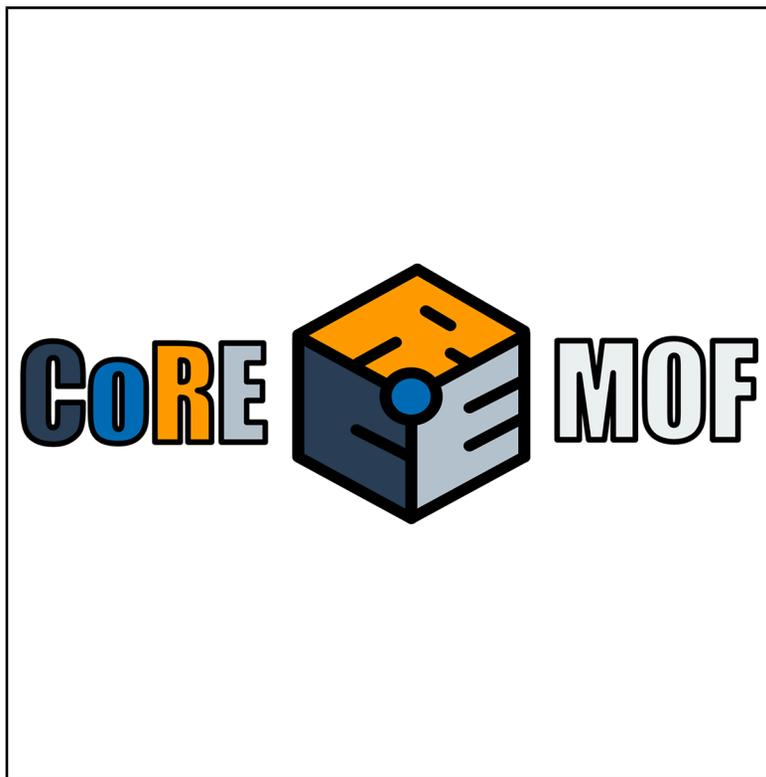


CoRE MOF DB: A curated experimental metal-organic framework database with machine-learned properties for integrated material-process screening

Graphical abstract



Highlights

- The updated CoRE MOF DB includes 40,000 experimental MOFs reported by early 2024
- Geometric and ML-predicted properties have been assigned to MOFs in the database
- TSA simulations were used to evaluate MOFs for carbon-capture scenarios

Authors

Guobin Zhao, Logan M. Brabson, Saumil Chheda, ..., J. Ilja Siepmann, Randall.Q. Snurr, Yongchul G. Chung

Correspondence

drygchung@gmail.com

In brief

A computational workflow was developed to streamline the curation and featurization of experimental metal-organic framework (MOF) structures, improving and updating the widely accessed CoRE MOF database. This updated CoRE MOF DB was screened computationally to identify top-performing MOFs for carbon capture across diverse CO₂ concentrations. An interactive website was developed to allow users to curate and evaluate their own MOF structures. This integrated platform accelerates computational MOF design and discovery for carbon capture and beyond.



Demonstrate

Proof-of-concept of performance with intended application/response

Zhao et al., 2025, Matter 8, 102140
June 4, 2025 © 2025 Elsevier Inc. All rights are reserved, including those for text and data mining, AI training, and similar technologies.
<https://doi.org/10.1016/j.matt.2025.102140>

Article

CoRE MOF DB: A curated experimental metal-organic framework database with machine-learned properties for integrated material-process screening

Guobin Zhao,¹ Logan M. Brabson,² Saamil Chheda,^{3,4,5} Ju Huang,⁶ Haewon Kim,¹ Kunhuan Liu,⁷ Kenji Mochida,⁷ Thang D. Pham,⁷ Prerna,^{3,4} Gianmarco G. Terrones,⁸ Sunghyun Yoon,¹ Lionel Zoubritzky,^{9,10} François-Xavier Coudert,^{9,15} Maciej Haranczyk,^{11,15} Heather J. Kulik,^{8,12,15} Seyed Mohamad Moosavi,^{6,15} David S. Sholl,^{13,15} J. Ilja Siepmann,^{3,4,15} Randall.Q. Snurr,^{7,15} and Yongchul G. Chung^{1,14,15,16,*}

¹School of Chemical and Biomolecular Engineering, Pusan National University, Busan 46241, South Korea

²School of Chemical and Biomolecular Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA

³Department of Chemistry and Chemical Theory Center, University of Minnesota, 207 Pleasant Street SE, Minneapolis, MN 55455, USA

⁴Department of Chemical Engineering and Materials Science, University of Minnesota, 421 Washington Avenue SE, Minneapolis, MN 55455, USA

⁵Department of Chemistry, Kavli Energy Nanoscience Institute, and Bakar Institute of Digital Materials for the Planet, College of Computing, Data Science, and Society, University of California, Berkeley, Berkeley, CA 94720, USA

⁶Chemical Engineering and Applied Chemistry, University of Toronto, Toronto, ON M5S 3E5, Canada

⁷Department of Chemical and Biological Engineering, Northwestern University, Evanston, IL 60208, USA

⁸Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

⁹Chimie ParisTech, PSL University, CNRS, Institut de Recherche de Chimie Paris, 75005 Paris, France

¹⁰Air Liquide, Campus Innovation Paris, Les Loges-en-Josas 78350, France

¹¹IMDEA Materials Institute, Getafe, 28906 Madrid, Spain

¹²Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

¹³Oak Ridge National Laboratory, Oak Ridge, TN 37830, USA

¹⁴Graduate School of Data Science, Pusan National University, Busan 46241, South Korea

¹⁵Senior author

¹⁶Lead contact

*Correspondence: drygchung@gmail.com

<https://doi.org/10.1016/j.matt.2025.102140>

PROGRESS AND POTENTIAL The availability of high-quality materials datasets is foundational for accelerating materials discovery. Metal-organic frameworks (MOFs) are a class of crystalline solids where the number of experimentally reported MOFs continues to grow in the literature. We report a computational workflow to curate a database of experimentally reported MOFs (CoRE MOF DB) that includes over 40,000 MOFs reported by early 2024. The updated CoRE MOF DB includes pre-computed material properties that are useful for high-throughput material-process screening. We demonstrate the utility of the database by performing integrated material-process screening for carbon-capture applications.

SUMMARY

We present an updated version of the Computation-Ready, Experimental (CoRE) Metal-Organic Framework (MOF) database, which includes a curated set of computation-ready MOF crystal structures designed for high-throughput computational materials discovery. Data collection and curation procedures were improved from the previous version to enable more frequent updates in the future. Machine-learning-predicted properties, such as stability metrics and heat capacities, are included in the dataset to streamline screening activities. An updated version of MOFid was developed to provide detailed information on metal nodes, organic linkers, and topologies of an MOF structure. DDEC6 partial atomic charges of MOFs were assigned based on a machine-learning model. Gibbs ensemble Monte Carlo simulations were used to classify the hydrophobicity of MOFs. The finalized dataset was subsequently used to perform integrated material-process screening for various carbon-capture conditions using high-fidelity temperature-swing adsorption (TSA) simulations. Our workflow identified multiple MOF candidates that are predicted to outperform CALF-20 for these applications.



INTRODUCTION

Metal-organic frameworks (MOFs) are well-established nanoporous materials with potential commercial applications in various technologically important utilizations, such as carbon capture,¹ hydrogen storage,² catalysis,³ and beyond.^{4–7} A diverse range of MOFs can be synthesized by selecting appropriate secondary building units (SBUs) and organic linkers.^{7,8} There are now more than 100,000 experimentally reported MOF and MOF-like materials in the literature, and the number continues to grow due to the large MOF chemical space. Additionally, MOFs are attracting interest in critical research areas, such as direct air capture (DAC).^{9,10}

In recent years, there has been a significant interest in virtual high-throughput screening activities to identify high-performing MOFs for applications such as hydrogen storage and carbon capture.¹¹ Large databases of MOFs play a critical role in enabling these screenings as the number of both experimental and hypothetical MOFs continues to grow. These databases have been the foundation of accelerated MOF discovery using machine learning (ML) and artificial intelligence.^{12–15} However, direct use of MOF crystal structures from experimental X-ray diffraction poses challenges in accurately modeling MOF performance for targeted applications due to the presence of crystallographic disorders and residual solvents in the framework. Crystallographic disorders arise often due to structural flexibility of the material, leading to multiple conformations within the same structure, which can result in partial occupancies, distorted bond lengths, or disordered solvents (ions), all of which can affect the computational predictions.^{16–21}

Motivated by these issues, in 2014, we developed the Computation-Ready, Experimental Metal-Organic Framework (CoRE MOF 2014) database,²² which included approximately 5,000 3D experimentally reported MOFs. That work used a graph-based algorithm to remove both weakly bound (free) solvent molecules and strongly bound solvent molecules that are coordinated to unsaturated metal sites, similar to the activation or de-gassing procedures widely used in experiments prior to gas adsorption measurements. In some cases, the removal of solvents that are coordinated to unsaturated metal sites could lead to structural collapse, as these solvents can contribute to framework stability. The CoRE MOF 2014 collection did not account for these effects. In 2019, we expanded the CoRE MOF database by collecting missing structures in the Cambridge Structural Database (CSD) from the open literature while providing in-depth structural analysis of the dataset. One of the key advancements in this database is the introduction of a separate dataset in which coordinated solvent molecules are not removed. The CoRE MOF 2019 database includes three types of structures—all solvent removed, free solvent removed, and ionic structures (structures that contain charge-balancing ions)—expanding the CoRE MOF 2014 database from 5,000 up to 14,142 all-solvent-removed MOF structures.²³ Despite these advancements, there have been several reports in the literature^{17,21,24} related to the shortcomings of the CoRE MOF database and related experimental MOF databases (CSD MOF subset),²⁵ particularly due to the presence of disordered or

not-computation-ready (NCR) structures in the dataset. These shortcomings impose several practical limitations in materials-discovery activities, such as potentially unrealistic structure-property relationships and identification of unphysical structures as top-performing materials, which could hamper the utility of the databases and experimental realization of the predicted high-performance MOFs.²⁶

In this work, we present an updated CoRE MOF database with more than 40,000 experimental MOFs, addressing many of the previously reported limitations and incorporating new MOF structures published since CoRE MOF 2019. The collected and curated MOF crystal structures were classified to distinguish computation-ready (CR) from NCR MOFs, following a systematic approach using two independent validation methods to detect coordination errors, geometric inconsistencies, and missing structural elements. To further enhance the utility of the dataset, we pre-computed ML-predicted properties related to framework stability (water stability, solvent-removal stability, and thermal stability) and heat capacity. Gibbs ensemble Monte Carlo (GEMC) simulations were carried out to further classify MOFs based on their hydrophobicity. We demonstrate the utility of the dataset by performing integrated material-process screening for carbon-capture applications combining molecular simulation and a temperature-swing adsorption (TSA) process. From the material screening, we identified 34 MOFs with improved CO₂/N₂ selectivity compared to CALF-20 based on the equilibrium performance metrics (selectivity and working capacity) predictions. Top materials emerging from the screening were subsequently evaluated under various carbon-capture conditions via TSA simulation. We identified several Zn-based and Co-based MOFs with relatively inexpensive ligands that meet 90% purity and recovery requirements for CO₂ capture even at 1% gas-phase CO₂ feed concentration. This work demonstrates a powerful integration of high-fidelity MOF datasets, ML models, and multi-scale modeling for accelerated MOF exploration and discovery.

RESULTS

Database construction and property analysis Structure acquisition and curation procedure

We identified 65,999 digital object identifiers (DOIs) of research papers related to “metal-organic frameworks” or “porous organic polymers” from 2016/01/01 to 2023/12/31 by querying these two phrases in the Web of Science (WOS). For structures published before 2016, we re-downloaded the original crystallographic information files (CIFs) from the CSD and publication supplemental information (SI) files based on information from the CoRE MOF 2019 database²³ (REFCODE and DOIs). To avoid missing any crystal structures due to the confusion between publication and pre-proof dates, we collected data from 2016/01/01 rather than 2016/04/01 (the CoRE MOF 2019 search cutoff date). After removing duplicate structures based on DOI and the filenames from the original crystal files, we obtained 37,565 original CIF files, which form the basis of the dataset (Figure 1A). This step was necessary due to a 3-month overlap between CoRE MOF 2019 and CoRE MOF 2024, leading to repeated collection of the same structures.

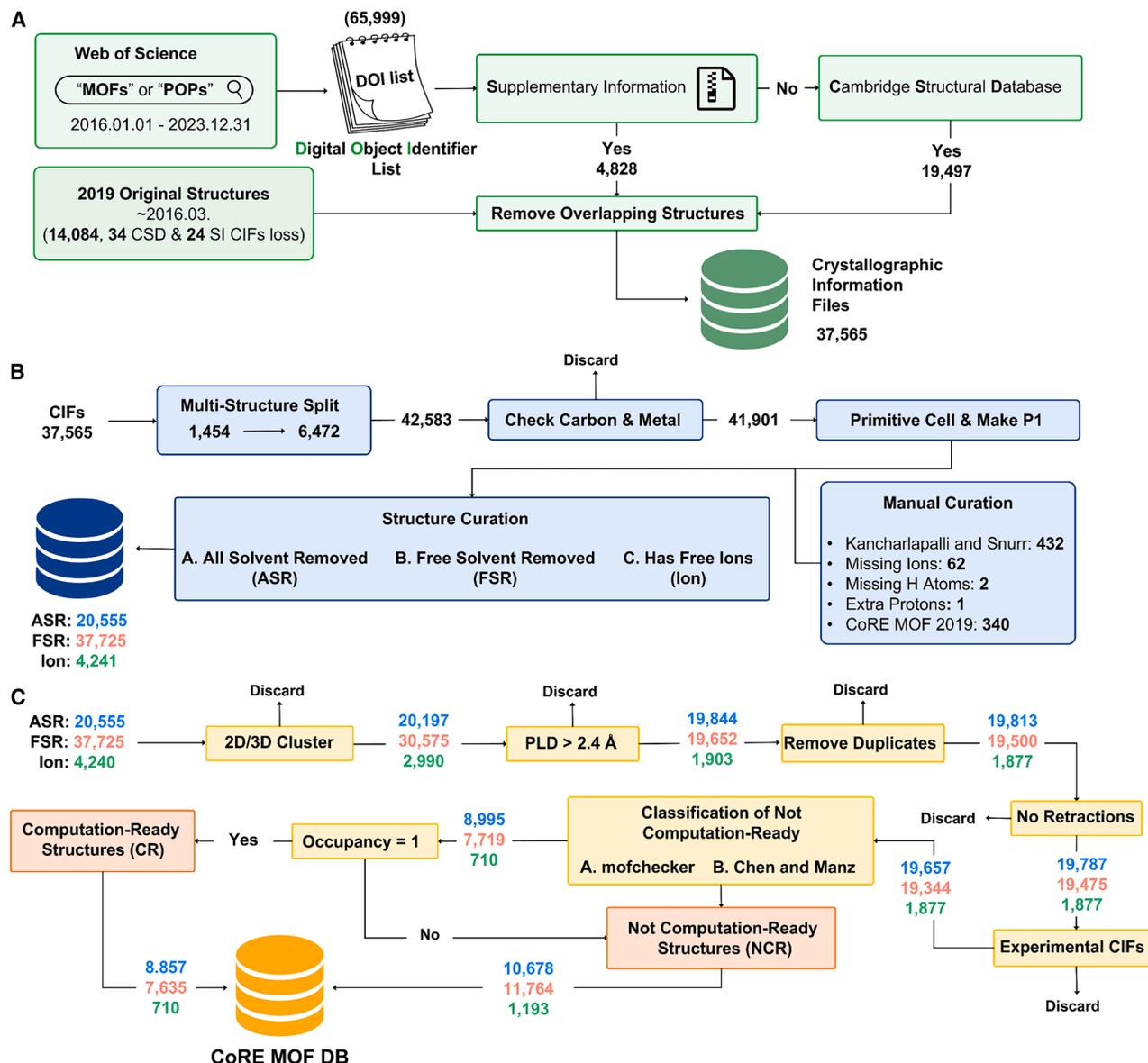


Figure 1. Overall workflow of CoRE MOF DB construction

(A–C) Schematic for MOF (A) CIF collection, (B) curation, and (C) classification. All numbers represent counts of CIFs except for (65,999) from (A), which represents the number of papers.

(A) Procedure of collecting CIF files from literature. The search date refers to when a paper is available online, including pre-proof and as soon as publishable (ASAP) of publications. The date format used in this work is YYYY/MM/DD.

(B) Procedure to curate and classify collected crystal structures. MOFs are classified into datasets of all solvent removed (ASR), free solvent removed (FSR), and MOFs with ions (Ion). The CIF structures of Ion datasets do not overlap with that of the ASR or FSR datasets. For assignment of ASR, FSR, and Ion structures, the workflow is shown in Figures S5–S7.

(C) Procedure to filter and classify MOFs as computation-ready (CR) and not computation-ready (NCR).

We subsequently performed a series of curation and classification procedures to create a more uniform and computation-ready MOF dataset. First, we identified CIFs that contained multiple MOF structures and split them into different files and labeled the new files with an “n” at the end of the filename, where n is the integer number identifier starting from 1. An example of such splitting is shown in Figure S1. From the new collected CIFs,

1,454 files contained more than one structure, and splitting these files increased the total number of CIFs to 42,583.

The next step of the curation procedure involved removing structures that do not contain carbon or metal atoms, which reduced the total number of structures to 41,901 (Figure 1B). Here, we first created the primitive cell of the initial structure and enforced P1 symmetry of the primitive cell to remove

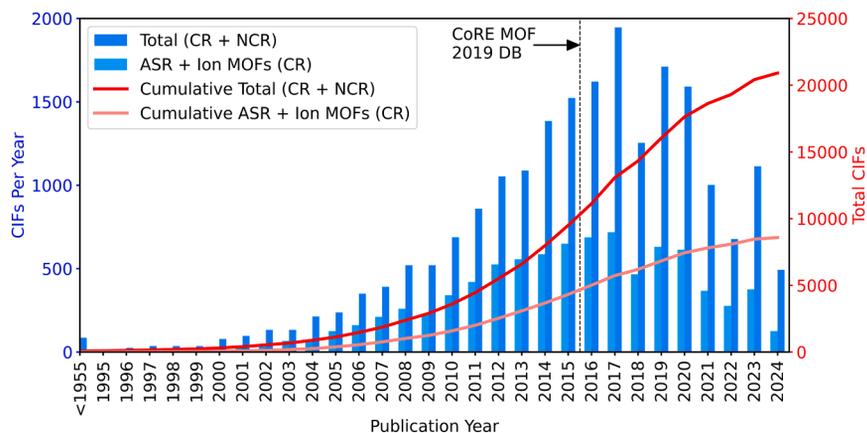


Figure 2. The number of experimental MOFs based on the publication year in the CoRE MOF DB

The deep-blue histogram represents the total number of structures in CoRE MOF DB reported by year, where the total is the sum of CR and NCR datasets minus FSR datasets. The number of structures is obtained as the sum of the ASR and Ion datasets, since there is a significant overlap in the ASR and FSR datasets. The NCR dataset is mutually exclusive of the CR dataset in the histogram. There are a few papers with a print publication year of 2024 because some papers were published online on 2023/12/31 but appeared later in print form. "Total" is calculated after the "Experimental CIFs" step (original structures) in Figure 1C. Figure S36 shows the publication year breakdown for the ASR, FSR, and Ion datasets.

symmetry (an example comparing the method used in this study with other methods can be found in Figure S2).²⁷ We then applied an improved solvent-removal algorithm to remove solvents while classifying structures with ions (based on a pre-tabulated stoichiometry of ions; Table S1) to create three separate datasets: all solvent removed (ASR), free solvent removed (FSR), and ionic MOFs (Ion). Details of the analyses and curation steps are provided in the [crystal structure cleaning and curation](#) section. The differences in CO₂ and CH₄ isotherms for original, FSR, and ASR structures (for the same MOF), as shown in Figures S10–S12, highlight the importance of solvent removal for molecular simulation. To further validate our solvent-removal algorithm, we applied structural activation via metal oxidation state analysis (SAMOSA)²⁸ to analyze all CR structures, and the results showed that no solvents were detected. These datasets also include structures that have been manually curated and reported in the literature. Note that the structures with free-floating ions are overall charge neutral. This created a baseline dataset that could be split to create CR and NCR datasets.

Similar to the CoRE MOF 2019 dataset, we only considered 2D and 3D MOFs²⁹ with pore-limiting diameter (PLD) larger than 2.4 Å (van der Waals [vdW] diameter of a hydrogen molecule) from the baseline dataset. While MOFs with smaller PLD may still be useful for mechanical-property studies, our focus here is on structures suitable for gas adsorption and storage applications. In order to maintain the quality of the CoRE MOF database (DB), we removed structures that have been retracted from the CSD³⁰ (i.e., those structures corresponding to a retracted paper or a paper that is under an expression of concern). This is of great importance given that several hundreds of MOF structures in the published literature are the result of paper mills, injecting fabricated results into the scientific literature.³¹ However, many publishers of the retracted papers are not indexed by the WOS. We will continue to monitor and remove retracted structures in future updates to improve data integrity. We also removed simulated CIF files (constructed from automated computer algorithms) so that every structure in the dataset is from experimental data. This was done by manually checking the reporting literature with more than two CIF files. Finally, we classified the remaining dataset into CR and NCR datasets. The classification was done primarily using two cheminformatic methods

available in the literature (MOFChecker³² and the Chen and Manz method¹⁷). Details of the analyses and steps are provided in the [classification of NCR structures](#) section. This created the two datasets (CR and NCR) that form the CoRE MOF 2024 DB v1. We retroactively applied the same classification steps to the CoRE MOF 2019 DB to create a curated collection denoted CoRE MOF 2019 v2.

We checked the number of MOF structures that we modified by the improved solvent-removal algorithm and manual-curation procedure. Figure S9 and Table S4 show that about 66% (56%) of the structures in the CR (NCR) dataset were modified. Table S4 gives a breakdown of the NCR dataset based on different cheminformatic classification methods. Structures in the NCR dataset have remaining issues related to metal nodes or linkers that cannot be resolved automatically. We noticed that the majority of the structures with problematic atomic occupancy originated from files obtained from the SI of literature, as these structures did not go through the Cambridge Crystallographic Data Centre (CCDC) manual-curation procedure. An important property of the CR dataset is that it only includes structures where the reported occupancy of every atom is 1 (see Figure 1). This is in contrast to the previous CoRE MOF datasets where we attempted to fix the problem either semi-automatically (by deleting atoms that are marked as partially occupied) or manually. The cheminformatics filters applied in our classification procedure detected many cases in which partial occupancies would have to be resolved before a structure can become CR (see Figure S14 for an example), but we did not attempt to further resolve such structures.

We found the majority of the structures in the CoRE MOF dataset have fewer than 500 atoms in the unit cell. In the ASR and FSR datasets, 92% of the structures contain fewer than 500 atoms (Figure S37). For the Ion dataset, 85% of the structures have fewer than 500 atoms. We found two MOFs (the well-known MIL-101(Cr)³³ and PCN-21³⁴) that have more than 3,500 atoms in their unit cells. These two structures exhibit high porosities and exceptionally large pore sizes (Figure S38).

Figure 2 shows the number of experimental MOFs reported in the literature over time. In this figure, we only show the number of experimental MOF structures by excluding structures in FSR datasets because these structures are similar to the ASR dataset.

Table 1. Number of CIF files in the Public CoRE MOF 2019 DB (CoRE MOF 2019 v1), CoRE MOF 2019 DB with new NCR classifications (CoRE MOF 2019 v2), CSD MOF set, QMOF DB, and CoRE MOF DB

DB	ASR	FSR	Ion	CR	NCR ^a
CoRE MOF 2024 v1	8,857 (6,963) ^b	7,635 (4,978) ^b	710 (558) ^b	17,202 (12,499) ^b	23,635 (16,077) ^b
CoRE MOF 2019 v2 ^c	6,603	3,764	–	10,367	8,714
CoRE MOF 2019 v1 ^d	10,143	6,008	–	16,151	2,930
QMOF v14 ^e	–	14,390	–	14,390	5,985
CSD MOF subset v5.43 ^e	–	–	–	7,656	4,849

^aThe NCR category includes both the ASR and FSR structures.

^bThe numbers in parentheses represent the number of publicly available structures in the CoRE MOF DB (CoRE MOF SI + CoRE MOF Modified CSD).

^cCoRE MOF 2019 v2 DB was obtained through the same cheminformatic checks during the NCR classification process depicted in Figure 1C. CoRE MOF 2019 v2 DB is based on the public version of the database. We created a new classification called Ion for structures containing ions in the ASR dataset from the CoRE MOF 2019 database.

^dCoRE MOF 2019 v1 is based on the internal version of the database.

^eResults for other databases (QMOF,³⁶ CSD MOF subset²⁵) using the same method are shown in Figures S31–S35.

There has been a significant increase in the overall number of reported structures throughout the years, with the number of experimental MOFs reported from 2014 to 2020 reaching 1,200 annually. The number of NCR structures from our classification accounts for 60% of the total reported structures, indicating a need to improve the current CIF curation procedure by developing automated cheminformatic tools or community-driven effort in curating CR datasets.

Table 1 shows the number of ASR structures in the CoRE MOF 2024 DB along with other experimental MOF datasets available in the literature. The Chen and Manz method detects overcoordinated and undercoordinated atoms, while MOFChecker identifies structures with geometrically exposed metals and missing counterions, among other structural inconsistencies. These methods go beyond simple interatomic distance checks, enabling more accurate error detection. The NCR and CR classification of the CoRE MOF 2019 DB v1 reveals that approximately 50% of the structures belong to the NCR dataset. This observation is consistent with previous analyses by Chen and Manz¹⁷ and others.^{9,24} These cheminformatics methods have some overlapping algorithms to check for NCR structures but differ in the covalent radii used to define the crystal network (Figures S3 and S13). Figure S19 shows the comparison of two NCR classification methods on CoRE MOF 2019 and 2024 DBs. There is a large overlap between the two methods. To achieve the most accurate classification possible for the CoRE MOF DB, we retain only structures that passed both the Chen and Manz and MOFChecker validation methods. While this classification is systematic and structured, we acknowledge it is not perfect, and further improvements in NCR detection remain an ongoing research challenge in the field. Figures S20–S27 show the frequency of various NCR cases identified by each checking method. Among all types of NCR, unsaturated bonds are the most common, accounting for more than 8% of the total NCR instances. For instance, where two NCR cases occur simultaneously, we found “atom overlapping” or “has_atomic_overlaps” and “has_overcoordinated” or “Over bonded carbon” are the most common types of NCR, accounting for more than 30% of the NCR instances. For the Ion dataset, we did not check for isolated molecules, as ions are present as free-floating spe-

cies. While these two methods (MOFChecker and Chen and Manz) are effective, there remains room for future algorithmic improvement. While we do not have precise false-positive/negative rates, we acknowledge that these methods are imperfect, as demonstrated by the undetected errors in BAKGIF and FIBKEI. These methods may not reliably detect incorrect protonation states or incorrectly classify correct structures as incorrect. We discuss two examples of such cases in supplemental information and S2.9. For example, SBMOF-1, which is known for its excellent Xe/Kr separation performance,³⁵ was classified as NCR with a “has isolated molecule” flag by the Chen and Manz method. This is due to the covalent radius used for Ca atoms when creating an adjacency matrix (Figure S28). Similarly, several HKUST-1 structures were classified by MOFChecker as having a “geometrically exposed metal” while others were not. These incorrectly classified MOFs are classified into the NCR dataset, but we believe there could be similar instances like this which we have not found yet.

We note that the current CR and NCR classification scheme is not perfect. While bond-level and geometry-based NCR checks are fast, they do not guarantee accurate classification. Perfect classification of the NCR and CR datasets is currently an unsolved problem in this area, but emerging approaches in the community could aid in future development of the dataset.^{24,28} We compared CoRE MOF 2024 ASR structures with Metal Oxidation State Automated Error Checker (MOSAEC)-DB³⁷ and found that 112 MOFs had successfully removed solvents in CoRE MOF 2024 (such as DMF, acetonitrile), whereas MOSAEC-DB failed to remove these. Conversely, MOSAEC-DB successfully removed solvents from 25 MOFs (such as pyridine) that CoRE MOF 2024 could not. We note that even after running the SAMOSA script on these structures, the solvents could not be removed. Manual curation of crystal structures has previously been used to classify and curate individual MOF structures, but, without careful inspection of the original literature, this approach could also potentially lead to errors. For example, we found that one manually curated structure from the CoRE MOF 2019 DB v1 did not match the information reported in the literature, as the number of protons was incorrect because of an error by the hydrogenation protocol in the

Table 2. The percentage of structures with successful calculations or available data

Properties ^a	ASR	FSR	Ion
Geometric_Zeo++ ^b	100.00%	100.00%	100.00%
Geometric_Topology	97.38%	97.29%	98.60%
Geometric_OMS	100.00%	100.00%	100.00%
MOFid_v1	82.70%	95.87%	89.06%
MOFid_v2 ^c	76.00%	–	1.8%
ML_ q^d	100.00%	100.00%	100.00%
ML_ C_p^e	100.00%	100.00%	100.00%
ML_ T_d^f	99.03%	97.64%	5.75%
ML_ $P_{solvent}^g$	99.03%	97.64%	5.75%
ML_ P_{water}^h	99.03%	97.64%	5.75%
DOI ⁱ	99.79%	99.87%	100.00%

^aDigital objective identifier.

^bIf the property calculations fail, we record “unknown” in the excel spreadsheet.

^cZeo++ calculation includes pore sizes (PLD, pore-limiting diameter; LCD, largest cavity diameter), pore volume, surface area, and framework dimensions.

^dMOFid-v2 was only applied to the ASR and Ion dataset.

^ePartial atomic charge based on PACMAN density-derived electrostatic and chemical (DDEC6) charges.

^fHeat capacity.

^gDecomposition temperature.

^hThe probability of solvent-removal stability.

ⁱThe probability of water stability.

Materials Studio software. Even after locating the problem, it is difficult to assign which proton topology is correct based on the available stoichiometry information from experiments. To address this issue, we generated nine possible conformers and identified the most stable configuration through density functional theory (DFT) calculations. We observed significant differences, up to an order of magnitude, in the Henry's coefficient for water among the different conformations (see Figures S29 and S30; Table S5).

MOFid and geometric and ML-predicted properties

We computed commonly used geometric features,²⁹ the MOFid,³⁸ and ML-predicted properties for CR structures. Details of how these properties were calculated are provided in the methods section. Table 2 summarizes the success rates of each set of property calculations. Within the category of geometric properties, topology was the most challenging aspect to automate; however, we successfully matched over 97% of the structures with their corresponding topologies—failures being due to unstable nets or interpenetrated nets with mixed topologies. MOFid-v2, compared to MOFid-v1, provides more detailed information about metal nodes but results in fewer successful cases. We observed that, even though we were able to assign MOFid-v2 for 76% of the structures, many of these structures have unknown topologies (i.e., “nan”). An unknown topology is a valid net that has not been named in the Reticular Chemistry Structure Resource (RCSR) database. It would be interesting in the future to further classify the underlying unknown topologies using detailed topology analysis and statistically compare the different methods of topological analysis.

The ML-predicted properties are more successfully determined largely because many of these models rely on pre-computed chemical and geometric features. We notice that the stability predictions for the Ion dataset could not proceed due to failures in feature generation, primarily due to the presence of free ions far from the framework. Lastly, less than 1% of structures are missing their DOI. These structures were either directly deposited to the CCDC or are from publishers that do not assign a DOI to their articles.

We examined the distribution of crystal space groups in different datasets. Figure S39 shows that the distribution of space groups of the structures remains almost unchanged after a series of structure curation procedures. Geometric feature analysis (Figure S40) shows that the removal of solvents contributes to an increase in porosity (as shown by wider distribution of pore volume, void fraction, and accessible surface areas). The Ion dataset tends to have smaller pores than the ASR and FSR datasets because of the presence of free-floating ions that block the pores (Figure S40). Compared to FSR and Ion, ASR structures exhibit a slightly broader distribution of pore features, with maximum accessible gravimetric and volumetric surface areas of 6,511 m²/g and 3,562 m²/cm³ (when N₂ is used as the probe), respectively. The open metal site (OMS) analysis shows Zn and Cu are the most frequently occurring metal atoms in the dataset, with over 92% of MOFs having only one metal (Figures S41 and S42). Bimetallic MOFs account for about 7% of the total structures, while MOFs with more than two metal elements are even rarer; only nine structures have more than two metal elements. This is similar to the statistics from a database of the 1D MOFs reported in 2022, where 10% of the MOFs include two or more metals.²⁹ Among the 23 multi-metallic MOFs in the CoRE MOF 2024 DB, nine structures have assigned topologies: **sql**, **pth**, and **dia** from the ASR dataset, and **srs** from the Ion dataset. Notably, all of these structures have OMSs, and only one is interpenetrating. The most frequent metal combinations are Cu/Mn/Mo, Zn/Mn/Mo, Co/Pt/Pd, and Zn/Pt/Au (Table S6). The number of structures with OMSs in FSR and Ion is much lower than that in ASR due to the removal of solvent molecules coordinated to metals in the case of the ASR dataset. OMS analysis shows the dataset includes both transition metals and most of the lanthanides (Figure S42). With the decision to include 2D MOFs in this update, we found 2D MOFs account for 14% of the entire database (Table S7; Figure S43).

Topology

As shown in Figure 3, the relative frequencies of the most common topologies are similar to those in the CoRE MOF 2019 v2 DB: primitive cubic lattice (**pcu**) and diamond (**dia**) remain by far the most common 3D topologies, and square linear (**sql**) and honeycomb (**hcb**) are the most common 2D topologies. The distribution of topologies has a fat tail (see Figure S44). 4% of the structures have a topology encountered fewer than five times, whereas the five most common topologies encompass 24% of all structures, highlighting the large diversity of the MOF topological space. Ionic structures show a comparatively higher fraction of SrSi (**srs**), UCl₃ (**gfy**), and **nuc** 3D topologies, but this may simply be a statistical variation due to the low number of structures in the Ion dataset (Figure S45). Interpenetrating topologies follow a roughly similar distribution

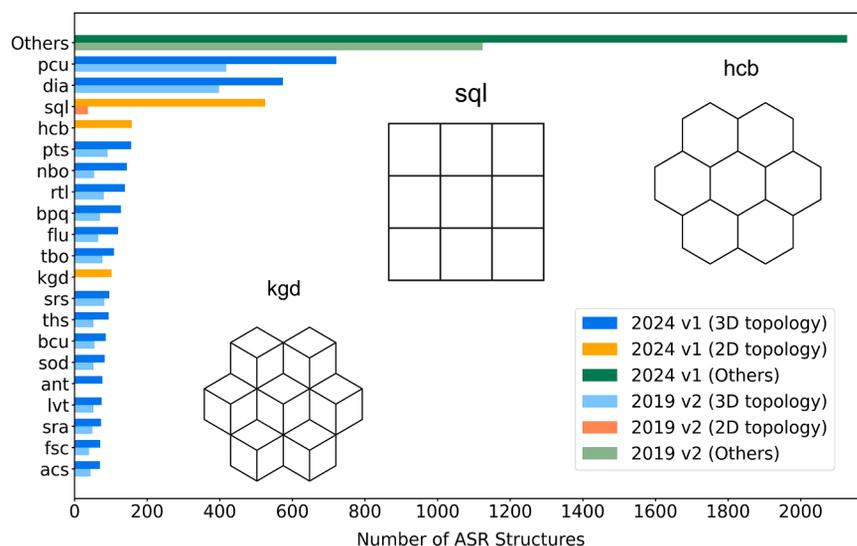


Figure 3. Distribution of the most frequent MOF topologies (top 20) among ASR structures from CoRE MOF 2019 and CoRE MOF 2024 DBs

The CoRE MOF 2019 v2 is a subset of the CoRE MOF 2024 DB. 3D topologies and 2D topologies are color coded in blue and orange, respectively. Since previous studies only collected 3D MOFs, there are fewer structures with 2D topologies in the CoRE MOF 2019 DB. The “Others” category excludes unnamed cases and encompasses the topologies beyond the most common 20—highlighting the important diversity of topologies in the database. The three most frequent 2D topologies are shown. The topologies are arranged in descending order of frequency in the CoRE MOF 2024 DB.

(Figure S46), with **dia** and **pcu** the most common 3D interpenetrated topologies, followed by **srs** and ThSi_2 (**ths**), similar to the observations of Proserpio and coworkers.³⁹ We note that **dia** shows a greater range of interpenetration, from 2- to 8-fold in more than 10 structures, and also features cases of 9-fold (2016[Cu][dia]2[ASR]1) and 10-fold interpenetration (2015[Cd][dia]3[ASR]6), whereas **pcu** is only present up to 4-fold in more than 10 structures and 5-fold in one structure (2018[Zn][pcu]3[ASR]2). Using CrystalNets.jl,⁴⁰ we identified a significant number of structures in the ASR dataset that have an unnamed topology, i.e., a topology we can clearly identify but that is not listed in the RCSR database and does not have an assigned three-letter code. We have analyzed by hand the six most common topologies of this kind, confirmed that they are indeed quite common, and highlighted some details in Table 3 and Property analysis.

MOFid-v2

MOFid³⁸ is a cheminformatics tool designed to provide a textual identifier based on the composition and topology of MOFs. The original MOFid algorithm breaks down MOFs into their constituent components (organic linkers and SBUs) and identifies their topology and degree of catenation. We generated MOFid-v1 for the CoRE MOF 2024 ASR (7,322), FSR (7,320), and Ion (635

CIFs) datasets as well as an updated MOFid-v2 (Note S5) on the CoRE MOF 2024 ASR (6,732) and Ion (13) datasets.

Using MOFid-v2, we determined the most common building blocks (SBUs and organic linkers) as well as the number of MOFs associated with each type of SBU and organic linker finding 678 unique organic linkers and 1,081 unique SBUs. The most common SBU is the single-atom Zn metal node ([Zn_Type1]). This SBU is a key component in various ZIF structures, as well as other MOFs including MFM-520 and kag-MOF-1. The next most common SBU is the Cu paddlewheel ([C4Cu2O8_Type-1]), which is present in HKUST-1, PCN-12, PCN-21, UTSA-28, UTSA-60, MFM-136, and MFM-180 (Figure S49). The 15 most common SBUs account for 43% of all nodes in these 6,861 structures (successfully assigned MOFid-v2 with each structure containing only a single type of SBU). The most common linkers are benzene-1,3,5-tricarboxylate (BTC), benzene-1,4-dicarboxylate (TPA/BDC), and 2,5-dicarboxybenzene-1,4-diolate (DCBD) (Figure S50B). The top 15 linkers account for 28% of all the unique linkers in the structures analyzed. Figure S51 and Table S14 show the 3D structures of the most common SBUs, and Table S13 shows the structures of the most common linkers.

Figure 4 shows the most frequently occurring combinations of building blocks. For this analysis, we focused on combinations of one SBU and one organic linker. The most common SBU-linker combinations are (1) C4Cu2O8_Type-1 and benzene-1,3,5-tricarboxylate, with 82 occurrences, and (2) C6O13Zn4_Type-1 and benzene-1,4-dicarboxylate, with 53 occurrences. The first combination corresponds to HKUST-1 (or Cu-BTC),⁴¹ and the second combination corresponds to IRMOF-1 (or MOF-5).⁴² SBUs in the form of single metal ions are very common in MOF structures and can often coordinate to the same linker even if the metal element changes, such as in MOF-74⁴³ and $\text{M}_i(\text{HCOO})_j$.^{44,45}

Structures were checked for similarity using MOFid-v2. This is different than the “duplicate” check, as the “duplicate” check only verifies whether the structures are completely identical, while the “unique” check considers whether the structures have the same metal node, linker, and topology. This means

Table 3. Six most common “unnamed” topologies found in the CoRE MOF 2024 DB, identified by CrystalsNets.jl but without a corresponding net and three-letter code in the RCSR database

	N^a	Common MOFs with topology	Related/proposed topologies	TopCryst code
#1	117	MOF-74/CPO-27	<i>bnn, etb, pts, msf</i>	4,8T1
#2	82	NOTT-300/MFM-300	–	6,8T21
#3	58	–	–	5,5T13
#4	51	CUK-1	<i>fee</i> , “ Δ -chain”	3,5,5,6T3
#5	47	$\text{Fe}_2(\text{BDP})_3/\text{Sr}_2(\text{BDC})_3$	<i>sct</i>	2,4C4
#6	26	Cu-SIP-3	<i>sql</i>	3,4,5,6T8

^aNumber of occurrences.

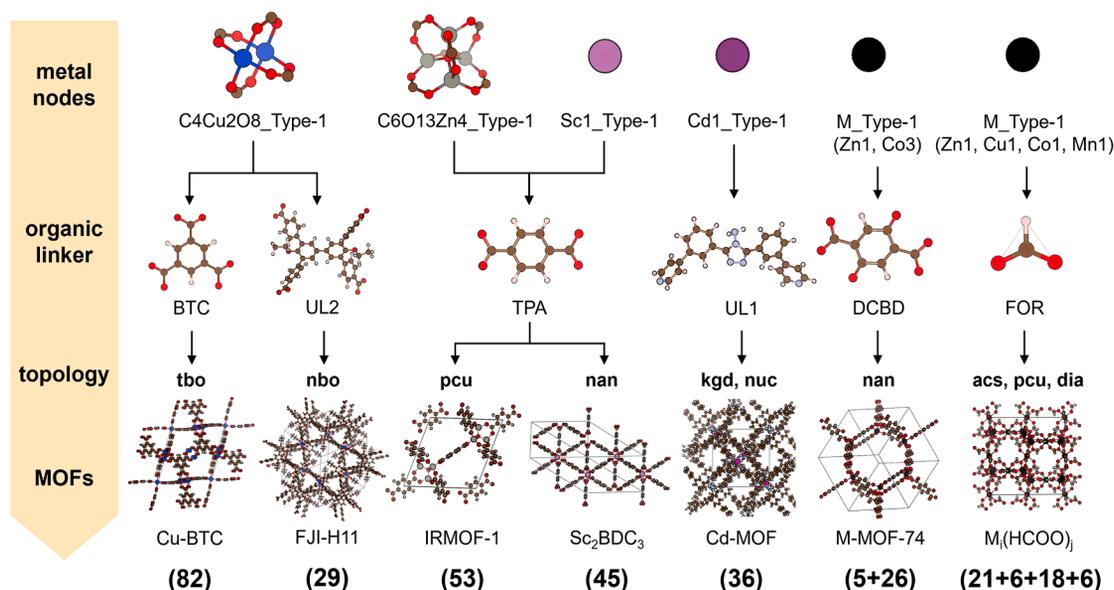


Figure 4. Combinations of the most common SBUs (metal nodes) and organic linkers in the ASR datasets having only one unique node and linker

Combinations with different topologies ("nan" indicates that the topology cannot be assigned by MOFid-v2.), where M indicates a type of metal. *i* and *j* of M_i(HCOO)_j indicate the number of different components in each structure. The detail of organic linker is listed in Table S13. Numbers in the parentheses represent the number of structures in the ASR dataset.

that even the same MOF synthesized by different experimental researchers will be detected as the same, as their MOFid-v2 should be identical, despite differences in their experimental crystal parameters. As shown in Table S16, many well-known MOFs were identified as having been uploaded multiple times to the ASR dataset. Ultimately, a total of 6,984 unique MOFs were obtained, and 706 MOFs were identified as having been reported multiple times. The list of unique MOFs can enhance ML training by preventing data leakage, where the same MOF appears in both the training set and the test set.

Diversity analysis

Diversity metrics and feature categories are explained in the methods section and Diversity analysis. Diversity in crystal structures is an important consideration for dataset preparation for ML tasks.⁴⁶ We performed diversity analysis to gain a deeper understanding of the CoRE MOF 2024 DB in the context of other databases reported in the literature. Across all feature categories (metal, linker, functional group, and geometry), the diversity metrics for CoRE MOF 2019 v2 ASR and CoRE MOF 2024 ASR are largely similar, but CoRE MOF 2024 has slightly larger balance, variety, and disparity, as might be expected due to the presence of new MOFs in CoRE MOF 2024 (Table S12). The set of MOFs pulled from 12 hypothetical datasets is more diverse than CoRE MOF 2024, except in the case of metal chemistry, in keeping with the previously observed lack of metal diversity in hypothetical MOF databases.⁴⁷ This indicates that the CoRE MOF DB contains a wider variety of metal node types than hypothetical MOF databases, which is consistent with previous work.

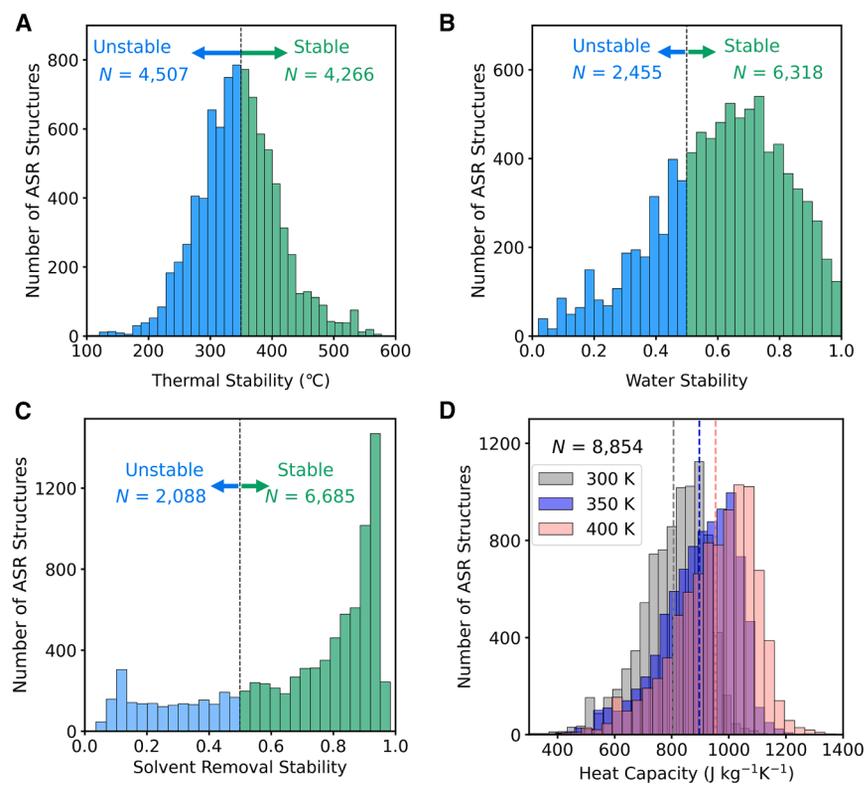
ML-predicted DDEC6 partial atomic charges

High-fidelity partial atomic charges are necessary to accurately model the interaction between quadrupolar and polar molecules,

such as CO₂ and H₂O, and the framework atoms.⁴⁸ For the updated CoRE MOF DB, we assigned DDEC6 partial atomic charges for all structures (including CR and NCR) using a recently developed ML model called PACMAN.⁴⁹ The charges are reported in the processed CIFs. After assigning the partial atomic charges, we analyzed the distribution of predicted charges in the CR and NCR datasets (see Figures S57 and S58). We observed that the charge distributions of certain metal elements (such as Ru, Nb, Ir, As, Pb, and Sn) are significantly different in CR and NCR structures. For other metals (such as In, Sb, Ge, Zr, Zn, V, Y, and lanthanides), the structures in the NCR dataset exhibited a similar charge distribution to the CR dataset. The partial atomic charges on metal atoms in SBMOF-1 and HKUST-1 from the NCR dataset (Ca, 1.4017; Cu, 0.8453) fall within the charge distribution range of their CR counterparts. We assigned partial atomic charges to the Ion dataset because the PACMAN method could correctly assign partial atomic charges to ions in a structure without being explicitly informed of the presence of ions. Figure S59 shows the distribution of the net charges on anions and cations. The 10 most frequently occurring ions in structures are listed in Table S17.

ML-predicted stability metrics and heat capacity

ML models can facilitate the fast and accurate estimation of properties for large numbers of materials without requiring computationally intensive calculations or experimentally challenging measurements. While many ML models have been developed to estimate adsorption properties^{50–52} in MOFs, ML models for evaluating stability or other material properties have not been reported until recently. Nandy et al. first developed stability models for solvent removal (i.e., activation) and thermal stability using natural language processing (NLP) to mine literature

**Figure 5. ML-predicted material properties**

Predictions are over MOFs in the CoRE MOF 2024 ASR dataset. Note that a predicted activation and water-stability value close to 0.5, as predicted by ANN and random forest, represents significant uncertainty, and filtering based on these values may lead to false exclusions.

(A) Thermal decomposition temperatures (dotted line is $T = 350^{\circ}\text{C}$).

(B) Probabilities of water stability (dotted line is $p = 0.5$).

(C) Probabilities of solvent-removal stability (dotted line is $p = 0.5$).

(D) Gravimetric heat capacities at $T = 300, 350$ and 400 K. Dashed lines are average predicted heat-capacity values at specified temperatures (section S6.3).

data.⁵³ Another relevant metric is the heat capacity, which is a useful property for accurately ranking the relative performance of MOFs in carbon-capture applications.⁵⁴ Predictions of the heat capacity of different materials can be used for more accurate process modeling compared to older work that simply assumed the same value ($985 \text{ J kg}^{-1} \text{ K}^{-1}$)⁵⁵ for all materials. To alleviate this, Moosavi et al. developed a regression model for accurately predicting MOF heat capacity, which has been corroborated with experimental data.⁵⁴ Terrones et al. created classification models to predict water and acid stability in MOFs based on literature data.⁵⁶ These models have been successfully incorporated into high-throughput computational screening workflows for MOF discovery.^{47,57,58}

We predicted stability metrics for MOFs in the CR MOFs based on the developed models.⁵³ We only discuss the results from the ASR dataset here as that dataset is used for our subsequent screening activity. The mean predicted thermal decomposition temperature (i.e., thermal stability temperature) is 349°C , with a standard deviation of 66°C . The predicted thermal decomposition temperature of ASR, FSR, and Ion datasets can be found in Figures 5A, S60A, and S60B. Of the MOFs for which solvent-removal predictions were obtained, 76% are predicted to be stable upon activation (see Figures 5C, S60E, and S60F). In the case of water stability, 28% of MOFs are predicted to be water unstable, and 72% are predicted to be water stable (see Figures 5B, S60C, and S60D).

In addition to stabilities, we predicted the heat capacities for all CR MOFs (distributions of these predictions are shown in Figures 5D and S62). The predicted heat capacities range from

200 to $1,300 \text{ J kg}^{-1} \text{ K}^{-1}$, although 38% of the ASR structures have predicted heat capacities within $200 \text{ J kg}^{-1} \text{ K}^{-1}$ of the commonly used heat-capacity value of $985 \text{ J kg}^{-1} \text{ K}^{-1}$. The average heat capacities of the ASR dataset at 300, 350, and 400 K are 807, 897, and $954 \text{ J kg}^{-1} \text{ K}^{-1}$, respectively, with standard deviations of 115, 132, and $141 \text{ J kg}^{-1} \text{ K}^{-1}$.

Figure S65 reveals that the updated CoRE MOF DB has a wider heat-capacity distribution compared to CoRE MOF 2019 v2. The distribution of heat capacity of CoRE MOF FSR and Ion datasets can be found in Figure S63.

We also performed predictions for the MOFs in the NCR dataset, with the results presented in Figure S61 and Table S18. We found that the prediction success rates in the NCR dataset were significantly lower than those in the CR dataset by comparing with the data in Table 2. This is due to the presence of various disorders in the structures, which prevented effective feature assignment. For stabilities prediction, the NCR dataset contained a higher proportion of thermally stable structures compared to the CR dataset, while the trends in water stability and solvent-removal stability remained consistent across both datasets. For heat capacity, the predicted values in the NCR dataset were slightly lower than those in the CR dataset at the same temperature, with the NCR dataset also exhibiting lower uncertainty.

Hydrophobicity classification

The affinity of MOFs for water at high and low pressure is an important consideration for their application in water harvesting and carbon capture under humid conditions. Strongly water-adsorbing MOFs (hydrophilic MOFs) may not be appropriate for carbon-capture applications because adsorption sites may be occupied by water molecules and/or the pores may be saturated with water instead of CO_2 . Water harvesting from air requires MOFs with tuned intermediate affinity for water, so that saturation occurs only in the range from 10% to 30% relative humidity. We performed both Widom ghost-molecule-insertion calculations and GEMC simulations over a wide range of

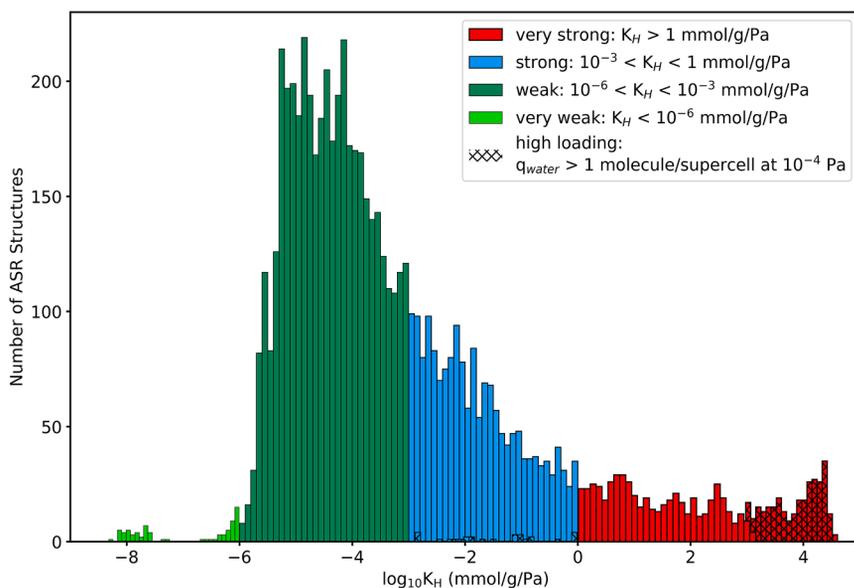


Figure 6. Analysis of water affinity for ASR MOFs by GEMC simulations

Histogram of the logarithm of K_H indicating MOFs with very strong, strong, weak, and very weak affinity for water. q_{water} represents the uptake of water.

pressures (≤ 0.1 Pa to $0.95 P_{sat}$) to assess water affinity. The K_H data from the Widom calculations indicate that the majority of the MOFs have very weak affinity for water (Figure S67) and that the removal of the strongly bound solvent molecules in the ASR set can lead to structures with significantly stronger affinity for water (Figure S68). The GEMC simulations are computationally much more expensive, and an iterative workflow (Figure S69) is used to find a pressure at which the water loading, q_{water} , per supercell (i.e., usually the simulation box contains multiple unit cells) falls into the range of $0.1 \leq q_{water} \leq 1$. Figures 6 and S71 show that the majority of structures exhibit weak interactions with water in Henry's law regime. The majority of structures classified as having strong adsorption sites for water require GEMC simulations at very low pressures to reach Henry's law regime, and the ASR set includes a larger fraction of MOFs with strong adsorption sites. The structures classified as very weak do not have pores of sufficient size to accommodate water molecules. Using the adsorption data from the higher pressures allows us to determine whether pore filling is attained at $0.2 P_{sat}$ and whether pore filling is associated with high capacity or with hydrate formation in defined pockets (i.e., small capacity). Note that we did not take the framework flexibility into account during the calculations, and the distribution could be changed if the framework flexibility is considered. Also, it is known that the presence of defects in MOFs can significantly change their hydrophilicity.⁵⁹ Efforts have been made based on earlier versions of the CoRE MOF DB to generate large collections of structures that include chemically plausible defects,^{9,60} but we have not pursued this direction in the current work. For DAC applications, where feed stream compositions often contain moderate level of humidity, very weak or weak MOFs may be advantageous, as they could avoid pore saturation by water. In contrast, for water-harvesting applications, strong and very strong MOFs may be preferable due to their high water affinity and ability to perform well in low-humidity environments. Nevertheless, iden-

tifying optimal adsorption enthalpy is essential to strike a balance between water uptake and the efficient release of water molecules from the pores across different applications.

Integrated material-process screening for carbon capture

With the increasing usage of natural gas-fired power plants to generate electricity, it is important to develop carbon-capture solutions at lower CO_2 concentrations than 10% CO_2 gas-phase concentration at the feed. Among the MOFs proposed

for carbon-capture applications, CALF-20 has recently emerged as a highly promising candidate⁶¹ due to its excellent CO_2 selectivity, water tolerance, and stability under harsh conditions.⁶¹ Several pilot-plant demonstration sites are now using an adsorption-based CO_2 -capture process, including the rotary adsorption machine (RAM) developed by Svante, which incorporates CALF-20 in its proprietary structured solid sorbent filters.⁶² The success of CALF-20 demonstrates the importance of integrating selectivity and stability metrics into the material evaluation workflow.

To demonstrate the utility of the compiled datasets and ML-predicted properties, we performed an integrated material-process screening for carbon-capture applications with various gas-phase CO_2 concentrations under dry conditions. Firstly, we reduced the total number of candidate materials from 8,854 to 891 by filtering MOFs based on economic feasibility (no expensive nor rare metals; see section [adsorption isotherm calculation](#) and Pre-screening for GCMC simulation), material stability (predicted thermal decomposition temperature above 350°C , which aligns with the average decomposition temperature ($359^\circ\text{C} \pm 87^\circ\text{C}$) of the CoRE MOF 2019 DB,⁴⁵ ensuring retention of thermally stable structures), and hydrophobicity (no OMSs and not a high K_H of water). CALF-20 was selected as a benchmark material for comparison. Note that the predicted solvent removal (activation) stability of the MOF was not used as a pre-filtering criterion due to the low predicted solvent-removal stability of CALF-20, which is 0.4. We focused on the ASR dataset since this dataset is representative of the FSR dataset but includes more structures. Pre-screening for GCMC simulation. We did not consider the Ion dataset.

From 891 candidate materials, we identified 34 candidate MOFs with higher or similar performance-metric values than CALF-20 (Figures 7A and 7B) based on the equilibrium material-based metrics, such as CO_2/N_2 selectivity (>100) and CO_2 working capacity (2 mmol/g). The water-adsorption characteristics at low H_2O relative-humidity conditions of these 35

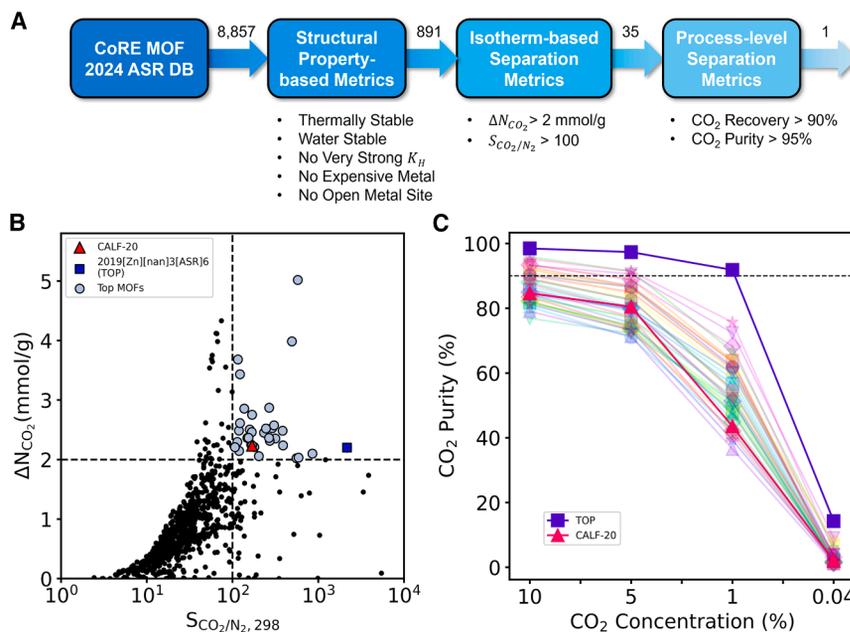


Figure 7. Results of integrated material-process screening of CoRE MOF DB ASR dataset

(A) Integrated material-process screening workflow to identify high-performing MOFs for various carbon-capture scenarios. The numbers over arrows represent the remaining number of structures after completing a step. “No Very Strong K_H ” represents the exclusion of MOFs with K_H greater than 1 mmol/g/Pa by GEMC.

(B) CO_2 working capacity (at adsorption conditions of 298 K and 1–0.1 bar swing) versus CO_2 selectivity (at adsorption conditions of 298 K and 1 bar) for 891 MOFs. Each point corresponds to a single MOF. The points above vertical and horizontal dashed lines indicate selectivity ≥ 100 and working capacity ≥ 2 mmol/g, respectively. The red triangle and blue square represent CALF-20 and the top MOF (CoRE MOF ID, 2019[Zn][nan]3[ASR]6; CoRE MOF REFCODE, CIXDIZ_ASR_pacman) from TSA results.

(C) Rankings of CO_2 purity at different feed compositions. CO_2 purities were obtained from the TSA-cycle optimization. The purity ranking was determined based on CO_2 purity with the constraint of 90% CO_2 recovery. Names of the MOFs are listed in Table S23. The horizontal dashed line represents 90% CO_2 purity.

MOFs were analyzed based on the GEMC simulation data (Figure S72). Among these, structure 33 (CALF-20) demonstrates high hydrophobicity, as it begins to adsorb water at relative humidity above 10%. This is in line with the previous investigation into water adsorption in CALF-20. Only structures 9, 5, and 32 exhibit better hydrophobicity than CALF-20 from the top materials. We also examined the structural characteristics of these 35 top-performing MOFs. 17 out of the 35 MOFs have a PLD smaller than the diameter of N_2 (based on the TraPPE force field used for grand canonical Monte Carlo (GCMC) calculations), and all 35 MOFs have pore volumes below 0.53 cm^3/g . The effect of ML-based partial atomic charge assignment on CO_2 isotherms was analyzed in Figures S74 and S78, with PACMAN-DDEC6 showing excellent agreement with experimental adsorption isotherms and simulated isotherms based on DDEC6 (see Note S8.3).

TSA simulations were carried out to evaluate these MOFs for carbon capture at various CO_2 -capture scenarios (10%, 5%, 1%, and 400 ppm feed concentration of CO_2 ; balance N_2). From our screening, we discovered that 2019[Zn][nan]3[ASR]6 demonstrates promising CO_2 -capture performance, achieving CO_2 purity above 90% at 90% CO_2 recovery under feed conditions of 10%, 5%, and 1% CO_2 . This MOF is composed of tri-triazole ligands and single Zn ions as the nodes and was initially synthesized and tested for its fluorescence properties.⁶³ The material has not been tested for its adsorption properties. In comparison, CALF-20 achieved an 80% CO_2 purity level at 10% and 5% CO_2 concentration at the feed, and <50% purity level at 1% CO_2 concentration at the feed. The top-performing MOF, 2019[Zn][nan]3[ASR]6, has the highest heat of adsorption for CO_2 between 298 and 423 K, although its K_H for H_2O is moderate. The Pareto frontier of CO_2 purity and recovery during full-

scale optimization is shown in Figure S77. Note that, under DAC conditions (0.04% CO_2 feed), even the best-performing 2019[Zn][nan]3[ASR]6 could not reach a CO_2 purity of 20%.

The top material that emerged from the screening is 2019[Zn][nan]3[ASR]6, which has a narrow 1D channel with preferential adsorption sites for CO_2 over H_2O and N_2 . GCMC simulations show extremely low uptakes of N_2 at all pressures and temperatures (Figures 8 and S77), which is a characteristic for high CO_2 purity from process modeling. To explore the impact of solvent removal on the framework flexibility, we recalculated geometric features and single-component isotherms and TSA simulations on the DFT-optimized structure (see Note S10.3, Table S31 shows the comparison of geometric features, TSA inputs, and TSA results of DFT-optimized 2019[Zn][nan]3[ASR]6 with unoptimized 2019[Zn][nan]3[ASR]6; Figure S81 shows the energy histograms for the insertion of a single CO_2 , N_2 , and H_2O molecule into the DFT-optimized structure of 2019[Zn][nan]3[ASR]6). The results indicate that the optimized structure exhibits excellent CO_2 -capture capabilities, with improved CO_2 purity at various CO_2 feed concentrations compared to the unoptimized structure. The rigid optimized-framework’s window diameter from the DFT-optimized structure is 3.2 Å, which lies between the vdW diameters of CO_2 (3.05 Å) and N_2 (3.31 Å), contributing to its excellent separation performance.

We further considered the impact of framework flexibility by calculating pore sizes from the last 5 ps of 10-ps NVT *ab initio* molecular dynamics (AIMD) simulations for two systems: (1) framework with three H_2O molecules, and (2) framework with three CO_2 molecules. Following the AIMD simulations, H_2O and CO_2 molecules were removed, and the pore sizes were calculated based on the trajectory of loaded frameworks. We

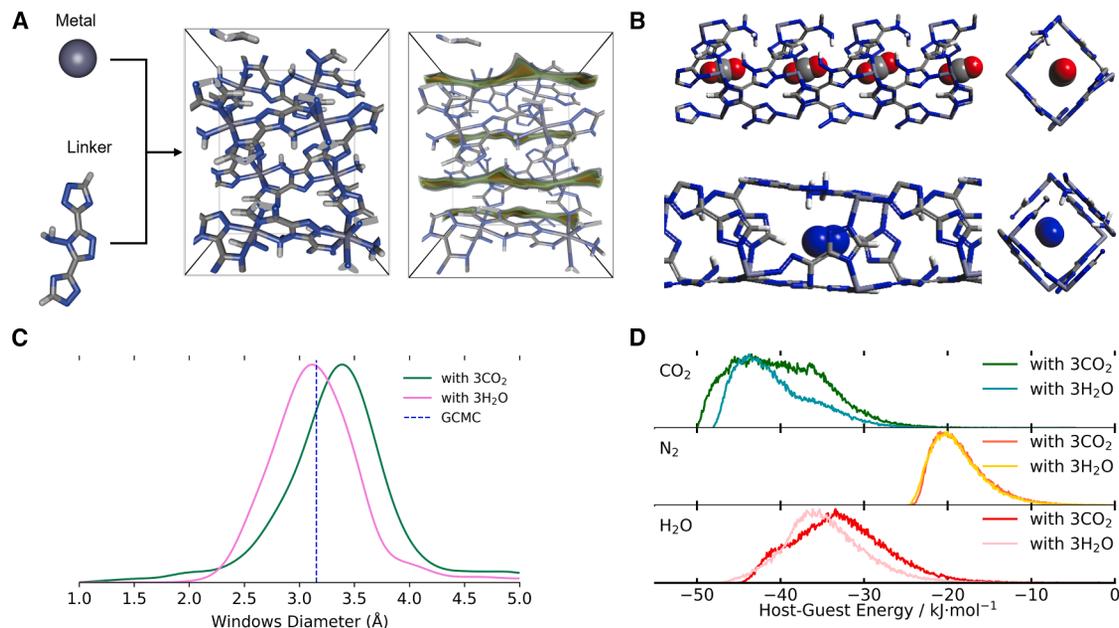


Figure 8. Analysis of the porosity and interaction of 2019[Zn][nan]3[ASR]6 with adsorbates

(A) Visualization of the metal node, linker, structure, and channel of DFT-optimized 2019[Zn][nan]3[ASR]6. The force-field-based nitrogen-energy isocontour levels of DFT-optimized 2019[Zn][nan]3[ASR]6 are 0 kJ/mol (gray), -12.5 kJ/mol (green), and -24.9 kJ/mol (brown).

(B) Snapshots of CO₂ and N₂ in the top MOF taken from different angles at 1 bar and 298 K.

(C) Distribution of pore-window diameters from snapshots of AIMD simulations. The blue dashed line represents the window diameter of the DFT-optimized structure used for GCMC simulations. Since our GCMC simulations assume a rigid structure, the window does not vary.

(D) Energy histogram of force-field energies for inserting a single CO₂, N₂, and H₂O into the snapshots of 2019[Zn][nan]3[ASR]6 from AIMD simulation at 298 K. Adsorbates were removed from the structures prior to calculating the energy histograms with force fields and window diameters.

found that the presence of CO₂ slightly increases the pore size, while the presence of H₂O leads to slight decreases in the pore size (Figure 8C), demonstrating that the pore size of the rigid (DFT-optimized) structure for GCMC simulations is representative of the structure.

The competitive physical adsorption between CO₂ and H₂O, which has been recognized for over a decade,^{48,64,65} was investigated using energy-histogram analysis based on force field combined with AIMD data. The interaction energies from the force field showed CO₂ having a stronger affinity to the framework compared to H₂O and N₂. To probe the accuracy of these force fields, we performed DFT calculations on different orientations of CO₂ and H₂O in the MOF. We observed DFT binding energies of ~30 kJ/mol for CO₂ and 25–51 kJ/mol for H₂O. The higher value of H₂O-binding energy, which was not observed in our force field calculation, likely indicates hydrogen bonding with nitrogen groups in the ligand, which our force-field models do not fully capture. This qualitative difference between DFT results and force fields suggests that a more precise force field might be necessary in the future to make more quantitative predictions about adsorption in this MOF. In contrast to CO₂ and H₂O, the host-guest energy histogram for N₂ essentially remains the same for different sets of structures from AIMD, suggesting that the preferential adsorption sites for N₂ do not change with changing energy landscape within the channel.

We also evaluated CO₂ and H₂O uptakes under ternary mixture conditions (CO₂/N₂/H₂O) at varying humidity levels

(Figure S80) using force-field-based calculations for the original rigid framework. The material maintained high CO₂ uptake even at conditions of 100% relative humidity, which we attribute to the 1D channel preventing water cluster formation. However, even in 1D channels, water adsorption can still occur, which can lead to competitive adsorption with CO₂ and reduced CO₂ uptake. As reported in both simulation and experimental studies,^{66,67} CALF-20 exhibits good CO₂ adsorption performance under low-humidity conditions (<30% RH) but shows a significant decline in CO₂ uptake when the relative humidity exceeds 30%–40%. Therefore, while 1D channels inhibit bulk water cluster formation, they do not prevent all water adsorption effects, which explains the CO₂ uptake reduction at higher humidity levels. Overall, our molecular modeling indicates that the material is a promising candidate for CO₂ capture and warrants further experimental testing and more detailed computational modeling. Note that the predicted solvent-removal stability of the material is 0.35 (see Table S23), indicative of possible framework collapse during the experimental activation process.

Accumulating evidence in the literature shows that equilibrium-based metrics, such as CO₂/N₂ selectivity and CO₂ working capacities, show poor correlation with process-level performance metrics, such as CO₂ purity and recovery.⁵⁸ We analyzed the relationship between separation metrics obtained from molecular-level simulations and CO₂ purity from process-level simulations for the 35 candidate MOFs (Figure 9). For the feed mole fractions of 10:90, 5:95, 1:99, and 0.04, 99.96 CO₂ to N₂,

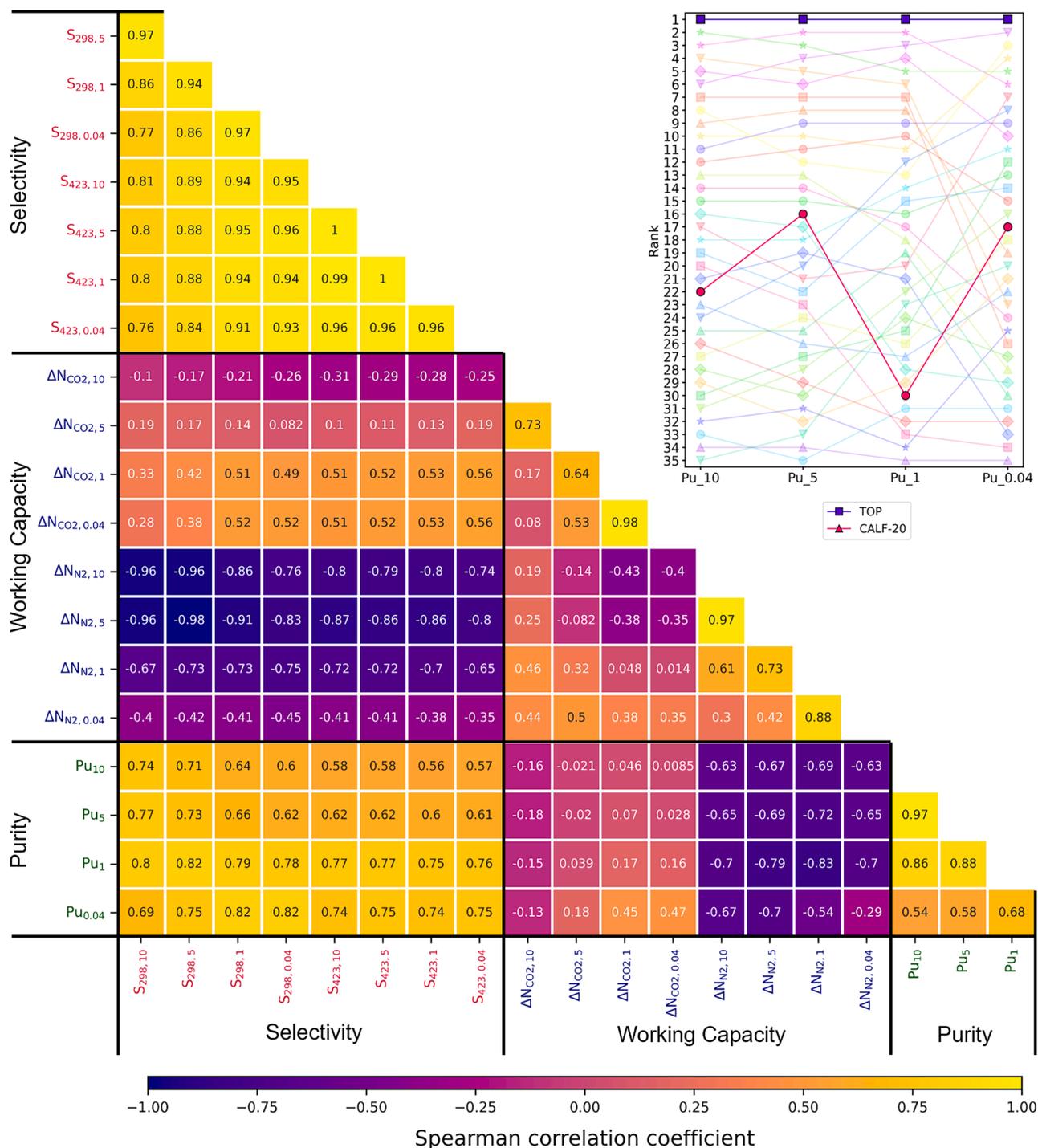


Figure 9. Correlation plots of equilibrium and process performance metrics

Explanation of symbols in the figure. $S_{298,10}$, CO_2/N_2 selectivity at 298 K with CO_2/N_2 mole fraction as 10:90; $\Delta N_{\text{CO}_2,10}$, CO_2 working capacity with CO_2/N_2 mole fraction as 10:90; $\Delta N_{\text{N}_2,10}$, N_2 working capacity with CO_2/N_2 mole fraction as 10:90; Pu_{10} , CO_2 purity with CO_2/N_2 mole fraction as 10:90. All CO_2 purities were obtained from TSA simulation with a 90% CO_2 recovery constraint. The color bar represents the Spearman's rank correlation coefficient (SRCC) values. SRCC quantifies the degree of ranking agreement between properties, where a value of 1 indicates PP correlation and -1 indicates PN correlation. The SRCC values were calculated by Eq. 10-1 in the supplemental information. (Inset) Ranking of the 35 candidate MOFs based on CO_2 purity with different feed compositions.

the Spearman's rank correlation coefficient (SRCC) ranking reveals that CO₂/N₂ selectivity at 298 K (feed temperature) has a moderate correlation with CO₂ purity, which is higher than the selectivity correlation at 423 K (desorption temperature) with SRCC values ranging from 0.82 to 0.60. This relatively poor correlation is in line with previous reports in the literature.^{50,69,70} Among the 35 MOFs, we observed no correlation between CO₂ working capacity and CO₂ purity, while N₂ working capacity exhibits a relatively strong correlation with CO₂ purity. Since CO₂ purity is impacted by the presence of other gas species, such as N₂, a moderate to strong correlation between the N₂ working capacity and CO₂ purity is expected. The correlations between TSA simulation results at different feed ratios range from 0.97 to 0.54 between CO₂ purity at 10% feed concentration and CO₂ purity at feed concentrations of 5%, 1%, and 400 ppm. This is consistent with the ranking data for selectivity, suggesting that, when there are significant changes in CO₂ concentration, it is necessary to run relevant simulations. It should be noted that the results of this work are evaluated based on the performance of top candidate MOFs rather than the entire dataset. Finally, we present the ligand costs of the top-performing MOFs and CALF-20 in Table S30. Figure S82 demonstrates how to obtain detailed information for a single structure (CALF-20).

DISCUSSION

We report an updated CoRE MOF DB, which includes over 40,000 experimentally reported MOF crystal structures (17,202 CR and 23,635 NCR). The updated dataset includes many properties useful for future high-throughput computational screening and ML model development for various MOF-discovery activities. Potential future applications of the dataset include leveraging the provided DOI for each structure to retrieve details on synthesis conditions and experimental data to train a large language model (LLM) for synthesis predictions and optimizations.^{71,72} More importantly, highly streamlined curation procedures will provide more timely access to the latest curated CR experimental MOF datasets with a planned annual update. Finally, we demonstrated the utility of the compiled dataset for carbon-capture applications and identified 2019[Zn][nan]3 [ASR]6, which shows near-complete rejection of N₂ in the temperature range of 298–423 K while showing excellent CO₂ uptakes, leading to 90% purity and recovery even at 1% CO₂ feed conditions. The exercise demonstrated in this work provides a robust starting point for subsequent targeted MOF exploration for gas separation applications and beyond.

METHODS

Data collection

For newly synthesized MOFs, it is common practice for authors to upload a CIF as part of the SI of the paper or to the CSD as mandated by a few publishers (RSC, AAAS). In this work, we opted to download the CIFs from the SI manually if the CIFs were available as part of the SI. Otherwise, we obtained the CIFs from the CSD using the Python application-programming interface (API) provided by CCDC. We collected 4,828 CIFs from SIs and 19,497 CIFs from the CSD. We make the modified

structures derived from the CSD publicly available, while a list will be provided for unmodified structures sourced directly from the CSD.

Crystal structure cleaning and curation

Based on our previous study,^{22,23} we adopted similar cleaning and curation algorithms for database construction with crystal graphs constructed from adjacency matrices based on the covalent radii⁷³ as in Figure S3. The structure cleaning procedure then continues with two steps: (1) FSR and (2) ASR. The free-solvent-removal step removes solvent molecules, i.e., molecules that are not connected to the largest connected crystal. We define the adjacency matrix using a default skin distance of 0.25 Å and remove portions of the crystal structures that are not part of the largest connected graph. Here, two atoms are considered bonded if the distance between them is less than the sum of their covalent radii plus the default skin distance (0.25 Å). The resulting structures are labeled with “FSR” in their CoRE MOF ID. During the removal process, we checked if the solvents were ions by checking their stoichiometry against a pre-tabulated list of ions (Table S1) obtained from the CoRE MOF 2019 v1 DB and labeled any structures with ion solvent with an Ion flag in their CoRE MOF ID. A specific example of this implementation can be seen in Figure S8B, where the structure has both ammonium (NH₄⁺) and water molecules in the pores. By checking the stoichiometry of the independent solvents, we only removed water molecules while keeping the NH₄⁺ ions in the structures. Note that a structure flagged with Ion is charge-neutral.

The second step (all solvent removal [ASR]) involves removal of solvents coordinated to the metal center in the framework. During the construction of the previous version of the CoRE MOF DBs, we used a default skin distance of 0.3 Å to construct the adjacency matrix of a structure. This approach occasionally led to the unintended removal of metal atoms from some structures as reported by others in the literature.^{16,18–21} To address this issue, we improved our coordinated solvent removal procedure with variable skin distance. Since different bonds have different lengths, we tested the cleaning results with different skin values for some cases in Figure S4. The new cleaning process starts with a skin distance of 0.25 Å, and monitors whether the removed atoms are metal. If metal atoms are identified among the atoms classified as solvent, the skin distance is increased by 0.05 Å and a new adjacency matrix is calculated. This procedure is repeated until the identified solvent molecules no longer include metal atoms. The new algorithm may still lead to residual solvents (e.g., H₂O or OH⁻) in a limited number of structures. However, this is preferable to the unintended removal of metal nodes from a structure. If structure files were modified during this step, we labeled the structures with an “ASR” flag in their CoRE MOF ID. Note that the structures with ions do not go through the coordinated solvent removal step, as such steps may create a charge imbalance in the structure. We also replaced the structures that had been modified manually in the CoRE MOF 2019 v1 DB (340 structures) and from the literature⁷⁴ (432 structures) to the database. We additionally manually corrected 65 structure files (62 structures needed charge-balancing ions, two structures were missing hydrogen atoms, and one structure had the wrong number of protons), and these

structures replaced the structures that went through the cleaning procedure. At the end of the structure-cleaning step, we had 20,552 ASR, 37,725 FSR, and 4,244 Ion CIF files for subsequent analysis.

Classification of NCR structures

Although the previous CoRE MOF DBs were meant to provide CR structures that could be used with some confidence for high-throughput computational screening studies, we have found that some of the structures were, although technically speaking CR, not fit for the intended purpose; that is, these structures contained atoms in non-viable positions or in unexpected oxidation states. These deficiencies sometimes resulted in structures falsely identified as high performing for a given application. In this work, we have developed new procedures to improve the fraction of structures that are CR. This was done in two ways: (1) improving the solvent-removal algorithm with variable skin distances to prevent the algorithm from removing metal clusters, and (2) introducing better schemes for classifying whether our procedures would be able to clean a given structure and, if not, simply removing such structures and classifying them as NCR. Previously, we denoted such structures with a “disorder” flag, but, since this terminology has specific meaning in crystallography, we opted to create a new terminology that is easier to understand. To achieve this goal, we ran two automated methods reported in the literature to detect structures. The first method was reported by Chen and Manz in 2020,¹⁷ and the second method is based on a Python program (MOFChecker³²) developed by Kevin M. Jablonka. The method by Chen and Manz considers four criteria, while MOFChecker considers 12 criteria when assigning MOFs to be NCR (Tables S2 and S3). There are four shared criteria between the methods, namely atom overlap, under coordinated carbon, over coordinated carbon, and lone molecules (or isolated atoms). We outline the implementation details of the two methods below. Note that the NCR structures are provided as a separate part of the database.

(1) Chen and Manz method

Chen and Manz reported a classification scheme to detect disorder types in CoRE MOF 2019, and full details can be found in that work.¹⁷ As discussed above, “disorder” refers to a specific type of NCR structure that arises from crystallographic disorder, such as partial occupancies of atoms. The Chen and Manz method considers four criteria for classifying an MOF as disordered (Table S2). Briefly, an atom is considered isolated if it is not connected to any other atom, and atoms are considered overlapping if the distance between them is less than half the sum of their radii. Lastly, the bond order is estimated for each bond including a carbon atom using an empirical distance-based calculation described by Chen and Manz as follows:

$$\log_{10}(BO) = A \times d + C \quad (\text{Equation 1})$$

where BO is the bond order, $A (\text{\AA}^{-1})$ is the slope of bond order with bond length, d is the distance between two atoms, and C is a constant. A carbon atom is considered under-bonded if

the sum of its bond orders is less than 3.3 and over-bonded if the sum is greater than or equal to 5.5. The atom typing radii (ATRs) used in the above analyses were provided by Chen and Manz in a previous study⁷⁵ and are also shown in Figure S13. Note that these effective metal radii are generally smaller than the corresponding covalent radii used in our solvent-removal algorithm (Figure S3) due to positive atomic charges on MOF metal nodes.

MOFs were classified as having unphysical disorder if they included isolated atoms, overlapping atoms, and under- and over-bonded carbon atoms. Any structures containing at least one type of disorder except isolated atoms were moved from the CR dataset to the NCR dataset. We flagged but did not remove MOFs containing isolated atoms that were otherwise physical because isolated atoms may be caused by the solvent-removal procedures outlined above (section [crystal structure cleaning and curation](#)).

(2) MOFChecker method

We used a Python program developed by Kevin M. Jablonka called MOFChecker 0.9.6 (<https://github.com/kjappelbaum/mofchecker>) to identify NCR structures. We selected 12 criteria to filter structures, which are listed in Table S3. The radii used in the method are based on the covalent radii reported by the CCDC,⁷³ and the details are available on the website above. However, we briefly explain these criteria for completeness in section S2.2 and the supplemental information. Figures S14–S18 show some disorder cases identified from the MOFChecker program.

(3) Removing MOFs from retracted papers, simulations, and MOSAEC

To ensure that all structures originated from experimental synthesis, we examined papers that reported more than two crystal structures to remove any structures from simulation. We removed DFT-optimized, first-principles-predicted, and hypothetical MOFs from the database, including 130 ASR and 131 FSR structures. Regarding retracted papers flagged by the CSD, we removed 26 ASR, 25 FSR, and 30 NCR structures based on the DOI list provided in the CSD report.³⁰ Based on the report from MOSAEC,²⁴ we removed one NCR structure (REFCODE, FIBKEI_ASR_pacman; CoRE MOF ID, 2013[Cu][nan]3[ASR]5). Note that we kept the structures that were curated from literature, which may involve some level of structure relaxation.

OMs

We employed two approaches based on the geometry of the atoms to identify the presence of OMs within the structures. The first method is based on a binary decision criterion that determines whether a metal site is open or closed. The second method quantifies the deviation of the first coordination sphere of a metal site from an ideal closed configuration by measuring its geometrical distortion. The details can be found in Note S4.4.

Duplicate detection

We assessed the structures for instances where the volume, chemical composition, lattice constants, and atom positions,

or their textural properties, were identical. This occurrence was attributed to authors submitting the same structures to paper SIs as well as to the CSD. In cases where such duplications were identified, we opted to randomly retain only one structure from each set of duplicates, as they are structurally identical and interchangeable. Retaining any one of them ensures consistency without altering dataset composition. We have found that some structures were not detected as duplicates before solvent removal but were identified as the same structure after removing the solvents. This is because the solvent molecules are disordered, and the disorder varies among them, yet, in fact, structures with these solvents are the same. The matching method between structures is the same as in [Note S1.8](#).

Topology analysis

The underlying connectivity between organic ligands and metal nodes in MOFs can be described mathematically by periodic graphs⁷⁶ that are often referred as crystal nets^{77,78} or the topology of the crystal structures.^{79,80} We used CrystalNets.jl 0.5.0 (<https://github.com/coudertlab/CrystalNets.jl>)⁴⁰ to identify the topology of the structures in the dataset/new database. The identification of underlying topologies of MOFs sometimes is not unique,^{40,80,81} and we choose to report the topological results obtained by the “single nodes” algorithm in [Figure 3](#) for ASR structures. There are three sources for the referral names of the net topologies: (1) the RCSR database,⁸² (2) the IZA-SC database,^{83,84} and (3) the EPINET (the codes mixing alphabetic letters and numbers).⁸⁵ We denote net topologies that are not recorded in any of the above databases as “unnamed” topologies analyzed in detail in [Note S4.6](#). In the analysis in the SI, interpenetrating structures with heterogeneous topologies, such as UNAZIT, were discarded (about 22% of the database). Overall, slightly over 60% of structures have a known topology identified by CrystalNets.jl.

Textural properties

We conducted textural property calculations using the Zeo++ 0.3 software^{86–90} utilizing nitrogen molecules as probe molecules (diameter of 3.31 Å) for the calculation of accessible surface area (ASA). The density, framework dimensionality, pore-size parameters (PLD, LCD), surface area, and porosity parameters (pore volume and porosity) of the structures were also calculated. Structures with a dimensionality of 0 or 1 (non-MOF) and PLD < 2.4 Å (non-porous material) were excluded. All calculations are performed with a high-accuracy flag in Zeo++ (“-ha”). Note that the default settings for the high-accuracy mode failed for a handful of structures due to numeric floating issues and can be resolved by specifying an alternative flag option (e.g., “-ha S50”).

Diversity metrics

We sought to compare the chemical space spanned by CoRE MOF 2024 relative to CoRE MOF 2019 v2 ASR and hypothetical MOFs. To represent the space of hypothetical MOFs, we considered 12 hypothetical datasets curated by Boyd et al.,⁹¹ Lan et al.,⁹² Colón et al.,⁹³ Anderson et al.,⁹⁴ Gómez-Gualdrón et al.,⁹⁵ Chung et al.,⁹⁶ Li et al.,⁹⁷ Majumdar et al.,⁹⁸ Anderson and Gómez-Gualdrón,⁹⁹ Anderson et al.,¹⁰⁰ Bao et al.,¹⁰¹ and

Burner et al.¹⁰² To quantify the diversity of the datasets, we employed three diversity metrics: balance, variety, and disparity. After using *k*-means clustering to divide MOFs into bins based on feature similarity, we use relative entropy to quantify balance, and we define variety as the number of *k*-means bins occupied by a dataset over the total number of bins for the full space (i.e., the union of datasets; [Note S4.7.3](#)). To calculate disparity, we take a ratio of concave hulls occupied by datasets on a 2D projection, for which we use principal-component analysis (PCA). Both variety and disparity measure the fraction of chemical space covered by a dataset. The three metrics have previously been used to quantify the diversity of MOF datasets and indicate the evenness or the spread of data, albeit sometimes with different 2D projection methods for disparity calculation such as UMAP.^{46,102} Furthermore, we visualized the chemical space of CoRE MOF 2024, CoRE MOF 2019, and the hypothetical MOFs using PCA on MOF descriptors encoding information about metal, linker, and functional-group chemical environments (the detail can be found in [Tables S8–S11](#)), as well as descriptors conveying geometric information ([Figure S47](#)). Diversity metrics are reported in [Table S12](#).

MOFid-v2

We computed the MOFid-v1 identifiers using the original MOFid code³⁸ for all 8,854 ASR, 7,635 FSR, and 713 Ion MOF structures in CoRE MOF 2024. While MOFid has become a useful cheminformatics tool for analyzing MOFs, it exhibits certain limitations. First, it simplifies metal nodes to their bare metal cores, predominantly identifying only the element(s) of the metals present. This simplification fails to adequately capture the detailed chemistry, composition, and atomic arrangement of the metal nodes. Additionally, MOFid uses Simplified Molecular Input Line Entry System (SMILES) notation^{103,104} to represent the organic linkers, which, despite its utility and widespread usage, has some drawbacks. For example, not all symbol combinations in SMILES represent valid chemical structures, which constrains the utility of MOFid in ML, especially for generative ML applications.^{104–106} To address these limitations, we have developed an updated version, MOFid-v2. MOFid-v2 code is a postprocessing workflow that utilizes input data from MOFid-v1 code. [Figure S48](#) shows a comparison of MOFid-v1 and MOFid-v2 for the example of IRMOF-1. There are two major differences (the format of MOFid-v1 and MOFid-v2 as shown in [Table S15](#)). First, the metal node is represented as a categorical variable that denotes both its chemical composition and geometric configuration. Thus, the metal node representation in MOFid-v1, “[Zn][O]([Zn])([Zn])[Zn]” is replaced with “[C6O13Zn4_Type-1]” in MOFid-v2, where the “type” is from a lookup JSON file (refcode_groupby_node.json) for the categorical representations of metal nodes. The lookup JSON file provides sample coordinates from one of the MOFs containing each metal node composition and type. This allows, for example, researchers to easily find MOFs with similar metal nodes to IRMOF-1. The second change in MOFid-v2 is that the SMILES representation of the organic linker is replaced with a SELFIES string.¹⁰⁵ The GitHub repository for the MOFid code (<https://github.com/snurr-group/mofid>) contains utilities for not only MOFid-v1 but also MOFid-v2. Details of the MOFid-v2 construction are provided in [Note S5](#)

(Figures S52 and S53 present an upset plot of the MOFid outputs on the CoRE MOF ASR and Ion datasets; Figure S54 shows the detailed workflow of the MOFid algorithm; Figure S55 illustrates the different classifications of metal nodes with identical chemical formulas).

Nomenclature

The CoRE MOF DB pulls structures from the SI files of research papers and the commercial CSD. Naming MOFs based on CSD REFCODE or the DOI poses challenges because such a name does not contain useful information. To address this, we developed a new nomenclature for the structures found in the CoRE MOF DB. We define structure names as, for example, “1993[CuCo][tfk]3[ASR]1.cif.” This naming convention is composed of the year when the structure was reported (“1993”), the metal types (“[CuCo]”), topology (“[tfk]”), dimensionality (“3”), crystal type (“[ASR]”), and a unique identifier (to prevent duplicate names). For structures with an unknown year of deposition (commonly found in the CSD dataset) we use “0000” for the year. We explain the details and give examples in Figure S56. Publication dates are obtained through the Crossref API based on DOIs, where a small number of DOIs may have dates that lag behind the actual publication dates. This delay is due to variations in license timing across different publishers. The “license date” in Crossref metadata indicates the date on which the paper was officially authorized for publication under specific licensing terms.

ML-predicted DDEC6 partial atomic charges

We assigned DDEC6 partial atomic charges predicted by the PACMAN method.⁴⁹ For the CoRE MOF 2024 DB, we used the PACMAN charge v1.1 model for DDEC6 with 10 decimal places to estimate partial atomic charges of MOF structures. The method is based on a crystal graph convolution neural network trained on the QMOF database and shows excellent performance across the periodic table. The same partial atomic charges are assigned for the same atom types. The code can be found on GitHub (<https://github.com/mtap-research/PACMAN-charge>).

Thermal, solvent removal, and water stabilities

We applied ANNs developed on a subset of CoRE MOF 2019 v1 ASR^{53,107} in prior work to predict the thermal and solvent-removal stability of MOFs in CoRE MOF 2024. We also applied random-forest models to predict two-class water stability,⁵⁶ though we note alternative models predictive of MOF-water stability exist.^{108,109} The models we apply use revised autocorrelation (RAC)¹¹⁰ and geometric features of MOFs as inputs, which are generated from CIF structure files by molSimplify 1.7.3^{46,111} and Zeo++ 0.3,⁸⁶ respectively (Note S4.7.1). The models were trained on experimental data text mined from literature in the case of solvent removal and water stability, and they were trained on decomposition temperatures extracted from digitized thermogravimetric analysis (TGA) curves for thermal stability. The thermal-stability model predicts MOF decomposition temperatures through regression, while the solvent-removal (activation) and water-stability models perform binary classification, outputting values between 0 (unstable) and 1 (stable). In classification, confidence increases as the predicted value moves further from 0.5, but this

uncertainty quantification is unavailable in regression tasks. The accuracy of the model increases when only high-confidence data points (>0.7) are considered, demonstrating that the model reliably predicts stability metrics when trained on high-quality datasets. The models can be found on GitHub (<https://github.com/hjkgrp/MOFSimplify/tree/main/model>). The thermal and solvent-removal stability models permit uncertainty quantification through latent space distance and latent space entropy, respectively.⁵³ These metrics measure the similarity of a new MOF to model training data in the artificial neural network latent space, i.e., the learned representations by the model, and have been shown to correlate with prediction accuracy.¹¹²

Heat capacities

The heat capacities of MOFs in the ASR, FSR, and Ion datasets were directly predicted by XGBoost models on CIFs; these models were reported by Moosavi et al. (<https://github.com/SeyedMohamadMoosavi/tools-cp-porousmat>).⁵⁴ The models were trained from a diverse dataset containing 232 structures. These structures were sampled from experimental structures in the CoRE MOF 2019 v1 DB,²³ experimental covalent-organic frameworks (COFs) from the Clean, Uniform, and Refined with Automatic Tracking from Experimental Database COF (CURATED-COF) database,¹¹³ and experimental all-silica zeolites from the International Zeolite Association (IZA) database.⁵⁵ 120 structures were used for training and 112 structures were used for testing. The temperatures associated with the heat capacity in this dataset are 300, 350, and 400 K. The boosting ensemble mechanism was applied to quantify model uncertainty. An ensemble of 100 gradient-boosted decision tree models was trained. The standard deviation of the prediction was used as a metric of uncertainty. The total performance of trained models on the test dataset shows a mean absolute error (MAE) of $20 \text{ J kg}^{-1} \text{ K}^{-1}$. For erroneous predictions, the high uncertainty (high standard deviation) of predictions is signed, and the purple square markers in Figure S62 are uncertain predictions of the test dataset with standard deviations of higher than $100 \text{ J kg}^{-1} \text{ K}^{-1}$. Figure S64 displays the uncertainty from the trained models as a function of the average atomic mass from the trained models of an MOF. The overall uncertainty in the predictions is in the range of $0\text{--}130 \text{ J kg}^{-1} \text{ K}^{-1}$ at 300 K. By setting an error threshold of $70 \text{ J kg}^{-1} \text{ K}^{-1}$, we find that only 44 ASR, 45 FSR, and four Ion MOFs exceed this threshold at 300 K. The low number of uncertain predictions instills confidence in the predicted heat-capacity values. The large uncertainty in predictions for the 93 MOFs indicates that the chemical environments of these MOF structures differ significantly from those in the training set. These MOF structures can help improve the model performance through incorporation of the frameworks and their calculated heat capacities into the training dataset in further research. We performed linear fitting of the heat-capacity data to extend it to different temperatures (see Figure S66).

Classification of hydrophilicity and hydrophobicity

We computed the water Henry's law constant K_H of all CR MOFs at 298 K by using the Widom particle insertion method with 10^5 Monte Carlo cycles to categorize hydrophilicity and hydrophobicity for MOFs.¹¹⁴ For the Widom particle-insertion calculation,

we truncated the Lennard-Jones (LJ) interactions at 14 Å with tail corrections. The framework LJ parameters were taken from the DREIDING force field (DFF)¹¹⁵ and universal force field (UFF).¹¹⁶ We used the TIP4P force field model¹¹⁷ for H₂O.

The nonbonded interactions between the adsorbate and adsorbent and the adsorbate and adsorbate were defined by the LJ 12-6 potential¹¹⁸ model plus the coulomb potential model as follows:

$$U_{ij} = U_{LJ} + U_{Coulomb} = 4\epsilon_{ij} \left[\left(\frac{\sigma}{r_{ij}} \right)^{12} - \left(\frac{\sigma}{r_{ij}} \right)^6 \right] + \frac{1}{4\pi\epsilon_0} \frac{q_i q_j}{r_{ij}} \quad (\text{Equation 2})$$

where U and r are the interaction energy and the distance between two atoms, and ϵ and σ are the LJ well depth and diameter. q represents partial atomic charge, and the atomic charges of the framework were derived from PACMAN (DDEC6 model). Electrostatic interactions between MOF structures and adsorbed molecules were assessed using the Ewald summation method with a precision of 10^{-5} . ϵ_0 is the vacuum permittivity constant. Interaction parameters for atom i and atom j were approximated by the Lorentz-Berthelot (LB) mixing rules,¹¹⁹ as follows:

$$\epsilon_{ij} = \sqrt{\epsilon_i \epsilon_j} \quad (\text{Equation 3})$$

$$\sigma_{ij} = (\sigma_i + \sigma_j)/2 \quad (\text{Equation 4})$$

K_H and the heat of adsorption (ΔH) can be determined from the following equations:

$$K_H = \frac{1}{RT\rho_f} \frac{\langle W \rangle}{\langle W^{IG} \rangle} \quad (\text{Equation 5})$$

$$\Delta H = \Delta U - RT = \langle U_{hg} \rangle - \langle U_h \rangle - \langle U_g \rangle - RT \quad (\text{Equation 6})$$

where ρ_f is the density of the framework, $\langle W \rangle$ is the Rosenbluth weight, and $\langle W^{IG} \rangle$ is the Rosenbluth weight of a single adsorbate in the ideal gas phase (1 used for this study). The internal energy of the system (ΔU) consists of the average potential energy of the guest molecule inside the host framework ($\langle U_{hg} \rangle$), the host-framework energy ($\langle U_h \rangle$) (for a rigid framework, the value is zero), and the guest-molecule energy ($\langle U_g \rangle$) (for a rigid molecule, the value is zero). R and T are the universal gas constant ($8.314, m^3 \cdot Pa \cdot K^{-1} \cdot mol^{-1}$) and the temperature of the system (K). The simulation details (force-field parameters and critical constants) are provided in [Note S8.1](#).

We computed water adsorption isotherms using isobaric-isothermal (NpT) GEMC simulations^{120,121} with the Monte Carlo for Complex Chemical Systems-Minnesota (MCCCS-MN) software.¹²² We used the TIP4P model¹¹⁷ to represent the intermolecular interaction of the rigid water molecules and UFF4MOF¹²³ for MOFs. The LB mixing rules¹²⁴ described the MOF-water LJ interactions. The charges for MOF atoms were provided by the

PACMAN⁴⁹ 1.1 Python package in the MOFs' CIF file. The long-range Coulombic interactions were implemented using Ewald summation,¹²⁵ and analytical tail corrections¹²⁵ were used for long-range LJ interactions. The pairwise LJ and coulomb interactions were truncated at a spherical cutoff of $r_{cut} = 9$ Å for the sorbent and 40% of the box length for the vapor phase. The size of the MOF supercell was based on satisfying the condition that the supercell is sufficiently large to fully enclose a sphere with a radius larger than r_{cut} , and MOF structures were kept rigid for the simulations. We used the LAMMPS interface 0.2.1¹²⁶ Python package to get the UFF4MOF potential parameters for the MCCCS-MN input files. The various MC moves employed in our water-adsorption simulations are translation, rotation, adsorbate transfer¹²⁷ between the sorbent and vapor phase, and volume moves on the vapor box. To capture the K_H values for the entire CoRE MOF DB, we conducted the NpT -GEMC water adsorption simulations for a large pressure (p) range of 10^{-1} to 10^3 Pa. The water vapor box was pre-equilibrated for 5×10^3 cycles without swap moves to reach the target pressure. The two-box system was equilibrated for 1.5×10^4 cycles, followed by 2.5×10^4 cycles of production run. The details and parameters used in this work are provided in [Note S7.2](#). [Figure S69](#) shows the flowchart to get K_H using the high-throughput water adsorption simulations in the 10^{-1} - to 10^3 -Pa range. For computationally efficient simulation, the number of water molecules in the vapor phase was set to four times the number of unit cells in the MOF supercell, as we target that, in the linear pressure regime used to obtain K_H , each supercell should have less than one water molecule. We performed water adsorption using NpT -GEMC for $\sim 17,000$ MOFs and obtained the water loading (q_{water}) over the considered pressure range. A simulation is regarded as complete if the production run is finished without runtime errors. From the completed simulations, we utilized q_{water} in the range of 0.1–1 molecules/supercell to calculate the K_H for an MOF (labeled K_H). However, if $q_{water} < 0.1$ at the highest pressure (10^3 Pa), we performed GEMC simulation at 0.95 of the saturation pressure, p_{sat} . Additionally, if $q_{water} > 1$ at the lowest pressure (10^{-1} Pa), we conducted GEMC at lower pressures to get a reliable K_H . The pie chart in [Figure S70](#) illustrates the distribution of GEMC simulation statuses and K_H calculations for MOFs in the CoRE MOF ASR and Ion datasets.

Adsorption isotherm calculation

We removed the structures with expensive metal atoms (Au, Ag, Dy, Eu, Ga, Gd, Hf, In, Ir, La, Mo, Nd, Pd, Pr, Pt, Rh, Ru, Se, Sm, Tb, Te, Tm, U, and Y) and rare metals ([Figure S73](#), Ag, Au, Be, Bi, Cs, Dy, Er, Eu, Gd, Hf, Hg, Ho, In, Ir, La, Lu, Nb, Nd, Os, Pr, Re, Rh, Sb, Sm, Ta, Tb, Th, Ti, Tm, U, W, and Yb) to improve economic viability. Generic force fields cannot accurately describe the OMSs in MOFs, which typically require specific, proprietary force-field parameters. Therefore, we do not consider MOF structures with OMSs. We removed the MOFs with poor water and thermal stability (as predicted by the ML models) based on results in Thermal, solvent removal, and water stability prediction.

This first step of screening left us with 891 structures based on pre-screening. We ran mixture GCMC simulations for CO₂/N₂

with a gas-phase mole ratio of 1:9 at 1 bar and at different temperatures (298 and 423 K).

The CO₂/N₂ selectivity (S_{CO_2/N_2}) and CO₂ working capacity (ΔN_{CO_2}) were defined as follows:

$$S_{CO_2/N_2} = \frac{N_{CO_2,298}}{N_{N_2,298}} \frac{y_{N_2}}{y_{CO_2}} \quad (\text{Equation 7})$$

$$\Delta N_{CO_2} = N_{CO_2,298} - N_{CO_2,423} \quad (\text{Equation 8})$$

where N and y are the uptake and mole fraction.

We also calculated Henry's selectivity based on Henry's coefficient at 298 K from Widom insertion calculation.

$$S_{Henry,CO_2/N_2} = \frac{K_{CO_2,298}}{K_{N_2,298}} \quad (\text{Equation 9})$$

N₂ and CO₂ were modeled by the 3-site TraPPE model.¹²⁸ The force-field parameters and critical constants used in this work are shown in Tables S20–S22. We used tail corrections for adsorbate-adsorbate interactions with a cutoff of 14 Å and no tail corrections for the adsorbate-framework interactions. We used 10,000 and 20,000 cycles for initializing the atomic positions of the system and collecting the statistical average of the system, respectively. RASPA 2.0.50 (<https://github.com/IRASPA/RASPA2>) was used to calculate single-component and binary-component adsorption of CO₂ and N₂, as well as Widom's particle insertion calculations.^{128–130}

We also calculated the several mixture isotherms of H₂O/CO₂/N₂ with a gas-phase CO₂/N₂ mole ratio of 1:9 at 1 bar and 298 K with different relative humidities (5%–100%) for the top candidate, 2019[Zn][nan]3[ASR]6. We set the saturation loading of H₂O as 4.1 kPa at 298 K based on the TIP4P model (the mole fraction of H₂O/CO₂/N₂ can be found in Table S32). For ternary isotherm calculations, we used 10⁷ cycles for initialization and 10⁷ cycles for production to collect the statistical average of the system. For ternary mixture calculations, we used gRASPA (<https://github.com/snurr-group/gRASPA>) v-093024.¹³¹ Table S19 lists the software used in this study for calculating uptake and thermodynamic parameters.

Adsorption equilibrium model

The extended dual-site Langmuir (EDSL) model was employed to describe the competitive CO₂ and N₂ isotherms:

$$q_i^* = \frac{M_i^1 \cdot B_i^1 \cdot P_i}{1 + B_{CO_2}^1 \cdot P_{CO_2} + B_{N_2}^1 \cdot P_{N_2}} + \frac{M_i^2 \cdot B_i^2 \cdot P_i}{1 + B_{CO_2}^2 \cdot P_{CO_2} + B_{N_2}^2 \cdot P_{N_2}} \quad (i = CO_2 \text{ or } N_2) \quad (\text{Equation 10})$$

$$B_i^j = b_i^j \cdot \exp\left[\frac{-\Delta U_i^j}{RT}\right] \quad (j = 1 \text{ or } 2) \quad (\text{Equation 11})$$

where q_i^* represents the equilibrium loading (mmol/g) of component i on the solid phase, P_i denotes the gas-phase partial pres-

sure (Pa) of component i , M_i^1 and M_i^2 represent the solid phase saturation loadings (mmol/g) of component i at sites 1 and 2, ΔU_i^1 and ΔU_i^2 are the internal energies (J/mol) of adsorption of component i at site 1 and 2, b_i^1 and b_i^2 represent the isotherm parameter (1/Pa) of component i at site 1 and 2, T is the gas-phase temperature (K), and R is the gas constant (J mol⁻¹ K⁻¹).

There are several methods to estimate the valid isotherm parameters that accurately predict the CO₂ and N₂ competitive isotherms, such as the perfect-positive (PP) scheme, perfect-negative (PN) scheme, and equal-energy-site (EES) scheme.^{132,133} Among these methods, we employed the EES scheme, which accurately predicted the experimental CO₂ and N₂ competitive adsorption isotherms on CALF-20.¹³² In the procedure for estimating the isotherm parameters using the EES scheme, the single-component CO₂ isotherms are fitted first to the dual-site Langmuir (DSL) model:

$$q_i^* = \frac{M_i^1 \cdot B_i^1 \cdot P_i}{1 + B_i^1 \cdot P_i} + \frac{M_i^2 \cdot B_i^2 \cdot P_i}{1 + B_i^2 \cdot P_i} \quad (i = CO_2 \text{ or } N_2) \quad (\text{Equation 12})$$

In the EES approach, N₂ is distributed to two equal-energy sites (sites 1 and 2), and the solid-phase saturation loading at each site is the same as that of CO₂. Thus, the single-component N₂ isotherms are fitted to the DSL model with the following constraints:

$$M_{CO_2}^j = M_{N_2}^j \quad (j = 1 \text{ or } 2) \quad (\text{Equation 13})$$

$$b_{N_2}^1 = b_{N_2}^2 \quad (\text{Equation 14})$$

$$\Delta U_{N_2}^1 = \Delta U_{N_2}^2 \quad (\text{Equation 15})$$

The fitted DSL parameters are listed in Tables S24 and S25. The fitted R² and root-mean-square error (RMSE), the fitted isotherms, and visualizations of the structures are shown in Figure S75 for the 35 candidate MOFs.

TSA cycle

We considered a six-step TSA cycle (Figure S76), which is based on a previously reported five-step TSA cycle.^{134,135} In our study, we found a pressure drop inside the column after the closed cooling step, which would lead to backflow if directly followed by the adsorption step. To address this, we added a pressurization step after closed cooling, resulting in a six-step TSA cycle. Given the low CO₂ feed concentration of less than 10% in our study, pressure-swing adsorption (PSA) would require either a high adsorption pressure or low desorption pressure to meet the CO₂ purity and recovery requirements. PSA uses pumps or compressors that convert electricity to heat, which introduces significant energy disadvantages, such as conversion loss. Therefore, we considered TSA, which offers energy advantages by directly using heat.

The column was initially saturated with pure N₂ at the feed pressure and temperature. The individual steps of the cycle are briefly explained below.

- (1) Adsorption (Ads): both ends of the column are open, allowing the feed gas to enter the column, where CO₂ is

adsorbed, and the light product (N_2) exits the column as effluent.

- (2) Heavy reflux (HR): the feed-gas flows stop, and the heavy product (CO_2) from the light reflux is introduced into the column. This heavy product has a higher CO_2 concentration than the feed, increasing the CO_2 concentration at the column inlet and enhancing CO_2 purity and recovery.
- (3) Heating (heat): the end of the column is closed, and the column is heated indirectly by a high-temperature thermostatic fluid. As the temperature increases, CO_2 desorbs and exits through the column entrance.
- (4) Light reflux (LR): the column end is opened, and the light product collected from the adsorption step is introduced through the column end, displacing the heavy product through the column entrance. During this step, the column begins to cool through the low-temperature thermostatic fluid.
- (5) Closed cooling (Cool): both ends of the column are closed, and the column is further cooled at a low temperature. During this process, the temperature and pressure inside the column continue to decrease.
- (6) Pressurization (Pres): the column entrance is opened, and the feed gas is reintroduced into the column. The pressure is brought back to the adsorption pressure during this step.

We developed a TSA model to evaluate the performance of the adsorbent in the TSA cycle, based on the 1D pressure vacuum swing adsorption (PVSA) model developed by Leperi et al.¹³⁶ and Yancy-Caballero et al.⁶⁹ The model fundamentally incorporates mass, energy, and momentum balances to account for all transport phenomena occurring within the column. Additionally, since the six-step TSA cycle involves adsorption and desorption driven by temperature changes via convective heat transfer, a wall-energy balance was also included. These governing equations, combined with the adsorption-equilibrium model and the linear-driving-force (LDF) model, form a set of partial differential equations (PDEs), which are detailed in Table S26. To ensure numerical stability, the PDEs were converted into a dimensionless form and then discretized using the finite-volume method (FVM)¹³⁷ with the weighted essentially non-oscillatory (WENO) scheme.¹³⁸ The column was discretized into 30 finite-volume elements. The resulting ordinary differential equations (ODEs) were solved using the ode15s solver¹³⁹ in MATLAB R2023a.

Different boundary conditions were applied to model each step of the six-step TSA cycle, which are summarized in Table S27. The TSA cycle was simulated using an uni-bed approach, where a single column was modeled to sequentially undergo all steps of the cycle. The TSA-cycle simulation was repeated until cyclic steady state (CSS) was achieved, which was attained when two criteria were satisfied: (1) the error between the initial condition in the first step (adsorption step) and final condition in the last step (pressurization step) of the six-step TSA cycle was less than 0.1%, and (2) the mass balance, defined as the ratio of gas entering and leaving the column over the entire cycle, was satisfied to within 99%. The cycle simulation was iterated up to a maximum of 700 times, and, if

CSS was not attained within this limit, the simulation was considered not converged and discarded.

Once CSS was achieved, CO_2 purity and recovery were calculated as follows:

$$CO_2 \text{ Purity} = \frac{n_{CO_2}^{Out \text{ from Heat}}}{n_{total}^{Out \text{ from Heat}}} \quad (\text{Equation 16})$$

$$CO_2 \text{ Recovery} = \frac{n_{CO_2}^{Out \text{ from Heat}}}{n_{CO_2}^{Into \text{ Pres}} + n_{CO_2}^{Into \text{ Ads}}} \quad (\text{Equation 17})$$

where $n_{CO_2}^{Out \text{ from Heat}}$ and $n_{total}^{Out \text{ from Heat}}$ are the number of CO_2 moles and the total number of moles in the outlet stream from the heating (heat) step, respectively; $n_{CO_2}^{Into \text{ Pres}}$ and $n_{CO_2}^{Into \text{ Ads}}$ are the number of CO_2 moles in the inlet streams from the pressurization (Pres) and adsorption steps (Ads), respectively. All parameters used to model the TSA cycle are listed in Tables S28 and S29.

Optimization of TSA cycle

Since the operating parameters of the process cycle have a significant impact on the process performance of the adsorbent, the TSA cycle was optimized using these parameters to determine the maximum achievable CO_2 purity and recovery for the adsorbent. Because purity and recovery are conflicting objectives, a multi-objective optimization was performed by setting each as an objective function. The optimization problem is defined as follows:

$$\text{minimize } J_1 = - CO_2 \text{ Purity}$$

$$\text{minimize } J_2 = - CO_2 \text{ Recovery}$$

$$\text{s.t. } CO_2 \text{ Purity} \geq y_{CO_2,0} \quad (\text{Equation 18})$$

where $y_{CO_2,0}$ is the CO_2 mole fraction in the feed gas.

The Thompson sampling efficient multi-objective optimization (TSEMO) algorithm,¹⁴⁰ as implemented in MATLAB R2023a,¹⁴¹ was used to solve this optimization problem. The TSEMO algorithm, based on a state-of-the-art Bayesian optimization framework, uses only a single CPU and solves problems much faster than the non-dominated sorting genetic algorithm (NSGA), a genetic algorithm-based optimization method, when using the same number of CPUs.¹⁴² This makes TSEMO an efficient tool for evaluating multiple adsorbents. In the TSEMO algorithm, the initial dataset size was set to 120, and 200 consecutive TSEMO iterations were performed. The decision variables along with the lower and upper bounds used in the optimization are listed in Table S28.

DFT calculation

We used the Vienna Ab initio Simulation Package (VASP)^{143,144} 6.4.1 to perform cell optimizations for 2019[Zn][nan]3[ASR]6. The interactions between core and valence electrons were modeled using the projector augmented wave (PAW) method. The Perdew-Burke-Ernzerhof (PBE)¹⁴⁵ formulation within the

generalized gradient approximation (GGA) was employed as the exchange-correlation functional with an energy cutoff of 520 eV. For the integration over the Brillouin zone, the Gaussian smearing method was utilized with a sigma value of 0.01 eV. Structural optimizations were performed using the conjugate gradient (CG) method, which allowed for the simultaneous relaxation of atomic positions, cell shape, and cell volume. The maximum number of ionic iterations was set to 500 to ensure full convergence. To accurately resolve the wave functions and charge densities, all calculations were performed with high precision. The electronic self-consistency field (SCF) iterations were conducted with an energy convergence criterion set to $10e^{-4}$ eV. Symmetry was turned off during the calculations to avoid any constraints on the atomic positions. Additionally, the DFT-D3 method with Becke-Johnson damping (BJ)^{146,147} was employed to account for van der Waals interactions. The Brillouin zone was sampled using a $1 \times 1 \times 1$ Monkhorst-Pack k-point grid. The fast algorithm was selected to accelerate the convergence of electronic steps, and the projection functions were kept in reciprocal space to enhance computational efficiency. We fitted the DDEC6 charges with CHARGEMOL 2017-09-26 software.^{148–151} AIMD simulations were employed to investigate the window diameter of the structure. Three CO₂ or three H₂O molecules were randomly placed inside the pores of the framework, and the framework, framework-3CO₂, and framework-3H₂O systems were maintained at constant temperatures of 300 K with the Nose-Hoover thermostat in NVT ensembles. The simulation time step was set to 2 fs, with a total simulation duration of 10 ps (5,000 steps). CO₂ and H₂O molecules were removed from the trajectory files, and the window diameter for 500 snapshots (starting with the last five ps, one snapshot every five steps) was calculated using the pywindow 0.0.2 package,¹⁵² with atomic radii set to CCDC radii.⁷³

Web interface and Python API

We developed a web interface (<https://mof-db.pusan.ac.kr/>) for the CoRE MOF DB to facilitate community engagement and error reporting. The web interface includes information for the latest version of CoRE MOF DB, including the number of unique MOFs (the sum of ASR and Ion datasets) per year. Users can search for MOF structures in the database (CoRE MOF SI dataset) by entering keywords and can filter structures by setting cut-off values for computed properties. Moreover, the website enables users to upload their structures for visualization and process the structures for solvent removal and NCR classification following the same methodologies used in this work. Users can also upload their structure for visualization and predict both geometric and ML-predicted properties developed in the community. These include structural descriptors (pore diameters, geometric surface areas, dimensionality of channels, and pore volume), OMS analysis, RACs calculations (highlight the position of uploaded structure relative to the CoRE MOF 2024 ASR dataset in *t*-SNE plot), heat capacity (at 300, 350, and 400 K), water stability, and PACMAN partial atomic charges. A dedicated webpage was developed for the users to submit issues and suggestions for the dataset, as well as to upload corrected crystal structures. Users can upload their structure to be included in our annual updates in the future by contributing their own experimental crystal structures.

Python API was also developed to demonstrate the usage of CoRE MOF-Tools, a Python package with example workflow to curate MOF crystals and compute properties. Detailed API documentation is available at <https://coremof-tools.readthedocs.io/en/latest/>.

Limitations of simulation methodologies in this work

For the GCMC simulations, we assumed a rigid framework, and the impact of structural flexibility on the results was not considered. The geometries of structures, except for 2019[Zn][nan]3[ASR]6, were optimized with neither DFT nor force fields. If the structures were optimized, the isotherms and the performance ranking based on process simulations might change (Figure S79).

For process simulations, we fixed some of the process parameters such as the mass-transfer coefficient and heat-transfer coefficient for all adsorbents during the process simulations. These parameters can typically be estimated through experimental measurements and may vary depending on MOFs. Additional evaluation metrics, such as productivity, process scheduling, scaling up, and a cost estimation model, have not been considered in this work.

RESOURCE AVAILABILITY

Lead contact

The lead contact is Yongchul G. Chung (drygchung@gmail.com).

Materials availability

This study did not generate new, unique reagents.

Data and code availability

This section outlines the availability of datasets and scripts associated with this study, adhering to the CSD license agreement and its addendum. All resources are publicly accessible as of March 8, 2025.

- (1) CIFs
 - 8,300 structures from CoRE MOF SI. Derived from the supporting information of published literature, these are freely available on the Zenodo web server at <https://zenodo.org/records/15055758>.
 - 20,276 structures from CoRE MOF CSD Modified. Adapted from the CSD database, these can be accessed freely via the CSD website with a valid email registration: <https://www.ccdc.cam.ac.uk/support-and-resources/downloads/>.
 - 12,261 structures from CoRE MOF CSD Unmodified. Available through the CCDC GitHub repository using a Python API script (<https://github.com/ccdc-opensource/csd-python-api-scripts/tree/main/notebooks/CoRE-MOF>), provided the user has a valid CSD license (CSD-Core or better). These structures require additional processing (e.g., conversion to primitive cells and imposing P1 symmetry) and assignment of PACMAN DDEC6 partial atomic charges. The workflow for these operations is detailed in the same repository.
- (2) Excel spreadsheets
 - Split into two parts:
 - Available on Zenodo: <https://zenodo.org/records/15055758>.
 - Available on the CSD website: <https://www.ccdc.cam.ac.uk/support-and-resources/downloads/>.
- (3) Analysis scripts
 - All scripts used in this work are publicly available in the GitHub repository: <https://github.com/mtap-research/CoRE-MOF-Tools>.

ACKNOWLEDGMENTS

G.Z., S.Y., H.K., and Y.G.C. are supported by National Research Foundation of Korea (NRF) grants funded by the Korean government (RS-2021-NR057842,

RS-2023-00242528, and RS-2024-00449431). L.M.B. thanks Meta for financial support. D.S.S. received support from the US Department of Energy, Office of Science Energy Earthshot Initiative as part of NEETER at ORNL under FWP #ERKCG70. The work performed by S.C., P., J.I.S., K.L., K.M., P.D.P., and R.Q.S. is supported by the US Department of Energy, Office of Science, Basic Energy Sciences, under award DE-SC0023454, as part of the Computational and Theoretical Chemistry program. S.M.M. and J.H. acknowledge support from the National Research Council Canada under the Materials for Clean Fuel program, as well as the Mitacs Elevate program. This work was supported under the France 2030 Framework by Agence Nationale de la Recherche (project ANR-22-PEXD-0009 DIADEM). M.H. is supported by the MAD2D-CM project funded by the Community of Madrid; the Recovery, Transformation and Resilience Plan, Spain; and NextGenerationEU from the European Union. G.G.T. and H.J.K. were supported by a seed grant from the Abdul Latif Jameel Water and Food Systems Lab at the Massachusetts Institute of Technology as well as by the Center for Enhanced Nanofluidic Transport, an Energy Frontier Research Center funded by the US Department of Energy, Office of Science, Basic Energy Sciences under award DE-SC0019112.

The authors acknowledge the computational time provided by KISTI (KSC-2024-CRE-0283), GENCI (A0150807069). Computational resources for the GEMC simulations of water adsorption were provided by the Minnesota Supercomputing Institute at the University of Minnesota. We thank Yu Chen for his contribution to the design of the CoRE MOF logo.

AUTHOR CONTRIBUTIONS

CIF collection and curation, G.Z. and Y.G.C.; NCR classification, L.M.B., G.Z., and Y.G.C.; geometric features calculation and diversity analysis, L.Z., G.G.T., M.H., and G.Z.; MOFid, K.L., K.M., P.D.P., and R.Q.S.; ML-based properties prediction, J.H., G.G.T., G.Z., and Y.G.C.; water GEMC calculations, S.C., P., and J.I.S.; multi-scale screening, S.Y. and G.Z.; AIMD simulations, G.Z. and Y.G.C.; web interface and Python API, G.Z. and Y.G.C.; conceptualization, Y.G.C.; methodology, G.Z., L.M.B., S.C., M.H., J.H., K.L., K.M., T.P., P.P., G.G.T., S.Y., and L.Z.; software, G.Z., L.M.B., S.C., M.H., J.H., K.L., K.M., T.P., P.P., G.G.T., S.Y., and L.Z.; validation, G.Z.; formal analysis, G.Z., L.M.B., S.C., J.H., K.L., K.M., T.P., P.P., G.G.T., S.Y., and L.Z.; investigation, G.Z.; resources, G.Z., F.-X.C., M.H., H.J.K., S.M.M., D.S.S., J.I.S., R.Q.S., and Y.G.C.; data curation, G.Z., L.M.B., S.C., J.H., K.L., K.M., T.P., P.P., G.G.T., S.Y., and L.Z.; writing – original draft, G.Z., L.M.B., S.C., J.H., K.L., K.M., T.P., P.P., G.G.T., S.Y., and L.Z.; writing – review and editing, F.-X.C., H.J.K., S.M.M., D.S.S., J.I.S., R.Q.S., and Y.G.C.; visualization, G.Z., S.C., J.H., H.K., K.L., K.M., T.P., P.P., G.G.T., S.Y., and L.Z.; supervision, F.-X.C., M.H., H.J.K., S.M.M., D.S.S., J.I.S., R.Q.S., and Y.G.C.; project administration, F.-X.C., M.H., H.J.K., S.M.M., D.S.S., J.I.S., R.Q.S., and Y.G.C.; funding acquisition, F.-X.C., M.H., H.J.K., S.M.M., D.S.S., J.I.S., R.Q.S., and Y.G.C.

DECLARATION OF INTERESTS

R.Q.S. has a financial interest in the startup company NuMat Technologies, which is commercializing MOFs.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.matt.2025.102140>.

Received: December 5, 2024

Revised: March 7, 2025

Accepted: April 8, 2025

Published: May 5, 2025

REFERENCES

- Ding, M., Flaig, R.W., Jiang, H.-L., and Yaghi, O.M. (2019). Carbon capture and conversion using metal–organic frameworks and MOF-based materials. *Chem. Soc. Rev.* *48*, 2783–2828. <https://doi.org/10.1039/C8CS00829A>.
- Chen, Z., Kirlikovali, K.O., Idrees, K.B., Wasson, M.C., and Farha, O.K. (2022). Porous materials for hydrogen storage. *Chem* *8*, 693–716. <https://doi.org/10.1016/j.chempr.2022.01.012>.
- Ma, S., Han, W., Han, W., Dong, F., and Tang, Z. (2023). Recent advances and future perspectives in MOF-derived single-atom catalysts and their application: a review. *J. Mater. Chem. A* *11*, 3315–3363. <https://doi.org/10.1039/D2TA08735A>.
- Navalón, S., Dhakshinamoorthy, A., Álvaro, M., Ferrer, B., and García, H. (2023). Metal–Organic Frameworks as Photocatalysts for Solar-Driven Overall Water Splitting. *Chem. Rev.* *123*, 445–490. <https://doi.org/10.1021/acs.chemrev.2c00460>.
- Kreno, L.E., Leong, K., Farha, O.K., Allendorf, M., Van Duyne, R.P., and Hupp, J.T. (2012). Metal–Organic Framework Materials as Chemical Sensors. *Chem. Rev.* *112*, 1105–1125. <https://doi.org/10.1021/cr200324t>.
- Peng, Y., Krungleviciute, V., Eryazici, I., Hupp, J.T., Farha, O.K., and Yildirim, T. (2013). Methane Storage in Metal–Organic Frameworks: Current Records, Surprise Findings, and Challenges. *J. Am. Chem. Soc.* *135*, 11887–11894. <https://doi.org/10.1021/ja4045289>.
- Lu, X.F., Fang, Y., Luan, D., and Lou, X.W.D. (2021). Metal–Organic Frameworks Derived Functional Materials for Electrochemical Energy Storage and Conversion: A Mini Review. *Nano Lett.* *21*, 1555–1565. <https://doi.org/10.1021/acs.nanolett.0c04898>.
- Moghadam, P.Z., Chung, Y.G., and Snurr, R.Q. (2024). Progress toward the computational discovery of new metal–organic framework adsorbents for energy applications. *Nat. Energy* *9*, 121–133. <https://doi.org/10.1038/s41560-023-01417-2>.
- Sriram, A., Choi, S., Yu, X., Brabson, L.M., Das, A., Ulissi, Z., Uytendaele, M., Medford, A.J., and Sholl, D.S. (2024). The Open DAC 2023 Dataset and Challenges for Sorbent Discovery in Direct Air Capture. *ACS Cent. Sci.* *10*, 923–941. <https://doi.org/10.1021/acscentsci.3c01629>.
- Bose, S., Sengupta, D., Rayder, T.M., Wang, X., Kirlikovali, K.O., Sekizkardes, A.K., Islamoglu, T., and Farha, O.K. (2023). Challenges and Opportunities: Metal–Organic Frameworks for Direct Air Capture. *Adv. Funct. Mater.* *34*, 2307478. <https://doi.org/10.1002/adfm.202307478>.
- Colón, Y.J., and Snurr, R.Q. (2014). High-throughput computational screening of metal–organic frameworks. *Chem. Soc. Rev.* *43*, 5735–5749. <https://doi.org/10.1039/C4CS00070F>.
- Jablonka, K.M., Ongari, D., Moosavi, S.M., and Smit, B. (2020). Big-Data Science in Porous Materials: Materials Genomics and Machine Learning. *Chem. Rev.* *120*, 8066–8129. <https://doi.org/10.1021/acs.chemrev.0c00004>.
- Moosavi, S.M., Jablonka, K.M., and Smit, B. (2020). The Role of Machine Learning in the Understanding and Design of Materials. *J. Am. Chem. Soc.* *142*, 20273–20287. <https://doi.org/10.1021/jacs.0c09105>.
- Chibani, S., and Coudert, F.-X. (2020). Machine learning approaches for the prediction of materials properties. *APL Mater.* *8*, 245. <https://doi.org/10.1063/5.0018384>.
- Park, J., Kim, H., Kang, Y., Lim, Y., and Kim, J. (2024). From Data to Discovery: Recent Trends of Machine Learning in Metal–Organic Frameworks. *JACS Au* *4*, 3727–3743. <https://doi.org/10.1021/jacsau.4c00618>.
- Altintas, C., Avci, G., Daglar, H., Nemati Vesali Azar, A., Erucar, I., Velioglu, S., and Keskin, S. (2019). An extensive comparative analysis of two MOF databases: high-throughput screening of computation-ready MOFs for CH₄ and H₂ adsorption. *J. Mater. Chem. A* *7*, 9593–9608. <https://doi.org/10.1039/C9TA01378D>.
- Chen, T., and Manz, T.A. (2020). Identifying misbonded atoms in the 2019 CoRE metal–organic framework database. *RSC Adv.* *10*, 26944–26951. <https://doi.org/10.1039/D0RA02498H>.
- Nurhuda, M., Perry, C.C., and Addicoat, M.A. (2022). Performance of GFN1-xTB for periodic optimization of metal organic frameworks. *Phys. Chem. Chem. Phys.* *24*, 10906–10914. <https://doi.org/10.1039/D2CP00184E>.

19. Coupury, D.E., Addicoat, M.A., and Heine, T. (2016). Extension of the Universal Force Field for Metal–Organic Frameworks. *J. Chem. Theory Comput.* *12*, 5215–5225. <https://doi.org/10.1021/acs.jctc.6b00664>.
20. Ongari, D., Talirz, L., and Smit, B. (2020). Too Many Materials and Too Many Applications: An Experimental Problem Waiting for a Computational Solution. *ACS Cent. Sci.* *6*, 1890–1900. <https://doi.org/10.1021/acscentsci.0c00988>.
21. Daglar, H., Gulbalkan, H.C., Avci, G., Aksu, G.O., Altundal, O.F., Altintas, C., Erucar, I., and Keskin, S. (2021). Effect of Metal–Organic Framework (MOF) Database Selection on the Assessment of Gas Storage and Separation Potentials of MOFs. *Angew. Chem. Int. Ed. Engl.* *60*, 7828–7837. <https://doi.org/10.1002/anie.202015250>.
22. Chung, Y.G., Camp, J., Haranczyk, M., Sikora, B.J., Bury, W., Krungleviciute, V., Yildirim, T., Farha, O.K., Sholl, D.S., and Snurr, R.Q. (2014). Computation-Ready, Experimental Metal–Organic Frameworks: A Tool To Enable High-Throughput Screening of Nanoporous Crystals. *Chem. Mater.* *26*, 6185–6192. <https://doi.org/10.1021/cm502594j>.
23. Chung, Y.G., Haldoupis, E., Bucior, B.J., Haranczyk, M., Lee, S., Zhang, H., Vogiatzis, K.D., Milisavljevic, M., Ling, S., Camp, J.S., et al. (2019). Advances, Updates, and Analytics for the Computation-Ready, Experimental Metal–Organic Framework Database: CoRE MOF 2019. *J. Chem. Eng. Data* *64*, 5985–5998. <https://doi.org/10.1021/acs.jced.9b00835>.
24. White, A., Gibaldi, M., Burner, J., Mayo, R.A., and Woo, T. (2024). Alarming structural error rates in MOF databases used in data driven workflows identified via a novel metal oxidation state-based method. Preprint at ChemRxiv. <https://doi.org/10.26434/chemrxiv-2024-ftsv3>.
25. Li, A., Perez, R.B., Wiggin, S., Ward, S.C., Wood, P.A., and Fairen-Jimeñez, D. (2021). The launch of a freely accessible MOF CIF collection from the CSD. *Matter* *4*, 1105–1106. <https://doi.org/10.1016/j.matt.2021.03.006>.
26. Daglar, H., Gulbalkan, H.C., Aksu, G.O., and Keskin, S. (2024). Computational Simulations of Metal–Organic Frameworks to Enhance Adsorption Applications. *Adv. Mater.* 2405532. <https://doi.org/10.1002/adma.202405532>.
27. Setyawan, W., and Curtarolo, S. (2010). High-throughput electronic band structure calculations: Challenges and tools. *Comput. Mater. Sci.* *49*, 299–312. <https://doi.org/10.1016/j.commatsci.2010.05.010>.
28. Gibaldi, M., Kapeliukha, A., White, A., and Woo, T.K. (2025). Incorporation of Ligand Charge and Metal Oxidation State Considerations into the Computational Solvent Removal and Activation of Experimental Crystal Structures Preceding Molecular Simulation. *J. Chem. Inf. Model.* *65*, 275–287. <https://doi.org/10.1021/acs.jcim.4c01897>.
29. Gharagheizi, F., Yu, Z., and Sholl, D.S. (2022). Curated Collection of More than 20,000 Experimentally Reported One-Dimensional Metal–Organic Frameworks. *ACS Appl. Mater. Interfaces* *14*, 42258–42266. <https://doi.org/10.1021/acscami.2c12485>.
30. Retractions in the Cambridge Structural Database. <https://www.ccdc.cam.ac.uk/support-and-resources/support/case/?caseid=819cfd76-c25d-40a2-ac9b-b4cf20d775a7>.
31. Bimler, D. (2022). Better Living through Coordination Chemistry: A descriptive study of a prolific papermill that combines crystallography and medicine. *Res. Sq* *50*, 234. <https://doi.org/10.21203/rs.3.rs-1537438/v1>.
32. Xin, J., Jablonka, K., Moubarak, E., Li, Y., and Smit, B. (2025). MOFChecker: An algorithm for Validating and Correcting Metal–Organic Framework (MOF) Structures. <https://doi.org/10.26434/chemrxiv-2025-bh607>.
33. Férey, G., Mellot-Draznieks, C., Serre, C., Millange, F., Dutour, J., Surblé, S., and Margiolaki, I. (2005). A Chromium Terephthalate-Based Solid with Unusually Large Pore Volumes and Surface Area. *Science* *309*, 2040–2042. <https://doi.org/10.1126/science.1116275>.
34. Clegg, J.K., Iremonger, S.S., Hayter, M.J., Southon, P.D., Macquart, R. B., Duriska, M.B., Jensen, P., Turner, P., Jolliffe, K.A., Kepert, C.J., et al. (2010). Hierarchical Self-Assembly of a Chiral Metal–Organic Framework Displaying Pronounced Porosity. *Angew. Chem. Int. Ed. Engl.* *49*, 1075–1078. <https://doi.org/10.1002/anie.200905497>.
35. Banerjee, D., Simon, C.M., Plonka, a.m., Motkuri, R.K., Liu, J., Chen, X., Smit, B., Parise, J.B., Haranczyk, M., and Thallapally, P.K. (2016). Metal–organic framework with optimally selective xenon adsorption and separation. *Nat. Commun.* *7*, ncomms11831. <https://doi.org/10.1038/ncomms11831>.
36. Rosen, A.S., Iyer, S.M., Ray, D., Yao, Z., Aspuru-Guzik, A., Gagliardi, L., Notestein, J.M., and Snurr, R.Q. (2021). Machine learning the quantum-chemical properties of metal–organic frameworks for accelerated materials discovery. *Matter* *4*, 1578–1597. <https://doi.org/10.1016/j.matt.2021.02.015>.
37. Gibaldi, M., Kapeliukha, A., White, A., Luo, J., Mayo, R.A., Burner, J., and Woo, T.K. (2025). MOSAEC-DB: a comprehensive database of experimental metal–organic frameworks with verified chemical accuracy suitable for molecular simulations. *Chem. Sci.* *16*, 4085–4100. <https://doi.org/10.1039/D4SC07438F>.
38. Bucior, B.J., Rosen, A.S., Haranczyk, M., Yao, Z., Ziebel, M.E., Farha, O. K., Hupp, J.T., Siepmann, J.I., Aspuru-Guzik, A., and Snurr, R.Q. (2019). Identification Schemes for Metal–Organic Frameworks To Enable Rapid Search and Cheminformatics Analysis. *Cryst. Growth Des.* *19*, 6682–6697. <https://doi.org/10.1021/acs.cgd.9b01050>.
39. Alexandrov, E.V., Blatov, V.A., and Proserpio, D.M. (2014). Interpenetration of three-periodic networks in crystal structures: Description and classification methods, geometrical-topological conditions of implementation. *J. Struct. Chem.* *55*, 1308–1325. <https://doi.org/10.1134/S0022476614070130>.
40. Zoubritsky, L., and Coudert, F.-X. (2022). CrystalNets.jl: Identification of Crystal Topologies. *SciPost Chem.* *1*, 005. <https://doi.org/10.21468/SciPostChem.1.2.005>.
41. Chui, S., Lo, S., Charmant, J., Orpen, A.G., and Williams, I.D. (1999). A Chemically Functionalizable Nanoporous Material [Cu₃(TMA)₂(H₂O)₃]_n. *Science* *283*, 1148–1150. <https://doi.org/10.1126/science.283.5405.1148>.
42. Rosi, N.L., Eckert, J., Eddaoudi, M., Vodak, D.T., Kim, J., O’Keeffe, M., and Yaghi, O.M. (2003). Hydrogen Storage in Microporous Metal–Organic Frameworks. *Science* *300*, 1127–1129. <https://doi.org/10.1126/science.1083440>.
43. Wong-Ng, W., Kaduk, J.A., Wu, H., and Suchomel, M. (2012). Synchrotron X-ray studies of metal–organic framework M₂(2,5-dihydroxyterephthalate), M = (Mn, Co, Ni, Zn) (MOF74). *Powder Diffr.* *27*, 256–262. <https://doi.org/10.1017/S0885715612000863>.
44. Wang, Z., Zhang, B., Kurmoo, M., Green, M.A., Fujiwara, H., Otsuka, T., and Kobayashi, H. (2005). Synthesis and Characterization of a Porous Magnetic Diamond Framework, Co₃(HCOO)₆, and Its N₂ Sorption Characteristic. *Inorg. Chem.* *44*, 1230–1237. <https://doi.org/10.1021/ic048986r>.
45. Wang, Z.-M., Zhang, Y.-J., Liu, T., Kurmoo, M., and Gao, S. (2007). [Fe₃(HCOO)₆]: A Permanent Porous Diamond Framework Displaying H₂/N₂ Adsorption, Guest Inclusion, and Guest-Dependent Magnetism. *Adv. Funct. Mater.* *17*, 1523–1536. <https://doi.org/10.1002/adfm.200600676>.
46. Moosavi, S.M., Nandy, A., Jablonka, K.M., Ongari, D., Janet, J.P., Boyd, P.G., Lee, Y., Smit, B., and Kulik, H.J. (2020). Understanding the diversity of the metal–organic framework ecosystem. *Nat. Commun.* *11*, 4068. <https://doi.org/10.1038/s41467-020-17755-8>.
47. Nandy, A., Yue, S., Oh, C., Duan, C., Terrones, G.G., Chung, Y.G., and Kulik, H.J. (2023). A database of ultrastable MOFs reassembled from stable fragments with machine learning models. *Matter* *6*, 1585–1603. <https://doi.org/10.1016/j.matt.2023.03.009>.
48. Rajendran, A., Shimizu, G.K.H., and Woo, T.K. (2024). The Challenge of Water Competition in Physical Adsorption of CO₂ by Porous Solids for Carbon Capture Applications – A Short Perspective. *Adv. Mater.* *36*, 2301730. <https://doi.org/10.1002/adma.202301730>.

49. Zhao, G., and Chung, Y.G. (2024). PACMAN: A Robust Partial Atomic Charge Predictor for Nanoporous Materials Based on Crystal Graph Convolution Networks. *J. Chem. Theory Comput.* *20*, 5368–5380. <https://doi.org/10.1021/acs.jctc.4c00434>.
50. Zhao, G., Chen, Y., and Chung, Y.G. (2023). High-Throughput, Multiscale Computational Screening of Metal–Organic Frameworks for Xe/Kr Separation with Machine-Learned Parameters. *Ind. Eng. Chem. Res.* *62*, 15176–15189. <https://doi.org/10.1021/acs.iecr.3c02211>.
51. Anderson, R., Biong, A., and Gómez-Gualdrón, D.A. (2020). Adsorption Isotherm Predictions for Multiple Molecules in MOFs Using the Same Deep Learning Model. *J. Chem. Theory Comput.* *16*, 1271–1283. <https://doi.org/10.1021/acs.jctc.9b00940>.
52. Charalambous, C., Moubarak, E., Schilling, J., Sanchez Fernandez, E., Wang, J.-Y., Herraiz, L., McIlwaine, F., Peh, S.B., Garvin, M., Jablonka, K.M., et al. (2024). A holistic platform for accelerating sorbent-based carbon capture. *Nature* *632*, 89–94. <https://doi.org/10.1038/s41586-024-07683-8>.
53. Nandy, A., Duan, C., and Kulik, H.J. (2021). Using Machine Learning and Data Mining to Leverage Community Knowledge for the Engineering of Stable Metal–Organic Frameworks. *J. Am. Chem. Soc.* *143*, 17535–17547. <https://doi.org/10.1021/jacs.1c07217>.
54. Moosavi, S.M., Novotny, B.Á., Ongari, D., Moubarak, E., Asgari, M., Kadioglu, Ö., Charalambous, C., Ortega-Guerrero, A., Farmahini, A.H., Sarkisov, L., et al. (2022). A data-science approach to predict the heat capacity of nanoporous materials. *Nat. Mater.* *21*, 1419–1425. <https://doi.org/10.1038/s41563-022-01374-3>.
55. Huck, J.M., Lin, L.-C., Berger, A.H., Shahrak, M.N., Martin, R.L., Bhowan, A.S., Haranczyk, M., Reuter, K., and Smit, B. (2014). Evaluating different classes of porous materials for carbon capture. *Energy Environ. Sci.* *7*, 4132–4146. <https://doi.org/10.1039/C4EE02636E>.
56. Terrones, G.G., Huang, S.-P., Rivera, M.P., Yue, S., Hernandez, A., and Kulik, H.J. (2024). Metal–Organic Framework Stability in Water and Harsh Environments from Data-Driven Models Trained on the Diverse WS24 Dataset. *J. Am. Chem. Soc.* *146*, 20333–20348. <https://doi.org/10.1021/jacs.4c05879>.
57. Mohamed, S.A., Zhao, D., and Jiang, J. (2023). Integrating stability metrics with high-throughput computational screening of metal–organic frameworks for CO₂ capture. *Commun. Mater.* *4*, 79. <https://doi.org/10.1038/s43246-023-00409-9>.
58. Moghadam, P.Z., Rogge, S.M.J., Li, A., Chow, C.-M., Wieme, J., Mohararami, N., Aragones-Anglada, M., Conduit, G., Gomez-Gualdrón, D.A., Van Speybroeck, V., and Fairen-Jimenez, D. (2019). Structure-Mechanical Stability Relations of Metal-Organic Frameworks via Machine Learning. *Matter* *1*, 219–234. <https://doi.org/10.1016/j.matt.2019.03.002>.
59. Jamdade, S., Yu, Z., Boulfelfel, S.E., Cai, X., Thyagarajan, R., Fang, H., and Sholl, D.S. (2024). Probing Structural Defects in MOFs Using Water Stability. *J. Phys. Chem. C* *128*, 3975–3984. <https://doi.org/10.1021/acs.jpcc.3c07497>.
60. Yu, Z., Jamdade, S., Yu, X., Cai, X., and Sholl, D.S. (2023). Efficient Generation of Large Collections of Metal–Organic Framework Structures Containing Well-Defined Point Defects. *J. Phys. Chem. Lett.* *14*, 6658–6665. <https://doi.org/10.1021/acs.jpcllett.3c01524>.
61. Lin, J.-B., Nguyen, T.T.T., Vaidhyanathan, R., Burner, J., Taylor, J.M., Durekova, H., Akhtar, F., Mah, R.K., Ghaffari-Nik, O., Marx, S., et al. (2021). A scalable metal-organic framework as a durable physisorbent for carbon dioxide capture. *Science* *374*, 1464–1469. <https://doi.org/10.1126/science.abi7281>.
62. Hovington, P., Ghaffari-Nik, O., Mariac, L., Liu, A., Henkel, B., and Marx, S. (2022). Rapid cycle temperature swing adsorption process using solid structured sorbent for CO₂ capture from cement flue gas. In *Proceedings of the 16th Greenhouse Gas Control Technologies Conference (GHGT-16)*.
63. Zhang, Y., Shi, D.-W., Ma, Y.-Q., Li, H.-Y., Bian, H.-D., Liu, H.-F., Tang, J., and Huang, F.-P. (2019). Crystal structures and fluorescence properties of two architecture-correlated Zn(II)/Cd(II) MOFs constructed from an unprecedented tri-triazole ligand. *J. Coord. Chem.* *72*, 1002–1012. <https://doi.org/10.1080/00958972.2019.1580363>.
64. Keskin, S., van Heest, T.M., and Sholl, D.S. (2010). Can Metal–Organic Framework Materials Play a Useful Role in Large-Scale Carbon Dioxide Separations? *ChemSusChem* *3*, 879–891. <https://doi.org/10.1002/cssc.201000114>.
65. Nguyen, T.T.T., Balasubramaniam, B.M., Fylstra, N., Huynh, R.P.S., Shimizu, G.K.H., and Rajendran, A. (2024). Competitive CO₂/H₂O Adsorption on CALF-20. *Ind. Eng. Chem. Res.* *63*, 3265–3281. <https://doi.org/10.1021/acs.iecr.3c04266>.
66. Gopalsamy, K., Fan, D., Naskar, S., Magnin, Y., and Maurin, G. (2024). Engineering of an Isoreticular Series of CALF-20 Metal–Organic Frameworks for CO₂ Capture. *ACS Appl. Eng. Mater.* *2*, 96–103. <https://doi.org/10.1021/acsaenm.3c00622>.
67. Ho, C.-H., and Paesani, F. (2023). Elucidating the Competitive Adsorption of H₂O and CO₂ in CALF-20: New Insights for Enhanced Carbon Capture Metal–Organic Frameworks. *ACS Appl. Mater. Interfaces* *15*, 48287–48295. <https://doi.org/10.1021/acsaami.3c11092>.
68. Park, J., Rubiera Landa, H.O., Kawajiri, Y., Realf, M.J., Lively, R.P., and Sholl, D.S. (2020). How Well Do Approximate Models of Adsorption-Based CO₂ Capture Processes Predict Results of Detailed Process Models? *Ind. Eng. Chem. Res.* *59*, 7097–7108. <https://doi.org/10.1021/acs.iecr.9b05363>.
69. Yancy-Caballero, D., Leperi, K.T., Bucior, B.J., Richardson, R.K., Islamoglu, T., Farha, O.K., You, F., and Snurr, R.Q. (2020). Process-level modelling and optimization to evaluate metal-organic frameworks for post-combustion capture of CO₂. *Mol. Syst. Des. Eng.* *5*, 1205–1218. <https://doi.org/10.1039/D0ME00060D>.
70. Leperi, K.T., Chung, Y.G., You, F., and Snurr, R.Q. (2019). Development of a General Evaluation Metric for Rapid Screening of Adsorbent Materials for Postcombustion CO₂ Capture. *ACS Sustain. Chem. Eng.* *7*, 11529–11539. <https://doi.org/10.1021/acssuschemeng.9b01418>.
71. Zheng, Z., Zhang, O., Nguyen, H.L., Rampal, N., Alawadhi, A.H., Rong, Z., Head-Gordon, T., Borgs, C., Chayes, J.T., and Yaghi, O.M. (2023). ChatGPT Research Group for Optimizing the Crystallinity of MOFs and COFs. *ACS Cent. Sci.* *9*, 2161–2170. <https://doi.org/10.1021/acscentsci.3c01087>.
72. Zheng, Z., Alawadhi, A.H., Chheda, S., Neumann, S.E., Rampal, N., Liu, S., Nguyen, H.L., Lin, Y.-h., Rong, Z., Siepmann, J.I., et al. (2023). Shaping the Water-Harvesting Behavior of Metal–Organic Frameworks Aid by Fine-Tuned GPT Models. *J. Am. Chem. Soc.* *145*, 28284–28295. <https://doi.org/10.1021/jacs.3c12086>.
73. Cordero, B., Gómez, V., Platero-Prats, A.E., Revés, M., Echeverría, J., Cremades, E., Barragán, F., and Alvarez, S. (2008). Covalent radii revisited. *Dalton Trans.*, 2832–2838. <https://doi.org/10.1039/B801115J>.
74. Kancharlapalli, S., and Snurr, R.Q. (2023). High-Throughput Screening of the CoRE-MOF-2019 Database for CO₂ Capture from Wet Flue Gas: A Multi-Scale Modeling Strategy. *ACS Appl. Mater. Interfaces* *15*, 28084–28092. <https://doi.org/10.1021/acsami.3c04079>.
75. Chen, T., and Manz, T.A. (2019). A collection of forcefield precursors for metal–organic frameworks. *RSC Adv.* *9*, 36492–36507. <https://doi.org/10.1039/C9RA07327B>.
76. Blatov, V.A., and Proserpio, D.M. (2010). Periodic-Graph Approaches in Crystal Structure Prediction. In *Modern Methods of Crystal Structure Prediction*, Artem R. Oganov, ed. (WILEY), pp. 1–28. <https://doi.org/10.1002/9783527632831.ch1>.
77. Delgado-Friedrichs, O., and O’Keeffe, M. (2005). Crystal nets as graphs: Terminology and definitions. *J. Solid State Chem.* *178*, 2480–2485. <https://doi.org/10.1016/j.jssc.2005.06.011>.

78. Li, M., Li, D., O’Keeffe, M., and Yaghi, O.M. (2014). Topological Analysis of Metal–Organic Frameworks with Polytopic Linkers and/or Multiple Building Units and the Minimal Transitivity Principle. *Chem. Rev.* *114*, 1343–1370. <https://doi.org/10.1021/cr400392k>.
79. Blatov, V.A., Delgado-Friedrichs, O., O’Keeffe, M., and Proserpio, D.M. (2007). Three-periodic nets and tilings: natural tilings for nets. *Acta Crystallogr. A* *63*, 418–425. <https://doi.org/10.1107/S0108767307038287>.
80. Bonneau, C., O’Keeffe, M., Proserpio, D.M., Blatov, V.A., Batten, S.R., Bourne, S.A., Lah, M.S., Eon, J.-G., Hyde, S.T., Wiggin, S.B., and Öhrström, L. (2018). Deconstruction of Crystalline Networks into Underlying Nets: Relevance for Terminology Guidelines and Crystallographic Databases. *Cryst. Growth Des.* *18*, 3411–3418. <https://doi.org/10.1021/acs.cgd.8b00126>.
81. Blatov, V.A. (2016). A method for topological analysis of rod packings. *Struct. Chem.* *27*, 1605–1611. <https://doi.org/10.1007/s11224-016-0774-1>.
82. O’Keeffe, M., Peskov, M.A., Ramsden, S.J., and Yaghi, O.M. (2008). The Reticular Chemistry Structure Resource (RCSR) Database of, and Symbols for, Crystal Nets. *Acc. Chem. Res.* *41*, 1782–1789. <https://doi.org/10.1021/ar800124u>.
83. Deem, M.W., Pophale, R., Cheeseman, P.A., and Earl, D.J. (2009). Computational Discovery of New Zeolite-Like Materials. *J. Phys. Chem. C* *113*, 21353–21360. <https://doi.org/10.1021/jp906984z>.
84. Pophale, R., Cheeseman, P.A., and Deem, M.W. (2011). A database of new zeolite-like materials. *Phys. Chem. Chem. Phys.* *13*, 12407–12412. <https://doi.org/10.1039/C0CP02255A>.
85. Robins, V., Ramsden, S., and Hyde, S. (2006). EPINET: euclidean patterns in non-euclidean tilings. ACM SIGGRAPH 2006 Research Posters. Association for Computing Machinery.
86. Willems, T.F., Rycroft, C.H., Kazi, M., Meza, J.C., and Haranczyk, M. (2012). Algorithms and tools for high-throughput geometry-based analysis of crystalline porous materials. *Microporous Mesoporous Mater.* *149*, 134–141.
87. Martin, R.L., Smit, B., and Haranczyk, M. (2012). Addressing Challenges of Identifying Geometrically Diverse Sets of Crystalline Porous Materials. *J. Chem. Inf. Model.* *52*, 308–318. <https://doi.org/10.1021/ci200386x>.
88. Pinheiro, M., Martin, R.L., Rycroft, C.H., Jones, A., Iglesia, E., and Haranczyk, M. (2013). Characterization and comparison of pore landscapes in crystalline porous materials. *J. Mol. Graph. Model.* *44*, 208–219. <https://doi.org/10.1016/j.jmkgm.2013.05.007>.
89. Pinheiro, M., Martin, R.L., Rycroft, C.H., and Haranczyk, M. (2013). High accuracy geometric analysis of crystalline porous materials. *CrystEngComm* *15*, 7531–7538. <https://doi.org/10.1039/C3CE41057A>.
90. Martin, R.L., and Haranczyk, M. (2014). Construction and Characterization of Structure Models of Crystalline Porous Polymers. *Cryst. Growth Des.* *14*, 2431–2440. <https://doi.org/10.1021/cg500158c>.
91. Boyd, P.G., Chidambaram, A., García-Díez, E., Ireland, C.P., Daff, T.D., Bounds, R., Gladysiak, A., Schouwink, P., Moosavi, S.M., Maroto-Valer, M.M., et al. (2019). Data-driven design of metal–organic frameworks for wet flue gas CO₂ capture. *Nature* *576*, 253–256. <https://doi.org/10.1038/s41586-019-1798-7>.
92. Lan, Y., Yan, T., Tong, M., and Zhong, C. (2019). Large-scale computational assembly of ionic liquid/MOF composites: synergistic effect in the wire-tube conformation for efficient CO₂/CH₄ separation. *J. Mater. Chem. A* *7*, 12556–12564. <https://doi.org/10.1039/C9TA01752F>.
93. Colón, Y.J., Gómez-Gualdrón, D.A., and Snurr, R.Q. (2017). Topologically Guided, Automated Construction of Metal–Organic Frameworks and Their Evaluation for Energy-Related Applications. *Cryst. Growth Des.* *17*, 5801–5810. <https://doi.org/10.1021/acs.cgd.7b00848>.
94. Anderson, R., Rodgers, J., Argueta, E., Biong, A., and Gómez-Gualdrón, D.A. (2018). Role of Pore Chemistry and Topology in the CO₂ Capture Capabilities of MOFs: From Molecular Simulation to Machine Learning. *Chem. Mater.* *30*, 6325–6337. <https://doi.org/10.1021/acs.chemmater.8b02257>.
95. Gomez-Gualdrón, D.A., Gutov, O.V., Krungleviciute, V., Borah, B., Mondloch, J.E., Hupp, J.T., Yildirim, T., Farha, O.K., and Snurr, R.Q. (2014). Computational Design of Metal–Organic Frameworks Based on Stable Zirconium Building Units for Storage and Delivery of Methane. *Chem. Mater.* *26*, 5632–5639. <https://doi.org/10.1021/cm502304e>.
96. Chung, Y.G., Gómez-Gualdrón, D.A., Li, P., Leperi, K.T., Deria, P., Zhang, H., Vermeulen, N.A., Stoddart, J.F., You, F., Hupp, J.T., et al. (2016). In silico discovery of metal-organic frameworks for precombustion CO₂ capture using a genetic algorithm. *Sci. Adv.* *2*, e1600909. <https://doi.org/10.1126/sciadv.1600909>.
97. Li, S., Chung, Y.G., Simon, C.M., and Snurr, R.Q. (2017). High-Throughput Computational Screening of Multivariate Metal–Organic Frameworks (MTV-MOFs) for CO₂ Capture. *J. Phys. Chem. Lett.* *8*, 6135–6141. <https://doi.org/10.1021/acs.jpclett.7b02700>.
98. Majumdar, S., Moosavi, S.M., Jablonka, K.M., Ongari, D., and Smit, B. (2021). Diversifying Databases of Metal Organic Frameworks for High-Throughput Computational Screening. *ACS Appl. Mater. Interfaces* *13*, 61004–61014. <https://doi.org/10.1021/acsami.1c16220>.
99. Anderson, R., and Gómez-Gualdrón, D.A. (2019). Increasing topological diversity during computational “synthesis” of porous crystals: how and why. *CrystEngComm* *21*, 1653–1665. <https://doi.org/10.1039/C8CE01637B>.
100. Anderson, G., Schweitzer, B., Anderson, R., and Gómez-Gualdrón, D.A. (2019). Attainable Volumetric Targets for Adsorption-Based Hydrogen Storage in Porous Crystals: Molecular Simulation and Machine Learning. *J. Phys. Chem. C* *123*, 120–130. <https://doi.org/10.1021/acs.jpcc.8b09420>.
101. Bao, Y., Martin, R.L., Haranczyk, M., and Deem, M.W. (2015). In silico prediction of MOFs with high deliverable capacity or internal surface area. *Phys. Chem. Chem. Phys.* *17*, 11962–11973. <https://doi.org/10.1039/C5CP00002E>.
102. Burner, J., Luo, J., White, A., Mirmiran, A., Kwon, O., Boyd, P.G., Maley, S., Gibaldi, M., Simrod, S., Ogden, V., and Woo, T.K. (2023). ARC-MOF: A Diverse Database of Metal–Organic Frameworks with DFT-Derived Partial Atomic Charges and Descriptors for Machine Learning. *Chem. Mater.* *35*, 900–916. <https://doi.org/10.1021/acs.chemmater.2c02485>.
103. Weininger, D., Weininger, A., and Weininger, J.L. (1989). SMILES. 2. Algorithm for generation of unique SMILES notation. *J. Chem. Inf. Comput. Sci.* *29*, 97–101. <https://doi.org/10.1021/ci00062a008>.
104. Weininger, D. (1988). SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Comput. Sci.* *28*, 31–36. <https://doi.org/10.1021/ci00057a005>.
105. Krenn, M., Häse, F., Nigam, A., Friederich, P., and Aspuru-Guzik, A. (2020). Self-referencing embedded strings (SELFIES): A 100% robust molecular string representation. *Mach. Learn. Sci. Technol.* *1*, 045024. <https://doi.org/10.1088/2632-2153/aba947>.
106. Krenn, M., Ai, Q., Barthel, S., Carson, N., Frei, A., Frey, N.C., Friederich, P., Gaudin, T., Gayle, A.A., Jablonka, K.M., et al. (2022). SELFIES and the future of molecular string representations. *Patterns* *3*, 100588. <https://doi.org/10.1016/j.patter.2022.100588>.
107. Nandy, A., Terrones, G., Arunachalam, N., Duan, C., Kastner, D.W., and Kulik, H.J. (2022). MOFSimplify, machine learning models with extracted stability data of three thousand metal-organic frameworks. *Sci. Data* *9*, 74. <https://doi.org/10.1038/s41597-022-01181-0>.
108. Zhang, Z., Pan, F., Mohamed, S.A., Ji, C., Zhang, K., Jiang, J., and Jiang, Z. (2024). Accelerating Discovery of Water Stable Metal–Organic Frameworks by Machine Learning. *Small* *20*, 2405087. <https://doi.org/10.1002/smll.202405087>.
109. Batra, R., Chen, C., Evans, T.G., Walton, K.S., and Ramprasad, R. (2020). Prediction of water stability of metal–organic frameworks using machine learning. *Nat. Mach. Intell.* *2*, 704–710. <https://doi.org/10.1038/s42256-020-00249-z>.

110. Janet, J.P., and Kulik, H.J. (2017). Resolving Transition Metal Chemical Space: Feature Selection for Machine Learning and Structure–Property Relationships. *J. Phys. Chem. A* *121*, 8939–8954. <https://doi.org/10.1021/acs.jpca.7b08750>.
111. Ioannidis, E.I., Gani, T.Z.H., and Kulik, H.J. (2016). molSimplify: A toolkit for automating discovery in inorganic chemistry. *J. Comput. Chem.* *37*, 2106–2117. <https://doi.org/10.1002/jcc.24437>.
112. Janet, J.P., Duan, C., Yang, T., Nandy, A., and Kulik, H.J. (2019). A quantitative uncertainty metric controls error in neural network-driven chemical discovery. *Chem. Sci.* *10*, 7913–7922. <https://doi.org/10.1039/C9SC02298H>.
113. Ongari, D., Yakutovich, A.V., Talirz, L., and Smit, B. (2019). Building a Consistent and Reproducible Database for Adsorption Evaluation in Covalent–Organic Frameworks. *ACS Cent. Sci.* *5*, 1663–1675. <https://doi.org/10.1021/acscentsci.9b00619>.
114. Widom, B. (1963). Some topics in the theory of fluids. *J. Chem. Phys.* *39*, 2808–2812. <https://doi.org/10.1063/1.1734110>.
115. Mayo, S.L., Olafson, B.D., and Goddard, W.A. (1990). DREIDING: a generic force field for molecular simulations. *J. Phys. Chem.* *94*, 8897–8909. <https://doi.org/10.1021/j100389a010>.
116. Rappe, A.K., Casewit, C.J., Colwell, K.S., Goddard, W.A., III, and Skiff, W.M. (1992). UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *J. Am. Chem. Soc.* *114*, 10024–10035. <https://doi.org/10.1021/ja00051a040>.
117. Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W., and Klein, M.L. (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* *79*, 926–935. <https://doi.org/10.1063/1.445869>.
118. Lennard-Jones, J.E. (1931). Cohesion. *Proc. Phys. Soc.* *43*, 461–482. <https://doi.org/10.1088/0959-5309/43/5/301>.
119. Lorentz, H.A. (1881). Ueber die Anwendung des Satzes vom Virial in der kinetischen Theorie der Gase. *Ann. Phys.* *248*, 127–136. <https://doi.org/10.1002/andp.18812480110>.
120. Panagiotopoulos, A.Z. (1987). Direct determination of phase coexistence properties of fluids by Monte Carlo simulation in a new ensemble. *Mol. Phys.* *61*, 813–826. <https://doi.org/10.1080/00268978700101491>.
121. Panagiotopoulos, A.Z., Quirke, N., Stapleton, M., and Tildesley, D.J. (1988). Phase equilibria by simulation in the Gibbs ensemble. *Mol. Phys.* *63*, 527–545. <https://doi.org/10.1080/00268978800100361>.
122. Siepmann, J.I., Martin, M.G., Chen, B., Wick, C., Stubbs, J.M., Potoff, J. J., Eggimann, B.L., McGrath, M.J., Zhao, X.S., Anderson, K.E., et al. (2020). Monte Carlo for Complex Chemical Systems–Minnesota, version 20.2. University of Minnesota.
123. Addicoat, M.A., Vankova, N., Akter, I.F., and Heine, T. (2014). Extension of the Universal Force Field to Metal–Organic Frameworks. *J. Chem. Theory Comput.* *10*, 880–891. <https://doi.org/10.1021/ct400952t>.
124. Maitland, G.C. (1981). Intermolecular Forces: Their Origin and Determination, International Series of Monographs on Chemistry No. 3.
125. Allen, M.P., and Tildesley, D.J. (2017). *Computer Simulation of Liquids* (Oxford university press).
126. Boyd, P.G., Moosavi, S.M., Witman, M., and Smit, B. (2017). Force-Field Prediction of Materials Properties in Metal–Organic Frameworks. *J. Phys. Chem. Lett.* *8*, 357–363. <https://doi.org/10.1021/acs.jpclett.6b02532>.
127. Siepmann, J.I., and Frenkel, D. (1992). Configurational bias Monte Carlo: a new sampling scheme for flexible chains. *Mol. Phys.* *75*, 59–70. <https://doi.org/10.1080/00268979200100061>.
128. Potoff, J.J., and Siepmann, J.I. (2001). Vapor–liquid equilibria of mixtures containing alkanes, carbon dioxide, and nitrogen. *AIChE J.* *47*, 1676–1682. <https://doi.org/10.1002/aic.690470719>.
129. Dubbeldam, D., Torres-Knoop, A., and Walton, K.S. (2013). On the inner workings of Monte Carlo codes. *Mol. Simul.* *39*, 1253–1292. <https://doi.org/10.1080/08927022.2013.819102>.
130. Dubbeldam, D., Calero, S., Ellis, D.E., and Snurr, R.Q. (2016). RASPA: molecular simulation software for adsorption and diffusion in flexible nanoporous materials. *Mol. Simul.* *42*, 81–101. <https://doi.org/10.1080/08927022.2015.1010082>.
131. Li, Z., Shi, K., Dubbeldam, D., Dewing, M., Knight, C., Vázquez-Mayagoitia, Á., and Snurr, R.Q. (2024). Efficient Implementation of Monte Carlo Algorithms on Graphical Processing Units for Simulation of Adsorption in Porous Materials. *J. Chem. Theory Comput.* *20*, 10649–10666. <https://doi.org/10.1021/acs.jctc.4c01058>.
132. Nguyen, T.T.T., Lin, J.-B., Shimizu, G.K.H., and Rajendran, A. (2022). Separation of CO₂ and N₂ on a hydrophobic metal organic framework CALF-20. *Chem. Eng. J.* *442*, 136263. <https://doi.org/10.1016/j.ccej.2022.136263>.
133. Wilkins, N.S., and Rajendran, A. (2019). Measurement of competitive CO₂ and N₂ adsorption on Zeolite 13X for post-combustion CO₂ capture. *Adsorption* *25*, 115–133. <https://doi.org/10.1007/s10450-018-00004-2>.
134. Peh, S.B., Farooq, S., and Zhao, D. (2022). A metal-organic framework (MOF)-based temperature swing adsorption cycle for postcombustion CO₂ capture from wet flue gas. *Chem. Eng. Sci.* *250*, 117399. <https://doi.org/10.1016/j.ces.2021.117399>.
135. Peh, S.B., Farooq, S., and Zhao, D. (2023). Techno-economic analysis of MOF-based adsorption cycles for postcombustion CO₂ capture from wet flue gas. *Chem. Eng. Sci.* *268*, 118390. <https://doi.org/10.1016/j.ces.2022.118390>.
136. Leperi, K.T., Snurr, R.Q., and You, F. (2016). Optimization of Two-Stage Pressure/Vacuum Swing Adsorption with Variable Dehydration Level for Postcombustion Carbon Capture. *Ind. Eng. Chem. Res.* *55*, 3338–3350. <https://doi.org/10.1021/acs.iecr.5b03122>.
137. Eymard, R., Gallouët, T., and Herbin, R. (2000). Finite volume methods. In *Handbook of Numerical Analysis*, Du Qiang, Hintermüller Michael, Mishra Siddhartha, and Townsend Alex, eds. (Elsevier), pp. 713–1018. [https://doi.org/10.1016/S1570-8659\(00\)07005-8](https://doi.org/10.1016/S1570-8659(00)07005-8).
138. Liu, X.-D., Osher, S., and Chan, T. (1994). Weighted Essentially Non-oscillatory Schemes. *J. Comput. Phys.* *115*, 200–212. <https://doi.org/10.1006/jcph.1994.1187>.
139. Shampine, L.F., and Reichelt, M.W. (1997). The MATLAB ODE Suite. *SIAM J. Sci. Comput.* *18*, 1–22. <https://doi.org/10.1137/s1064827594276424>.
140. Bradford, E., Schweidtmann, a.m., and Lapkin, A. (2018). Efficient multiobjective optimization employing Gaussian processes, spectral sampling and a genetic algorithm. *J. Glob. Optim.* *71*, 407–438. <https://doi.org/10.1007/s10898-018-0609-2>.
141. The MathWorks, I. (2023). *MATLAB Version: 9.14.0 (R2023a) (The MathWorks Inc.)*.
142. Ward, A., and Pini, R. (2022). Efficient Bayesian Optimization of Industrial-Scale Pressure–Vacuum Swing Adsorption Processes for CO₂ Capture. *Ind. Eng. Chem. Res.* *61*, 13650–13668. <https://doi.org/10.1021/acs.iecr.2c02313>.
143. Kresse, G., and Joubert, D. (1999). From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B* *59*, 1758–1775. <https://doi.org/10.1103/PhysRevB.59.1758>.
144. Kresse, G., and Furthmüller, J. (1996). Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B* *54*, 11169–11186. <https://doi.org/10.1103/PhysRevB.54.11169>.
145. Perdew, J.P., Burke, K., and Ernzerhof, M. (1996). Generalized gradient approximation made simple. *Phys. Rev. Lett.* *77*, 3865–3868. <https://doi.org/10.1103/PhysRevLett.77.3865>.
146. Grimme, S., Antony, J., Ehrlich, S., and Krieg, H. (2010). A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H–Pu. *J. Chem. Phys.* *132*, 154104. <https://doi.org/10.1063/1.3382344>.

147. Grimme, S., Ehrlich, S., and Goerigk, L. (2011). Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* 32, 1456–1465. <https://doi.org/10.1002/jcc.21759>.
148. Manz, T.A., and Limas, N.G. (2016). Introducing DDEC6 atomic population analysis: part 1. Charge partitioning theory and methodology. *RSC Adv.* 6, 47771–47801. <https://doi.org/10.1039/C6RA04656H>.
149. Limas, N.G., and Manz, T.A. (2016). Introducing DDEC6 atomic population analysis: part 2. Computed results for a wide range of periodic and nonperiodic materials. *RSC Adv.* 6, 45727–45747. <https://doi.org/10.1039/C6RA05507A>.
150. Manz, T.A. (2017). Introducing DDEC6 atomic population analysis: part 3. Comprehensive method to compute bond orders. *RSC Adv.* 7, 45552–45581. <https://doi.org/10.1039/C7RA07400J>.
151. Limas, N.G., and Manz, T.A. (2018). Introducing DDEC6 atomic population analysis: part 4. Efficient parallel computation of net atomic charges, atomic spin moments, bond orders, and more. *RSC Adv.* 8, 2678–2707. <https://doi.org/10.1039/C7RA11829E>.
152. Miklitz, M., and Jelfs, K.E. (2018). pywindow: Automated Structural Analysis of Molecular Pores. *J. Chem. Inf. Model.* 58, 2387–2391. <https://doi.org/10.1021/acs.jcim.8b00490>.