

Computer-Aided Design of Covalent Organic Frameworks for SF₆ Capture: The Combination of High-Throughput Screening and Machine Learning

Junjie Ning, Kun Shen, Rui Zhao,* Kunqi Gao, Jingyu Cai,* and Linxi Hou*



Cite This: *J. Phys. Chem. C* 2024, 128, 11355–11366



Read Online

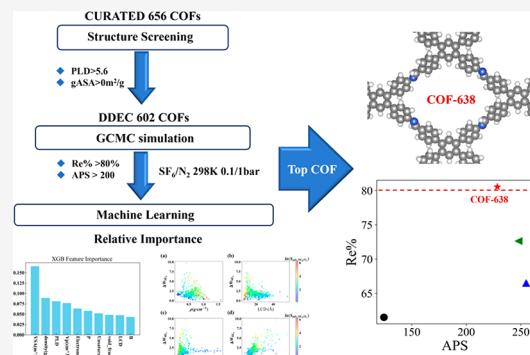
ACCESS |

Metrics & More

Article Recommendations

Supporting Information

ABSTRACT: Efficiently separating sulfur hexafluoride/nitrogen (SF₆/N₂) poses an urgent challenge. Four covalent organic frameworks (COFs) (Re % > 80%) with greater performance in SF₆/N₂ separation experiment were selected from the CURATED database by high-throughput screening in this paper. XGB was selected among four machine learning models (SVM, RF, GBRT, and XGB) and this model had good fitting effects in terms of both regeneration (Re %, $R^2 = 0.809$) and ln(S_{ads}). Relative importance analyses of XGB described that porosity and infinite dilution heat are the most key features for Re % and ln(S_{ads}). The binding energy, charge density difference, Bader charge, and independent gradient model based on Hirshfeld partition (IGMH) analysis were all calculated to investigate the adsorption mechanisms. GCMC simulations combined with density functional theory calculations revealed that COF-638 exhibited an excellent SF₆/N₂ separation performance. The probability distribution diagram of the center of mass illustrates the adsorption sites of SF₆ in coadsorption.



1. INTRODUCTION

Sulfur hexafluoride (SF₆) finds widespread utility in various industrial applications owing to its exceptional arc extinguishing and insulating properties and its notable chemical stability. For instance, within the semiconductor industry, high-purity SF₆ enjoys extensive utilization in the fabrication of computer chips and liquid crystal displays.¹ In the electrical equipment sector, SF₆ is primarily employed in gas-insulated switchgear (GIS).² To reduce manufacturing costs, SF₆/N₂ gas mixtures are commonly utilized instead of pure SF₆ in high-voltage electrical equipment, such as GIS.^{3,4} In contemporary industrial applications, SF₆ typically exists in combinations or compositions. Specific reports assert that SF₆ discharged from electrical equipment contributes to 70% of global SF₆ emissions.^{5–8} SF₆ has a substantial global warming potential (GWP) and has been acknowledged as a significant greenhouse gas in the Paris Agreement (~23,900 times greater than carbon dioxide). The release of SF₆ into the atmosphere poses a substantial challenge with respect to resource depletion and environmental contamination.^{8,9} Therefore, it is of great significance to consider how to effectively recover SF₆ from SF₆/N₂ gas mixtures.

Compared to the conventional methods of SF₆ purification, namely, deep-cooling distillation and liquefaction, vacuum swing adsorption using porous materials exhibits distinct advantages, including reduced energy expenditure, straightforward operation, and enhanced separation efficiency.^{7,10} As a porous material, the structure of the covalent organic

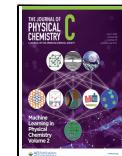
frameworks (COFs) has many advantages, including low density, a substantial surface area, customizable pore size and structure, facile functionalization, and robust stability. These attributes are crucial for their potential application as industrial adsorbents.¹¹ Research on the utilization of COFs in this context has been relatively limited with a predominant emphasis on experimental investigations. Before experimental investigations, molecular modeling is pivotal in assessing their gas adsorption and separation properties.^{12–15} Likewise, Hasell et al. synthesized flexible molecular organic cages that exhibited impressive SF₆/N₂ selectivity (76.5) but had a relatively modest working capacity of 1.49 mmol/g.⁹ Until now, Liao et al. have made notable advancements in improving the efficacy of SF₆ adsorbents by employing Suzuki–Miyaura cross-coupling techniques to modify the functional groups of BrCOF-1.¹³ At a subsequent point, Liao et al. synthesized RCOF-1 by adjusting the ligand, resulting in remarkable SF₆/N₂ selectivity (83), although the material displayed a relatively low working capacity of 1.96 mmol/g.¹⁴ In past experimental studies, shortcomings were identified, including extended research cycles and limited sample sizes, which typically

Received: March 20, 2024

Revised: June 21, 2024

Accepted: June 21, 2024

Published: July 2, 2024



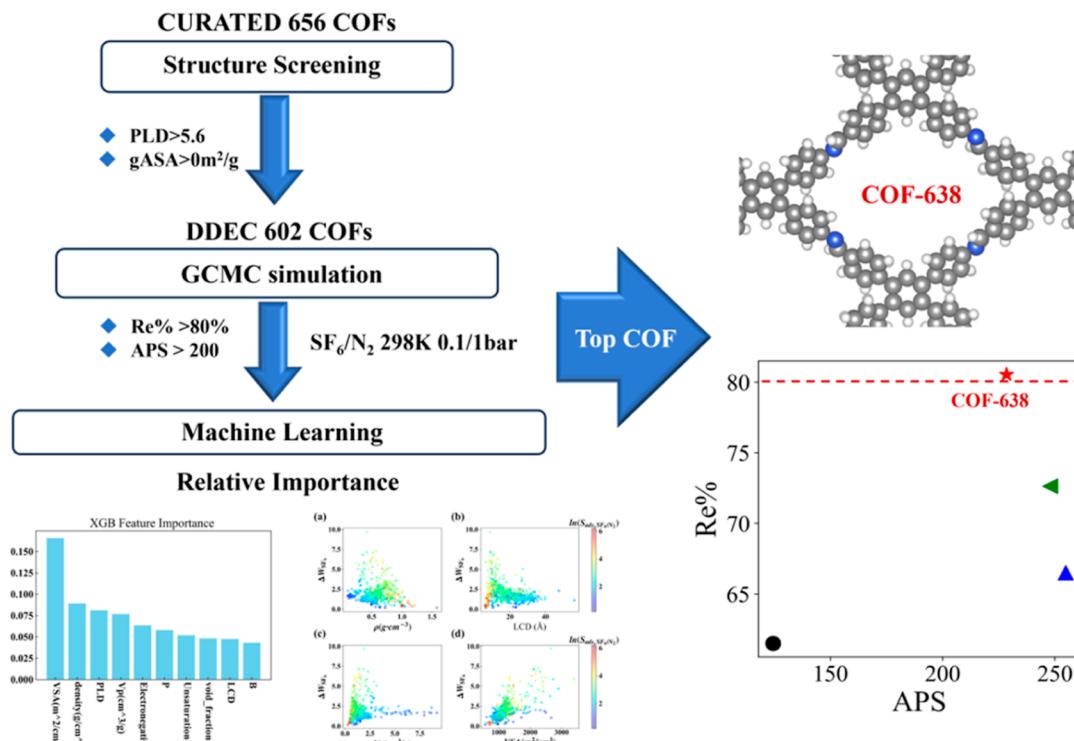


Figure 1. Schematic representation of the numerical screening protocol.

necessitate higher costs to obtain adsorbents with ideal performance characteristics. High-throughput computational screening (HTCS) is an effective computational simulation method that can be used to assess the gas separation potential of thousands of materials. It is integrated with experiments to investigate optimal materials.¹⁶ This approach offers the advantages of accelerating the experimental cycle and reducing the scale of the experiments. The combination of machine learning (ML) and molecular simulations by Cao et al. enabled the pre-experimental prediction of reverse separation of C₂H₆/C₂H₄ for COFs.¹⁷ Aksu et al. employed computational screening methods to identify promising hypoCOF adsorbents for the capture of CO₂ and H₂ purification. The selectivity and working capacity of these adsorbents surpassed those experimentally synthesized COFs under VSA conditions.¹⁸ The examples adequately demonstrate the significance of high-throughput screening studies, as they can expedite progress in experimental research and provide guidance for experimental work.

In view of a vast database, ML research has gradually been combined with high-throughput screening. For instance, Sun et al. utilized a materials genomics approach to predict key factors for the assembly of adsorbents for isobutene/isobutane by screening the CoREMOF database. They successfully obtained Cu₂O₈-BTC_B-core-4_No1, which demonstrated a higher isobutene uptake rate.¹⁹ The prediction of propane/propylene separation through ML methods was carried out by Tang and colleagues. The CoREMOF database was used as a training set, and CSDMOF was employed as a test set. Nine CSDMOFs with performance superior to CoREMOF were successfully predicted.²⁰ Luo et al. combined ML with experimental synthesis by establishing a crystal synthesis database to predict new experimental synthesis conditions for MOFs.²¹ The ability to predict unknown structures and discover potential relation-

ships between structure and performance had gradually become popular through ML.

In this paper, a high-throughput screening technique was employed to screen 656 COFs from the CURATED database, identifying COF materials with optimal adsorption performance on the GCMC simulation (Figure 1). Features and simulation results of COFs were used to train and test for four ML algorithms. The relative importance (RI) was used to analyze key factors influencing working capacity, selectivity, and regenerability by a prediction model. Based on the relationship between structure feature and performance indicators, the mechanisms behind the high adsorption performance of the best-performing COFs were also explored by density functional theory (DFT) calculation.

2. COMPUTATIONAL DETAILS

2.1. COF Database. The high-throughput screening was utilized in the CURATED database, which included 656 COFs. These COFs had meticulously optimized geometries and high-quality point charges derived from CP2K software by Smit et al.²² In this study, we used Zeo++ 0.3 software to calculate various structural features of the COFs, including parameters such as the largest cavity diameter (LCD), pore limiting diameter (PLD), volume surface area (VSA), pore volume (V_p), density (ρ), and void fraction (φ).²³ Determining the surface area in porous materials is commonly achieved through the Brunauer–Emmett–Teller (BET) method, which relies on nitrogen adsorption within these materials. In this investigation, we employed a probe radius of 1.86 Å, corresponding to the radius of a nitrogen molecule, to estimate the accessible surface area of the COF materials. Considering both accuracy and computational cost, a COF with a PLD greater than 5.6 Å (slightly larger than molecular dynamics diameter of SF₆, 5.13 Å) is used as a reference standard in this paper.^{16,24} The structural diversity of the COFs obtained from

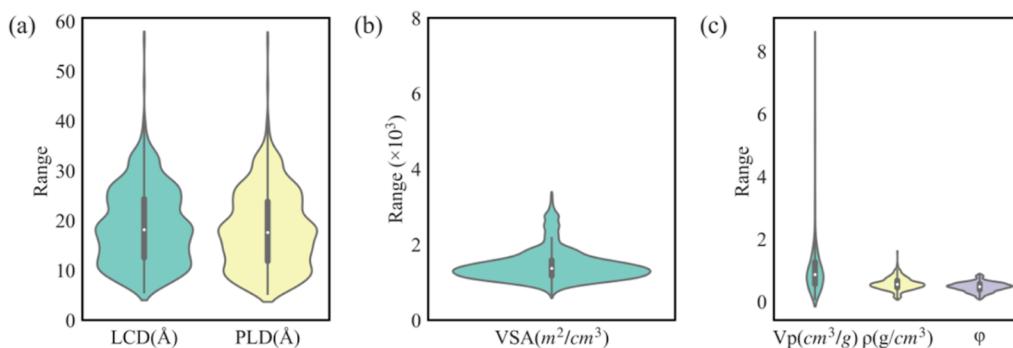


Figure 2. (a) PLD and LCD, (b) VSA, and (c) V_p , ρ , and φ of the 602 COFs. The height and width of the violin plots show the range and probability density distribution of COF structural features of the CURATED database.

this initial screening was visually displayed through a violin plot and a structure table (Figure 2 and Table S1).

2.2. Molecular Simulations. To begin our investigation, we initiated grand canonical Monte Carlo (GCMC) simulations using RASPA software version 2.0.35 to calculate the gas uptake of SF_6 and N_2 in the COFs at 0.1/1 bar and 298 K.^{25,26} Studies indicated that a 10:90 ratio of SF_6/N_2 yielded higher insulating properties than pure SF_6 gas.^{7,27} To match industrial SF_6/N_2 mixture condition and literature condition, we set the ratio of SF_6/N_2 to 10:90.^{28–32} The SF_6 molecular force field model was represented as a rigid six-point molecular model, as established by Dellis and Samios.³³ This model encompassed six fluorine atoms, with one sulfur atom serving as the center of mass. The N_2 molecular force field model was depicted as a three-point rigid model developed by Martin-Calvo et al.³⁴ This model comprises two nitrogen atoms and a virtual atom acting as the center of mass (COM). The interactions between the host–gas and gas–gas molecules were described using the Lennard-Jones 12–6 (LJ) potential and are typically represented by eq 1. The interaction parameters for the COF atoms were obtained from the generic force field known as DREIDING.³⁵ The interaction parameters for some of the metal atoms within the COFs were derived from the universal force field (UFF).³⁶ The LJ interactions were truncated at a cutoff distance of 12 Å, and long-range corrections were not applied. The box side lengths in the simulations were all greater than 24 Å. The Lorentz–Berthelot mixing rule was employed to determine the force field parameters for interactions between different types of atoms. The Coulomb potential was utilized to compute the electrostatic interactions among SF_6 , N_2 , and COF atoms in the GCMC simulations. To account for long-range electrostatic interactions, the Ewald summation method was applied. During the GCMC simulation, 20,000 cycles were executed for initialization purposes, and additional 40,000 cycles were conducted to determine the overall mean value of gas absorption. Each loop utilized five different types of moves: translation, rotation, reinsertion, exchange, and molecular identity exchange. The Peng–Robinson equation of state determined the gas fugacity coefficient. To verify the reliability of the above method and matching the experimental isotherms, the SF_6 isotherms obtained from GCMC simulations were compared with experimental results from ACOF-1 and BrCOF-1 (Figure S1)^{13,14}

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

In eq 1, q_i and q_j represent the partial charges associated with atom i and atom j , respectively, and ϵ_0 denotes the vacuum dielectric constant, equal to $8.8542 \times 10^{-12} \text{ C}^2 \text{ N}^{-1} \text{ m}^{-2}$. Moreover, ϵ_{ij} and r_{ij} signify the depth of the potential well and the collision volume associated with atoms i and j , respectively. Furthermore, the infinite dilution heat (H_{id}) was calculated using the subsequent equation (eq 2) on GCMC simulation²⁵

$$H_{id} = \langle U_{hg} \rangle - \langle U_h \rangle - \langle U_g \rangle - RT \quad (2)$$

where U_{hg} , U_h , and U_g are the average energy of the guest molecule inside the host framework, the average energy of the host framework, and the average energy of the guest molecule, respectively. In simulations, a common approximation was used to assume that the framework is rigid, and in this case, the enthalpy of adsorption at infinite dilution could be understood to be the difference in the internal energy of a single molecule outside and inside the confinement of the host framework. In the limit of zero temperature, the enthalpy of adsorption becomes the binding energy.

The gas working capacity (ΔW_{SF_6}) was defined as the disparity between the adsorption capacity and desorption capacity of the gas during adsorption, as represented by eq 3. Additionally, the selectivity ($S_{ads,SF_6/N_2}$) quantified the propensity of the COFs to adsorb a specific gas in SF_6 and N_2 and could be expressed using eq 4. The regenerability (Re %) was a measure of the ability of an adsorbent material to be regenerated, as expressed by eq 5. W_{ads,SF_6} signified the adsorption capacity of SF_6 under adsorption conditions at 1 bar, while W_{des,SF_6} presented the adsorption capacity of SF_6 under desorption conditions at 0.1 bar, where y_{SF_6} denoted the mole fraction of SF_6 within the mixture gas, and y_{N_2} signified the mole fraction of N_2 within the same mixture gas. To provide a comprehensive assessment of the efficacy of the adsorption separation process, the adsorbent performance score (APS) was computed using eq 6

$$\Delta W_{SF_6} = W_{ads,SF_6} - W_{des,SF_6} \quad (3)$$

$$S_{ads,SF_6/N_2} = \frac{W_{ads,SF_6}/W_{ads,N_2}}{y_{SF_6}/y_{N_2}} \quad (4)$$

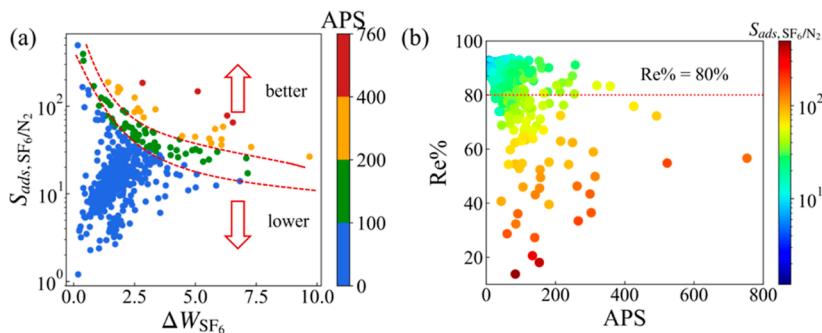


Figure 3. (a) Relationship between $S_{\text{ads},\text{SF}_6/\text{N}_2}$, ΔW_{SF_6} , and APS of COFs for SF_6/N_2 (10:90) separation. (b) Relationship between $\text{Re}\%$, $S_{\text{ads},\text{SF}_6/\text{N}_2}$, and APS. Red line represents $\text{Re}\% = 80\%$.

$$\text{Re}\% = \frac{\Delta W_{\text{SF}_6}}{W_{\text{ads},\text{SF}_6}} \times 100 \quad (5)$$

$$\text{APS} = S_{\text{ads},\text{SF}_6/\text{N}_2} \times \Delta W_{\text{SF}_6} \quad (6)$$

2.3. Machine Learning. All computed results from GCMC simulations were utilized for training and testing various ML models, including support vector machines (SVMs), random forests (RFs), gradient boosting regression tree (GBRT), and extreme gradient boosting (XGB). All ML models were implemented by the scikit-learn (sklearn) module in Python.^{37,38} All parameters were presented (Table S3). In the end, ML models were trained and tested using 11 geometric features (such as φ , PLD, LCD, V_p , VSA, and ρ), chemical features (element density, electronegativity, unsaturation, nitrogen-to-oxygen ratio, carbon-to-oxygen ratio, etc.), and energy features (infinite dilution heat) from the CURATED database (Table S4). The database was randomly divided into an 80% training set and 20% testing set. Fivefold cross-validation was employed to further get the average performance of the model. The accuracy of the models was assessed using the coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE).

2.4. DFT Calculations. The structural units of the COFs were determined using the Vienna Ab initio Simulation Package (VASP).^{39–42} The electron–ion interaction was described by the projector-augmented wave generalized gradient approximation (PAW-GGA) method and the Perdew–Burke–Ernzerhof (PBE) function to compute the exchange–correlation energy.^{43,44} The GGA/PBE functional with dispersion correction (DFT-D3) was utilized to optimize the COF structures and calculate adsorption interactions.⁴⁵ Geometrical relaxation of the structure of the COFs involved a plane-wave cutoff energy of 500 eV and a Gaussian smearing energy of 0.05 eV, ensuring optimization of both cell volume and atomic positions. Utilizing the conjugate gradient algorithm, the self-consistent field (SCF) convergence was achieved when the change in electronic energy was less than 10^{-5} eV. The convergence criterion for ionic relaxation was set at forces on all atoms below 0.01 eV/Å. The Brillouin zone was sampled with a set of K -point ($3 \times 1 \times 3$) according to the Monkhorst–Pack scheme. The electronic energy of the optimized structure was less than 10^{-7} eV in single-point calculations. The DDEC charges of the optimized structure was obtained using the CHARGEMOL software.^{46–48} The interaction between COFs and molecules was analyzed using the independent gradient model based on Hirshfeld partition (IGMH).^{49,50} GCMC simulations were conducted to obtain

the COM probability distributions within the identified COFs. The adsorbed molecules were positioned at the preferred adsorption sites determined by GCMC simulations, followed by a thorough structural relaxation using VASP. The binding energy (E_b) for each adsorption site was calculated using the following formula eq 7

$$E_b = E_{\text{gas+COF}} - E_{\text{COF}} - E_{\text{gas}} \quad (7)$$

Among these terms, $E_{\text{gas+COF}}$ represents the total energy of structural optimization after the gas loading into the COF, E_{COF} represents the total energy with structural optimization of the COF, and E_{gas} represents the total energy with structural optimization of the gas.

Differential charge density ($\Delta\rho$, e/Bohr³) and Bader charge (ΔQ , e) for charge transfer between adsorbent molecules (SF_6 , N_2) and COFs were calculated using the following formula eqs 8 and 9

$$\Delta\rho = \rho_{\text{gas+COF}} - \rho_{\text{COF}} - \rho_{\text{gas}} \quad (8)$$

$$\Delta Q = Q_{\text{ads}} - Q_{\text{vac}} \quad (9)$$

The $\rho_{\text{gas+COF}}$, ρ_{COF} , and ρ_{gas} (e/Bohr³) (1 Bohr = 0.05291772 nm) were the charge densities of the adsorption system, COFs, and adsorbent molecules (SF_6 , N_2), respectively; Q_{ads} and Q_{vac} (e) were the Bader charges of atoms in the stable adsorption system and under vacuum conditions, respectively.

3. RESULTS AND DISCUSSION

3.1. Screening High-Performance COFs. We evaluated the adsorption performance of 602 COFs from the CURATED database through GCMC simulations. The relationship between the $S_{\text{ads},\text{SF}_6/\text{N}_2}$ and ΔW_{SF_6} of the database illustrated that COFs with high $S_{\text{ads},\text{SF}_6/\text{N}_2}$ were not characterized by high ΔW_{SF_6} , and vice versa (Figure 3a). For example, 19371N2 had high $S_{\text{ads},\text{SF}_6/\text{N}_2}$ and low ΔW_{SF_6} ($S_{\text{ads},\text{SF}_6/\text{N}_2} = 184.730$, $\Delta W_{\text{SF}_6} = 2.828$ mmol/g), and 19001N3 had high ΔW_{SF_6} and low $S_{\text{ads},\text{SF}_6/\text{N}_2}$ ($S_{\text{ads},\text{SF}_6/\text{N}_2} = 26.528$, $\Delta W_{\text{SF}_6} = 9.690$ mmol/g) (Table S5). To balance COFs with $S_{\text{ads},\text{SF}_6/\text{N}_2}$ and ΔW_{SF_6} , we employed the APS to screen COFs.^{18,19,51} The color code intuitively revealed that red signifies COFs with the best performance ($400 < \text{APS} < 760$), yellow indicates COFs with better performance ($200 < \text{APS} < 400$), green denotes COFs with moderate performance ($100 < \text{APS} < 200$), and blue represents lower adsorption ($\text{APS} < 100$) (Figure 3a). The

20 COFs with best and better APS were selected in order of APS from high to low (Tables S5 and S6). The Re % served as a critical indicator in assessing the potential practicality of gas separation adsorbents, with higher Re % values signifying suitability for cyclic production. The correlation between the $R\%$, $S_{ads,SF_6/N_2}$, and APS had been shown (Figure 3b), where the red line represents the minimum value for high Re %. Approximately, 86% of the COF materials within the database exhibited high Re % ($Re\% > 80\%$), and these materials were considered to possess the potential to be efficient and economical in gas separation adsorption. However, it was noteworthy that only a few COF materials had both high Re % and high APS. Furthermore, our observations revealed that those COF materials with high $S_{ads,SF_6/N_2}$ and APS values tended to exhibit low Re % values (Figure 3b). As an illustration, 16400N2 ($S_{ads,SF_6/N_2} = 125.415$, APS = 314.391, Re % = 49.854) and 19371N2 ($S_{ads,SF_6/N_2} = 184.730$, APS = 522.499, Re % = 54.794) had high $S_{ads,SF_6/N_2}$ and APS but relatively lower Re %.

Our finding revealed an inverse relationship between Re % and APS of adsorbent materials in the experimental data (Figure 4). This finding implied a trade-off between Re % and

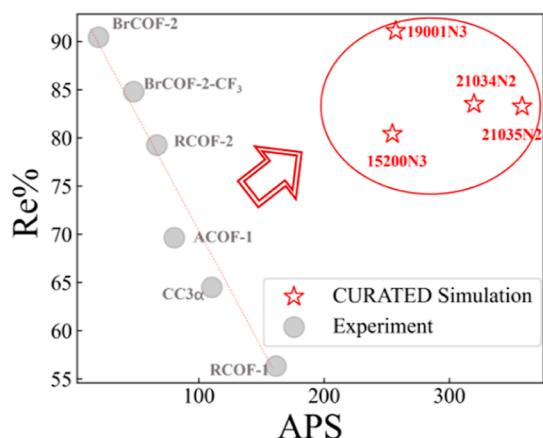


Figure 4. Comparison of experimental and simulation data with Re % and APS. The red star represents COFs with high APS in GCMC simulation. The gray scatter represents adsorbent materials on the SF₆/N₂ separation experiment (Table S7).

APS of the adsorbent materials in SF₆/N₂ adsorption separation experiments, which posed a challenge for researchers to obtain COFs with both high Re % and APS. The top four COFs (21035N2, 21034N2, 19001N3, and 15200N3) with high Re (Re % > 80%) and APS were identified among the 20 COFs (Figure 4 and Table S7). The APS of four COFs (Re % > 80%) were higher than the adsorbent materials on the SF₆/N₂ separation experiment (Figure 4). It was further stated that those COFs not only possessed low cycle production costs but also exhibited excellent SF₆/N₂ separation potential. The excellent SF₆/N₂ separation performance of COFs would be analyzed by GCMC simulations and DFT calculations in the subsequent study of the separation mechanism.

3.2. Machine Learning Analysis. In the previous section, COFs with higher APS (Re % > 80%) had been selected. We employed four ML models, namely, SVM, RF, GBRT, and XGB, to conduct multivariate analysis and unveil the

relationship between features and performance indicators within COFs. The SVM model utilized kernel functions to diagram nonlinear models to feature space for regression.⁵² Unlike linear regression, SVM could be used to avoid difficulties of using linear functions in the high-dimensional feature space, and optimization problem was transformed into dual convex quadratic programs. RF, GBRT, and XGB were ensemble learnings, built upon decision trees (DTs). Ensemble learning amalgamated multiple weak learners to combine them into one strong learner. This approach enhances the result accuracy compared to individual weak classifiers. Ensemble learning could be categorized into parallelization and serialization methods: RF exemplified parallelization methods, generating individual learners simultaneously, while GBRT and XGB represented serialization methods, wherein weak learners were dependent and generated sequentially. While RF achieved high accuracy due to its pseudorandom sampling and random averaging, it might be overfitted in some noisy classification or regression problems. GBRT built upon RF's accuracy but lacked parallel processing capability due to the sequential nature of weak learner results.⁵³ XGB employed second-order Taylor expansion to optimize the cost function, surpassing GBRT's reliance solely on first-order derivatives.³⁸ In addition to selecting six structural features (φ , PLD, LCD, V_p , VSA, and ρ), we also considered the characteristics of element density, electronegativity, unsaturation, carbon-to-nitrogen ratio, nitrogen-to-oxygen ratio, infinite dilution heat (H_{id}), etc. The H_{id} obtained from GCMC simulations, represented the enthalpy of adsorption when loading single SF₆ molecules into COFs at equilibrium. This feature was crucial in comprehending microporous adsorption behavior within porous materials. To ensure model accuracy, we employed the K-fold cross-validation method, running the ML model five times and averaging the results (Table S8). It is worth noting that XGB had significantly better performance indicators than the other ML models, followed by GBRT and RF, while SVM exhibited the worst performance indicators (Table 1). The results indicated that XGB achieved high accuracy in various performance indicators including Re % ($R^2 = 0.809$, RMSE = 2.440, MAE = 4.882), $S_{ads,SF_6/N_2}$ ($R^2 = 0.521$, RMSE = 8.634, MAE = 25.350), $\ln(S_{ads,SF_6/N_2})$ ($R^2 = 0.855$, RMSE = 0.192, MAE = 0.319), APS ($R^2 = 0.594$, RMSE = 20.279, MAE = 45.908), $\ln(APS)$ ($R^2 = 0.808$, RMSE = 0.327, MAE = 0.484), and ΔW_{SF_6} ($R^2 = 0.706$, RMSE = 0.401, MAE = 0.613) (Table 1 and Figures S2–S7). It was important to highlight that using $\ln(S_{ads,SF_6/N_2})$ and $\ln(APS)$ within the same ML models resulted in enhanced predictive performance compared to the direct use of $S_{ads,SF_6/N_2}$ and APS (Figures S3, S5–S7). This improvement could be attributed to the narrow data intervals provided by the logarithmic transformation of these variables, which likely ameliorated issues associated with skewed distributions or uneven data intervals, reflected in other literature.^{54,55} The accuracy of the XGB model for Re % and $\ln(S_{ads,SF_6/N_2})$, respectively, was 0.809 and 0.855, suggesting a strong fit to those indicators (Figure 5). In contrast to traditional greedy algorithms, XGB utilized an exact greedy algorithm to explore all best possible feature split. Consequently, XGB demonstrated superior generalization ability, as evidenced by the analysis of the predicted four ML algorithms in various performance indicators.

Table 1. Evaluation Metrics for Four ML Algorithms

performance indicators	ML models	database	average		
			R ²	MAE	RMSE
$S_{\text{adsSF}_6/\text{N}_2}$	SVR	test	0.445	9.578	30.122
	RF		0.539	9.565	25.140
	GBRT		0.549	8.737	25.457
	XGB		0.521	8.634	25.350
	SVR	train	0.405	7.254	32.790
	RF		0.883	6.971	14.412
	GBRT		0.971	2.690	7.200
	XGB		0.998	0.980	1.642
	SVR		0.569	0.453	0.746
ΔW_{SF_6}	RF	test	0.447	0.561	0.845
	GBRT		0.622	0.437	0.691
	XGB		0.706	0.401	0.613
	SVR		0.923	0.171	0.315
	RF		0.629	0.472	0.693
	GBRT	train	0.961	0.161	0.225
	XGB		0.982	0.104	0.152
	SVR		0.605	3.286	7.390
	RF		0.802	2.563	4.945
	GBRT		0.813	2.349	4.812
Re %	XGB	test	0.809	2.440	4.882
	SVR		0.805	1.702	5.284
	RF		0.913	2.044	3.528
	GBRT		0.984	0.895	1.496
	XGB		0.999	0.269	0.392
	SVR	train	0.348	23.955	58.086
	RF		0.469	24.265	52.180
	GBRT		0.538	20.734	48.588
	XGB		0.594	20.279	45.908
	SVR		0.386	20.112	56.962
ΔW_{SF_6}	RF	test	0.723	19.856	38.291
	GBRT		0.959	6.872	14.772
	XGB		0.999	1.602	2.252
	SVR		0.770	0.242	0.398
	RF		0.836	0.229	0.340
	GBRT	train	0.852	0.196	0.323
	XGB		0.855	0.192	0.319
	SVR		0.984	0.085	0.108
	RF		0.882	0.196	0.289
	GBRT		0.982	0.081	0.113
ln(APS)	XGB		0.982	0.071	0.113
	SVR	test	0.729	0.372	0.571
	RF		0.723	0.403	0.580
	GBRT		0.770	0.349	0.529
	XGB		0.808	0.327	0.484
	SVR	train	0.972	0.128	0.184
	RF		0.795	0.359	0.502
	GBRT		0.975	0.132	0.174
	XGB		0.984	0.096	0.142

Single indicators alone could not accurately identify high-performance COFs. This was because COFs with high adsorption capabilities typically required a combination of several indicators, such as $S_{\text{adsSF}_6/\text{N}_2}$ and ΔW_{SF_6} . Therefore, to accurately assess COFs with superior adsorption performance, we employed APS as a comprehensive adsorbent performance, which also was consistent with previous studies.^{18,19,51} The subsets including TOP20%, TOP40%, TOP60%, and

TOP80% were obtained by ordering APS from high to low (Figure 6a,b). Each subset utilized the XGB model to predict the RI of Re %, $S_{\text{adsSF}_6/\text{N}_2}$, and ΔW_{SF_6} (Figures 6b and S8). As the subsets from all to TOP20%, the proportion of energy features decreased, while the proportion of structural features gradually increased. In the subsets of TOP20%, structural features (42.380%) of the total importance indicated their significant impact on Re % (Figure 6a). In the entire database, H_{id} (31.122%) exhibited the highest RI, which could also be reflected in the Pearson correlation coefficient (0.736) (Figure 6b,c). Furthermore, for the subsets from all to TOP20%, the RI of φ gradually increased (from 10.364 to 22.627%), suggesting that it had a significant effect on the Re % of COFs with high adsorption performance (Figure 6b). Similarly, φ and H_{id} exhibited a significant effect on $\ln(S_{\text{adsSF}_6/\text{N}_2})$ (Figure S8a). Specifically, as subsets moved from all to TOP20%, the RI of φ increased (from 8.375 to 24.545%), and H_{id} decreased (from 43.004 to 17.139%).

The strong correlation observed between two structural features (φ and H_{id}) and performance indicators (Re % and $S_{\text{adsSF}_6/\text{N}_2}$) had verified those RI of those features in 20 COFs with high APS screened by high-throughput screening (Figure 7). The range of φ (0.207–0.527), H_{id} (19.868–43.611 kJ/mol), and ρ (0.419–0.919 g/cm³) for the 20 COFs of high APS was displayed (Table S6). Compared to all COFs, the structural features of the 20 COFs with high APS exhibited a narrower range, which suggested the likelihood for predicting COFs with excellent performance solely from structural features (Table S1). The negative correlation of Re % with H_{id} and the positive correlation with φ further emphasized the significance of these structural features in predicting the performance of the COFs (Figure 7a,b). In the subset of TOP20% COFs, H_{id} (16.446%), ρ (19.398%), and φ (15.186%) had a significant effect on ΔW_{SF_6} (Figure S8b). Additionally, the ΔW_{SF_6} of COFs with high APS demonstrated a negative correlation with H_{id} and ρ and positive correlation with φ (Figure S9). Notably, the RI analysis revealed a clearer relationship compared to the Pearson correlation analysis between ΔW_{SF_6} and ρ (-0.044) or φ (-0.086) (Figures 6c and S9a,b). Compared to the entire data set, analyzing the RI of the TOP20% COFs proved it a more effective predictor among high-APS COFs and it is facilitated to search similar structures with high-APS COFs.

3.3. Adsorption Mechanism of the Top COFs. In view of the 21035N2 material exhibiting a high Re % and APS (Re % = 83.27%, APS = 357.89), it was necessary to further investigate its adsorption mechanism. We found four isomeric frameworks within the 21035N2 material, synthesized from A₂B₂-type tetraphenyl benzene monomers (*p*-TetPB, *m*-TetPB, *o*-TetPB, TetPB-CHO, and TetPB-4NH₂), which had the para-, meta-, and ortho-substituted isomeric structures.⁵⁶ We could obtain COF materials with different structures and properties by combining different monomers (Figure S10). We utilized molecular simulations to explore the properties of these four COFs and named them COF-635 (*p*-TetPB-COF-M), COF-637 (*m*-TetPB-COF-M), COF-638 (*o*-TetPB-COF-M), and COF-639 (21035N2, TetPB-TetPB-COF). In the original literature, researchers confirmed through thermogravimetric analysis (TGA) that all TetPB-COFs displayed exceptional thermal stability, with no significant weight loss observed until 535 °C under a N₂ atmosphere.⁵⁶ Structural

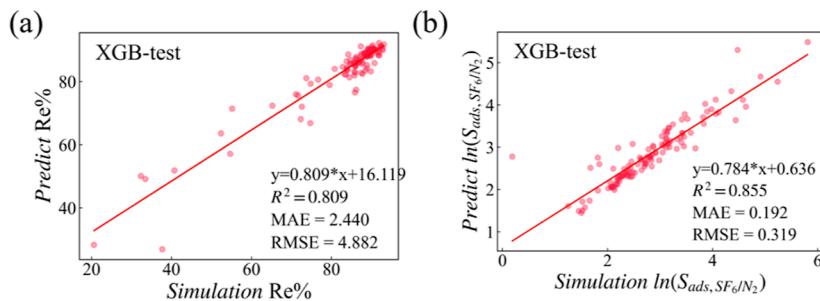


Figure 5. Comparison of prediction and simulation of (a) $\ln(S_{ads,SF_6/N_2})$ and (b) Re %.

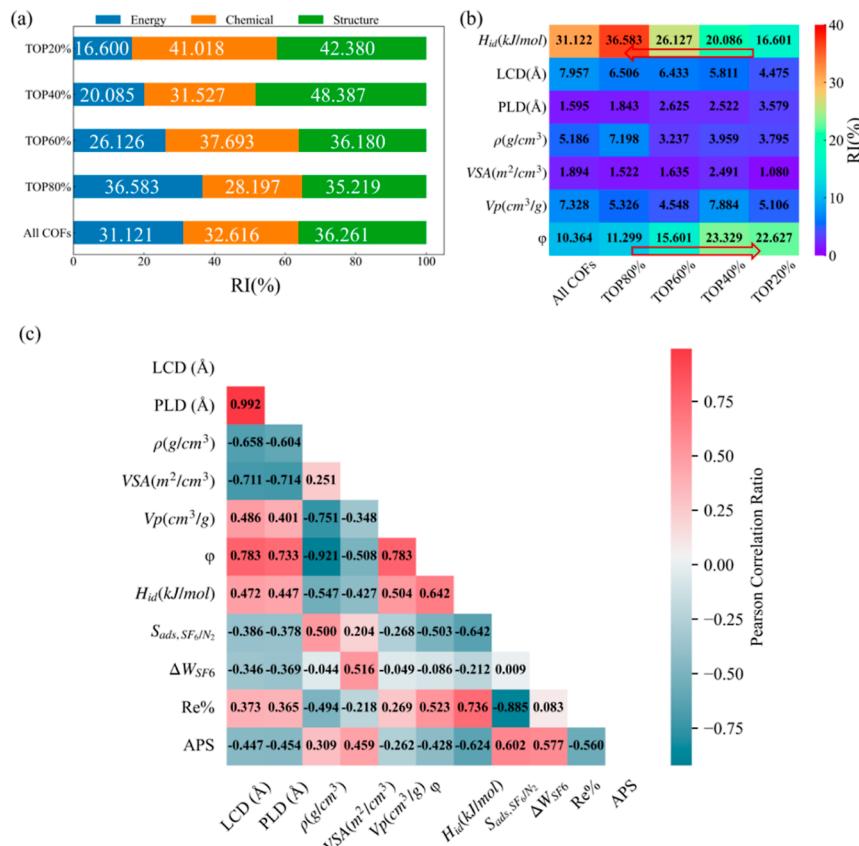


Figure 6. (a) RI of energy, chemical, and structural features. (b) RI of different features predicted by XGB models for Re % with the range from all to TOP20% COFs. (c) The Pearson correlation analyses of LCD, PLD, ρ , VSA, V_p , H_{id} , $S_{ads,SF_6/N_2}$, ΔW_{SF_6} , Re %, and APS.

features of the four COFs (COF-635, COF-637, COF-638, and COF-639) were optimized using VASP software, and then, DDEC charges were applied to the COFs with structural optimization.

To explore the adsorption mechanism, the COM density distribution diagrams of SF₆ and N₂ on the four COFs were plotted (Figures 8 and S11–S13). Comparing SF₆ and N₂ sites in COF-638, the COM diagrams revealed that the SF₆ site exhibited high intensity and a concentrated distribution in GCMC simulations (Figure 8). This phenomenon occurred because the molecular dynamics diameter of SF₆ (5.13 Å) was closer to the pore size of COF-638 than that of N₂ (3.64 Å), which lent a significant spatial site-blocking effect on the adsorption of SF₆ (Figure 8c). Conversely, the smaller molecular dynamics diameter of N₂ lent to a more dispersed distribution within the rhombic pore, and the distribution of SF₆ was more concentrated (Figure 8b,d). In the rhombic pore, SF₆ exhibited a high adsorption strength and was inclined

to adsorb in the acute region. This phenomenon was recognized by Liao et al. through SF₆/N₂ adsorption separation experiment.¹⁴ The same phenomenon was presented in the COM diagrams of COF-635, COF-637, and COF-639 (Figures S11–S13). To further illustrate the dominance of SF₆ in coadsorption, we visualized the coadsorption of SF₆ and N₂ by movie distribution diagrams in GCMC simulation. Although the amount of SF₆ was less than N₂ in the initial state (SF₆/N₂ = 10:90), movie distribution diagrams displayed a higher number of SF₆ than N₂ during adsorption equilibrium (Figure 9). The adsorption process was also visually demonstrated, indicating that SF₆ is dominant in coadsorption with N₂.

Isothermal adsorption curves for pure SF₆ and N₂ showed that COF-637 (1.553 mmol/g) exhibited the highest SF₆ adsorption at 0.1 bar, while COF-638 (0.999 mmol/g) displayed the lowest SF₆ adsorption (Figure 10a). However, COF-638 (5.136 mmol/g) showed the most significant

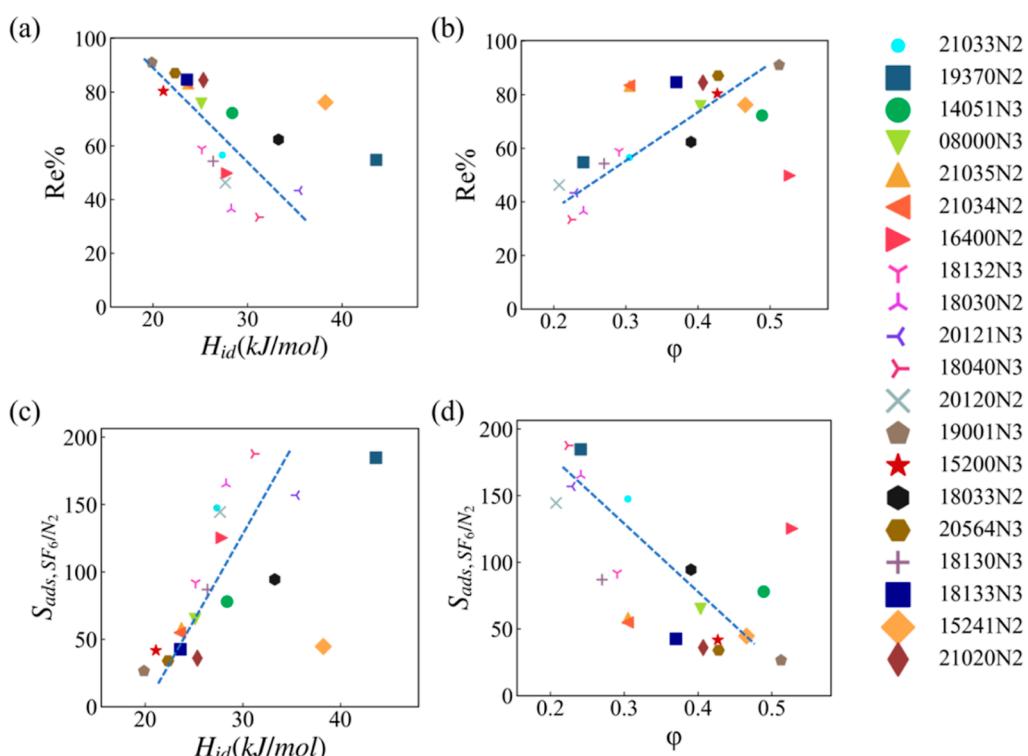


Figure 7. Relationship between (a) $Re\%$ and H_{id} , (b) $Re\%$ and ϕ , (c) $S_{ads, SF_6/N_2}$ and H_{id} , and (d) $S_{ads, SF_6/N_2}$ and ϕ of 20 COFs with higher APS screened in the CURATED database.

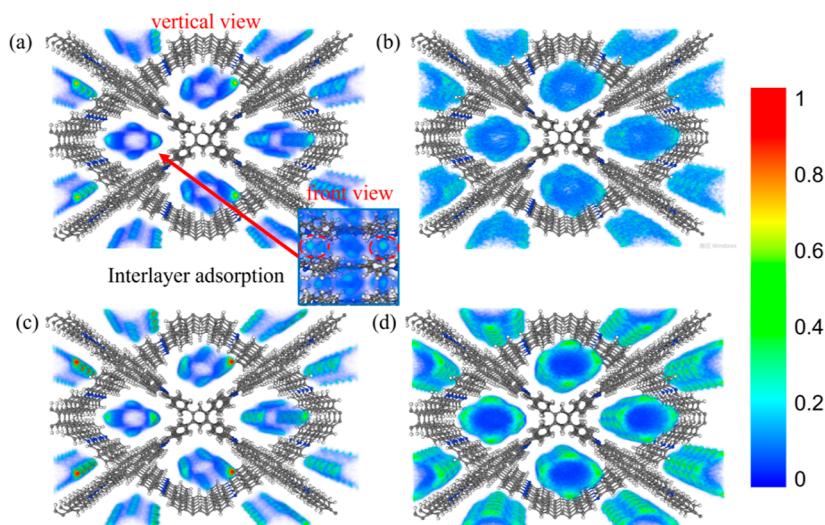


Figure 8. COM probability distribution diagrams of (a) SF₆ and (b) N₂ at 0.1 bar and of (c) SF₆ and (d) N₂ at 1 bar in COF-638 (SF₆/N₂ ratio is 10:90).

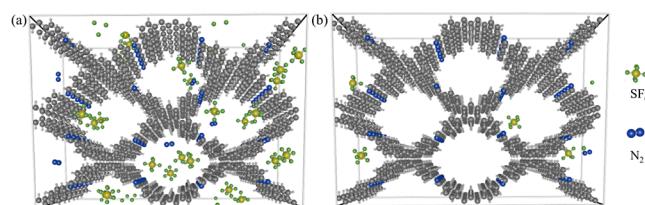


Figure 9. SF₆ and N₂ movie distribution diagrams at (a) 1 and (b) 0.1 bar (SF₆/N₂ = 10:90).

adsorption of SF₆ at 1 bar, followed by COF-639 (4.769 mmol/g), COF-637 (4.637 mmol/g), and COF-635 (2.872

mmol/g) (Table S10). This observation indicated that COF-638 had the highest ΔW_{SF_6} while COF-635 had the lowest ΔW_{SF_6} . The RI of the XGB model showed that ΔW_{SF_6} exhibited a significant correlation with ϕ , ρ , and H_{id} (Figure S8b). For the above significant correlation, COF-638 exhibited the highest ϕ and lowest ρ in the four COFs (Table S9). The $S_{ads, SF_6/N_2}$ of the four COFs increased with increasing pressure (Figure 10c). At 1 bar, COF-637 demonstrated the highest $S_{ads, SF_6/N_2}$ (82.607) in four COFs, suggesting that COF-637 possessed the strongest ability of SF₆/N₂ selective separation. Similarly, the adsorption curve directly demonstrated that SF₆

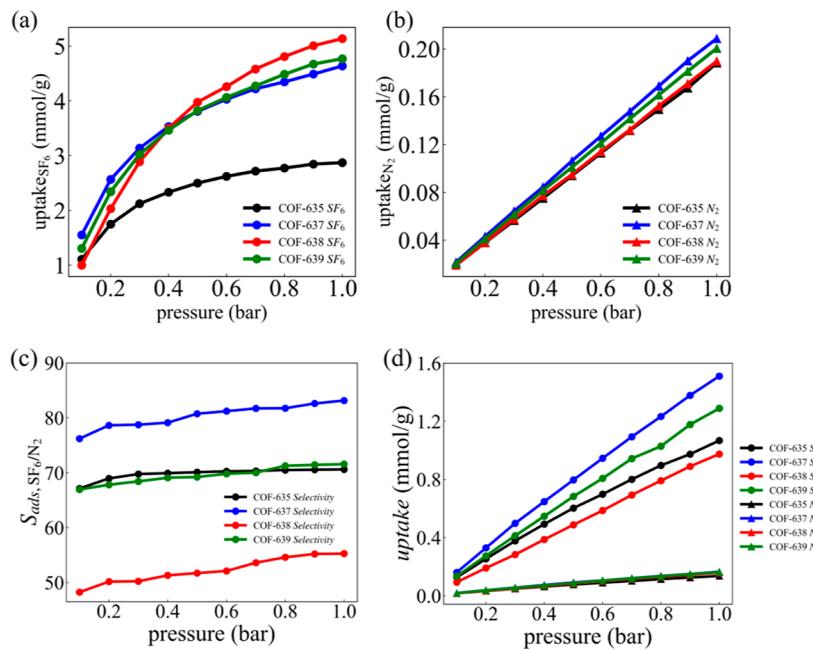


Figure 10. Adsorption curves of pure (a) SF₆ and (b) N₂ for COF-635, COF-637, COF-638, and COF-639. (c) Adsorption curve of selectivity. (d) Coadsorption curves of SF₆ and N₂ (SF₆/N₂ = 10:90, 298 K).

is adsorbed much more than N₂ during the coadsorption process of SF₆ and N₂ (Figure 10d and Table S10). RI analysis of the XGB model found that H_{id} (17.139%) and φ (24.545%) had a significant effect on S_{ads,SF₆/N₂} (Figure S8a). The increase in H_{id} increased the ability of the SF₆/N₂ selective separation. Further discussion of the differences in binding energies of COFs would be essential for a comprehensive understanding of their adsorption mechanisms.

The selective adsorption mechanism of SF₆/N₂ on promising COFs was investigated by a detailed study of the binding energy (E_b) on DFT. Based on the sites shown in the previous COM diagram, we labeled the acute region of the rhombic pore as site A and the obtuse region as site B (Figure 8a). Different binding sites (A and B) and their binding energies of different adsorbents (SF₆ and N₂) in different COFs were calculated (Figures S14–S17). The outcomes indicated significantly higher binding energies of SF₆ than those of N₂ on the COFs, consistent with the results obtained from the GCMC calculations. Specifically, the binding energies of four COFs (COF-635, COF-637, COF-638, and COF-639) loaded with SF₆ were -30.87, -29.15, -27.05, and -28.67 kJ/mol at site A, while at site B, the binding energies were -26.77, -24.49, -25.06, and -24.63 kJ/mol at site B. This observation suggested that site A with the highest adsorption affinity was the primary adsorption site, while site B was the secondary adsorption site. Despite that COF-635 showed the highest binding energies for SF₆ (-30.87 kJ/mol) and N₂ (-17.62 kJ/mol) at site A, it did not yield the highest S_{ads,SF₆/N₂} during GCMC simulations (Figure S8). This was due to the fact that COF-635 had the highest adsorption energy for N₂ (-17.62 kJ/mol), followed by COF-637 (-14.79 kJ/mol), COF-638 (-14.45 kJ/mol), and COF-639 (-15.13 kJ/mol), resulting in lower S_{ads,SF₆/N₂} than COF-637 (Table 2). Conversely, COF-637 exhibited superior S_{ads,SF₆/N₂} in GCMC simulations due to its higher binding energy with SF₆ (-29.15 kJ/mol) at site A and lower binding energy with N₂ (-14.79 kJ/mol) (Figure

Table 2. Binding Energy (E_b, kJ/mol) at the A and B Sites for SF₆ and N₂ for the COFs: COF-635, COF-637, COF-638, and COF-639

number	E _{b,A} (kJ/mol)		E _{b,B} (kJ/mol)	
	SF ₆	N ₂	SF ₆	N ₂
COF-635	-30.87	-17.62	-26.77	-13.81
COF-637	-29.15	-14.79	-24.49	-13.08
COF-638	-27.05	-14.45	-25.06	-13.48
COF-639	-28.67	-15.13	-24.63	-13.24

S9). Although the binding energy of COF-638 at site A (-27.05 kJ/mol) was lower than that of other COFs, it was due to the lower binding energy that COF-638 had higher ΔW_{SF_6} and Re %.

In the study of adsorbents with high S_{ads,SF₆/N₂} and Re % properties, the assessment of Re % emerged as a key indicator of SF₆/N₂ separation. Despite its moderate selectivity, COF-638 exhibited the highest ΔW_{SF_6} and Re % (80.549%) (Figure 11b). This high Re % suggested that COF-638 could be used

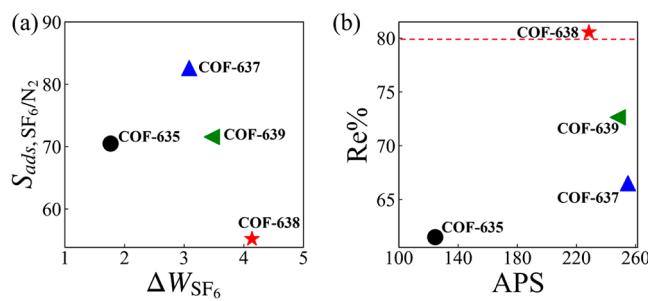


Figure 11. (a) Relationship between S_{ads,SF₆/N₂} and ΔW_{SF_6} of COF-635, COF-637, COF-638, and COF-639. (b) Relationship between the Re % and APS. The dashed line indicates that the Re % is above 80%.

iteratively, maintaining its performance over multiple cycles, and consequently leading to cost-saving benefits.¹⁷ In practical industrial separation applications where efficiency and cost effectiveness were critical indicators, ΔW_{SF_6} and Re % often had more significance than $S_{ads,SF_6/N_2}$. In comparison to other COFs, COF-638 was better suited for use in practical circular production (Figure 11b). In the final selection process, COF-638 (Re > 80%, $\Delta W_{SF_6} = 4.137$ mmol/g) was considered as an adsorbent material to meet the demands of practical industrial separation.

COF-638 was employed as a case study to investigate the adsorption mechanism of SF₆/N₂ separation in depth. The charge distributions of SF₆ and N₂ at site A and B were analyzed by charge density difference diagrams (Figure 12).

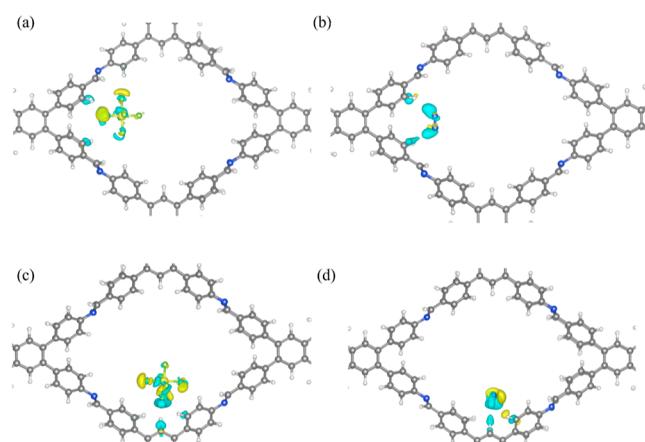


Figure 12. Charge density difference of SF₆ loaded to (a) acute region (site A) and (b) obtuse region (site B) and N₂ loaded to (c) acute region (site A) and (d) obtuse region (site B). Yellow represents gain of charge and blue represents loss of charge. The iso-surface is set to 0.002 e/Bohr³.

The charge density of H atoms on COF-638 near SF₆ appeared in blue color, indicating that the framework was losing charge (Figure 12a,c). Conversely, the yellow coloration around the F atom of SF₆ suggested a directional movement of charge toward the F atom (Figure 12a,c). The state of SF₆ losing and gaining charge was quantitatively described by Bader charge analysis. Comparison of Bader charges of F (7.000 e) and S (6.000 e) atoms of SF₆ revealed that six F atoms gained charge and S atoms lost charge. The Bader charge analysis indicated that the molecules acquired charge at A_{SF₆} (0.024e), B_{SF₆} (0.018e), A_{N₂} (0.008e), and B_{N₂} (0.008e) (Tables S11 and S12). Overall, the charge of SF₆ and N₂ increased while the charge of COF-638 decreased. Specifically, the charge transfer at site A_{SF₆} (0.024e) was higher at site B_{SF₆} (0.018e), suggesting that site A_{SF₆} had the strongest affinity for SF₆, which was consistent with the maximum binding energy of site A_{SF₆} (-27.05 kJ/mol) (Table S11). Meanwhile, SF₆ (0.024e) gained a higher charge than N₂ (0.008e) at site A. This phenomenon implied that SF₆ interacted more significantly with COF-638 compared to N₂.

To visualize the contrasting adsorption behaviors of SF₆ and N₂ within COF-638, COF-638 clusters were extracted and loaded with SF₆ and N₂ molecules. The independent gradient model based on IGMH was employed to analyze inter- and

intrafragment interactions.^{49,50} The green part between the adsorbent molecules (SF₆ and N₂) and COF-638 clusters suggested that the adsorption was a van der Waals interaction (Figure 13). Comparing the adsorption configurations of SF₆

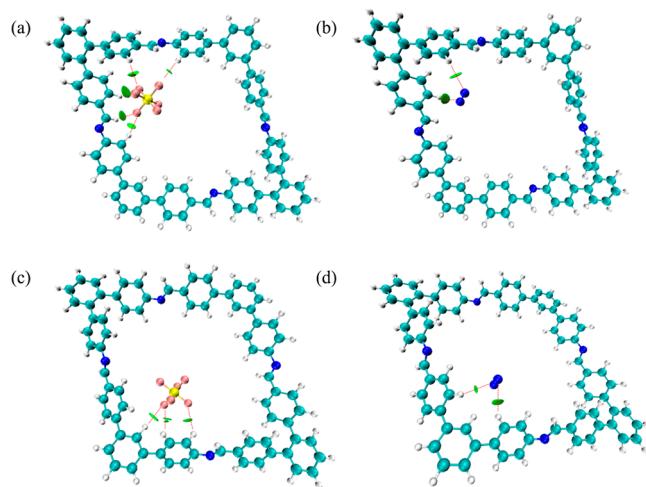


Figure 13. IGMH diagrams of the COF-638 cluster with (a) SF₆ and (b) N₂ at site A and with (c) SF₆ and (d) N₂ at site B. The iso-surfaces are set as 0.005 au.

at site A and site B, it was found that SF₆ at site B only interacted with the H atoms on the two adjacent benzene rings of the COF-638 cluster (Figure 13a, 13c). The lower attraction of the obtuse region (site B) of COF-638 to SF₆ is due to the lower van der Waals interaction between its H atoms and SF₆. Moreover, IGMH analysis showed that SF₆ at site A had van der Waals interaction with H atoms of benzene rings and imine bonds from both sides of COF-638, whereas N₂ merely interacted from one side of COF-638, which suggested that the affinity of COF-638 to SF₆ was stronger (Figure 13a,b). The weak interaction between N₂ and COF-638 may be due to their lower molecular dynamic diameter (3.64 Å). Based on the above analysis, N₂ only interacted with one side of COF-638 (Figure 13b,d). Therefore, SF₆ exhibited stronger adsorption capacity in the acute region (site A) compared to that in the obtuse region (site B) of COF-638.

4. CONCLUSIONS

In summary, four COFs (Re % > 80%) were selected from the CURATED database by high-throughput screening and their performance in the SF₆/N₂ separation experiment is excellent, which demonstrated higher SF₆/N₂ separation efficiency and lower industrial recycling costs. XGB was selected among four ML models (SVM, RF, GBRT, and XGB) to predict the optimal accuracy of Re % ($R^2 = 0.809$, RMSE = 2.440, MAE = 4.882) and $\ln(S_{ads,SF_6/N_2})$ ($R^2 = 0.855$, RMSE = 0.192, MAE = 0.319). RI analysis was employed to analyze the key features of Re % and $\ln(S_{ads,SF_6/N_2})$. From all to TOP20% of COFs, there was a gradual increase in the RI of φ and a decrease in the RI of H_{id} . φ and H_{id} had significant importance in Re % and $\ln(S_{ads,SF_6/N_2})$ in all database. 21035N2 (Re > 80%, APS = 357.890) was selected by high-throughput screening and three other COFs with similar structures were found (COF-635, COF-637, COF-638, and COF-639 (21035N2)). Among the four COFs, COF-638 showed the best adsorption performance (Re > 80%, $\Delta W_{SF_6} = 4.137$ mmol/g, $S_{ads,SF_6/N_2} = 55.212$) and

was considered as a material to meet the demands of practical industrial separation. Bader charge revealed that SF₆ gained higher charge than N₂ in site A. IGMH analysis illustrated that COF-638 had a stronger affinity for SF₆, because SF₆ at site A interacted with H atoms of benzene rings and imine bonds from both sides of COF-638, whereas N₂ merely interacted from one side of COF-638. The above analyses efficiently demonstrate that COF-638 exhibited an excellent SF₆/N₂ adsorption separation performance. Based on the relationship between structure and performance discussed in this study, it is hoped that the structure of COFs can be reasonably adjusted to meet specific application requirements and provide theoretical basis for future experimental research.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jpcc.4c01825>.

LJ parameters of SF₆ and N₂; descriptors and optimization parameters of the ML model; RI of the performance indicators; and adsorption site of the SF₆ and N₂ on the COFs ([PDF](#))

AUTHOR INFORMATION

Corresponding Authors

Rui Zhao — *Qingyuan Innovation Laboratory, Quanzhou 362801, P. R. China;*  orcid.org/0009-0004-1866-4240; Email: 1610977686@qq.com

Jingyu Cai — *Department of Materials-Oriented Chemical Engineering, College of Chemical Engineering, Fuzhou University, Fuzhou 350116, P. R. China;*  orcid.org/0000-0002-0233-3007; Email: jyc1991@fzu.edu.cn

Linxi Hou — *Department of Materials-Oriented Chemical Engineering, College of Chemical Engineering, Fuzhou University, Fuzhou 350116, P. R. China; Qingyuan Innovation Laboratory, Quanzhou 362801, P. R. China; Fujian Key Laboratory of Advanced Manufacturing Technology of Specialty Chemicals, Fuzhou University, Fuzhou 350116, PR China;*  orcid.org/0000-0002-7962-0936; Email: lzhou@fzu.edu.cn

Authors

Junjie Ning — *Department of Materials-Oriented Chemical Engineering, College of Chemical Engineering, Fuzhou University, Fuzhou 350116, P. R. China*

Kun Shen — *Department of Materials-Oriented Chemical Engineering, College of Chemical Engineering, Fuzhou University, Fuzhou 350116, P. R. China*

Kunqi Gao — *School of Science, College of Art and Science, Shanghai Polytechnic University, Shanghai 201209, PR China*

Complete contact information is available at: <https://pubs.acs.org/doi/10.1021/acs.jpcc.4c01825>

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was financially supported by the National Natural Science Foundation of China (grant no. 22208059), Natural Science Foundation of Fujian Province (grant no.

2022J05128), and Starting Research Fund of Qingyuan Innovation Laboratory (grant no. 00522006).

REFERENCES

- (1) Lee, H. M.; Chang, M. B.; Wu, K. Y. Abatement of sulfur hexafluoride emissions from the semiconductor manufacturing process by atmospheric pressure plasmas. *J. Air Waste Manage. Assoc.* **2004**, *54* (8), 960–970.
- (2) Kim, S.; Nagorny, P. Electrochemical synthesis of glycosyl fluorides using sulfur(vi) hexafluoride as the fluorinating agent. *Org. Lett.* **2022**, *24* (12), 2294–2298.
- (3) Montzka, S. A.; Dlugokencky, E. J.; Butler, J. H. Non-co2 greenhouse gases and climate change. *Nature* **2011**, *476* (7358), 43–50.
- (4) Hiziroglu, H. R.; Bian, X.; Lu, T. Experimental study of synergism in n2 and sf6 gas mixtures. In *2020 IEEE 3rd International Conference on Dielectrics (ICD)*, 2020; pp 814–817.
- (5) Jacobs, J. D. Climate change 1995—economic and social dimensions of climate change: Contribution of working group iii to the second assessment report of the intergovernmental panel on climate change. *Can. Geogr.* **1998**, *42* (1), 105–106.
- (6) Jacobs, J. D. Climate change 1995—impacts, adaptations and mitigation of climate change: Scientific-technical analyses: Contribution of working group ii to the second assessment report of the intergovernmental panel on climate change. *Can. Geogr.* **1998**, *42* (1), 105–106.
- (7) Chuah, C. Y.; Lee, Y.; Bae, T. H. Potential of adsorbents and membranes for sf6 capture and recovery: A review. *Chem. Eng. J.* **2021**, *404*, 126577.
- (8) Fang, X. K.; Hu, X.; Janssens-Maenhout, G.; Wu, J.; Han, J. R.; Su, S. S.; Zhang, J. B.; Hu, J. X. Sulfur hexafluoride (sf6) emission estimates for china: An inventory for 1990–2010 and a projection to 2020. *Environ. Sci. Technol.* **2013**, *47* (8), 3848–3855.
- (9) Hasell, T.; Miklitz, M.; Stephenson, A.; Little, M. A.; Chong, S. Y.; Clowes, R.; Chen, L. J.; Holden, D.; Tribello, G. A.; Jelfs, K. E.; et al. Porous organic cages for sulfur hexafluoride separation. *J. Am. Chem. Soc.* **2016**, *138* (5), 1653–1659.
- (10) Wang, S. M.; Mu, X. T.; Liu, H. R.; Zheng, S. T.; Yang, Q. Y. Pore-structure control in metal-organic frameworks (mofs) for capture of the greenhouse gas (sf6) with record separation. *Angew. Chem., Int. Ed. Engl.* **2022**, *61* (33), No. e202207066.
- (11) Yuan, S.; Li, X.; Zhu, J.; Zhang, G.; Van Puyvelde, P.; Van der Bruggen, B. Covalent organic frameworks for membrane separation. *Chem. Soc. Rev.* **2019**, *48* (10), 2665–2681.
- (12) Zheng, X. Q.; Shen, Y. L.; Wang, S. T.; Huang, K.; Cao, D. P. Selective adsorption of sf6 in covalent- and metal-organic frameworks. *Chin. J. Chem. Eng.* **2021**, *39*, 88–95.
- (13) Liao, Q.; Ke, C.; Huang, X.; Wang, D.; Han, Q.; Zhang, Y.; Zhang, Y.; Xi, K. A versatile method for functionalization of covalent organic frameworks via suzuki-miyaura cross-coupling. *Angew. Chem., Int. Ed. Engl.* **2021**, *60* (3), 1411–1416.
- (14) Liao, Q.; Xu, H.; Ke, C.; Zhang, Y.; Han, Q.; Zhang, Y.; Xu, Y.; Wang, D.; Xi, K. Rational regulating pore structures of covalent organic frameworks for sulfur hexafluoride capture and separation. *Sep. Purif. Technol.* **2023**, *306*, 122595.
- (15) Li, X.-D.; Wang, Y.-D.; Guo, F.; Feng, S.-Q.; Liu, X.-Y.; Yuan, J.-N.; Chen, Z. Computer-aided design of high-connectivity covalent organic frameworks as ch4/h2 adsorption and separation media. *Int. J. Hydrogen Energy* **2023**, *48* (34), 12753–12766.
- (16) Ren, J.; Chang, M.; Zeng, W.; Xia, Y.; Liu, D.; Maurin, G.; Yang, Q. Computer-aided discovery of mofs with calixarene-analogous microenvironment for exceptional sf6 capture. *Chem. Mater.* **2021**, *33* (13), 5108–5114.
- (17) Cao, X.; Zhang, Z.; He, Y.; Xue, W.; Huang, H.; Zhong, C. Machine-learning-aided computational study of covalent organic frameworks for reversed c2h6/c2h4 separation. *Ind. Eng. Chem. Res.* **2022**, *61* (30), 11116–11123.

- (18) Aksu, G. O.; Erucar, I.; Haslak, Z. P.; Keskin, S. Accelerating discovery of cofs for co₂ capture and h₂ purification using structurally guided computational screening. *Chem. Eng. J.* **2022**, *427*, 131574.
- (19) Sun, X.; Lin, W.; Jiang, K.; Liang, H.; Chen, G. Accelerated screening and assembly of promising mofs with open cu sites for isobutene/isobutane separation using a data-driven approach. *Phys. Chem. Chem. Phys.* **2023**, *25* (12), 8608–8623.
- (20) Tang, H.; Xu, Q.; Wang, M.; Jiang, J. Rapid screening of metal–organic frameworks for propane/propane separation by synergizing molecular simulation and machine learning. *ACS Appl. Mater.* **2021**, *13* (45), 53454–53467.
- (21) Luo, Y.; Bag, S.; Zaremba, O.; Cierpka, A.; Andreo, J.; Wuttke, S.; Friederich, P.; Tsotsalas, M. Mof synthesis prediction enabled by automatic data mining and machine learning. *Angew. Chem., Int. Ed.* **2022**, *61* (19), No. e202200242.
- (22) Ongari, D.; Yakutovich, A. V.; Talirz, L.; Smit, B. Building a consistent and reproducible database for adsorption evaluation in covalent-organic frameworks. *ACS Cent. Sci.* **2019**, *5* (10), 1663–1675.
- (23) Willems, T. F.; Rycroft, C.; Kazi, M.; Meza, J. C.; Haranczyk, M. Algorithms and tools for high-throughput geometry-based analysis of crystalline porous materials. *Microporous Mesoporous Mater.* **2012**, *149* (1), 134–141.
- (24) Cha, J.; Ga, S.; Lee, S.-j.; Nam, S.; Bae, Y.-S.; Chung, Y. G. Integrated material and process evaluation of metal–organic frameworks database for energy-efficient sf₆/n₂ separation. *Chem. Eng. J.* **2021**, *426*, 131787.
- (25) Dubbeldam, D.; Calero, S.; Ellis, D. E.; Snurr, R. Q. R. RASPA: molecular simulation software for adsorption and diffusion in flexible nanoporous materials. *Mol. Simul.* **2016**, *42* (2), 81–101.
- (26) Panagiotopoulos, A. Z. Direct determination of phase coexistence properties of fluids by monte-carlo simulation in a new ensemble. *Mol. Phys.* **1987**, *61* (4), 813–826.
- (27) Inami, K.; Maeda, Y.; Habuchi, Y.; Yoshimura, M.; Hamano, S.; Hama, H. Problems of the application of n₂/sf₆ mixtures to gas-insulated bus. *Electr. Eng. Jpn.* **2001**, *137* (4), 25–31.
- (28) Chang, M.; Yan, T.; Wei, Y.; Wang, J.-X.; Liu, D.; Chen, J.-F. Metal–organic framework-based single-molecule sf₆ trap for record sf₆ capture. *Chem. Mater.* **2022**, *34* (20), 9134–9143.
- (29) Wu, Y.; Wang, S.; Zhang, W.; Chen, S.; Zhang, Z.; Yang, B.; Li, S.; Ma, H. Enhancing perfluorinated electron specialty gases separation selectivity in ultra-microporous metal organic framework. *Sep. Purif. Technol.* **2022**, *289*, 120739.
- (30) Yang, M.; Chang, M.; Yan, T.; Liu, D. A nickel-based metal–organic framework for efficient sf₆/n₂ separation with record sf₆ uptake and sf₆/n₂ selectivity. *Sep. Purif. Technol.* **2022**, *295*, 121340.
- (31) Wu, Y.; Li, X.; Li, Y.; Tao, C.; Tang, Y.; Wang, S.; Fu, Y.; Ma, W.; Zhang, W.; Ma, H. Porous aromatic frameworks as hf resistant adsorbents for sf₆ separation at elevated pressure. *Sep. Purif. Technol.* **2023**, *315*, 123657.
- (32) Zheng, S.-T.; Jiang, R.-Y.; Jiang, Y.; Ni, S.; Guan, G.-W.; Shao, S.-Q.; Wang, Y.-C.; Wang, S.-M.; Yang, Q.-Y. Methyl-functionalized microporous metal–organic framework for efficient sf₆/n₂ separation. *Sep. Purif. Technol.* **2023**, *318*, 123957.
- (33) Dellis, D.; Samios, J. Molecular force field investigation for sulfur hexafluoride: A computer simulation study. *Fluid Phase Equilif.* **2010**, *291* (1), 81–89.
- (34) Martin-Calvo, A.; Garcia-Perez, E.; Garcia-Sanchez, A.; Bueno-Perez, R.; Hamad, S.; Calero, S. Effect of air humidity on the removal of carbon tetrachloride from air using cu-btc metal–organic framework. *Phys. Chem. Chem. Phys.* **2011**, *13* (23), 11165–11174.
- (35) Mayo, S. L.; Olafson, B. D.; Goddard, W. A. Dreiding - a generic force-field for molecular simulations. *J. Phys. Chem.* **1990**, *94* (26), 8897–8909.
- (36) Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. Uff, a full periodic-table force-field for molecular mechanics and molecular-dynamics simulations. *J. Am. Chem. Soc.* **1992**, *114* (25), 10024–10035.
- (37) Pedregosa, F. V. G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V. S. Scikit-learn: Machine learning in python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
- (38) Chen, T.; Guestrin, C. Xgboost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016; pp 785–794.
- (39) Kresse, G.; Furthmuller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1996**, *54* (16), 11169–11186.
- (40) Kresse, G.; Furthmuller, J. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Comput. Mater. Sci.* **1996**, *6* (1), 15–50.
- (41) Kresse, G.; Hafner, J. Abinitio molecular-dynamics for liquid-metals. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1993**, *47* (1), 558–561.
- (42) Kresse, G.; Hafner, J. Ab-initio molecular-dynamics simulation of the liquid-metal amorphous-semiconductor transition in germanium. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1994**, *49* (20), 14251–14269.
- (43) Blochl, P. E.; Forst, C. J.; Schimpl, J. Projector augmented wave method: Ab initio molecular dynamics with full wave functions. *Bull. Mater. Sci.* **2003**, *26* (1), 33–41.
- (44) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **1996**, *77* (18), 3865–3868.
- (45) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* **2011**, *32* (7), 1456–1465.
- (46) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (dft-d) for the 94 elements h-pu. *J. Chem. Phys.* **2010**, *132* (15), 154104.
- (47) Limas, N. G.; Manz, T. A. Introducing ddec6 atomic population analysis: Part 2. Computed results for a wide range of periodic and nonperiodic materials. *RSC Adv.* **2016**, *6* (51), 45727–45747.
- (48) Manz, T. A. Introducing ddec6 atomic population analysis: Part 3. Comprehensive method to compute bond orders. *RSC Adv.* **2017**, *7* (72), 45552–45581.
- (49) Lu, T.; Chen, Q. Independent gradient model based on hirshfeld partition: A new method for visual study of interactions in chemical systems. *J. Comput. Chem.* **2022**, *43* (8), 539–555.
- (50) Lu, T.; Chen, F. M. Multiwfns: A multifunctional wavefunction analyzer. *J. Comput. Chem.* **2011**, *33* (5), 580–592.
- (51) Altundal, O. F.; Altintas, C.; Keskin, S. Can cofs replace mofs in flue gas separation? High-throughput computational screening of cofs for co₂/n₂ separation. *J. Mater. Chem. A* **2020**, *8* (29), 14609–14623.
- (52) Hearst, M. A. a. D. S. T.; Dumais, S.; Osuna, E.; Platt, J.; Scholkopf, B. Support vector machines. *IEEE Intell. Syst. Appl.* **1998**, *13* (4), 18–28.
- (53) Friedman, J. H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29* (5), 1189–1232.
- (54) Zhang, Z.; Cao, X.; Geng, C.; Sun, Y.; He, Y.; Qiao, Z.; Zhong, C. Machine learning aided high-throughput prediction of ionic liquid@mof composites for membrane-based co₂ capture. *J. Membr. Sci.* **2022**, *650*, 120399.
- (55) Yan, Y.; Shi, Z.; Li, H.; Li, L.; Yang, X.; Li, S.; Liang, H.; Qiao, Z. Machine learning and in-silico screening of metal–organic frameworks for o₂/n₂ dynamic adsorption and separation. *Chem. Eng. J.* **2022**, *427*, 131604.
- (56) Li, Y.; Guo, L.; Lv, Y.; Zhao, Z.; Ma, Y.; Chen, W.; Xing, G.; Jiang, D.; Chen, L. Polymorphism of 2d imine covalent organic frameworks. *Angew. Chem., Int. Ed. Engl.* **2021**, *60* (10), 5363–5369.