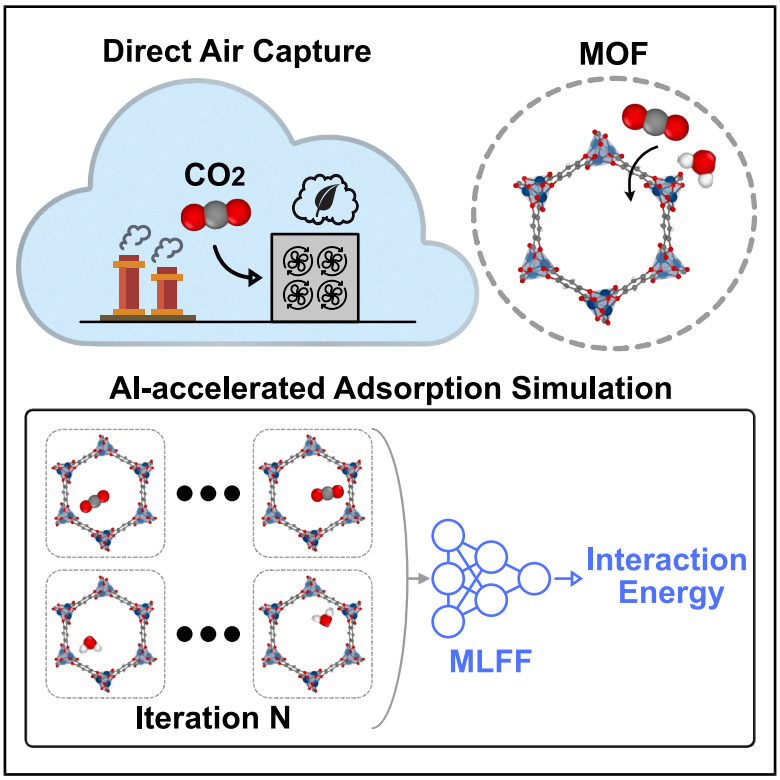
Matter

Article

Accelerating CO_2 direct air capture screening for metal-organic frameworks with a transferable machine learning force field

Graphical abstract



Highlights

- MLFF improves modeling CO_2 and H_2O interaction in MOFs
- Enhanced predictions with significantly lower computational cost can be achieved
- Curated dataset and continual learning strategy improve the fine-tuning process
- More than 100 MOFs missed by conventional methods were found suitable for DAC

Authors

Yunsung Lim, Hyunsoo Park, Aron Walsh, Jihan Kim

Correspondence

a.walsh@imperial.ac.uk (A.W.), jihankim@kaist.ac.kr (J.K.)

In brief

DAC of CO_2 using porous materials holds promise for achieving net-zero emissions but is hindered by slow, inaccurate simulations. We integrate machine learning models into simulation workflows to enable fast and accurate simulations, allowing rapid screening of thousands of materials. This framework is integrated into a user-friendly Python package, offering a scalable approach for data-driven materials discovery in carbon capture and related energy and environmental applications.

Improvement
Enhanced performance with innovative design or material control

Lim et al., 2025, Matter *8*, 102203
July 2, 2025
https://doi.org/10.1016/j.matt.2025.102203

CellPress

Please cite this article in press as: Lim et al., Accelerating CO_2 direct air capture screening for metal-organic frameworks with a transferable machine learning force field, Matter (2025), https://doi.org/10.1016/j.matt.2025.102203

Matter

CellPress

Article

Accelerating CO_2 direct air capture screening for metal-organic frameworks with a transferable machine learning force field

Yunsung Lim,[1,3] Hyunsoo Park,[2,3] Aron Walsh,[2,*] and Jihan Kim[1,4,*]
[1]Department of Chemical and Biomolecular Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141, Republic of Korea
[2]Department of Materials, Imperial College London, London SW7 2AZ, UK
[3]These authors contributed equally
[4]Lead contact
*Correspondence: a.walsh@imperial.ac.uk (A.W.), jihankim@kaist.ac.kr (J.K.)
https://doi.org/10.1016/j.matt.2025.102203

PROGRESS AND POTENTIAL Capturing CO_2 directly from ambient air, known as direct air capture (DAC), is essential for achieving net-zero emissions. Metal-organic frameworks (MOFs) are promising porous solids for DAC, but existing computational methods either take a long time or lack the accuracy needed to model CO_2 and H_2O adsorption within the MOFs. As a result, high-throughput screenings via computational simulation struggle to identify the most effective materials for DAC. To overcome this challenge, we developed an artificial intelligence-accelerated adsorption simulation framework built on a machine learning force field (MLFF) that captures MOF-CO_2 and MOF-H_2O interactions both quickly and accurately. It enabled the screening of more than 8,000 MOFs, uncovering more than 100 high-potential candidates that conventional simulations missed. This computational platform can easily be adopted by researchers and industry to accelerate the discovery of porous materials for DAC and other gas separation needs. Our approach could enable more efficient, cost-effective carbon capture technologies and support global emission reduction goals.

SUMMARY

Direct air capture (DAC) of CO_2 is necessary for climate change mitigation, but it faces challenges from low CO_2 concentrations and competition from water vapor. Metal-organic frameworks (MOFs) hold significant promise for DAC owing to their high surface area and adsorption-based capture processes. However, identifying optimal MOFs is hindered by structural complexity and vast chemical diversity. Here, we introduced a machine learning force field (MLFF) tailored for CO_2 and H_2O interactions in MOFs by fine-tuning a foundation model. To address smoothing issues and catastrophic forgetting, we curated the GoldDAC dataset and introduced a continual learning scheme. We further developed DAC-SIM, a molecular simulation package integrated with MLFF, including a Widom insertion. Then, we screened an extensive MOF database, uncovering high-performing MOFs and identifying chemical features for DAC applications. This approach overcomes prior limitations in describing MOF-CO_2 and MOF-H_2O interactions, providing a scalable and accurate framework for DAC research of porous materials.

INTRODUCTION

The average global temperature has been steadily rising since the late 19th century and was approximately 1.2°C degree warmer in 2023 than the pre-industrial average.[1] This warming is largely driven by greenhouse gas emissions, with CO_2 accounting for more than 60% of the effect.[2] Among various strategies to manage CO_2 emission levels, direct air capture (DAC) offers a solution by capturing CO_2 directly from ambient air.[3,4] This process is challenging due to trace amounts of CO_2 concentration (approximately 400 ppm) and low selective capturing of CO_2 over atmospheric water vapor.

To date, DAC technologies have predominantly relied on chemisorption of CO_2 using amines and solid alkali hydroxides, due to their strong affinity for CO_2.[5,6] However, the regeneration of adsorbents is the dominant source of the total operational costs.[7,8] To mitigate these issues, physisorption-based materials with weaker CO_2 interactions have been explored to reduce

Please cite this article in press as: Lim et al., Accelerating CO_2 direct air capture screening for metal-organic frameworks with a transferable machine learning force field, Matter (2025), https://doi.org/10.1016/j.matt.2025.102203

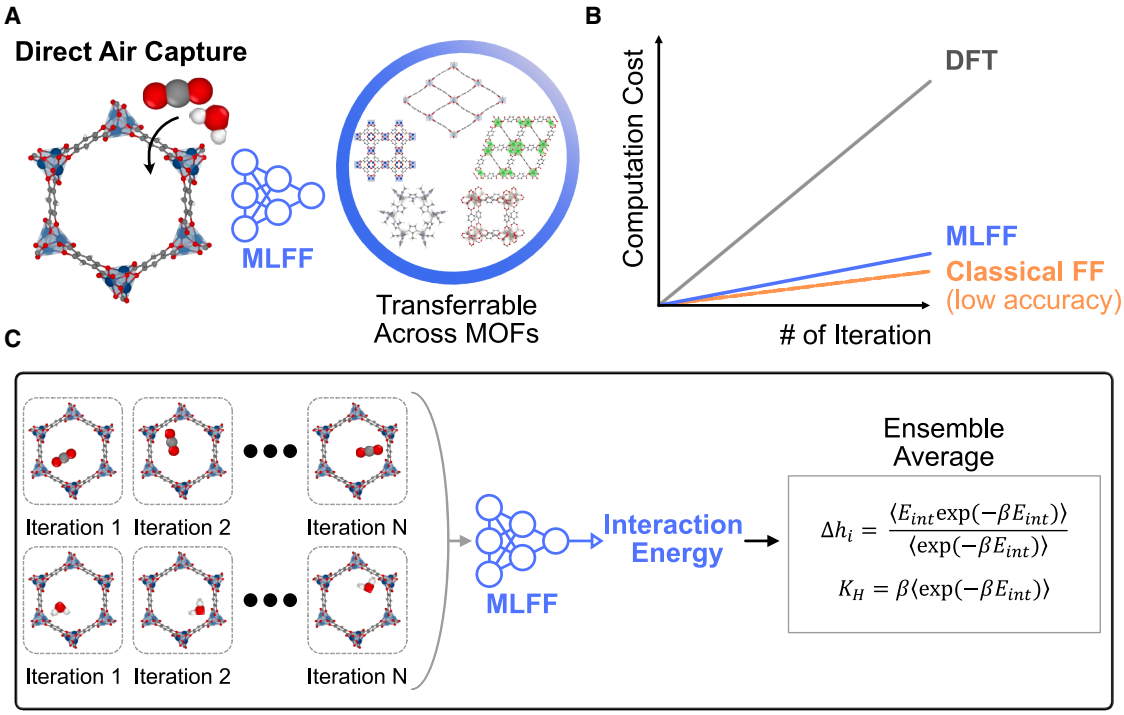


Figure 1. Overview of our computational approach for CO_2 capture implemented in DAC-SIM

(A) Schematic representation of the transferable MLFF across MOFs for DAC applications.

(B) Comparison of computational cost in terms of the number of iterations required in MC simulations for DFT, MLFF, and classical force fields.

(C) Workflow for calculating the Q_{st} and K_H using MC simulations, leveraging Widom insertion integrated with MLFFs.

regeneration costs with repeated usage.[9] Among these, metal-organic frameworks (MOFs) stand out due to their high surface area and remarkable tunable properties.[10,11] However, the chemical space for forming MOFs is vast due to the large number of possible chemical building blocks and topologies.

To expedite the identification and design of promising MOFs for CO_2 capture, large-scale screening has been widely applied.[12–17] These are typically based on classical molecular simulations with force fields, such as universal force fields (UFFs)[18] with point charges (e.g., Qeq,[19] EQeq, [20] and density-derived electrostatic and chemical [DDEC][21]). For CO_2 capture, Boyd et al.[15] designed new MOFs suitable for flue gas CO_2 capture by introducing CO_2-responsive pockets acquired from screening results. Findley and Sholl[16] screened the CoRE MOF database to evaluate whether MOFs are suitable for DAC application. While these methods are time efficient, they struggle to accurately describe the interaction between the MOF frameworks and the gas molecules, such as CO_2 and H_2O, due to the diversity of chemical environments and the complexity of the short-range potential energy landscape.[22,23] While *ab initio* calculations offer a more accurate alternative by explicitly accounting for electron density distributions and have been shown to closely match experimental adsorption behavior in many cases,[22,24] the computational cost increases substantially and are impractical for large-scale screening or extensive data accumulation for porous solids.

Machine learning force fields (MLFFs) have come to the forefront to tackle this challenge (Figures 1A and 1B), combining the low computational cost of classical force fields with the high accuracy of *ab initio* methods.[25–27] While various materials including alloys, metal oxides, and perovskites have been widely studied,[28–30] the applications of MLFF for MOFs have still been limited, especially for heterogeneous systems that contain the MOF material itself and the guest molecules. Although a few studies have explored the development of MLFFs for the combined system of MOF with gas molecules such as CO_2 or H_2,[31–34] the transferability across different MOF systems remains significantly limited. This limitation is primarily due to the lack of comprehensive databases that contain density functional theory (DFT) simulation data involving the combined system of MOFs with gas molecules.

In this work, we introduce an MLFF that is both accurate for DAC applications and transferable across a wide range of MOFs, as shown in Figure 1A. We used the term transferable to highlight that the MLFF models in this work are applicable to various MOFs for adsorption modeling, especially for CO_2 and H_2O. By including configurations of MOFs interacting with CO_2 and H_2O gas molecules contained in our curated GoldDAC dataset, we fine-tuned a foundation model[35] that was pre-trained on diverse crystalline compounds. To deploy this model, we have developed the DAC-SIM package, which combines our MLFF with the statistical Widom insertion method,[36] molecular dynamics, and structure optimization. We used DAC-SIM to conduct high-throughput screening of 8,131 experimental MOFs in the CoRE MOF database.[37] Comparison of the

Please cite this article in press as: Lim et al., Accelerating CO_2 direct air capture screening for metal-organic frameworks with a transferable machine learning force field, Matter (2025), https://doi.org/10.1016/j.matt.2025.102203

distributions in CO_2 adsorption and selectivity highlights several limitations in the traditional screening approach and we identify several novel candidates and design principles.

RESULTS

Monte Carlo simulation for DAC application

A key challenge for DAC lies in identifying materials that exhibit strong CO_2 affinity and high selectivity over H_2O. To address this, Monte Carlo (MC) simulations such as Widom insertion and grand canonical MC (GCMC) under rigid framework assumptions are used to determine gas adsorption properties of materials by calculating the heat of adsorption (Q_{st}) and Henry's law coefficient (K_H).[38] These ensemble-averaged properties enable a more comprehensive evaluation than relying solely on single configuration properties, such as adsorption energies at the most likely stable binding sites.

The Widom insertion method, as outlined in Figure 1C, is well suited for DAC under the condition of low CO_2 concentration,[16,39,40] where modeling a dilute system is applicable. The simulation involves numerous individual calculations to build the ensemble, where single gas molecules are randomly inserted into MOF frameworks to compute the ensemble-averaged interaction energy. The interaction energy E_{int} is determined from:

$$E_{int} = E(system) - E(MOF) - E(gas),$$

where $E(system)$ represents the potential energy of the MOF with the adsorbed gas, and $E(MOF)$ and $E(gas)$ correspond with the potential energies of the MOF framework and gas molecules (e.g., CO_2 or H_2O), respectively. This approach allows for the calculation of both Q_{st} and K_H from statistical mechanics:

$$\Delta H_{ads} = -Q_{st} = \frac{\langle E_{int} \exp(-\beta E_{int})\rangle}{\langle \exp(-\beta E_{int})\rangle} - k_\beta T$$

$$K_H = \beta \langle \exp(-\beta E_{int})\rangle,$$

where ΔH_{ads} is the enthalpy of adsorption, β represents the inverse of the thermal energy, defined as $\beta = \frac{1}{k_\beta T}$, where k_β is the Boltzmann constant and T is the temperature. Without the contribution from the thermal term $k_\beta T$ in Q_{st}, the value represents the ensemble-averaged interaction energy, denoted as Δh_i.

Fine-tuning of foundation model

A variety of advanced MLFFs (e.g., MACE,[41] CHGNet,[42] M3GNet,[43] and SevenNet[44]) have been developed that accurately reproduce the DFT results for a diverse set of crystalline materials. These models have undergone extensive pre-training on datasets with total DFT energies (E) and atomic forces (F), mainly sampling from geometry optimization trajectories such as MPtrj.[42] Among these, MACE-MP-0 model (referred to hereafter as MACE-MP for simplicity)[35] stands out with its combination of the atomic cluster expansion for structure representation with equivariant and high body-order message passing graph neural networks for regression. The pre-training datasets for these models are dominated by inorganic compositions with fewer organic and organic-inorganic solids; however, the MACE-MP model has shown reasonable preliminary results on MOF systems.[35] Therefore, we have built on this work to fine-tune a transferable MLFF, termed MACE-DAC, tailored to MOFs and their interaction with relevant gases.

GoldDAC dataset

One of the primary challenges in fine-tuning models is the risk of severe overfitting to the limited fine-tuning dataset. When fine-tuning the pre-trained foundation models for MLFFs, Deng et al.[45] highlight a critical issue referred to as smoothing (or softening), which refers to the degradation in a model's ability to accurately predict the non-equilibrium regions of potential energy surface (PES), as shown in Figure 2A. It arises from the reason that the training data used for MLFF development is predominantly concentrated around 0K equilibrium configurations (hereafter referred to as equilibrium configurations), which correspond with the local minimum on the PES. The smoothing becomes more pronounced during fine-tuning than the pre-training stage, as this process typically involves updating the pre-training model with a limited dataset.

Accurately predicting the entire PES is particularly crucial for calculating gas adsorption properties via the Widom insertion method. Properties such as Q_{st} and K_H values are sensitive to small errors in the PES, especially in regions where the interaction energy is less than zero. These errors, stemming from the exponential weighting of energy terms, can lead to significant deviations in predicted adsorption properties. For instance, while equilibrium regions of the PES are generally well modeled, inaccuracies in the repulsion region can alter the slope of the PES, leading to notable errors in derived thermodynamic properties.

To overcome this challenge, we created the GoldDAC dataset that involves DFT results from CO_2 + MOF, H_2O + MOF, MOF, and gas molecule calculations. This dataset enables fine-tuning to enhance model performance across the entire PES, encompassing not only equilibrium regions but also the repulsive and weak-attraction regions. The training and validation sets were constructed from 60 structures, with three structures randomly sampled for each of the 20 most frequently occurring metals in the ODAC23 dataset.[39] The ODAC23 database is a comprehensive database including adsorption calculations of an MOF system for DAC applications at the DFT level.

Since the interaction energy depends not only on the energy of the combined system (the MOF with gas molecule) but also on the individual energies of the framework and the gas molecules (e.g., CO_2 and H_2O), both configurations were included in the training and validation sets. For robust training, PES data from two of these structures were randomly allocated to the training set, while data from the remaining structure were used for validation. Since the dataset was developed for DAC application, the test set comprises 26 MOFs identified in the literature as promising for CO_2 capture. Overall, as shown in Table S1, 784 DFT calculations (720 for MOF + gas, 60 for MOF, and 4 for gas) were conducted to construct the training and validation sets, and an additional 312 DFT calculations were conducted to

Please cite this article in press as: Lim et al., Accelerating CO_2 direct air capture screening for metal-organic frameworks with a transferable machine learning force field, Matter (2025), https://doi.org/10.1016/j.matt.2025.102203

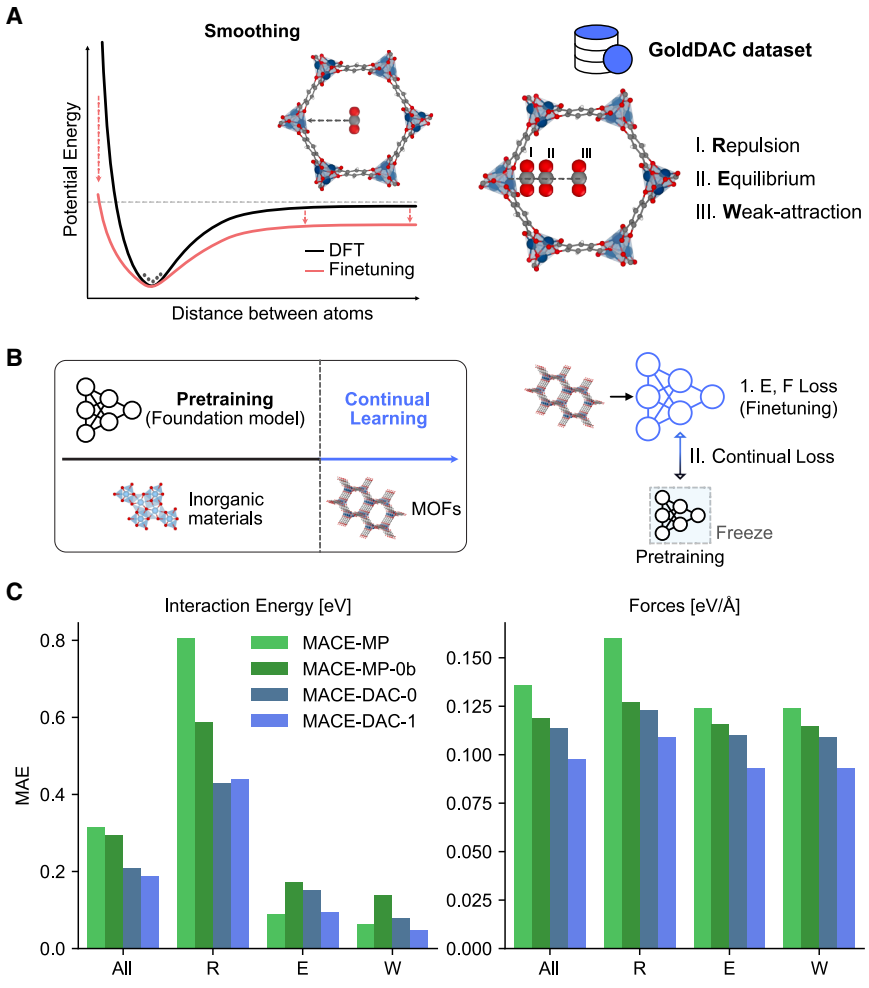


Figure 2. Fine-tuning strategy for the MACE-DAC force field

(A) Smoothing problem during the fine-tuning stage, highlighting the GoldDAC dataset which samples repulsive (R), equilibrium (E), and weak-attraction near pore center (W) regions of the PES.

(B) Continual learning framework transitioning from pre-training on inorganic materials to MOFs while preserving the pre-training model's knowledge.

(C) MAE values for the test set of the GoldDAC dataset, comparing the performance of MACE-MP, MACE-MP-0b, MACE-DAC-0 (fine-tuning), and MACE-DAC-1 (continual learning) models in terms of interaction energy and force predictions.

construct a test set. Details of data construction for GoldDAC dataset are provided in the Methods section.

Continual learning

Figure 2C presents the performance evaluation of the fine-tuning model using the GoldDAC dataset. This analysis encompasses both the foundational MACE-MP model and its variant, MACE-MP-0b that incorporates the analytic Ziegler-Biersack-Littmark repulsive potential to remove pathologies in the PES at short distances.[46] The mean absolute error (MAE) for the average of total PES data, denoted as 'All' in Figures 2C and Table 1, exhibits a notable reduction compared with the original foundation models with respect to interaction energy and force predictions. This improvement is mainly attributed to enhance predictive accuracy in the repulsive region of the fine-tuning model, assisted by configurations sampled by GoldDAC dataset.

The fine-tuning model does decline in performance within the equilibrium and weak attraction regions compared with the pre-training MACE-MP models. This degradation can be attributed to the complexity in the interaction energy, which is not determined solely by the combined system's total energy but by its energy relative to the isolated framework and gas molecule. The fine-tuning process can disrupt the delicate balance originally established with biases toward the data represented in the fine-tuning dataset. This issue aligns with the phenomenon of "catastrophic forgetting," wherein knowledge acquired during pre-training is disrupted or overwritten during the subsequent fine-tuning process.[47,48]

To remedy this issue, we implemented a refined fine-tuning strategy with continual learning loss.[49,50] This additional loss term is designed to preserve the pre-training model's knowledge during the fine-tuning. Specifically, the continual learning loss is implemented by freezing the weights of the pre-training model and minimizing discrepancies in energy and force predictions between the pre-training and the fine-tuning model during the fine-tuning process, as illustrated in Figure 2B. This approach ensures that the fine-tuning process does not introduce significant deviations from the pre-training model, thereby ensuring the original energy balance between components in the interaction energy calculation.

By introducing this regularization loss, the fine-tuning process is guided to adapt more gradually to the new MOF data while retaining the knowledge learned from pre-training on inorganic materials. As shown in Figure 2C, this method prevents performance degradation in the equilibrium and weak-attraction regions, observed in the fine-tuning model without the continual learning loss (termed as MACE-DAC-0), while maintaining the trend that significantly reduces loss in the repulsion region. Therefore, the continual learning model (termed as MACE-DAC-1) achieved the lowest MAE across all PES data. Furthermore, continual learning contributes to a stable reduction in force prediction loss, thereby enabling the model to achieve the highest performance in force predictions as well. In MACE-DAC-1, some configurations with 3D transition metals, open metal sites, and repulsion regions showed high force errors, indicating still minor challenges in capturing unusual chemical environments. However, our MLFF outperformed the MACE-MP

Please cite this article in press as: Lim et al., Accelerating CO_2 direct air capture screening for metal-organic frameworks with a transferable machine learning force field, Matter (2025), https://doi.org/10.1016/j.matt.2025.102203

Table 1. Performance evaluation on the GoldDAC test set

Model	Property	All	Repulsion	Equilibrium	Weak attraction
MACE-MP	interaction energy	0.315	0.805	**0.090**	*0.063*
	force	0.136	0.160	0.124	0.124
MACE-MP-0b	interaction energy	0.294	0.587	0.173	0.140
	force	0.119	0.127	0.116	0.115
MACE-DAC-0 (fine-tuning)	interaction energy	*0.208*	**0.429**	0.151	0.079
	force	*0.114*	*0.123*	*0.110*	*0.109*
MACE-DAC-1 (continual learning)	interaction energy	**0.188**	*0.441*	*0.096*	**0.049**
	force	**0.098**	**0.109**	**0.093**	**0.093**

MAE values for MACE-MP, MACE-MP-0b, MACE-DAC-0 (fine-tuning), and MACE-DAC-1 (continual learning) models are calculated with the GoldDAC test set, evaluated in terms of interaction energy and forces. The units of MAEs are provided as eV and eV/Å.
The best-performing value is indicated in bold, and the second-best (runner-up) is indicated in *italic*.

model, demonstrating its practical advantages (Figures S3 and S4). The corresponding parity plots and MAE values for all four models are provided in Figures S2–S5 and Table S2, respectively.

As shown in Figure S5, the MACE-DAC-0 model shows lower MAE for the total energy of the system. Nonetheless, the MACE-DAC-0 model underperforms compared with the MACE-DAC-1 model in terms of interaction energy. Continual learning regularization effectively maintains the balance among different energy components during fine-tuning.

High-throughput MOF screening

To facilitate the deployment of MLFFs for DAC applications, we have developed a Python package, DAC-SIM, to perform the necessary molecular simulations in a computationally efficient manner. Using this package, Widom insertion simulations were conducted to calculate the Q_{st} at zero loading and K_H for structures within the CoRE MOF database. During the simulations, the framework atoms were fixed.

Initially, 8,131 MOFs were selected for Widom insertion simulation, after excluding MOFs containing more than 250 atoms in their unit cells for computational efficiency during large-scale screening. Subsequently, geometry optimizations were carried out using MACE-DAC-1, where the details of geometry optimization are provided in the methods section. Following these optimizations, Q_{st} and K_H values were calculated by the Widom insertion simulations for 6,461 MOFs, after excluding MOFs encountering convergence issues and consistency problems (where the standard deviation from two repeated folds of averaged interaction energy calculations exceeded 0.05 eV, detailed analysis on statistical noise was provided in Note S2 and Figure S30).

Figure 3A shows the calculated Q_{st} values for CO_2 and H_2O, which are predominantly located between 20 and 40 kJ/mol. Previous studies suggest that MOFs with CO_2 Q_{st} exceeding 50 kJ/mol are regarded as promising for DAC application within physisorption mechanism.[16,39] In line with this criterion, it is observed that 364 MOFs (5.63%) exhibit CO_2 Q_{st} values exceeding 50 kJ/mol, and 167 of them even surpass those for H_2O. Figure 3B presents the relationship between Q_{st} for CO_2 and the selectivity derived from the ratio of K_H for CO_2 to H_2O. Although K_H generally follows the trend of Q_{st} as shown in Figure S6, given that selectivity is typically calculated using K_H rather than Q_{st}, this study uses the K_H ratio to define selectivity. Consequently, a total of 161 MOFs were identified as promising candidates for DAC application, characterized by CO_2 Q_{st} values of greater than 50 kJ/mol and demonstrating high selectivity for CO_2 over H_2O. While both Q_{st} and K_H have limitations in capturing the strong H_2O-H_2O interactions that become significant at high relative humidity, our analysis in Figure S8A shows that more than 67% of the candidate materials possess relatively small pores (largest cavity diameter [LCD] of <5.3 Å, approximately twice the kinetic diameter of a H_2O molecule). These confined pore sizes are likely to suppress the formation of water clusters, thereby making Q_{st} and K_H suitable as initial thermodynamic descriptors for screening candidates in DAC applications.

Single interaction energy versus ensemble-averaged property

In contrast with the MC method that computes ensemble-averaged properties for gas adsorption, an alternative approach can use single configuration properties such as the lowest interaction energy at the most favorable binding site within a given MOF framework. This approach, used in the ODAC23,[39] involved calculating the lowest interaction energy (adsorption energy in original paper) at the most favorable binding site within a given MOF framework using DFT. To identify the most stable adsorption site, it begins with generating potential sites through classical force fields or random selection, yielding multiple candidate configurations per structure. MOFs with fewer than four configurations were excluded in this work. Each configuration then undergoes geometry optimization via DFT to identify the most energetically favorable binding site.

Although this approach is effective for identifying the optimal interaction site, it may not represent the overall gas adsorption performance of MOFs. Indeed, the lowest CO_2 interaction energy of 2,794 MOFs derived from the ODAC23 database and the corresponding Q_{st} values for CO_2 calculated by DAC-SIM do not exhibit a strong correlation, as shown in the scatterplot in Figure 3C. The same trend was observed for the H_2O molecule in Figure S7.

The ensemble-averaged gas adsorption properties are influenced not only by chemical environment but also by various geometric features such as accessible pore diameter, void fraction,

Please cite this article in press as: Lim et al., Accelerating CO_2 direct air capture screening for metal-organic frameworks with a transferable machine learning force field, Matter (2025), https://doi.org/10.1016/j.matt.2025.102203

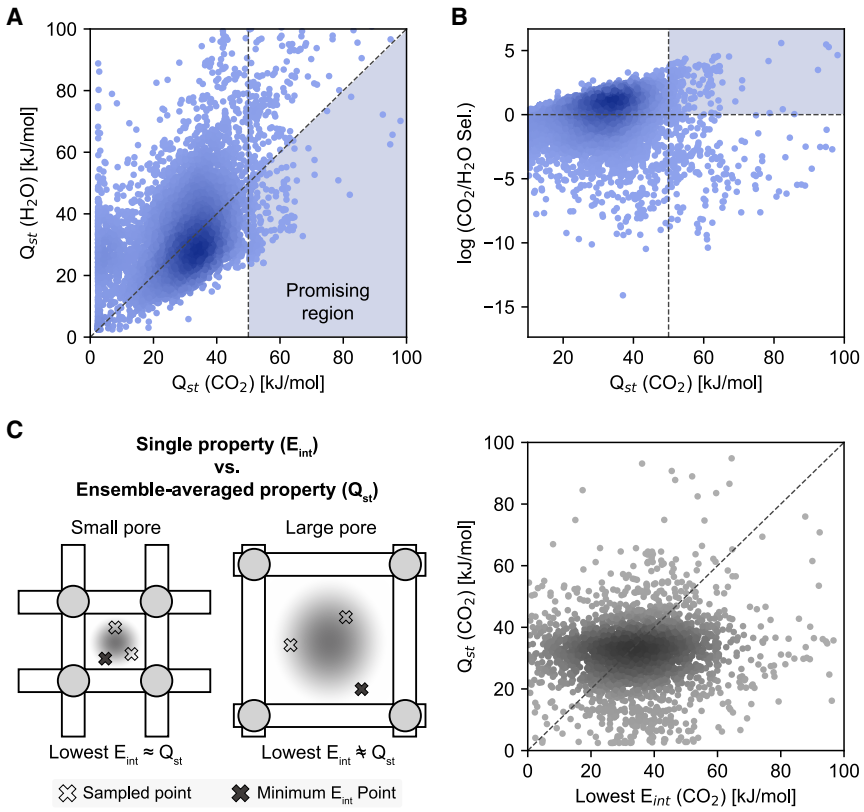


Figure 3. Overall screening results using DAC-SIM

(A and B) Scatterplots showing the relationship between (A) CO_2 Q_{st} and H_2O Q_{st}, and (B) CO_2 Q_{st} and selectivity, with selectivity depicted on a log scale for convenience. The shaded regions in (A) and (B) indicate the promising areas for DAC of CO_2 (CO_2 Q_{st} > 50 kJ/mol and CO_2 Q_{st} > H_2O Q_{st} for [A]; CO_2 Q_{st} > 50 kJ/mol and selectivity >1 for [B]).

(C) A schematic and a scatterplot illustrating discrepancies between the lowest interaction energy and Q_{st}. Q_{st} were calculated using MACE-DAC-1 with DAC-SIM package and the interaction energies were directly obtained from ODAC23 database. Possible insertion areas are schematically depicted as gray shade areas within the pore.

and surface area, which are not accounted for in the single configuration approaches. Figure S8 demonstrates that Q_{st} and selectivity values based on K_H have strong correlations with these geometric features. For instance, when the LCD is close to the kinetic diameter of CO_2 (approximately 3.3 Å), both the CO_2 Q_{st} and CO_2 selectivity tend to be high. Therefore, particularly in MOFs with larger pores, using ensemble-averaged properties is more effective in assessing gas adsorption performance.

Comparison with classical force field simulations

The Widom insertion simulation results derived from our developed MLFF (MACE-DAC-1) using DAC-SIM were systematically compared with those obtained from a classical force field approach. In the classical method, the UFF[18] combined with DDEC charges[21] (hereinafter referred to as UFF + DDEC) were used to simulate the CO_2 and the H_2O adsorption properties in the MOFs.[16,17,51] UFF is interatomic potential consisting of empirical parameters and DDEC charges are derived from DFT calculations. UFF + DDEC were only used for framework atoms and customized force fields provided for CO_2 and H_2O molecules were used to model them. A total of 2,654 MOFs, having the calculated DDEC charges provided by Kancharlapalli et al.[52] were used for this analysis.

Figures 4A, 4B, and S9 present the Q_{st} distributions for CO_2 and H_2O, respectively, as computed using the DAC-SIM package with the developed MLFF and the RASPA software[53] with the UFF + DDEC method. The overall distribution of CO_2 Q_{st} values from DAC-SIM closely resembles that obtained with UFF + DDEC; however, a notable difference is a minor peak at the lower range near 0 kJ/mol. As shown in Figure S10, the LCD distribution of these MOFs exhibiting CO_2 Q_{st} values below 2.5 kJ/mol shifts to a lower range compared with the structures before undergoing geometry optimization. That is, the shrinkage in pore size occurred during the prior geometry optimization stage before Widom insertion simulation, thereby preventing them from accommodating CO_2 molecules. This highlights the advantage of DAC-SIM, enabling computationally efficient geometry optimization with accuracy near that of DFT prior to Widom insertion simulation, by improving force predictions in the fine-tuning stage.

Notably, the distributions of H_2O Q_{st} reveal substantial differences, particularly in the high Q_{st} range. While most MOFs have H_2O Q_{st} values of less than 100 kJ/mol in the DAC-SIM calculations, more than 17.8% of MOFs exceed 100 kJ/mol when using UFF + DDEC. These higher Q_{st} values in the UFF + DDEC calculations are primarily associated with MOFs containing lanthanide metals. Figure S11 shows the most frequently observed metal types in MOFs with H_2O Q_{st} values exceeding 100 kJ/mol in UFF + DDEC calculations, largely comprising lanthanide elements such as Gd, Nd, Tb, La, Er, Pr, Dy, Ce, Sm, Ho, and Tm.

Figure S12 further indicates that there is no significant deviation in distribution between the overall and lanthanide MOFs regarding the lowest interaction energies from the ODAC23 database based on DFT calculations, while a pronounced increase in Q_{st} values is observed for lanthanide MOFs in UFF + DDEC calculations. It suggests that the UFF + DDEC method implemented to the framework atoms tends to overestimate Q_{st} values for H_2O in lanthanide MOFs. Figure S13 illustrates that UFF + DDEC calculations produce significantly higher H_2O Q_{st} values compared with CO_2 for MOFs containing lanthanide metals. In contrast, the DAC-SIM results, similar to the DFT results of interaction from the ODAC23 database, do not exhibit

Please cite this article in press as: Lim et al., Accelerating CO_2 direct air capture screening for metal-organic frameworks with a transferable machine learning force field, Matter (2025), https://doi.org/10.1016/j.matt.2025.102203

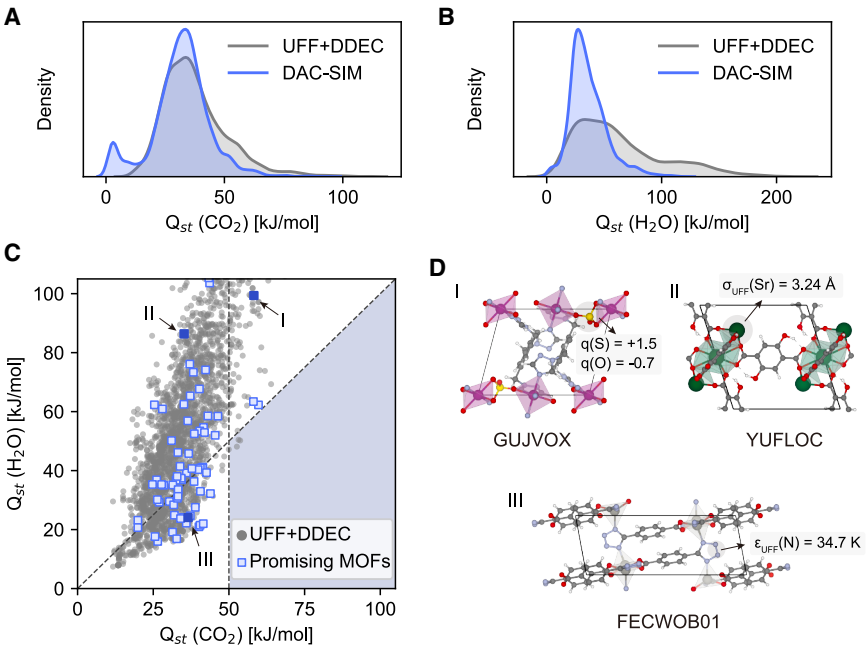


Figure 4. Comparison between DAC-SIM and classical force field calculations

(A and B) Distribution of (A) CO_2 Q_{st} and (B) H_2O Q_{st}, calculated using RASPA[53] (with UFF + DDEC) and DAC-SIM (with MACE-DAC-1).

(C) Q_{st} values of promising MOFs were calculated using the classical force field. Three filled squares denote three representative MOFs shown in (D).

(D) Illustration of the three representative MOFs, highlighting the point charges or UFF parameters of possible active sites for CO_2 and H_2O adsorption. Atom colors: C (gray), H (white), O (red), N (light blue), S (yellow), Mn (purple), Sr (dark green), and Zn (dark gray).

such a pronounced difference between lanthanide-containing MOFs and those with other metals.

This inconsistency between classical force field and DFT calculations arises from the chemically complex inorganic-organic hybrid compositions of the MOFs. Simulating interactions involving metals and neighboring organic atoms (e.g., oxygen or nitrogen atoms bonded to metal) within these hybrid environments requires careful parameter fitting regarding each individual chemical context. However, UFF employs fixed parameters, which limit its ability to accurately represent varying chemical environments. While point charges offer an additional degree of freedom for simulating interactions, UFF was originally developed without incorporating point charges. Thus, it remains uncertain whether point charges effectively compensate for or exacerbate simulation errors.

Given that the unique chemical environments of MOFs are primarily attributed to the presence of metals, the role of metals was investigated. Figure S14 illustrates the distribution of point charges across various metals, revealing that metals with higher point charges tend to exhibit greater discrepancies between UFF + DDEC and DAC-SIM results in Figure S13C. Metals with elevated point charges can function as hard Lewis acids and polarize neighboring atoms nearby metals (typically oxygen or nitrogen atoms within MOFs). This polarization effect decreases the polarizability of adjacent atoms relative to similar atoms in different environments, while the metal's polarizability remains largely unchanged.[22] However, since UFF uses fixed parameters, it is unable to accurately account for this variation in polarizability. Although point charges such as DDEC are generally applied to enhance interaction modeling by providing additional electrostatic potential, they are only partially effective in specific cases and often exacerbate discrepancies. Therefore, the overestimation of Q_{st} in UFF + DDEC calculations is particularly pronounced for H_2O molecule, which can induce stronger electrostatic potential compared with CO_2 molecule.

Due to the overestimation of H_2O Q_{st}, few candidates were identified as suitable for DAC application from the screening of the CoRE MOF database using the classical force field. Likewise, promising MOFs identified in DAC-SIM simulations similarly fail to exhibit high performance using the UFF + DDEC model classical as shown in Figure 4C. We conducted an in-depth investigation into the origin of this difference using three MOFs, whose REFCODEs are GUJVOX ($Mn_2(ttmb)_2(SO_4)_2$), YUFLOC (AEMOF-5), and FECWOB01 (Zn-TBC), as case studies (filled square points in Figure 4C).

First, in the case of GUJVOX, while the CO_2 Q_{st} remained above 50 kJ/mol, the H_2O Q_{st} exhibits a substantial increase in the UFF + DDEC calculation. The presence of the electronegative functional group (SO_4^{2-}) within the framework, which acts as a strong binding site for H_2O, appears to be a major factor in this discrepancy. The point charges for S in SO_4^{2-} are +1.5, while for O in SO_4^{2-}, they are −0.7, indicating significant polarization within this moiety (see Figure 4D). As discussed with Figure 4B, UFF + DDEC encounters challenges in accurately simulating highly polarized moieties, leading to a general overestimation of interactions with H_2O compared with CO_2.

In the case of YUFLOC, the CO_2 Q_{st} was underestimated, which contains an open metal site (OMS), which is generally challenging to simulate accurately using UFF + DDEC. The optimal intermolecular distance between the MOF and guest molecules is primarily governed by the sigma values in UFF parameters. However, the sigma values for some metal atoms tend to be overestimated, preventing guest molecules from approaching the OMS closely enough. Consequently, the interaction between the MOF and the guest molecule is underestimated. Unlike H_2O, CO_2 cannot generate sufficient electrostatic potential to counterbalance this underestimation in interaction energy.

For FECWOB01, the Q_{st} values for both CO_2 and H_2O are underestimated, although the trend of CO_2 Q_{st} being higher than H_2O Q_{st} remained consistent. FECWOB01 contains an azole linker, and since the electronegativity of azoles varies with the nitrogen content, the force field parameters of nitrogen

Please cite this article in press as: Lim et al., Accelerating CO_2 direct air capture screening for metal-organic frameworks with a transferable machine learning force field, Matter (2025), https://doi.org/10.1016/j.matt.2025.102203

require careful consideration. For example, a study on ZIF-8,[54] which contains imidazole as an organic linker, derived a tailored force field using nitrogen parameters from a generalized amber force field.[55] In this case, the epsilon value was set at 85.5 K, which is 2.5 times larger than that in UFF, while the sigma value remained nearly identical. Similarly, applying a larger epsilon for nitrogen in azoles would likely increase the interaction strength for both CO_2 and H_2O.

To conduct a systematic analysis, interaction energies between MOFs and gas molecules were compared by examining the PES using UFF + DDEC and MACE-DAC-1 for the GoldDAC test set as provided in Figures S25–29 and Table S4. Overall, the MLFF model exhibits more accurately capturing the interactions between MOFs and CO_2 or H_2O molecules, aligned with the DFT calculations. Moreover, Figure S15 and Table S3 showed that CO_2 Q_{st} from MACE-DAC-1 align more closely with experimental data than those from both UFF and UFF+DDEC, particularly in high Q_{st} region. While extensive parameter fitting within classical force fields can yield high accuracy, such methods are typically confined to a limited set of similar MOFs and lack broad transferability. Therefore, using our DAC-SIM process to screen adsorption properties offers a more precise and time-efficient alternative to traditional calculations based on classical force fields.

DISCUSSION

The screening results obtained using MACE-DAC-1 with DAC-SIM allowed for a systematic analysis of chemical features within promising MOFs. All 161 promising candidates in the CoRE MOF database were further investigated. We excluded 43 chemically incorrect MOFs (e.g., missing hydrogen atoms or overlapping atoms due to symmetry issues, see Figure S16).

Before proceeding with the analysis of the remaining MOFs, we evaluated the necessity of fine-tuning by comparing MACE-DAC-1 with MACE-MP. As shown in Figures S3, S4, and S17, MACE-MP tended to overestimate interaction energies, leading to an overestimation of H_2O Q_{st} compared with MACE-DAC-1. In this context, similar to the observation that not all promising MOFs demonstrate desirable performance in UFF + DDEC calculations, 46 of 118 MOFs (approximately 40%) fail to be identified as promising when using MACE-MP instead of MACE-DAC-1.

As shown in Figure 5A, seven chemical features, including OMS, parallel benzene ring (PAR), metal-electronegative atom-metal (M-X-M), uncoordinated nitrogen, azole linker, methyl functional group (–CH_3), and electronegative functional group (–EN), were identified in the promising MOFs. Four of them (OMS, PAR, M-X-M, and uncoordinated nitrogen) were suggested in previous screening works,[15,39] and others (azole linker, –CH_3, and –EN) were intermittently described in experimental works.[56–59] Among these, OMS and the azole linker appeared in nearly one-half, or more, of the MOF candidates. The OMS, a well-known chemical moiety, facilitates strong interactions between positively charged metal atoms and negatively charged oxygen atoms in gas molecules. However, the stronger interaction between the more negatively charged O atom in H_2O, compared with that in CO_2, leads to H_2O typically interacting more strongly with OMS than CO_2. This suggests that OMS use should be approached with caution.

Additionally, the azole linker, commonly used in MOF synthesis due to its robust bonding with transition metals,[60] contains nitrogen and other non-carbon atoms (e.g., nitrogen, sulfur, or oxygen) that actively interact with CO_2 molecules. Given the frequent application of nitrogen-rich MOFs and porous materials in DAC application, azole linkers may serve a similar function within MOFs for facilitating CO_2 capture in humid conditions.

Apart from the two frequently observed features, the performance of promising MOFs containing one of the remaining five chemical features is shown in Figure 5B to examine the role of those features in DAC application. Among them, PAR, identified through a screening study using classical force fields as one of the effective features for CO_2 capture in humid conditions,[15] appeared frequently in the results from the DAC-SIM. The strong dispersion forces created between PARs form selective pockets conducive to capturing CO_2. Due to its larger molecular size and unique pi bonds between C and O atoms, CO_2 is more responsive to the chemical environment dominated by dispersion forces, unlike H_2O. Furthermore, PAR is less affected by the stronger electrostatic interactions induced by H_2O in comparison with CO_2. As a result, 9 of 11 promising MOFs containing PAR show high selectivity, with K_H ratio for CO_2 to H_2O exceeding 100.

The M-X-M, a generalized form of metal-oxygen-metal (M-*O*-M) bridges, was proposed as a feature for strong CO_2 adsorption from screening study with the classical force field approach.[15] However, this feature was not considered essential for constructing MOFs specifically for selective CO_2 capture, as many MOFs with M-*O*-M bridges interact strongly with both H_2O and CO_2. However, given that the overestimation of H_2O interaction, which was observed in calculations using UFF and point charges, was corrected in this study, M-X-M still appeared in more than 20% of the promising MOFs often associated with high CO_2 selectivity over H_2O.

Uncoordinated nitrogen, as defined in this study, refers to nitrogen atoms that are not part of metal coordination complexes or azole linkers, such as those in amine groups or azines. They have been experimentally recognized as promising functional groups for DAC applications.[61–64] The lone pair on the nitrogen atom create localized pockets within the framework that act as a nucleophilic site, facilitating strong interactions with electrophilic CO_2 molecules. Consequently, 87.5% of MOFs containing uncoordinated nitrogen exhibit CO_2 Q_{st} values exceeding 58.9 kJ/mol (the top 25% CO_2 Q_{st} value of the candidates) and selectivity greater than 100. Notably, selective CO_2 capture associated with amine groups occurs generally via chemical reaction, highlighting the need for further research to fully understand this selective behavior.

For the –CH_3 group, known for its hydrophobic properties, it is commonly used in MOFs to impart hydrophobic characteristics.[57,59] As the smallest hydrocarbon unit, –CH_3 becomes increasingly hydrophobic with greater carbon chain length. Previous study has demonstrated that organic ligands with extended hydrocarbon chains are particularly suitable for DAC application when integrated into MOFs.[40] Likewise, promising MOFs functionalized with –CH_3 groups exhibit

Please cite this article in press as: Lim et al., Accelerating CO_2 direct air capture screening for metal-organic frameworks with a transferable machine learning force field, Matter (2025), https://doi.org/10.1016/j.matt.2025.102203

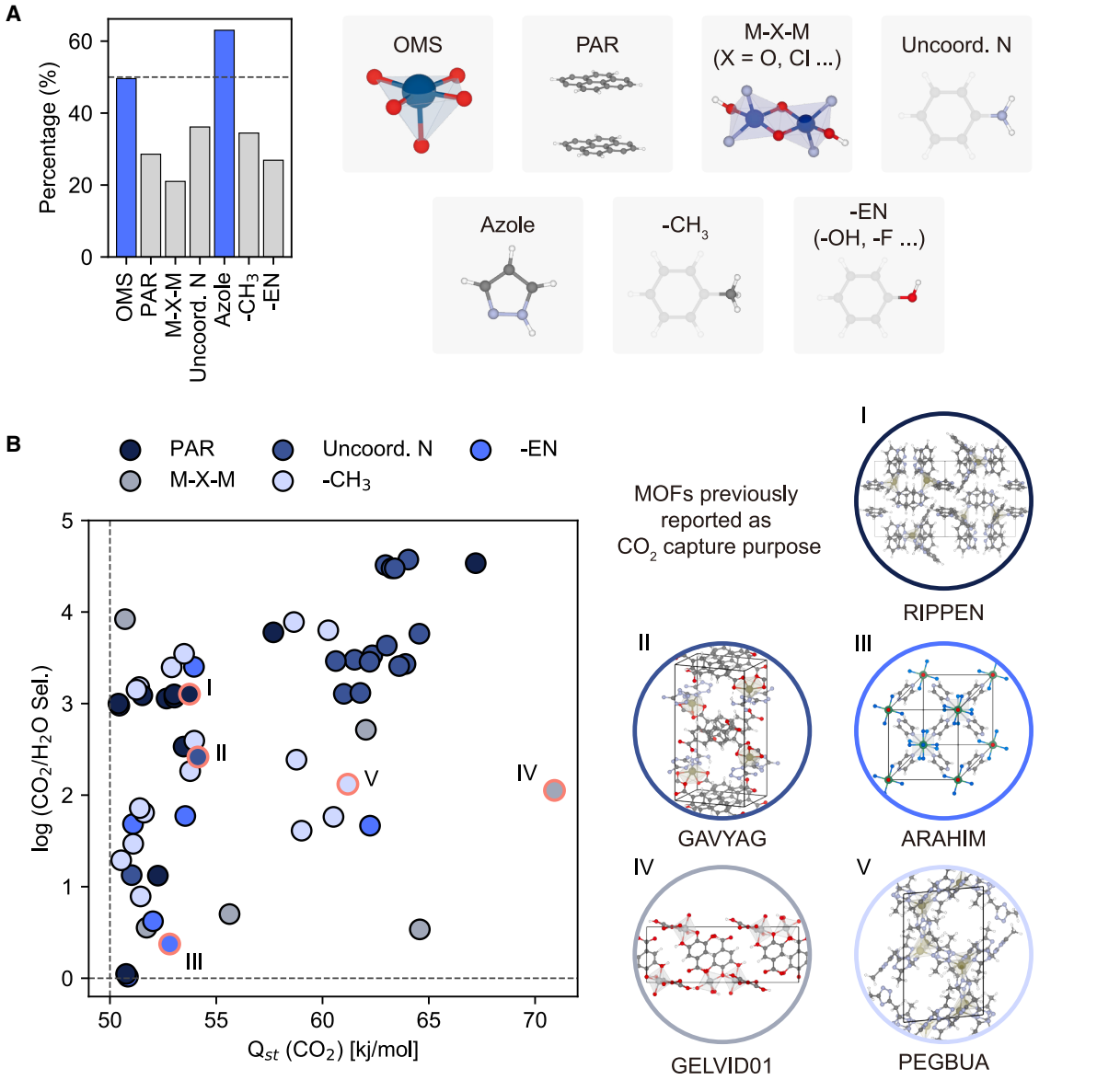


Figure 5. Chemical features analysis from promising MOFs

(A) Introduction of the seven chemical features discovered in candidate MOFs. The number of MOFs containing each feature is depicted as a bar plot on the left side, while illustrations of the seven chemical features are shown on the right side.

(B) A scatterplot showing the relationship between CO_2 Q_{st} and selectivity (in log scale for convenience) for candidate MOFs containing a single feature (excluding OMS and azole). Five representative MOFs for each feature are highlighted with red edges, and corresponding illustrations are provided on the right side. Atom colors: C (gray), H (white), N (light blue), O (red), Ni (light gray), Zn (khaki), Nb (green), and F (dodger blue).

selectivity values consistently above 7.5, with 59% of them exceeding 100.

Last, the –EN group, capable of readily accumulating electrons, can function as a nucleophilic site, selectively interacting with the electrophilic C atom within CO_2 molecule. Due to the high electronegativity of fluorine, fluoride functional groups are frequently used to design optimal pockets for selective CO_2 capture within MOFs.[56,65,66] A few MOFs incorporating fluoride functional groups have been experimentally validated to maintain their CO_2 capture efficiency under humid conditions.[56,66,67]

Additionally, five representative MOFs (RIPPEN,[78] GAVYAG,[68] ARAHIM,[56] GELVID01,[69] and PEGBUA[70]), which were discovered during our screening process and have been computationally or experimentally studied for CO_2 capture purpose, are illustrated in Figure 5B. Among these MOFs, ARAHIM, also known as NbOFFIVE-1-Ni or KAUST-7,[71] has even been proven to maintain trace CO_2 capture performance in humid conditions.[56] Moreover, the common names of promising MOFs were examined to enhance accessibility. As shown in Table 2, 12 MOFs were presented with the corresponding common name. Given

Please cite this article in press as: Lim et al., Accelerating CO_2 direct air capture screening for metal-organic frameworks with a transferable machine learning force field, Matter (2025), https://doi.org/10.1016/j.matt.2025.102203

Table 2. Collection of promising MOFs for DAC with common names

REFCODE	Common name	Metal	Features	Q_{st} (CO_2) [kJ/mol]	Q_{st} (H_2O) [kJ/mol]	K_H (CO_2) [mol/kg/Pa]	K_H (H_2O) [mol/kg/Pa]
ARAHIM[71]	KAUST-7	Nb, Ni	–EN	52.8	46.8	2.97×10^{-2}	1.25×10^{-2}
CISMEY[72]	JUC-115	Cd	OMS, PAR, M-X-M, azole	53.9	52.8	5.86×10^{-2}	5.43×10^{-2}
COXCUP[73]	MAF-42	Cu	OMS, azole, $-CH_3$	53.8	37.0	1.37×10^{-1}	7.51×10^{-4}
FECWOB01[74]	Zn-BTC	Zn	OMS, PAR, azole	57.7	31.5	2.44	4.06×10^{-4}
HOJJUN[75]	ZJNU-40	Cu	azole, –EN	51.1	41.1	1.82×10^{-1}	3.76×10^{-3}
NOFQIK[76]	TCuI	Cu	OMS, M-X-M azole, $-CH_3$	51.9	29.9	8.77×10^{-1}	3.08×10^{-4}
PEGBUA[70]	MAF-23	Zn	OMS, azole, $-CH_3$	61.2	50.4	9.49	7.21×10^{-2}
RIPNOV[77]	D-ZIF-7	Zn	PAR, azole	52.7	31.9	3.96×10^{-1}	3.49×10^{-4}
VEJZEQ[78]	ZIF-9	Co	PAR, azole	50.4	28.5	1.60×10^{-1}	1.60×10^{-4}
YIZJEY[79]	IFMC-3	Zn	azole, $-CH_3$	51.4	48.5	5.20×10^{-1}	7.21×10^{-3}
YUFLOC[80]	AEMOF-5	Sr	OMS, PAR, M-X-M, –EN	82.5	78.0	2.51×10^4	9.57×10^3
YUZCED[81]	MAF-49	Zn	uncoordinated nitrogen, azole	63.4	35.5	7.48×10	2.50×10^{-3}

A total of 12 promising MOFs, along with their respective common names, are presented. The REFCODE, corresponding common name, metal type, associated chemical features, Q_{st} and K_H are listed. The Q_{st} and K_H values were calculated using the MACE-DAC-1 model with the DAC-SIM package.

the extensive tunability of MOFs, which allows for a wide range of chemical environments, all structures except ARAHIM exhibit multiple chemical features. Thus, the diverse chemical attributes of these MOFs are anticipated to synergistically contribute to DAC.

In the same vein, combining these features offers an effective strategy for designing MOFs. For example, SIFSIX-18-Ni,[57] which demonstrated exceptional trace CO_2 capture capabilities in humid conditions, incorporates three key features: azole linker, $-CH_3$, and –EN. As shown in Figure S18, SIFSIX-18-Ni consists of a tetramethyl bipyrazole organic linker (a methyl functionalized azole-based linker) and a SiF_6^{2-} pillar (an electronegative functional group). The methyl groups on the organic linker act as water-blocking groups and the F atoms of the SiF_6^{2-} pillar are working as nucleophile sites to interact with the electrophilic C atom of the CO_2 molecule. The hydrophobicity of the methyl groups creates a water-limiting pocket near the electronegative groups, making the pocket suitable for DAC application. By combining features, a range of MOFs can be designed to meet specific DAC needs, and various combinations can be possible from a set of seven distinct features.

Since our work focused on the large-scale screening and evaluated performance using thermodynamic descriptors (Q_{st} and K_H) as a primary step, further analysis should be conducted on our candidates to confirm that their practical applicability for the real-world DAC applications. For example, given that DAC processes are typically conducted under humid conditions, it is essential to consider the adsorption performance in the presence of both CO_2 and H_2O for more realistic simulations, a consideration not addressed in this study. In practice, the adsorption of H_2O can introduce new adsorption sites for CO_2 capture. Meanwhile, at high relative humidity, H_2O molecules can cluster due to strong H_2O-H_2O interactions, potentially blocking the pores within MOFs. Furthermore, this study assumes MOFs to be rigid, precluding the modeling of chemical reactions (e.g., CO_2 conversion to carbamic acid) commonly used in porous materials for DAC application. Therefore, further study should focus on simulations that more accurately reflect practical DAC scenarios (e.g., GCMC simulation and simulations considering flexibility) by tackling those limitations. Beyond investigating the equilibrium phase at the atomistic scale, kinetic factors (e.g., thermal and mechanical stability, long-term stability, and adsorption kinetics) and process-level metrics (e.g., regeneration energy) should also be considered to identify practically viable candidates for DAC applications.

Conclusions

In this work, we presented a scalable computational approach for investigating gas adsorption properties of materials. We developed a transferable machine learning forcefield, MACE-DAC, for CO_2 interactions with MOFs and integrated it with a simulation workflow to calculate Q_{st} and K_H with high accuracy and efficiency. This workflow is provided as a Python package, DAC-SIM, for ease of use for various user purposes. We screened the CO_2 and H_2O adsorption properties of more than 8,000 experimentally synthesized frameworks and identified 161 promising candidates that were overlooked by conventional molecular simulation methods based on classical force fields. Additionally, we categorized features that can be introduced to design porous frameworks for DAC applications. Our workflow addresses prior limitations associated with classical force fields and the computational demands of direct *ab initio* methods, thus providing a solid foundation for accurate and fast virtual materials screening.

METHODS

GoldDAC database construction

To encompass the full range of the PES of interaction energy between the framework and gas molecules (CO_2 and H_2O), interaction energies were measured by adjusting the positions of the molecules within the structure. Starting from the equilibrium

Please cite this article in press as: Lim et al., Accelerating CO_2 direct air capture screening for metal-organic frameworks with a transferable machine learning force field, Matter (2025), https://doi.org/10.1016/j.matt.2025.102203

state with the presumably most stable configuration, two repulsive configurations were collected by moving the gas molecule closer to the structure, while three weak attraction configurations were collected by moving the guest molecule toward the pore center (details on configuration construction are provided in Note S1 and Figure S19). Interaction energies were calculated while not considering geometry optimization. DFT calculations were used to compute energies for each system, and the gas molecule was isolated in a 10 Å cubic box to calculate its energy. Thus, single self-consistent field calculation was conducted for each system.

The equilibrium state, regarded as the starting point for configuration construction, was collected from two different data sources: the ODAC23 database and manually calculated adsorption results from 26 MOFs that were identified as promising for CO_2 capture or DAC applications.[39,82] The MLFF model was trained based on optimized configurations collected from the ODAC23 database (training and validation set). First, the lowest CO_2 and H_2O interaction energies (interaction energies in this case were calculated after geometry optimization of MOF + gas system) were gathered for MOFs using the IS2RE database from the ODAC23 database. MOFs that met the following two conditions were considered: (1) the lowest CO_2 and H_2O binding energy is smaller than −0.1 eV, and (2) the number of atoms is less than 250. Last, given that the top 20 most frequently appearing metal atoms in the ODAC23 database were provided in the original paper (see Figure S1), 3 MOFs were randomly selected from those containing each metal type. Therefore, 60 MOFs were considered, and 120 equilibrium configurations were collected from the ODAC23 database, considering both CO_2 and H_2O molecules. For evaluation of the trained MLFF model, an additional database was constructed based on the equilibrium states obtained from the manually selected 26 MOFs. A single guest molecule was introduced into the framework, where it was allowed to relax while the framework remained fixed. Three to five different initial points were sampled, considering active adsorption sites for the guest molecules, and the configuration with the lowest energy was selected as the equilibrium state. In total, 52 equilibrium configurations were collected for both CO_2 and H_2O molecules. Further details and illustrations of these 26 MOFs are provided in Figure S20–S24.

Training details

The foundation model, MACE-MP,[35] was used for fine-tuning with the GoldDAC dataset, where the detailed data distribution of training, validation, and test sets can be found in Table S1. Fine-tuning was carried out over 100 epochs with a batch size of 4. The Adam optimizer was adopted for parameter updates,[83] featuring a learning rate of 1e−4 and zero weight decay. The loss function ($L_{E,F}$) for fine-tuning was balanced equally between total energy and force predictions, with a ratio set of 1:1.

For the continual learning, an additional regularization loss ($L_{continual}$) was introduced, computed as the root-mean-square error between energy and force predictions of the frozen pre-trained model and the corresponding fine-tuned model. Consequently, the total loss function for continual learning can be defined as:

$$L = L_{E,F} + \lambda L_{continual},$$

where λ represents the weighting coefficient for the continual loss, which was empirically set to 1.0 to achieve a balance between retaining pre-trained knowledge and adapting to new task-specific features.

Molecular simulation with MLFFs

The DAC-SIM package is a python package integrated with the developed MLFF through the ASE calculator interface,[84] for molecular simulations including Widom insertion, molecular dynamics, and geometry optimization. The Widom insertion simulations were conducted on the CoRE MOF 2019 dataset[37] to determine the K_H and Q_{st}. Before performing the simulation, geometry optimization was implemented for each MOF structure. The optimization process includes a total of 30 steps, each consisting of 50 iterations of cell relaxation followed by 50 iterations of internal relaxation with the cell held fixed. The convergence criterion for the maximum force was set at 0.05 eV/Å to balance computational efficiency.

For the following Widom insertion simulation, a grid with a spacing of 0.15 Å was generated to define potential insertion points across the unit cells. Grid points located within 1.5 Å of the central positions of gas molecules, such as the C atom in CO_2 and the O atom in H_2O, were excluded to avoid inserting molecules into blocked regions. In each iteration, a gas molecule was randomly rotated and placed at grid points, and their interaction energies between frameworks and gas molecules were calculated using MLFFs. To enhance statistical reliability, two independent Widom insertion simulations were performed, with each consisting of 10,000 insertions. As shown in the convergence test conducted in three independent iterations for Q_{st} and K_H values in Figures S31 and 32, we set 10,000 insertions as the standard for our simulations, as it provides a good balance between computational efficiency and convergence reliability. The results from these simulations were averaged to calculate the final K_H and Q_{st}. The temperature was set as 300 K.

Details of molecular simulations

DFT calculations were conducted using Vienna Ab initio Simulation Package (VASP).[85–87] The exchange-correlation terms were calculated using generalized gradient approximation with the Perdew-Burke-Ernzerhof functional.[88] The recommended pseudopotentials from VASP manual except Eu atom (Eu_3 was used instead of Eu_2 following pseudopotential selection in QMOF database[89]) were used to describe the ion-electron interactions. The energy cutoff was selected as 600 eV and the electronic relaxations were iterated until the energy changes below 10^{-5} eV. The gamma points were only used during calculation to ensure computational efficiency along with the Gaussian smearing method (sigma = 0.05 Å). However, for improved accuracy, denser k-point sampling may be necessary for MOFs with small lattice constants or significant electronic and dispersion interactions. Dispersion correction using the D3 (Becke-Johnson) scheme[90,91] were applied during preparing equilibrium state configurations for test set and ionic relaxations were iterated

until the forces were less than 0.03 eV/Å. During construction of GoldDAC dataset, the additional calculations were conducted without dispersion correction. In case of labeled as "+D3," the separate calculations were conducted with dispersion correction using D3 (Becke-Johnson) scheme. Spin-polarization calculations were only considered for frameworks with open-shell metal atoms, and U corrections were applied for d orbitals of Cu, Mn, Ni, Fe, Co, V, Nb, and Ti. Since these DFT settings for fine-tuning database did not fully match those used in the pre-training database for MACE-MP model, this discrepancy may contribute to the initial errors observed in MACE-MP model during evaluation (more details on Note S3). However, as interaction energies are derived from three different systems (MOF + gas, MOF, and gas), they are relatively insensitive to subtle variations in DFT settings.

Q_{st} and K_H using classical force field were calculated with RASPA2 software.[53] Both properties were calculated via Widom insertion MC method that we tried to implement in this work. The cutoff was set as 12.8 Å and the Ewald summation method was adopted to consider the electrostatic potential. The pore blocking algorithm was not considered. The 10,000 MC cycles were iterated for both CO_2 and H_2O molecules and temperature was set as 300 K to keep consistency with DAC-SIM calculations.

The interaction energies using a classical force field of MOFs in the test set were computed using Large-scale Atomic/Molecular Massively Parallel Simulator software.[92] Consistent with the approach taken in adsorption property calculations, a cutoff distance of 12.8 Å was used, and the Ewald summation method was used to account for long-range coulombic interactions with an accuracy threshold of 1×10^{-5}.

For classical force field calculations, framework atoms were modeled using the UFF,[18] with point charges assigned based on the DDEC[21] scheme. If the .cif file for a structure contained pre-defined charges, these values were retained without further modification. For guest molecules, specifically CO_2 and H_2O, the TraPPE[93] and SPC/E[94] force fields were used, respectively.

Geometric properties such as LCD, surface area (SA), and volume fraction (VF) were calculated using Zeo++ software.[95] Probe radius for calculation of SA and VF were selected as 1.4 Å, which is the van der Waals radius of Helium. We conducted 2k and 50k MC samplings to calculate the SA and VF, respectively.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and materials should be directed to and will be fulfilled by the lead contact, Jihan Kim (jihankim@kaist.ac.kr).

Materials availability

This study did not generate new, unique reagents.

Data and code availability

The DAC-sim library is available at https://github.com/hspark1212/DAC-SIM. Documentation for the library is available at https://hspark1212.github.io/DAC-SIM. The implementation of continual learning with MACE is available at https://github.com/hspark1212/mace/tree/develop. Associated files for screening results and the GoldDAC dataset are available via Figshare (https://doi.org/10.6084/m9.figshare.27978474.v3).

ACKNOWLEDGMENTS

We thank Jisu Jung and Honghui Kim for insightful discussions. Y.L. and J.K. acknowledge funding from the Saudi Aramco-KAIST CO_2 Management Center (Projects: G01230260 and N11230051) and the National Research Foundation of Korea under grant No. RS-2025-02218827. H.P. and A.W. acknowledge funding from this work was supported by EPSRC project EP/X037754/1 and are grateful to the UK Materials and Molecular Modelling Hub for computational resources, which is partially funded by EPSRC (EP/T022213/1, EP/W032260/1, and EP/P020194/1).

AUTHOR CONTRIBUTIONS

Conceptualization, Y.L. and H.P.; methodology, Y.L. and H.P.; investigation, Y.L. and H.P.; writing – original draft, Y.L, H.P., A.W., and J.K.; writing – review and editing, Y.L., H.P., A.W., and J.K.; funding acquisition, A.W. and J.K.; supervision, A.W. and J.K.

DECLARATION OF INTERESTS

The authors declare no competing interests.

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During the preparation of this work, the authors used ChatGPT to edit the final draft of the manuscript. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.matt.2025.102203.

Received: December 11, 2024
Revised: March 19, 2025
Accepted: May 16, 2025

REFERENCES

1. NASA's Goddard Institute for Space Studies (GISS). Global Temperature. https://climate.nasa.gov/vital-signs/global-temperature/?intent=121%60.
2. Olivier, J.G., Schure, K., and Peters, J. (2017). Trends in global CO2 and total greenhouse gas emissions. PBL Netherlands Environmental Assessment Agency *5*, 1–11.
3. Lackner, K.S., Brennan, S., Matter, J.M., Park, A.-H.A., Wright, A., and Van Der Zwaan, B. (2012). The urgency of the development of CO2 capture from ambient air. Proc. Natl. Acad. Sci. USA *109*, 13156–13162.
4. Sabatino, F., Grimm, A., Gallucci, F., van Sint Annaland, M., Kramer, G.J., and Gazzani, M. (2021). A comparative energy and costs assessment and optimization for direct air capture technologies. Joule *5*, 2047–2076.
5. Didas, S.A., Choi, S., Chaikittisilp, W., and Jones, C.W. (2015). Amine–oxide hybrid materials for CO2 capture from ambient air. Acc. Chem. Res. *48*, 2680–2687.
6. Stolaroff, J.K., Keith, D.W., and Lowry, G.V. (2008). Carbon dioxide capture from atmospheric air using sodium hydroxide spray. Environ. Sci. Technol. *42*, 2728–2735.
7. Zhang, W., Liu, H., Sun, Y., Cakstins, J., Sun, C., and Snape, C.E. (2016). Parametric study on the regeneration heat requirement of an amine-based solid adsorbent process for post-combustion carbon capture. Applied energy *168*, 394–405.
8. Keith, D.W., Holmes, G., St Angelo, D., and Heidel, K. (2018). A process for capturing CO2 from the atmosphere. Joule *2*, 1573–1594.

9. Khraisheh, M., Mukherjee, S., Kumar, A., Al Momani, F., Walker, G., and Zaworotko, M.J. (2020). An overview on trace CO2 removal by advanced physisorbent materials. J. Environ. Manage. *255*, 109874.
10. Zhou, H.-C., Long, J.R., and Yaghi, O.M. (2012). Introduction to metal–organic frameworks. Chem. Rev. *112*, 673–674.
11. Furukawa, H., Cordova, K.E., O'Keeffe, M., and Yaghi, O.M. (2013). The chemistry and applications of metal-organic frameworks. Science *341*, 1230444.
12. Lin, L.-C., Berger, A.H., Martin, R.L., Kim, J., Swisher, J.A., Jariwala, K., Rycroft, C.H., Bhown, A.S., Deem, M.W., Haranczyk, M., and Smit, B. (2012). In silico screening of carbon-capture materials. Nat. Mater. *11*, 633–641.
13. Fernandez, M., Boyd, P.G., Daff, T.D., Aghaji, M.Z., and Woo, T.K. (2014). Rapid and accurate machine learning recognition of high performing metal organic frameworks for CO2 capture. J. Phys. Chem. Lett. *5*, 3056–3060.
14. Li, S., Chung, Y.G., and Snurr, R.Q. (2016). High-throughput screening of metal–organic frameworks for CO2 capture in the presence of water. Langmuir *32*, 10368–10376.
15. Boyd, P.G., Chidambaram, A., García-Díez, E., Ireland, C.P., Daff, T.D., Bounds, R., Gładysiak, 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.
16. Findley, J.M., and Sholl, D.S. (2021). Computational screening of MOFs and zeolites for direct air capture of carbon dioxide under humid conditions. J. Phys. Chem. C *125*, 24630–24639.
17. 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.
18. Rappé, A.K., Casewit, C.J., Colwell, K., Goddard III, W.A., 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.
19. Rappe, A.K., and Goddard, W.A., III. (1991). Charge equilibration for molecular dynamics simulations. J. Phys. Chem. *95*, 3358–3363.
20. Wilmer, C.E., Kim, K.C., and Snurr, R.Q. (2012). An extended charge equilibration method. J. Phys. Chem. Lett. *3*, 2506–2511.
21. Manz, T.A., and Sholl, D.S. (2010). Chemically meaningful atomic charges that reproduce the electrostatic potential in periodic and nonperiodic materials. J. Chem. Theory Comput. *6*, 2455–2468.
22. Li, Y., Jin, X., Moubarak, E., and Smit, B. (2024). A refined set of universal force field parameters for some metal nodes in metal–organic frameworks. J. Chem. Theory Comput. *20*, 10540–10552.
23. Dzubak, A.L., Lin, L.-C., Kim, J., Swisher, J.A., Poloni, R., Maximoff, S.N., Smit, B., and Gagliardi, L. (2012). Ab initio carbon capture in open-site metal–organic frameworks. Nat. Chem. *4*, 810–816.
24. Lee, K., Howe, J.D., Lin, L.-C., Smit, B., and Neaton, J.B. (2015). Small-molecule adsorption in open-site metal–organic frameworks: a systematic density functional theory study for rational design. Chem. Mater. *27*, 668–678.
25. von Lilienfeld, O.A., Müller, K.-R., and Tkatchenko, A. (2020). Exploring chemical compound space with quantum-based machine learning. Nat. Rev. Chem *4*, 347–358.
26. Unke, O.T., Chmiela, S., Sauceda, H.E., Gastegger, M., Poltavsky, I., Schütt, K.T., Tkatchenko, A., and Müller, K.R. (2021). Machine learning force fields. Chem. Rev. *121*, 10142–10186.
27. Keith, J.A., Vassilev-Galindo, V., Cheng, B., Chmiela, S., Gastegger, M., Müller, K.R., and Tkatchenko, A. (2021). Combining machine learning and computational chemistry for predictive insights into chemical systems. Chem. Rev. *121*, 9816–9872.
28. Hart, G.L.W., Mueller, T., Toher, C., and Curtarolo, S. (2021). Machine learning for alloys. Nat. Rev. Mater. *6*, 730–755.
29. Hwang, S., Jung, J., Hong, C., Jeong, W., Kang, S., and Han, S. (2023). Stability and equilibrium structures of unknown ternary metal oxides explored by machine-learned potentials. J. Am. Chem. Soc. *145*, 19378–19386.
30. Mosquera-Lois, I., Kavanagh, S.R., Ganose, A.M., and Walsh, A. (2024). Machine-learning structural reconstructions for accelerated point defect calculations. npj Comput. Mater. *10*, 121.
31. Goeminne, R., Vanduyfhuys, L., Van Speybroeck, V., and Verstraelen, T. (2023). DFT-Quality adsorption simulations in metal–organic frameworks enabled by machine learning Potentials. J. Chem. Theory Comput. *19*, 6313–6325.
32. Zheng, B., Oliveira, F.L., Neumann Barros Ferreira, R., Steiner, M., Hamann, H., Gu, G.X., and Luan, B. (2023). Quantum informed machine-learning potentials for molecular dynamics simulations of CO2's chemisorption and diffusion in Mg-MOF-74. ACS Nano *17*, 5579–5587.
33. Shaidu, Y., Smith, A., Taw, E., and Neaton, J.B. (2023). Carbon capture phenomena in metal-organic frameworks with neural network potentials. PRX Energy *2*, 023005.
34. Liu, S., Dupuis, R., Fan, D., Benzaria, S., Bonneau, M., Bhatt, P., Eddaoudi, M., and Maurin, G. (2024). Machine learning potential for modelling H 2 adsorption/diffusion in MOFs with open metal sites. Chem. Sci. *15*, 5294–5302.
35. Batatia, I., Benner, P., Chiang, Y., Elena, A.M., Kovács, D.P., Riebesell, J., Advincula, X.R., Asta, M., and Avaylon, M. (2023). A foundation model for atomistic materials chemistry. Preprint at arXiv. https://arxiv.org/abs/2401.00096.
36. Widom, B. (1963). Some topics in the theory of fluids. J. Chem. Phys. *39*, 2808–2812.
37. 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.
38. Frenkel, D., and Smit, B. (2023). Understanding Molecular Simulation: From Algorithms to Applications (Elsevier).
39. Sriram, A., Choi, S., Yu, X., Brabson, L.M., Das, A., Ulissi, Z., Uyttendaele, 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.
40. Park, H., Majumdar, S., Zhang, X., Kim, J., and Smit, B. (2024). Inverse design of metal–organic frameworks for direct air capture of CO 2 via deep reinforcement learning. Digital Discovery *3*, 728–741.
41. Batatia, I., Kovacs, D.P., Simm, G., Ortner, C., and Csányi, G. (2022). MACE: Higher order equivariant message passing neural networks for fast and accurate force fields. Adv. Neural Inf. Process. Syst. *35*, 11423–11436.
42. Deng, B., Zhong, P., Jun, K., Riebesell, J., Han, K., Bartel, C.J., and Ceder, G. (2023). CHGNet as a pretrained universal neural network potential for charge-informed atomistic modelling. Nat. Mach. Intell. *5*, 1031–1041.
43. Chen, C., and Ong, S.P. (2022). A universal graph deep learning interatomic potential for the periodic table. Nat. Comput. Sci. *2*, 718–728.
44. Park, Y., Kim, J., Hwang, S., and Han, S. (2024). Scalable Parallel Algorithm for Graph Neural Network Interatomic Potentials in Molecular Dynamics Simulations. J. Chem. Theor. Comput. *20*, 4857–4868.
45. Deng, B., Choi, Y., Zhong, P., Riebesell, J., Anand, S., Li, Z., Jun, K., Persson, K.A., and Ceder, G. (2024). Overcoming systematic softening in universal machine learning interatomic potentials by fine-tuning. Preprint at arXiv. https://arxiv.org/abs/2405.07105.
46. Witt, W.C., van der Oord, C., Gelžinytė, E., Järvinen, T., Ross, A., Darby, J. P., Ho, C.H., Baldwin, W.J., and Sachs, M. (2023). ACEpotentials. jl: A Julia implementation of the atomic cluster expansion. J. Chem. Phys. *159*, 16.

47. McCloskey, M., and Cohen, N.J. (1989). Catastrophic interference in connectionist networks: The sequential learning problem. Psychol. Learn. Motiv. *24*, 109–165.

48. Luo, Y., Yang, Z., Meng, F., Li, Y., Zhou, J., and Zhang, Y. (2023). An empirical study of catastrophic forgetting in large language models during continual fine-tuning. Preprint at arXiv. https://arxiv.org/abs/2308.08747.

49. Kudithipudi, D., Aguilar-Simon, M., Babb, J., Bazhenov, M., Blackiston, D., Bongard, J., Brna, A.P., Chakravarthi Raja, S., Cheney, N., Clune, J., et al. (2022). Biological underpinnings for lifelong learning machines. Nat. Mach. Intell. *4*, 196–210.

50. Chen, Z., and Liu, B. (2018). Lifelong Machine Learning (Morgan & Claypool Publishers).

51. Li, W., Rao, Z., Chung, Y.G., and Li, S. (2017). The Role of Partial Atomic Charge Assignment Methods on the Computational Screening of Metal-Organic Frameworks for CO2 Capture under Humid Conditions. ChemistrySelect *2*, 9458–9465.

52. Kancharlapalli, S., Gopalan, A., Haranczyk, M., and Snurr, R.Q. (2021). Fast and accurate machine learning strategy for calculating partial atomic charges in metal–organic frameworks. J. Chem. Theory Comput. *17*, 3052–3064.

53. 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.

54. Zheng, B., Sant, M., Demontis, P., and Suffritti, G.B. (2012). Force field for molecular dynamics computations in flexible ZIF-8 framework. J. Phys. Chem. C *116*, 933–938.

55. Wang, J., Wolf, R.M., Caldwell, J.W., Kollman, P.A., and Case, D.A. (2004). Development and testing of a general amber force field. J. Comput. Chem. *25*, 1157–1174.

56. Bhatt, P.M., Belmabkhout, Y., Cadiau, A., Adil, K., Shekhah, O., Shkurenko, A., Barbour, L.J., and Eddaoudi, M. (2016). A fine-tuned fluorinated MOF addresses the needs for trace CO2 removal and air capture using physisorption. J. Am. Chem. Soc. *138*, 9301–9307.

57. Mukherjee, S., Sikdar, N., O'Nolan, D., Franz, D.M., Gascón, V., Kumar, A., Kumar, N., Scott, H.S., Madden, D.G., Kruger, P.E., et al. (2019). Trace CO2 capture by an ultramicroporous physisorbent with low water affinity. Sci. Adv. *5*, eaax9171.

58. 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.

59. Wang, X., Alzayer, M., Shih, A.J., Bose, S., Xie, H., Vornholt, S.M., Malliakas, C.D., Alhashem, H., Joodaki, F., Marzouk, S., et al. (2024). Tailoring Hydrophobicity and Pore Environment in Physisorbents for Improved Carbon Dioxide Capture under High Humidity. J. Am. Chem. Soc. *146*, 3943–3954.

60. Lu, W., Wei, Z., Gu, Z.-Y., Liu, T.-F., Park, J., Park, J., Tian, J., Zhang, M., Zhang, Q., Gentle, T., 3rd., et al. (2014). Tuning the structure and function of metal–organic frameworks via linker design. Chem. Soc. Rev. *43*, 5561–5593.

61. Demessence, A., D'Alessandro, D.M., Foo, M.L., and Long, J.R. (2009). Strong CO2 binding in a water-stable, triazolate-bridged metal– organic framework functionalized with ethylenediamine. J. Am. Chem. Soc. *131*, 8784–8786.

62. McDonald, T.M., Lee, W.R., Mason, J.A., Wiers, B.M., Hong, C.S., and Long, J.R. (2012). Capture of carbon dioxide from air and flue gas in the alkylamine-appended metal–organic framework mmen-Mg2 (dobpdc). J. Am. Chem. Soc. *134*, 7056–7065.

63. McDonald, T.M., Mason, J.A., Kong, X., Bloch, E.D., Gygi, D., Dani, A., Crocellà, V., Giordanino, F., Odoh, S.O., Drisdell, W.S., et al. (2015). Cooperative insertion of CO2 in diamine-appended metal-organic frameworks. Nature *519*, 303–308.

64. Zhou, Z., Ma, T., Zhang, H., Chheda, S., Li, H., Wang, K., Ehrling, S., Giovine, R., and Li, C. (2024). Carbon dioxide capture from open air using covalent organic frameworks. Nature *635*, 96–101.

65. Shekhah, O., Belmabkhout, Y., Chen, Z., Guillerm, V., Cairns, A., Adil, K., and Eddaoudi, M. (2014). Made-to-order metal-organic frameworks for trace carbon dioxide removal and air capture. Nat. Commun. *5*, 4228.

66. Ullah, S., Tan, K., Sensharma, D., Kumar, N., Mukherjee, S., Bezrukov, A. A., Li, J., Zaworotko, M.J., and Thonhauser, T. (2022). CO2 Capture by Hybrid Ultramicroporous TIFSIX-3-Ni under Humid Conditions Using Non-Equilibrium Cycling. Angew. Chem. *61*, e202206613.

67. Low, M.-Y.A., Danaci, D., Azzan, H., Jiayi, A.L., Yong, G.W.S., Itskou, I., and Petit, C. (2024). Physicochemical Properties, Equilibrium Adsorption Performance, Manufacturability, and Stability of TIFSIX-3-Ni for Direct Air Capture of CO2. Energy & Fuels *38*, 11947–11965.

68. Liu, X.-J., Wang, X., Xu, J.-L., Tian, D., Chen, R.-Y., Xu, J., and Bu, X.-H. (2017). Selective gas adsorption and fluorescence sensing response of a Zn(ii) metal–organic framework constructed by a mixed-ligand strategy. Dalton Trans. *46*, 4893–4897.

69. Heymans, N., Bourrelly, S., Normand, P., Bloch, E., Mkhadder, H., Cooper, L., Gorman, M., Bouzidi, I., Guillou, N., De Weireld, G., and Devic, T. (2020). Small-Pore Gallates MOFs for Environmental Applications: Sorption Behaviors and Structural Elucidation of Their High Affinity for CO2. J. Phys. Chem. C *124*, 3188–3195.

70. Liao, P.-Q., Zhou, D.-D., Zhu, A.-X., Jiang, L., Lin, R.-B., Zhang, J.-P., and Chen, X.-M. (2012). Strong and dynamic CO2 sorption in a flexible porous framework possessing guest chelating claws. J. Am. Chem. Soc. *134*, 17380–17383.

71. Cadiau, A., Adil, K., Bhatt, P.M., Belmabkhout, Y., and Eddaoudi, M. (2016). A metal-organic framework–based splitter for separating propylene from propane. Science *353*, 137–140.

72. He, H., Sun, F., Su, H., Jia, J., Li, Q., and Zhu, G. (2014). Syntheses, structures and luminescence properties of three metal–organic frameworks based on 5-(4-(2 H-tetrazol-5-yl) phenoxy) isophthalic acid. CrystEngComm (Camb.) *16*, 339–343.

73. Liao, P.-Q., Zhu, A.-X., Zhang, W.-X., Zhang, J.-P., and Chen, X.-M. (2015). Self-catalysed aerobic oxidization of organic linker in porous crystal for on-demand regulation of sorption behaviours. Nat. Commun. *6*, 6350.

74. Wei, Q., Yang, D., Larson, T.E., Kinnibrugh, T.L., Zou, R., Henson, N.J., Timofeeva, T., Xu, H., Zhao, Y., and Mattes, B.R. (2012). Kinetic hysteresis in gas adsorption behavior for a rigid MOF arising from zig-zag channel structures. J. Mater. Chem. *22*, 10166–10171.

75. Song, C., He, Y., Li, B., Ling, Y., Wang, H., Feng, Y., Krishna, R., and Chen, B. (2014). Enhanced CO 2 sorption and selectivity by functionalization of a NbO-type metal–organic framework with polarized benzothiadiazole moieties. Chem. Commun. *50*, 12105–12108.

76. He, J., Duan, J., Shi, H., Huang, J., Huang, J., Yu, L., Zeller, M., Hunter, A. D., and Xu, Z. (2014). Immobilization of volatile and corrosive iodine monochloride (ICl) and I2 reagents in a stable metal–organic framework. Inorg. Chem. *53*, 6837–6843.

77. Zhao, P., Lampronti, G.I., Lloyd, G.O., Suard, E., and Redfern, S.A.T. (2014). Direct visualisation of carbon dioxide adsorption in gate-opening zeolitic imidazolate framework ZIF-7. J. Mater. Chem. A *2*, 620–623.

78. Park, K.S., Ni, Z., Côté, A.P., Choi, J.Y., Huang, R., Uribe-Romo, F.J., Chae, H.K., O'Keeffe, M., and Yaghi, O.M. (2006). Exceptional chemical and thermal stability of zeolitic imidazolate frameworks. Proc. Natl. Acad. Sci. USA *103*, 10186–10191.

79. Qin, J.S., Bao, S.J., Li, P., Xie, W., Du, D.Y., Zhao, L., Lan, Y.Q., and Su, Z. M. (2014). A stable porous anionic metal–organic framework for luminescence sensing of Ln3+ ions and detection of nitrobenzene. Chem. Asian J. *9*, 749–753.

80. Douvali, A., Papaefstathiou, G.S., Gullo, M.P., Barbieri, A., Tsipis, A.C., Malliakas, C.D., Kanatzidis, M.G., Papadas, I., Armatas, G.S., Hatzidimitriou, A.

G., et al. (2015). Alkaline earth metal ion/dihydroxy–terephthalate MOFs: structural diversity and unusual luminescent properties. Inorg. Chem. *54*, 5813–5826.

81. Liao, P.-Q., Zhang, W.-X., Zhang, J.-P., and Chen, X.-M. (2015). Efficient purification of ethene by an ethane-trapping metal-organic framework. Nat. Commun. *6*, 8697.
82. Bose, S., Sengupta, D., Rayder, T.M., Wang, X., Kirlikovali, K.O., Sekizkardes, A.K., Islamoglu, T., and Farha, O.K. (2024). Challenges and opportunities: Metal–organic frameworks for direct air capture. Adv. Funct. Mater. *34*, 2307478.
83. Kingma, D.P. (2014). Adam: A method for stochastic optimization. Preprint at arXiv. https://arxiv.org/abs/1412.6980.
84. Hjorth Larsen, A., Jørgen Mortensen, J., Blomqvist, J., Castelli, I.E., Christensen, R., Dułak, M., Friis, J., Groves, M.N., Hammer, B., Hargus, C., et al. (2017). The atomic simulation environment—a Python library for working with atoms. J. Phys. Condens. Matter *29*, 273002.
85. Kresse, G., and Hafner, J. (1993). Ab initio molecular dynamics for liquid metals. Physical review B *47*, 558–561.
86. Kresse, G., and Furthmüller, J. (1996). Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. Comput. Mater. Sci. *6*, 15–50.
87. 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.
88. Perdew, J.P., Burke, K., and Ernzerhof, M. (1996). Generalized gradient approximation made simple. Phys. Rev. Lett. *77*, 3865–3868.
89. 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.
90. 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.
91. 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.
92. Plimpton, S. (1995). Fast parallel algorithms for short-range molecular dynamics. J. Comput. Phys. *117*, 1–19.
93. Eggimann, B.L., Sunnarborg, A.J., Stern, H.D., Bliss, A.P., and Siepmann, J.I. (2014). An online parameter and property database for the TraPPE force field. Mol. Simul. *40*, 101–105.
94. Berendsen, H.J.C., Grigera, J.-R., and Straatsma, T.P. (1987). The missing term in effective pair potentials. J. Phys. Chem. *91*, 6269–6271.
95. 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.