.B. and H.Z. reconstructed disordered MOFs using ToBaCCo and carried out topological analyses. M.H. carried out geometric property calculations. M.M. searched for the ligand structures in CoRE MOFs under the guidance of D.S.S. S.L. and B.S. contributed reconstructed MOFs based on the

CoRE MOF 2014 Database. Y.G.C., E.H., B.B., J.I.S and R.Q.S., wrote the manuscript based on the inputs from other co-authors.

‡These authors contributed equally.

ACKNOWLEDGMENT

REFERENCES

1.   Furukawa, H.; Cordova, K. E.; O'Keeffe, M.; Yaghi, O. M., The Chemistry and Applications of Metal-Organic Frameworks. *Science* **2013,** *341* (6149), 974-+.
2.   Slater, A. G.; Cooper, A. I., Function-led design of new porous materials. *Science* **2015,** *348* (6238).
3.   Morris, R. E.; Wheatley, P. S., Gas storage in nanoporous materials. *Angew Chem Int Edit* **2008,** *47* (27), 4966-4981.
4.   Li, J. R.; Kuppler, R. J.; Zhou, H. C., Selective gas adsorption and separation in metal-organic frameworks. *Chem Soc Rev* **2009,** *38* (5), 1477-1504.
5.   Li, J. R.; Sculley, J.; Zhou, H. C., Metal-Organic Frameworks for Separations. *Chem Rev* **2012,** *112* (2), 869-932.
6.   Lee, J.; Farha, O. K.; Roberts, J.; Scheidt, K. A.; Nguyen, S. T.; Hupp, J. T., Metal-organic framework materials as catalysts. *Chem Soc Rev* **2009,** *38* (5), 1450-1459.
7.   Horcajada, P.; Gref, R.; Baati, T.; Allan, P. K.; Maurin, G.; Couvreur, P.; Ferey, G.; Morris, R. E.; Serre, C., Metal-Organic Frameworks in Biomedicine. *Chem Rev* **2012,** *112* (2), 1232-1268.
8.   Kreno, L. E.; Leong, K.; Farha, O. K.; Allendorf, M.; Van Duyne, R. P.; Hupp, J. T., Metal-Organic Framework Materials as Chemical Sensors. *Chem Rev* **2012,** *112* (2), 1105-1125.
9.   Moghadam, P. Z.; Li, A.; Wiggin, S. B.; Tao, A.; Maloney, A. G. P.; Wood, P. A.; Ward, S. C.; Fairen-Jimenez, D., Development of a Cambridge Structural Database Subset: A Collection of Metal-Organic Frameworks for Past, Present, and Future. *Chem Mater* **2017,** *29* (7), 2618-2625.
10.   Chung, Y. G.; Camp, J.; Haranczyk, M.; Sikora, B. J.; Bury, W.; Krungleviciute, V.; Yildirim, T.; Farha, O. K.; Sholl, D. S.; Snurr, R. Q., Computation-Ready, Experimental Metal-Organic Frameworks: A Tool To Enable High-Throughput Screening of Nanoporous Crystals. *Chemistry of Materials* **2014,** *26* (21), 6185-6192.
11.   Groom, C. R.; Bruno, I. J.; Lightfoot, M. P.; Ward, S. C., The Cambridge Structural Database. *Acta Crystallogr B* **2016,** *72*, 171-179.
12.   Wilmer, C. E.; Leaf, M.; Lee, C. Y.; Farha, O. K.; Hauser, B. G.; Hupp, J. T.; Snurr, R. Q., Large-scale screening of hypothetical metal-organic frameworks. *Nat Chem* **2012,** *4* (2), 83-89.
13.   Simon, C. M.; Kim, J.; Gomez-Gualdron, D. A.; Camp, J. S.; Chung, Y. G.; Martin, R. L.; Mercado, R.; Deem, M. W.; Gunter, D.; Haranczyk, M.; Sholl, D. S.; Snurr, R. Q.; Smit, B., The materials genome in action: identifying the performance limits for methane storage. *Energ Environ Sci* **2015,** *8* (4), 1190-1199.
14.   Zhang, H. D.; Deria, P.; Farha, O. K.; Hupp, J. T.; Snurr, R. Q., A thermodynamic tank model for studying the effect of higher hydrocarbons on natural gas storage in metal-organic frameworks. *Energ Environ Sci* **2015,** *8* (5), 1501-1510.
15.   Basdogan, Y.; Sezginel, K. B.; Keskin, S., Identifying Highly Selective Metal Organic Frameworks for CH4/H-2 Separations Using Computational Tools. *Ind Eng Chem Res* **2015,** *54* (34), 8479-8491.
16.   Li, S.; Chung, Y. G.; Snurr, R. Q., High-Throughput Screening of Metal-Organic Frameworks for CO2 Capture in the Presence of Water. *Langmuir* **2016,** *32* (40), 10368-10376.

17.   Qiao, Z. W.; Zhang, K.; Jiang, J. W., In silico screening of 4764 computation-ready, experimental metal-organic frameworks for CO2 separation. *J Mater Chem A* **2016,** *4* (6), 2105-2114.

18.   Chung, Y. G.; Gomez-Gualdron, D. A.; Li, P.; Leperi, K. T.; Deria, P.; Zhang, H. D.; Vermeulen, N. A.; Stoddart, J. F.; You, F. Q.; Hupp, J. T.; Farha, O. K.; Snurr, R. Q., In silico discovery of metal-organic frameworks for precombustion CO2 capture using a genetic algorithm. *Science Advances* **2016,** *2* (10).

19.   Simon, C. M.; Mercado, R.; Schnell, S. K.; Smit, B.; Haranczyk, M., What Are the Best Materials To Separate a Xenon/Krypton Mixture? *Chem Mater* **2015,** *27* (12), 4459-4475.

20.   Gee, J. A.; Zhang, K.; Bhattacharyya, S.; Bentley, J.; Rungta, M.; Abichandani, J. S.; Sholl, D. S.; Nair, S., Computational Identification and Experimental Evaluation of Metal-Organic Frameworks for Xylene Enrichment. *J Phys Chem C* **2016,** *120* (22), 12075-12082.

21.   Altintas, C.; Keskin, S., Computational screening of MOFs for C2H6/C2H4 and C2H6/CH4 separations. *Chem Eng Sci* **2016,** *139*, 49-60.

22.   Kulkarni, A. R.; Sholl, D. S., Screening of Copper Open Metal Site MOFs for Olefin/Paraffin Separations Using DFT-Derived Force Fields. *J Phys Chem C* **2016,** *120* (40), 23044-23054.

23.   Nie, X. W.; Kulkarni, A.; Sholl, D. S., Computational Prediction of Metal Organic Frameworks Suitable for Molecular Infiltration as a Route to Development of Conductive Materials. *J Phys Chem Lett* **2015,** *6* (9), 1586-1591.

24.   Vogiatzis, K. D.; Haldoupis, E.; Xiao, D. J.; Long, J. R.; Siepmann, J. I.; Gagliardi, L., Accelerated Computational Analysis of Metal–Organic Frameworks for Oxidation Catalysis. *The Journal of Physical Chemistry C* **2016,** *120* (33), 18707-18712.

25.   Chung, Y. G.; Bai, P.; Haranczyk, M.; Leperi, K. T.; Li, P.; Zhang, H.; Wang, T. C.; Duerinck, T.; You, F.; Hupp, J. T.; Farha, O. K.; Siepmann, J. I.; Snurr, R. Q., Computational Screening of Nanoporous Materials for Hexane and Heptane Isomer Separation. *Chem Mater* **2017**.

26.   Jeong, W.; Lim, D.-W.; Kim, S.; Harale, A.; Yoon, M.; Suh, M. P.; Kim, J., Modeling adsorption properties of structurally deformed metal–organic frameworks using structure–property map. *Proceedings of the National Academy of Sciences* **2017,** *114* (30), 7923-7928.

27.   Chong, S.; Thiele, G.; Kim, J., Excavating hidden adsorption sites in metal-organic frameworks using rational defect engineering. *Nat Commun* **2017,** *8*.

28.   Yang, L.; McNichols, B. W.; Davidson, M.; Schweitzer, B.; Gomez-Gualdron, D. A.; Trewyn, B. G.; Sellinger, A.; Carreon, M. A., Noble metal-free catalytic decarboxylation of oleic acid to n-heptadecane on nickel-based metal-organic frameworks (MOFs). *Catal Sci Technol* **2017,** *7* (14), 3027-3035.

29.   Nazarian, D.; Camp, J. S.; Sholl, D. S., A Comprehensive Set of High-Quality Point Charges for Simulations of Metal-Organic Frameworks. *Chem Mater* **2016,** *28* (3), 785-793.

30.   Nazarian, D.; Camp, J. S.; Chung, Y. G.; Snurr, R. Q.; Sholl, D. S., Large-Scale Refinement of Metal-Organic Framework Structures Using Density Functional Theory. *Chem Mater* **2017,** *29* (6), 2521-2528.

31.   Altintas, C.; Avci, G.; Daglar, H.; Azar, A. N. V.; Erucar, I.; Velioglu, S.; Keskin, S., An extensive comparative analysis of two MOF databases: high-throughput screening of computation-ready MOFs for CH4 and H-2 adsorption. *J Mater Chem A* **2019,** *7* (16), 9593-9608.

32.   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. *J Am Chem Soc* **2010**, *132* (21), 7528-7539.

33.   Cordero, B.; Gomez, V.; Platero-Prats, A. E.; Reves, M.; Echeverria, J.; Cremades, E.; Barragan, F.; Alvarez, S., Covalent radii revisited. *Dalton T* **2008**, (21), 2832-2838.

34.   PorousMaterials.jl.

35.   Allen, F. H., The Cambridge Structural Database: a quarter of a million crystal structures and rising. *Acta Crystallographica Section B-Structural Science* **2002**, *58*, 380-388.

36.   Taylor, R.; Wood, P. A., A Million Crystal Structures: The Whole Is Greater than the Sum of Its Parts. *Chem Rev* **2019**, *119* (16), 9427-9477.

37.   Ong, S. P.; Richards, W. D.; Jain, A.; Hautier, G.; Kocher, M.; Cholia, S.; Gunter, D.; Chevrier, V. L.; Persson, K. A.; Ceder, G., Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis. *Comp Mater Sci* **2013**, *68*, 314-319.

38.   Haldoupis, E. A Python program to analyze collections of Metal Organic Frameworks (MOFs) for open metal sites. https://github.com/emmhald/open_metal_detector.

39.   Colon, Y. J.; Gomez-Gualdron, D. A.; Snurr, R. Q., Topologically Guided, Automated Construction of Metal-Organic Frameworks and Their Evaluation for Energy-Related Applications. *Cryst Growth Des* **2017**, *17* (11), 5801-5810.

40.   O'Keeffe, M.; Peskov, M. A.; Ramsden, S. J.; Yaghi, O. M., The Reticular Chemistry Structure Resource (RCSR) Database of, and Symbols for, Crystal Nets. *Accounts Chem Res* **2008**, *41* (12), 1782-1789.

41.   Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M., UFF, a Full Periodic-Table Force-Field for Molecular Mechanics and Molecular-Dynamics Simulations. *J Am Chem Soc* **1992**, *114* (25), 10024-10035.

42.   Willems, T. F.; Rycroft, C.; Kazi, M.; Meza, J. C.; Haranczyk, M., Algorithms and tools for high-throughput geometry-based analysis of crystalline porous materials. *Micropor Mesopor Mat* **2012**, *149* (1), 134-141.

43.   Bae, Y. S.; Yazaydin, A. O.; Snurr, R. Q., Evaluation of the BET Method for Determining Surface Areas of MOFs and Zeolites that Contain Ultra-Micropores. *Langmuir* **2010**, *26* (8), 5475-5483.

44.   Alexandrov, E. V.; Blatov, V. A.; Kochetkov, A. V.; Proserpio, D. M., Underlying nets in three-periodic coordination polymers: topology, taxonomy and prediction from a computer-aided analysis of the Cambridge Structural Database. *Crystengcomm* **2011**, *13* (12), 3947-3958.

45.   Blatov, V. A.; Shevchenko, A. P.; Proserpio, D. M., Applied Topological Analysis of Crystal Structures with the Program Package ToposPro. *Cryst Growth Des* **2014**, *14* (7), 3576-3586.

46.   Ramsden, S. J.; Robins, V.; Hyde, S. T., Three-dimensional Euclidean nets from two-dimensional hyperbolic tilings: kaleidoscopic examples. *Acta Crystallogr A* **2009**, *65*, 81-108.

47.   Bruno, I. J.; Cole, J. C.; Edgington, P. R.; Kessler, M.; Macrae, C. F.; McCabe, P.; Pearson, J.; Taylor, R., New software for searching the Cambridge Structural Database and visualizing crystal structures. *Acta Crystallographica Section B-Structural Science* **2002**, *58*, 389-397.

48.   Yang, L.; Powell, D. R.; Houser, R. P., Structural variation in copper(I) complexes with pyridylmethylamide ligands: structural analysis with a new four-coordinate geometry index, tau(4). *Dalton Transactions* **2007**, (9), 955-964.

49.   Addison, A. W.; Rao, T. N.; Reedijk, J.; Vanrijn, J.; Verschoor, G. C., Synthesis, Structure, and Spectroscopic Properties of Copper(Ii) Compounds Containing Nitrogen Sulfur Donor Ligands - the Crystal and Molecular-Structure of Aqua[1,7-Bis(N-Methylbenzimidazol-2'-Yl)-2,6-Dithiaheptane]Copper(Ii) Perchlorate. *J Chem Soc Dalton* **1984,** (7), 1349-1356.

50.   Rosi, N. L.; Kim, J.; Eddaoudi, M.; Chen, B. L.; O'Keeffe, M.; Yaghi, O. M., Rod packings and metal-organic frameworks constructed from rod-shaped secondary building units. *J Am Chem Soc* **2005,** *127* (5), 1504-1518.

51.   Meier, W. M.; Moeck, H. J., The topology of three-dimensional 4-connected nets: Classification of zeolite framework types using coordination sequences. *Journal of Solid State Chemistry* **1979,** *27* (3), 349-355.

52.   Rutkai, G.; Thol, M.; Span, R.; Vrabec, J., How well does the Lennard-Jones potential represent the thermodynamic properties of noble gases? *Mol Phys* **2017,** *115* (9-12), 1104-1121.

53.   Lorentz, H. A., Ueber die Anwendung des Satzes vom Virial in der kinetischen Theorie der Gase. *Annalen der Physik* **1881,** *248* (1), 127-136.

54.   Peng, D.-Y.; Robinson, D. B., A New Two-Constant Equation of State. *Industrial & Engineering Chemistry Fundamentals* **1976,** *15* (1), 59-64.

55.   Dubbeldam, D.; Calero, S.; Ellis, D. E.; Snurr, R. Q., RASPA: molecular simulation software for adsorption and diffusion in flexible nanoporous materials. *Molecular Simulation* **2015,** *42* (2), 81-101.

56.   McDonald, T. M.; Lee, W. R.; Mason, J. A.; Wiers, B. M.; Hong, C. S.; Long, J. R., Capture of Carbon Dioxide from Air and Flue Gas in the Alkylamine-Appended Metal-Organic Framework mmen-Mg-2(dobpdc). *J Am Chem Soc* **2012,** *134* (16), 7056-7065.

57.   VandeVondele, J.; Krack, M.; Mohamed, F.; Parrinello, M.; Chassaing, T.; Hutter, J., QUICKSTEP: Fast and accurate density functional calculations using a mixed Gaussian and plane waves approach. *Comput Phys Commun* **2005,** *167* (2), 103-128.

58.   Hutter, J.; Iannuzzi, M.; Schiffmann, F.; VandeVondele, J., CP2K: atomistic simulations of condensed matter systems. *Wires Comput Mol Sci* **2014,** *4* (1), 15-25.

59.   VandeVondele, J.; Hutter, J., Gaussian basis sets for accurate calculations on molecular systems in gas and condensed phases. *Journal of Chemical Physics* **2007,** *127* (11).

60.   Goedecker, S.; Teter, M.; Hutter, J., Separable dual-space Gaussian pseudopotentials. *Phys Rev B* **1996,** *54* (3), 1703-1710.

61.   Krack, M., Pseudopotentials for H to Kr optimized for gradient-corrected exchange-correlation functionals. *Theor Chem Acc* **2005,** *114* (1-3), 145-152.

62.   Perdew, J. P.; Burke, K.; Ernzerhof, M., Generalized gradient approximation made simple. *Phys Rev Lett* **1996,** *77* (18), 3865-3868.

63.   Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H., A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *Journal of Chemical Physics* **2010,** *132* (15).

64.   Shang, R.; Chen, S.; Wang, B. W.; Wang, Z. M.; Gao, S., Temperature-Induced Irreversible Phase Transition From Perovskite to Diamond But Pressure-Driven Back-Transition in an Ammonium Copper Formate. *Angew Chem Int Edit* **2016,** *55* (6), 2097-2100.

65.   Banerjee, D.; Simon, C. M.; Plonka, A. M.; Motkuri, R. K.; Liu, J.; Chen, X. Y.; Smit, B.; Parise, J. B.; Haranczyk, M.; Thallapally, P. K., Metal-organic framework with optimally selective xenon adsorption and separation. *Nat Commun* **2016,** *7*.

66.   Dodson, R. A.; Wong-Foy, A. G.; Matzger, A. J., The Metal-Organic Framework Collapse Continuum: Insights from Two-Dimensional Powder X-ray Diffraction. *Chem Mater* **2018,** *30* (18), 6559-6565.

67.   Farha, O. K.; Hupp, J. T., Rational Design, Synthesis, Purification, and Activation of Metal-Organic Framework Materials. *Accounts Chem Res* **2010,** *43* (8), 1166-1175.

68.   Mondloch, J. E.; Karagiaridi, O.; Farha, O. K.; Hupp, J. T., Activation of metal-organic framework materials. *Crystengcomm* **2013,** *15* (45), 9258-9264.

69.   Friedrichs, O. D.; O'Keeffe, M.; Yaghi, O. M., Three-periodic nets and tilings: regular and quasiregular nets. *Acta Crystallographica Section A* **2003,** *59*, 22-27.

70.   O'Keeffe, M.; Yaghi, O. M., Deconstructing the Crystal Structures of Metal-Organic Frameworks and Related Materials into Their Underlying Nets. *Chem Rev* **2012,** *112* (2), 675-702.

71.   Chui, S. S. Y.; Lo, S. M. F.; Charmant, J. P. H.; Orpen, A. G.; Williams, I. D., A chemically functionalizable nanoporous material [Cu-3(TMA)(2)(H2O)(3)](n). *Science* **1999,** *283* (5405), 1148-1150.

72.   Barthel, S.; Alexandrov, E. V.; Proserpio, D. M.; Smit, B., Distinguishing Metal-Organic Frameworks. *Cryst Growth Des* **2018,** *18* (3), 1738-1747.

73.   Baerlocher, C. M., L. B Database of Zeolite Structures.

74.   DeJaco, R. F.; Elyassi, B.; Dorneles de Mello, M.; Mittal, N.; Tsapatsis, M.; Siepmann, J. I., Understanding the unique sorption of alkane-alpha, omega-diols in silicalite-1. *J Chem Phys* **2018,** *149* (7), 072331.

75.   Shah, M. S.; Tsapatsis, M.; Siepmann, J. I., Identifying Optimal Zeolitic Sorbents for Sweetening of Highly Sour Natural Gas. *Angew Chem Int Edit* **2016,** *55* (20), 5938-+.

Figures
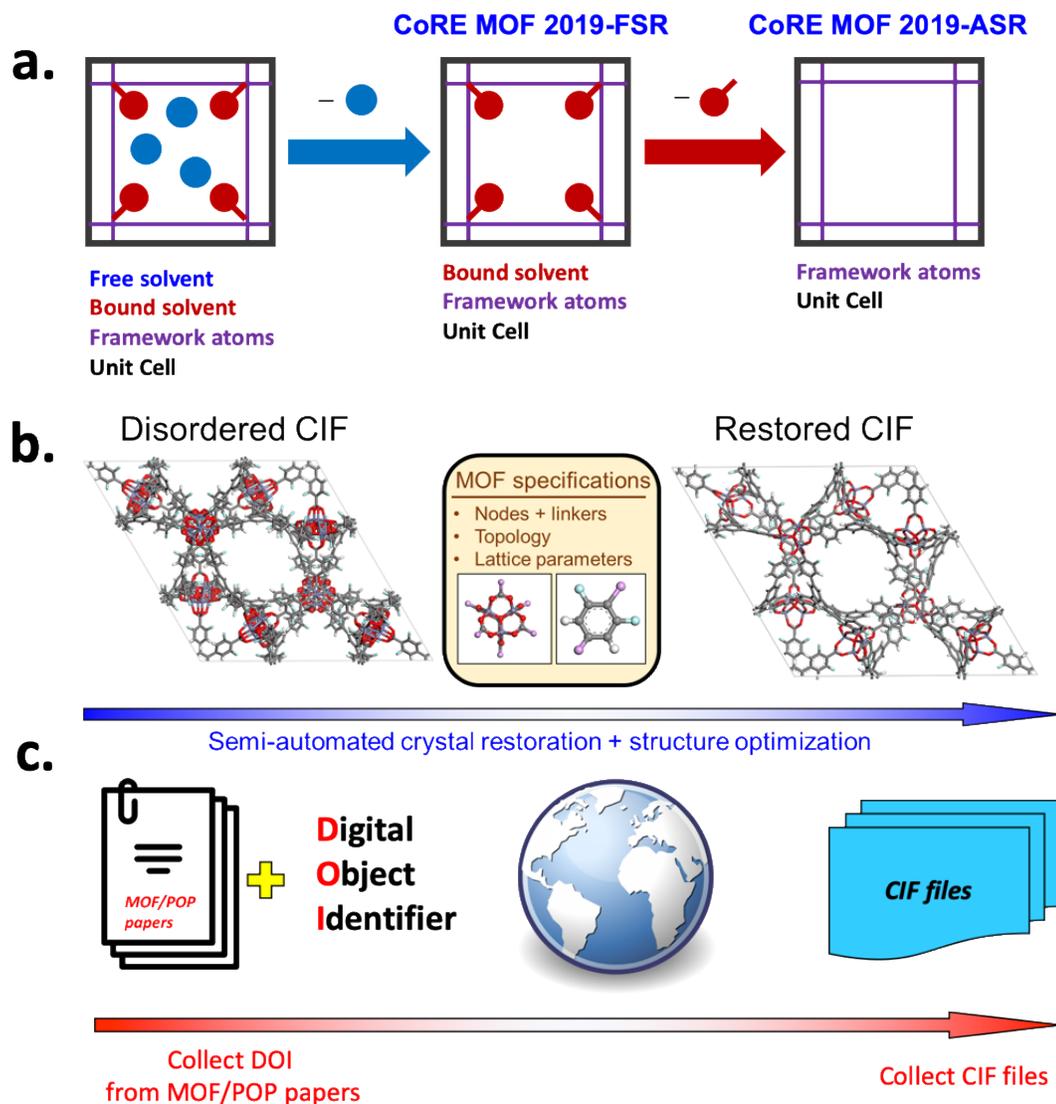
Figure 1. Preparation methods for structures in the CoRE MOF 2019 database: **a.** Solvent removal is performed in two stages: first removing "free" solvents (CoRE MOF 2019-FSR), then also removing bound solvent molecules (CoRE MOF 2019-ASR); **b.** Semi-automated restoration of 28 disordered MOFs using a crystal generator[39] and structure optimizations using the Materials Studio Forcite Module. The cartoon shows the restoration procedure of MOF-177 with the **qom** net (ja512311a_si_003_auto); **c.** Schematic of MOF structure collection from the literature.
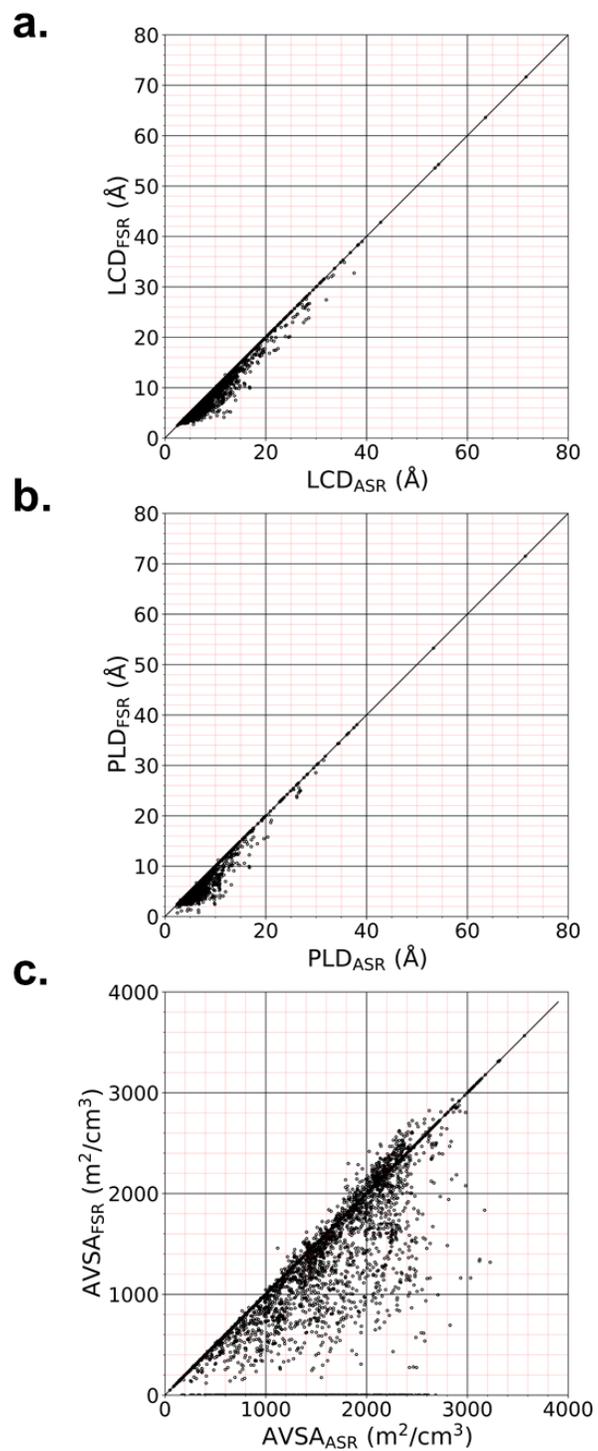
Figure 2. Comparison between the geometric properties of CoRE MOF 2019-ASR and 2019-FSR datasets (N = 8,815): **a.** Largest Cavity Diameter (LCD), **b.** Pore Limiting Diameter (PLD), **c.** Accessible Volumetric Surface Area (AVSA).
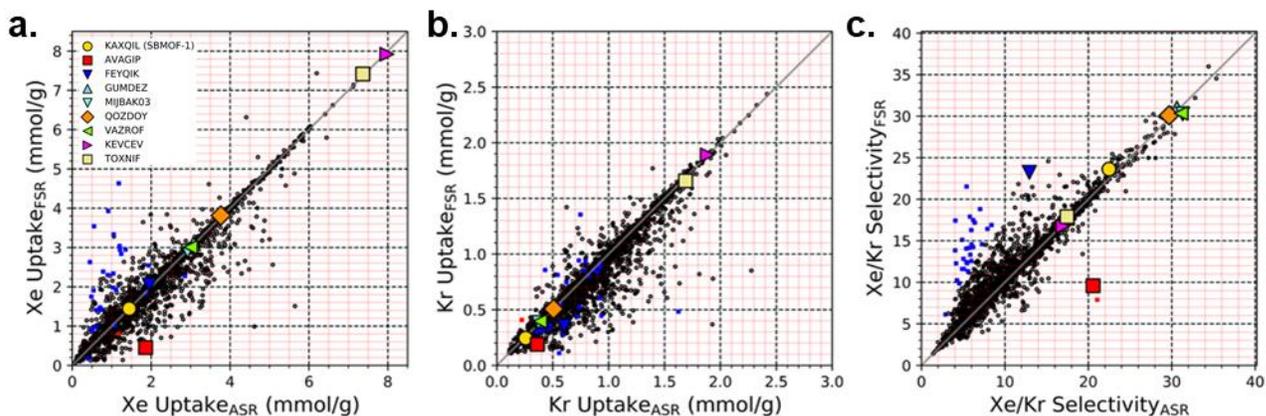
Figure 3. Comparison of 20:80 xenon:krypton mixture adsorption properties at $T = 298$ K and $P = 1$ bar for ASR and FSR structures ($N = 4{,}517$ for the subset of CoRE MOF 2019 ASR and FSR overlap with PLD > 3.97 Å, which is the vdW diameter of a xenon atom). **a.** Xenon uptake; **b.** Krypton uptake; and **c.** Xe/Kr Selectivity. Blue data points: Xe/Kr Selectivity$_{FSR}$ / Xe/Kr Selectivity$_{ASR}$ > 2; Red data points: Xe/Kr Selectivity$_{ASR}$ / Xe/Kr Selectivity$_{FSR}$ > 2.
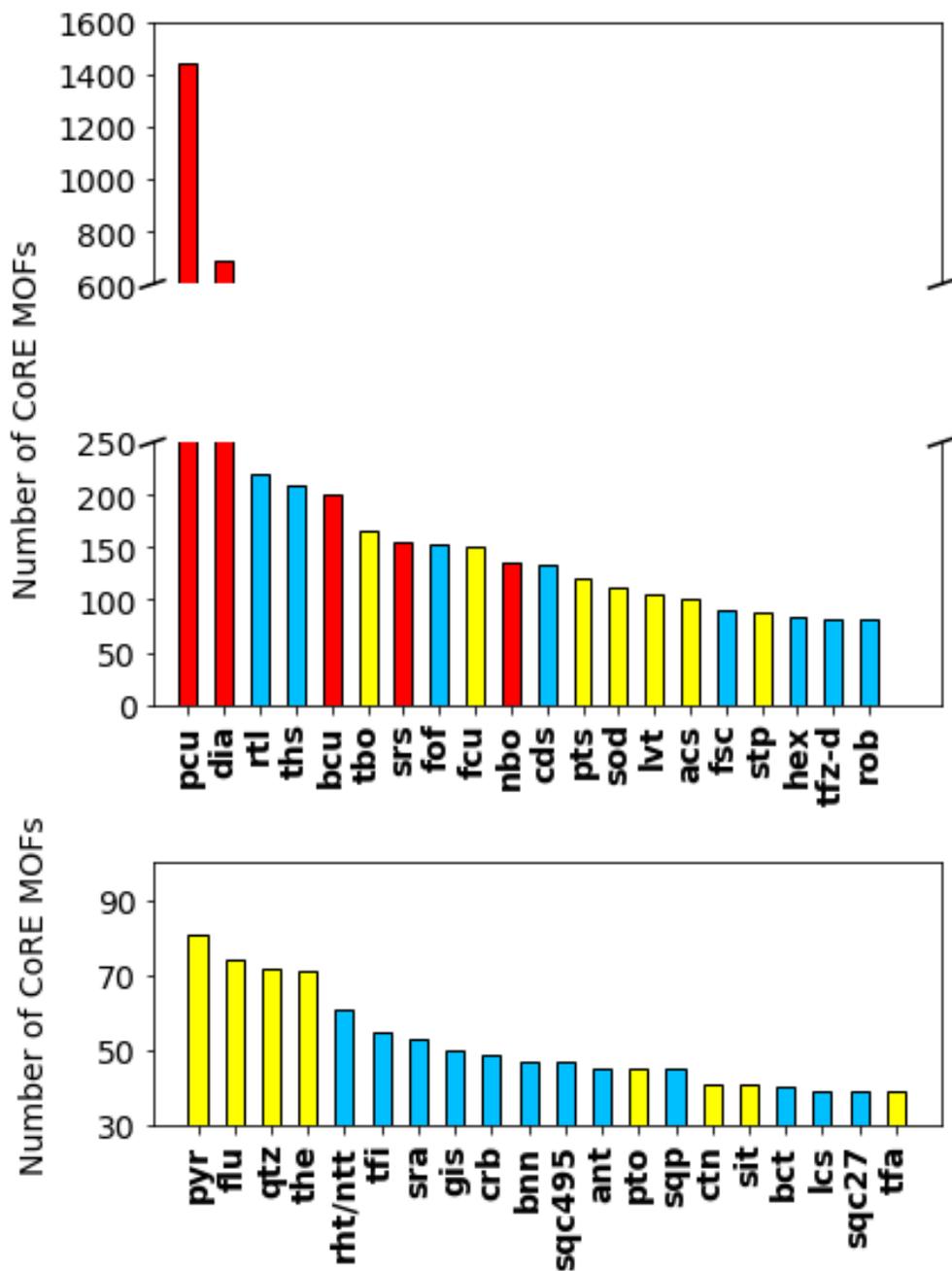
Figure 4. Histogram of 40 most common topologies in the CoRE MOF 2019-ASR database. Regular nets with transitivity *1111* (**pcu**, **dia**, **bcu**, **nbo**, **srs**) are highlighted in red; edge-transitive nets with transitivity *p1rs* (where *p* can be 1 or 2) are highlighted in yellow.
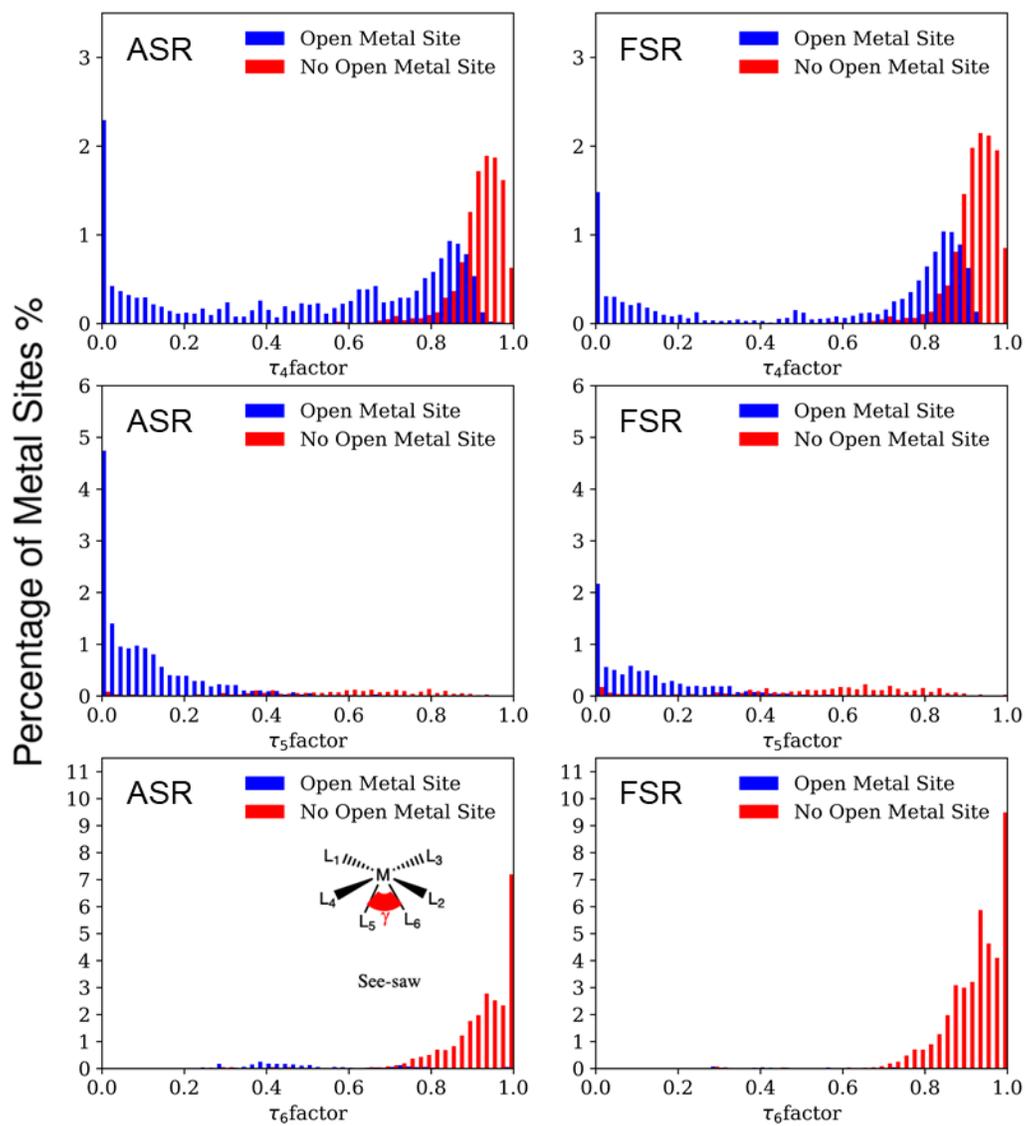
Figure 5. Percentage of unique metal sites with τ_4, τ_5, and τ_6 factors for CoRE MOF 2019-ASR (left) and 2019-FSR (right) datasets.
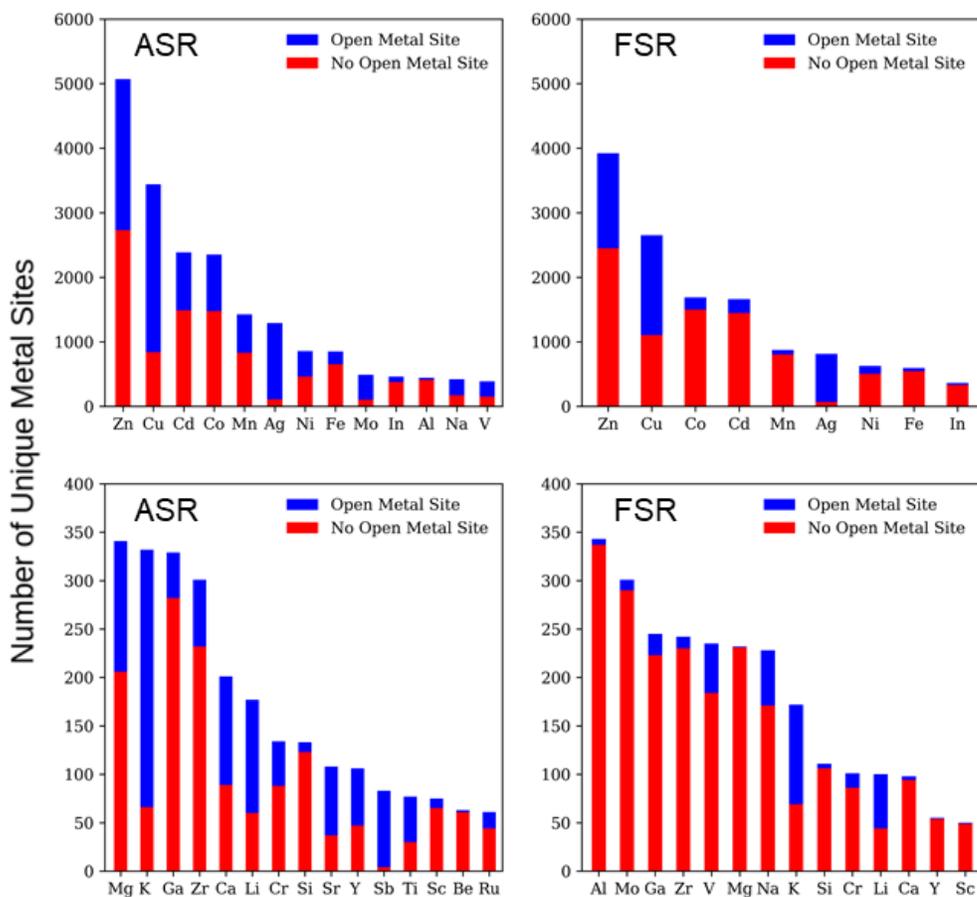
Figure 6. Breakdown by element of unique open and fully coordinated metal sites in the CoRE MOF 2019-ASR (right) and 2019-FSR (left) datasets. Only elements up to the fifth row of the periodic table having more than 50 unique sites are shown.

Tables

Table 1. Summary of the Computation-Ready, Experimental Metal-Organic Framework Databases[1]

Datasets	Cambridge Structural Database	Web of Science	Number of MOFs	Reference
CoRE MOF 2014-ASR	5,109	0	5,109 (4,764)	10
CoRE MOF 2014-DDEC	2,932	0	2,932 (2,932)	29
CoRE MOF 2014-DFT-optimized	832	0	832 (832)	30
CoRE MOF 2019-FSR	9,317	552	9,869 (7,061)	This work
CoRE MOF 2019-ASR	13,596	546[2]	14,142 (12,020)	This work

[1]The CoRE MOF 2014-ASR dataset is reported by Chung et al.,[10] the CoRE MOF 2014-DDEC dataset is reported by Nazarian et al.,[29] and the CoRE MOF 2014-DFT-optimized dataset is reported by Nazarian et al.[30] All structures in the CoRE MOF 2019-FSR and CoRE MOF 2019-ASR datasets have a pore limiting diameter (PLD) greater than 2.4 Å, which corresponds to approximately the van der Waals diameter of a hydrogen molecule. The numbers in parenthesis provide the number of MOFs for which atomic coordinates are provided in the CoRE MOF databases; the remaining structures are defined via a CSD REFCODE. [2]The number of Web of Science structures decreases from 552 to 546 because some of the MOFs are no longer 3D following the bound solvent removal.

Table 2. Summary of Xe/Kr Adsorption Properties of Selected MOFs from the CoRE MOF 2019-ASR Dataset[1]

MOF ID	Xe Uptake at 1 bar (mmol/g)	Kr Uptake at 1 bar (mmol/g)	Selectivity at 1 bar	LCD (Å)	PLD (Å)	GSA (m2/g)	VSA (m2/cm3)	Pore Volume (cm³/g)
KAXQIL[2]	1.46 (1.45)	0.25 (0.25)	23 (24)	4.1 (4.1)	4.5 (4.5)	254 (276)	397 (431)	0.22 (0.22)
KEVCEV	7.98 (7.92)	1.88 (1.89)	17 (17)	5.7 (5.7)	4.3 (4.3)	2521 (2521)	2230 (2 230)	0.60 (0.60)
TOXNIF	7.37 (7.42)	1.69 (1.66)	17 (18)	7.0 (7.0)	6.8 (6.8)	1458 (1458)	1556 (1556)	0.58 (0.58)
QOZDOY	3.80 (3.80)	0.50 (0.50)	30 (30)	5.2 (5.2)	4.7 (4.7)	937 (937)	1090 (1090)	0.45 (0.45)
GUMDEZ	3.09 (3.09)	0.41 (0.41)	31 (31)	5.3 (5.3)	4.8 (4.8)	885 (885)	1198 (1198)	0.38 (0.38)
VAZROF	3.00 (3.00)	0.39 (0.40)	31 (31)	5.0 (5.0)	4.6 (4.6)	663 (663)	820 (820)	0.37 (0.37)
MIJBAK03	2.91 (2.91)	0.38 (0.38)	31 (31)	6.1 (6.1)	5.5 (5.5)	658 (658)	953 (953)	0.29 (0.29)
FEYQIK	1.94 (2.04)	0.59 (0.35)	13 (23)	5.7 (5.7)	4.2 (4.2)	1102 (858)	1433 (1229)	0.44 (0.36)
AVAGIP	1.86 (0.44)	0.36 (0.18)	21 (9)	4.8 (4.8)	4.7 (4.6)	571 (264)	803 (387)	0.33 (0.29)

[1]Values in parentheses are for FSR structures; adsorption properties are reported at 298 K and 1 bar for a gas-phase composition of 20:80 Xe:Kr [2]SBMOF-1 (CoRE-MOF ID: ncomm11831-s3)

Table 3. Summary of the DFT Energies and the Changes in the Cell Volumes Upon Bound Solvent Removal for Some MOFs.

MOF ID	Volume from CSD (Ang^3)	Volume After Cell Optimization (Ang^3)	Energy Cost to Remove Solvent Molecules (kJ/mol)	Relative Volume Change (%)
FIJDOS_clean	2654.31	2736.44		
			113.7	0.3
FIJDOS_freeONLY	2654.31	2729.39		
ja300034j_clean	5601.36	5738.97		
			123.6	−1.4
ja300034j_freeONLY	5601.36	5818.31		
AVAGIP_clean	15443.51	15192.3		
			44.6	2.4
AVAGIP_freeONLY	15443.51	14836.94		

Table 4. Some Organic Linkers Found in a Subset of the CoRE MOF 2019-ASR dataset.

Common Name	Molecular Structure	# of entries
Terephthalic acid		344
4,4-Bipyridine		265
Trimesic acid		209
Isonicotinic acid		179
2,6-Naphthalenedicaboxylic acid (H2NDC)		77
1,3,5-Tris(4-carboxyphenyl)benzene (H3BTB)		54
4,5-Imidazoledicarboxylic acid		45
Oxalic acid		45
Phthalic acid		30
Biphenyl-3,4',5-tricaboxylic acid		22
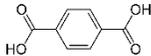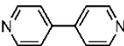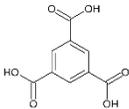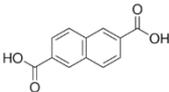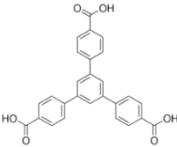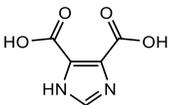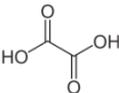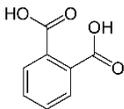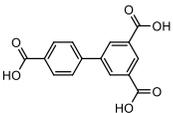
Table 5. The Number of Unique MOFs Estimated Based on the MOF Fingerprint Using +/- 10%

Threshold for Textural Properties

Methods	Number of Groups
Molecular formula (no hydrogens)	9,593 (out of 14,845)
Molecular formula with hydrogen	10,667 (out of 14,845)
Molecular formula with hydrogen and crystal density	11,209 (out of 14,845)
Molecular formula with hydrogen, crystal density, and other textural properties	11,923 (out of 14,845)
Individual CIF files	14,845 (out of 14,845)

Table of Contents Graphic

~5,000 3D MOFs   ~14,000 3D MOFs