Chemical Engineering Journal Advances 25 (2026) 100984

Contents lists available at ScienceDirect

Chemical Engineering Journal Advances

journal homepage: www.sciencedirect.com/journal/chemical-engineering-journal-advances

Leveraging molecular simulations and machine learning to assess CO_2, O_2, and N_2 adsorption and separation performances of diverse MOF databases

Hasan Can Gulbalkan, Seda Keskin *

Department of Chemical and Biological Engineering, Koç University, Rumelifeneri Yolu, Sariyer, Istanbul, 34450, Turkey

ARTICLE INFO

Keywords:
Metal-organic framework (MOF)
Gas adsorption
Molecular simulations
Machine learning (ML)

ABSTRACT

We integrated molecular simulations and machine learning (ML) to comprehensively explore the gas adsorption and separation performances of both synthesized and hypothetical metal-organic frameworks (MOFs) available in five different MOF databases. Following the generation of CO_2, O_2, and N_2 adsorption data for synthesized MOFs at varying pressures through grand canonical Monte Carlo (GCMC) simulations, we developed ML models that can swiftly and accurately predict the gas adsorption properties of any MOF based on its structural, chemical, and energetic characteristics. These ML models were then transferred to four distinct hypothetical MOF databases consisting of nearly 130,000 structures to assess their CO_2, O_2, and N_2 adsorption properties in addition to CO_2/N_2 and O_2/N_2 separation performances as a very efficient alternative to computationally time and resource demanding molecular simulations. We identified the top-performing materials from each database to uncover their structural, chemical, and topological properties leading to high selectivities and concluded that synthesized MOFs with narrow pores, lanthanide metals, and linkers featuring oxalate, pyridine dicarboxylate, and fumarate offer the highest CO_2/N_2 selectivities. Our work presents the most extensive dataset produced for CO_2, O_2, and N_2 gas adsorption in MOFs, composed of ~3.9 million data points for materials' structural, chemical, and energetic features, gas adsorption properties, and selectivities computed at different pressures to accelerate the materials design and discovery for CO_2, O_2, and N_2 adsorption and separation.

1. Introduction

Adsorption of CO_2, O_2, and N_2, along with efficient CO_2/N_2 and O_2/N_2 separations, plays a critical role in addressing environmental and industrial challenges. CO_2 separation is essential for mitigating greenhouse gas emissions, while O_2/N_2 separation is crucial for oxygen production in medical and industrial applications [1]. Traditional adsorbents, such as zeolites and activated carbons, often suffer from low adsorption capacities and selectivities, limiting their effectiveness in these applications [2]. Therefore, the discovery of new porous materials that can achieve high performance in gas adsorption and separation processes is critical.

Metal-organic frameworks (MOFs) have emerged as promising porous materials for adsorption-based gas separations thanks to their diverse pore sizes, high surface areas, large porosities, and tunable physical and chemical properties [3–5]. The combinations of metals and organic linkers have resulted in a vast library of MOFs, including 128,319 distinct synthesized materials to date [6], in addition to several hundreds of thousands of hypothetical MOFs (hMOFs) [7–10] generated computationally but not yet synthesized. Experimental synthesis, characterization, and testing of even a single MOF material for the adsorption of a target gas molecule require significant resources, time, and effort. Computational approaches, specifically molecular simulations, accurately and efficiently evaluate the gas adsorption and separation properties of MOFs prior to experiments. High-throughput computational screening (HTCS) studies where thousands of materials are examined through Grand Canonical Monte Carlo (GCMC) simulations highlighted the significant potential of MOFs to replace conventional porous materials in various gas adsorption and separation applications in addition to identifying the most promising structures among many candidates [11,12]. Molecular simulations demand substantial computational time due to the need to calculate the electrostatic interactions between gas molecules and MOFs. This has limited HTCS studies on CO_2/N_2 and O_2/N_2 separation to a narrow subset of synthesized or hypothetical MOFs, representing only a small fraction of the broad MOF landscape [13–20].

Machine learning (ML) has been recently integrated into molecular simulations to efficiently predict the gas adsorption and separation

* Corresponding author.
E-mail address: skeskin@ku.edu.tr (S. Keskin).

https://doi.org/10.1016/j.ceja.2025.100984

Available online 2 December 2025
2666-8211/© 2025 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

performance of an enormous number and variety of MOFs, while uncovering the hidden correlations between structural and chemical properties of materials and their gas adsorption performances. For example, the CO_2 adsorption capacities of 20 synthesized MOFs were predicted by ML using temperature, pressure, surface area, and pore volume as the input data and experimental gas uptakes as the target data [21]. The CO_2 adsorption capacities of MOFs were found to be highly dependent on their surface areas. ML models were developed to predict the single-component CO_2 adsorption in 2932 synthesized MOFs at various pressures using their structural and chemical properties together with the results of GCMC simulations [22]. Similarly, structural and chemical properties of 5000 synthesized MOFs were utilized together with the simulated gas adsorption data to train ML models for predicting the single-component O_2 and N_2 uptakes at 1 bar, 298 K, and MOFs were found to exhibit O_2/N_2 selectivities between 0.5–2.0 [23].

ML models were also developed to predict the CO_2 adsorption and separation performances of hMOFs. Over 400 hMOFs with distinct functional groups were computationally constructed and studied by GCMC simulations to compute their CO_2 and N_2 adsorption at 1 bar, 298 K [24]. ML models were subsequently developed to predict CO_2/N_2 mixture selectivities using the structural and chemical features of hMOFs and selectivities were found to be primarily influenced by the pore size and void fraction of materials. Single-component CO_2 adsorption properties of 100,240 hMOFs were assessed by ML utilizing

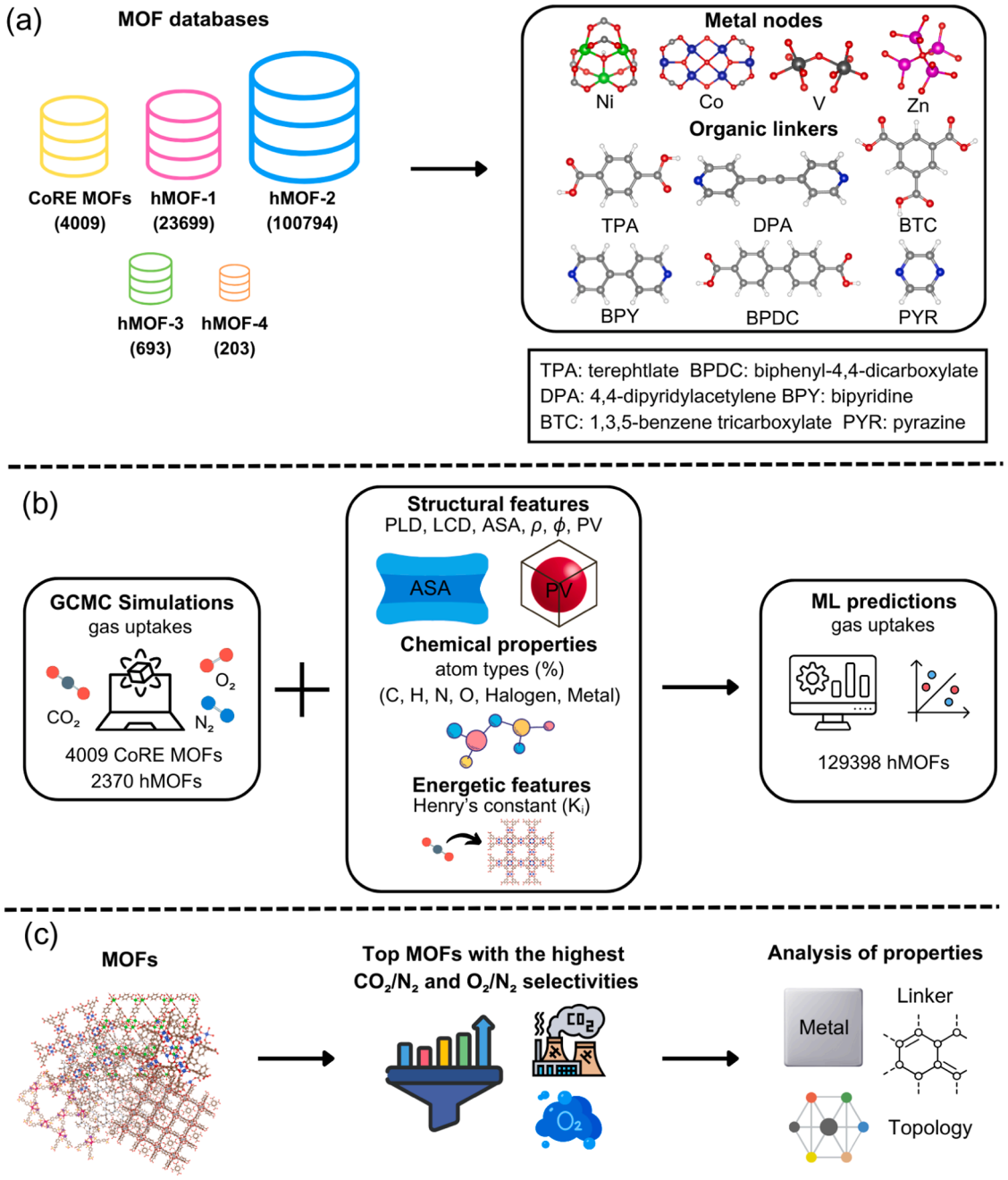


Fig. 1. A schematic overview of the computational workflow. (a) MOF databases: Numbers in parentheses represent the number of materials that we studied in each database. C, H, N, O, Ni, Co, V, and Zn atoms are represented with gray, white, blue, red, green, dark blue, dark gray, and magenta colors, respectively. (b) GCMC simulations were performed for 6379 MOFs to calculate their CO_2, O_2, and N_2 uptakes at three different pressures and then used to develop ML models to predict gas adsorption and separation performances of 129,398 hMOFs when their structural, chemical, and energetic features are given to the models. (c) The most promising MOFs for CO_2/N_2 and O_2/N_2 separations were identified and analyzed at the molecular level.

the CO_2 uptakes obtained from GCMC simulations as target data, and results showed that surface area, pore size, and void fraction are the most critical factors influencing CO_2 adsorption [25]. GCMC and deep learning were used to generate the CO_2 working capacities and CO_2/N_2 mixture selectivities of hMOFs at 1 bar, 298 K [26]. Results showed that ML-predicted CO_2/N_2 selectivities of hMOFs can reach up to 300.

This literature review highlights that current studies have focused on either synthesized MOFs or one of the available hypothetical MOF databases, emphasizing the need to fully explore the potential of both real and hypothetical MOF space in a comparative manner for CO_2, O_2, and N_2 adsorption, as well as for CO_2/N_2 and O_2/N_2 separations. Motivated from this gap in the literature, in this work, we integrated HTCS and ML algorithms to assess the adsorption and separation performances of five different MOF databases presenting varying structural and chemical properties: (i) Computation-ready experimental MOF (CoRE MOF) [27] database consisting of 4009 synthesized structures, (ii) the first publicly available hypothetical MOF database [7] comprising 100,794 hMOFs with five metal nodes and 84 linkers, (iii) a hypothetical MOF database [8] consisting of 23,699 structures possessing high metal diversity with 14 metal nodes and 95 linkers, (iv) an anion-pillared hypothetical MOF database [9] composed of 693 structures with five metal nodes and 12 linkers, and (v) a zirconium hypothetical MOF database [10] with 203 highly stable MOFs.

Single-component adsorption data for CO_2, O_2, and N_2 across all MOF databases were initially generated through HTCS, and these data were then used to develop ML models capable of accurately predicting gas adsorption for any given MOF at various pressures in just minutes. The CO_2/N_2 and O_2/N_2 separation performances of all MOFs existing in five different databases were assessed, and MOFs offering the highest selectivities among nearly 130,000 different structures were thoroughly analyzed to reveal the chemical, structural, and energetic properties that lead to high selectivities. Our work offers one of the most comprehensive and diverse datasets available for MOFs, composed of ~3.9 million data points for all synthesized and hypothetical MOFs' structural, chemical, and energetic features (1.9 million), their gas adsorption properties (1.2 million), and adsorption-based selectivities (0.8 million) computed at different conditions to significantly accelerate the materials discovery for gas adsorption and separation applications.

2. Computational methodology

Fig. 1 summarizes our computational methodology integrating HTCS and ML to evaluate the adsorption of CO_2, O_2, and N_2, as well as CO_2/N_2 and O_2/N_2 selectivities of MOFs. The structural properties of each MOF, including pore limiting diameter (PLD), the largest cavity diameter (LCD), accessible surface area (ASA), geometric pore volume (PV), density (ρ), and porosity (ϕ) were calculated using Zeo++ software (version 0.3) [28]. The crystallographic information files (CIFs) of the MOFs available in these five databases were utilized to extract the chemical descriptors and the elemental distribution of the structures (C %, H %, N %, O %, Metal %, and Halogen %). These percentages were calculated as the ratio of the number of atoms of a specific element to the total number of atoms in a unit cell of each MOF.

2.1. MOF databases

Synthesized MOFs: We examined the most recent CoRE MOF database [27] which includes 10,143 synthesized structures, and only used the computation-ready (CR) structures as listed in the updated version [29]. Early versions of this database have been used in HTCS studies for several gas storage and separation applications, including CH_4 storage [30], H_2 storage [31], and CH_4/H_2 [32], CO_2/CH_4 [33], CO_2/H_2 [34], CO_2/N_2 [14], O_2/N_2 [19,23], Xe/Kr [35], and xylene [36,37] separations. We selected CoRE MOFs having PLDs >3.64 Å and ASA >0 m^2/g to ensure the transport of CO_2 (3.30 Å), O_2 (3.46 Å), and N_2 (3.64 Å) molecules through their pores and as a result, we ended up with 4009 CoRE MOFs.

Hypothetical MOFs: Four hMOF databases, containing 23,891, 137,953, 936, and 204 structures, were studied and designated in this work as hMOF-1 [8], hMOF-2 [7], hMOF-3 [9], and hMOF-4 [10], respectively. hMOF-1 database consists of 14 metal nodes and 95 organic linkers with 52 distinct topologies. hMOF-2 database has five metal nodes, 84 organic linkers, and mostly pcu topologies. hMOF-3 database consists of five metal nodes, 12 organic linkers, and mostly pcu topologies. hMOF-4 database has only Zr-based metal nodes, 36 organic linkers, and four topologies. Similar to the CoRE MOF database, these hMOF databases were previously studied for CH_4 storage [7,10], H_2 storage [8,38], CO_2/CH_4 [15], C_2H_6/C_2H_4 [39], and Xe/Kr [9,40] separations. Following the structural filtering steps based on PLD >3.64 Å and ASA >0 m^2/g criteria, these databases were refined to include 23, 699, 100,794, 693, and 203 hMOFs, respectively.

2.2. Molecular simulations

GCMC simulations were performed using the RASPA [41] simulation code version 2.0.50 to study the single-component adsorption of CO_2, O_2, and N_2 in 4009 CoRE MOFs at pressures of 0.1, 1, 10 bar and a temperature of 298 K. These simulations were also conducted on representative subsets of 2370 and 2015 materials selected from hMOF-1 and hMOF-2 databases, respectively, for the development of ML models, making the total number of simulated MOFs as 8394. These subsets were generated by sorting the materials based on their PLDs in ascending order and selecting every 10th material for hMOF-1 and every 50th material for hMOF-2, as we will discuss in more detail in the ML section.

The dispersion interactions between MOFs and gas molecules were modeled using Lennard-Jones 12–6 (LJ) potentials. The potential parameters for Van der Waals interactions of the framework atoms were defined using the Universal Force Field (UFF) [42]. The charge equilibration (Qeq) method [43,44] was employed to assign the partial point charges to the framework atoms for computing the electrostatic interactions between the gas molecules and the frameworks through the Ewald summation [45]. CO_2 was modeled as a three-site rigid molecule with partial charges assigned to the center of each site, and its potential parameters were derived from the TraPPE force field [46]. O_2 and N_2 were represented as three-site rigid molecules with the two O and N atoms located at the first two sites and the center of mass at the third site, where partial charges were assigned [47,48]. These force fields, potential parameters, and the charge method were previously shown to accurately reproduce the experimental CO_2, O_2, and N_2 adsorption in MOFs [13,49,50].

In GCMC simulations, 10,000 cycles were used for initialization, followed by 20,000 cycles for ensemble averaging. The pair potentials between different atoms were calculated using the Lorentz-Berthelot mixing rules. A cut-off distance of 13 Å was applied for van der Waals interactions, and the cell lengths were extended to at least 26 Å in each dimension. Periodic boundary conditions were implemented in the three-dimensional system, and all simulations assumed a rigid framework. During the GCMC simulations, four types of trial moves were considered: translation, rotation, reinsertion, and swap of a molecule. These move probabilities are appropriately normalized by the RASPA code [41] and therefore given relative to each other. We assigned equal probabilities (0.5) to translation, rotation, and reinsertion to ensure that molecular displacements and reorientations are attempted with the same frequency. The swap probability (1.0) was set to be twice as high as the other moves to enhance particle exchange with the reservoir in the μVT ensemble, thereby improving the reliability of adsorption calculations. Ideal adsorption selectivities were calculated using the ratio of single-component gas uptakes obtained from GCMC simulations. The ratio of CO_2 (O_2) uptake to N_2 uptake was used to compute CO_2/N_2 (O_2/N_2) selectivities. We also performed molecular simulations to calculate the Henry's coefficients (K_i) of CO_2, O_2, and N_2 molecules for

all MOFs using the Widom particle insertion method [51]. A total of 25,000 cycles were performed, including 5000 initialization cycles for these simulations.

To show the accuracy of our simulation approach, we compared the gas uptakes obtained from our GCMC results with the experimentally reported CO_2, O_2, and N_2 uptakes in the literature for a set of widely studied MOFs (**Fig. S1**). This comparison revealed that the gas uptakes obtained from GCMC simulations using UFF generally agree well with the reported experimental uptake data, including the MOFs containing open metal sites (OMS). However, it is important to note that generic force fields like UFF may lack the specific parameters needed to accurately describe the strong interactions between OMS and adsorbates such as H_2O, CO_2, and O_2 [19,52]. In high-throughput computational screening of MOFs, the use of generic force fields is a well-established practice due to their broad transferability across diverse MOF structures [14,32,53,54]. Nevertheless, results related to MOFs with OMS should be interpreted with this limitation in mind.

2.3. Machine learning

The structural, chemical, and energetic descriptors of 4009 CoRE MOFs and 2370 hMOFs from hMOF-1 database were used as the input data, while their simulated CO_2, O_2, and N_2 uptakes at 0.1, 1, 10 bar obtained from GCMC were used as the target data to train the ML models. 13 different features of MOFs utilized as the input are listed in **Table S1**. Structural features include PLD, LCD, ASA, ρ, ϕ, and PV, whereas chemical features consist of the elemental composition percentages; C %, H %, N %, O %, Metal %, and Halogen %. As the energetic features, Henry's constants (K) for CO_2, O_2, or N_2 calculated at infinite dilution were used. The Tree-Based Pipeline Optimization Tool (TPOT) [55] (version 0.12.2) was employed to identify the optimal ML algorithms and fine-tune the model parameters. Regression algorithms from the scikit-learn package [56] were used for model selection. To maintain the consistent feature distributions between trained and test datasets, a stratified sampling approach was applied, dividing the data into an 80% trained set and a 20% test set. Additionally, five-fold cross-validation was implemented to avoid the risk of overfitting. The accuracy of the ML models was evaluated using the coefficient of determination (R^2), root-mean-square error (RMSE), and mean absolute error (MAE), with their respective expressions provided in **Table S2**. SHapley Additive exPlanations (SHAP) analysis [57] was performed to quantify the impact of materials' features on the ML predictions both in terms of magnitude and direction.

Nine different ML models were developed to predict CO_2, O_2, and N_2

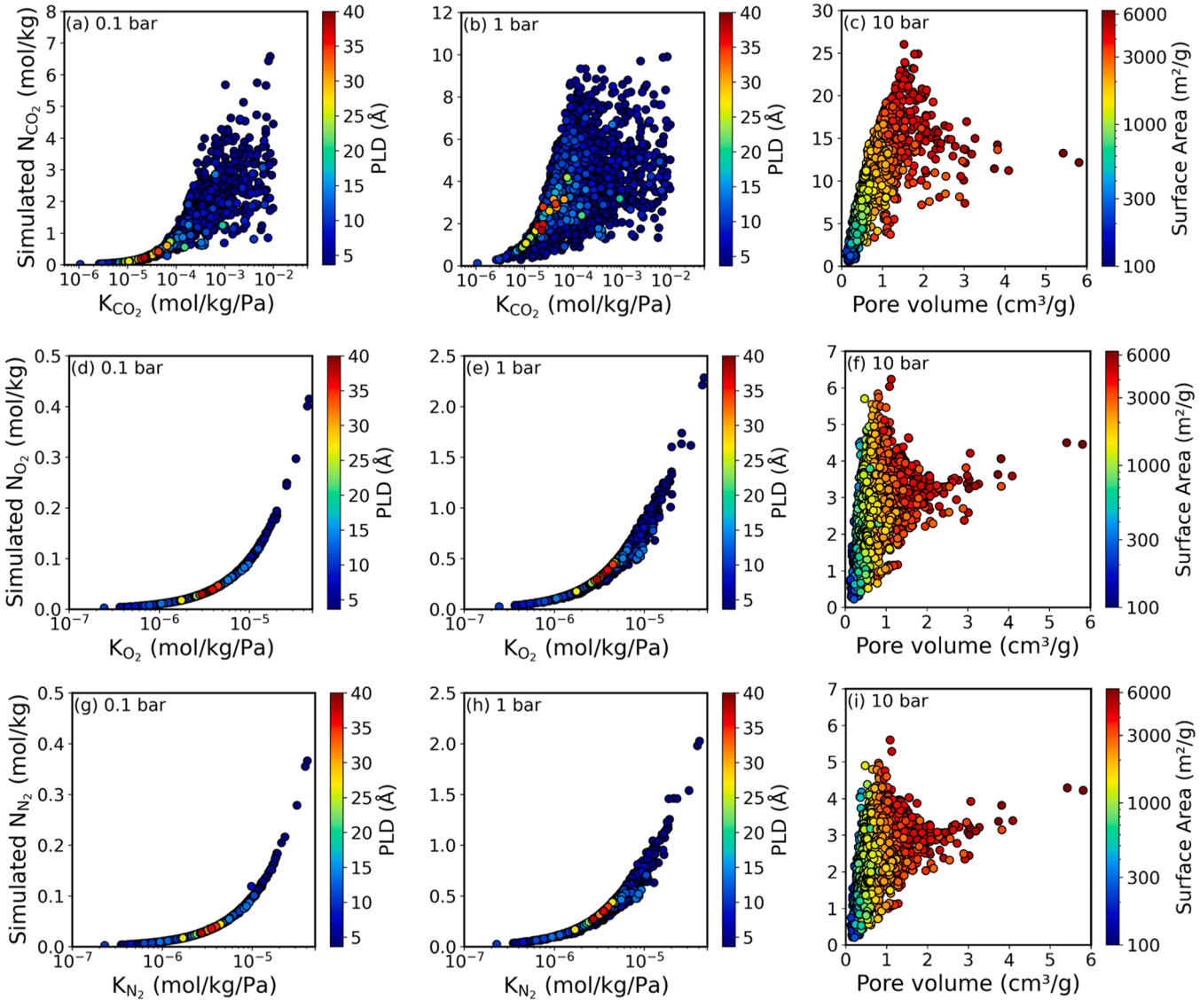


Fig. 2. Simulated CO_2, O_2, and N_2 uptakes (N) of 4009 CoRE MOFs at (a, d, g) 0.1 bar, (b, e, h) 1 bar, and (c, f, i) 10 bar.

uptakes in CoRE MOFs at three distinct pressures and these models were then tested on unseen 123,019 hMOFs available in four hypoMOF databases and demonstrated to be accurate for assessing their CO_2, O_2, and N_2 uptakes. The details and optimized hyperparameters of all ML models are provided in **Tables S3-S4**. ML-predicted CO_2, O_2, and N_2 uptakes were finally used to compute CO_2/N_2 and O_2/N_2 selectivities of all MOFs at 0.1, 1, and 10 bar, and several high-performing MOFs offering high selectivity were identified and examined to understand which molecular features of materials such as metal, linker type, pore sizes etc. lead to high selectivities.

3. Results and discussion

We first analyzed the correlations between the materials' features and their simulated gas uptakes at different pressures in **Fig. S2** and then showed how CO_2, O_2, and N_2 adsorption of CoRE MOFs change with the most correlated features in Fig. 2. At 0.1 and 1 bar, all gas uptakes are strongly correlated with their corresponding Henry's constants, as shown in Fig. 2(a, d, g) and (b, e, h), respectively, since Henry's constants represent the strength of interactions between the gas molecules and frameworks at low pressures. The Henry's constants of CO_2, O_2, and N_2 were computed between 1.1×10^{-6}–1.0×10^{-2}, 2.5×10^{-7}–4.6×10^{-5}, and 2.3×10^{-7}–4.1×10^{-5} mol/kg/Pa, respectively, indicating that CO_2 is much more strongly adsorbed in MOFs compared to O_2 and N_2, for which MOFs have similar adsorption affinities. At 10 bar, accessible surface area and pore volume of MOFs become dominant in determining the gas uptakes, and as they increase, higher gas uptakes are observed in Fig. 2(c, F, i).

Performing molecular simulations for CO_2, O_2, and N_2 adsorption for nearly 130,000 MOFs poses a substantial computational challenge, making ML an efficient alternative for rapidly and accurately predicting MOFs' gas adsorption properties. Nine different ML models were developed to predict CO_2, O_2, and N_2 adsorption in CoRE MOFs at 0.1, 1, and 10 bar. **Fig. S3** shows that R^2 values for the test sets of ML models predicting CO_2 uptakes are 0.926, 0.923, and 0.966 at 0.1, 1, and 10 bar, respectively, representing the models' high accuracy. Similarly, for O_2 and N_2 adsorption, R^2 values for test sets were calculated between 0.975–0.999. These values surpass the previously reported ML models' R^2 values for predicting CO_2, O_2, and N_2 uptakes of CoRE MOFs at 1 bar, 298 K, as presented in **Table S5**. Motivated from this, we used these ML models to predict the CO_2, O_2, and N_2 adsorption properties of the unseen hMOF-1 database. Here, unseen refers to materials that were not included in the dataset used for ML model training or testing.

Fig. S4 compares ML-predicted and simulated gas uptakes at 0.1, 1, and 10 bar for hMOF-1 subset consisting of 2370 hMOFs. This subset was created by sorting PLDs in ascending order and selecting every 10th structure to ensure that it adequately represents the feature distribution of the entire hMOF-1 database as shown in **Fig. S5**. For CO_2, the accuracy of the models is strong at 0.1 and 1 bar with high R^2 (0.89 and 0.83), low RMSE (0.15 and 0.69 mol/kg), and low MAE (0.04 and 0.28 mol/kg) values but the low R^2 (0.60) observed at 10 bar hints the risk of inaccurate CO_2 uptake predictions at high pressure. ML models accurately predicted O_2 and N_2 adsorption at all pressures except for 30 and 23 hMOFs, respectively, showing gas uptakes >5 mol/kg at 10 bar.

Based on these results, we concluded that the ML models trained using CoRE MOF database are not directly transferable to hMOF-1 database, especially for predicting high pressure gas uptakes. To develop new ML models that perform well on unseen databases at high pressures, we first analyzed the feature distributions of MOFs across five databases, including the trained and test set of CoRE MOFs and all hMOF databases, in **Fig. S5**. The trained and test sets of CoRE MOFs generally cover broader ranges for all the descriptors, capturing greater diversity compared to hMOF databases. However, the accessible surface area and pore volume ranges for the CoRE MOF database (up to 6362 m^2/g and 0.1–5.8 cm^3/g) are significantly narrower than that of hMOF-1 database (up to 7958 m^2/g and 0.3–18.0 cm^3/g), which is the key reason why our initial ML models trained on CoRE MOFs failed to accurately capture CO_2 adsorption properties of hMOF-1 database at 10 bar. As we discussed in Fig. 2, the accessible surface area and pore volume are critical in determining the gas uptakes at high pressure. Thus, we retrained the ML models by incorporating these 2370 hMOFs selected from the hMOF-1 database into the existing 4009 CoRE MOFs to better capture the full range of materials' features.

Fig. 3 presents the distributions of structural and energetic features for the new trained and test set of 6379 seen MOFs (4009 CoRE MOFs+2370 hMOFs) in comparison with all hMOF databases. The ranges for K_{CO_2}, K_{O_2}, and K_{N_2} for the trained sets now cover those of hMOF databases as shown in Fig. 3(a-c). For example, K_{CO_2} values for the trained set were computed in the range of 1×10^{-6}–1×10^{-2} mol/kg/Pa, covering most of the values in hMOF-1, hMOF-2, hMOF-3, hMOF-4 databases, which were computed as 3.9×10^{-6}–9.3×10^{-3}, 3.0×10^{-6}–6.0×10^{-2}, 9.4×10^{-6}–1.2×10^{-2}, and 4.6×10^{-6}–8.6×10^{-5} mol/kg/Pa, respectively. The only exceptions were 16 and two hMOFs from hMOF-2 and hMOF-3 databases, respectively. Furthermore, the most important features at high pressures are now better covered by the trained and test sets to improve the model's ability to generalize across diverse structures. For instance, the accessible surface area, pore volume, and density ranges for the trained MOFs (up to 7931 m^2/g, 0.1–13.6 cm^3/g, and 0.1–3.2 g/cm^3) closely align with those of the remaining hMOF-1 database (up to 7694 m^2/g, 0.3–18.0 cm^3/g, and 0.1–2.1 g/cm^3, respectively), as presented in Fig. 3(d-F), and only seven hMOFs with pore volumes of 13.9–18.0 cm^3/g were outside the range.

The new ML models predicting CO_2 uptakes utilized the Random Forest algorithm at 0.1 and 1 bar, and the XGB Regression algorithm at 10 bar. They all have high accuracy with R^2 values of 0.943, 0.925, and 0.939 for test sets at 0.1, 1, and 10 bar, respectively. The Ridge Regression algorithm was used for predicting O_2 and N_2 uptakes at 0.1 bar, the XGB Regression algorithm for O_2 uptakes at 1 and 10 bar, as well as N_2 uptakes at 10 bar, and the Extra Trees Regression algorithm for N_2 uptakes at 1 bar. These models demonstrated high accuracies with R^2 values ranging from 0.981 to 0.999 for the test sets. TPOT automatically evaluates a wide range of ML pipelines to identify the model that provides the best predictive performance. Based on the results given in Fig. 4 and **Table S6**, TPOT selected Random Forest, XGBoost, and ExtraTrees regression models for predicting CO_2, O_2, and N_2 uptakes at 1 bar, respectively, because these models offer the most favorable combination of high R^2 values together with low RMSE and MAE relative to the alternatives.

Fig. 4 shows the ML predictions for CO_2, O_2, and N_2 adsorption at 1 bar. ML-estimated CO_2, O_2, and N_2 uptakes at 1 bar were between 0.16–9.87, 0.03–2.27, and 0.03–1.87 mol/kg, respectively, in very good agreement with the simulated values, 0.11–9.90, 0.02–2.28, and 0.02–2.02 mol/kg, respectively. Fig. 4(a) illustrates that CO_2 uptakes are mostly concentrated in the range of 1–4 mol/kg for 3990 MOFs out of 6379. Fig. 4(b, c) demonstrate that O_2 and N_2 uptakes are predominantly distributed between 0.2–0.6 mol/kg, corresponding to 4640 and 4670 MOFs, respectively. For O_2 and N_2, the ML predictions are nearly identical to the simulated uptakes. The deviations between simulation and ML for CO_2 uptakes are more pronounced compared to O_2 and N_2, which can be attributed to the inherently complex adsorption behavior of CO_2 because of the strong electrostatic interactions with MOFs. Results for 0.1 and 10 bar are given in **Figs. S6-S7. Fig. S6** shows that ML-predicted CO_2, O_2, and N_2 uptakes at 0.1 bar (0.03–5.69, 0.01–0.42, and 0.01–0.37 mol/kg) align closely with the simulation results (0.01–6.57, 0.01–0.41, and 0.01–0.37 mol/kg, respectively). At 10 bar, ML-predicted CO_2, O_2, and N_2 uptakes (0.66–23.76, 0.25–6.86, and 0.22–6.63 mol/kg) are in very good agreement with the simulated uptakes (0.56–26.00, 0.22–7.05, and 0.20–6.85 mol/kg, respectively), as presented in **Fig. S7**.

Developing ML models provides not only accurate and fast property predictions but also the ability to derive molecular-level insights into the contributions of various features to the target properties. Fig. 4

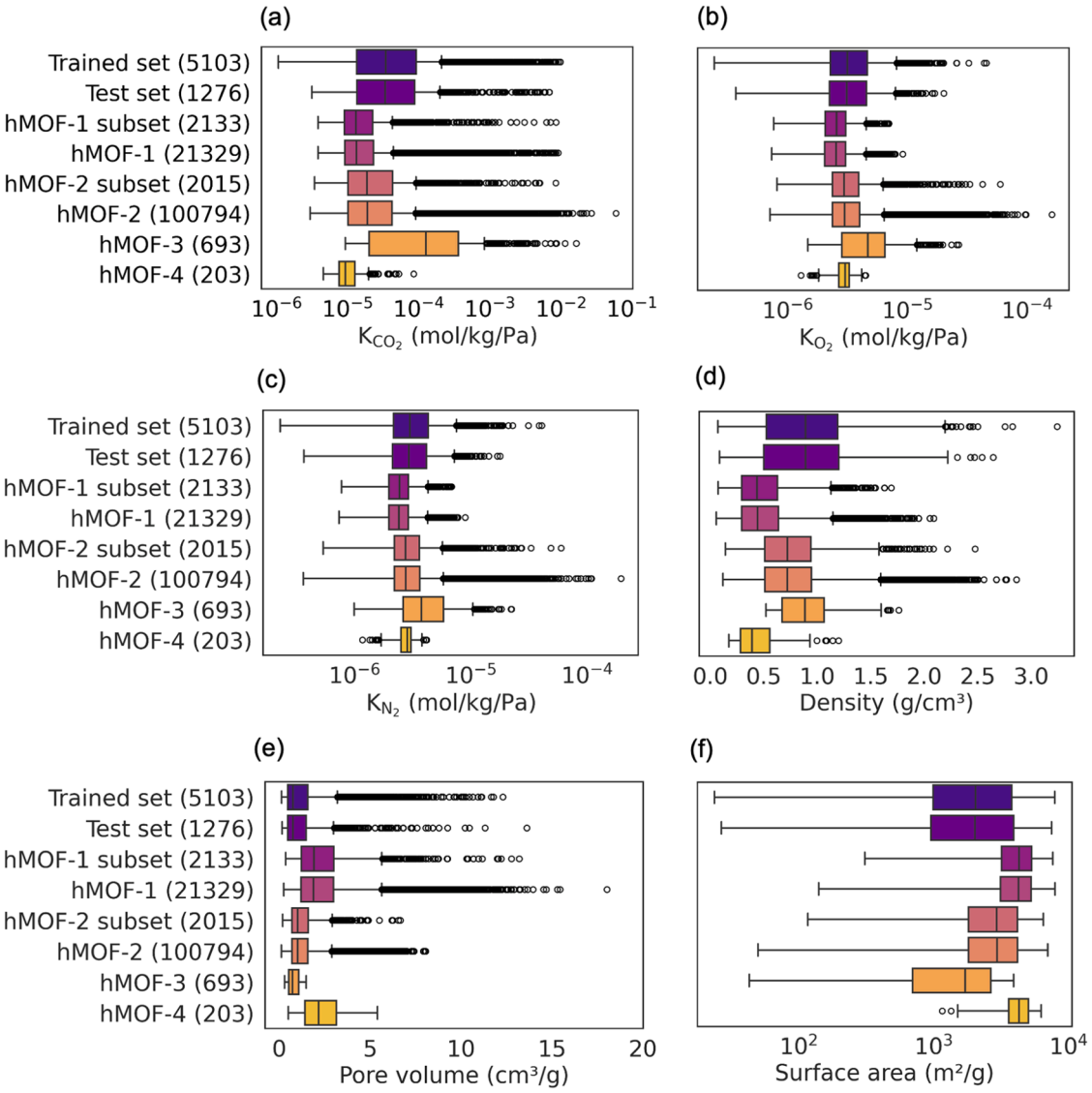


Fig. 3. Box plots comparing the key features of different MOF databases. (a) K_{CO_2}, (b) K_{O_2}, (c) K_{N_2}, (d) density, (e) pore volume, and (f) accessible surface area. Number of materials present in each database is given in parentheses.

demonstrates the results of SHAP analysis for predicting CO_2, O_2, and N_2 uptakes of MOFs at 1 bar, whereas results at 0.1 and 10 bar are given in **Figs. S6-S7**. The SHAP value plots provide a detailed breakdown of how individual features influence the model's predictions both in terms of magnitude and direction. For CO_2, K_{CO_2} and LCD are the most important descriptors at 0.1 and 1 bar. High K_{CO_2} and low LCD values are associated with increased CO_2 uptakes. Among the chemical features, the percentage of halogen atoms is the most important one influencing CO_2 adsorption at low pressures, with higher halogen content enhancing adsorption through electrostatic interactions. High accessible surface area and high pore volume generally yield high CO_2 uptakes at 10 bar, as shown in **Fig. S7.** For O_2 and N_2, K_{O_2} and K_{N_2} have profound effects at all pressures. At 10 bar, the importance of density, pore volume, and accessible surface area increases substantially, with MOFs having high accessible surface area and low density emerging as promising candidates for O_2 and N_2 adsorption.

Using these ML models, we predicted CO_2, O_2, and N_2 adsorption data for all MOFs available in four different hMOF databases. We first demonstrated the transferability of our ML models to two large hMOF databases (hMOF-1 and hMOF-2) by comparing ML-predicted gas uptakes with the simulation results. ML predictions for 2133 hMOF-1 materials are closely aligned with the simulation data along the parity line for CO_2 adsorption at all pressures (high R^2 between 0.87–0.91), as shown in **Fig. S8(a-c)**. ML predictions for CO_2 uptake in the majority of hMOF-1 subset (1846, 1608, and 1579 materials at 0.1, 1, and 10 bar, respectively) are within the ±10 % error margin. For O_2 and N_2 adsorption, there is a perfect agreement at all pressures (R^2: 0.99), as shown in **Fig. S8(d-f). Fig. S9** compares ML-predicted and simulated gas uptakes for hMOF-2 subset consisting of 2015 structures. **Fig. S9(a-c)** shows that predictions agree with the simulated CO_2 adsorption data at all pressures (high R^2 between 0.89–0.93). ML-predicted CO_2 uptakes of hMOF-2 subset are within the ±10 % error margin for 1515, 1061, and 1086 materials at 0.1, 1, and 10 bar, respectively. ML-predicted and simulated O_2 and N_2 adsorption data show strong agreement across all pressures, with R^2 values ranging from 0.97 to 0.99, as depicted in **Fig. S9(d-f)**. Overall, we concluded that the new ML models are transferable to hMOF databases and capable of accurately predicting their CO_2, O_2, and N_2 adsorption properties.

Fig. 5 represents the gas adsorption properties of all five MOF databases computed by ML models. At 1 bar, CO_2 uptakes for CoRE MOF, hMOF-1, hMOF-2, hMOF-3, and hMOF-4 databases were observed to be between 0.2–9.9, 0.4–7.7, 0.5–8.4, 0.9–6.4, and 0.6–4.2 mol/kg,

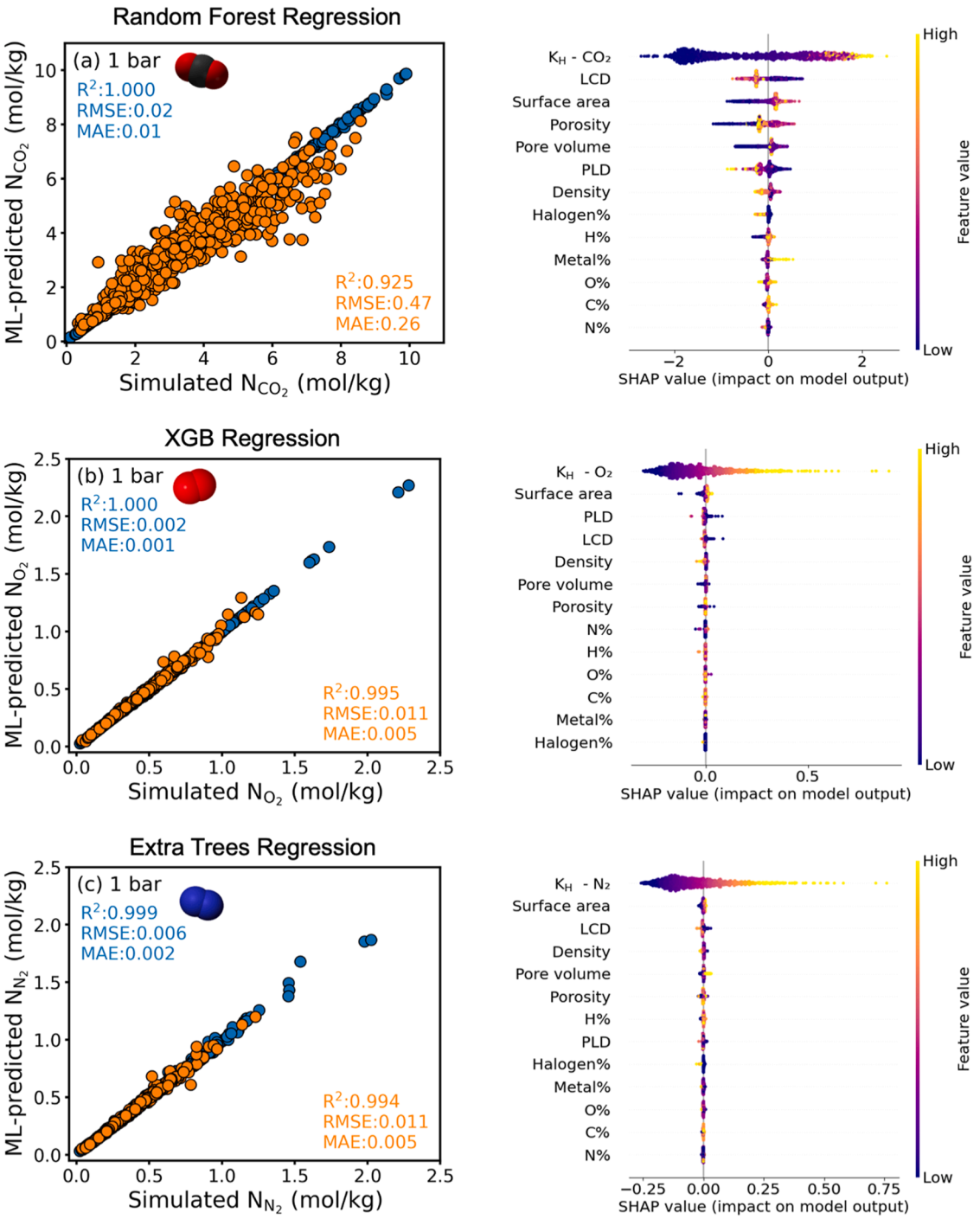


Fig. 4. Comparison of ML-predicted and simulated (a) CO_2, (b) O_2, (c) N_2 uptakes of 6379 MOFs at 1 bar. Blue (orange) points represent trained (test) sets. SHAP value analysis highlighting the impact of materials' features on the predictions is given on the right-hand side.

respectively, as shown in Fig. 5(a). Uptakes are significantly higher in MOFs with narrow pore sizes (PLD <10 Å) as CO_2 strongly interacts with the pore walls. Materials in CoRE MOF and hMOF-3 databases predominantly fall within PLD <10 Å range, but CoRE MOFs exhibit greater variation in CO_2 uptake, which can be attributed to their metal, linker, and topological diversities. When we ranked the materials based on their CO_2 uptakes at 1 bar, we observed that 845 of the top 1000 MOFs (uptakes between 6.7–9.9 mol/kg) originate from hMOF-2 database followed by 144 MOFs from CoRE MOF database. Fig. 5(b) shows that the highest CO_2 uptakes observed in CoRE MOF, hMOF-1, hMOF-2, and hMOF-4 databases are quite similar at 10 bar, reaching to 23.8, 23.2, 22.7, and 21.6 mol/kg, respectively, except for hMOF-3 database for which the highest uptake was 18.1 mol/kg. A closer look at the top 1000 MOFs at 10 bar suggests that hMOF-2 database remains the dominant with 697 materials (uptakes between 18.8–23.8 mol/kg), followed by 229 and 71 MOFs from hMOF-1 and CoRE MOF databases, respectively.

Fig. 5(c) shows that the highest O_2 uptakes are observed for CoRE MOFs up to 2.3 mol/kg at 1 bar, followed by hMOF-2, hMOF-3, hMOF-1, and hMOF-4 databases exhibiting uptakes up to 1.9, 1.7, 0.7, and 0.4 mol/kg, respectively. Among the top 1000 MOFs (uptakes between 1.1–2.3 mol/kg), hMOF-2 database has 945 structures. Fig. 5(d) demonstrates that at 10 bar, hMOF-1 database shows high O_2 uptakes up to

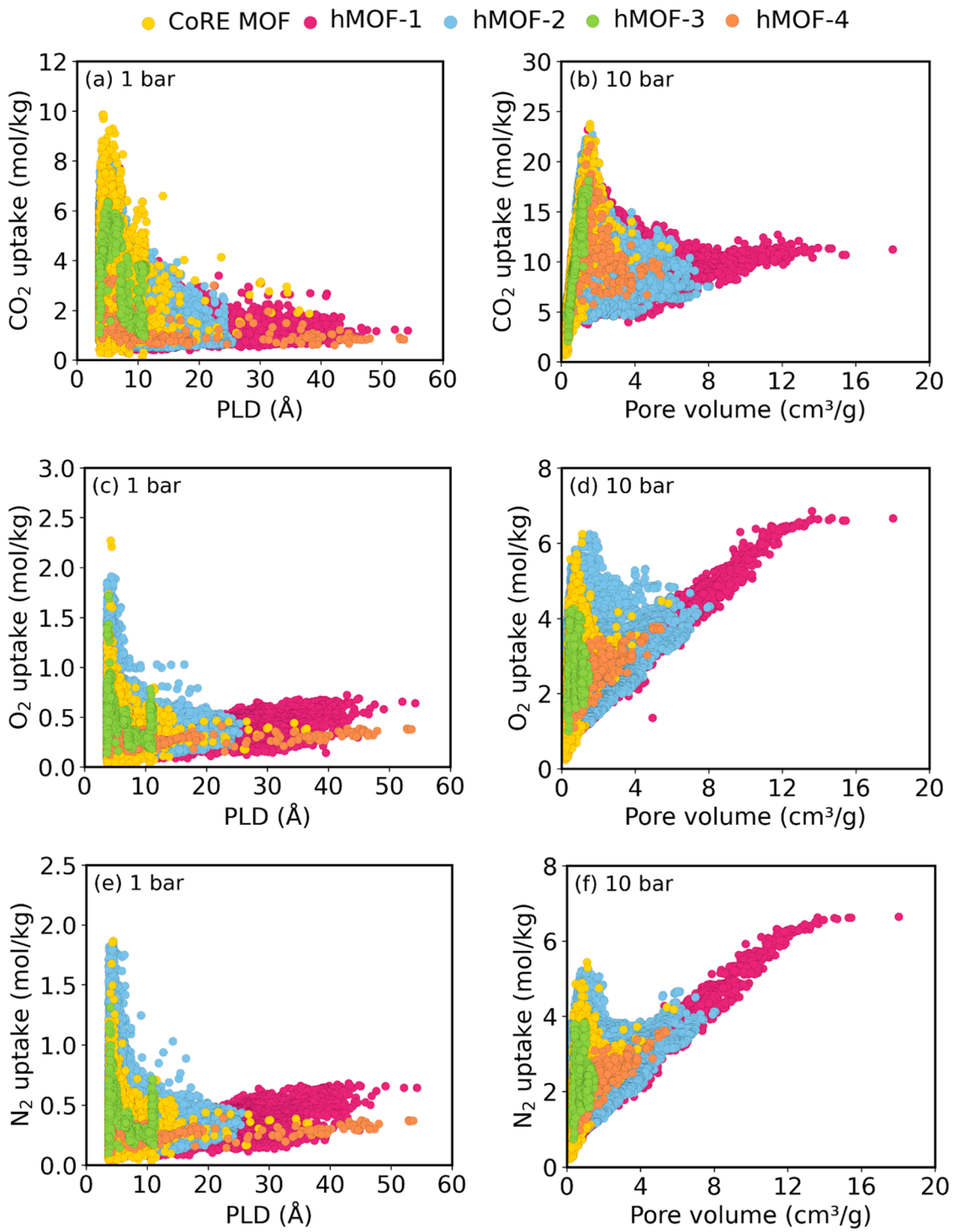


Fig. 5. CO_2, O_2, and N_2 uptakes of 129,398 MOFs available in five different databases at (a, c, e) 1 bar, (b, d, f) 10 bar.

6.9 mol/kg, followed by hMOF-2, CoRE MOF, hMOF-3, and hMOF-4 databases with uptakes up to 6.2, 6.2, 4.3, and 3.8 mol/kg, respectively. hMOF-1 contributes 275 materials among the top 1000 (uptakes between 5.0–6.9 mol/kg) despite not appearing at 1 bar. hMOF-2 remains the most dominant database with 704 MOFs, while CoRE MOFs comprise 21 materials. Results for N_2 adsorption are presented in Fig. 5 (e, F), reflecting similar trends to those seen for O_2. The highest N_2 uptakes are observed for CoRE MOF and hMOF-2 databases reaching1.9 and 1.8 mol/kg at 1 bar, respectively. Among the top 1000 MOFs (uptakes between 1.0–1.9 mol/kg)), hMOF-2 and CoRE databases have 959 and 31 materials, respectively. Fig. 5(f) shows that hMOF-1 database shows high N_2 uptakes up to 6.6 mol/kg, followed by CoRE MOF, hMOF-2, hMOF-3, and hMOF-4 databases with uptakes up to 5.4, 5.2, 3.8, and 3.6 mol/kg, respectively. We note that ML-predicted gas uptakes of all MOFs at 0.1 bar are given in **Fig. S10**. CoRE MOF and hMOF-2 databases exhibit CO_2 uptakes between 0.03–5.7 and 0.05–4.0 mol/kg, respectively. hMOF-2 database also exhibits the highest O_2 and N_2 uptakes, between 0.01–1.5 and 0.01–1.0 mol/kg, respectively, at low pressure of 0.1 bar.

Our ML models were further utilized to assess the CO_2/N_2 and O_2/N_2

selectivities of all MOFs. To check the power of our ML models in accurately assessing MOFs' selectivities, we first compared ML-predicted and simulated selectivities of 6379 MOFs (4009 CoRE MOFs and 2370 hMOF-1) for which we have the gas uptakes from the molecular simulations in **Figs. S11-S12**. Strong agreement was observed between ML and simulation data. For example, at 0.1 bar, ML-predicted CO_2/N_2 and O_2/N_2 selectivities, 1.6–345.1 and 0.4–2.3, closely matched the simulation results, 1.6–374.3 and 0.3–2.3, respectively. At 1 bar, ML-predicted (simulated) CO_2/N_2 selectivities decreased to 1.6–96.1 (1.6–102.6), while O_2/N_2 selectivities remained almost unchanged, 0.6–2.2 by ML and 0.5–2.3 by simulations. The decrease in selectivities became more prominent as the pressure was increased to 10 bar since the most favorable adsorption sites are already saturated and adsorption shifts to less favorable sites. ML-predicted CO_2/N_2 and O_2/N_2 selectivities at 10 bar, 1.0–15.4 and 0.7–1.9, respectively, are in good agreement with the simulated ones, ranging from 1.1–17.8 and 0.7–1.9.

The ML-predicted selectivities of all five MOF databases are presented in Fig. 6. CO_2/N_2 selectivities are between 1.5–345.1, 1.3–96.1, and 0.6–15.4 at 0.1, 1, and 10 bar, respectively, in Fig. 6(a-c). CoRE MOF and hMOF-1 databases exhibit the highest CO_2/N_2 selectivities, particularly at 0.1 and 1 bar, up to 345.1 and 96.1 for CoRE MOF and 245.5 and 57.5 for hMOF-1 database. 76, 26, and 70% of the structures

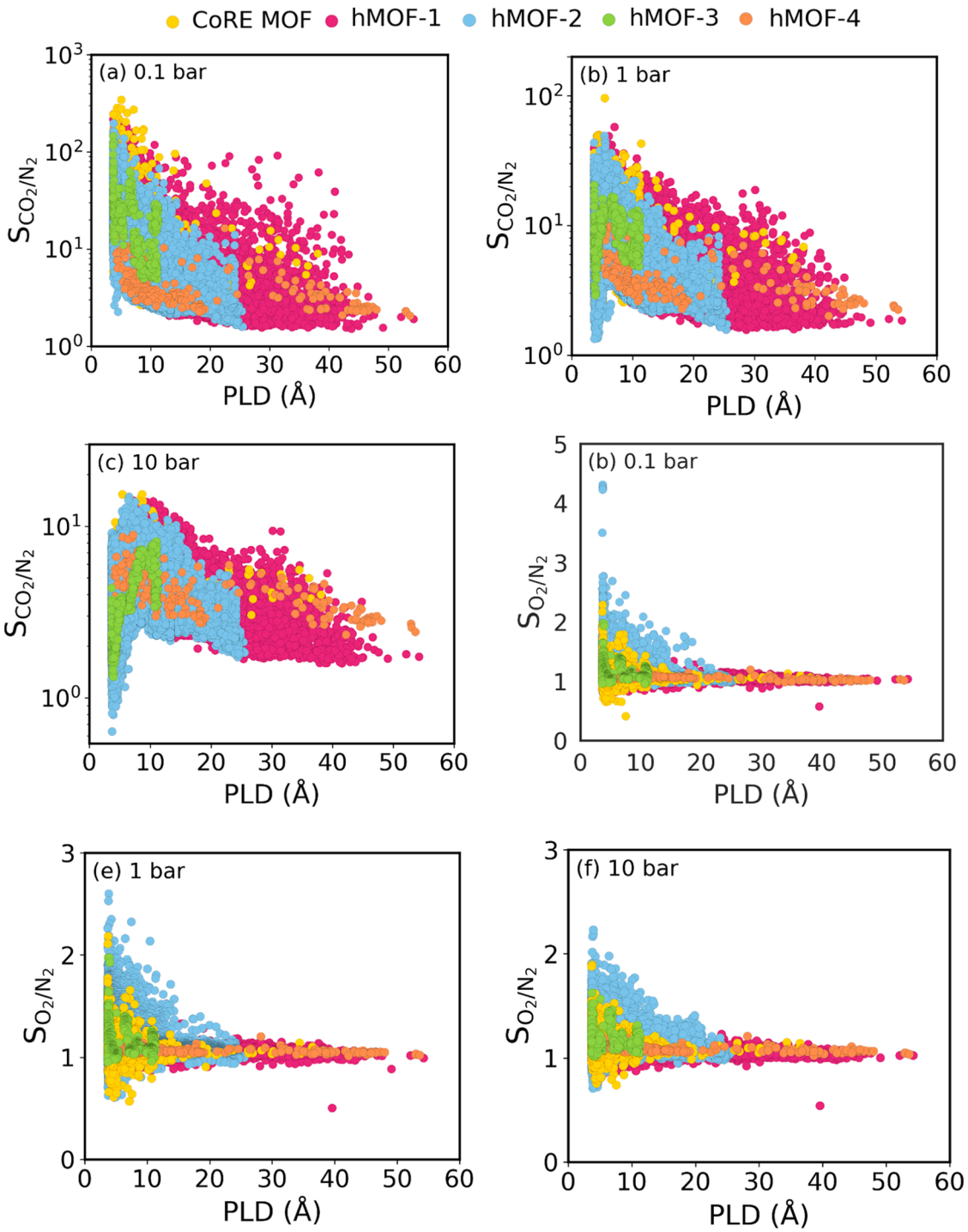


Fig. 6. CO_2/N_2 and O_2/N_2 selectivities of different MOF databases composed of 129,398 structures at (a, d) 0.1, (b, e) 1, and (c, f) 10 bar.

available in CoRE MOF, hMOF-1, and hMOF-3 databases exhibit CO_2/N_2 selectivity >10 at 0.1 bar, respectively. Higher CO_2/N_2 selectivities are generally associated with MOFs having narrower PLDs (<10 Å). hMOF-1 database exhibits a broad range of PLDs and selectivities compared to other databases, and hMOF-3 database shows clustering in selectivities associated with narrow PLDs. Fig. 6(d-f) shows that O_2/N_2 selectivities of MOFs are generally very low, between 0.4–4.3, 0.5–2.6, 0.5–2.3, at 0.1, 1, and 10 bar, respectively. However, an exception is observed for 2 and 28 MOFs from CoRE MOF and hMOF-2 databases, respectively, which exhibit selectivities greater than 2 at 1 bar. Overall, these results demonstrate the significant potential of MOFs for selective CO_2/N_2 separation while indicating MOF's limited capability to separate O_2 from N_2.

We also compared CO_2/N_2 and O_2/N_2 separation performances of CoRE MOFs and hMOFs with the experimentally reported data available in the literature. The single-component CO_2 and N_2 adsorption measurements were performed for several MOFs including ZIF-8 [58], NH_2-MIL-53(Al) [59], Ni-BDP-COOH [60], Cu-BTC [61], UiO-66-NH_2 [62], UiO-66 [62], and HNUST-5 [63], and their CO_2/N_2 selectivities were reported as 7.5, 12.5, 15.8, 18.0, 20.0, and 32.5, respectively, at 1 bar, 298 K. Based on the selectivities computed for CoRE MOFs up to 96.1 and for hMOFs up to 57.5, in this work, we infer that numerous MOFs with high CO_2/N_2 selectivities have either not been tested for this gas separation or have yet to be synthesized. Similarly, single-component O_2 and N_2 adsorption measurements were reported for Fe/Cu-BTC [64], UMCM-1 [65], MOF-177 [66], and $Co_2(OH)_2$(BBTA) [67], and their O_2/N_2 selectivities were reported as 1.1, 1.6, 1.8, and 5.2, respectively, at 1 bar, 298 K. O_2/N_2 selectivities we computed for the MOF space were mostly around 1.5–2, comparable to or better than those reported in the literature. However, we note that these O_2/N_2 selectivities are still low, showing that MOFs may not be very efficient adsorbents for this challenging gas separation.

We explored the use of ML as a faster alternative to molecular simulations for ranking materials and identifying the top-performing ones with the highest selectivities. We initially ranked 6379 MOFs for which we have the GCMC simulation data based on their selectivities and then re-ranked them based on ML-predicted selectivities to identify the top 20 materials. At 0.1 bar, ML successfully identified the top 19 (17) MOFs for CO_2/N_2 (O_2/N_2) separation, while at 1 and 10 bar, ML identified the top 17 (18) and 18 (11) MOFs for CO_2/N_2 (O_2/N_2) separation, respectively. The Spearman Rank Correlation Coefficient (SRCC) values were calculated for the 6379 MOFs used in the ML model development as 0.998, 0.992, and 0.989 (0.934, 0.972, and 0.961) at 0.1, 1, and 10 bar, respectively, for CO_2/N_2 (O_2/N_2) separation. Thus, we concluded that ML models can effectively capture material rankings and reliably identify the top-performing MOFs.

Motivated from this, we used the ML models to identify the top MOFs among 129,398 materials. We focused on 1 bar and highlighted the top 20 materials within each MOF database for CO_2/N_2 separation, whereas all databases were considered to identify the top 10 MOFs for O_2/N_2 separation due to the limited number of materials exhibiting O_2/N_2 selectivity >2. The most common metal and linker types, topologies, the range of pore sizes, surface areas, porosities, and pore volumes for the top 20 MOFs for CO_2/N_2 separation are presented in **Table S7**. CO_2/N_2 selectivities of the top 20 CoRE MOFs are between 35.0–96.1, and they have small-medium PLDs (3.7–11.4 Å), low-mediocre surface areas (367–2436 m^2/g), and mediocre porosities (0.5–0.8). 14 of them contain a variety of lanthanides (Er, Eu, Gd, La, Sm, Tb, and Tm), in agreement with previous studies showing that lanthanide-based MOFs are promising for CO_2 separation [13,14]. The remaining top six synthesized MOFs contain transition metals; Cu, V, Zn, and Ni. The top CoRE MOFs predominantly have oxalate, pyridine dicarboxylate, pyrazine dicarboxylate, benzene tetracarboxylate, and fumarate linkers, which were reported to be useful for CO_2 capture [68–71].

For the top 20 hMOFs in hMOF-1 database, CO_2/N_2 selectivities are between 37.5–57.5, and these hMOFs possess narrow pore sizes, PLDs (3.7–7.0 Å), mediocre surface areas (533–2650 m^2/g), and porosities (0.55–0.75). They all have Ni metal, except for one having Pb. Most of them have fluorine, bromine, and cyano-functionalized aromatic linkers, known to enhance CO_2 adsorption through electrostatic interactions [72]. The top 20 materials from hMOF-2 database also have narrow pore sizes, PLDs (3.7–6.0 Å), mediocre accessible surface areas (230–2382 m^2/g), and porosities (0.56–0.73). 16 of them contain Zn, with three featuring V, and one having Cu. Nearly all structures exhibit the pcu topology, which is unsurprising given that over 90 % of the materials in this database have this topology [73]. In terms of organic linkers, fluorine functionalized bicyclo[2.2.2]octane carboxylates are dominant, which were shown to be promising for other gas separations, such as Xe/Kr [74] and hexane separation [75]. The top 20 hMOFs identified from hMOF-3 database have lower CO_2/N_2 selectivities, 15.7–20.5, compared to other databases. These hMOFs possess PLDs, surface areas, and porosities between 3.7–7.0 Å, 224–1449 m^2/g, and 0.48–0.63, respectively. 17 of them have 1,4-bis(4-pyridyl)benzene linkers, while the rest have pyrazine linkers. They include all the metal types present in the database (Ni, Fe, Cu, Co, and Zn) and share the pcu topology. For hMOF-4 database, CO_2/N_2 selectivities of the top 20 materials are the lowest, 6.9–11.4, among all databases, which exhibit low to high PLDs (4.1–28.2 Å), mediocre to large surface areas between 1164–4537 m^2/g and mediocre porosities (0.60–0.80).

Table S8 summarizes the most common metals, linker types, and topologies, as well as the ranges of pore sizes, accessible surface areas, porosities, and pore volumes for the top 10 MOFs for O_2/N_2 separation. 8 of them are from hMOF-2 database, and the remaining are CoRE MOFs. The O_2/N_2 selectivities are between 2.1–2.6, and these MOFs possess narrow PLDs (3.7–7.4 Å), low accessible surface areas (up to 1075 m^2/g), and low porosities (0.43–0.62). These MOFs possess transition metals (Cu, Fe, V, and Zn), mostly benzene dicarboxylate or heterocyclic linkers (pyridine, tetrazole, imidazole), and distinct topologies (dia, hms, pcu, rna, and sqc).

We finally present the top three MOFs for each separation, in Fig. 7. The most selective candidate for CO_2/N_2 separation is OJONOQ [76], which is a synthesized MOF from CoRE MOF database having a selectivity of 96.1 at 1 bar. This MOF was previously shown to be promising in two different HTCS studies for CO_2/N_2 and CO_2/CH_4 mixture separations [13,14]. We identified this MOF as a promising candidate among 129,398 MOFs using ML without performing HTCS calculations for every single material. This result highlights both the accuracy and efficiency of integrating ML with HTCS to rapidly identify the most useful materials by saving enormous computational time and resources. The second top-performing MOF is a hypothetical material, ddmof_7590 from hMOF-1 database, with a selectivity of 57.5, showing the potential of hMOFs as promising adsorbents for gas separation and underscoring the necessity of their synthesis and testing. The third one is from CoRE MOF database, FAZFET [77], which has a CO_2/N_2 selectivity of 50. For O_2/N_2 separation, all top three MOFs are from the hMOF-2 database. These hMOFs, identified as hMOF-5029,561, hMOF-5080,140, and hMOF-1002,995, exhibit O_2/N_2 selectivities as 2.6, 2.5, and 2.3, respectively.

After analyzing the linkers of the top MOFs that we identified, a comprehensive literature review was done to understand the gas preference of these linkers. DFT studies have shown that CO_2 can interact strongly with oxalate-like environments, particularly on electron-rich surfaces or in the presence of metal centers that enhance electrostatic and polarization-based interactions [79]. For example, oxalate-linked frameworks can stabilize CO_2 via electrostatic interactions, which are not as favorable for N_2 due to its lower quadrupole moment and lack of polarizability [80]. The fluorination of benzene rings enhances their interaction with CO_2 molecules. The introduction of fluorine atoms increases the electron-withdrawing character of the aromatic ring, leading to a higher quadrupole moment and polarizability. For instance, benzene and its 12 fluorinated derivatives were analyzed in detail, revealing that increased fluorination correlates with enhanced CO_2 interactions

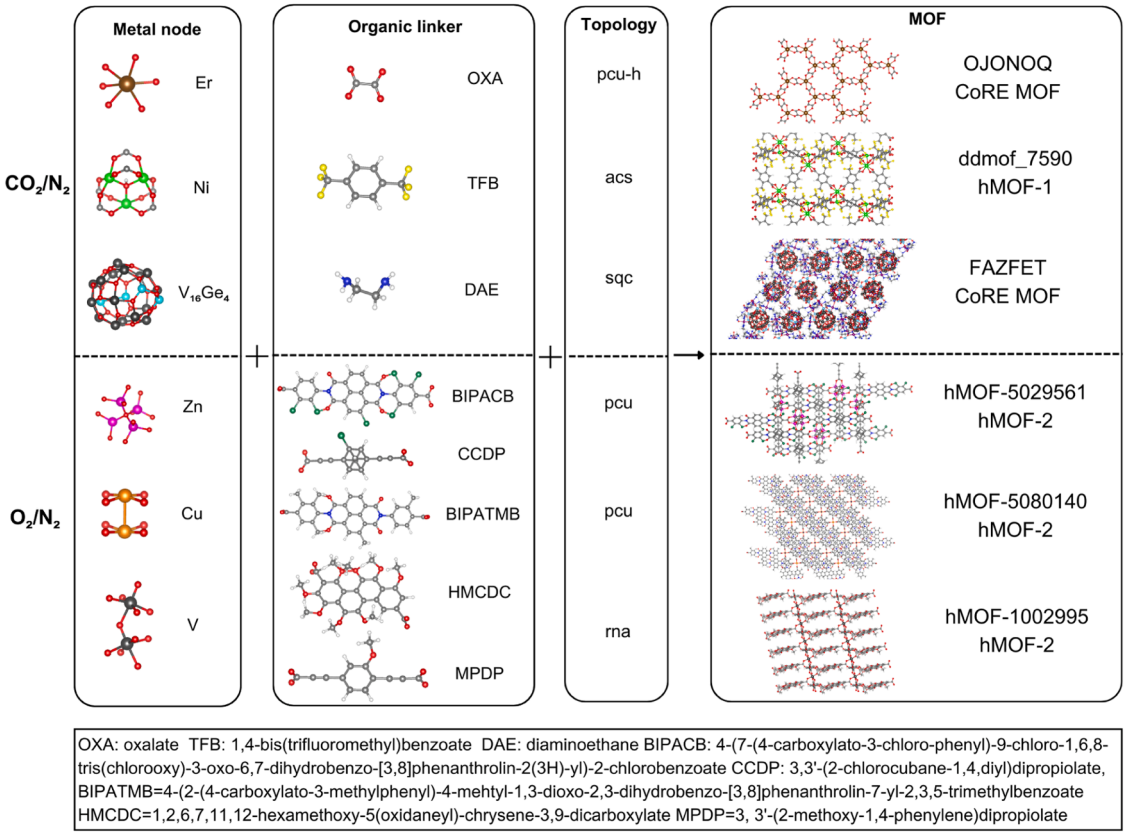


Fig. 7. Metal nodes, organic linkers, and topologies of the top three MOFs identified for CO_2/N_2 and O_2/N_2 separations at 1 bar, 298 K. Carbon, hydrogen, oxygen, nitrogen, fluorine, chlorine, erbium, nickel, cobalt, copper, germanium, zinc, and vanadium atoms in the framework are represented with gray, white, red, blue, yellow, dark green, brown, green, dark blue, orange, cyan, magenta, and dark gray colors, respectively. Linker names were identified using the ChemDraw software [78].

due to elevated quadrupole moments and dipole moments [81]. These findings suggest that incorporating fluorinated aromatic linkers into MOFs can improve CO_2 adsorption selectivity over N_2. DFT studies also reported that CO_2 interacts strongly with diaminoethane through hydrogen bonds and electrostatic interactions [82]. The two amino groups act as electron donors, engaging with the electrophilic carbon center and partially negative oxygen atoms of CO_2. These interactions stabilize the CO_2 molecule in close proximity to the diamine, significantly lowering the system's total energy compared to weakly interacting species like N_2. The directional nature and strength of these interactions contribute to high CO_2/N_2 selectivity.

In our computational study, we employed several methods and assumptions to investigate a large number of MOFs, and we concluded by evaluating the validity of these approaches: After showing the highly accurate single-component gas uptake predictions of our ML models, we used these models to compute the ideal selectivities of MOFs for CO_2/N_2 and O_2/N_2 separations. Ideal selectivity refers to the ratio of the single-component uptakes of two gases in each material under the same conditions. Ideal selectivities provide an important insight into the gas separation performance of MOFs, and they have been widely used to screen large numbers of MOFs in high-throughput computational screening studies [83–85]. However, mixture selectivities can be different than the ideal selectivities due to the competition effects between different guest molecules for the same adsorption sites. We therefore performed mixture GCMC simulations for the top-performing MOFs identified by our ML models for CO_2/N_2:15/85 and O_2/N_2:21/79 separations at 298 K, 1 bar. Results presented in **Fig. S13** demonstrated that MOFs exhibiting high ideal selectivities derived from single-component gas adsorption data also maintain high (or even higher) selectivity under mixture conditions, showing that screening materials based on their ideal selectivities is an efficient and reliable preliminary screening approach.

To assess the robustness of our simulation approach, we reassigned partial atomic charges for the top-performing MOFs using the PACMAN (Partial Atomic Charge Predictor for Nanoporous Materials) charges [86], which provides the density functional theory (DFT)-quality charges consistent with those derived from the density derived electrostatic and chemical (DDEC) method. We then performed additional GCMC simulations for the top MOFs with the PACMAN-assigned charges and compared the results with those obtained for the top MOFs with the Qeq-assigned charges. As shown in **Fig. S14**, O_2 and N_2 uptakes exhibited close agreement across both charge assignment schemes. For CO_2, some deviations in uptake were observed, which is attributed to its strong electrostatic interactions, but these differences did not significantly change the calculated CO_2/N_2 and O_2/N_2 selectivities, as the relative adsorption trends remained consistent. Furthermore, increasing the number of GCMC cycles (from 10,000 initialization and 20,000 production to 20,000 initialization and 40,000 production cycles) led to negligible changes in uptake values, further confirming the reliability of our initial simulation parameters.

While MOFs offer exceptional tunability and high performance for gas adsorption and separation applications, the transition from computational discovery to industrial usage should consider practical factors, such as cost, sustainability, and scalability. The production cost of MOFs is influenced by the choice of metal nodes and organic linkers. For instance, frameworks utilizing earth-abundant metals like iron and aluminum, and synthesized via optimized routes, can achieve production costs below $30 kg [87]. Developing sustainable synthesis approaches, including solvent-free methods and the use of green solvents, is important to minimize environmental impact [88]. Scalability

remains a challenge, but advancements, such as continuous-flow synthesis and mechanochemical methods, have demonstrated potential for large-scale MOF production [89,90]. Despite these challenges, examining hypothetical MOFs remains valuable, as prior literature has demonstrated that hypothetical MOFs are synthesizable [91,92] and can be just as promising as synthesized MOFs in terms of separation performances, suggesting that it is still valuable to explore the broader chemical design space, including structures that are not yet synthesized.

4. Conclusion

In this work, we integrated HTCS and ML to systematically analyze CO_2, O_2, and N_2 adsorption properties in addition to CO_2/N_2 and O_2/N_2 separation performances of nearly 130,000 MOFs available across five different MOF databases. While traditional HTCS of this large material spectrum can take several months to complete all the molecular simulations using 300 CPUs, the ML models we developed in this study predicted gas uptakes for all MOFs in just 15 minutes on a standard computer equipped with 12 CPUs. Our ML models predicted the gas adsorption and separation performances of MOFs using a very small number of descriptors, just 13 easily computable structural, chemical, and energetic features of materials. Therefore, these models can be further transferred to assess CO_2, O_2, and N_2 adsorption properties of new MOFs, either synthetic or hypothetic, in the future. Both the MOFs' features and our ML models were made publicly available to support further applications of our models in different gas adsorption studies. Our results revealed that synthesized MOFs exhibit the highest CO_2 uptakes and CO_2/N_2 selectivities, while hMOFs mostly demonstrate higher O_2 and N_2 uptakes along with higher O_2/N_2 selectivities, highlighting the critical role of a targeted design approach in the development of MOFs for efficient gas adsorption and separation processes. We concluded that synthesized MOFs with narrow pores, lanthanide metals, and linkers incorporating oxalate, pyridine dicarboxylate, and fumarate demonstrate the highest CO_2/N_2 selectivities, while hypothetical MOFs with narrow pores, transition metals, and benzene-based linkers also exhibit strong CO_2/N_2 separation performance. Our models will be useful for providing valuable insights for experimentalists in designing and synthesizing novel materials for CO_2 capture and air separation applications.

Funding

S.K. acknowledges funding by the European Union (ERC, STARLET, 101,124,002). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them.

CRediT authorship contribution statement

Hasan Can Gulbalkan: Writing – review & editing, Writing – original draft, Software, Methodology, Formal analysis, Data curation. **Seda Keskin:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Methodology, Funding acquisition, Data curation, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Seda Keskin reports financial support was provided by European Commission. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

S.K. acknowledges funding by the European Union (ERC, STARLET, 101124002). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ceja.2025.100984.

Data availability

The ML scripts developed in this study are openly available on GitHub at https://github.com/hasancangulbalkan/MOFs_CO2N2O2_ML

References

[1] C.Y. Chuah, W. Li, Y. Yang, T.-H. Bae, Evaluation of porous adsorbents for CO_2 capture under humid conditions: the importance of recyclability, Chem. Eng. J. Adv. 3 (2020) 100021.
[2] W. Gao, S. Liang, R. Wang, Q. Jiang, Y. Zhang, Q. Zheng, B. Xie, C.Y. Toe, X. Zhu, J. Wang, Industrial carbon dioxide capture and utilization: state of the art and future challenges, Chem. Soc. Rev. 49 (2020) 8584–8686.
[3] Z. Ji, H. Wang, S. Canossa, S. Wuttke, O.M. Yaghi, Pore chemistry of metal–organic frameworks, Adv. Funct. Mater. 30 (2020) 2000238.
[4] D.-W. Lim, H. Kitagawa, Rational strategies for proton-conductive metal–organic frameworks, Chem. Soc. Rev. 50 (2021) 6349–6368.
[5] V.J. Ariyamparambil, B. Kandasubramanian, A mini-review on the recent advancement of electrospun MOF-derived nanofibers for energy storage, Chem. Eng. J. Adv. 11 (2022) 100355.
[6] I.J. Bruno, J.C. Cole, P.R. Edgington, M. Kessler, C.F. Macrae, P. McCabe, J. Pearson, R. Taylor, New software for searching the Cambridge structural database and visualizing crystal structures, Acta Crystallogr. Sect. B Struct. Sci 58 (2002) 389–397.
[7] C.E. Wilmer, M. Leaf, C.Y. Lee, O.K. Farha, B.G. Hauser, J.T. Hupp, R.Q. Snurr, Large-scale screening of hypothetical metal–organic frameworks, Nat. Chem. 4 (2012) 83–89.
[8] S. Majumdar, S.M. Moosavi, K.M. Jablonka, D. Ongari, B. Smit, Diversifying databases of metal organic frameworks for high-throughput computational screening, ACS Appl. Mater. Interfaces 13 (2021) 61004–61014.
[9] C. Gu, Z. Yu, J. Liu, D.S. Sholl, Construction of an anion-pillared MOF database and the screening of MOFs suitable for Xe/Kr separation, ACS Appl. Mater. Interfaces 13 (2021) 11039–11049.
[10] D.A. Gomez-Gualdron, O.V. Gutov, V. Krungleviciute, B. Borah, J.E. Mondloch, J. T. Hupp, T. Yildirim, O.K. Farha, R.Q. Snurr, Computational design of metal–organic frameworks based on stable zirconium building units for storage and delivery of methane, Chem. Mater. 26 (2014) 5632–5639.
[11] E. Ren, P. Guilbaud, F.-X. Coudert, High-throughput computational screening of nanoporous materials in targeted applications, Digit. Discov. 1 (2022) 355–374.
[12] Y.J. Colón, R.Q. Snurr, High-throughput computational screening of metal–organic frameworks, Chem. Soc. Rev. 43 (2014) 5735–5749.
[13] C. Altintas, G. Avci, H. Daglar, A. Nemati Vesali Azar, S. Velioglu, I. Erucar, S. Keskin, Database for CO_2 separation performances of MOFs based on computational materials screening, ACS Appl. Mater. Interfaces 10 (2018) 17257–17268.
[14] Z. Qiao, K. Zhang, J. Jiang, In silico screening of 4764 computation-ready, experimental metal–organic frameworks for CO_2 separation, J. Mater. Chem. A 4 (2016) 2105–2114.
[15] C.E. Wilmer, O.K. Farha, Y.-S. Bae, J.T. Hupp, R.Q. Snurr, Structure–property relationships of porous materials for carbon dioxide separation and capture, Energy Environ. Sci. 5 (2012) 9849–9856.
[16] K.T. Leperi, Y.G. Chung, F. You, R.Q. Snurr, Development of a general evaluation metric for rapid screening of adsorbent materials for postcombustion CO_2 capture, ACS Sustain. Chem. Eng. 7 (2019) 11529–11539.
[17] S. Li, Y.G. Chung, C.M. Simon, R.Q. Snurr, High-throughput computational screening of multivariate metal–organic frameworks (MTV-MOFs) for CO_2 capture, J. Phys. Chem. Lett. 8 (2017) 6135–6141.
[18] H. Demir, S. Keskin, Hypothetical yet effective: computational identification of high-performing MOFs for CO_2 capture, Comput. Chem. Eng. 160 (2022) 107705.
[19] H. Demir, S.J. Stoneburner, W. Jeong, D. Ray, X. Zhang, O.K. Farha, C.J. Cramer, J. I. Siepmann, L. Gagliardi, Metal–organic frameworks with metal–catecholates for O_2/N_2 separation, J. Phys. Chem. C 123 (2019) 12935–12946.
[20] E. Gulcay-Ozcan, I. Erucar, Biocompatible MOFs for storage and separation of O_2: a molecular simulation study, Ind. Eng. Chem. Res. 58 (2019) 3225–3237.
[21] J. Abdi, F. Hadavimoghaddam, M. Hadipoor, A. Hemmati-Sarapardeh, Modeling of CO_2 adsorption capacity by porous metal organic frameworks using advanced decision tree-based models, Sci. Rep. 11 (2021) 24468.

[22] G.S. Fanourgakis, K. Gkagkas, E. Tylianakis, G. Froudakis, A generic machine learning algorithm for the prediction of gas adsorption in nanoporous materials, J. Phys. Chem. C 124 (2020) 7117–7126.

[23] I.B. Orhan, H. Daglar, S. Keskin, T.C. Le, R. Babarao, Prediction of O_2/N_2 selectivity in metal–organic frameworks via high-throughput computational screening and machine learning, ACS Appl. Mater. Interfaces 14 (2021) 736–749.

[24] R. Anderson, J. Rodgers, E. Argueta, A. Biong, D.A. Gómez-Gualdrón, Role of pore chemistry and topology in the CO_2 capture capabilities of MOFs: from molecular simulation to machine learning, Chem. Mater. 30 (2018) 6325–6337.

[25] Y. Teng, G. Shan, Interpretable machine learning for materials discovery: Predicting CO2 adsorption properties of metal–organic frameworks, APL Mater. 12 (2024) 081115.

[26] C. Lu, X. Wan, X. Ma, X. Guan, A. Zhu, Deep-learning-based end-to-end predictions of CO_2 capture in metal–Organic frameworks, J. Chem. Inf. Model. 62 (2022) 3281–3290.

[27] Y.G. Chung, E. Haldoupis, B.J. Bucior, M. Haranczyk, S. Lee, H. Zhang, K. D. Vogiatzis, M. Milisavljevic, S. Ling, J.S. Camp, Advances, updates, and analytics for the computation-ready, experimental metal–organic framework database: CoRE MOF 2019, J. Chem. Eng. Data 64 (2019) 5985–5998.

[28] T.F. Willems, C.H. Rycroft, M. Kazi, J.C. Meza, M. Haranczyk, Algorithms and tools for high-throughput geometry-based analysis of crystalline porous materials, Microporous Mesoporous Mater. 149 (2012) 134–141.

[29] Y.G. Chung, E. Haldoupis, B.J. Bucior, M. Haranczyk, S. Lee, H. Zhang, K. D. Vogiatzis, M. Milisavljevic, S. Ling, J.S. Camp, B. Slater, J.I. Siepmann, D. S. Sholl, R.Q. Snurr, Advances, Updates, and Analytics for the Computation-Ready, Experimental Metal–Organic Framework Database, CoRE MOF (2019). https://doi.org/10.5281/zenodo.14184621.

[30] S.-Y. Kim, S.-I. Kim, Y.-S. Bae, Machine-learning-based prediction of methane adsorption isotherms at varied temperatures for experimental adsorbents, J. Phys. Chem. C 124 (2020) 19538–19547.

[31] A.W. Thornton, C.M. Simon, J. Kim, O. Kwon, K.S. Deeg, K. Konstas, S.J. Pas, M. R. Hill, D.A. Winkler, M. Haranczyk, Materials genome in action: identifying the performance limits of physical hydrogen storage, Chem. Mater. 29 (2017) 2844–2854.

[32] S. Wang, M. Cheng, L. Luo, X. Ji, C. Liu, K. Bi, L. Zhou, High-throughput screening of metal–organic frameworks for hydrogen purification, Chem. Eng. J. 451 (2023) 138436.

[33] J. Rogacka, A. Seremak, A. Luna-Triguero, F. Formalik, I. Matito-Martos, L. Firlej, S. Calero, B. Kuchta, High-throughput screening of metal–organic frameworks for CO_2 and CH_4 separation in the presence of water, Chem. Eng. J. 403 (2021) 126392.

[34] Y.G. Chung, D.A. Gómez-Gualdrón, P. Li, K.T. Leperi, P. Deria, H. Zhang, N. A. Vermeulen, J.F. Stoddart, F. You, J.T. Hupp, In silico discovery of metal-organic frameworks for precombustion CO_2 capture using a genetic algorithm, Sci. Adv. 2 (2016) e1600909.

[35] Z. Cai, W. Li, Y.G. Chung, S. Li, T. Liang, T. Wu, Transfer learning-assisted computational screening of metal-organic frameworks and covalent-organic frameworks for the separation of Xe/Kr noble gas, Sep. Purif. Technol. 348 (2024) 127752.

[36] J.A. Gee, K. Zhang, S. Bhattacharyya, J. Bentley, M. Rungta, J.S. Abichandani, D. S. Sholl, S. Nair, Computational identification and experimental evaluation of metal–organic frameworks for xylene enrichment, J. Phys. Chem. C 120 (2016) 12075–12082.

[37] Z. Qiao, Y. Yan, Y. Tang, H. Liang, J. Jiang, Metal–organic frameworks for xylene separation: from computational screening to machine learning, J. Phys. Chem. C 125 (2021) 7839–7848.

[38] B.J. Bucior, N.S. Bobbitt, T. Islamoglu, S. Goswami, A. Gopalan, T. Yildirim, O. K. Farha, N. Bagheri, R.Q. Snurr, Energy-based descriptors to rapidly predict hydrogen storage in metal–organic frameworks, Mol. Syst. Des. Eng. 4 (2019) 162–174.

[39] V. Daoo, J.K. Singh, Accelerating in silico discovery of metal–organic frameworks for ethane/ethylene and propane/propylene separation: a synergistic approach integrating molecular simulation, machine learning, and active learning, ACS Appl. Mater. Interfaces 16 (2024) 6971–6987.

[40] B.J. Sikora, C.E. Wilmer, M.L. Greenfield, R.Q. Snurr, Thermodynamic analysis of Xe/Kr selectivity in over 137000 hypothetical metal–organic frameworks, Chem. Sci. 3 (2012) 2217–2223.

[41] D. Dubbeldam, S. Calero, D.E. Ellis, R.Q. Snurr, RASPA: molecular simulation software for adsorption and diffusion in flexible nanoporous materials, Mol. Simul. 42 (2016) 81–101.

[42] A.K. Rappé, C.J. Casewit, K. Colwell, W.A. Goddard III, W.M. Skiff, UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations, J. Am. Chem. Soc. 114 (1992) 10024–10035.

[43] A.K. Rappe, W.A. Goddard III, Charge equilibration for molecular dynamics simulations, J. Phys. Chem. 95 (1991) 3358–3363.

[44] C.E. Wilmer, R.Q. Snurr, Towards rapid computational screening of metal-organic frameworks for carbon dioxide capture: calculation of framework charges via charge equilibration, Chem. Eng. J. 171 (2011) 775–781.

[45] P.P. Ewald, Die berechnung optischer und elektrostatischer gitterpotentiale, Ann. Phys. 369 (1921) 253–287.

[46] B.L. Eggimann, A.J. Sunnarborg, H.D. Stern, A.P. Bliss, J.I. Siepmann, An online parameter and property database for the TraPPE force field, Mol. Simul. 40 (2014) 101–105.

[47] C. Mellot, J. Lignieres, Monte Carlo simulations of N_2 and O_2 adsorption in silicalite and CaLSX zeolites, Mol. Simul. 18 (1997) 349–365.

[48] C. Murthy, K. Singer, M. Klein, I. McDonald, Pairwise additive effective potentials for nitrogen, Mol. Phys. 41 (1980) 1387–1399.

[49] Z. Sumer, S. Keskin, Adsorption-and membrane-based CH_4/N_2 separation performances of MOFs, Ind. Eng. Chem. Res. 56 (2017) 8713–8722.

[50] O. Durak, N. Habib, H.C. Gulbalkan, A.S. Aydogdu, R. Ozdemir, S. Keskin, A. Uzun, Molecular engineering of a MOF with an ionic liquid sheath for direct air separation at ambient conditions, Chem. Eng. J. (2024) 152608.

[51] D. Frenkel, B. Smit, Understanding Molecular Simulation: from Algorithms to Applications, Elsevier, 2023.

[52] C. Campbell, J.R. Gomes, M. Fischer, M. Jorge, New model for predicting adsorption of polar molecules in metal–organic frameworks with unsaturated metal sites, J. Phys. Chem. Lett. 9 (2018) 3544–3553.

[53] S. Kancharlapalli, R.Q. Snurr, High-throughput screening of the CoRE-MOF-2019 database for CO_2 capture from wet flue gas: a multi-scale modeling strategy, ACS Appl. Mater. Interfaces 15 (2023) 28084–28092.

[54] Y. Yan, Z. Shi, H. Li, L. Li, X. Yang, S. Li, H. Liang, Z. Qiao, Machine learning and in-silico screening of metal–organic frameworks for O_2/N_2 dynamic adsorption and separation, Chem. Eng. J. 427 (2022) 131604.

[55] T.T. Le, W. Fu, J.H. Moore, Scaling tree-based automated machine learning to biomedical big data with a feature set selector, Bioinformatics 36 (2020) 250–256.

[56] P.-G. Martinsson, V. Rokhlin, M. Tygert, A randomized algorithm for the decomposition of matrices, Appl. Comput. Harmon. Anal. 30 (2011) 47–68.

[57] S.M. Lundberg, G. Erion, H. Chen, A. DeGrave, J.M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, S.-I. Lee, From local explanations to global understanding with explainable AI for trees, Nat. Mach. Intell. 2 (2020) 56–67.

[58] F.P. Kinik, C. Altintas, V. Balci, B. Koyuturk, A. Uzun, S. Keskin, [BMIM][PF_6] incorporation doubles CO_2 selectivity of ZIF-8: elucidation of interactions and their consequences on performance, ACS Appl. Mater. Interfaces 8 (2016) 30992–31005.

[59] E. Stavitski, E.A. Pidko, S. Couck, T. Remy, E.J. Hensen, B.M. Weckhuysen, J. Denayer, J. Gascon, F. Kapteijn, Complexity behind CO2 capture on NH2-MIL-53 (Al), Langmuir, 27 (2011) 3970–3976.

[60] T. He, Y.Z. Zhang, H. Wu, X.J. Kong, X.M. Liu, L.H. Xie, Y. Dou, J.R. Li, Functionalized base-stable metal–Organic frameworks for selective CO_2 adsorption and proton conduction, ChemPhysChem 18 (2017) 3245–3252.

[61] K.B. Sezginel, S. Keskin, A. Uzun, Tuning the gas separation performance of CuBTC by ionic liquid incorporation, Langmuir 32 (2016) 1139–1147.

[62] W. Zhang, H. Huang, C. Zhong, D. Liu, Cooperative effect of temperature and linker functionality on CO_2 capture from industrial gas mixtures in metal–organic frameworks: a combined experimental and molecular simulation study, Phys. Chem. Chem. Phys. 14 (2012) 2317–2325.

[63] B. Zheng, H. Wang, Z. Wang, N. Ozaki, C. Hang, X. Luo, L. Huang, W. Zeng, M. Yang, J. Duan, A highly porous rht-type acylamide-functionalized metal–organic framework exhibiting large CO_2 uptake capabilities, Chem. Commun. 52 (2016) 12988–12991.

[64] D.F. Sava Gallis, M.V. Parkes, J.A. Greathouse, X. Zhang, T.M. Nenoff, Enhanced O_2 selectivity versus N_2 by partial metal substitution in Cu-BTC, Chem. Mater. 27 (2015) 2018–2025.

[65] B. Mu, P.M. Schoenecker, K.S. Walton, Gas adsorption study on mesoporous metal−organic framework UMCM-1, J. Phys. Chem. C 114 (2010) 6464–6471.

[66] Y. Li, R.T. Yang, Gas adsorption and storage in metal− organic framework MOF-177, Langmuir 23 (2007) 12937–12944.

[67] A.S. Rosen, M.R. Mian, T. Islamoglu, H. Chen, O.K. Farha, J.M. Notestein, R. Q. Snurr, Tuning the redox activity of metal–organic frameworks for enhanced, selective O_2 binding: design rules and ambient temperature O_2 chemisorption in a cobalt–triazolate framework, J. Am. Chem. Soc. 142 (2020) 4317–4328.

[68] A.A. Barkhordarian, C.J. Kepert, Two new porous UiO-66-type zirconium frameworks; open aromatic N-donor sites and their post-synthetic methylation and metallation, J. Mater. Chem. A 5 (2017) 5612–5618.

[69] J. Kim, A.G. Oliver, G.T. Neumann, J.C. Hicks, Zn-MOFs containing pyridine and bipyridine carboxylate organic linkers and open Zn^{2+} sites, Eur. J. Inorg. Chem. 2015 (2015) 3011–3018.

[70] Y. Zhang, B.E. Lucier, S.M. McKenzie, M. Arhangelskis, A.J. Morris, T. Friščić, J. W. Reid, V.V. Terskikh, M. Chen, Y. Huang, Welcoming gallium-and indium-fumarate MOFs to the family: synthesis, comprehensive characterization, observation of porous hydrophobicity, and CO_2 dynamics, ACS Appl. Mater. Interfaces 10 (2018) 28582–28596.

[71] Q. Yang, S. Vaesen, F. Ragon, A.D. Wiersum, D. Wu, A. Lago, T. Devic, C. Martineau, F. Taulelle, P.L. Llewellyn, A Water Stable Metal-Organic Framework with Optimal Features for CO_2 Capture, Angew. Chem. Int. Ed. 52 (2013) 10316–10320.

[72] A.E. Amooghin, H. Sanaeepur, R. Luque, H. Garcia, B. Chen, Fluorinated metal–organic frameworks for gas separation, Chem. Soc. Rev. 51 (2022) 7427–7508.

[73] B.J. Sikora, R. Winnegar, D.M. Proserpio, R.Q. Snurr, Textural properties of a large collection of computationally constructed MOFs and zeolites, Microporous Mesoporous Mater. 186 (2014) 207–213.

[74] K.B. Idrees, Z. Chen, X. Zhang, M.R. Mian, R.J. Drout, T. Islamoglu, O.K. Farha, Tailoring pore aperture and structural defects in zirconium-based metal–organic frameworks for krypton/xenon separation, Chem. Mater. 32 (2020) 3776–3782.

[75] K.B. Idrees, K.O. Kirlikovali, C. Setter, H. Xie, H. Brand, B. Lal, F. Sha, C.S. Smoljan, X. Wang, T. Islamoglu, Robust carborane-based metal–organic frameworks for hexane separation, J. Am. Chem. Soc. 145 (2023) 23433–23441.

[76] M. Camara, C. Daiguebonne, K. Boubekeur, T. Roisnel, Y. Gérault, C. Baux, F. Le Dret, O. Guillou, *Re*-investigation of the Er^{3+}–$C_2O_4^{2-}$–H_2O system: from the classical ceramic precursor to a new nanoporous molecular material potential precursor, Comptes Rendus Chim. 6 (2003) 405–415.

[77] Y. Gao, Y. Xu, K. Huang, Z. Han, C. Hu, Two three-dimensional $\{V_{16}Ge_4\}$-based open frameworks stabilized by diverse types of Co II-amine bridges and magnetic properties, Dalton Trans. 41 (2012) 6122–6129.
[78] PerkinElmer Informatics, Inc.: Waltham, MA, ChemDraw Prime 21, (1998).
[79] I.C. Popoola, B.X. Shi, F. Berger, A. Michaelides, A. Zen, D. Alfè, Y.S. Al-Hamdani, Cooperative CO_2 capture via oxalate formation on metal-decorated graphene, Phys. Rev. Mater. 9 (2025) 015401.
[80] P. Singh, H.D. Singh, P. Shekhar, C. Jain, P. Song, M. Loloei, M. Edwin, Y. Gao, N. Masoumifard, R. Vaidhyanathan, Ultra-microporous zinc-diaminotriazolate-oxalate metal organic framework with nonlayered-pillared structure showing high carbon dioxide uptakes at low partial pressures, Chem. Mater. 37 (2025) 2496–2505.
[81] D. Ruivo, A.B. Pereiro, J.M. Esperança, J.N. Canongia Lopes, L.P. Rebelo, Rationalizing the diverse solid− liquid equilibria of binary mixtures of benzene and its fluorinated derivatives, J. Phys. Chem. B 114 (2010) 12589–12596.
[82] E. Orestes, C.M. Ronconi, J.W. de Mesquita Carneiro, Insights into the interactions of CO_2 with amines: a DFT benchmark study, Phys. Chem. Chem. Phys. 16 (2014) 17213–17219.
[83] Z. Qiao, C. Peng, J. Zhou, J. Jiang, High-throughput computational screening of 137953 metal–organic frameworks for membrane separation of a $CO_2/N_2/CH_4$ mixture, J. Mater. Chem. A 4 (2016) 15904–15912.
[84] S. Li, Y.G. Chung, R.Q. Snurr, High-throughput screening of metal–organic frameworks for CO2 capture in the presence of water, Langmuir 32 (2016) 10368–10376.
[85] J.S. De Vos, S. Ravichandran, S. Borgmans, L. Vanduyfhuys, P. Van Der Voort, S. M. Rogge, V. Van Speybroeck, High-throughput screening of covalent organic frameworks for carbon capture using machine learning, Chem. Mater. 36 (2024) 4315–4330.
[86] G. Zhao, Y.G. Chung, PACMAN: a robust partial atomic charge predicter for nanoporous materials based on crystal graph convolution networks, J. Chem. Theory Comput. 20 (2024) 5368–5380.
[87] M.I. Severino, C. Freitas, V. Pimenta, F. Nouar, M.L. Pinto, C. Serre, Cost Estimation of the Production of MIL-100 (Fe) at Industrial Scale from Two Upscaled Sustainable Synthesis Routes, Ind. Eng. Chem. Res. 64 (2025) 2708–2718.
[88] D. Chakraborty, A. Yurdusen, G. Mouchaham, F. Nouar, C. Serre, Large-scale production of metal–organic frameworks, Adv. Funct. Mater. 34 (2024) 2309089.
[89] E.G. Rasmussen, J. Kramlich, I.V. Novosselov, Scalable continuous flow metal–organic framework (MOF) synthesis using supercritical CO_2, ACS Sustain. Chem. Eng. 8 (2020) 9680–9689.
[90] T. Paul, A. Juma, R. Alqerem, G. Karanikolos, H.A. Arafat, L.F. Dumée, Scale-up of metal-organic frameworks production: engineering strategies and prospects towards sustainable manufacturing, J. Environ. Chem. Eng. 11 (2023) 111112.
[91] M. Witman, S. Ling, S. Anderson, L. Tong, K.C. Stylianou, B. Slater, B. Smit, M. Haranczyk, In silico design and screening of hypothetical MOF-74 analogs and their experimental synthesis, Chem. Sci. 7 (2016) 6263–6272.
[92] P.G. Boyd, A. Chidambaram, E. García-Díez, C.P. Ireland, T.D. Daff, R. Bounds, A. Gładysiak, P. Schouwink, S.M. Moosavi, M.M. Maroto-Valer, Data-driven design of metal–organic frameworks for wet flue gas CO_2 capture, Nature 576 (2019) 253–256.