

Combining Interpretable Machine Learning and Molecular Simulation to Advance the Discovery of COF-Based Membranes for Acid Gas Separation

Bingru Xin, Minggao Feng, Min Cheng, Zhongde Dai, Shiyang Ye, Li Zhou, Yiyang Dai,* and Xu Ji*



Cite This: <https://doi.org/10.1021/acs.iecr.4c00855>



Read Online

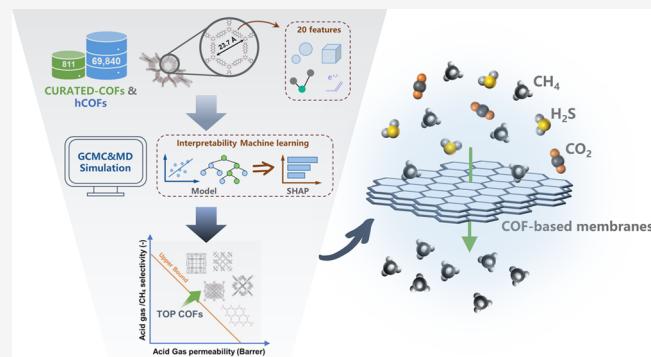
ACCESS |

Metrics & More

Article Recommendations

Supporting Information

ABSTRACT: Membrane technology can effectively remove acidic gases (H_2S and CO_2) from natural gas. Covalent organic frames (COFs) have been widely used as membrane materials due to their large pore size and pure organic properties. This work combines machine learning (ML) and molecular simulation (MS) to develop a method for rapidly screening and discovering high-performance COF-based membranes. The ML model is first trained on MS data, using the structural and chemical features obtained from 20 calculations as inputs. Characteristic contributions were obtained through interpretable analytical models, and nearly 70,000 COFs were quickly screened. Finally, the top 10 high-performance COFs were selected by MS under the mixed gas condition, and the properties of the mixed matrix membranes (MMMs) obtained by combining them with six polymers were analyzed. The results show that the highest acid gas permeability of COF-based membranes reaches 8×10^5 Barrer, and the void fraction is the key factor determining the separation performance. The top COFs serve as effective fillers to enhance the performance of polymer membranes, which surpass the capabilities of existing MOF fillers. This paper provides an efficient and rapid method for the discovery of COF-based membranes for natural gas deacidification.



COF-based membranes

1. INTRODUCTION

Natural gas is a viable alternative to traditional fossil fuels as an acceptable energy source in transition to clean energy systems and economies.¹ It plays a pivotal role in the global energy transition due to its abundant reserves, wide distribution, low pollution levels, and high calorific value.² Nevertheless, it is important to note that natural gas comprises not only CH_4 but also carbon dioxide (CO_2), hydrogen sulfide (H_2S), N_2 , and other impurities. Acidic impurities (e.g., H_2S and CO_2) in natural gas not only reduce the calorific value but also cause potential pipeline corrosion. Consequently, the separation of acid gases from natural gas becomes critical for its effective utilization.³ Presently, various technologies are employed for the desulphurization of natural gas, including liquid absorption,⁴ solid adsorption,⁵ and membrane separation.⁶ Among these methods, membrane separation has the advantages of small footprint, low energy consumption, and ease of operation.⁷ Conventional amorphous polymer membranes are limited by “permeability-selectivity” constraints.⁸ Over the past two decades, nanoporous materials (NPMs) have become key materials for enhancing membrane separation performance. Zeolites,⁹ metal–organic frameworks (MOFs),¹⁰ and porous polymer networks (PNMs)¹¹ have been widely investigated and exhibited good separation performance.

Covalent organic frameworks (COFs) are an emerging class of crystal network porous materials. Constructed by assembling organic joints composed of reversible covalent bonds, they have the advantages of low density, tunable structure and high porosity, etc.¹² They are widely applied in many fields, such as gas storage,¹³ separation,¹⁴ and catalysis.¹⁵ COFs can be prepared as pure COF-based membranes,¹⁶ as fillers added into polymer substrates to be prepared as COF-based mixed-matrix membranes (MMMs),¹⁷ and hybrid membranes (with MOFs¹⁸ and GO¹⁹). Like MOFs, the number of COFs that have been experimentally synthesized and documented has steadily increased over time, reaching a count in the thousands. The production of COF-based membranes is a complex process, requiring stringent conditions and lengthy gas separation testing cycles.²⁰ Consequently, examining every COF material through conventional trial-and-error methods in a laboratory is impractical.

Received: March 4, 2024

Revised: April 15, 2024

Accepted: April 17, 2024

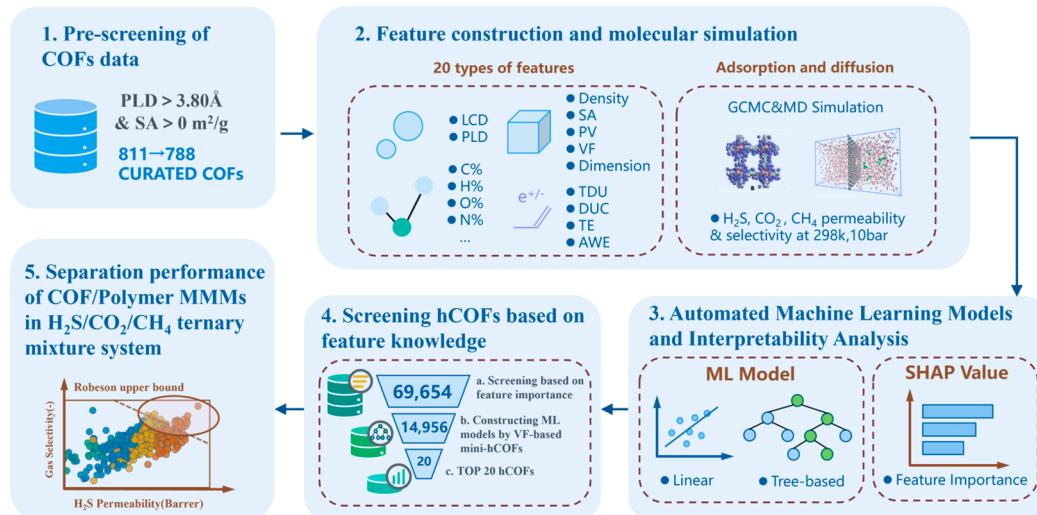


Figure 1. Workflow for interpretable ML-assisted discovery of innovative COF-based membrane materials with high acid gas separation performance.

With the rapid development of materials science and computer science, high throughput computational screening (HTCS) has become an effective method to evaluate material properties.^{21–23} For porous membranes, Grand Canonical Monte Carlo (GCMC) and molecular dynamics (MD) simulations are the basis for high-throughput calculations. Numerous computational systems with HTCS have been established to study the gas separation properties of COFs. Feng et al.²⁴ examined the performance of 688 COFs that were synthesized experimentally for He/CH_4 and He/N_2 membrane separation. The evaluation was conducted using both infinitely diluted and real gas mixtures, leading to the identification of the top five COF materials with high-performance membrane separation properties. Regarding the membrane separation of acid gas from natural gas, the main research focuses on MOFs materials. Qiao et al.²⁵ established the conformational relationship between structural descriptors and $\text{CO}_2/\text{N}_2/\text{CH}_4$ membrane separation performance by molecular simulation of 4764 CoRE MOFs, and identified the membrane screening pathway for high-performance MOFs through decision tree modeling. Similarly, Glover et al.²⁶ calculated the $\text{H}_2\text{S}/\text{CH}_4$ and CO_2/CH_4 membrane separation performance of 7909 MOFs under infinite dilution conditions and reported the hydrophobic structures of specific MOFs, which provided an effective screening method for novel MOFs for biogas upgrading.

High-throughput calculations based on MS are costly and inefficient for screening materials in large databases. As an alternative, machine learning (ML) is increasingly popular for understanding complex structure–property relationships, predicting material properties, and expediting material discovery.²⁷ Many researchers are now utilizing ML and molecular simulation (MS) to assess the performance of new materials.^{28–30} This approach serves as a valuable tool to guide experimental efforts in discovering promising COF materials. In the realm of material discovery, the interpretability of ML models holds significant importance. It enables us to establish trust in the model, analyze the impact of features on predicted results, and gain valuable chemical insights. Currently, interpretability for material property prediction primarily relies on ex post facto interpretability. For instance, Daglar et al.³¹

achieves model interpretability by computing internal feature importance through a tree-based machine learning model, while Yang et al.³² calculate fingerprint feature importance for polymer gas separation membranes by using Shapley additive explanations(SHAP). Drawing insights from interpreting ML, greater attention should be devoted to these crucial factors when designing new membrane materials.

Despite many material screening efforts, the screening of COF membranes for acid gas separation has not been systematically investigated. To fill this gap, the present work is based on interpretable machine learning and molecular simulation to study the latest 811 CURATED-COFs as well as 69840 hypothetical COFs (hCOFs) for acid gas membrane separation. First, 811 CURATED-COFs were prescreened based on structural knowledge. Second, molecular simulation calculations were performed on the prescreened CURATED-COFs and structure–property relationship analyses were carried out on this basis. Then, rapid screening of hCOFs was performed by the feature knowledge of top COFs obtained from SHAP analysis with machine learning. Finally, the obtained high-performance COFs were again subjected to rigorous ternary mixture molecular simulations to evaluate their gas permeability and selectivity and the performance of MMMs obtained in combination with six polymers was investigated.

2. METHODS

This section describes the proposed workflow combining interpretable machine learning and molecular simulation for COF membrane discovery for acid gas separation in natural gas, as shown in Figure 1.

2.1. COFs Data Sets and Polymers. In this work, the COF data were obtained from the latest Clean, Uniform and Refined with Automatic Tracking from Experimental Database (CURATED) COFs (containing 811 COFs)³³ and the computer virtual synthesis of the hCOFs database (containing 69,840 COFs).³⁴ Notably, considering that three gases (H_2S , CO_2 , and CH_4) are involved in the gas separation system, we focus on COF materials with pore size limiting diameters $\text{PLD} > 3.8 \text{ \AA}$ and accessible specific surface area $\text{SA} > 0 \text{ m}^2/\text{g}$ to allow all the gas molecules to pass through the pores of the

Table 1. Calculation Formula for COF-Based Membrane^a

Condition	Property	Formula
Single component gas	Permeability (Barrer)	$P_i = \frac{\phi \times N_i \times D_i \times \rho_i}{f_i}$ $P_i = N_i \times D_i \times \rho_i f_i$
	Membrane selectivity (-)	$S_{i/j} = P_i / P_j$ $S_{i/j} = P_i / P_j$
Mixture gas	Permeability (Barrer)	$P_i^{Mix} = \frac{\phi \times N_i \times D_i \times \rho_i}{f_i^{Mix}}$ $P_i^{Mix} = N_i \times D_i \times \rho_i f_i^{Mix}$
	Membrane selectivity (-)	$S_{i/j}^{Mix} = P_i^{Mix} / P_j^{Mix}$ $S_{i/j}^{Mix} = P_i^{Mix} / P_j^{Mix}$
MMMs	Permeability (Barrer)	$P^{MMM} = P^P \times \left[\frac{2 \times (1 - \varphi) + (1 + 2\varphi) \times (P^{COF}/P^P)}{(1 + \varphi) + (1 - \varphi) \times (P^{COF}/P^P)} \right]$ $P^{MMM} = P^P \times [2 \times (1 - \varphi) + (1 + 2\varphi) \times (P^{COF}/P^P)(1 + \varphi) + (1 - \varphi) \times (P^{COF}/P^P)]$

^a*i* and *j*: gas species (H_2S , CO_2 or CH_4); N_i : gas uptake (mol/kg); D_i : self-diffusivity (m^2/s); ϕ : the void fraction of COF; ρ_i : the density of COF (g/cm^3); f_i : partial fugacity of gas (Pa); P^{MMM} : gas permeability of COF/polymer MMM (Barrer); P^P : gas permeability of polymer (Barrer); P^{COF} : gas permeability of COF membranes (Barrer); φ : volume fraction of COF filler in MMM (≤ 0.2); 1 Barrer = 3.348×10^{-16} momol · m/($m^2 \cdot s \cdot Pa$).

materials. Finally, 788 CURATED-COFs and 69,654 hCOFs were used for further screening.

Based on the gas permeability, the experiment gas permeation test results of polymeric membranes were collected from literature, considering the representative polymeric membrane materials with separation performances located low (e.g., cellulose acetate), medium (e.g., Polyimide) and high region (e.g., PIM). We selected six typical polymers from the literature that have been widely used in the preparation of H_2S and CO_2 separation membranes and whose gas separation data are listed in Table S1.

2.2. Molecular Simulation and Calculation of Membrane. Next, GCMC and MD simulations were carried out to simulate the adsorption and diffusion behavior of gas molecules in COF-based membranes. First, for the start of the screening, the adsorption coefficients (N_i) and self-diffusion coefficients (D_i) were calculated for single component gases at 298 K and 10 bar for H_2S , CO_2 and CH_4 , respectively. All simulations were performed using RASPA2.0.³⁵ CH_4 was modeled as a single spherical nonpolar atom.³⁶ CO_2 was modeled as a three-point rigid model.³⁷ H_2S was modeled as a four-site model where the LJ charge sites are located on a pseudoatom, two hydrogen atoms, and a pseudoatom.³⁸ COF-gas and gas-gas interactions were performed using the Lennard-Jones (LJ) potential description. The Dreiding force field is used to describe the elemental composition of COFs,³⁹ and the Universal Force Field (UFF) is used to describe other metallic elements,⁴⁰ and the work of Wang et al.⁴¹ have demonstrated that the “UFF_Dreiding” hybrid force field is in good agreement with experiment. Due to the presence of polar gas molecules, charge equilibrium (Qeq) calculations are used to quickly compute the partial charges of the atomic framework of COFs.⁴² The Lorentz–Berthelot mixing rules were implemented to obtain pair potentials between different atoms and Ewald summation is used to compute long-range electrostatic interactions.⁴³ Tables S2 and S3 provide details of the atomic parameters for the modeling of the gas molecules as well as the force field

simulation parameters, respectively. In GCMC simulations, we used 1×10^4 cycles for initialization and another 1×10^4 cycles for taking the ensemble averages. The number of GCMC cycles used was also tested to ensure simulation accuracy. Figure S1 shows that the COFs with different gas adsorption levels remain stable in the simulation with different number of cycles, thus demonstrating that further increasing the number of cycles has little effect on the simulation results.

The end-state results of the GCMC simulation were used as the initial state for the MD simulation after 1×10^3 cycles of initialization and 1×10^4 cycles for the equilibration of each COF. Considering computational efficiency, the mean-square displacements (MSD) of the single-component gas in the NVT system were obtained using 1×10^6 cycles at a time step of 1 fs. Subsequently, we performed molecular simulations of ternary gas ($H_2S/CO_2/CH_4 = 20/20/60$) mixtures for the screened high-performance COFs at 308 K and 14 bar. The simulation conditions are consistent with those we have collected from the literature for polymer gas tests (shown in Table S1).

Daglar et al. considered flexible structures and found that there was little effect on the calculation results compared to rigidity.⁴⁴ So rigid structure is used for all the COF structures in this calculation. The cutoff is set to 13 Å, and the number of crystal cells is automatically generated which means that the size of the crystal cell along the x, y and z dimensions should at least exceed twice the cutoff.

According to the solution-diffusion model,⁴⁵ permeability and membrane selectivity can be determined by gas adsorption and gas diffusivity. Detailed calculation formulas are listed in Table 1, including the formulas for single-component, multicomponent gases and MMMs. Here, N_i , D_i , and f_i are the uptake, self-diffusivity of gas *i*, and feed side pressure of the membrane, respectively, and the permeate was assumed to be at vacuum.⁴⁶ For the calculation of MMMs, Maxwell model is the most widely used permeation model, especially under the condition of low packing coincidence (volume fraction ≤ 0.2).

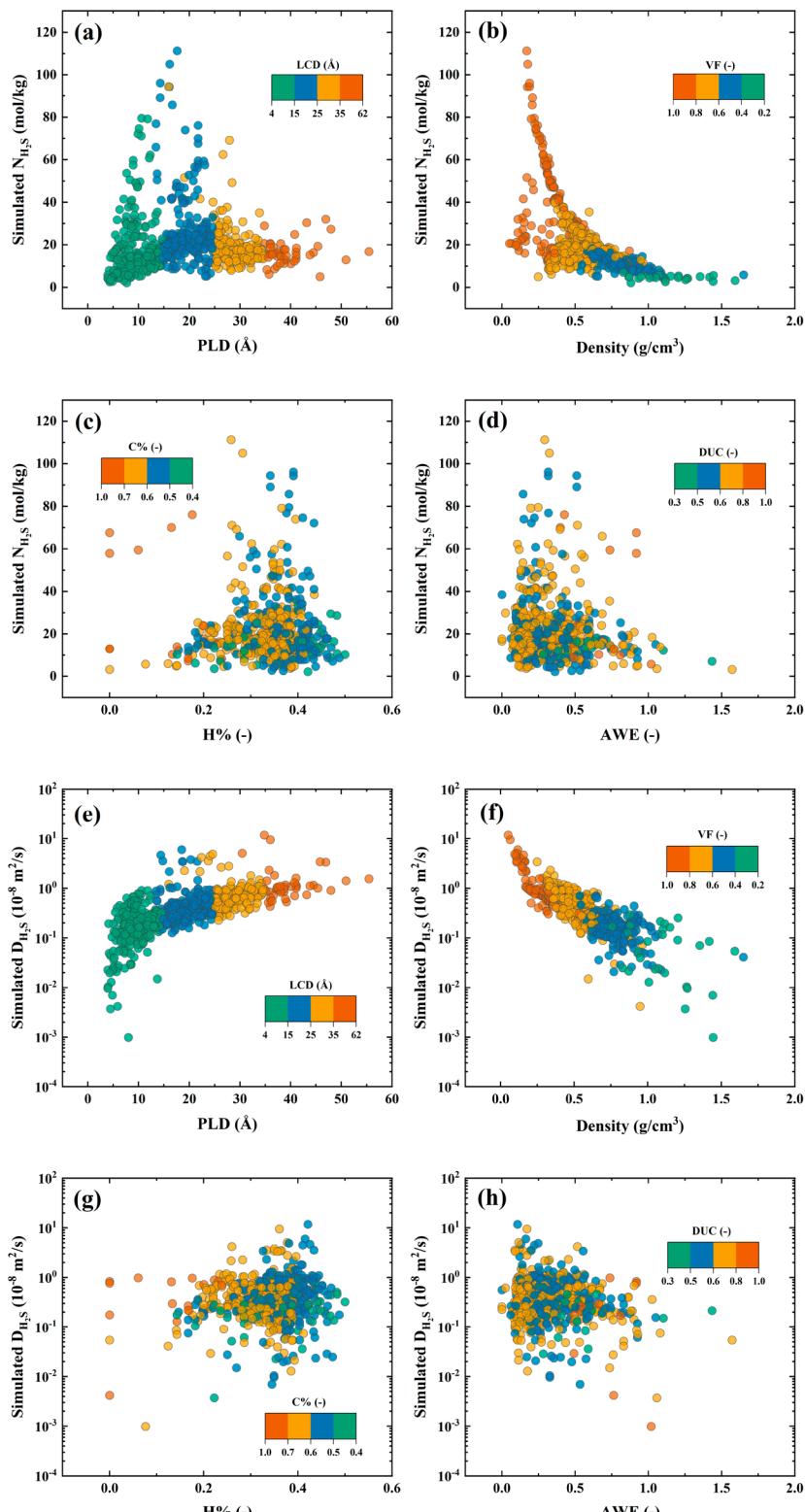


Figure 2. Bivariate analysis of the effect of features on H_2S adsorption (a, b, c, d) and diffusion (e, f, g, h) in 773 COFs. (a, e) Pore size (PLD and LCD); (b, f) Structure (Density and VF); (c, g) Atomic (H% and C%); (d, h) Chemical (AWE and DUC).

We used a filler loading (φ) of 0.2 to maximize MMMs performance while maintaining Maxwell model's range.⁴⁷

2.3. Feature Analysis of COFs. In this study, four types of COF descriptors that can be quickly computed are developed. For the pore size and structure descriptors, LCD, PLD, ρ , SA, and VF were calculated by the open-source software Zeo+

+0.3,⁴⁸ where SA and PV were calculated considering a spherical molecular probe with a radius of 1.86 \AA (corresponding to the N_2 kinetic diameter) and 0 \AA , respectively. The VF was estimated using He with a radius of 1.32 \AA as a probe⁴⁹ using the RASPA 2.0. The Dimensions aims to distinguish between 2D/3D structural COFs as a

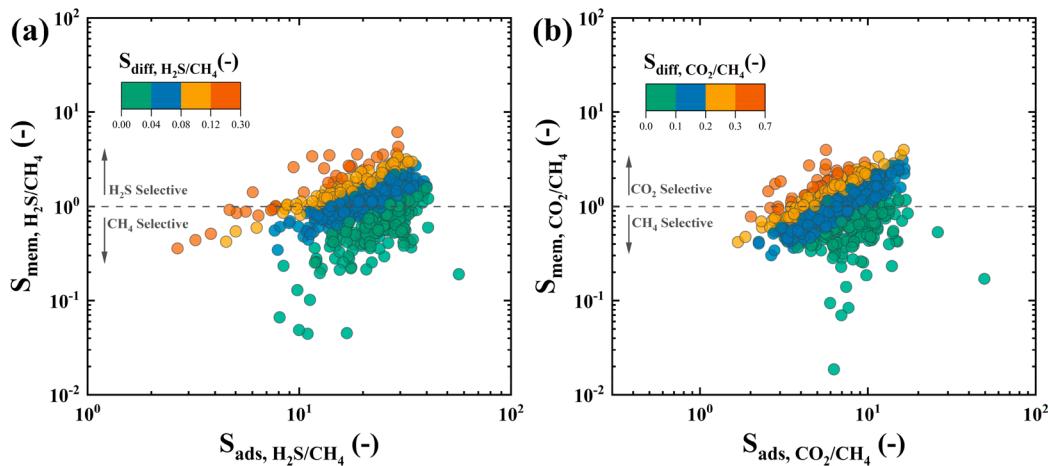


Figure 3. Scattered distribution of adsorption, diffusion, and membrane selectivity.

dichotomous feature. In addition, atomic descriptors and chemical descriptors have been added to further refine the feature space of the COFs. Specifying the number of elements divided by the total number of atoms in the COF to get C%, H%, Halogen%, nitrogen–oxygen pairs (N-to-O), and so on to get atomic descriptors. Since nitrogen-containing functional groups are known to be part of the organic linkers in NPMs, and oxygen atoms are mostly doped into the metal clusters, the ratio of nitrogen atoms to oxygen atoms reflects the distribution of adsorption sites between linkers and clusters.⁵⁰ We introduced this concept into the COF descriptors, where the nitrogen–oxygen ratio of the COF without oxygen atoms is labeled as 0. The chemical bond unsaturation (TDU, DUC) and electronegativity (TE, AWE) of the compounds were calculated to get the chemical property descriptors and the detailed calculation formula is as follows.⁵¹ The descriptions and data distribution of all 20 classes of features are shown in Table S4.

Highly correlated features may lead to problems such as overfitting, poor model generalization, and low prediction accuracy.⁵² Pearson's correlation coefficient (r) is used in mathematics to measure the linear correlation between two sets of variables X and Y. It takes a value between -1 and 1 . The larger the absolute value means the stronger the linear correlation between the two variables. We calculated the r value between each feature of the two databases (CURATED-COFs and hCOFs) respectively. And select the features in the two databases that have high correlation in common, in which LCD/PLD and TE/AWE have strong correlation, respectively 0.98 and 0.99, so choose to delete LCD and TE descriptors, and the rest of the features correlation are lower than 0.95. The heatmap of feature correlation is shown in Figure S2.

2.4. Automated Machine Learning and Interpretable Analysis. The membrane properties obtained from molecular simulation are used as the prediction target, and four sets of descriptors are used as input features for machine learning model construction. The automated machine learning tool we used is the Tree-based Pipeline Optimization Tool (TPOT), which is an automated machine learning tool that integrates feature processing, model selection, and hyperparameter optimization.⁵³ The feature processing and model selection are performed by calling the scikit-learn toolkit for pipeline optimization and genetic algorithm for optimal prediction. The data is divided into 80% training set and 20% test set, and

stratified sampling is followed to ensure that the feature distribution of the training set and test set is consistent, to prevent overfitting we use a 5-fold cross-validation to avoid overfitting. Four metrics, coefficient of determination (R^2), root-mean-square error (RMSE), mean absolute error (MAE), and Spearman's rank correlation coefficient (SRCC) are used to evaluate the model's effectiveness. Detailed formulas are given in Table S5.

Interpretability is currently an important development in machine learning, which avoids black-box models leading to ambiguous guidance for materials science. In this work, we also investigate feature interpretability based on ML models. We use the SHAP interpretability tool. This is an additive feature explanation model inspired by cooperative game theory, where all features are considered as “contributors” to the predicted value, resulting in SHAP value.⁵⁴ SHAP value can reflect the influence of features in each sample and show the positivity and negativity of the influence. The best machine learning algorithm given by TPOT in this research work is different for different prediction tasks. We use the corresponding SHAP interpreter for each ML model, mainly the tree-based explainer and the linear-based explainer.

3. RESULTS AND DISCUSSION

3.1. Structure-Performance Analysis and Membrane Properties. First, structure-performance relationships between four classes of descriptors and simulated data on gas adsorption and diffusion were established. Two features from each class of descriptors were selected, namely LCD and PLD for the pore size descriptors, ρ and VF for the structural descriptors, H% and C% for the atomic descriptors, AWE and DUC for the chemical property descriptors. Figure 2 illustrates the relevance of these features for the adsorption and diffusion of H_2S . From Figure 2a, it can be seen that COF with pore size between 10 and 25 Å can adsorb H_2S more, however, Figure 2e shows that larger pore size leads to higher self-diffusion coefficients; The low density ($0.10\text{--}0.25\text{ g/cm}^3$) of COF is often accompanied by high void fraction ($0.8\text{--}0.95$), which also implies high adsorption of H_2S (Figure 2b), and a similar pattern exists in diffusion of H_2S (Figure 2f); For atomic descriptors and chemical property descriptors, it can be seen from Figure 2 that high adsorption and diffusion are distributed in COFs with medium H% ($0.3\text{--}0.4$), medium C% ($0.5\text{--}0.7$), low AWE ($0\text{--}0.5$), and medium DUC ($0.5\text{--}0.8$).

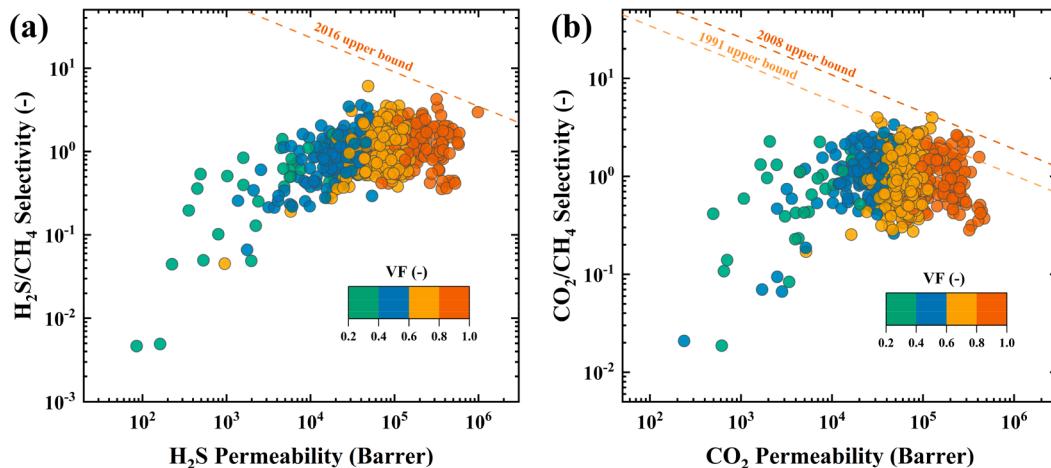


Figure 4. H₂S permeability and H₂S/CH₄ selectivity (a), CO₂ permeability and CO₂/CH₄ selectivity (b) of COF membranes at 10 bar, 298 K. The VF indicates void fraction.

Figures S3 and S4 also show the structure–property relationship between the adsorption and diffusion of CO₂ and CH₄ with the four types of descriptors, respectively. Taken together, a similar pattern can be obtained: for gas adsorption, COFs with PLDs between 5 and 25 Å have higher H₂S and CO₂ adsorption, but for CH₄ adsorption there is a small positive correlation with pore size (Figure S4a). The effect of structural descriptors on adsorption is more pronounced, with low density accompanied by high void fraction determining higher gas adsorption. For gas diffusion, the pore sizes of COFs all showed positive correlations, and similarly, the densities showed negative correlations. Moreover, neither the atomic descriptors nor the chemical property descriptors showed significant conformational correlations. Considering that the time-scale-based MD simulation is more complex and time-consuming than GCMC, the correlation of the above feature relationships is slightly less obvious compared to the adsorption amount.

After calculating the gas adsorption and diffusion at 10 bar and 298 K, the distribution of adsorption selectivity, self-diffusion selectivity, and membrane selectivity of COFs can be obtained as shown in Figure 3. Based on the selectivity calculation equation, high membrane selectivity is obtained by the combination of high adsorption selectivity and high diffusion selectivity. However, the trend of the two tends to be opposite, which is caused by the incompatibility between the adsorption and diffusion movement of gas molecules, i.e., the molecules that can be easily adsorbed diffuse more slowly. Therefore, the highly selective COFs have $S_{ads,H_2S/CH_4}$ ($S_{ads,CO_2/CH_4}$) in the range of 9.7–39.8 (4.1–16.8), and $S_{diff,H_2S/CH_4}$ ($S_{diff,CO_2/CH_4}$) in the range of 0.04–0.21 (0.1–0.7). In addition to this, Figure S5 compares the H₂S/CH₄ selectivity of COF membranes with CO₂/CH₄ selectivity. In real natural gas where acid gases are often present at the same time, we should pay equal attention to COFs with both high H₂S selectivity and CO₂ selectivity. COFs with H₂S permeability greater than 2.5×10^5 and CO₂ permeability greater than 1.8×10^5 were selected to have high acid gas permeability. For them, the H₂S/CH₄ selectivity is distributed in the range of 0.36–4.26 and the CO₂/CH₄ selectivity is distributed in the range of 0.28–2.61. It can be found that the H₂S selectivity is higher in the acid gas regime, which is also

consistent with the findings of the subsequent mixture simulations.

To better understand the performance of COF membranes. Figure 4 shows the H₂S permeability and H₂S/CH₄ selectivity as well as CO₂ permeability and CO₂/CH₄ selectivity of the COF membranes. Similar to polymers, COF membranes also have permeability-selectivity trade-offs, and only a few COFs can approach the upper bound curve. A gradient classification of COFs based on void fraction clearly shows the effect of the high void fraction on the performance of the membrane. This is consistent with the findings of our structure–performance relationship analysis described above and is further verified in subsequent interpretable machine learning.

3.2. ML Model Prediction and Interpretability Analysis.

Based on the 18 fast computational features developed above, we developed corresponding machine learning models for each type of task to predict gas adsorption, diffusion, and permeability. After removing the redundant features of LCD and TE that possess high correlation properties. All other features can be used as input variables. It is worth noting that TPOT has the capability of feature processing. Therefore, there is no need to normalize the features, and the only thing we need to do is to construct a “feature-attribute” data structure. In this summary, LassoLars CV, Extra Trees Regressor, Gradient Boosting Regressor, XGB Regressor, Ridge CV, and Random Forest Regressor are used for the regression task of COF membrane performance prediction. Among them, Lasso,⁵⁵ Random Forest,⁵⁶ and Extreme Gradient Boosting Tree⁵⁷ are used to train ML models for MOFs. Based on the number of iterations of the genetic algorithm and the size of the population evolution, different feature processing methods and model pipelines may be found. Table S6. shows the best ML pipeline and hyperparameters based on gas adsorption, diffusion, and permeability properties.

By directly inputting all features, highly accurate prediction results can be achieved. Figure S6 shows that the gas adsorption and diffusion predicted by the ML model compared to the result of molecular simulation. The accuracy of gas adsorption prediction is notably high, particularly for CH₄ adsorption, where the R² values for the training and test sets are 0.999 and 0.985, respectively, and the SRCC values are 0.984 and 0.981, respectively. The slightly lower accuracy in

Table 2. Best ML Model and Evaluation Indicators^a

Target	Training set					Test set		
	R ²	RMSE	MAE	SRCC	R ²	RMSE	MAE	SRCC
N_{H_2S}	0.999	0.274	0.099	0.999	0.904	4.342	2.203	0.900
D_{H_2S}	0.892	0.244	0.139	0.879	0.914	0.287	0.153	0.852
P_{H_2S}	0.986	1.165	0.440	0.997	0.942	2.240	1.481	0.952
N_{CO_2}	0.948	0.812	0.235	0.999	0.852	1.437	0.879	0.895
D_{CO_2}	0.965	0.289	0.213	0.941	0.873	0.352	0.769	0.907
P_{CO_2}	0.994	0.511	0.383	0.987	0.928	1.770	1.163	0.901
N_{CH_4}	0.993	0.059	0.040	0.984	0.985	0.065	0.045	0.981
D_{CH_4}	0.999	0.342	0.202	0.999	0.767	3.982	2.792	0.750
P_{CH_4}	0.999	0.420	0.215	0.999	0.856	4.366	2.885	0.837

^aThe unit of gas adsorption (N_i), diffusion (D_i), and permeability (P_i) are mol/kg, 10^{-8} m²/s, and 10^4 Barrer, respectively.

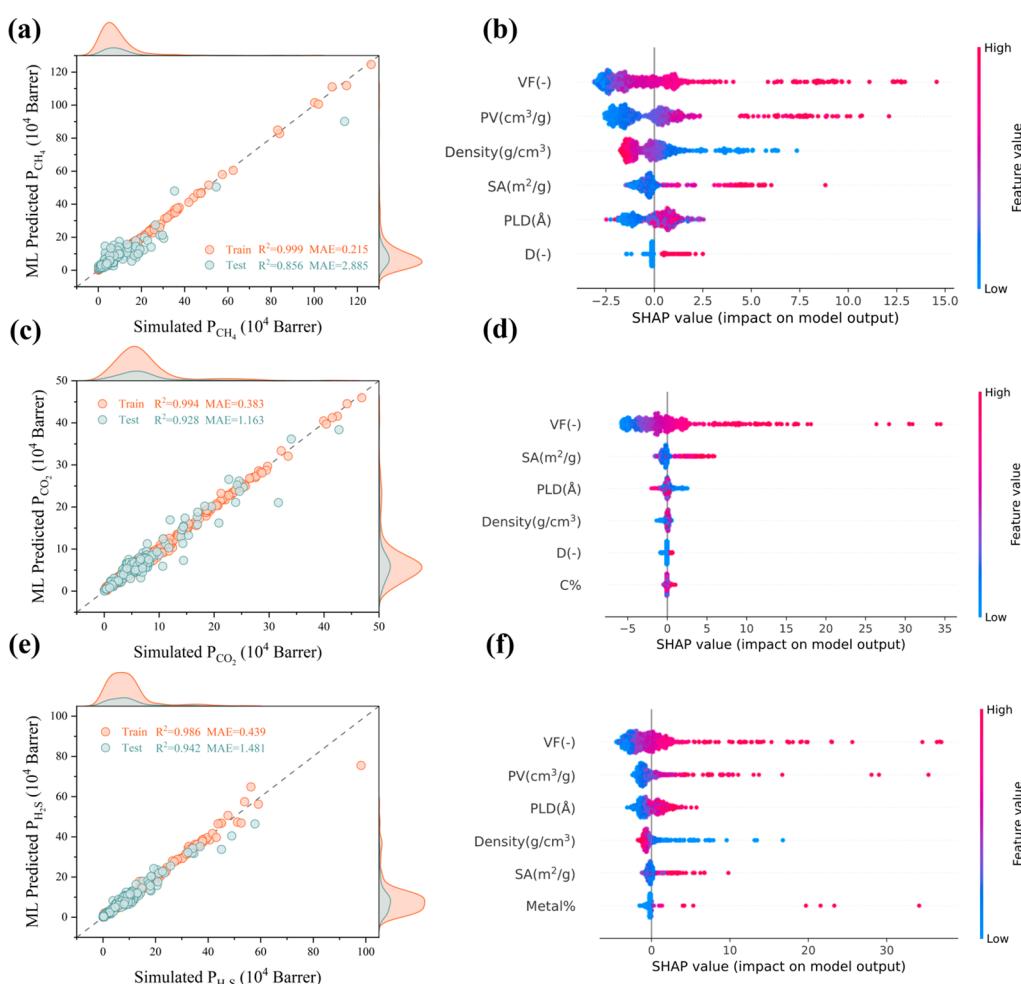


Figure 5. Marginal distribution scatter plots of gas (a, c, e) permeability data predicted by the ML model compared with molecular simulation results and SHAP feature importance distribution (b, d, f) of the corresponding ML model. Features are listed in top-down descending order based on SHAP values, and only the top six important features are shown in the figure.

predicting CO₂ and H₂S adsorption can be attributed to the greater complexity of the multiatom modeling compared to the spherical modeling used for CH₄, which makes the mathematical relationship between CH₄ adsorption properties and features more apparent. In contrast to GCMC, MD modeling is more intricate. Before this study, ML modeling for predicting COF diffusivity had not been extensively investigated. Seda et al.³¹ examined diffusion ML modeling based on MOF membranes and reported similar findings, where the prediction accuracy of gas diffusion was relatively low but consistently maintained an R² value above 0.85, except for CH₄ diffusion with an R² of 0.77, which still demonstrates a high level of prediction accuracy.

Considering that the gas permeability is calculated from the adsorption and diffusion data, the ML model, which is directly used to predict the gas permeability calculated from the simulated data, was constructed. Table 2 presents the model evaluation metrics, namely R², RMSE, MAE, and SRCC, along with the best ML model. The tree model is used more as the best model for gas adsorption and permeability prediction and the CV model is used more for diffusion prediction. Figure 5a, 5c, and 5e show the gas permeability obtained from ML prediction compared with the data obtained from molecular simulation calculations, respectively. It can be seen that the two are in good agreement, and the direct construction of the gas permeability prediction model is a more direct and efficient prediction method. The R² of the ML model for H₂S permeability reaches 0.942, and the SRCC reaches 0.952, which is more accurate than the prediction of adsorption and diffusion successively and then the calculation of permeability.

The ML models developed to predict gas adsorption, diffusion, and permeability vary in the importance of various types of features. Figure 5b, 5d and 5f lists the top six important features based on the three gas permeability prediction models, with VF being the most contributing feature for all of them and appearing mostly for the pore size and structure descriptors. For specific SHAP feature contribution values, see Figure S7. The contribution of all features was calculated using SHAP and presented in Figure 6 through radar plots. Pore size and structure descriptors

contribute most of the feature influence for almost all models. VF has been found to make the greatest contribution overall. The influence of PLD on adsorption is found to be more significant than that of diffusion. PV, on the other hand, plays a larger role in the diffusion of H₂S and CO₂ compared to VF, accounting for 57% and 55% respectively. SA only exhibits a significant impact on the prediction of CO₂ adsorption, contributing 28% to its overall performance. This further underscores the dependability of the extracted feature knowledge. For atomic descriptors and chemical property descriptors, their contributions are generally low. Only the prediction of CH₄ adsorption and diffusion has a higher contribution than the other tasks. We note that for the three prediction tasks of H₂S diffusion, CO₂ diffusion, and CH₄ adsorption, some of the atomic descriptors and chemical property descriptors have a SHAP value of 0. This is because they use the linear regression models LassoLarCV and RidgeCV, which are based on compression estimation with different regularizations. They compress the unimportant features for high-dimensional data. So, it will get the features with the contribution of 0 SHAP value are calculated by the linear explainer.

It is important to note that similar findings have been observed in experimental studies regarding the influence of porosity and pore size on the effectiveness of NPMs membranes.^{58–61} Compared with the pore size of COF, void fraction and pore volume have more influence on gas permeability. By analyzing the pore size of the top COF base membrane, it can be found that the pore size of the high-performance COF membrane is maintained at 15–25 Å, which indicates that neither too small nor too large pore size is suitable for the formation of high-performance COF-based membranes. From the perspective of membrane analysis, too small pore size will obstruct the gas transport channel and lead to decreased permeability, while too large pore size will limit the gas fraction efficiency and lead to pore blockage in the polymer compatibility process.^{62,63} This directly proves that our data-driven interpretability agrees with physical experimental experience.

In this subsection, highly accurate ML models are constructed for the adsorption, diffusion, and permeability of three gases. For the evaluation of membrane performance, the direct construction of prediction models for gas permeability is an effective method. The feature contribution analysis is performed by the machine learning interpretability tool of SHAP. It was found that VF contributes the most to the model, and not only that, the rest of the pore size and structural descriptors have a significant contribution. In comparison atomic and chemical property descriptors have less impact on the prediction task. In the next section, we will use the obtained feature knowledge for fast screening of the hCOFs database.

3.3. Rapid Screening of the hCOFs Database. Based on molecular simulation calculations to obtain the membrane performance of CURATED-COFs, we first focus on the top COFs with high gas permeability COF membranes. On this basis, COFs with H₂S/CH₄ selectivity >1 and CO₂/CH₄ selectivity >0.7 were selected, resulting in TOP20 COFs. Leveraging the chemical knowledge extracted from the machine learning model discussed in the preceding subsection, we identified key features, namely PLD, Density, SA, PV, and VF, as primary metrics for rapid screening of hCOFs in the subsequent subsection. Additionally, we incorporated the key

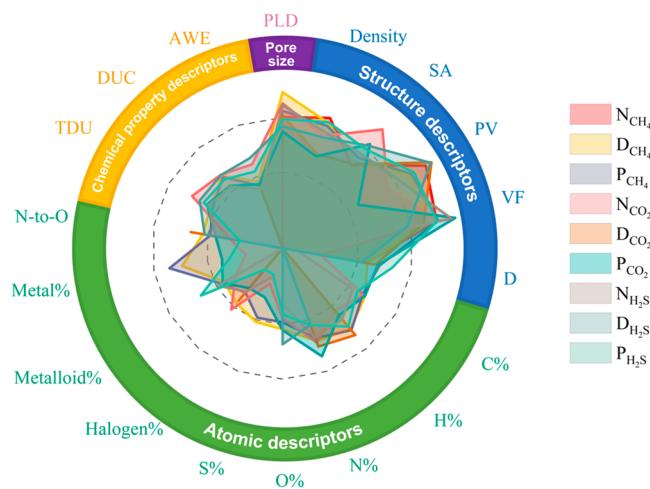


Figure 6. Radar plots of feature importance for the gas adsorption, diffusion, and permeability of COFs. The SHAP values for each type of explainer have been normalized, and the scale range shows a Log distribution from 0.01 to 1.

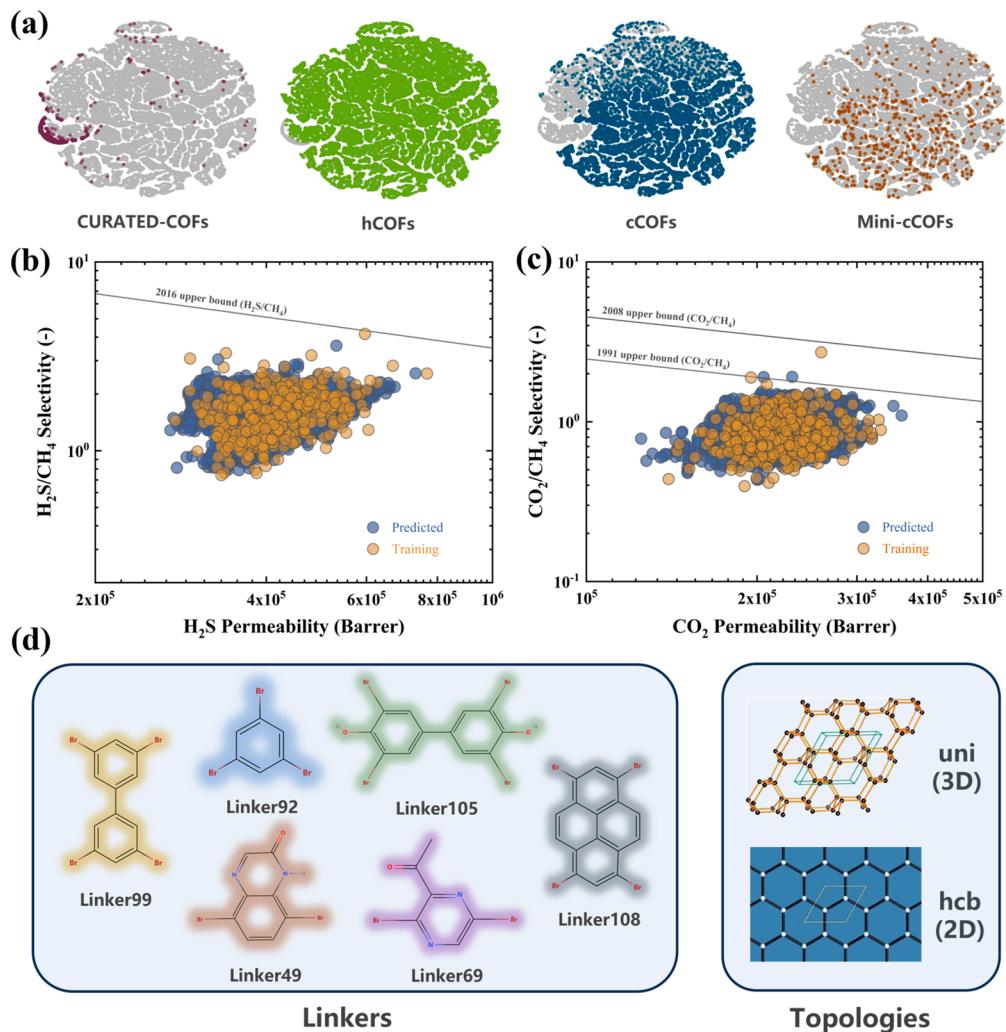


Figure 7. (a) Projections of the feature distribution of the data based on t-SNE. (b and c) H_2S/CH_4 and CO_2/CH_4 permeability-selectivity distributions obtained from the hCOFs database after ML projection. (d) Main linkers and topologies of top hCOFs.

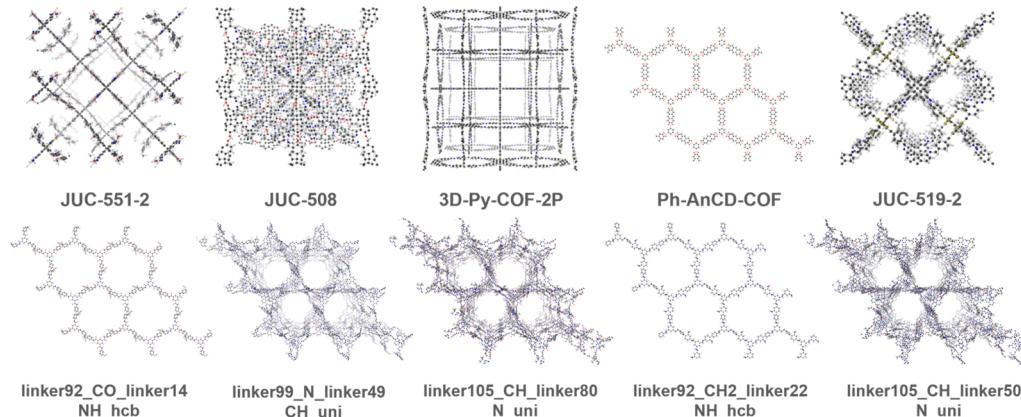
pore size and structure descriptors as prescreening features. To ensure a representative distribution of data, hCOFs were swiftly screened based on specific intervals: PLD (15–25 Å), Density (0.15–0.31 g/cm³), SA (5100–8300 m²/g), PV (2.7–6 cm³/g), and VF (0.88–0.95). The large pore size and large void fraction facilitate the passage of gas molecules, and the low density and high pore volume ensure the framework structure and the adsorption of gas molecules, resulting in highly permeable and selective COFs. This screening process yielded a database of 14,957 potential Candidate hCOFs (cCOFs).

Predicting hCOFs directly using machine learning models obtained by training with CURATED-COFs leads to a decrease in accuracy, considering the feature differences before the database. Here we consider the use of data augmentation to supplement the training data to improve the model's prediction accuracy for hCOFs. The cCOFs were uniformly sampled based on VF to obtain mini-cCOFs containing 500 COFs, which were computed by molecular simulation and merged with the previous 788 CURATED-COFs (around 1300 data points in total) to construct ML models to optimize the model's generalization to the unknown chemical space. To assess the uniformity of the COF feature space, the t-distributed stochastic neighbor embedding (t-SNE) dimensionality reduction method serves as an effective means of visualization. This method has previously been extensively employed in the comparative analysis of multiple MOF databases.⁶⁴ Figure 7a displays the visualization of CURATED-COFs, hCOFs, cCOFs, and mini-cCOFs using t-SNE dimensionality reduction, which projects the high-dimensional feature distributions onto a 2D plane. It is evident that cCOFs predominantly occupy the main chemical space, and the preliminary screening showcases some degree of diversity. Furthermore, the mini-cCOFs are more evenly distributed and representatively sampled. The chemical space obtained by the combination of CURATED-COFs and mini-cCOFs basically covers the entire prediction space, which ensures the prediction accuracy of rapid prediction screening. The color distribution in Figure S8 indicates the gradient of void fraction in cCOFs. The upper right to lower left direction on the projection plane signifies low to high porosity, illustrating a certain linear pattern within the intricate chemical space. This further confirms the rationale behind uniform sampling based on VF.

After training the best machine learning (ML) model for each prediction task (Table S7 for model training results), we utilized ML models to directly predict the gas permeability of tens of thousands of hypothetical COFs in the cCOF. Similar

Table 3. Structural Parameters and Gas Separation Performance of TOP 10 COFs

COFs	PLD (Å)	VF (-)	Di (-)	P_{H_2S} (Barrer)	P_{CO_2} (Barrer)	S_{H_2S/CH_4} (-)	S_{CO_2/CH_4} (-)
JUC-5S1-2	10.1	0.91	3D	4.19×10^5	3.02×10^5	5.46	3.93
JUC-508	8.9	0.87	3D	6.01×10^5	1.85×10^5	5.19	1.60
3D-Py-COF-2P	12.3	0.91	3D	5.31×10^5	2.88×10^5	4.23	2.29
Ph-An _{CD} -COF	21.1	0.91	2D	3.70×10^5	2.30×10^5	4.70	2.92
JUC-519-2	8.9	0.89	3D	3.83×10^5	3.12×10^5	3.54	2.90
linker92_CO_linker14_NH_hcb_relaxed	19.1	0.89	2D	3.28×10^5	2.17×10^5	4.61	3.05
linker99_N_linker49_CH_uni_relaxed	23.4	0.93	3D	4.16×10^5	2.13×10^5	4.21	2.16
linker105_CH_linker80_N_uni_relaxed	22.9	0.90	3D	3.91×10^5	2.45×10^5	3.69	2.31
linker92_CH2_linker22_NH_hcb_relaxed	19.0	0.90	2D	3.90×10^5	2.22×10^5	3.49	1.99
linker105_CH_linker50_N_uni_relaxed	22.4	0.91	3D	3.44×10^5	2.62×10^5	3.13	2.38

**Figure 8.** Structural visualization of top 5 COFs in CURATED-COFs and hCOFs. All snapshots are taken looking down the crystallographic *c*-axis of each structure.

to CURATED-COFs, direct permeability prediction provides more accurate and generalizable observations. The R^2 values for H_2S and CO_2 permeability reached 0.96, while the SRCC for CH_4 permeability reached 0.95, indicating the accuracy of the prediction. All three gas permeability prediction models are tree models (Table S8 shows the optimal pipeline parameters for different tasks), ensuring that the predicted permeability falls within the expected region in the permeability-selectivity space. The predicted permeabilities for the ML model are illustrated in Figure 7b and c for H_2S/CH_4 and CO_2/CH_4 separations. Compared to CURATED-COFs, we found highly permeable COFs materials beyond the original chemical space. Thirty of the predicted cCOFs have H_2S permeabilities exceeding 6×10^5 Barrer and up to 7.3×10^5 Barrer. In terms of selectivity, the highest H_2S/CH_4 and CO_2/CH_4 selectivities are 3.6 and 1.9, respectively, which are similar to the distribution of the top CURATED-COFs. Although most of the cCOFs were in the lower left part of the upper bound, we still found a few high-performance COF membranes close to or even beyond the upper bound.

Considering the permeability-selectivity trade-off, we examined the structural characteristics of the top cCOFs. Most of the top COFs possess a 3D structure, with most of their PLDs centered around 20 Å and exhibiting high void fractions (>0.9). Additionally, Figure 7d showcases the primary linkers and topologies employed in their composition. Among these, linker92 (1,3,5-tribromobenzene), linker99 (3,3,5-tetrabromo-1,1-biphenyl), linker105 (3,3',5,5'-tetrabromo-4,4'-biphenyldiol), and linker108 (1,3,6,8-tetrabromopyrene) are the most frequently occurring linkers. Linker49 (5,8-dibromoquinoxalin-2(1H)-one) and linker69 (1-(3,6-dibromopyrazin-2-

yl)ethan-1-one), on the other hand, are linkers that have each appeared twice. In terms of topology, uni is the dominant structure. Additionally, dmp, jea, and qtz also appear among them. A small number of top-level 2D hCOFs exhibit the hcb topology. The results show that based on the feature discovery of the top CURATED-COFs, our proposed ML method based on porosity gradient sampling accurately discovers promising COF-based membranes in hCOFs.

3.4. TOP COFs and COF/Polymer MMMs. Further, we carefully selected a total of 40 top-COFs, with 20 each from CURATED-COFs and hCOFs. The obtained membrane properties from the simulations are presented in Figure 9a, where the acid gas permeability is determined by the combined permeability of H_2S and CO_2 , and the acid gas selectivity is obtained by the sum of the H_2S/CH_4 selectivity and CO_2/CH_4 selectivity. It is important for a membrane to have both high gas permeability and selectivity in order to be considered high-performance. By evaluating the product of selectivity and permeability, we identified the top five optimal COFs for both CURATED-COFs and hCOFs, which are indicated with asterisks. Detailed structural information and membrane properties of these COFs can be found in Table 3, respectively. In addition to this, Figure 8 visualizes the structure of these top COFs (taken looking down the crystallographic *c*-axis). Most of these COFs exhibited 3D structures, such as JUC-5S1-2,⁶⁵ JUC-508,⁶⁶ and 3D-Py-COF-2P,⁶⁷ and were characterized by large pore sizes (>10 Å) and high void fractions (>0.9). Notably, these COFs surpass the Robeson upper bound and demonstrate exceptional performance in acid gas separation. However, hCOFs have not yet been experimentally synthesized and verified. None-

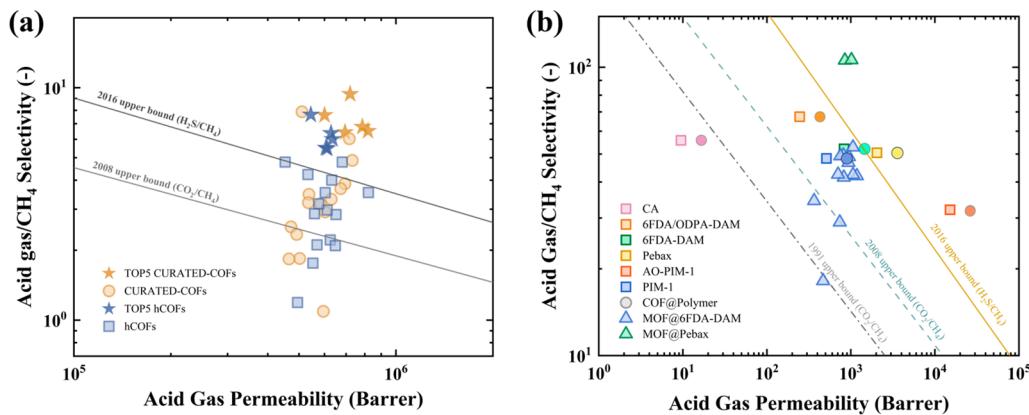


Figure 9. (a) Molecular simulation results of top 40 COFs under ternary mixture conditions at 14 bar, 308 K, where acid gas permeability represents the sum of H₂S permeability and CO₂ permeability. (b) Acid gas separation performance of MMMs.

theless, the structural characteristics of highly permeable and selective COF-based membranes can provide computational and experimental membrane scientists with a reference perspective for in-depth study of acid gas separation membrane materials.

The incorporation of COFs into polymers to create MMMs is an effective strategy for enhancing industrial applications and improving the performance of polymer membranes compared to pure COF membranes. In addition to the advantages similar to other porous materials, COF has better miscibility affinity with polymers due to its pure organic nature.⁶⁸ Therefore, the incorporation of COFs as fillers into polymer substrates provides higher stability and compatibility. We selected six representative polymer membranes for acid gas separation from the literature and maintained consistent simulation conditions for the ternary mixtures of top COFs as reported in the literature. Before predicting the performance of mm, we first verified the accuracy of the method. As shown in Figure S9. Our simulation data is in good agreement with the data collected by the experiment. Figure 9b presents data on 60 MMMs resulting from the combination of the top 10 COFs with six polymers. The circular data in the figure indicates COF/Polymer MMMs and COF filler loading is 20%. The incorporation of COF-based fillers significantly enhances gas permeability without compromising polymer membrane selectivity. Furthermore, comparison with MOF/polymer MMMs from the literature (data displayed in Table S9) revealed that COF-based fillers outperform MOFs in terms of permeability enhancement. These findings suggest that the improved acid gas separation performance achieved by adding COF-based fillers to the polymer matrix primarily stems from increased permeability, with the selectivity of the polymer and the permeability of the COFs predominantly influencing the selectivity and permeability of the resulting MMMs, respectively. However, COF can constitute up to 50 wt % of the filler weight in polymers due to its organic nature.⁵⁸ The potential of improving the separation performance of polymer membranes is greater than that of MOFs. Therefore, researchers should prioritize COFs with high permeability when selecting MMMs for acid gas membrane separation.

4. CONCLUSIONS

In this study, a new method combining interpretable ML and MS was proposed to evaluate the separation performance of COF based membranes for H₂S/CH₄ and CO₂/CH₄. First, the

molecular simulation of 788 COFs in CURATED-COFs database was performed, and the structure and chemical properties of the best COFs were qualitatively analyzed. Then, by constructing automated machine learning models, most models predict that R² is greater than 0.9. Moreover, directly building a model for gas permeability is a more effective prediction method, and the R² for H₂S gas permeability prediction reaches 0.94, which is higher than the adsorption capacity and diffusion coefficient respectively predicted. After that, SHAP analysis was used to calculate the characteristic contribution degree of all the features, and the results showed that the porosity contributed the most to the membrane gas permeability. Next, we quickly screened nearly 70,000 hCOFs using five important characteristics of high-performance CURATED-COFs. Considering database differences, 500 representative hCOFs were further enhanced by molecular simulation to train a new machine learning model to predict the remaining hCOFs quickly. Finally, the mixture molecules of 40 top-level COF were simulated. The exploration of COF/polymer MMMs shows that high performance COFs can make up for the lack of polymer permeability and effectively improve the separation performance of acid gas. Our results not only reveal the membrane separation potential of the vast chemical space in the COFs database, but also help accelerate the design and discovery of membrane materials in experimental work. It provides important theoretical guidance for the development of COFs-based membranes in the field of simultaneous separation of H₂S and CO₂ from natural gas. At the same time, it is also very important to experimentally synthesize usable high-performance COF-based membranes, which can be a subject of in-depth study by computational chemists and membrane scientists.

■ ASSOCIATED CONTENT

Data Availability Statement

COF molecular simulation computational data, automated machine learning models, and interpretability analysis can be viewed at <https://github.com/Xinbingru/COFsMembraneML>.

SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.iecr.4c00855>.

Permeability of polymer; Gas molecular modeling; Force field parameters; COF descriptor; Comparison of simulation results of GCMC cycle times; ML evaluation

index; Characteristic heat map; Bivariate analysis of CO₂ adsorption and diffusion; CH₄ adsorption and diffusion bivariate analysis; Acid gas selectivity analysis; Prediction results of adsorption and diffusion ML; CURATED-COFs best machine learning pipeline; Global feature contribution degree; t-SNE projection based on VF; ML prediction results of cCOFs; Best ML pipeline for cCOFs; Comparison of MMMs simulation and experimental data; MMMs data collected from the experimental literature ([PDF](#))

AUTHOR INFORMATION

Corresponding Authors

Yiyang Dai — School of Chemical Engineering, Sichuan University, Chengdu 610065, China;  [orcid.org/0000-0001-6000-9160](#); Email: daiyy@scu.edu.cn

Xu Ji — School of Chemical Engineering, Sichuan University, Chengdu 610065, China;  [orcid.org/0000-0002-1005-7401](#); Email: jixu@scu.edu.cn

Authors

Bingru Xin — School of Chemical Engineering, Sichuan University, Chengdu 610065, China

Minggao Feng — Department of Chemistry and Materials Innovation Factory, University of Liverpool, Liverpool L7 3NY, U.K.

Min Cheng — BYD Company Ltd., Shenzhen, Guangdong 518122, China

Zhongde Dai — School of Carbon Neutrality Future Technology, Sichuan University, Chengdu 610065, China;  [orcid.org/0000-0002-3558-5403](#)

Shiyang Ye — School of Chemical Engineering, Sichuan University, Chengdu 610065, China

Li Zhou — School of Chemical Engineering, Sichuan University, Chengdu 610065, China;  [orcid.org/0000-0003-3539-7573](#)

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acs.iecr.4c00855>

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The financial support provided by the Young Scientists Fund of the National Natural Science Foundation of China (No. 52170112) and the National Key Research and Development Program of China (No. 2023YFF0720102) is gratefully acknowledged.

ABBREVIATIONS

AWE, Atomic weighted electronegativity; COFs, Covalent organic frameworks; cCOFs, Candidate hCOFs; CoRE COFs, Computation-Ready Experimental COFs; CURATED-COFs, Clean, Uniform and Refined with Automatic Tracking from Experimental Database COFs; DUC, Degree of unsaturation per carbon; GCMC, Grand canonical Monte Carlo; GO, Graphene oxide; hCOFs, hypothetical COFs; HTCS, High-throughput computational screening; LCD, Largest cavity diameter; LJ, Lennard-Jones; MAE, Absolute percentage error; MD, Molecular dynamics; ML, Machine learning; MMMs, Mixed-matrix membranes; MOFs, Metal–organic frameworks; MS, Molecular simulation; NPMs, Nano porous materials;

PNNs, Porous polymer networks; PLD, Pore limiting diameter; PV, Pore volume; RMSE, Root mean square error; SA, Surface area; SHAP, Shapley additive explanations; SRCC, Spearman's rank correlation coefficient; TPOT, Tree-based Pipeline Optimization Tool; TDU, Total degree of unsaturation; UFF, Universal Force Field; t-SNE, t-distributed stochastic neighbor embedding; VF, Void fraction; ZIFs, Zeolitic imidazolate frameworks

REFERENCES

- (1) Faramawy, S.; Zaki, T.; Sakr, A. A.-E. Natural Gas Origin, Composition, and Processing: A Review. *J. Nat. Gas Sci. Eng.* **2016**, *34*, 34–54.
- (2) Ma, Y.; Guo, H.; Selyanchyn, R.; Wang, B.; Deng, L.; Dai, Z.; Jiang, X. Hydrogen Sulfide Removal from Natural Gas Using Membrane Technology: A Review. *J. Mater. Chem. A* **2021**, *9* (36), 20211–20240.
- (3) Pudi, A.; Rezaei, M.; Signorini, V.; Andersson, M. P.; Baschetti, M. G.; Mansouri, S. S. Hydrogen Sulfide Capture and Removal Technologies: A Comprehensive Review of Recent Developments and Emerging Trends. *Sep. Purif. Technol.* **2022**, *298*, 121448.
- (4) Cui, G.; Wang, J.; Zhang, S. Active Chemisorption Sites in Functionalized Ionic Liquids for Carbon Capture. *Chem. Soc. Rev.* **2016**, *45* (15), 4307–4339.
- (5) Li, C.; Yu, Z.; Cui, K.; Schmidt, J. R.; Sholl, D. S.; Lively, R. P. Assessment of Acid Gas Adsorption Selectivities in MIL-125-NH₂. *J. Phys. Chem. C* **2022**, *126* (50), 21414–21425.
- (6) Xu, Q.; Xin, B.; Wei, J.; Ma, Y.; Qing, Z.; Feng, C.; Yi, S.; Li, N.; Li, K.; Wang, F.; Zhao, J.; Yang, L.; Yao, L.; Jiang, W.; Dai, Y.; Dai, Z. Troger's Base Polymeric Membranes for CO₂ Separation: A Review. *J. Mater. Chem. A* **2023**, *11* (29), 15600–15634.
- (7) Dai, Z.; Deng, L. Membranes for CO₂ Capture and Separation: Progress in Research and Development for Industrial Applications. *Sep. Purif. Technol.* **2024**, *335*, 126022.
- (8) Freeman, B. D. Basis of Permeability/Selectivity Tradeoff Relations in Polymeric Gas Separation Membranes. *Macromolecules* **1999**, *32* (2), 375–380.
- (9) Chen, G.; Liu, G.; Pan, Y.; Liu, G.; Gu, X.; Jin, W.; Xu, N. Zeolites and Metal-Organic Frameworks for Gas Separation: The Possibility of Translating Adsorbents into Membranes. *Chem. Soc. Rev.* **2023**, *52* (14), 4586–4602.
- (10) Qin, Z.; Ma, Y.; Wei, J.; Guo, H.; Wang, B.; Deng, J.; Yi, C.; Li, N.; Yi, S.; Deng, Y.; Du, W.; Shen, J.; Jiang, W.; Yao, L.; Yang, L.; Dai, Z. Recent Progress in Ternary Mixed Matrix Membranes for CO₂ Separation. *Green Energy Environ.* **2024**, *9* (5), 831–858.
- (11) Hou, S.; Hu, J.; Liang, X.; Zhang, D.; Tan, B. Carbonate-Based Hyper-Cross-Linked Polymers with Pendant Versatile Electron-Withdrawing Functional Groups for CO₂ Adsorption and Separation. *J. Mater. Chem. A* **2022**, *10* (28), 15062–15073.
- (12) Wang, Z.; Zhang, S.; Chen, Y.; Zhang, Z.; Ma, S. Covalent Organic Frameworks for Separation Applications. *Chem. Soc. Rev.* **2020**, *49* (3), 708–735.
- (13) Gao, Q.; Li, X.; Ning, G.-H.; Xu, H.-S.; Liu, C.; Tian, B.; Tang, W.; Loh, K. P. Covalent Organic Framework with Frustrated Bonding Network for Enhanced Carbon Dioxide Storage. *Chem. Mater.* **2018**, *30* (5), 1762–1768.
- (14) Fan, H.; Mundstock, A.; Feldhoff, A.; Knebel, A.; Gu, J.; Meng, H.; Caro, J. Covalent Organic Framework-Covalent Organic Framework Bilayer Membranes for Highly Selective Gas Separation. *J. Am. Chem. Soc.* **2018**, *140* (32), 10094–10098.
- (15) Yusran, Y.; Li, H.; Guan, X.; Fang, Q.; Qiu, S. Covalent Organic Frameworks for Catalysis. *EnergyChem.* **2020**, *2* (3), 100035.
- (16) Fan, H.; Wang, H.; Peng, M.; Meng, H.; Mundstock, A.; Knebel, A.; Caro, J. Pore-in-Pore Engineering in a Covalent Organic Framework Membrane for Gas Separation. *ACS Nano* **2023**, *17* (8), 7584–7594.
- (17) Wang, S.; Wei, X.; Li, Z.; Liu, Y.; Wang, H.; Zou, L.; Lu, D.; Hassan Akhtar, F.; Wang, X.; Wu, C.; Luo, S. Recent Advances in

- Developing Mixed Matrix Membranes Based on Covalent Organic Frameworks. *Sep. Purif. Technol.* **2022**, *301*, 122004.
- (18) Cheng, Y.; Ying, Y.; Zhai, L.; Liu, G.; Dong, J.; Wang, Y.; Christopher, M. P.; Long, S.; Wang, Y.; Zhao, D. Mixed Matrix Membranes Containing MOF@COF Hybrid Fillers for Efficient CO₂/CH₄ Separation. *J. Membr. Sci.* **2019**, *573*, 97–106.
- (19) Tang, Y.; Feng, S.; Fan, L.; Pang, J.; Fan, W.; Kong, G.; Kang, Z.; Sun, D. Covalent Organic Frameworks Combined with Graphene Oxide to Fabricate Membranes for H₂/CO₂ Separation. *Sep. Purif. Technol.* **2019**, *223*, 10–16.
- (20) He, G.; Zhang, R.; Jiang, Z. Engineering Covalent Organic Framework Membranes. *Acc. Mater. Res.* **2021**, *2* (8), 630–643.
- (21) 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.
- (22) Daglar, H.; Aydin, S.; Keskin, S. MOF-Based MMMs Breaking the Upper Bounds of Polymers for a Large Variety of Gas Separations. *Sep. Purif. Technol.* **2022**, *281*, 119811.
- (23) Aydin, S.; Altintas, C.; Keskin, S. High-Throughput Screening of COF Membranes and COF/Polymer MMMs for Helium Separation and Hydrogen Purification. *ACS Appl. Mater. Interfaces* **2022**, *14* (18), 21738–21749.
- (24) Feng, M.; Cheng, M.; Deng, J.; Ji, X.; Zhou, L.; Dang, Y.; Bi, K.; Dai, Z.; Dai, Y. High-Throughput Computational Screening of Covalent-Organic Framework Membranes for Helium Purification. *Results Eng.* **2022**, *15*, 100538.
- (25) Qiao, Z.; Xu, Q.; Jiang, J. High-Throughput Computational Screening of Metal-Organic Framework Membranes for Upgrading of Natural Gas. *J. Membr. Sci.* **2018**, *551*, 47–54.
- (26) Glover, J.; Besley, E. A High-Throughput Screening of Metal-Organic Framework Based Membranes for Biogas Upgrading. *Faraday Discuss.* **2021**, *231*, 235–257.
- (27) Jablonka, K. M.; Ongari, D.; Moosavi, S. M.; Smit, B. Big-Data Science in Porous Materials: Materials Genomics and Machine Learning. *Chem. Rev.* **2020**, *120* (16), 8066–8129.
- (28) Feng, M.; Cheng, M.; Ji, X.; Zhou, L.; Dang, Y.; Bi, K.; Dai, Z.; Dai, Y. Finding the Optimal CO₂ Adsorption Material: Prediction of Multi-Properties of Metal-Organic Frameworks (MOFs) Based on DeepFM. *Sep. Purif. Technol.* **2022**, *302*, 122111.
- (29) Cheng, M.; Zhang, Z.; Wang, S.; Bi, K.; Hu, K.; Dai, Z.; Dai, Y.; Liu, C.; Zhou, L.; Ji, X.; Shi, W. A Large-Scale Screening of Metal-Organic Frameworks for Iodine Capture Combining Molecular Simulation and Machine Learning. *Front. Environ. Sci. Eng.* **2023**, *17* (12), 148.
- (30) 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.
- (31) Daglar, H.; Keskin, S. Combining Machine Learning and Molecular Simulations to Unlock Gas Separation Potentials of MOF Membranes and MOF/Polymer MMMs. *ACS Appl. Mater. Interfaces* **2022**, *14* (28), 32134–32148.
- (32) Yang, J.; Tao, L.; He, J.; McCutcheon, J. R.; Li, Y. Machine Learning Enables Interpretable Discovery of Innovative Polymers for Gas Separation Membranes. *Sci. Adv.* **2022**, *8* (29), No. eabn9545.
- (33) 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.
- (34) Mercado, R.; Fu, R.-S.; Yakutovich, A. V.; Talirz, L.; Haranczyk, M.; Smit, B. In Silico Design of 2D and 3D Covalent Organic Frameworks for Methane Storage Applications. *Chem. Mater.* **2018**, *30* (15), 5069–5086.
- (35) Dubbeldam, D.; Calero, S.; Ellis, D. E.; Snurr, R. Q. RASPA: Molecular Simulation Software for Adsorption and Diffusion in Flexible Nanoporous Materials. *Mol. Simul.* **2016**, *42* (2), 81–101.
- (36) Martin, M. G.; Siepmann, J. I. Transferable Potentials for Phase Equilibria. 1. United-Atom Description of n-Alkanes. *J. Phys. Chem. B* **1998**, *102* (14), 2569–2577.
- (37) Potoff, J. J.; Siepmann, J. I. Vapor-Liquid Equilibria of Mixtures Containing Alkanes, Carbon Dioxide, and Nitrogen. *AIChE J.* **2001**, *47* (7), 1676–1682.
- (38) Shah, M. S.; Tsapatsis, M.; Siepmann, J. I. Development of the Transferable Potentials for Phase Equilibria Model for Hydrogen Sulfide. *J. Phys. Chem. B* **2015**, *119* (23), 7041–7052.
- (39) 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.
- (40) 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.
- (41) Wang, Y.; Wang, G.; Liu, Y.; Zheng, B.; Wang, Z.; Yang, Q. Identifying Promising Covalent-Organic Frameworks for Decarburation and Desulfurization from Biogas via Computational Screening. *ACS Sustain. Chem. Eng.* **2021**, *9* (26), 8858–8867.
- (42) Wilmer, C. E.; Snurr, R. Q. Towards Rapid Computational Screening of Metal-Organic Frameworks for Carbon Dioxide Capture: Calculation of Framework Charges via Charge Equilibration. *Chem. Eng. J.* **2011**, *171* (3), 775–781.
- (43) Ewald, P. P. Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Ann. Phys.* **1921**, *369* (3), 253–287.
- (44) Daglar, H.; Erucar, I.; Keskin, S. Exploring the Performance Limits of MOF/Polymer MMMs for O₂/N₂ Separation Using Computational Screening. *J. Membr. Sci.* **2021**, *618*, 118555.
- (45) Wijmans, J. G.; Baker, R. W. The Solution-Diffusion Model: A Review. *J. Membr. Sci.* **1995**, *107* (1), 1–21.
- (46) Keskin, S.; Sholl, D. S. Assessment of a Metal-Organic Framework Membrane for Gas Separations Using Atomically Detailed Calculations: CO₂, CH₄, N₂, H₂ Mixtures in MOF-5. *Ind. Eng. Chem. Res.* **2009**, *48* (2), 914–922.
- (47) Maxwell, J. C. *A Treatise on Electricity and Magnetism*; Clarendon Press, 1873.
- (48) Willems, T. F.; Rycroft, C. H.; 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.
- (49) Ongari, D.; Boyd, P. G.; Barthel, S.; Witman, M.; Haranczyk, M.; Smit, B. Accurate Characterization of the Pore Volume in Microporous Crystalline Materials. *Langmuir* **2017**, *33* (51), 14529–14538.
- (50) Pardakhti, M.; Nanda, P.; Srivastava, R. Impact of Chemical Features on Methane Adsorption by Porous Materials at Varying Pressures. *J. Phys. Chem. C* **2020**, *124* (8), 4534–4544.
- (51) Pardakhti, M.; Moharreri, E.; Wanik, D.; Suib, S. L.; Srivastava, R. Machine Learning Using Combined Structural and Chemical Descriptors for Prediction of Methane Adsorption Performance of Metal Organic Frameworks (MOFs). *ACS Comb. Sci.* **2017**, *19* (10), 640–645.
- (52) Benesty, J.; Chen, J.; Huang, Y.; Cohen, I. Pearson Correlation Coefficient. In *Noise Reduction in Speech Processing*; Springer Topics in Signal Processing; Springer Berlin Heidelberg: Berlin, Heidelberg, 2009; Vol. 2, pp 1–4. DOI: [10.1007/978-3-642-00296-0_5](https://doi.org/10.1007/978-3-642-00296-0_5).
- (53) Olson, R. S.; Moore, J. H. TPOT: A Tree-Based Pipeline Optimization Tool for Automating Machine Learning. In *Automated Machine Learning*; Hutter, F.; Kotthoff, L.; Vanschoren, J., Eds.; The Springer Series on Challenges in Machine Learning; Springer International Publishing: Cham, 2019; pp 151–160. DOI: [10.1007/978-3-030-05318-5_8](https://doi.org/10.1007/978-3-030-05318-5_8).
- (54) Lundberg, S. M.; Lee, S.-I. A Unified Approach to Interpreting Model Predictions. In *Advances in Neural Information Processing Systems*; Guyon, I.; Luxburg, U. V.; Bengio, S.; Wallach, H.; Fergus, R.; Vishwanathan, S.; Garnett, R., Eds.; Curran Associates, Inc., 2017; Vol. 30.

- (55) Bucior, B. J.; Bobbitt, N. S.; Islamoglu, T.; Goswami, S.; Gopalan, A.; Yildirim, T.; Farha, O. K.; Bagheri, N.; Snurr, R. Q. Energy-Based Descriptors to Rapidly Predict Hydrogen Storage in Metal-Organic Frameworks. *Mol. Syst. Des. Eng.* **2019**, *4* (1), 162–174.
- (56) Burner, J.; Schwiedrzik, L.; Krykunov, M.; Luo, J.; Boyd, P. G.; Woo, T. K. High-Performing Deep Learning Regression Models for Predicting Low-Pressure CO₂ Adsorption Properties of Metal-Organic Frameworks. *J. Phys. Chem. C* **2020**, *124* (51), 27996–28005.
- (57) Dureckova, H.; Krykunov, M.; Aghaji, M. Z.; Woo, T. K. Robust Machine Learning Models for Predicting High CO₂ Working Capacity and CO₂ /H₂ Selectivity of Gas Adsorption in Metal Organic Frameworks for Precombustion Carbon Capture. *J. Phys. Chem. C* **2019**, *123* (7), 4133–4139.
- (58) 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.
- (59) Biswal, B. P.; Chaudhari, H. D.; Banerjee, R.; Kharul, U. K. Chemically Stable Covalent Organic Framework (COF)-Polybenzimidazole Hybrid Membranes: Enhanced Gas Separation through Pore Modulation. *Chem. - Eur. J.* **2016**, *22* (14), 4695–4699.
- (60) Ying, Y.; Peh, S. B.; Yang, H.; Yang, Z.; Zhao, D. Ultrathin Covalent Organic Framework Membranes via a Multi-Interfacial Engineering Strategy for Gas Separation. *Adv. Mater.* **2022**, *34* (25), 2104946.
- (61) Fan, L.; Kang, Z.; Shen, Y.; Wang, S.; Zhao, H.; Sun, H.; Hu, X.; Sun, H.; Wang, R.; Sun, D. Mixed Matrix Membranes Based on Metal-Organic Frameworks with Tunable Pore Size for CO₂ Separation. *Cryst. Growth Des.* **2018**, *18* (8), 4365–4371.
- (62) Wang, S.; Wei, X.; Li, Z.; Liu, Y.; Wang, H.; Zou, L.; Lu, D.; Hassan Akhtar, F.; Wang, X.; Wu, C.; Luo, S. Recent Advances in Developing Mixed Matrix Membranes Based on Covalent Organic Frameworks. *Sep. Purif. Technol.* **2022**, *301*, 122004.
- (63) Liu, Y.; Wu, H.; Wu, S.; Song, S.; Guo, Z.; Ren, Y.; Zhao, R.; Yang, L.; Wu, Y.; Jiang, Z. Multifunctional Covalent Organic Framework (COF)-Based Mixed Matrix Membranes for Enhanced CO₂ Separation. *J. Membr. Sci.* **2021**, *618*, 118693.
- (64) Moosavi, S. M.; Nandy, A.; Jablonka, K. M.; Ongari, D.; Janet, J. P.; Boyd, P. G.; Lee, Y.; Smit, B.; Kulik, H. J. Understanding the Diversity of the Metal-Organic Framework Ecosystem. *Nat. Commun.* **2020**, *11* (1), 4068.
- (65) Wang, Y.; Liu, Y.; Li, H.; Guan, X.; Xue, M.; Yan, Y.; Valtchev, V.; Qiu, S.; Fang, Q. Three-Dimensional Mesoporous Covalent Organic Frameworks through Steric Hindrance Engineering. *J. Am. Chem. Soc.* **2020**, *142* (8), 3736–3741.
- (66) Yan, S.; Guan, X.; Li, H.; Li, D.; Xue, M.; Yan, Y.; Valtchev, V.; Qiu, S.; Fang, Q. Three-Dimensional Salphen-Based Covalent-Organic Frameworks as Catalytic Antioxidants. *J. Am. Chem. Soc.* **2019**, *141* (7), 2920–2924.
- (67) Lin, G.; Ding, H.; Yuan, D.; Wang, B.; Wang, C. A Pyrene-Based, Fluorescent Three-Dimensional Covalent Organic Framework. *J. Am. Chem. Soc.* **2016**, *138* (10), 3302–3305.
- (68) Hosseini Monjezi, B.; Kutonova, K.; Tsotsalas, M.; Henke, S.; Knebel, A. Current Trends in Metal-Organic and Covalent Organic Framework Membrane Materials. *Angew. Chem., Int. Ed.* **2021**, *60* (28), 15153–15164.