JPhys Materials

PURPOSE-LED PUBLISHING™

PAPER • OPEN ACCESS

Finding high-performance MOFs for effective SF_6/N_2 separation through high-throughput computational screening and machine learning

To cite this article: Pelin Sezgin *et al* 2024 *J. Phys. Mater.* **7** 045006

View the article online for updates and enhancements.

You may also like

- Application and Recycling of SF6 gas Mixtures in Gas-insulated Circuit Breaker in Northern China
Wei Liu, Zhenxi Su, Jiong Qi et al.
- Effect of electrode surface roughness on breakdown in SF_6-N_2 and SF_6-CO_2 gas mixtures
Y Qiu and I D Chalmers
- Theoretical study of the decomposition mechanism of SF_6/Cu gas mixtures
Yuwei Fu, Xiaohua Wang, Aijun Yang et al.



This content was downloaded from IP address 88.255.99.39 on 21/08/2025 at 08:13

Journal of Physics: Materials

PAPER

Finding high-performance MOFs for effective SF_6/N_2 separation through high-throughput computational screening and machine learning

Pelin Sezgin, Hasan Can Gulbalkan and Seda Keskin*

Department of Chemical and Biological Engineering, Koç University, Rumelifeneri Yolu, Sariyer, 34450 Istanbul, Turkey
* Author to whom any correspondence should be addressed.

E-mail: skeskin@ku.edu.tr

Keywords: mof, molecular simulation, gas separation, machine learning

Supplementary material for this article is available online

OPEN ACCESS

RECEIVED
19 July 2024

REVISED
24 September 2024

ACCEPTED FOR PUBLICATION
27 September 2024

PUBLISHED
10 October 2024

Original content from this work may be used under the terms of the Creative Commons Attribution 4.0 licence.

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

Abstract

Given the rapidly expanding pool of synthesized and hypothetical metal–organic frameworks (MOFs), testing every single material for SF_6/N_2 separation by iterative experimental methods or computationally demanding molecular simulations is not practical. In this study, we integrated high-throughput computational screening and machine learning (ML) approaches to evaluate SF_6/N_2 mixture adsorption and separation performances of over 25 000 different types of synthesized and hypothetical MOFs (hypoMOFs), representing the largest set of structures studied for SF_6/N_2 separation to date. SF_6/N_2 mixture adsorption data that we produced for synthesized MOFs using molecular simulations were utilized to develop ML models to accurately and quickly predict SF_6 and N_2 uptakes, SF_6/N_2 selectivities, SF_6 working capacities, adsorbent performance scores, and regenerabilities of both synthesized and hypoMOFs. Results showed the MOF space that we studied exhibits very high SF_6/N_2 selectivities in the range of 1.8–4204 at 1 bar in addition to high SF_6 working capacities between 0.04–5.68 mol kg^{-1} at an adsorption pressure of 1 bar and desorption pressure of 0.1 bar at room temperature. The top-performing MOF adsorbents for SF_6/N_2 mixture separation were identified to have Zn, Cu, Ni metals; terphenyl, pyridine, naphthalene linkers; and medium pore sizes. Our comprehensive computational approach offers a highly efficient alternative to brute-force computer simulations by enabling the rapid assessment of the MOF adsorbents for SF_6/N_2 separation and provides molecular insights into the key structural features of the most promising adsorbents.

1. Introduction

Sulfur hexafluoride (SF_6) is extensively used in the electrical industry thanks to its excellent insulating properties and chemical stability, which make it highly effective for use in circuit breakers and transformers [1, 2]. Despite its beneficial applications, SF_6 poses significant environmental concerns as a potent greenhouse gas with a global warming potential that is 23 500 times greater than carbon dioxide (CO_2) when measured over a 100-year period [3]. Consequently, even small concentrations of SF_6 released into the atmosphere can significantly contribute to global warming. In industrial processes, SF_6 is often found mixed with nitrogen (N_2), which is used for the dilution of SF_6. The presence of N_2 in mixtures with SF_6 can complicate recovery and recycling processes, therefore they need to be separated [4]. Compared to conventional cryogenic fractional distillation, which demands significant energy input, adsorption-based SF_6/N_2 separation is an energy-efficient and cost-effective alternative. Thus, identifying promising adsorbent materials for effective SF_6/N_2 separation is essential for mitigating the environmental impact of SF_6.

Advanced materials that can selectively separate SF_6 from N_2 offer a promising solution to mitigate the environmental impact of SF_6. Among these materials, metal–organic frameworks (MOFs) have been recognized as promising adsorbents thanks to their wide range of pore sizes, high surface areas, exceptionally

© 2024 The Author(s). Published by IOP Publishing Ltd

high porosities, and tunable physical and chemical properties [5, 6]. The potential to combine an almost limitless number of metals and organic linkers has resulted in the synthesis of 125 383 different types of MOFs (obtained from the Cambridge Structural Database using the Conquest [7] tool as of July 2024). Similar to the experimentally synthesized MOFs, a large number of hypothetical MOFs (hypoMOFs) have been computationally generated by iterating through possible combinations of building units in various topologies, resulting in millions of structures. Given that the experimental synthesis, characterization, and testing of every single MOF for a target application demand extensive resources, time, and effort, computational simulations play a critical role. Specifically, molecular simulations offer significant potential by accurately computing the gas adsorption and separation properties of MOFs prior to experimental efforts. High-throughput computational screening (HTCS) of thousands of MOFs not only aids in assessing the gas adsorption and separation capabilities of a vast array of materials, thereby guiding experimental pursuits toward the most promising candidates but also helps in elucidating the relationships between a material's structure and its performance [8–10]. While numerous HTCS studies have examined MOFs for various gas separations like CO_2/N_2 [11, 12], CO_2/CH_4 [13, 14], CH_4/H_2 [15, 16], and O_2/N_2 [17, 18] the separation of SF_6 from N_2 has often been under looked.

Only a limited number of HTCS studies have focused on the SF_6/N_2 separation capabilities of MOFs by utilizing grand canonical Monte Carlo (GCMC) simulations. In the first study, 2513 synthesized MOFs were screened for adsorption-based SF_6/N_2 mixture separation using GCMC simulations and a promising MOF, Cu-MOF-NH_2, with a high SF_6/N_2 selectivity ($\sim$254) was identified at 1 bar, 298 K [19]. The second study assessed the SF_6/N_2 mixture separation performances of 2890 synthesized MOFs by GCMC simulations and the promising MOFs were shown to have SF_6/N_2 selectivities between 138–953 at 1 bar and SF_6 working capacities between 1.6–5.3 mol kg^{-1} at an adsorption pressure of 1 bar and desorption pressure of 0.1 bar, 298 K [20]. That work performed molecular simulations combined with ideal adsorption process simulations to screen only experimentally synthesized MOFs. These two HTCS studies highlighted the significant potential of MOF adsorbents to outperform conventional porous materials in SF_6/N_2 separation. However, conducting computationally intensive molecular simulations for the increasing number of experimentally synthesized and computer-generated MOFs is a challenging task. Molecular simulation of SF_6 adsorption is time and resource-consuming due to the presence of high number of interaction sites present in the SF_6 molecule compared to most other gases and their strong electrostatic interactions with the adsorbent material.

Recently, machine learning (ML) has been combined with HTCS to efficiently predict the gas adsorption and separation performances of a vast number of MOFs. This approach does not only save time but also uncovers the hidden correlations between the structural properties of adsorbent materials and their gas separation performances. For example, ML models identified the degree of unsaturation and the number of hydrogen atoms present in the MOF structures as critical factors for identifying the promising MOFs for SF_6/N_2 separation by studying 2877 synthesized MOFs [21] the top-performing MOFs were shown to have the tradeoff between selectivity and working capacity, TSN $>$29.3 mol kg^{-1} and SF_6/N_2 selectivity $>$200 at 1 bar, 298 K. ML models were also developed to accurately predict the tradeoff between selectivity and adsorption quantity, TSQ, for 623 synthesized MOFs and 252 covalent organic frameworks (COFs) at different pressures [22]. HTCS and ML were combined to examine 656 COFs and COF-638 was found to have the most promising results with SF_6/N_2 selectivity of 55.2, SF_6 working capacity of 4.1 mol kg^{-1}, and regenerability $>$80% at an adsorption and desorption pressures of 1 and 0.1 bar, respectively [23]. All these HTCS and ML studies concentrated on a single synthesized MOF database, which only constitutes a small segment of the vast material space. Considering the rapidly expanding pool of hypoMOFs, there can be many other adsorbent candidates that offer more efficient SF_6/N_2 separation compared to the synthesized MOFs.

Inspired by this necessity, we introduced a computational strategy that combines HTCS and ML to evaluate the adsorption and separation performances of both synthesized and hypoMOFs for SF_6/N_2 separation. We analyzed two distinct MOF databases: the computation-ready experimental MOF (CoRE MOF) database [24], composed of 2336 structures, and the hypothetical MOF (hypoMOF) database [25], which comprises 23 393 MOFs. The adsorption data for SF_6/N_2:10/90 mixtures was first produced for CoRE MOFs using GCMC simulations, and then this data was utilized to train ML models that can accurately predict the SF_6/N_2 mixture adsorption data and several adsorbent performance evaluation metrics (including selectivity, working capacity, adsorbent performance score, regenerability) of both CoRE MOFs and hypoMOFs for which computationally demanding molecular simulations were not directly performed. To identify the most promising CoRE and hypoMOFs, all adsorbents were ranked based on a combination of various adsorbent performance evaluation metrics. Detailed analyses of the structural and chemical properties of the top-performing materials highlighted the characteristics of effective adsorbents, which is crucial to accelerate the design of new MOFs with enhanced SF_6/N_2 separation potentials. Our comprehensive computational approach will be a highly efficient alternative to the brute-force computer

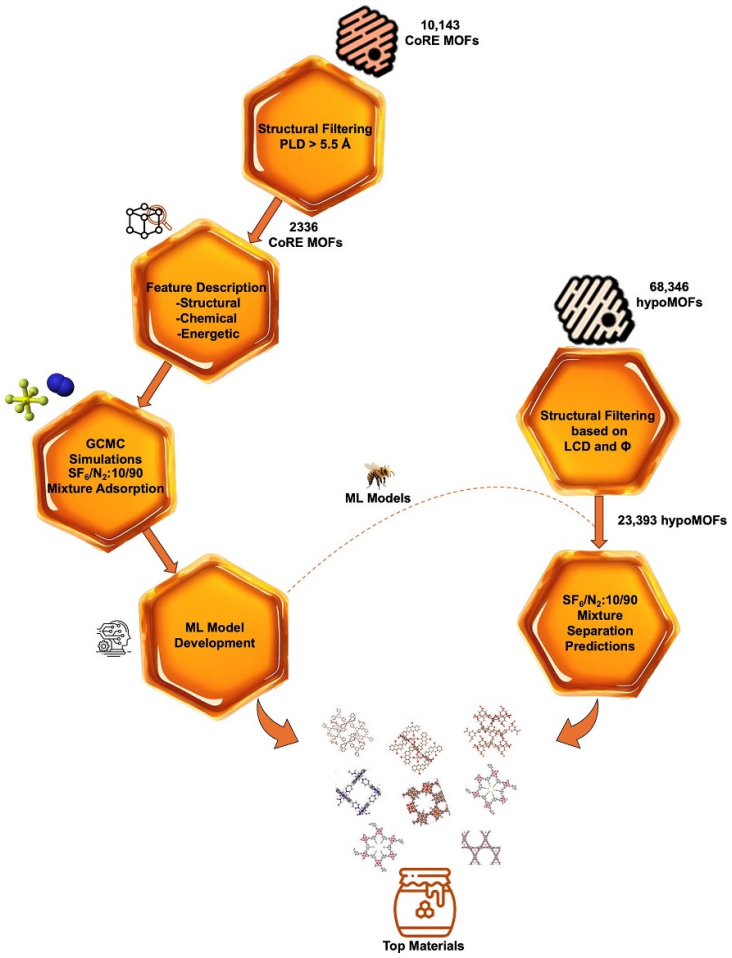


Figure 1. Representation of our computational approach: Structural filtering of CoRE MOFs, descriptor calculations, GCMC simulations to compute SF_6/N_2:10/90 mixture adsorption data, ML model development to predict adsorption and separation properties of CoRE MOFs, transfer of the ML models to assess SF_6/N_2:10/90 mixture adsorption and separation properties of 23 393 hypoMOFs., identification of the best adsorbents among CoRE and hypothetical MOFs for SF_6/N_2:10/90 mixture separation.

simulations for evaluating the SF_6/N_2 mixture adsorption and separation capabilities of large numbers of MOFs and guide the experimental efforts to the most promising materials.

2. Computational methodology

2.1. MOFs

Figure 1 outlines our computational methodology, which integrates molecular simulations and ML to evaluate the gas adsorption and separation capabilities of CoRE MOFs and hypoMOFs. Structural properties of MOFs, such as pore limiting diameter (PLD), the largest cavity diameter (LCD), accessible surface area (ASA), geometric pore volume (PV), and porosity (ϕ) were computed using Zeo++ software (version 0.3) [26]. ASA and PV were calculated using the Monte Carlo sampling procedure with a probe radius of 1.82 Å (representing N_2) and 0 Å, respectively. The most recent CoRE MOF database [24] which includes 10 143 CoRE MOFs and the hypoMOF database constructed by Wilmer *et al* [25] which includes 137 953 hypoMOFs were examined.

We focused on CoRE MOF and hypoMOF databases because they offer the most comprehensive and well-curated computation-ready datasets for experimentally synthesized and hypothetical MOFs, respectively. The CoRE MOF database is the only purely experimental MOF database with CIFs curated for molecular simulation software. The hypoMOF database, constructed by Wilmer *et al* [25], was the first of its kind, featuring high chemical diversity with a wide range of applications in gas storage and separation. A total of 2336 CoRE MOFs and 68 346 hypoMOFs was obtained by selecting the structures having PLDs greater than 5.50 Å, which guaranteed the adsorption of both SF_6 (5.50 Å) and N_2 (3.64 Å) molecules. HypoMOFs were then refined to have the same LCD and porosity range (6.71 Å $<$ LCD $<$ 17.07 Å and $0.55 < \phi < 0.75$) with the top 30 CoRE MOFs that we identified for SF_6/N_2 mixture separation as we will discuss in detail below. As a result, 2336 CoRE MOFs that we studied in this work have the following PLD,

LCD, ASA, PV, and ϕ values of 5.5–71.5 Å, 5.6–71.6 Å, 158.9–8772.1 $m^2\ g^{-1}$, 0.1–17.7 $cm^3\ g^{-1}$, and 0.26–0.97, respectively. The ranges for PLD, LCD, ASA, PV, and ϕ of 23 393 hypoMOFs were computed to be between 5.5–14.6 Å, 6.7–16.9 Å, 615.0–5048.0 $m^2\ g^{-1}$, 0.3–1.5 $cm^3\ g^{-1}$, and 0.55–0.75, respectively.

2.2. Molecular simulations

We used the RASPA [27] simulation code (version 2.0.3) for GCMC simulations to compute the adsorption of SF_6/N_2:10/90 mixture within 2336 CoRE MOFs at 298 K at 0.1 and 1 bar. The treatment of guest/MOF interactions involved the summation of Lennard-Jones (LJ)-type van der Waals (vdW) and Coulombic contributions [28], with the LJ parameters for MOF atoms obtained from the Universal Force Field (UFF) [29]. The long-range Coulombic interactions were computed using the Ewald summation method [30], and all the short-range LJ interactions were calculated with a cutoff distance of 13.0 Å to ensure that the simulation boxes have a minimum of twice the designated cutoff distance in each direction. Pair potentials between separate atoms were obtained by employing the Lorentz–Berthelot combining rules. The charge equilibration (Qeq) method [28, 31] was used to calculate the atomic charges of MOFs.

N_2 was represented as a three-site rigid molecule, with nitrogen atoms occupying two sites and the center of mass with partial charges at the third site [32]. The Dellis and Samios [33] model was used to describe the SF_6 molecule with seven sites. Our simulations utilizing this model for SF_6 adsorption were found to be in good agreement with the experimentally reported uptakes in three MOFs, MIL-47 [34], UiO-66-NH_2 [35], and ZIF-8 [34]: The adsorbed amount of SF_6 was reported as 1.64, 4.07, and 4.46 mol kg^{-1} for UiO-66-NH_2, MIL-47, and ZIF-8 at 1 bar, 298 K, respectively. Our simulation results computed under the same conditions with experiments resulted in uptakes of 1.67, 3.71 mol kg^{-1}, and 4.82 mol kg^{-1} for those three MOFs, respectively. The Widom particle insertion method [36] was used to compute the isosteric heat of adsorption (Q_{st}), the difference in the partial molar enthalpy of adsorbate between the bulk and adsorbed phases, and the Henry's constants of gas molecules at the limit of zero coverage (infinite dilution). We used 10^4 cycles for initialization and a further 5×10^4 cycles for taking the ensemble averages.

Molecular simulation results for the SF_6/N_2 mixture adsorption were used to compute several adsorbent performance evaluation metrics. Table S1 in the Supporting Information presents the formulas for SF_6/N_2 mixture selectivity (S_{mix}), SF_6 working capacity (ΔN_{SF_6}), regenerability (R%), and APS. Some studies examined SF_6/N_2 separation performance of MOFs using TSN [21] and TSQ [22], which are defined as the logarithm of selectivity multiplied by working capacity and the logarithm of selectivity multiplied by the adsorption amount of the desired component, respectively. We focused on selectivity, working capacity, APS, and R%. The mixture selectivity was calculated by taking the ratio of the adsorbed gases to their composition in the bulk mixture. The disparity in gas uptakes at adsorption and desorption pressures was computed as working capacity (ΔN). APS was calculated by multiplying selectivity and working capacity to consider the trade-off between these properties and regenerability (R%) was computed by dividing the SF_6 working capacity by its uptake at the adsorption pressure. A high R% is crucial since adsorption-based separation generally requires cyclic usage of the adsorbent [37, 38]. Therefore, we selected the materials having R%$>$ 80% and ranked them based on their APSs from the highest to the lowest to identify the top-performing MOFs for SF_6/N_2 mixture separation.

2.3. Machine Learning

The input data of the ML models were set as structural, chemical, and energetic descriptors of the CoRE MOFs, as listed in table S2. The structural descriptors were PLD (Å), LCD (Å), ASA ($m^2\ g^{-1}$), ϕ, and PV ($cm^3\ g^{-1}$), and the chemical descriptors were elemental percentages in the extended unit cell structures; C%, H%, N%, O%, Metal%, Halogen%, Ametal%, total degree of unsaturation, O to M ratio. The energetic descriptors were set as the isosteric heats of adsorption (Q_{st}° (kJ mol^{-1})) of SF_6 or N_2 computed at infinite dilution. We specifically used materials' features as the input data to link materials properties to their performances. ΔQ_{st} (kJ mol^{-1}), which is the difference between the heat of adsorption of SF_6 from N_2, was used as a feature only for predicting selectivity and APS since they both include the SF_6 and N_2 uptakes in their formulas. Eight different ML models were developed: (1) SF_6 uptakes at 0.1 bar, (2) SF_6 uptakes at 1 bar, (3) N_2 uptakes at 0.1 bar, (4) N_2 uptakes at 1 bar, (5) selectivity ($\log S_{SF_6/N_2}$), (6) working capacity (ΔN_{SF_6}), (7) logAPS, and (8) regenerability (R%). The Pearson correlation coefficients (r) were calculated, as shown in figure S1, and we eliminated the descriptors with a high correlation ($|r| > 0.90$) to prevent overtraining ML models [39]. The simulated SF_6 and N_2 uptakes for the SF_6/N_2:10/90 mixture at 0.1 and 1 bar were utilized as the target data to train the ML models. SHapley Additive exPlanations (SHAP) analysis [40] was used to assess the impact of features on ML predictions in terms of magnitude and direction. It provides a way to explain the predictions of a model by assigning each feature an importance value (referred to as a SHAP value) for a certain prediction. The major purpose of SHAP is to explain how various features affect model prediction. To better understand the underlying behavior of the model, SHAP breaks down a prediction into contributions from each feature. Positive or negative SHAP values were calculated by subtracting the model's

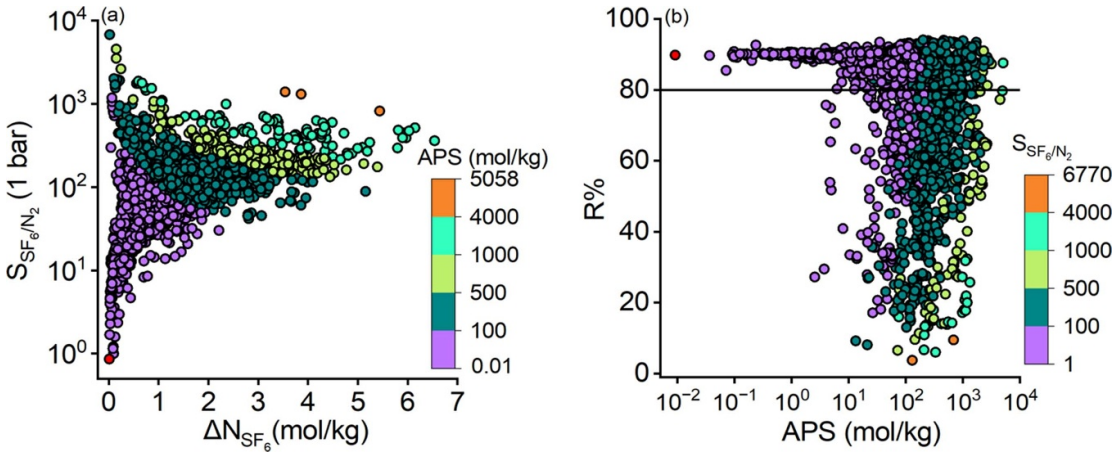


Figure 2. (a) S_{SF_6/N_2} and ΔN_{SF_6}, (b) R% and APS values of 2336 CoRE MOFs calculated from the results of molecular simulations at an adsorption (desorption) pressure of 1 (0.1) bar. The red point represents a MOF with a SF_6/N_2 selectivity of 0.9.

output for simulated SF_6 and N_2 uptakes from the average prediction for SF_6 and N_2 uptakes, as discussed in detail in the next section.

In our study, we used TPOT (Tree-based Pipeline Optimization Tool, version 0.12.2) [41], which is an automated ML tool that optimizes the ML pipelines using genetic programming. It iteratively tests and refines different combinations of ML algorithms to identify the best-performing pipeline for a given dataset. TPOT employs a variety of regression algorithms from scikit-learn toolkit [42], not just Random Forest, but also includes algorithms like Gradient Boosting, Decision Tree, Extra Tree, AdaBoost, Ridge, Lasso, and ElasticNet regressors. This approach ensures a thorough search for the optimal models and hyperparameters. Among these models, Random Forest regressors were offered as the best models by TPOT to predict gas uptakes and separation metrics of MOFs in our study. Therefore, we utilized Random Forest Regressor to make predictions for gas adsorption and separation potential of thousands of hypoMOFs for SF_6/N_2 mixture. A stratified sampling strategy was employed, with 80% of the data serving as a training set and 20% as a test set, to maintain the most uniform feature distribution possible in the training and test sets. In addition, we conducted a five-fold cross-validation to avoid overfitting. The accuracy of ML models was evaluated using the coefficient of determination (R^2), root-mean-square error, and mean absolute error, using the formulas given in table S3. As will be discussed in the upcoming sections, the Random Forest algorithm [43] was selected because of its accuracy in predicting SF_6 and N_2 adsorption in CoRE MOFs. Table S4 lists the optimized hyperparameters of the selected ML models.

To test the transferability of the ML models, a representative hypoMOF set was constructed and utilized as the unseen data [44]. The pore properties of the top 30 CoRE MOFs were identified as $6.71\ \text{Å} < \text{LCD} < 17.07\ \text{Å}$ and $0.55 < \phi < 0.75$, therefore we selected the hypoMOFs that have the same LCD and porosity range. As a result, 23 393 hypoMOFs were chosen and ML models were applied to predict their SF_6 and N_2 adsorption and adsorbent performance evaluation metrics. We also carried out GCMC simulations for a subset of 350 hypoMOFs to verify the accuracy of the ML models for an unseen dataset.

3. Results and discussion

3.1. Separation of SF_6/N_2 mixture using CoRE MOFs

The SF_6/N_2 mixture separation performances of 2336 CoRE MOFs were evaluated based on the results of GCMC simulations. Figure 2 demonstrates the adsorbent performance evaluation metrics of CoRE MOFs computed from the simulated gas uptakes at an adsorption pressure of 1 bar and desorption pressure of 0.1 bar. Figure 2(a) shows that the mixture selectivity (S_{SF_6/N_2}) and working capacity (ΔN_{SF_6}) of CoRE MOFs were computed as 0.9–6769.6 and 0.01–6.54 mol kg^{-1}, respectively. Almost all MOFs (except the MOF colored in red, which has a selectivity of 0.9) are SF_6 selective despite N_2 being mainly present in the mixture (SF_6/N_2:10/90). This is because SF_6 has seven interaction sites with high energetic interaction parameters for the sulfur atom ($\varepsilon/k_B = 165.1$ K) and six fluorine atoms (each with $\varepsilon/k_B = 27.0$ K) compared to N_2 ($\varepsilon/k_B = 36.0$ K). In addition, SF_6 possesses a high hexadecapole moment (35.8 cm^4) [45] compared to the quadrupole moment of N_2 (4.7 Cm^2) [46] which leads to stronger electrostatic interactions with MOFs. There are 100 MOFs exhibiting very high selectivity >1000 with a wide range of working capacities between 0.02–6.15 mol kg^{-1}. We colored the MOFs based on their APSs, the multiplication of selectivity and working capacity, to identify the materials offering high selectivity and working capacity in figure 2(a). MOFs with

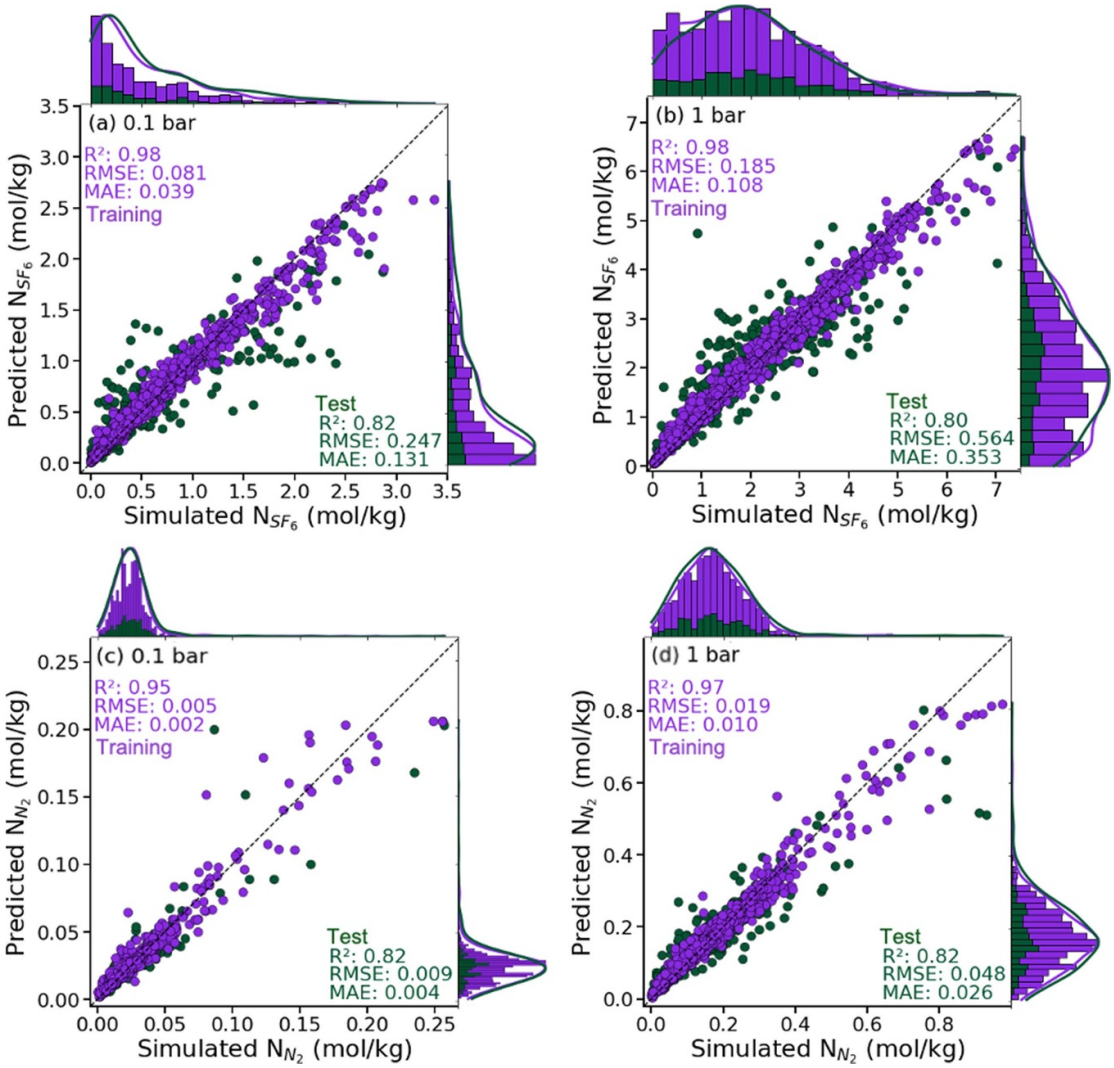


Figure 3. Comparison of ML-predicted and simulated (a), (b) SF_6 and (c), (d) N_2 uptakes of 2336 CoRE MOFs at 0.1 bar and 1 bar.

APS <100 mol kg^{-1} (purple data points) are unpromising because of either low working capacity or low selectivity, whereas the dark green points represent MOFs that perform mediocrely with 100 < APS < 500 mol kg^{-1}. The CoRE MOFs with the highest APS (>4000 mol kg), indicated by orange data points, typically offer a high working capacity together with excellent selectivity.

Figure 2(b) shows the R% and APS of CoRE MOFs at an adsorption pressure of 1 bar and desorption pressure of 0.1 bar. R% values varied from 3.8 to 94.1%, and APSs ranged from 0.01 to 5057.75 mol kg^{-1}. MOFs with high APS generally exhibited low R%. 144 MOFs with APS >500 mol kg^{-1} can still achieve a sufficient R%> 80%. We identified a total of 1363 CoRE MOFs that satisfy the condition of R%> 80%. The top 20 CoRE MOFs having R%> 80% offer high mixture selectivities of 179.7–654.9. These SF_6/N_2 selectivities of the top-performing MOFs that we identified are comparable to or even higher than those of the previously studied porous materials, including MOFs, amorphous activated carbons, zeolite 13X, and porous carbon nanotubes which exhibited SF_6/N_2 selectivities of 138–953 [20], 2.9–44.6 [47], 56.0 [48], and 211 [49] for the separation of SF_6/N_2: 10/90 mixture at 1 bar, 298 K, respectively.

3.2. ML predictions for SF_6/N_2 separation in CoRE MOFs

To provide a highly efficient alternative to the computationally demanding molecular simulations, we generated ML models that rapidly and accurately predict the SF_6/N_2 mixture adsorption data and adsorbent performance evaluation metrics of CoRE MOFs. We used the structural, chemical, and energetic features of the MOFs as the input data and the simulated SF_6 and N_2 uptakes of 2336 CoRE MOFs as the target data to train four distinct ML models to predict SF_6 and N_2 uptakes at 0.1 and 1 bar. ML predictions for mixture gas uptakes were compared with the simulation results to evaluate these models' accuracy, as shown in figure 3, which represents scatter plots with marginal histograms. The distributions for all ML-predicted metrics in

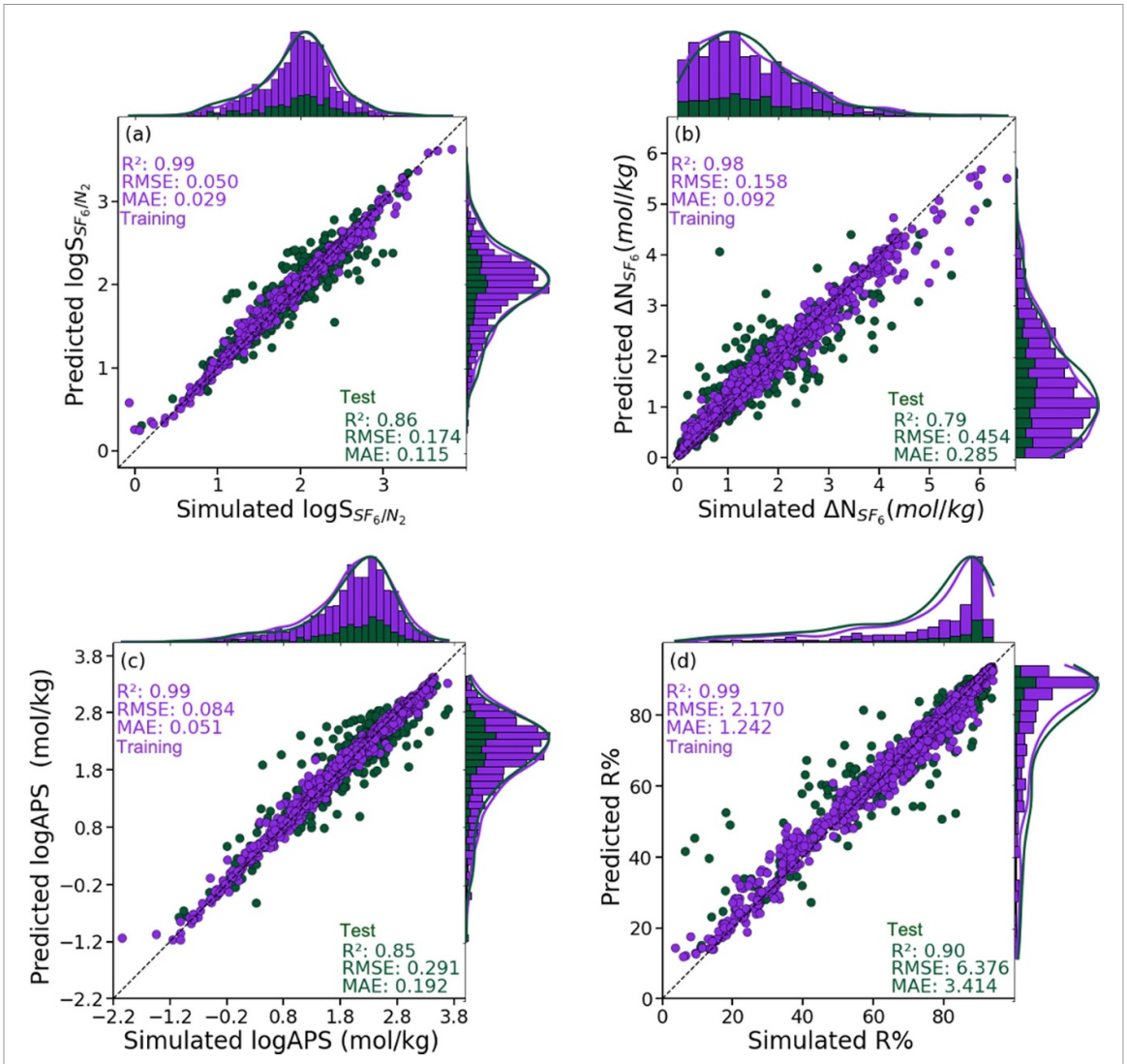


Figure 4. Comparison of ML-predicted and simulated (a) logS_{SF_6/N_2}, (b) ΔN_{SF_6}, (c) logAPS, and (d) R% of 2336 CoRE MOFs.

the training and test sets are similar to simulated ones, demonstrating that our models are capable of capturing and generalizing patterns from the training data to the corresponding test sets. ML models exhibit a fair prediction for SF_6 uptakes at 0.1 bar and N_2 uptakes at 0.1 bar and 1 bar with R^2 values of 0.82 for the test sets, and for SF_6 uptakes at 1 bar with an R^2 value of 0.80 for the test set.

A major advantage of developing ML models is gaining molecular insights into how different material features influence the gas adsorption and separation performance of the adsorbents. Figure S4 shows the feature importance analysis of SF_6 and N_2 mixture uptakes at 0.1 and 1 bar, which provides a molecular understanding of how distinct material properties affect the corresponding gas adsorption potentials of MOFs. All the distributions have Q_{st} as the most dominant feature. From low to high pressure, the effect of Q_{st} decreases while the effect of PV and ASA increases for SF_6 and N_2 adsorption (figures S4(b) and (d)). Gas-MOF interactions dominate at low pressures, which makes Q_{st} more significant. At high pressures, on the other hand, as more gas molecules are filled into the pores of MOFs, PV, and ASA become more important for accurately predicting SF_6 and N_2 uptakes. We also conducted SHAP analysis to examine how various features influence both the magnitude and direction of the model's output, as shown in figure S5. Higher Q_{st} and ASA values and lower PLDs and porosities are generally associated with higher gas uptake predictions at 0.1 and 1 bar.

Since some of the ML-predicted SF_6 uptakes deviated from the simulated uptakes, particularly at 0.1 bar, the adsorbent performance evaluation metrics computed from ML-predicted gas uptakes did not result in a very good agreement with the simulation results, as shown in figure S2. Therefore, we developed additional ML models to directly predict the adsorbent performance evaluation metrics. To train these ML models, we used the metrics computed from the results of GCMC simulations of 2336 CoRE MOFs as the target data and their structural, chemical, and energetic features as the input data. The comparison of ML-predicted and simulated performance metrics is shown in figure 4. Since S_{SF_6/N_2} and APS values were computed to be

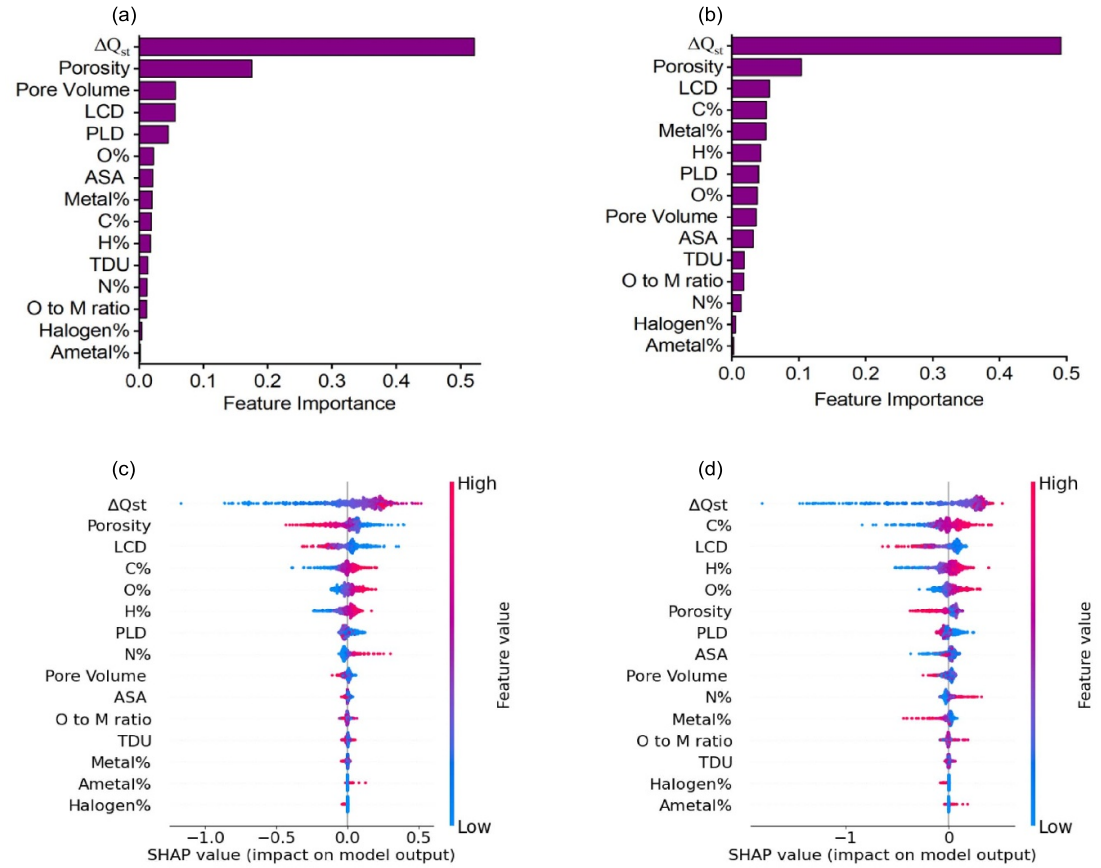


Figure 5. Feature importance analysis for (a) $\log S_{SF_6/N_2}$, (b) logAPS predictions. SHAP feature distributions used in predicting (c) $\log S_{SF_6/N_2}$ and (d) logAPS.

within a wide range (0.9–6769.6 and 0.01–5057.8 mol kg^{-1}, respectively), we directly predicted those metrics in the logarithmic scale. There is a strong agreement between ML predictions and simulations as demonstrated by R^2 values of 0.86, 0.79, 0.85, and 0.90 for $\log S_{SF_6/N_2}$, ΔN_{SF_6}, logAPS, and R%, respectively, for the test sets. ML predictions for 2336 CoRE MOFs have a similar range of ΔN_{SF_6} (0.04–5.68 mol kg^{-1}), and R% (11.8–93.5%) values compared to simulation results (0.01–6.54 mol kg^{-1} and 3.8–94.1%). Figure S3 compares ML-predicted and simulated performance metrics for selectivity vs. working capacity and APS vs R% for all CoRE MOFs. ML-predicted selectivities and working capacities (1.8–4204.2 and 0.04–5.68 mol kg^{-1}, respectively) were in good agreement with the simulation results (0.9–6769.6 and 0.01–6.54 mol kg^{-1}). Similarly, ML-predicted APS and R% values (0.07–2606.7 mol kg $^{-1}$ and 11.8–93.5%) agreed with the simulated ones (0.01–5057.8 mol kg^{-1} and 3.8–94.1% respectively).

Figure 5 demonstrates the feature importance analysis of $\log S_{SF_6/N_2}$ and logAPS predictions for SF_6/N_2:10/90 mixture. For both metrics, ΔQ_{st} is the most dominant feature (figures 5(a) and (b)), which is the difference between the heat of adsorption values of SF_6 from N_2. Structural features, including porosity, PV, LCD are also crucial for predicting these metrics. Feature importance as determined by the SHAP analysis demonstrated that high ΔQ_{st} and C% values positively affect the predictions, whereas high porosities and LCDs resulted in lower selectivities and APSs (figures 5(c) and (d)). As SF_6 has stronger energetic interactions with the MOF atoms than N_2, a higher C% in the frameworks increases the vdW interactions between the MOFs and SF_6 molecules, resulting in selective adsorption of SF_6 over N_2. For working capacity, the feature importances are generally evenly distributed, while the most important feature is Q_{st} of SF_6, as shown in Figure S6a. For R%, the effect of Q_{st} is more pronounced compared to LCD, porosity, and surface area. High Q_{st} values of SF_6 resulted in high working capacities but low R%. The strong interaction between SF_6 and the adsorbent is responsible for the high SF_6 uptake at the adsorption pressure, however, this interaction results in a low number of molecules being desorbed at the desorption pressure. Therefore, it is important to consider MOFs that can simultaneously provide a sufficient R% for efficient SF_6/N_2 separation.

3.3. Evaluating SF_6/N_2 separation performance of HypoMOFs by ML

Performing HTCS for evaluating SF_6/N_2 mixture adsorption and separation performances of thousands of MOFs has a very high computational cost. Therefore, ML models that we developed by training 2336 CoRE MOFs were used for predicting SF_6/N_2 mixture adsorption data and separation performance metrics of unseen 23 393 hypoMOFs. For testing the accuracy of the unseen predictions, GCMC simulations of a representative subset of 350 hypoMOFs were also performed. This subset was selected from the unseen MOFs, considering the most common metals (Zn, Cu) of the top CoRE MOFs as we discuss in more detail below. While the transferability of ML models to unseen data, such as hypoMOFs, is a promising approach, there are potential challenges that should be addressed. If the hypoMOFs have different structural or chemical characteristics that were not well-represented in the training dataset, the model may not generalize well, and this may lead to inaccurate predictions (overestimations or underestimations) when applied to unseen MOFs. In this case, expanding and diversifying training datasets may improve model generalization. Therefore, it is crucial to ensure that the training dataset is balanced and representative. Figure S7 displays simulation results and ML predictions for SF_6 and N_2 uptakes of 350 hypoMOFs at 0.1 and 1 bar. R^2 values of 0.71 and 0.82 for test sets show a relatively good prediction power for SF_6 adsorption at 0.1 and 1 bar. ML predictions for N_2 uptake at 1 bar display a scattered distribution since the ML models were transferred from the CoRE MOF uptake data, which showed a similar trend. We inferred that calculating the adsorbent performance evaluation metrics from the ML-predicted gas uptakes may lead to the expansion of the error margins from the uptake predictions. Therefore, we directly predicted the adsorbent performance evaluation metrics for the hypoMOFs. Figure S8 shows the comparison of metrics computed from molecular simulations and those predicted by ML models. ML predictions agreed well with the simulation results, and R^2 values of 0.75, 0.75, 0.79, and 0.81 were obtained for $\log S_{SF_6/N_2}$, ΔN_{SF_6}, logAPS and R%, respectively.

ML-predicted performance metrics for the whole set of material space covering 2336 CoRE MOFs and 23 393 hypoMOFs that we focused on in this work are represented in figure 6. Figure 6(a) shows that many MOFs with high selectivities >100 also have high working capacities >3 mol kg^{-1}, but there are also exceptions where MOFs with very high selectivities offer very low working capacities. For the whole material space, ML-predicted selectivities and working capacities change in between 1.8–4204.2 and 0.04–5.68 mol kg^{-1}, respectively. Figure 6(b) demonstrates that ML predictions for APS and R% vary from 0.07 to 2606.74 mol kg^{-1} and 11.8–93.5%, respectively. There is an inverse relation between APS and R%: MOFs with high APS generally possess R% lower than 80%, making them unpromising for practical applications. 950 MOFs out of 25 729 materials that we studied exhibit both high APS >500 mol kg^{-1} and high R%> 80%. The red region in figure 6(b) shows the top-performing adsorbent materials determined by the criteria of R%> 80% and APS >1000 mol kg^{-1}. Accordingly, there are 86 MOFs in this region and 20 of them are CoRE MOFs. LCDs and porosities of these top 20 CoRE MOFs vary from 6.7 to 11.6 Å and 0.59–0.73, respectively. For 66 hypoMOFs which are in the red region, similar LCDs and porosities were observed in the range of 6.9–9.2 Å and 0.65–0.72, respectively.

Figure 6 shows that SF_6/N_2 separation performances of the hypoMOFs could not surpass those of CoRE MOFs and this could be attributed to the nature of the ML models which have limitations in extrapolating beyond the trained dataset [50] and we trained the ML models using the CoRE MOFs' data. Top performing hypoMOFs and CoRE MOFs were accurately identified by our selected ML models. HypoMOF namely hMOF-5035 596-(id_95175) was the 36th MOF in the top performing 50 MOF list, and its predicted APS was 972.5 mol kg^{-1}, while the calculated APS from molecular simulations was 961.9 mol kg^{-1}, which shows that ML accurately predicted the performance of that hypoMOF, and it could be identified in the top material list. Similarly, hMOF-36 068-(id_50970), hMOF-5054 377-(id_113413) and hMOF-5060 769-(id_119569) were among the top performing 50 MOF list. The ML-predicted APSs of these hypoMOFs were 936.1, 958.1, and 963.8 mol kg^{-1}, in a very good agreement with the values obtained from molecular simulations, 1017.7, 1042.7 and 1082.5 mol kg^{-1}, respectively. One of the CoRE MOF, namely ATAYAX in the top 5 MOF has APS of 2650 mol kg^{-1} calculated from molecular simulations and 2593 mol kg^{-1} from ML predictions, which is almost twice of the top performing hypoMOF performance. Therefore, the reason that hypoMOFs could not outperform CoRE MOFs is not only due to the limitation of the ML model, but also because of the nature of hypoMOFs. For example, many hypoMOFs showing lower separation performances than CoRE MOFs have generally larger pore sizes, and materials with narrow pore sizes achieve higher selectivity because of the strong interaction between the adsorbate molecules and the pore walls.

Therefore, we specifically focused on the CoRE MOFs and examined the chemical properties of the top 20 CoRE MOFs that appeared in the red region in figure 6(b). The heatmap illustrating the linker subunit and metals of these top 20 CoRE MOFs is given in figure S9. The most common metals are Zn, Cu, Ni, and Co, and the most common linker subunits are terphenyl, pyridine, and naphthalene. These findings agree with the experimental reports in the literature: For example, a Ni-based MOF having multiple naphthalene rings was experimentally studied for SF_6/N_2:10/90 mixture separation at 1 bar and its selectivity calculated by

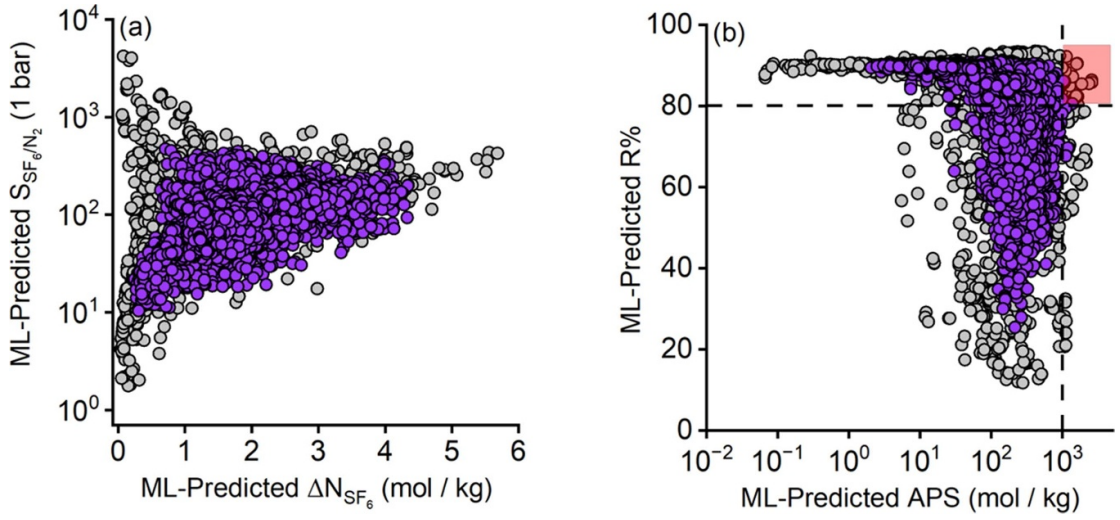


Figure 6. Analysis of 2336 CoRE MOFs (grey) and 23 393 hypoMOFs (purple) for SF_6/N_2:10/90 mixture separation: (a) S_{SF_6/N_2} and ΔN_{SF_6}, (b) R% and APS computed at an adsorption pressure of 1 bar and desorption pressure of 0.1 bar. The red region represents the top adsorbent zone.

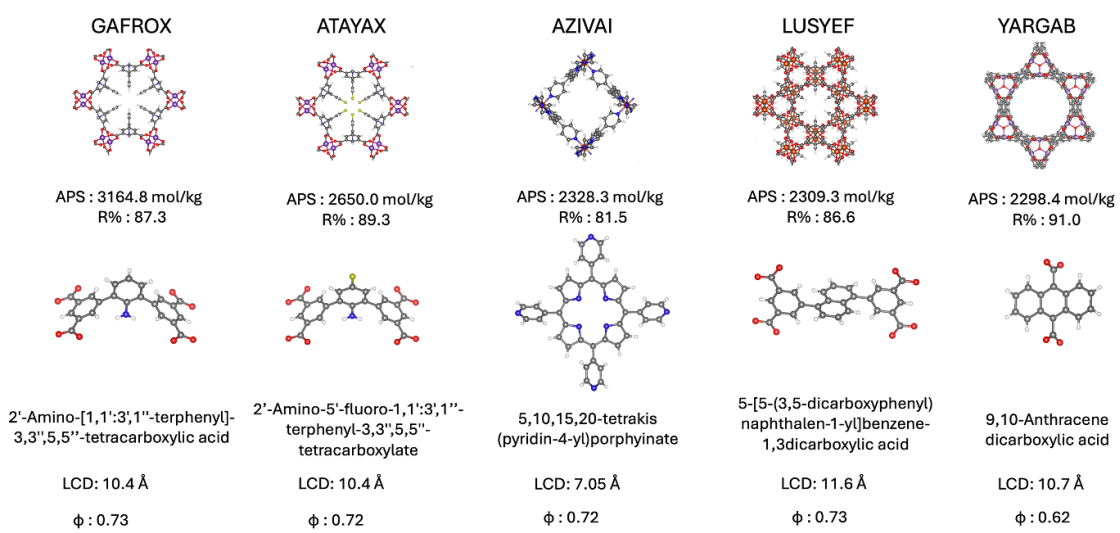


Figure 7. Structures of the top five CoRE MOFs for SF_6/N_2:10/90 mixture separation. Gray, white, red, purple, blue, yellow, green, and orange spheres represent C, H, O, Zn, N, F, Cl, and Cu atoms, respectively. Linker names and structural properties of MOFs are provided.

the Ideal Adsorbed Solution Theory (IAST) was reported to be high, 750 [51]. Similarly, a MOF consisting of Ni metal and pyridine linker exhibited a relatively SF_6/N_2:10/90 mixture separation performance with an IAST-predicted selectivity of 200.6 at 1 bar [52].

We finally refined the list of the top MOFs down to five, as shown in figure 7, with their detailed structural analysis showing their metal and linker types. Their refcodes are GAFROX, ATAYAX, AZIVAI, LUSYEF, and YARGAB, and these MOFs were accurately identified as the top five adsorbents both by molecular simulations and ML models. Based on the simulation results, those MOFs have high SF_6/N_2 separation performance with APSs ranging from 2298.4–3164.8 mol kg^{-1} and R% values ranging from 81.5 to 91.0%. ML predictions have very similar ranges for APS (1628.2–2593.0 mol kg^{-1}) and R% (80.3–90.5%). Similar to the top 20 CoRE MOFs that we discussed above, the top five MOFs have terphenyl, pyridine, naphthalene, and anthracene linkers. MOFs with terphenyl [19] and anthracene [53] linkers have been reported to exhibit effective SF_6 adsorption in the literature. MOFs including Zn metal (ATAYAX, GAFROX, and YARGAB) have high selectivities of 449, 515, and 531 at 1 bar, respectively. These selectivities are much higher than those reported in the experimental studies. For instance, two Zn-based MOFs, $Zn(BDC)(DABCO)_{0.5}$ and $Zn(DMBDC)(DABCO)_{0.5}$, were reported to have IAST-based selectivities of 31.8 and 118.9, respectively, for SF_6/N_2:10/90 mixture separation at 1 bar, 298 K [54]. The top-performing five CoRE MOFs also have

comparable selectivities with a carbon nanosheet, namely PC-750 having a high IAST-based selectivity of 436.4 for SF_6/N_2:10/90 mixture at 1 bar [55]. Thus, our results have the potential to direct the experimental efforts towards the promising MOFs that we identified for high-performance SF_6/N_2 separation. As we previously observed for CO_2/N_2 separation [12], high selectivities are generally obtained with MOFs having low to moderate pore sizes for SF_6/N_2 separation because of the strong confinement of SF_6 molecules in the pores. LCD and porosities of the top five MOFs are in between 7.05–11.6 Å and 0.62–0.73, respectively. The largest surface area is 3753.6 $m^2\ g^{-1}$ and the lowest surface area is 1359.2 $m^2\ g^{-1}$ among the top five MOFs, which can be interpreted as moderate compared the whole set of 2336 CoRE MOFs that we studied.

4. Conclusion

In this work, we presented a methodology that integrates molecular simulations with ML to develop models that accurately predict SF_6/N_2 mixture adsorption data for over 25 000 different types of synthesized and hypoMOFs. Computationally demanding molecular simulations for SF_6/N_2 mixture adsorption were performed only for 2336 synthesized MOFs and then this data was employed to train ML models which accurately and quickly predicted SF_6 and N_2 mixture adsorption in these MOFs. ML analysis showed that heat of adsorption of gases is the most dominant feature to predict SF_6 and N_2 adsorption in MOFs at 0.1 bar, whereas at 1 bar, surface area and pore volume become important for accurate assessment of SF_6 and N_2 adsorption in MOFs. We also developed ML models that evaluate MOF adsorbents' performance evaluation metrics such as selectivity, working capacity, APS, and regenerability for SF_6/N_2 mixture separation. ML analysis showed that the difference between heat of adsorption values of gases is the most dominant feature in predicting the SF_6 selectivity of MOFs. All these ML models were then transferred to thousands of hypoMOFs to assess their SF_6/N_2 mixture adsorption and separation performances to eliminate the need for performing computationally intensive simulations for a very large number of structures. Results showed that the whole MOF material space that we focused on has very high SF_6 selectivities up to 4204 at 1 bar and working capacities up to 5.68 mol kg^{-1} at an adsorption (desorption) pressure of 1 bar (0.1 bar). Finally, we examined the chemical and structural properties of the top-performing MOF adsorbents and found out that the best MOF candidates generally have Zn, Cu, Ni metals; terphenyl, pyridine, and naphthalene linkers; moderate pore sizes and surface areas. Our approach will be useful for predicting the SF_6/N_2 adsorption and separation performances of MOFs in an accurate and quick way as a strong alternative to the time and resource-intensive computer simulations and will guide the experimental efforts to high-performance materials for SF_6/N_2 separation. While this study demonstrates the effectiveness of integrating molecular simulations with ML models to predict SF_6/N_2 adsorption and separation performances in MOFs, there are several avenues for future research. One potential direction is to explore the application of this methodology to other gas mixtures, thereby expanding the utility of the developed models. Examining the transferability of our approach to different porous materials may also lead to the discovery of new materials for high-performance SF_6/N_2 separation.

Data availability statement

All data that support the findings of this study are included within the article (and any supplementary files) and the following GitHub address: https://github.com/pelinsezgin/MOFs_SF6N2_ML.

Acknowledgment

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.

Conflict of interest

The authors declare no competing interests.

ORCID iDs

Pelin Sezgin https://orcid.org/0009-0002-1204-7159
Hasan Can Gulbalkan https://orcid.org/0000-0001-8001-6018
Seda Keskin https://orcid.org/0000-0001-5968-0336

References

[1] Rabie M and Franck C M 2018 Assessment of eco-friendly gases for electrical insulation to replace the most potent industrial greenhouse gas SF6 *Environ. Sci. Technol.* **52** 369–80
[2] Zhang B, Xiong J, Chen L, Li X and Murphy A B 2020 Fundamental physicochemical properties of SF_6-alternative gases: a review of recent progress *J. Phys. D: Appl. Phys.* **53** 173001
[3] Zhao G, Kim H, Yang C and Chung Y G 2024 Leveraging machine learning to predict the atmospheric lifetime and the global warming potential of SF_6 replacement gases *J. Phys. Chem.* A **128** 2399–408
[4] Chuah C Y, Lee Y and Bae T-H 2021 Potential of adsorbents and membranes for SF_6 capture and recovery: a review *Chem. Eng. J.* **404** 126577
[5] Cai G, Yan P, Zhang L, Zhou H-C and Jiang H-L 2021 Metal–organic framework-based hierarchically porous materials: synthesis and applications *Chem. Rev.* **121** 12278–326
[6] Khalil I E, Fonseca J, Reithofer M R, Eder T and Chin J M 2023 Tackling orientation of metal-organic frameworks (MOFs): the quest to enhance MOF performance *Coord. Chem. Rev.* **481** 215043
[7] Bruno I J, Cole J C, Edgington P R, Kessler M, Macrae C F, McCabe P, Pearson J and Taylor R 2002 New software for searching the Cambridge Structural Database and visualizing crystal structures *Acta Crystallogr.* B **58** 389–97
[8] Daglar H and Keskin S 2020 Recent advances, opportunities, and challenges in high-throughput computational screening of MOFs for gas separations *Coord. Chem. Rev.* **422** 213470
[9] Boyd P G, Lee Y and Smit B 2017 Computational development of the nanoporous materials genome *Nat. Rev. Mater.* **2** 1–15
[10] Ren E, Guilbaud P and Coudert F-X 2022 High-throughput computational screening of nanoporous materials in targeted applications *Digit. Discovery* **1** 355–74
[11] Qiao Z, Zhang K and Jiang J 2016 In silico screening of 4764 computation-ready, experimental metal–organic frameworks for CO_2 separation *J. Mater. Chem.* A **4** 2105–14
[12] Altintas C, Avci G, Daglar H, Nemati Vesali Azar A, Velioglu S, Erucar I and Keskin S 2018 Database for CO_2 separation performances of MOFs based on computational materials screening *ACS Appl. Mater. Interfaces* **10** 17257–68
[13] Rogacka J, Seremak A, Luna-Triguero A, Formalik F, Matito-Martos I, Firlej L, Calero S and Kuchta B 2021 High-throughput screening of metal–organic frameworks for CO_2 and CH_4 separation in the presence of water *Chem. Eng. J.* **403** 126392
[14] Altintas C and Keskin S 2018 Molecular simulations of MOF membranes and performance predictions of MOF/polymer mixed matrix membranes for CO_2/CH_4 separations *ACS Sustain. Chem. Eng.* **7** 2739–50
[15] Altintas C, Erucar I and Keskin S 2018 High-throughput computational screening of the metal organic framework database for CH_4/H_2 separations *ACS Appl. Mater. Interfaces* **10** 3668–79
[16] Altintas C, Avci G, Daglar H, Azar A N V, Erucar I, Velioglu S and Keskin S 2019 An extensive comparative analysis of two MOF databases: high-throughput screening of computation-ready MOFs for CH_4 and H_2 adsorption *J. Mater. Chem.* A **7** 9593–608
[17] Yan Y, Shi Z, Li H, Li L, Yang X, Li S, Liang H and Qiao Z 2022 Machine learning and in-silico screening of metal–organic frameworks for O_2/N_2 dynamic adsorption and separation *Chem. Eng. J.* **427** 131604
[18] Orhan I B, Daglar H, Keskin S, Le T C and Babarao R 2021 Prediction of O_2/N_2 selectivity in metal–organic frameworks via high-throughput computational screening and machine learning *ACS Appl. Mater. Interfaces* **14** 736–49
[19] Ren J, Chang M, Zeng W, Xia Y, Liu D, Maurin G and Yang Q 2021 Computer-aided discovery of MOFs with calixarene-analogous microenvironment for exceptional SF_6 capture *Chem. Mater.* **33** 5108–14
[20] Cha J, Ga S, Lee S-J, Nam S, Bae Y-S and Chung Y G 2021 Integrated material and process evaluation of metal–organic frameworks database for energy-efficient SF_6/N_2 separation *Chem. Eng. J.* **426** 131787
[21] He Y, Cao X, Zhang Z, Jiang Z, Huang H, Zhang S, Han Q and Zhong C 2023 Discovery of high-performing metal–organic frameworks for efficient SF_6/N_2 separation: a combined computational screening, machine learning, and experimental study *Ind. Eng. Chem. Res.* **62** 7642–9
[22] Peng X and Wang H 2024 Machine learning based screening of organic frameworks for separation of CF_4/N_2, C_2F_6/N_2, and SF_6/N_2 *Chem. Eng. Sci.* **296** 120280
[23] Junjie N, Kun S, Rui Z, Kunqi G, Jingyu C and Linxi H 2024 Computer-aided design of covalent organic frameworks for SF_6 capture: the combination of high-throughput screening and machine learning *J. Phys. Chem.* C **128** 11355–66
[24] Chung Y G *et al* 2019 Advances, updates, and analytics for the computation-ready, experimental metal–organic framework database: CoRE MOF 2019 *J. Chem. Eng. Data* **64** 5985–98
[25] Wilmer C E, Leaf M, Lee C Y, Farha O K, Hauser B G, Hupp J T and Snurr R Q 2012 Large-scale screening of hypothetical metal–organic frameworks *Nat. Chem.* **4** 83–89
[26] 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** 134–41
[27] 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
[28] Wilmer C E and Snurr R Q 2011 Towards rapid computational screening of metal-organic frameworks for carbon dioxide capture: calculation of framework charges via charge equilibration *Chem. Eng. J.* **171** 775–81
[29] Rappé A K, Casewit C J, Colwell K, Goddard I I I 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** 10024–35
[30] Ewald P P 1921 Die Berechnung optischer und elektrostatischer Gitterpotentiale *Ann. Phys.* **369** 253–87
[31] Rappe A K and Goddard W A I I I 1991 Charge equilibration for molecular dynamics simulations *J. Phys. Chem.* **95** 3358–63
[32] Makrodimitris K, Papadopoulos G K and Theodorou D N 2001 Prediction of permeation properties of CO_2 and N_2 through silicalite via molecular simulations *J. Phys. Chem.* B **105** 777–88
[33] Dellis D and Samios J 2010 Molecular force field investigation for Sulfur Hexafluoride: a computer simulation study *Fluid Phase Equilib.* **291** 81–89
[34] Zheng X, Shen Y, Wang S, Huang K and Cao D 2021 Selective adsorption of SF_6 in covalent-and metal–organic frameworks *Chin. J. Chem. Eng.* **39** 88–95
[35] Kim M-B, Kim K-M, Kim T-H, Yoon T-U, Kim E-J, Kim J-H and Bae Y-S 2018 Highly selective adsorption of SF_6 over N_2 in a bromine-functionalized zirconium-based metal-organic framework *Chem. Eng. J.* **339** 223–9
[36] Frenkel D and Smit B 2023 *Understanding Molecular Simulation: From Algorithms to Applications* (Elsevier)
[37] Bae Y S and Snurr R Q 2011 Development and evaluation of porous materials for carbon dioxide separation and capture *Angew. Chem. Int. Ed.* **50** 11586–96

[38] Gulbalkan H C, Haslak Z P, Altintas C, Uzun A and Keskin S 2022 Assessing CH_4/N_2 separation potential of MOFs, COFs, IL/MOF, MOF/polymer, and COF/polymer composites *Chem. Eng. J.* **428** 131239
[39] Yang P, Zhang H, Lai X, Wang K, Yang Q and Yu D 2021 Accelerating the selection of covalent organic frameworks with automated machine learning *ACS Omega* **6** 17149–61
[40] Lundberg S M, Erion G, Chen H, DeGrave A, Prutkin J M, Nair B, Katz R, Himmelfarb J, Bansal N and Lee S-I 2020 From local explanations to global understanding with explainable AI for trees *Nat. Mach. Intell.* **2** 56–67
[41] Le T T, Fu W and Moore J H 2020 Scaling tree-based automated machine learning to biomedical big data with a feature set selector *Bioinformatics* **36** 250–6
[42] Martinsson P-G, Rokhlin V and Tygert M 2011 A randomized algorithm for the decomposition of matrices *Appl. Comput. Harmon. Anal.* **30** 47–68
[43] Tang H, Xu Q, Wang M and Jiang J 2021 Rapid screening of metal–organic frameworks for propane/propylene separation by synergizing molecular simulation and machine learning *ACS Appl. Mater. Interfaces* **13** 53454–67
[44] Gulbalkan H C, Uzun A and Keskin S 2023 Evaluating CH4/N2 separation performances of hundreds of thousands of real and hypothetical MOFs by harnessing molecular modeling and machine learning *ACS Appl. Mater. Interfaces* (https://doi.org/10.1021/acsami.3c13533)
[45] Hourri A, St-Arnaud J and Bose T 1997 Dielectric and pressure virial coefficients of imperfect gases: CO_2–SF_6 mixtures *J. Chem. Phys.* **106** 1780–5
[46] Bae Y-S and Lee C-H 2005 Sorption kinetics of eight gases on a carbon molecular sieve at elevated pressure *Carbon* **43** 95–107
[47] Peng X, Vicent-Luna J M and Jin Q 2020 Separation of CF_4/N_2, C_2F_6/N_2, and SF_6/N_2 mixtures in amorphous activated carbons using molecular simulations *ACS Appl. Mater. Interfaces* **12** 20044–55
[48] Kim P-J, You Y-W, Park H, Chang J-S, Bae Y-S, Lee C-H and Suh J-K 2015 Separation of SF_6 from SF_6/N_2 mixture using metal–organic framework MIL-100 (Fe) granule *Chem. Eng. J.* **262** 683–90
[49] Skarmoutsos I, Koukaras E N, Galiotis C, Froudakis G E and Klontzas E 2020 Porous carbon nanotube networks and pillared graphene materials exhibiting high SF_6 adsorption uptake and separation selectivity of SF_6/N_2 fluid mixtures: a comparative molecular simulation study *Microporous Mesoporous Mater.* **307** 110464
[50] Zhang H, Nettleton D and Zhu Z 2019 Regression-enhanced random forests (arXiv:1904.10416)
[51] Yang M, Chang M, Yan T and Liu D 2022 A nickel-based metal-organic framework for efficient SF6/N2 separation with record SF_6 uptake and SF_6/N_2 selectivity *Sep. Purif. Technol.* **295** 121340
[52] Wang S M, Mu X T, Liu H R, Zheng S T and Yang Q Y 2022 Pore-structure control in metal–organic frameworks (MOFs) for capture of the greenhouse gas SF_6 with record separation *Angew. Chem., Int. Ed.* **134** e202207066
[53] Chang M, Yan T, Wei Y, Wang J-X, Liu D and Chen J-F 2022 Metal–organic framework-based single-molecule SF_6 trap for record SF_6 capture *Chem. Mater.* **34** 9134–43
[54] Yan L, Zheng H-T, Song L, Wei Z-W, Jiang J-J and Su C-Y 2023 Methyl-functionalized flexible ultra-microporous MOF for efficient SF_6/N_2 mixture separation *Chem. Eng. J.* **472** 145145
[55] Fu W, Wang J, Li Y, Sui Z, Xiao B and Xu X 2024 Highly-efficient separation of SF_6/N_2 and NF_3/N_2 with record selectivity on one-step synthesized carbon nanosheet *Sep. Purif. Technol.* **330** 125496