Machine learning guided design of covalent organic frameworks for CO_2 capture in wet flue gas

Shuna Yang[#,1], Weichen Zhu[#,1], Linbin Zhu[#,1], Xue Ma, Tongan Yan[2], Nengcui Gu[1], Youshi Lan*[,3], Yi Huang[1], Mingyuan Yuan[1], Minman Tong*[,1]

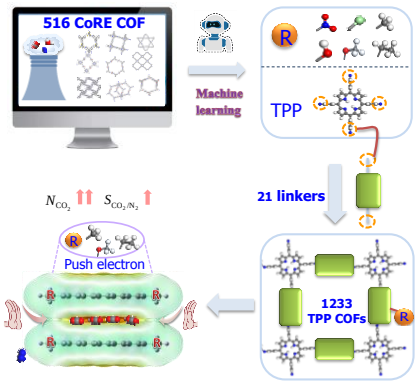
[1] *School of Chemistry and Materials Science, Jiangsu Normal University, Xuzhou 221116, China.*

[2] *State Key Laboratory of Organic−Inorganic Composites, Beijing University of Chemical Technology, Beijing 100029, China.*

[3] *China Institute of Atomic Energy, Department of Radiochemistry, Beijing 102413, China.*

E-mail: tongmm@jsnu.edu.cn; lanyoushi.dr@foxmail.com

Graphic Abstract



Abstract

Discovery of remarkable porous materials for CO_2 capture from wet flue gas is of great significance to reduce the CO_2 emissions, but elucidating the most critical structure features for boosting CO_2 capture capabilities remains a great challenge. Here, machine learning assisted computational screening on 516 experimental covalent organic frameworks (COFs) identify the superior secondary building units (SBUs) for wet flue gas separation, which are tetraphenylporphyrin unit in **sql**-type COFs and functional groups. Accordingly, 1233 COFs are assembled using the superior SBUs. Combined with DFT calculations, the "electron-donating induced vdW interaction" effect is discovered to design better-performing COFs with superior CO_2 uptake, which can achieve 4.4 mmol g^{-1} with CO_2/N_2 selectivity of 104.8; while the "electron-withdrawing induced vdW+electrostatic coupling interaction" effect is proposed to design better-performing COFs with superior CO_2/N_2 selectivity, which can arrive 277.6 with CO_2 uptake of 2.2 mmol g^{-1}, in this case, H_2O contributes to improving the CO_2/N_2 selectivity.

Keywords: covalent organic framework, wet flue gas separation, machine learning, material design

1.Introduction

In recent decades, the rapidly rising levels of CO_2 in the atmosphere have sparked public concerns around the world[1-3]. Many governments have formulated policies to reduce CO_2 emissions in the midterm and longterm plan including carbon capture and storage (CCS)[4]. However, the current CCS technology relies on high cost[5,6], energy inefficient and corrosive aqueous amine solutions[7]. Alternatively, porous adsorbent materials can be treated as a more effective CO_2 capture process[8-11]. A major source of CO_2 emissions comes from the generated flue gas after burning fossil fuels to produce electricity in power plants. The flue gas mainly contains N_2(70-75%) , CO_2(15-16%) and H_2O(5-7%)[11]. Although the H_2O in flue gas stream is of trace amount, the tolerance of adsorbent to H_2O is quite important to its actual use[12]. Most conventional adsorbent materials (e.g. zeolites and porous carbon) are subjected to water poisoning[13-15]. Hence, the development of new adsorbents represents a worth pursuing target[16].

Covalent organic framework (COF) represents a type of crystalline porous material combining organic linkers by strong covalent bonds. The covalent linkage in COFs ensures their application even in some harsh conditions[17]. The advantages like high specific surface area, high porosity and framework tenability endow COFs potentials as promising adsorbents. Li et al.[18] studied water adsorption and the impact on CO_2 capture by GCMC simulations of 18 chemically stable COFs, in which COF-300 exhibit excellent resistance to high relative humidity (p/p0 = 0.8) when used to capture CO_2. Zhao et al.[19] estimated a series of Schiff-base COFs for wet flue gas separation. It is indicated that keto-COFs of NUS-2 and TpPa-1 outperform the imine-COFs; moreover, they exhibit excellent hydrothermal and chemical stability. Keskin et al.[20] used high throughput computational screening to study the adsorption based flue gas separation performance of 295 COFs. It is found that COF adsorbents could compete with metal-organic frameworks (MOFs) in capturing CO_2 in flue gas, especially most COFs maintain high CO_2 selectivity in the presence of water.

In published works, hydrophobic framework is usually the most recommended structure feature as it is water-resistant. However, hydrophobic pores without sufficient polarity usually show poor separation performance on CO_2/N_2. Zhao et al.[19] suggested that COFs with

moderate hydrophobicity would be promising adsorbent candidates for practical post combustion CO_2 capture. An in depth and meticulous scientific issue is that what structural unit can commonly make COFs show both high CO_2/N_2 selectivity and CO_2 uptake in humid environment. To answer this question, we need to do two sub-works: 1. a thorough performance evaluation on the present synthesized COFs; 2. a deep mining of the key structure feature that can universally facilitates CO_2/N_2 separation.

Machine learning is an increasingly popular tool in material science due to its powerful ability in analyzing and extracting important feature from big data[21-30]. The machine learning obtained useful information can contribute to design materials in a more effective way. For example, Qiao et al.[28] combined machine learning and molecular fingerprint to identify the excellent bits (aromatic rings, double bonds, transition metals, halogens and oxygen heterocycles) that could promote the non-methane hydrocarbons capture performance of MOFs. Instead of counting the common characteristics that may benefit the property, Boyd et al.[12] used data mining to directly search for the water-resistant CO_2 adsorption sites, named adsorbaphores, from the top-ranked 8325 MOFs. Then, they generated a new set of hypothetical structures that contain the adsorbaphores and synthesized two hypothetical structures with expectant good performance. As is shown, the machine learning techniques develop fast in MOFs that specific substructure features are defined for guiding design of MOFs. By contrast, the application of machine learning in COFs is lately started and is mainly focused on establishing precise correlation model to forecast material properties[31,32]. Machine learning based mining of the excellent substructure features so as to design COFs with optimized performance are not yet reported as far as we know.

In this work, the superior secondary building units (SBUs) of COFs that can promote the CO_2 capture in wet flue gas are identified through machine learning analysis on the separation performance data of 516 computational ready experimental (CoRE) COFs. Then, 1233 potential COFs are constructed based on the superior SBUs. DFT calculations are applied to clarify the separation mechanism of the superior SBUs in facilitating $CO_2/N_2/H_2O$ separation on the electronic level. By accomplishing the research pipeline of superior substructure mining, material construction and mechanism analysis, we propose the “electron-donating induced vdW interaction” strategy to obtain the better-performing COFs with superior CO_2

uptake, and the "electron-withdrawing induced vdW+electrostatic coupling interaction" strategy to obtain the better-performing COFs with superior CO_2/N_2 selectivity. Through combining machine learning, materials genomics based material construction and multi-scale simulation, this work provide specific and in-depth theoretical guidance on designing COFs with both high CO_2/N_2 selectivity and high CO_2 uptake in wet environment.

2.Materials and Methods

2.1 **CoRE COF**

The number of experimentally synthesized COFs in our built CoRE COF database[33] is enlarged to 613. Among them, 97 boric-based COFs in the collected COFs are not calculated in this work considering their instability in the presence of water. Hence, 516 experimentally synthesized COFs (58 3D-COFs and 458 2D-COFs) are estimated for CO_2 capture in wet flue gas. Zeo++ software package[34,35] is used to calculate the structural properties of COFs materials, such as void fraction (V_f), surface area (S_{acc}), pore limiting diameter (PLD), largest cavity diameter (LCD) for each material, S_{acc} is calculated using a probe molecule with size equal to the kinetic diameter of N_2 (3.68 Å). V_{free} is computed with a probe size of 0.0 Å, which is the absolute amount of volume unoccupied by the framework atoms.

2.2 **Force fields**

The Lennard-Jones (LJ) interaction of COF framework atoms are taken from DREIDING force field[36] as it can accurately describe the gas adsorption behavior of COFs[37]. Partial point charges are assigned to frameworks using charge equilibration method (QEq)[38]. The LJ interaction and atomic charges for CO_2 and N_2 are taken from the TraPPE force field[39]. Previous studies have validated the reliability of the above force filed set in accurately describing CO_2 and N_2 adsorption behavior in COFs[37,40,41]. A set of four commonly used water models[18,42,43], including TIP4P, TIP4P_Ew, SPC and SPC/E, are evaluated in our study. The SPC/E model is chosen in our studies due to its computational efficiency and good accuracy to reproduce the experimental absorption performance of COFs (Figure S1).

2.3 **GCMC simulation**

The separation performances of CO_2/N_2 (0.1: 0.9) mixture under dry and wet condition

(relative humidity is ~90%) at 1 bar and 298K are studied by Grand canonical Monte Carlo (GCMC) simulations. Leonard Jones (LJ) 12-6 is used to calculate the van der Waals interaction (vdW) parameters between framework atoms and adsorbents. The Coulomb potential is used to consider the electrostatic interactions of N_2, H_2O, and CO_2 molecules. A cutoff radius is set to 14.0 Å for the simulation system. The long range interactions are handled by the Ewald summation technique. The fugacity of GCMC simulations are calculated by Peng-Robinson equation of state. For each state point, GCMC simulations consisted of 1×10^7 steps to confirm the equilibration, and the following 1×10^7 steps to sample the desired properties. The adsorption selectivity (S_{CO_2/N_2}) of adsorbent performance evaluation metrics from molecular simulation is calculated from the following definition, where *x* and *y* are the mole fractions of the gases in the adsorbed and bulk phases, respectively.

$$S_{CO_2/N_2} = (x_{CO_2} / x_{N_2})(y_{N_2} / y_{CO_2}) \quad (1)$$

2.4 DFT simulation

DFT calculations are performed by using D3[44] dispersion corrected M062x[45] /6-31+g(d)[46,47] as implemented in Gaussian 09[48]. The geometry and electronic structure optimizations of COF SBUs are carried out in gas phase with no constraints imposed beyond the multiplicity of the electronic state, but partial atoms are fixed to keep stacking modes of SBUs unchanged during optimizations. The stability of molecular structure of validated by frequency analysis, and the highest occupied molecular orbital (HOMO) and the electrostatic potential (ESP) of the SBUs are analyzed. Binding energy (BE) between CO_2 and the SBUs are evaluated with basis set superposition error (BSSE) correction by following equation:

$$BE = E_{SBU+CO_2} - E_{SBU} - E_{CO_2} - E_{BSSE} \quad (2)$$

2.5 Machine learning

In the machine learning process of this work, four algorithms are used to analyze the quantitative relationship between input descriptors and output separation performance of COFs. The algorithms are decision tree (DT), random forest (RF), Xtreme Gradient Boosting (XGBoost), and Categorical Boosting (CatBoost) in scikit-learn library function coded in Python[49]. Here, 90% of CoRE COFs are randomly selected as the training set, and the

remaining 10% are used as the test set. During the training process, each model is trained with Bayesian optimization algorithm to determine the hyper-parameters that could yield the best accuracy according to a 5-fold cross validation. After obtaining the optimal hyper-parameters, each model is retrained on the entire training dataset, and the final prediction performance evaluation is performed on the test dataset. The accuracy of the model is judged by the mean absolute error (MAE), root-mean-square error (RMSE) and R-squared (R^2), as shown in equations (3-5)[32,50,51].

$$R^2(x,y) = 1 - \frac{\sum_{i=1}^{n_{samples}} (x_i - y_i)^2}{\sum_{i=1}^{n_{samples}} (y_i - \bar{y})^2} \quad (3)$$

$$MAE(x,y) = \frac{1}{n_{samples}} \sum_{i=1}^{n_{samples}} |x_i - y_i| \quad (4)$$

$$RMSE(x,y) = \sqrt{\frac{1}{n_{samples}} \sum_{i=1}^{n_{samples}} (x_i - y_i)^2} \quad (5)$$

where $n_{samples}$, x_i, y_i and $\bar{y}$ denote the number of structures, simulated values from GCMC, machine learning predicted values, and average machine learning predicted values, respectively.

3. Results and Discussion

3.1 CO_2 capture from wet flue gas using CoRE COFs

Two key indicators of separation performance in wet flue gas, CO_2 adsorption uptake (N_{CO_2}) and CO_2/N_2 selectivity (S_{CO_2/N_2}), are calculated for the 516 COFs under the operating conditions of 298K and 1bar, as shown in Figure 1a. Among them, N_{CO_2} of 47 COFs exceeds 0.5 mmol g^{-1}, and S_{CO_2/N_2} of 38 COFs exceeds 50. 2D-COFs show better separation performance than 3D-COFs do. The highest N_{CO_2} and S_{CO_2/N_2} can achieve 2.87 mmol g^{-1} and 360, which are quite outstanding performance in porous adsorbents[52]. However, only 22 COFs outperform when using comprehensive indicators (N_{CO_2} > 0.5 mmol g^{-1} and S_{CO_2/N_2} > 50), due to the trade-off effect between N_{CO_2} and S_{CO_2/N_2}. To analyze the influence of structure features on CO_2 capture performance, four decision tree based machine learning algorithms (DT, RF, XGBoost, CatBoost) are used. We select seven structural parameters as descriptors that are functional group (FG), interlayer spacing (IS), V_f, S_{acc}, PLD, LCD and

density of framework (DF). FG is a category feature described by 1 or 0, representing the COF containing or not containing functional groups. Except for FG, other descriptors are numeric parameters. The energetic parameters (e.g. adsorption heat) are not used because they are the output parameters of GCMC simulations. Figure 1b,c present the prediction results of N_{CO_2} and S_{CO_2/N_2} obtained by machine learning and GCMC simulation. DT algorithm uses only one decision tree while other three algorithms use multiple trees, resulting the better fitting and generalization capability of the latter three algorithms. XGBoost and CatBoost are representatives of gradient boosting decision tree algorithms[53]. XGBoost is composed of multiple lift trees and develops trees in a level-wise way, while CatBoost uses oblivious trees as base predictor. Thus, CatBoost is good at dealing with category features, and outperform in this work (CatBoost has the highest R^2 for N_{CO_2} and S_{CO_2/N_2} during fitting).

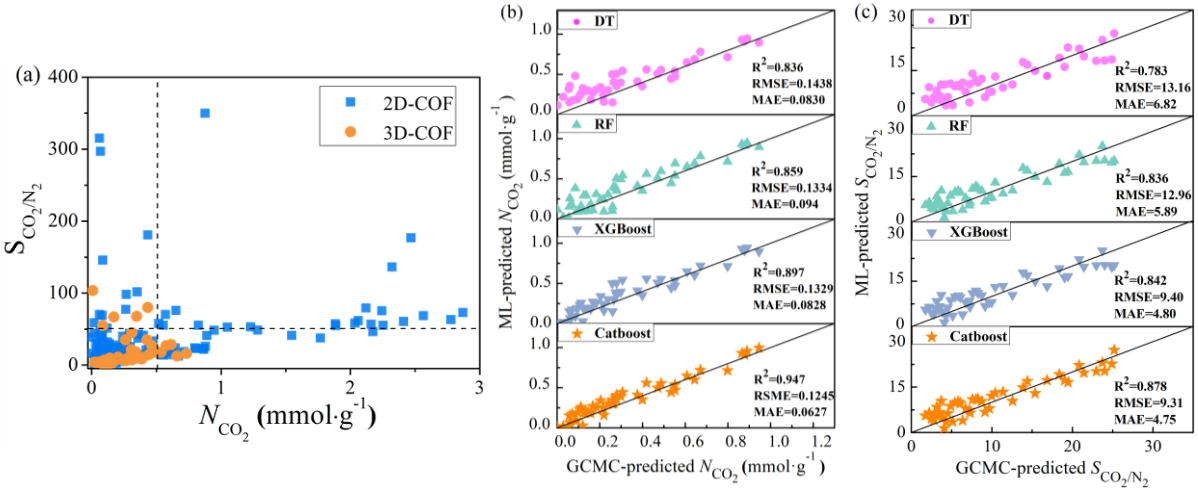


Figure 1. (a) Distribution diagram of the relationship between N_{CO_2} and S_{CO_2/N_2} of the 516 CoRE COFs under wet condition. (b) Accuracy comparison of four machine learning algorithm by using N_{CO_2} and (c) S_{CO_2/N_2} as separation performance indicators. DT, RF, XGBoost and CatBoost denote for decision tree, random forest, xtreme gradient boosting, and categorical boosting, respectively.

By using the Catboost model, the contribution (relative importance) of each descriptor in predicting the N_{CO_2} and S_{CO_2/N_2} can be determined. Here, we use the Shapley's additive interpretation method[54] to evaluate the feature importance values (Figure 2a,b). The higher the importance of a feature, the more it is used in building the boosting tree when predicting the target property. For N_{CO_2}, the descriptor importance follows the rank of IS > S_{acc} > PLD > LCD > DF > V_f > FG. Interlayer spacing is a structure parameter of 2D-COFs, a less studied

descriptor compared with the textural properties like S_{acc}, LCD and so on. A step-further analysis is performed to study the relationship between the adsorption capacity of materials and the interlayer spacing, as shown in Figure 2c. The COFs with high N_{CO_2} (> 1 mmol g^{-1}) are mainly in two types of structures: 1. AB-stacking structures; 2. tetraphenylporphyrin (TPP) contained COFs in **sql** topology with large interlayer spacing of 6.4~8.0 Å, which in turn explain the high contribution of large interlayer spacing in prediction of CO_2 uptake. For S_{CO_2/N_2}, the descriptor importance follows the rank of IS > DF > FG > S_{acc} > PLD > LCD > V_f. By checking the COFs with top 20 high S_{CO_2/N_2}, we find that fifteen of them are functionalized with polar functional groups like –C=O, –F, –Cl and $–NO_2$ (shown in Figure S4). The effectiveness of functionalization in improving S_{CO_2/N_2} is easy to understand because CO_2 is sensitive to electrostatic interaction[55].

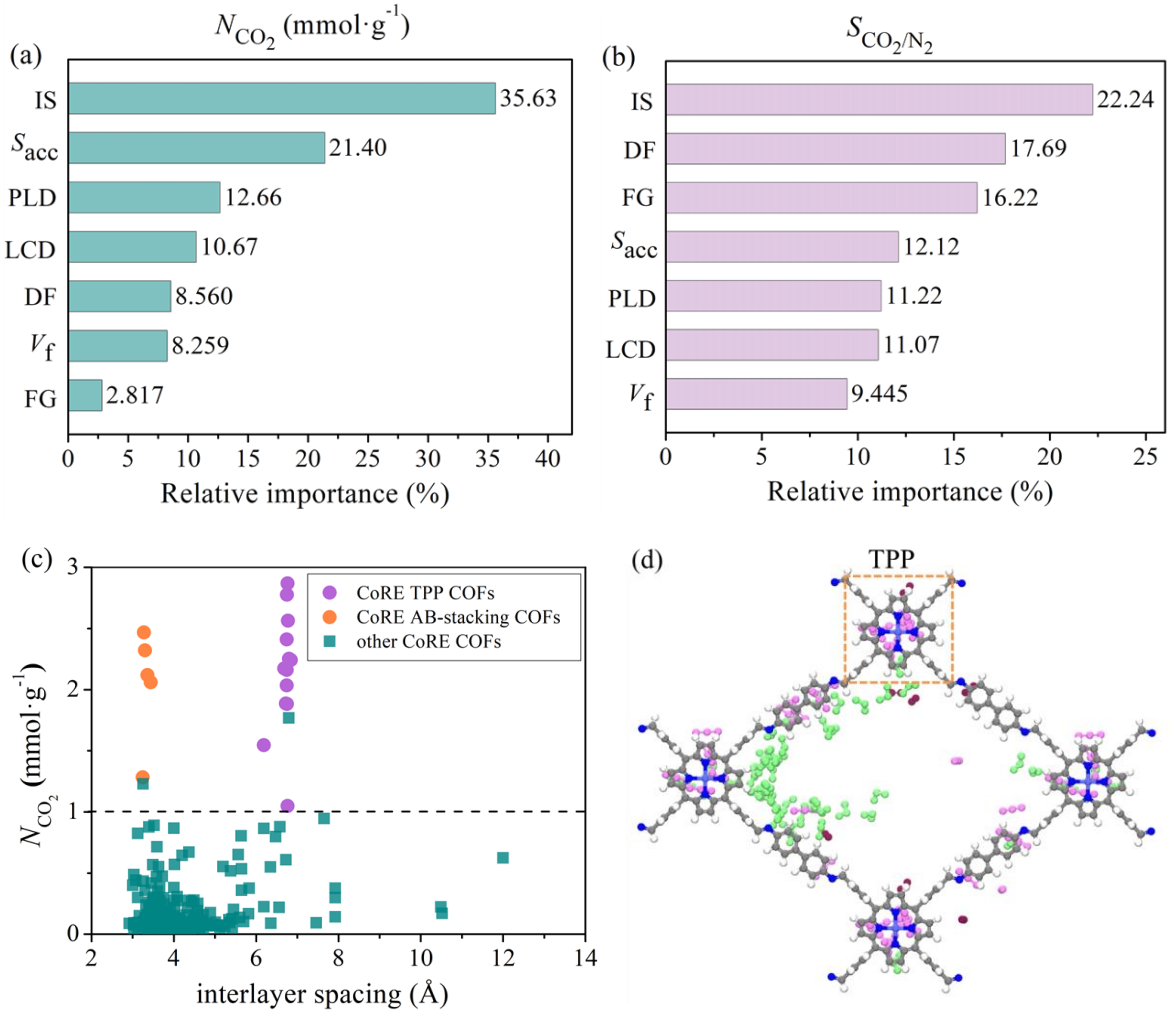


Figure 2. The relative importance of material influencing factors obtained from machine training on (a) N_{CO_2} and (b) S_{CO_2/N_2} under wet condition. (c) The relationship between the

interlayer spacing of the COFs and N_{CO_2}. (d) Humid CO_2/N_2 adsorption configuration of COF-367 containing TPP SBU. (Framework: N, blue; C, gray and H, white; gas molecules: CO_2, pink; N_2, brown; H_2O, green).

3.2 Machine learning guided design of TPP COFs

Machine learning assisted analysis indicates the TPP contained COFs in **sql** topology with interlayer spacing of 6.4~8.0 Å can bring high CO_2 uptake, while functionalization benefits the CO_2/N_2 separation. To solve the selectivity-uptake trade-off dilemma, the two favorable features for N_{CO_2} and S_{CO_2/N_2} are utilized to build COF adsorbents for wet flue gas separation, as shown in Figure 3a. First, in the parent TPP COFs construction step, TPP with 4-connection sites is used as the center unit, and 21 monomers with 2-connection sites are one by one used as the ligand unit to link with the TPP unit to build the **sql** network. We get 21 TPP COFs as the parent materials in this step. The structures of 21 parent TPP COFs are shown in Figure S5. Next, in the modified TPP COFs construction step, six common functional groups (–CH_2CH_3, –CH_3, –OCH_3, –OH, –Cl, –NO_2) are applied to decorate the 21 parent TPP COFs. Every parent COF is decorated by one type of functional group each time. The decoration site and number depend on the structure of the ligand unit. If the functional group has close contact with the framework atom or adjacent functional group, this modified structure is discarded. Finally, we get 1212 modified TPP COFs. The performance change of the TPP COFs is recorded in Figure 3b. In the parent COF construction step, the highest N_{CO_2} is increased to 3.4 mmol g^{-1} compared with the highest 2.9 mmol g^{-1} of the TPP contained CoRE COFs (labeled as CoRE TPP COFs in Figure 3b). In the functionalization step, the highest S_{CO_2/N_2} is improved to 277.6 (the N_{CO_2} of this modified TPP COF is 2.2 mmol g^{-1}); moreover, the highest N_{CO_2} is increased to 4.4 mmol g^{-1} (the S_{CO_2/N_2} of this modified TPP COF is 104.8), demonstrating the synergistic effect of the two designing steps in boosting CO_2/N_2 selectivity and CO_2 uptake. Among the six functional groups, –Cl and –NO_2 decorations contribute to build COFs with the highest CO_2/N_2 selectivity, while –CH_2CH_3, –CH_3, –OCH_3 and –OH are better at steadily increasing the CO_2 uptake, as shown in Figure 3c.

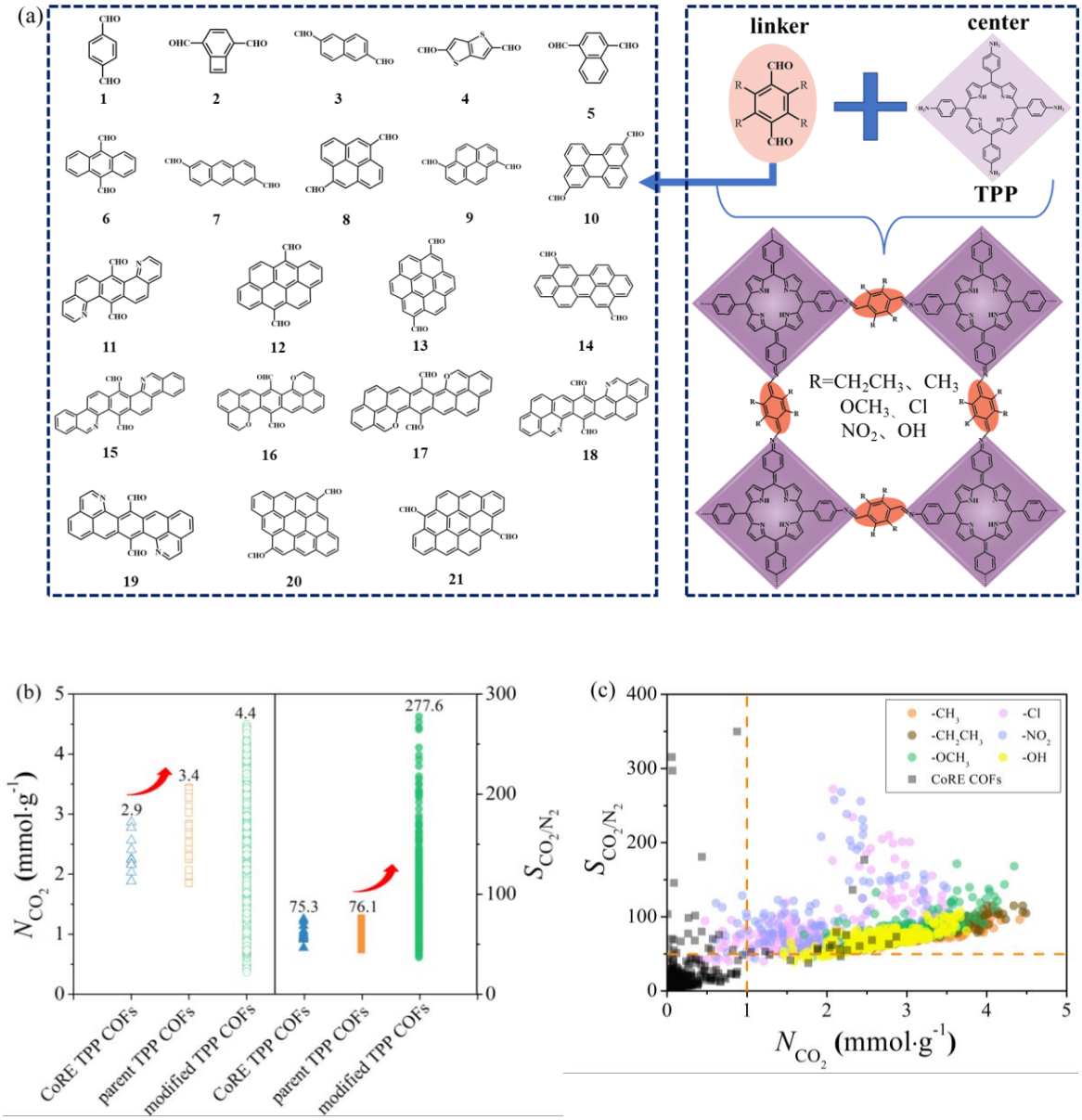


Figure 3. (a) The assembly scheme of the TPP COFs. TPP is the center unit, and 21 monomers are one by one used as the ligand units to build the parent TPP COFs. Six kinds of functional groups are postly modified to the 21 parent TPP COFs. (b) S_{CO_2/N_2} and N_{CO_2} of the 11 CoRE TPP COFs, 21 parent TPP COF and 1212 modified TPP COFs. (c) S_{CO_2/N_2} and N_{CO_2} of the CoRE COFs and designed TPP COFs. The performance of designed TPP COFs is classified by functional groups.

To have detailed knowledge of the improvement on both CO_2/N_2 selectivity and CO_2 uptake by functionalization on 21 parent TPP COFs, we calculate the proportion of these modified materials which outperform or underperform their parent COFs. The parameters are described as “proportion of outperformed COFs” (POC) and “proportion of underperformed

COFs" (PUC). The parameters can describe the universality of the functional group in decorating parent COFs with better performance. The corresponding calculation equations are shown as below,

$$\mathrm{POC} = \frac{\sum_{i=1}^{21} g(y_i \geq x_i)}{\sum_{i=1}^{21} g(y_i)} \qquad \mathrm{PUC} = \frac{\sum_{i=1}^{21} g(y_i < x_i)}{\sum_{i=1}^{21} g(y_i)}$$

where i denotes the 21 parent TPP COFs, x_i, y_i denote the target index (N_{CO_2} or S_{CO_2/N_2}) of the parent TPP COFs and the modified TPP COFs, $g(...)$ is the number of the modified TPP COFs.

Here, the dry flue gas separation performance is also estimated to clarify the effect of water on CO_2 capture in TPP COFs. The data classified by functional groups are present in Figure 4a,b. For CO_2 uptake (Figure 4a), in dry flue gas, $–CH_2CH_3$, $–CH_3$, and $–OCH_3$ can make 95%~99% decorated TPP COFs outperform the parent TPP COFs in CO_2 uptake. The same parameter for –OH, –Cl and $–NO_2$ is only 34%~49%. In wet flue gas, $–CH_3$ and $–CH_2CH_3$ basically maintain the POC as they do in dry flue gas; –OH, $–OCH_3$, –Cl and $–NO_2$ show obvious decline on POC. The H_2O uptake shown in Figure 4c can explain the sharp decrease of CO_2 uptake POC to almost zero in –Cl, $–NO_2$ decorated TPP COFs, because the largely adsorbed H_2O molecules will expel the CO_2 molecules. For CO_2/N_2 selectivity (Figure 4b), the presence of H_2O can promote the positive role of all functional groups on improving the CO_2/N_2 selectivity, especially for –Cl and $–NO_2$. The above data imply the following information: 1. enhancing CO_2 uptake is more difficult than enhancing CO_2/N_2 selectivity in wet flue gas; 2. $–CH_3$, $–CH_2CH_3$ and $–OCH_3$ are more universal than –OH, –Cl and $–NO_2$ in decorating different COFs to have higher CO_2 uptake in both dry and wet flue gas; 3. the performance enhancement mechanisms are quite different in $–CH_3$, $–CH_2CH_3$ and –Cl, $–NO_2$ decorated TPP COFs.

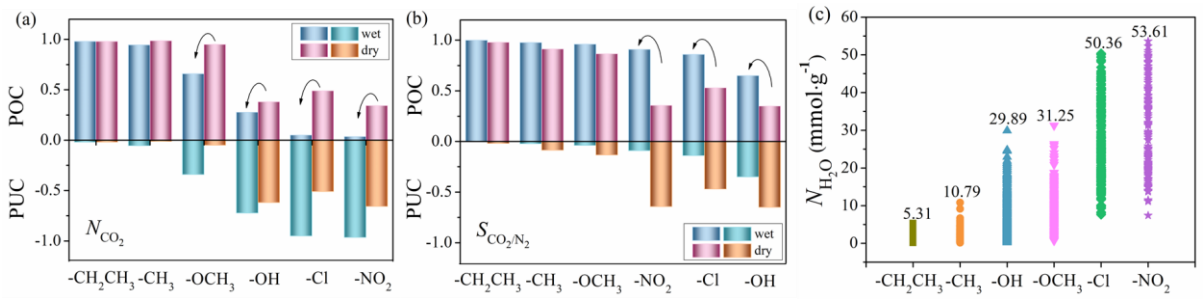


Figure 4. The proportion of the modified COFs which outperform (POC) or underperform (PUC) their parent COFs in (a) N_{CO_2} and (b) S_{CO_2/N_2}. (c) H_2O adsorption capacity of 1212 functionalized COFs classified by functional groups.

3.3 The role of superior SBUs on boosting separation performance

To clarify these issues, we further analyze the center of mass (COM) probability distributions of the adsorbed CO_2 molecules in the TPP COFs. The parent TPP-No.21 COF is used as an example. As shown in Figure 5a, the CO_2 molecules are mainly adsorbed in the interlayers of the TPP unit and the No.21 unit. This is the common feature of all TPP COFs. Specific to the No.21 unit, CO_2 distributions show differences according to the functional groups. In the $–CH_3$, $–CH_2CH_3$ and $–OCH_3$ decorated No.21 units, the CO_2 molecules are evenly and densely distributed (additional figures are shown in Figure S6). It is worth noting that $–CH_3$ and $–CH_2CH_3$ themselves are not the adsorption sites but the adsorbed CO_2 molecules in the interlayered multi-phenyl unit are increased. $–OCH_3$ can served as the weak adsorption sites as we can see CO_2 molecules around the $–OCH_3$. For –OH, –Cl and $–NO_2$ decorated No.21 unit, the functional groups themselves are strong CO_2 adsorption sites, while the density of CO_2 in the interlayered multi-phenyl unit is decreased.

How the functional groups affect the multi-phenyl unit adsorb CO_2 molecules? Density functional theory (DFT) calculations are performed to get the highest occupied molecular orbital (HOMO) to answer the question. No.21 unit with para modification by one type functional group is taken as an example to carry out the HOMO analysis. The HOMO energy levels of the functionalized No.21 unit follow the order of $–CH_2CH_3$ > $–CH_3$ > $–OCH_3$ > H > –OH > –Cl > $–NO_2$, as shown in Figure 5b. The less negative HOMO value, the stronger quadrupole-π electron interaction between the No.21 unir and CO_2 molecule. Therefore, the

HOMO analysis suggests the electron-donating effect of $–CH_2CH_3$, $–CH_3$, $–OCH_3$ and the electron-withdrawing effect of –OH, –Cl, $–NO_2$ on the No.21 unit. The adsorption of CO_2 molecules in the interlayer of No.21 unit relies on the vdW force interaction between molecular quadrupole of CO_2 and an aromatic system. The more the electron on the multi-phenyl unit, the stronger it is to interact with the CO_2 through vdW force. An intuitive observation on the electrostatic potential of the functionalized No.21 units clearly proves the speculation. $–CH_2CH_3$/$–NO_2$ functionalized No.21 units exhibit the most negative/positive electrostatic potential, indicating the most abundant/insufficient electron of the π aromatic system that can interact with CO_2 molecule (Figure 5c). The electrostatic potential of the $–CH_3$, $–OCH_3$, –OH, –Cl functionalized No.21 units are shown in Figure S7, which is in accordance with the order of HOMO for the functional groups. Then, the binding energy of CO_2 in the interlayer of the No.21 units is calculated to explore the effect of electron aggregation or loss in the No.21 units on CO_2 adsorption. Figure 5d presents the binding energy follows the order of $–CH_3$ > $–OCH_3$ > $–CH_2CH_3$ > H > –OH > –Cl > $–NO_2$. The results declare that utilizing electron-donating group to modify the aromatic unit facilitates the CO_2 adsorption through the quadrupole-π interaction.

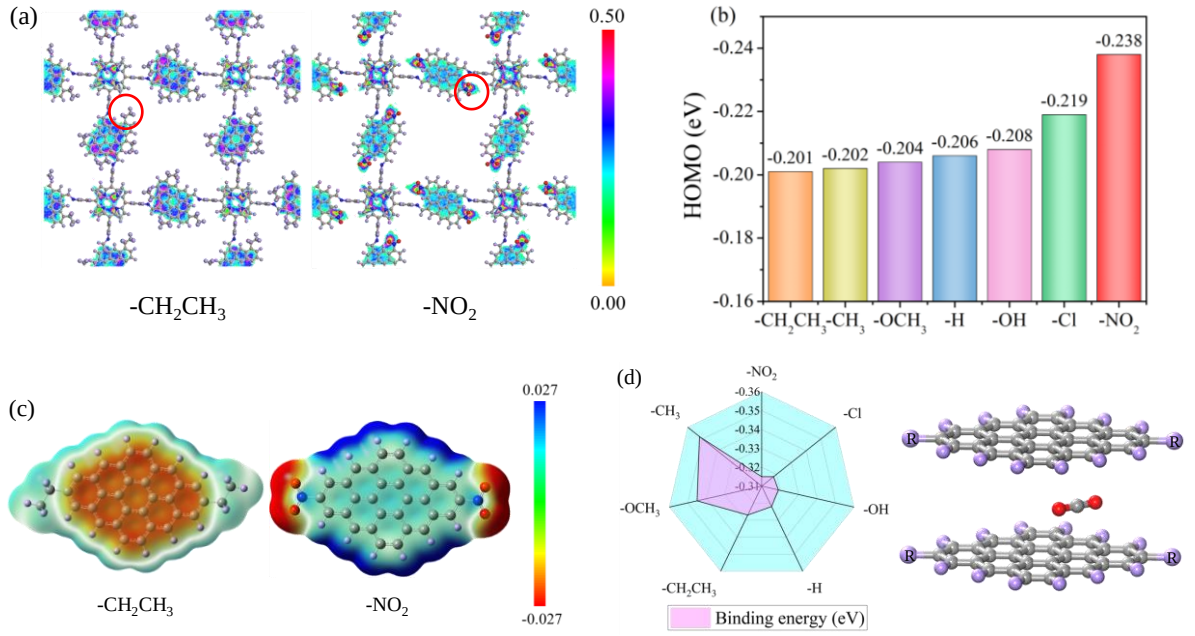


Figure 5. (a) Contour plots of the COM probability densities of CO_2 adsorbed in $–CH_2CH_3$ and $–NO_2$ modified TPP COF using No.21 unit (N, blue; O, red; C, gray and H, violet). Conditions of other functional groups are shown in Figure S6. (b) HOMO values of the

modified No.21 units (c) Electrostatic potential of the $–CH_2CH_3$ and $–NO_2$ modified No.21 unit. Conditions of other functional groups are shown in Figure S7. (d) Binding energy of CO_2 in the interlayer of the modified No.21 units (R = $-CH_2CH_3$, $-CH_3$, $-OCH_3$, -OH, -Cl, $-NO_2$).

Besides the interlayers of No.21 units, –OH, –Cl and $–NO_2$ themselves are strong adsorption sites for CO_2 relied on the polar electrostatic interaction (Figure 5a, Figure S6). Figure 6 depicts the CO_2/N_2 separation performance under dry and wet environment classified by functional groups. In dry flue gas, a certain part of the –Cl and $–NO_2$ functionalized TPP-COFs can achieve superior high CO_2 uptake of 4~6.3 mmol g^{-1}. In fact, –Cl and $–NO_2$ surpass $–NH_2$ and $–CH_3$ in CO_2 uptake under dry environment. In wet flue gas, H_2O strongly competes with CO_2 for the polar electrostatic interaction sites of –Cl and $–NO_2$ (Figure 4c), bringing the sharp decrease in CO_2 uptake. Even so, we should also note that the –Cl and $–NO_2$ functionalization can keep a high CO_2 uptake of 2~3 mmol g^{-1} and realize a much higher CO_2/N_2 selectivity between 180~300 in wet flue gas, because the largely adsorbed H_2O also impair the adsorption of N_2.

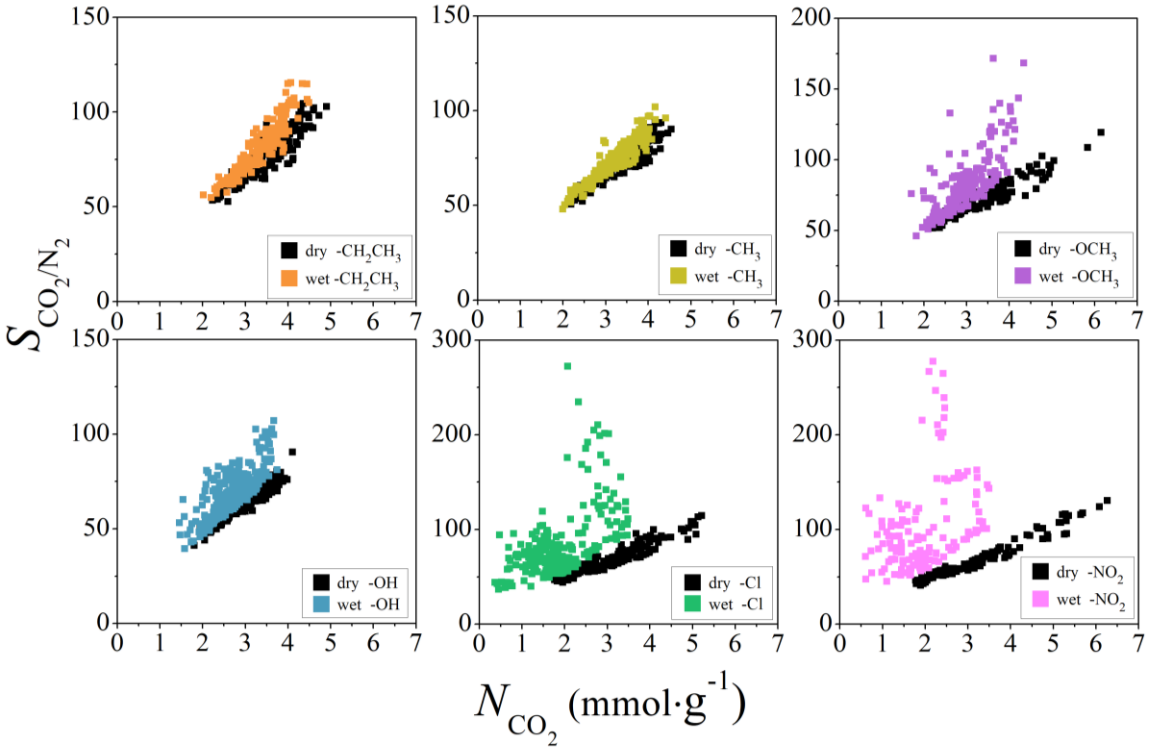


Figure 6. N_{CO_2} and S_{CO_2/N_2} of modified TPP COFs in dry and wet flue gas classified by functional groups.

That is to say, the introduction of functional groups initiates the electron transfer in the

aromatic unit, changing the way of CO_2 adsorption. As shown in Figure 7a, in the assembly of "electron-donating group + delocalized π aromatic system", delocalized π aromatic system is the CO_2 adsorption site, and the electron-donating group intensifies the π-quadrupole vdW interaction between the aromatic unit and CO_2 molecules, which is a water resistant adsorption site that can keep superior high CO_2 uptake under wet environment. In the assembly of "electron-withdrawing group + delocalized π aromatic system" shown in Figure 7b, the delocalized π aromatic system is still CO_2 adsorption sites but the vdW interaction is weakened, and the electron-withdrawing groups become new CO_2 adsorption sites. Under wet environment, the delocalized π aromatic system is kept as the adsorption site for CO_2, while the electron-withdrawing group turns into H_2O adsorption site which may also block the N_2 adsorption, and thus high CO_2 uptake and superior high CO_2/N_2 selectivity can be achieved.

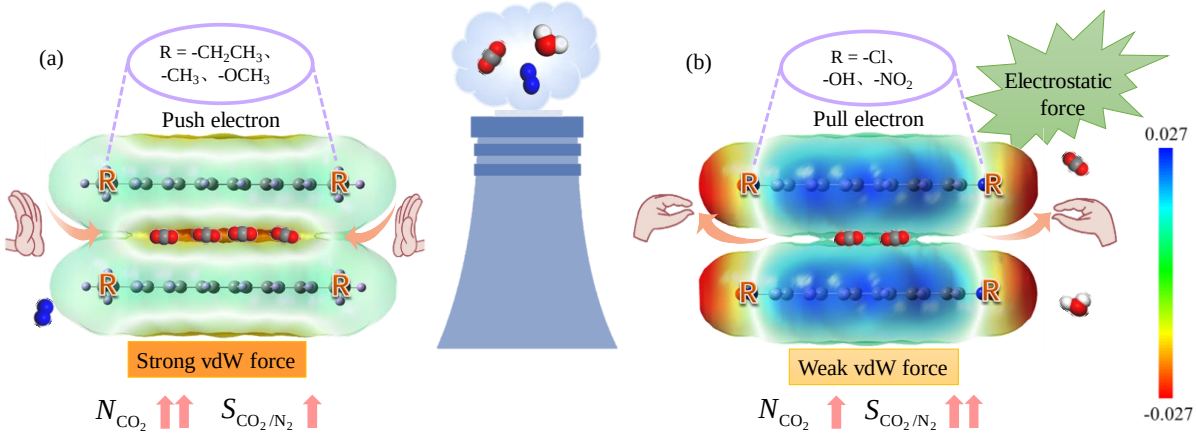


Figure 7. $CO_2/N_2/H_2O$ separation mechanism in modified TPP COFs (a) assembly of "electron-donating group + delocalized π aromatic system", (b) assembly of "electron-withdrawing group + delocalized π aromatic system". Surface of the functionalized No.21 units are mapped with the 2×10^{-5} electron density isosurface. Color of atom in gas molecules : C, gray; N, blue; O, red; H, white.

4. Conclusion

In summary, we use high throughput computational screening combining with machine learning analysis to identify the superior COF SBUs for facilitating CO_2/N_2 separation under wet environment. These superior SBUs, TPP center unit, 21 ligand units and 6 functional groups, are further adopted to construct 21 parent- and 1212 modified-TPP COFs with

potentially high N_{CO_2} and high S_{CO_2/N_2}. DFT calculation analysis on HOMO, electrostatic potential and binding energy reveal the influencing mechanism of functionalization on the wet flue gas separation performance of TPP COFs. The interlayers of the aromatic units in the TPP COFs are favorable and hydrophobic vdW adsorption sites for CO_2. Electron-donating functional groups (–CH_3, –CH_2CH_3 and –OCH_3) reinforce the vdW interaction between the CO_2 and the interlayered aromatic units, enabling the TPP COFs to show high S_{CO_2/N_2}(70~180) and very high N_{CO_2} (3~4.5 mmol g^{-1}) under wet environment. Electron-withdrawing functional groups (–OH, –Cl, –NO_2) reinforce the electrostatic interaction in the TPP COFs, which can bring superior high S_{CO_2/N_2}(150~280) and moderate high N_{CO_2} (2~3 mmol g^{-1}). Our work represents a multi-scale computational study on revealing and then utilizing the critical structure features for boosting CO_2 capture capabilities of COFs, and may provide useful information for designing optimized 2D-COFs adsorbents for wet flue gas separation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We gratefully thank the National Key R&D Program of China (No. 2021YFB3802200), the National Natural Science Foundation of China (No. 21706106), and the State Key Laboratory of Separation Membranes and Membrane Processes (Tiangong University, No. M202102) for the financial support.

References

1. Liu J, Thallapally PK, McGrail BP, Brown DR, Liu J. Progress in adsorption-based CO_2 capture by metal-organic frameworks. *Chem Soc Rev.* 2012;41(6):2308-2322.

2. Cox PM, Betts RA, Jones CD, Spall SA, Totterdell IJ. Acceleration of global warming due to carbon-cycle feedbacks in a coupled climate model. *Nature* 2000;408(6813):184-187.

3. Dindi A, Quang DV, Vega LF, Nashef E, Abu-Zahra MRM. Applications of fly ash for CO_2 capture, utilization, and storage. *J CO_2 Util.* 2019;29:82-102.

4. Hu ZG, Wang YX, Shah BB, Zhao D. CO_2 capture in metal-organic framework adsorbents: an engineering perspective. *Adv Sustain Syst.* 2019;3(1):1800080.

5. Benson SM, Surles T. Carbon dioxide capture and storage: an overview with emphasis on capture and storage in deep geological formations. *Proceedings of the IEEE* 2006;94(10):1795-1805.

6. Bui M, Adjiman CS, Bardow A, et al. Carbon capture and storage (CCS): the way forward. *Energy Environ Sci.* 2018;11(5):1062-1176.

7. Rochelle GT. Amine scrubbing for CO_2 capture. *Science* 2009;325(5948):1652-1654.

8. Lee J, Kim J, Hyeon T. Recent progress in the synthesis of porous carbon materials. *Adv Mater.* 2006;18(16):2073-2094.

9. Furukawa H, Cordova KE, O'Keeffe M, Yaghi OM. The chemistry and applications of metal-organic frameworks. *Science* 2013;341(6149):1230444.

10. Kenarsari SD, Yang D, Jiang G, et al. Review of recent advances in carbon dioxide separation and capture. *RSC Adv.* 2013;3(45):22739-22773.

11. McDonald TM, Lee WR, Mason JA, Wiers BM, Hong CS, Long JR. Capture of carbon dioxide from air and flue gas in the alkylamine-appended metal-organic framework mmen-Mg_2(dobpdc). *J Am Chem Soc*. 2012;134(16):7056-7065.

12. Boyd PG, Chidambaram A, Garcia-Diez E, et al. Data-driven design of metal-organic frameworks for wet flue gas CO_2 capture. *Nature* 2019;576(7786):253-256.

13. Jänchen J, Möhlmann D, Stach H. Water and carbon dioxide sorption properties of natural zeolites and clay minerals at martian surface temperature and pressure conditions. *Studies in surface science and catalysis*. Elsevier, 2007:2116-2121.

14. Sumida K, Rogow DL, Mason JA, et al. Carbon dioxide capture in metal-organic frameworks. *Chem Rev.* 2012;112(2):724-781.

15. Mukherjee A, Okolie JA, Abdelrasoul A, Niu C, Dalai AK. Review of post-combustion carbon dioxide capture technologies using activated carbon. *J Environ Sci.* 2019;83:46-63.

16. Kundu N, Sarkar S. Porous organic frameworks for carbon dioxide capture and storage. *J Environ Chem Eng.* 2021;9(1):105090.

17. Cote AP, Benin AI, Ockwig NW, O'Keeffe M, Matzger AJ, Yaghi OM. Porous, crystalline, covalent organic frameworks. *Science* 2005;310(5751):1166-1170.

18. Ge Y, Zhou H, Ji Y, et al. Understanding water adsorption and the impact on CO_2 capture in chemically stable covalent organic frameworks. *J Phys Chem C* 2018;122(48):27495-27506.

19. Wang Y, Kang C, Zhang Z, et al. Evaluation of schiff-base covalent organic frameworks for co_2 capture: structure–performance relationships, stability, and performance under wet conditions. *ACS Sustain Chem Eng.* 2021;10(1):332-341.

20. Altundal OF, Altintas C, Keskin S. Can COFs replace MOFs in flue gas separation? high-throughput computational screening of COFs for CO_2/N_2 separation. *J Mater Chem A* 2020;8(29):14609-14623.

21. Fernandez M, Boyd PG, Daff TD, Aghaji MZ, Woo TK. Rapid and accurate machine learning recognition of high performing metal organic frameworks for CO_2 capture. *J Phys Chem Lett*. 2014;5(17):3056-3060.

22. Anderson R, Rodgers J, Argueta E, Biong A, Gómez-Gualdrón DA. Role of pore chemistry and topology in the CO_2 capture capabilities of mofs: from molecular simulation to machine learning. *Chem Mater.* 2018;30(18):6325-6337.

23. Fanourgakis GS, Gkagkas K, Tylianakis E, Froudakis GE. A universal machine learning algorithm for large-scale screening of materials. *J Am Chem Soc.* 2020;142(8):3814-3822.

24. Moosavi SM, Jablonka KM, Smit B. The role of machine learning in the understanding and design of materials. *J Am Chem Soc.* 2020;142(48):20273-20287.

25. Wu Y, Duan H, Xi H. Machine learning-driven insights into defects of zirconium metal–organic frameworks for enhanced ethane–ethylene separation. *Chem Mater.* 2020;32(7):2986-2997.

26. Altintas C, Altundal OF, Keskin S, Yildirim R. Machine learning meets with metal organic frameworks for gas storage and separation. *J Chem Inf Model.* 2021;61(5):2131-2146.

27. Rosen AS, Iyer SM, Ray D, et al. Machine learning the quantum-chemical properties of metal–organic frameworks for accelerated materials discovery. *Matter* 2021;4(5):1578-1597.

28. Yuan X, Li L, Shi Z, Liang H, Li S, Qiao Z. Molecular-fingerprint machine-learning-assisted design and prediction for high-performance MOFs for capture of NMHCs from air. *Adv Powder Mater.* 2022;1(3):100026.

29. Yan Y, Shi Z, Li H, et al. Machine learning and in-silico screening of metal–organic

frameworks for O_2/N_2 dynamic adsorption and separation. *Chem Eng J.* 2022;427:131604.

30. Wang Z, Zhou T, Sundmacher K. Interpretable machine learning for accelerating the discovery of metal-organic frameworks for ethane/ethylene separation. *Chem Eng J.* 2022;444:136651.

31. Cao X, He Y, Zhang Z, et al. Predicting of covalent organic frameworks for membrane-based isobutene/1,3-butadiene separation: combining molecular simulation and machine learning. *Chem Res Chinese Universities* 2022;38(2):421-427.

32. Yang P, Zhang H, Lai X, Wang K, Yang Q, Yu D. Accelerating the selection of covalent organic frameworks with automated machine learning. *ACS Omega* 2021;6(27):17149-17161.

33. Tong M, Lan Y, Yang Q, Zhong C. Exploring the structure-property relationships of covalent organic frameworks for noble gas separations. *Chem Eng Sci.* 2017;168:456-464.

34. Ongari D, Boyd PG, Barthel S, Witman M, Haranczyk M, Smit B. Accurate characterization of the pore volume in microporous crystalline materials. *Langmuir* 2017;33(51):14529-14538.

35. Willems TF, Rycroft CH, Kazi M, Meza JC, Haranczyk MJM, Materials M. Algorithms and tools for high-throughput geometry-based analysis of crystalline porous materials. *Micropor Mesopor Mater.* 2012;149(1):134-141.

36. Mayo SL, Olafson BD, III WAG. DREIDING: a generic force field for molecular simulations. *J Phys chem.* 1990;94(26):8897-8909.

37. Yan T, Lan Y, Tong M, Zhong C. Screening and design of covalent organic framework membranes for CO_2/CH_4 separation. *ACS Sustain Chem Eng.* 2019;7(1):1220-1227.

38. Wilmer CE, Snurr RQ. 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.

39. Potoff JJ, Siepmann JI. Vapor-liquid equilibria of mixtures containing alkanes, carbon dioxide, and nitrogen. *AIChE J.* 2001;47(7):1676-1682.

40. Tong MM, Yang QY, Xiao YL, Zhong CL. Revealing the structure-property relationship of covalent organic frameworks for CO_2 capture from postcombustion gas: a multi-scale computational study. *Phys Chem Chem Phys.* 2014;16(29):15189-15198.

41. Xin H, Zhou S, Xu S, et al. Functionalized linker to form high-symmetry adsorption sites

in micropore COF for CO_2 capture and separation: insight from GCMC simulations. *J Mater Sci.* 2022;57(11):6282-6292.

42. Peng X, Lin L-C, Sun W, Smit B. Water adsorption in metal-organic frameworks with open-metal sites. *AIChE J.* 2015;61(2):677-687.

43. Horn HW, Swope WC, Pitera JW, et al. Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *J Chem Phys.* 2004;120(20):9665-9678.

44. 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.

45. Zhao Y, Truhlar DG. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor Chem Acc.* 2007;120(1-3):215-241.

46. Petersson GA, Bennett A, Tensfeldt TG. A complete basis set model chemistry. I. The total energies of closed‐shell atoms and hydrides of the first‐row elements. *J Chem Phys.* 1988;89(4):2193-2218.

47. Petersson GA, Al‐Laham MA. A complete basis set model chemistry. II. Open‐shell systems and the total energies of the first‐row atoms. *J Chem Phys.* 1991;94(9):6081-6090.

48. Frisch M, Trucks G, Schlegel H, et al. *Gaussian 09, Revision B.01*. Wallingford, CT: Gaussian, Inc., 2010.

49. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in python. *J Mach Learn Res.* 2011;12:2825-2830.

50. Wang R, Zhong Y, Bi L, Yang M, Xu D. Accelerating discovery of metal-organic frameworks for methane adsorption with hierarchical screening and deep learning. *ACS Appl Mater Interfaces* 2020;12(47):52797-52807.

51. Pardakhti M, Moharreri E, Wanik D, Suib SL, 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. González-Zamora E, Ibarra IA. CO_2 capture under humid conditions in metal–organic

frameworks. *Mater Chem Front.* 2017;1(8):1471-1484.

53. Zhang Y, Zhao Z, Zheng J. CatBoost: A new approach for estimating daily reference crop evapotranspiration in arid and semi-arid regions of Northern China. *J Hydrol.* 2020;588:125087.

54. Lundberg S.M, Lee S-I. A unified approach to interpreting model predictions, in: Proceedings of the 31st International Conference on Neural Information Processing Systems, 2017:4768–4777.

55. Lan Y, Yan T, Tong M, Zhong C. Large-scale computational assembly of ionic liquid/MOF composites: synergistic effect in the wire-tube conformation for efficient CO_2/CH_4 separation. *J Mater Chem A* 2019;7(20):12556-12564.