RSC Advances

PAPER

View Article Online
View Journal | View Issue

Cite this: *RSC Adv.*, 2025, **15**, 36778

Accurate DDEC-charge-guided screening of high-performance metal–organic frameworks for SF_6/N_2 separation

Ruihan Wang, Shiqi Wang and Qianji Han*

Achieving high SF_6 uptake and SF_6/N_2 selectivity is a key challenge in gas separation. High-throughput computational screening is an efficient strategy to identify high-performing adsorbents. However, these candidates may be overlooked because most studies rely on empirical partial charge assignments. In this study, we present a data-driven workflow that integrates accurate density-derived electrostatic and chemical (DDEC) partial atomic charges into grand canonical Monte Carlo (GCMC) simulations to accelerate the discovery of high-performance MOFs for SF_6/N_2 separation. By screening the quantum-chemical metal–organic framework (MOF) database, several top-performing candidates with high SF_6 uptake and selectivity were identified. The key features for efficient separation were open metal sites, parallel aromatic surfaces, uncoordinated nitrogen atoms, and metal–oxygen–metal bridges. A machine learning model trained on the DDEC-based GCMC results achieved excellent predictive performance (coefficient of determination = 0.968, mean absolute error = 0.281 mmol g^{-1}) and enabled rapid screening of 154 144 MOFs within 50 min. Zn-TCPP was selected for validation *via* density functional theory calculations, confirming the reliability of the proposed workflow. This study illustrates how quantum-chemical datasets facilitate high-throughput material discovery for challenging separations.

Received 23rd August 2025
Accepted 25th September 2025

DOI: 10.1039/d5ra06266g

rsc.li/rsc-advances

Introduction

Sulfur hexafluoride (SF_6) is widely used in the modern electrical industry because of its excellent chemical stability, nontoxicity, and arc-quenching capabilities.[1–3] However, there is growing concern about its environmental impacts. SF_6 is a potent greenhouse gas, with a global warming potential more than 23 000 times greater than that of carbon dioxide and an atmospheric lifetime longer than 3000 years.[4] To enhance the insulation system sustainability and cost-effectiveness while maintaining the good insulation performance, SF_6 is often diluted with nitrogen (N_2), with typical mixtures containing only 10–20% SF_6 by volume.[5] Nevertheless, efficient SF_6 recovery from its mixtures with N_2 is still a substantial technical challenge. Thus, advanced separation strategies are urgently needed for efficient recovery and mitigation of the environmental impacts of SF_6. In contrast to the energy-intensive cryogenic distillation, adsorption-based methods offer a more economical and energy-efficient solution for SF_6 separation from N_2. Therefore, identifying a high-performance adsorbent for effective SF_6/N_2 separation is essential.

Metal–organic frameworks (MOFs) are excellent candidates for high-performance gas separation applications because of their wide range of pore sizes, large surface areas, exceptionally high porosities, and tunable structures.[6–8] Using MOFs for SF_6/N_2 separation has been extensively studied.[9–11] Recent advances have demonstrated remarkable performance under industrially relevant conditions ($SF_6:N_2$ = 10 : 90, 1.0 bar). For example, BUT-53 exhibits a record selectivity of 2485 with an SF_6 uptake of 3.55 mmol g^{-1},[12] whereas Cu–MOF–NH_2 delivers a higher uptake of 7.88 mmol g^{-1} with a selectivity of 266.2.[5] Although the SF_6/N_2 selectivity of MOFs has reached satisfactory levels, SF_6 uptake should be further improved. However, there is an intrinsic trade-off relationship between capacity and selectivity, which is a key challenge in the design of porous materials. Achieving simultaneous optimization of both properties is the key for MOF application in industrial-scale SF_6/N_2 separation.

Currently, there are over 100 000 synthesized MOFs,[13] among which high-performance MOFs suitable for industrial-scale SF_6/N_2 separation should be identified. Nevertheless, experimental testing of this enormous number of MOFs is technically challenging and economically infeasible. To overcome this challenge, high-throughput computational screening can be used to efficiently identify high-performance MOFs from large material databases.[14–16] However, there are only few computational screening studies on SF_6/N_2 separation.[5,17,18] For example, Ren *et al.* employed grand canonical Monte Carlo (GCMC) simulations to evaluate 2513 experimentally synthesized MOFs and identified Cu–MOF–NH_2 as a promising candidate.[5] In another case, an integrated strategy combining high-throughput simulation with machine learning (ML) was used to screen more

Chemical Engineering College, Hebei Normal University of Science & Technology, Qinhuangdao, Hebei 066600, PR China. E-mail: hanqianji4193@hevttc.edu.cn

than 25 000 MOFs for SF_6/N_2 adsorption and separation.[18] These studies highlight the efficiency of computational and data-driven approaches in identifying high-performing MOFs. Moreover, the predictive accuracy of these computational screening methods help determine the success of subsequent experimental validations, particularly when experimental resources are limited.

To improve the reliability of the simulation-based screening, we previously optimized the force field parameters for CH_4 adsorption,[19] resulting in considerably improved computational accuracy. In another study, we applied data mining techniques to construct a high-quality experimental dataset for C_2H_6/C_2H_4 separation,[20] which facilitated the development of ML models with enhanced predictive performance. However, for small hydrocarbons, the commonly used transferable potentials for phase equilibria (TraPPE) force field[21] omits partial atomic charges because electrostatic interactions have a limited effect on their adsorption behavior. In contrast, in the cases of SO_2, CO_2, and SF_6, accurate treatment of electrostatic interactions is essential for reliable predictions.[22] This distinction underscores the critical need for accurate charge assignment schemes in the modeling of MOF–guest interactions involving these molecules.

Recent developments for overcoming this challenge introduced first-principles-based methods for the partial charge assignment in MOFs. Among these methods, the density-derived electrostatic and chemical (DDEC)[23–26] approach is the most accurate method because it can generate chemically consistent and transferable charges for both metal nodes and organic linkers.[27] Despite its recognized accuracy, DDEC remains underutilized in large-scale computational screenings. Most high-throughput computational screenings for SF_6/N_2 separation use empirical charge assignment methods such as charge equilibration[28] and its extended variant (EQeq).[29] These methods are computationally efficient and capable of producing reliable screening results; however, their empirical nature often limits their ability to accurately reproduce reference DDEC charges, especially in cases involving transition metals in MOFs.[30] Nazarian *et al.*[31] demonstrated that EQeq overestimates the charges of alkali and rare earth metals and underestimates those of alkaline earth metals, indicating the inherent bias introduced by parameterization.

To overcome these limitations, large-scale quantum chemical screenings have emerged for generating reliable and reproducible reference data. In particular, recent breakthroughs in this area have focused on constructing quantum chemical MOF databases that incorporate electronic structure information obtained using density functional theory (DFT) calculations. Major examples of these datasets include the Open DAC 2023 (ODAC23)[32] and Quantum MOF (QMOF)[33] datasets, which provide comprehensive annotations for properties such as DDEC partial atomic charges and band gaps. These quantum chemically derived resources offer a strong foundation for improving the reliability of electrostatic modeling in MOFs. Using these databases, accurate electrostatic interactions can be incorporated into high-throughput screenings which enhances the predictive accuracy of theorical methods for adsorption and separation performance of SF_6 on MOFs.

In this work, we propose a data-driven workflow for identifying high-performance MOFs for SF_6/N_2 separation, making use of the enhanced electrostatic accuracy provided by DDEC partial atomic charges (Fig. 1). GCMC simulations were performed on the QMOF dataset, leading to the identification of a subset of top-performing MOFs that exhibits high SF_6 uptake and selectivity. Subsequent analysis revealed that the top-performing MOFs exhibit several main structural features, including open metal sites (OMSs), parallel aromatic surfaces, uncoordinated nitrogen atoms, and metal–oxygen–metal bridges (MOMBs). Furthermore, we developed an ML model that can directly predict SF_6 uptake based on the MOF structure. The ML model showed strong predictive accuracy on the test set, with a coefficient of determination (R^2) of 0.968 and a mean absolute error (MAE) of 0.281 mmol g^{-1}. It was then applied to a database of 154 144 MOFs, predicting additional candidates with favorable properties. DFT calculations further demonstrated its practical potential for SF_6/N_2 separation.

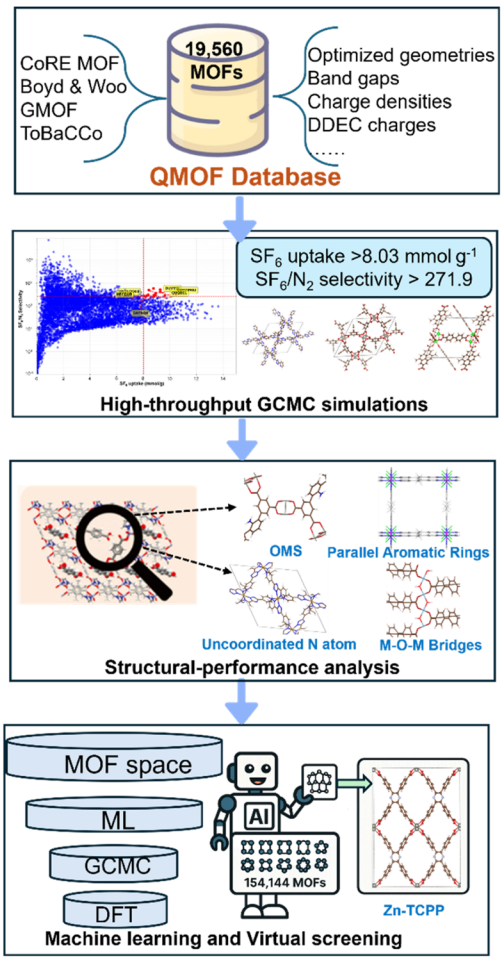

Fig. 1 Data-driven workflow of high-throughput GCMC simulations, structure–performance analysis, ML and virtual screening employed in this work.

Computational methods

MOF database

Herein, the QMOF dataset,[33] which initially comprised 20 372 MOF structures, was used as a structural source for the high-throughput computational screening. To improve the structural integrity, the metal oxidation state automated error checker (MOSAEC) algorithm,[34] which systematically identifies and removes chemically implausible structures based on the evaluation of the formal oxidation states of the metal centers. Following MOSAEC filtration, 19 560 structurally validated MOFs were selected from the QMOF dataset. The physical characteristics, such as the largest cavity diameter (LCD), gravimetric surface area (GSA), void fraction, and accessible volume (AV), of each MOF in the database were determined using the Zeo++ software package.[35] The presence of OMSs in each MOF was determined using the Python-based algorithm developed by Haldoupis *et al.*[36] The structural topology and organic linker identities of the MOF structures were analyzed using the MOFid package.[37]

GCMC simulations

In all molecular simulations, the MOF–guest interactions were modeled as the sum of the Lennard-Jones (LJ) and coulombic electrostatic contributions. The LJ parameters for the framework atoms were assigned based on the universal force field.[38] The N_2 molecule was modeled using the TraPPE force field[21] and the SF_6 molecule was represented based on the approach of Dellis and Samios,[39] which has been widely adopted for accurately capturing the properties of SF_6 in confined porous environments. A summary of the force field parameters is provided in Tables S1 and S2. The atomic partial charges for all MOF structures were assigned using the DDEC method based on data from the QMOF[33] database, ensuring consistency in the electrostatic interaction representation throughout all subsequent simulations.

RASPA-2.0 (ref. 40) was used to perform all GCMC simulations. In the simulations, 20 000 initialization cycles and 50 000 production cycles were conducted. During the simulations, four types of the Monte Carlo moves were defined: translation, rotation, insertion, and swap. The LJ interactions were determined using spherical cutoff of 14 Å with a long-range correction, whereas long-range electrostatic interactions were treated based on the Ewald summation[41] method. Lorentz–Berthelot mixing rules were employed to evaluate the interactions between different atom types. The simulation cell of each MOF was constructed by replicating the unit cell in the three dimensions so that its linear sizes were at least twice the spherical cutoff. The SF_6 uptake was obtained using single-component GCMC simulations conducted at 298 K and 1.0 bar. The SF_6/N_2 selectivity was calculated based on binary mixture adsorption simulations with a $SF_6:N_2$ molar composition of 10 : 90 under the same conditions, according eqn (1).

$$S_{SF_6/N_2} = \frac{(N_{SF_6}/N_{N_2})}{y_{SF_6}/y_{N_2}} \quad (1)$$

where N_{SF_6} and N_{N_2} are the adsorbed uptakes (mmol g^{-1}) of SF_6 and N_2, respectively, obtained based on the GCMC simulations of the binary gas mixture, and y_{SF_6} and y_{N_2} are the bulk mole fractions of SF_6 and N_2, respectively.

Machine learning model

To accurately predict the SF_6/N_2 separation properties of the MOFs, our previously developed ML architecture, *i.e.*, MOF-CGCNN, was employed.[19] This model was specifically designed to better capture the coordination geometry of the metal centers and the topological diversity present in the MOF structures. MOF-CGCNN incorporates chemically meaningful descriptors that reflect both local bonding environments and extended periodic connectivity, enabling more accurate learning of the structure–property relationships across numerous frameworks. A detailed description of the MOF-CGCNN model architecture and its theoretical basis is reported in our previous publication.[19] The model was trained to minimize the mean square error (MSE) between the predicted and GCMC-calculated values. Hyperparameter optimization was performed using Bayesian optimization, and the final optimized hyperparameters are summarized in Table S3. Model training was conducted using the PyTorch framework for 2000 epochs. Early stopping was applied based on the validation loss to prevent overfitting.

DFT calculations

To study the SF_6/N_2 selectivity of the representative MOF structure at the molecular level, dispersion-corrected DFT calculations were conducted. All periodic DFT computations were performed using the CP2K simulation package (version 2024.1),[42] employing the Perdew–Burke–Ernzerhof exchange-correlation functional combined with the hybrid Gaussian and plane-wave approaches. For the expansion of the valence electron wavefunctions, a double-ζ basis set augmented with polarization functions was used. Core electrons were treated using Goedecker–Teter–Hutter pseudopotentials.[43,44] Dispersion corrections were incorporated in all DFT calculations using the Grimme-D3 method.[45] The adsorption energies of the MOF-gas complexes ($\Delta E_{MOF\text{-}gas}$) were calculated using eqn (2).

$$\Delta E_{MOF\text{-}gas} = E_{MOF\text{-}gas} - E_{MOF} - E_{gas} \quad (2)$$

where $E_{MOF\text{-}gas}$ is the total energy of the MOF-gas complexes, E_{MOF} is the energy of the periodic framework structure, and E_{gas} is the energy of a gas molecule (SF_6 or N_2) within the same periodic box.

Results and discussion

Evaluation of the DDEC partial charges and force field parameters

To validate the reliability of our GCMC simulation protocol, the simulated performance was benchmarked against experimental data reported by Ren and co-workers.[5] They determined a top-performing candidate, *i.e.*, Cu-MOF-NH_2 (computation-ready,

experimental metal–organic framework (CoRE MOF) database code: GAFRUD), based on high-throughput GCMC simulations on the CoRE MOF 2019 database and experimentally confirmed its performance.[46] Under standard conditions (298 K, 1.0 bar) at an industrial gas composition ($SF_6 : N_2 = 10 : 90$), Cu–MOF–NH_2 demonstrated an SF_6 uptake of 7.88 mmol g^{-1} and an outstanding SF_6/N_2 selectivity of 266.2. Our GCMC simulations reproduced these values with high accuracy (the computed SF_6 uptake = 8.03 mmol g^{-1} and selectivity = 271.9), confirming the quantitative predictive power of the GCMC simulation workflow.

To further assess the applicability of the method to different MOF topologies and chemistries, we performed additional GCMC simulations under the same conditions for two well-characterized benchmark materials: UiO-66 (ref. 47) and HKUST-1.[48] The computed SF_6 uptakes were 1.79 and 5.22 mmol g^{-1}, respectively, which are close to the corresponding experimental values (1.67 and 4.98 mmol g^{-1}, respectively). The previously reported selectivities of UiO-66 and HKUST-1, which were obtained *via* IAST predictions, are 127.8 and 70.4, respectively, whereas our GCMC-derived values were 248.8 and 80.9, respectively. Although the GCMC-derived selectivity values are not fully consistent with the quantitative predictions from IAST, the computed SF_6/N_2 selectivity trend, in which Cu–MOF–NH_2 exhibits higher selectivity than UiO-66 and HKUST-1, agrees well with experimental observations. The observed discrepancy can be ascribed to the intrinsic limitations of IAST, which relies on the assumptions of homogeneous pore filling and uniform adsorbate accessibility. Collectively, these results confirm the robustness of the employed force field (Dellis–Samios model)[39] and the accuracy of the DDEC partial atomic charges in describing the MOF–SF_6 interactions. This benchmarking analysis confirms that the computational workflow provides a reliable foundation for large-scale computational screening of MOF candidates for SF_6/N_2 separation applications.

High-throughput computational screening

Building on this validated computational workflow, we conducted high-throughput GCMC screening on the QMOF database. The QMOF database incorporates both experimentally reported structures[36] and numerous hypothetical frameworks, including those from the Boyd & Woo,[49] Genomic MOF (GMOF),[50] and topologically based crystal constructor (ToBaCCo)[51] databases. The QMOF database contains structures that were optimized using high-throughput periodic DFT and include DDEC partial atomic charges. Structural quality control was conducted using the MOSAEC algorithm,[34] after which 19 560 MOFs were selected from the QMOF for screening.

To ensure consistency in the performance benchmarking, Cu–MOF–NH_2 was selected as the reference material. Its computed SF_6 uptake and SF_6/N_2 selectivity were 8.03 mmol g^{-1} and 271.9, respectively. These benchmark values are indicated by red dashed lines in Fig. 2 and were used to facilitate the identification of high-performance candidates. Cu–MOF–NH_2 achieves the objective of this study, which is to identify MOFs

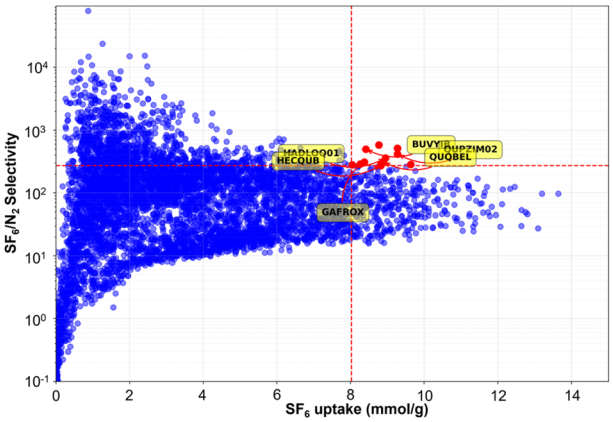

Fig. 2 (a) High-throughput GCMC screening of the QMOF dataset based on SF_6/N_2 selectivity and SF_6 uptake at 298 K and 1.0 bar.

with high SF_6 uptake and competitive separation performance. To identify MOFs with ultra-high selectivities, we also explored an alternative threshold region defined by a SF_6/N_2 selectivity higher than 1000 and uptake greater than 3.60 mmol g^{-1}. For example, the DOYBEA MOF[52] exhibits SF_6 uptake and SF_6/N_2 selectivity of 3.79 mmol g^{-1} and 1943.9, respectively. Fig. S1 shows the positions of all MOFs meeting this ultra-selective criterion, and their details are listed in Table S4.

The qmof-2634ae7 MOF, which was obtained from the GMOF subset in the QMOF dataset, exhibited the highest SF_6 uptake (13.65 mmol g^{-1}) in this database with a selectivity of 96.7. In comparison, the LEZPAJ MOF exhibited an exceptionally high selectivity of 79 019.1, with a limited uptake (0.88 mmol g^{-1}). This inverse relationship between the uptake and selectivity reflects a common trade-off relationship observed across the dataset and highlights the challenge in simultaneously optimizing both performance metrics. From the 19 560 MOFs evaluated in the QMOF dataset, 12 candidates fell within the upper-right quadrant of the uptake-selectivity plot (Fig. 2). For clarity, the structures are labeled with their CoRE MOF codes. HADLOQ01 (ref. 53) corresponds to Co(bdp), which is a flexible MOF that exhibits an expanded structure only under high-pressure conditions (~20 bar). However, under the pressure adopted for the SF_6 adsorption in this study (1 bar), the structure remains in a collapsed conformation with negligible accessible pore volume (AV), making it effectively inaccessible to SF_6 molecules. Therefore, HADLOQ01 represents a false positive in the screening process due to the failure to capture the pressure-dependent flexibility in the static structural model. Other flexible MOFs, such as QUPZIM02 (ref. 54) and QUQBEL,[54] were similarly flagged as top candidates but exhibited collapsed structures under the studied SF_6 adsorption conditions, leading to overestimated performances. BUVYIB and HECQUB were obtained from the experimental CoRE MOF database, while the others were drawn from the hypothetical subset. It should be emphasized that the use of rigid-framework GCMC can lead to false positives when dealing with pressure-responsive MOFs. For example, HADLOQ01, QUPZIM02, and QUQBEL were flagged as top candidates in our initial screening

but were later identified as collapsed structures under the studied SF_6 adsorption conditions, resulting in overestimated performances. This underscores the importance of carefully validating the structural stability of promising candidates after the high-throughput stage. For any individual candidate flagged as potentially flexible, we recommend performing one or more of the following validation steps: (i) literature or experimental checks for known open/closed phases; (ii) geometric screening of accessible pore volume and limiting pore diameter at the target conditions.

To further validate the screening results, we examined the top candidate MOFs in detail. None of these candidates have been experimentally reported to date. Interestingly, one candidate, GAFROX,[46] exhibits the same topological framework as the experimentally studied Cu–MOF–NH_2, with zinc replacing copper as the metal center. Given this close structural similarity, GAFROX is very likely to represent a high-performance material for SF_6/N_2 separation. A complete list of the top candidates identified *via* QMOF screenings is provided in Table 1.

To study the structural features affecting the performance of top-performing MOFs, we systematically analyzed the 12 candidates listed in Table 1. As summarized in Fig. 3, these MOFs consistently exhibit a set of structural features that likely affect their separation performance: (i) OMSs, (ii) parallel aromatic rings, (iii) uncoordinated nitrogen atoms, and (iv) MOMBs.

Among these structural features, OMSs is critical binding structural motifs that can enhance the gas–MOF interactions (Fig. 3a). OMSs facilitate the adsorption of gas molecules, such as CO_2 and C_2H_4, owing to their ability to engage in Lewis acid–base interactions and π-back bonding mechanisms.[55] The framework of GAFROX exhibits readily accessible OMSs, which can serve as strong SF_6 adsorption sites. HADLOQ01, QUPZIM02, and QUQBEL exhibit expanded structureonly under high pressure and highly collapsed pore structures with limited accessible porosity under ambient pressure. However, important structural characteristics can be retrieved from their high-pressure structures. These MOFs exhibit an arrangement of nearly parallel aromatic rings, with LCDs close to 10 Å (Fig. 3b). Comparable structural motifs with similar aromatic spacing are observed in qmof-1cc6d2b, qmof-7c9951b, and qmof-77a0b3f.

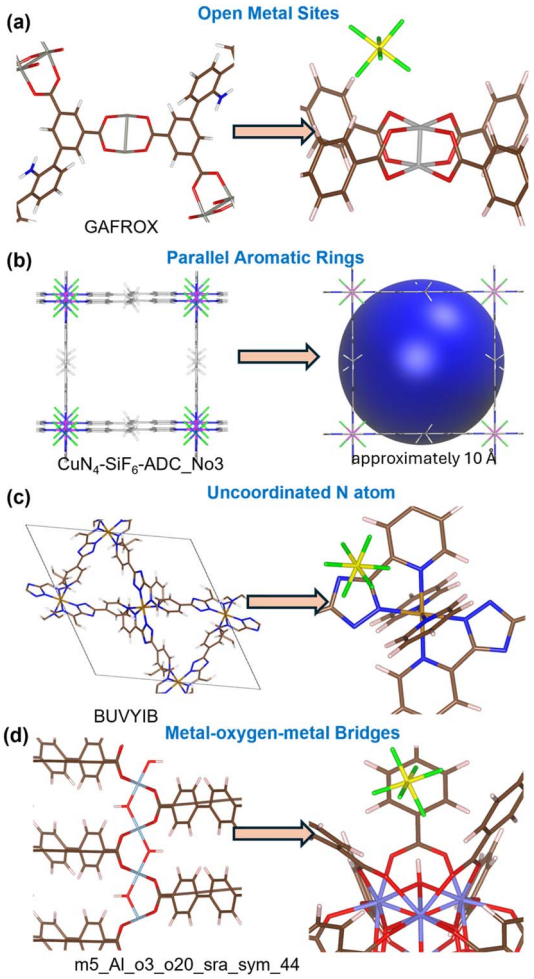

Fig. 3 Typical structural features commonly found in high-performing MOFs: (a) open metal sites; (b) parallel aromatic rings; (c) uncoordinated nitrogen atoms; and (d) metal–oxygen–metal bridges.

Table 1 Top-performing MOFs for SF_6 adsorption (with SF_6 uptakes > 8.03 mmol g^{-1} and SF_6/N_2 selectivity > 271.9) at 298 K and 1 bar. The structural features contributing to the high performance are as follows: OMSs = open metal sites, PAR = parallel aromatic rings (PAR 10.1 Å indicates a large cavity diameter of approximately 10.1 Å), UN = uncoordinated nitrogen atoms, MOMBs = metal–oxygen–metal bridges. All listed structures have passed structural validation using the MOSAEC algorithm

MOF name	SF_6 uptake (mmol g^{-1})	SF_6/N_2 selectivity	Structural features
qmof-1cc6d2b	9.28	512.4	PAR 10.3 Å
qmof-7c9951b	8.77	578.7	PAR 10.3 Å
qmof-77a0b3f	8.83	287.6	PAR 10.1 Å
QUPZIM02	9.28	414.8	PAR 10.4 Å
HADLOQ01	8.95	359.3	PAR 10.1 Å
QUQBEL	8.89	312.1	PAR 9.8 Å
GAFROX	8.38	308.6	OMSs
BUVYIB	8.42	498.2	UN
HECQUB	8.83	272.3	UN
qmof-b890eba	8.26	286.5	MOMBs
qmof-94124d1	8.05	280.0	MOMBs
qmof-6f5fabe	9.64	284.5	MOMBs

The recurrence of this geometry suggests that as eparation of approximately 10 Å between aromatic surfaces may serve as a reliable structural descriptor for efficient SF_6 accommodation and molecular packing in porous materials.

The presence of uncoordinated nitrogen atoms (Fig. 3c), such as in BUVYHE and HECQUB, is another recurring structural feature. Introduction of electron-deficient heteroaromatic rings reduces π-electron density of the ring plane and produces regions of positive local electrostatic potential distribution above the ring. These positively biased surface regions interact favorably with the negatively polarized fluorine shell of SF_6, thereby increasing host–guest binding.[56]

In addition, MOMBs (Fig. 3d) are frequently observed in structures including qmof-6f5fabe, qmof-b890eba, and qmof-94124d1. These MOMBs have previously been identified as performance-enhancing features in large-scale computational screenings for CO_2 capture[49] and may similarly contribute to SF_6 adsorption. Therefore, the simultaneous occurrence of OMSs, parallel aromatic surfaces with optimal spacing, uncoordinated nitrogen sites, and MOMBs can define the structural frameworks of high-performance SF_6/N_2 adsorbents.

Analysis of the structural–performance relationships

To clarify the structure effects on the SF_6/N_2 separation performance of MOFs, we systematically examined the relationships between critical performance metrics (SF_6 uptake and SF_6/N_2 selectivity) and four key pore descriptors: GSA, AV, LCD, and void fraction (Fig. 4). For instance, the highest SF_6 uptake was observed at GSA of 3700–4300 $m^2\ g^{-1}$ (Fig. 4a). Increasing the GSA beyond this range, does not enhance SF_6 adsorption. This plateau effect is likely due to a reduction in the effective interactions between the MOF and the SF_6 molecules in structures with extremely large surface areas.

In contrast, SF_6/N_2 selectivity reaches its maximum within a considerably lower GSA range (1190–1690 $m^2\ g^{-1}$) (Fig. 4a). Within this lower range, the compact pore structure enhances the molecular discrimination and reduces nonspecific N_2 adsorption. A similar contradiction was observed in the relationship between the performance and the AV. As shown in Fig. 4b, SF_6 uptake increased with AV, reaching an optimal range at 1.21–1.51 $cm^3\ g^{-1}$. Within this range, the pores provide sufficient space to accommodate numerous SF_6 molecules while maintaining favorable interactions. However, SF_6/N_2 selectivity was maximized at considerably lower AV values (approximately 0.19 $cm^3\ g^{-1}$, Fig. 4b). These tighter AV promote size exclusion and reduce N_2 adsorption, thereby enhancing selectivity. The effect of the LCD further confirms this inherent trade-off relationship. High SF_6 uptake is observed at LCD of 9.6–11.6 Å (Fig. 4c). In contrast, the highest selectivity is achieved at LCD of approximately 6.1 Å, where the pore size is only sufficient to admit SF_6 while restricting N_2 (Fig. 4c). A similar trade-off relationship was observed with the void fraction. As shown in Fig. 4d, MOFs with void fractions of approximately 0.75 exhibit high SF_6 uptake due to the availability of large internal free volumes. In contrast, selectivity is maximized in materials with void fractions between 0.37 and 0.47 because the tighter pore space enhances molecular sieving and restricts nitrogen transport. These results highlight an important challenge in the rational design of MOFs for SF_6/N_2 separation. Structural features that facilitate high adsorption capacity often decrease selectivity, whereas configurations that enhance selectivity tend to reduce uptake.

To assess whether high-performing MOFs exhibit distinct linker types or topological frameworks, we employed the MOFid package[37] to decompose each structure into its fundamental building units. Owing to the limited number of MOFs (only 12) that outperform Cu–MOF–NH_2 in the QMOF databases, statistical analysis at this level was not feasible. Therefore, we relaxed the selection criteria to include MOFs with an SF_6 uptake and SF_6/N_2 selectivity greater than 5.22 mmol g^{-1} and 80.9, respectively. These values are comparable to those of HKUST-1, a well-established benchmark material. Based on these revised criteria, 677 high-performing MOFs from the QMOF database were identified. Fig. 4e and f illustrate the most frequently occurring linkers and topologies among these high-performing materials in the QMOF dataset. Several linkers were functionalized with groups such as –NH_2, –F, and –Cl, which are commonly introduced to tune the chemical environment of the framework. These functional groups can affect factors such as polarity, electrostatic interactions, and adsorption site specificity. For example, the introduction of an amino group in YTU-29-NH_2 leads to a marked improvement in SF_6/N_2 separation efficiency compared to the YTU-29 parent framework.[57] The most common topologies among the high-performing MOFs were rna, pcu, acs, and sql. These results suggest that certain linkers and topologies are strongly associated with superior SF_6/N_2 separation performance, and hence, they can serve as fundamental structural motifs in inverse design frameworks.

Machine learning and virtual screening

To enable more extensive virtual screening of MOFs, it is essential to expand the number of structures with assigned DDEC atomic charges. Although collectively, the QMOF datasets provide DDEC-derived charges for tens of thousands of MOFs, a large fraction of the known MOF chemical space remains uncharacterized at the quantum mechanical level. The high computational cost associated with DFT-based charge assignment limit the feasibility of high-throughput screening and property prediction. To overcome this challenge, several recent studies have applied ML to predict atomic charges at the DDEC level,[58–61] thereby reducing the computational demand while retaining the charge assignment reliability. Herein, we adopted a different strategy involving directly predicting the SF_6 adsorption properties of the MOFs rather than using ML to reproduce the electrostatic parameters as an intermediate. This approach eliminates the need for a separate charge-assignment step and offers a more efficient and scalable solution for large-scale screening.

Building upon this strategy, we employed ML to develop predictive models that directly estimate SF_6/N_2 separation performance based on structural information. Our prior results indicate that the predictive accuracy and transferability of ML

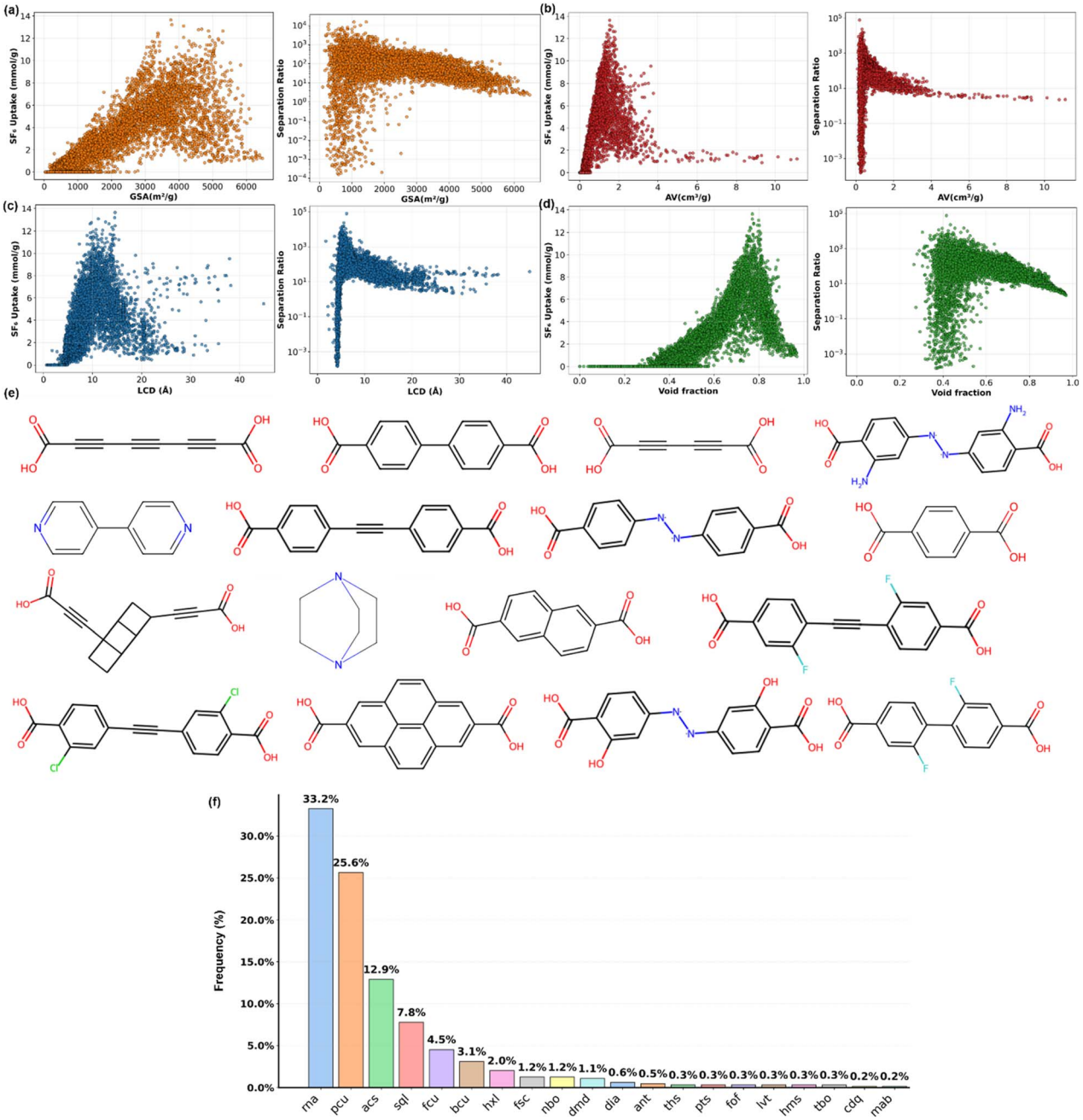

Fig. 4 Correlation between MOF structural features and their performance. SF_6 uptake and SF_6/N_2 selectivity analyzed against four key pore descriptors: (a) GSA, (b) AV, (c) LCD, and (d) void fraction. (e and f) Representative building unit statistics of the high-performance MOFs in the QMOF database. (e) Top 16 most frequent linkers and (f) top 20 most frequent topologies.

models strongly depend on the quality and reliability of the training set.[19] In this study, ML models were developed based on the DDEC-based GCMC results. The first model predicts pure-component SF_6 uptake, and the second estimates SF_6/N_2 selectivity for a 1 : 9-gas mixture. These models are hereafter referred to as the SF_6 uptake model and the SF_6/N_2 selectivity model, respectively. To ensure balanced coverage of adsorption values, the QMOF dataset (excluding structures with LCD < 4.7 Å) was split into training, validation, and test sets (80 : 10 : 10) using a label-based binning algorithm based on SF_6 adsorption uptake, rather than a purely random split. This strategy guarantees that all adsorption ranges are represented in each subset. The distribution after binning is shown in Fig. S2.

We employed MOF-CGCNN,[19] which has been shown to effectively capture local atomic environments. The predictive performance of each model was evaluated using standard regression metrics on the test set, including R^2, MAE, and root MSE (RMSE). As illustrated in Fig. 5, the SF_6 uptake model demonstrated excellent predictive performance, achieving an R^2

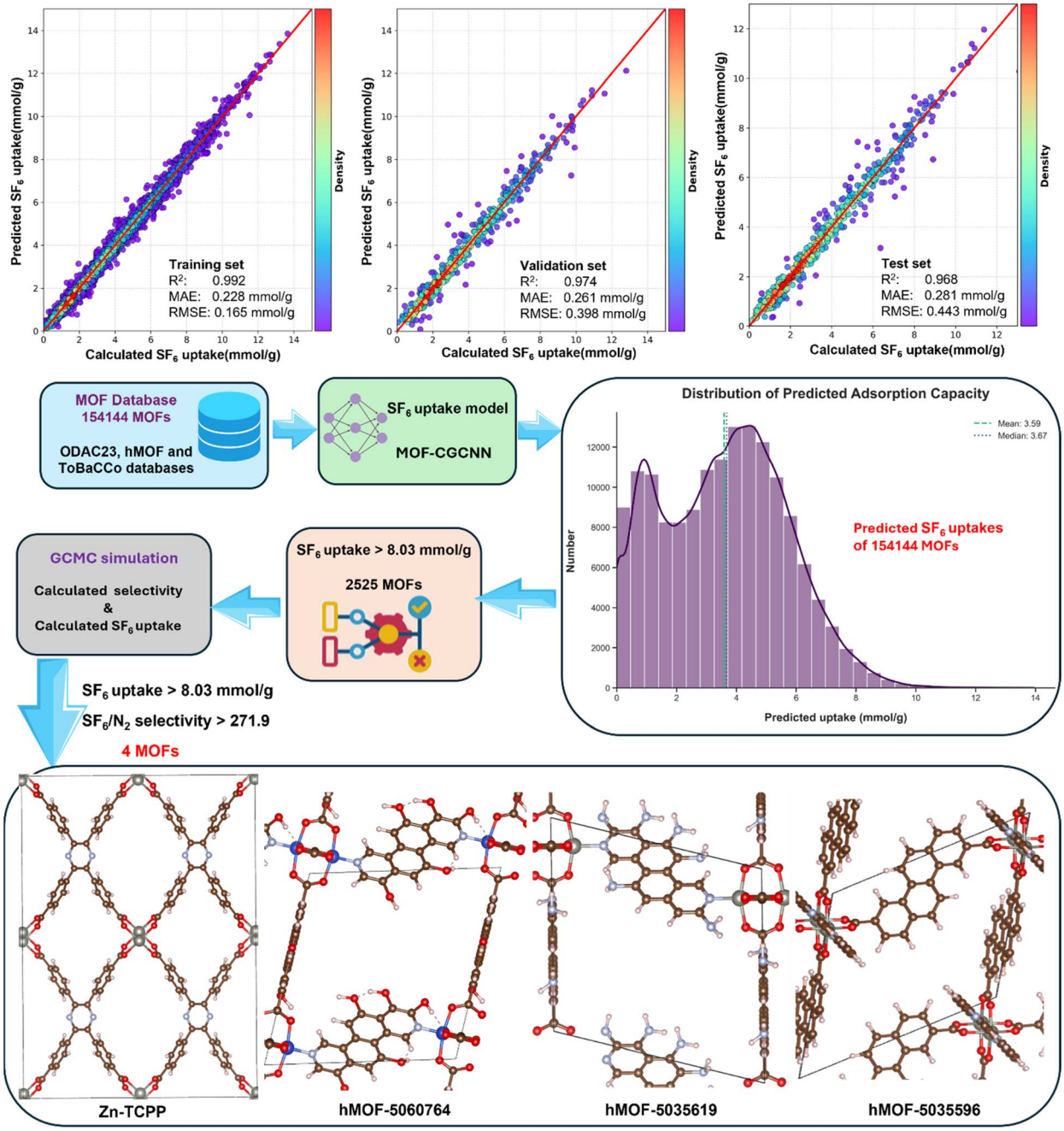

Fig. 5 Accuracy representations for the MOF-CGCNN models and a schematic of the hierarchical screening workflow.

of 0.968, with MAE and RMSE of 0.281 and 0.443 mmol g^{-1}, respectively.

In contrast, the SF_6/N_2 selectivity model exhibited considerably lower accuracy, showing an R^2 of 0.474 and considerably larger errors (MAE = 63.37 and RMSE = 86.26). Extensive hyperparameter optimization failed to improve the selectivity model to a level suitable for practical deployment. This may be attributed to the fact that selectivity is a ratio of two uptake values, which amplifies noise and compound errors from both predictions, rendering its accurate modeling inherently more difficult.

To demonstrate the practical applicability of the ML model, we performed ML-based virtual screening on a large-scale dataset comprising 154 144 MOFs from the ODAC23,[32] hMOF,[62] and ToBaCCo[51] databases. Owing to the limited predictive accuracy of the SF_6/N_2 selectivity model, we adopted a hierarchical screening workflow to efficiently identify high-performing candidates (Fig. 5). In the first stage, the SF_6 uptake model was applied to the full set of 154 144 MOFs. This large-scale prediction task was completed in approximately 50 min using a 40-core server and resulted in the identification of 2525 structures with predicted SF_6 uptakes exceeding

8.03 mmol g^{-1}. These 2525 candidates were then subjected to GCMC simulations to obtain more reliable estimates of their SF_6 uptake and SF_6/N_2 selectivity as shown in Fig. 5.

By filtering out over 98% of the structures before the first stage, this approach drastically reduced the computational cost. The candidates were then ranked according to their calculated SF_6 uptake and selectivity, identifying three hypothetical MOFs and one experimentally reported structure (CoRE MOF code: NAWXER, hereafter referred to as Zn-TCPP)[63] that exhibited SF_6 uptake greater than 8.03 mmol g^{-1} and selectivity higher than 271.9. Zn-TCPP is built from paddlewheel-type $Zn_2(COO)_4$ clusters, each coordinated equatorially by four 2,3,5,6-tetrakis(4-carboxyphenyl) pyrazine ($TCPP^{4-}$) linkers. Axial DMF ligands can be removed to generate OMSs. GCMC calculations for Zn-TCPP yielded SF_6 uptake of 9.67 mmol g^{-1} and SF_6/N_2 selectivity of 335.77, placing it among the top-performing materials in the 154 144 MOF dataset. Detailed SF_6 uptake and selectivity values for the three hypothetical MOFs are provided in Table S5. The corresponding GCMC calculated adsorption isotherms are displayed in Fig. 6. The integration of predictive modeling with large-scale screening offers a powerful route for exploring the chemical space. The full datasets and corresponding source code are available on GitHub at **https://github.com/ruihwang/SF6_MOFCGCNN/**.

Among the top-performing candidates, hMOF-5035619 and hMOF-5060764 feature electron-donating functional groups, such as $-NH_2$ and –OH, grafted onto their aromatic linkers. Both structures also exhibit parallel aromatic rings with linker center distances of 10.3 and 9.7 Å, respectively. These structural characteristics confirm the established structure–performance relationship, in which an optimal spacing of ~10 Å between aromatic surfaces correlates with enhanced SF_6/N_2 separation efficiency. The framework of Zn-TCPP prominently shows OMSs and uncoordinated nitrogen atoms, which are two structural elements previously identified as key contributors to SF_6/N_2 selectivity. Given that the synthesis of hypothetical MOFs remains to be further explored, we primarily focused on Zn-TCPP that has already been experimentally reported.[63]

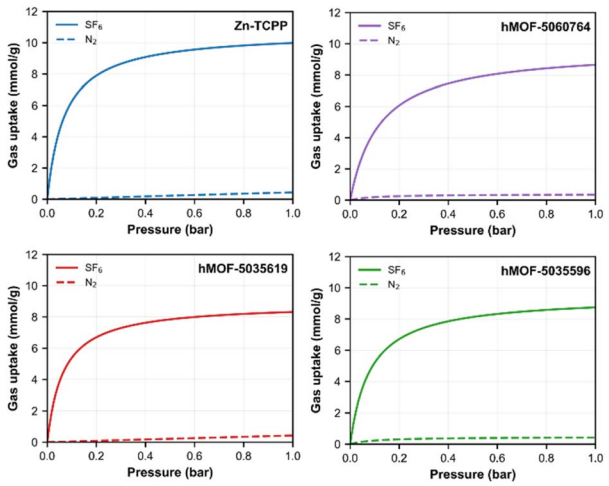

Fig. 6 GCMC calculated adsorption isotherms of the five top-performing MOFs identified from the ML-based virtual screening.

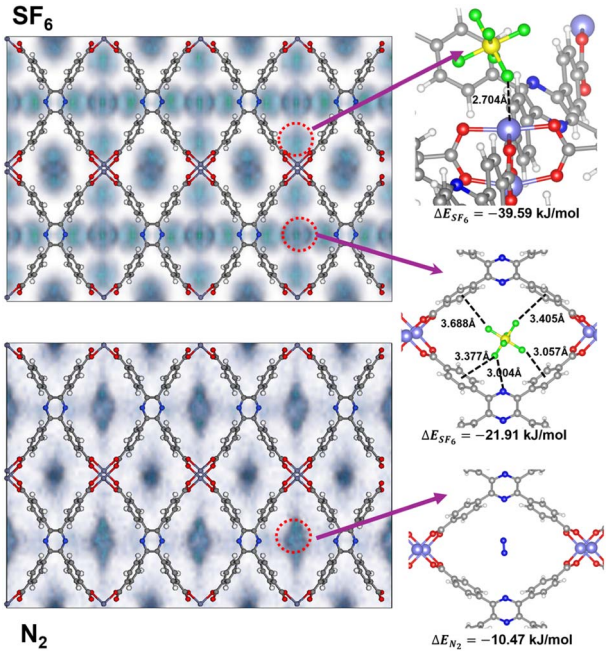

Fig. 7 Probability density distributions of SF_6 and N_2 molecules based on GCMC simulations. Binding sites of SF_6 and N_2 molecules in the Zn-TCPP pores based on the DFT optimization calculations.

To gain further molecular-level insights into the high selectivity of Zn-TCPP for SF_6 over N_2, periodic dispersion-corrected DFT calculations were conducted on representative configurations based on GCMC simulations (Fig. 7). These calculations revealed that the most favorable SF_6 adsorption site was located at the OMS, with a binding energy of −39.59 kJ mol^{-1}. At this site, the fluorine atom of SF_6 and the Zn center are 2.704 Å apart, indicating strong electrostatic attraction. A secondary binding site was identified within a confined pocket formed by four aromatic rings. In this environment, the distances from the fluorine atoms in SF_6 to the aromatic ring centroids are 3.057–3.688 Å. Moreover, an F···N contact of 3.004 Å further confirms the presence of noncovalent interactions. The total binding energy at this site was calculated to be −21.91 kJ mol^{-1}. In contrast, the binding energy of N_2 is substantially lower (−10.47 kJ mol^{-1}), underscoring the preferential affinity of the MOF for SF_6. Therefore, these DFT results validate the adsorption performance predicted by the GCMC simulation and provide a detailed energetic framework for understanding the origin of the selective SF_6 uptake of Zn-TCPP, revealing its potential for efficient SF_6/N_2 separation.

Conclusion

Computational screening based on DDEC charges was used to identify MOFs with superior performance for SF_6/N_2 separation. Through systematic evaluation of the QMOF database, we identified several top candidates with both high SF_6 uptake and favorable selectivity. Detailed structural analysis revealed that

high-performing MOFs share several key structural features, including OMSs, parallel aromatic surfaces, uncoordinated nitrogen atoms, and MOMBs.

A quantitative structure–performance relationship analysis confirmed the fundamental trade-off relationship between the uptake capacity and selectivity. The maximum SF_6 uptakes were observed in MOFs with GSA of 3700–4300 $m^2\ g^{-1}$, AV of 1.21–1.51 $cm^3\ g^{-1}$, void fractions close to 0.75, and LCD of 9.6–11.6 Å. In contrast, the highest SF_6/N_2 selectivities were associated with MOFs with considerably lower GSA values (1190–1690 $m^2\ g^{-1}$), low AV ($\sim$0.19 $cm^3\ g^{-1}$), void fractions of 0.37–0.47, and narrow LCDs centered around 6.1 Å. In addtion, the most common topologies among the high-performing MOFs were rna, pcu, acs, and sql.

Furthermore, the ML model for SF_6 uptake, which was trained on DDEC-based GCMC simulation data, achieved an R^2 of 0.968 and an MAE of 0.281 mmol g^{-1}. Leveraging this model, we efficiently screened 154 144 MOFs, among which Zn-TCPP was chosen as a representative candidate for proof-of-concept validation using DFT calculations. Overall, this integrated workflow provides a scalable and transferable model that can accelerate the development of advanced porous materials for challenging molecular separations.

Author contributions

Ruihan Wang: data curation, formal analysis, investigation, methodology, visualization, writing – review & editing. Shiqi Wang: data curation, investigation. Qianji Han: funding acquisition, project administration, resources, supervision, writing – review & editing.

Conflicts of interest

The authors declare no conflict of interest.

Data availability

All datasets, training codes, and trained models are publicly accessible in our GitHub repository (**https://github.com/ruihwang/SF6_MOFCGCNN**). Other data supporting the findings of this study are available from the corresponding author upon reasonable request.

The Supplementary information: includes the Lennard–Jones parameters employed in this work, the list of top-performing MOFs and hypothetical MOFs under specified conditions, and the distribution of SF_6 adsorption uptakes in the training, validation, and test sets after label-based binning. See DOI: **https://doi.org/10.1039/d5ra06266g**.

Acknowledgements

The authors acknowledge the financial support of the Natural Science Foundation of Hebei Province (B2024407001) and the Science Research Project of Hebei Education Department (BJ2025161).

References

1 C. Y. Chuah, Y. Lee and T.-H. Bae, *Chem. Eng. J.*, 2021, **404**, 126577.
2 T. Hasell, M. Miklitz, A. Stephenson, M. A. Little, S. Y. Chong, R. Clowes, L. Chen, D. Holden, G. A. Tribello, K. E. Jelfs and A. I. Cooper, *J. Am. Chem. Soc.*, 2016, **138**, 1653–1659.
3 I. Senkovska, E. Barea, J. A. R. Navarro and S. Kaskel, *Microporous Mesoporous Mater.*, 2012, **156**, 115–120.
4 W.-T. Tsai and J. Fluor, *Chem*, 2007, **128**, 1345–1352.
5 J. Ren, M. Chang, W. Zeng, Y. Xia, D. Liu, G. Maurin and Q. Yang, *Chem. Mater.*, 2021, **33**, 5108–5114.
6 H. Furukawa, K. E. Cordova, M. O'Keeffe and O. M. Yaghi, *Science*, 2013, **341**, 1230444.
7 H. Furukawa, N. Ko, Y. B. Go, N. Aratani, S. B. Choi, E. Choi, A. Ö. Yazaydin, R. Q. Snurr, M. O'Keeffe, J. Kim and O. M. Yaghi, *Science*, 2010, **329**, 424–428.
8 O. M. Yaghi, M. O'Keeffe, N. W. Ockwig, H. K. Chae, M. Eddaoudi and J. Kim, *Nature*, 2003, **423**, 705–714.
9 J.-W. Yan, Z.-Y. Liu, S. Song, Y.-N. Bian, R. Wang and J.-L. Du, *Chem. Eng. Sci.*, 2025, **302**, 120871.
10 Z.-Y. Liu, J. Yan, Q. Han, J.-C. Zhao, L. Liu, R. Wang, S.-Q. Yang and J.-L. Du, *Microporous Mesoporous Mater.*, 2025, **391**, 113622.
11 M. Chang, T. Yan, Y. Wei, J.-X. Wang, D. Liu and J.-F. Chen, *Chem. Mater.*, 2022, **34**, 9134–9143.
12 X. Zhang, Y.-L. Zhao, X.-Y. Li, X. Bai, Q. Chen and J.-R. Li, *J. Am. Chem. Soc.*, 2024, **146**, 19303–19309.
13 C. R. Groom, I. J. Bruno, M. P. Lightfoot and S. C. Ward, *Acta Crystallogr., Sect. B:Struct. Sci., Cryst. Eng. Mater.*, 2016, **72**, 171–179.
14 R. Wang, Y. Zhong, L. Bi, M. Yang and D. Xu, *ACS Appl. Mater. Interfaces*, 2020, **12**, 52797–52807.
15 C. M. Simon, J. Kim, D. A. Gomez-Gualdron, J. S. Camp, Y. G. Chung, R. L. Martin, R. Mercado, M. W. Deem, D. Gunter, M. Haranczyk, D. S. Sholl, R. Q. Snurr and B. Smit, *Energy Environ. Sci.*, 2015, **8**, 1190–1199.
16 X. Liu, R. Wang, X. Wang and D. Xu, *J. Phys. Chem. C*, 2023, **127**, 11268–11282.
17 J. Cha, S. Ga, S. Lee, S. Nam, Y.-S. Bae and Y. G. Chung, *Chem. Eng. J.*, 2021, **426**, 131787.
18 P. Sezgin, H. C. Gulbalkan and S. Keskin, *J. Phys. Mater.*, 2024, **7**, 045006.
19 R. Wang, Y. Zou, C. Zhang, X. Wang, M. Yang and D. Xu, *Microporous Mesoporous Mater.*, 2022, **331**, 111666.
20 S. Tang, H. Yuan, R. Wang and D. Xu, *Sep. Purif. Technol.*, 2025, **376**, 133864.
21 M. G. Martin and J. I. Siepmann, *J. Phys. Chem. B*, 1998, **102**, 2569–2577.
22 C. Altintas and S. Keskin, *Mol. Syst. Des. Eng.*, 2020, **5**, 532–543.
23 T. A. Manz and N. G. Limas, *RSC Adv.*, 2016, **6**, 47771–47801.
24 N. G. Limas and T. A. Manz, *RSC Adv.*, 2016, **6**, 45727–45747.
25 T. A. Manz, *RSC Adv.*, 2017, **7**, 45552–45581.
26 N. G. Limas and T. A. Manz, *RSC Adv.*, 2018, **8**, 2678–2707.

27 K. Sladekova, C. Campbell, C. Grant, A. J. Fletcher, J. R. B. Gomes and M. Jorge, *Adsorption*, 2020, **26**, 663–685.
28 A. K. Rappe and W. A. Goddard, *J. Phys. Chem.*, 1991, **95**, 3358–3363.
29 C. E. Wilmer, K. C. Kim and R. Q. Snurr, *J. Phys. Chem. Lett.*, 2012, **3**, 2506–2511.
30 S. Liu and B. Luan, *Nanoscale*, 2022, **14**, 9466–9473.
31 D. Nazarian, J. S. Camp, Y. G. Chung, R. Q. Snurr and D. S. Sholl, *Chem. Mater.*, 2017, **29**, 2521–2528.
32 A. Sriram, S. Choi, X. Yu, L. M. Brabson, A. Das, Z. Ulissi, M. Uyttendaele, A. J. Medford and D. S. Sholl, *ACS Cent. Sci.*, 2024, **10**, 923–941.
33 A. S. Rosen, S. M. Iyer, D. Ray, Z. Yao, A. Aspuru-Guzik, L. Gagliardi, J. M. Notestein and R. Q. Snurr, *Matter*, 2021, **4**, 1578–1597.
34 A. J. White, M. Gibaldi, J. Burner, R. A. Mayo and T. K. Woo, *J. Am. Chem. Soc.*, 2025, **147**, 17579–17583.
35 T. F. Willems, C. H. Rycroft, M. Kazi, J. C. Meza and M. Haranczyk, *Microporous Mesoporous Mater.*, 2012, **149**, 134–141.
36 Y. G. Chung, E. Haldoupis, B. J. Bucior, M. Haranczyk, S. Lee, H. Zhang, K. D. Vogiatzis, M. Milisavljevic, S. Ling, J. S. Camp, B. Slater, J. I. Siepmann, D. S. Sholl and R. Q. Snurr, *J. Chem. Eng. Data*, 2019, **64**, 5985–5998.
37 B. J. Bucior, A. S. Rosen, M. Haranczyk, Z. Yao, M. E. Ziebel, O. K. Farha, J. T. Hupp, J. I. Siepmann, A. Aspuru-Guzik and R. Q. Snurr, *Cryst. Growth Des.*, 2019, **19**, 6682–6697.
38 A. K. Rappe, C. J. Casewit, K. S. Colwell, W. A. Goddard and W. M. Skiff, *J. Am. Chem. Soc.*, 1992, **114**, 10024–10035.
39 D. Dellis and J. Samios, *Fluid Phase Equilib.*, 2010, **291**, 81–89.
40 D. Dubbeldam, S. Calero, D. E. Ellis and R. Q. Snurr, *Mol. Simul.*, 2016, **42**, 81–101.
41 U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, *J. Chem. Phys.*, 1995, **103**, 8577–8593.
42 J. VandeVondele, M. Krack, F. Mohamed, M. Parrinello, T. Chassaing and J. Hutter, *Comput. Phys. Commun.*, 2005, **167**, 103–128.
43 S. Goedecker, M. Teter and J. Hutter, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1996, **54**, 1703–1710.
44 C. Hartwigsen, S. Goedecker and J. Hutter, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1998, **58**, 3641–3662.
45 S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104.
46 T. K. Pal, D. De, S. Neogi, P. Pachfule, S. Senthilkumar, Q. Xu and P. K. Bharadwaj, *Chem.–Eur. J*, 2015, **21**, 19064–19070.
47 M.-B. Kim, K.-M. Kim, T.-H. Kim, T.-U. Yoon, E.-J. Kim, J.-H. Kim and Y.-S. Bae, *Chem. Eng. J.*, 2018, **339**, 223–229.
48 C. Y. Chuah, K. Goh and T.-H. Bae, *J. Phys. Chem. C*, 2017, **121**, 6748–6755.
49 P. G. Boyd, A. Chidambaram, E. García-Díez, C. P. Ireland, T. D. Daff, R. Bounds, A. Gładysiak, P. Schouwink, S. M. Moosavi, M. M. Maroto-Valer, J. A. Reimer, J. A. R. Navarro, T. K. Woo, S. Garcia, K. C. Stylianou and B. Smit, *Nature*, 2019, **576**, 253–256.
50 Y. Lan, T. Yan, M. Tong and C. Zhong, *J. Mater. Chem. A*, 2019, **7**, 12556–12564.
51 Y. J. Colón, D. A. Gómez-Gualdrón and R. Q. Snurr, *Cryst. Growth Des.*, 2017, **17**, 5801–5810.
52 E. Alvarez, N. Guillou, C. Martineau, B. Bueken, B. Van de Voorde, C. Le Guillouzer, P. Fabry, F. Nouar, F. Taulelle, D. de Vos, J. Chang, K. H. Cho, N. Ramsahye, T. Devic, M. Daturi, G. Maurin and C. Serre, *Angew. Chem., Int. Ed.*, 2015, **54**, 3664–3668.
53 M. K. Taylor, T. Runčevski, J. Oktawiec, M. I. Gonzalez, R. L. Siegelman, J. A. Mason, J. Ye, C. M. Brown and J. R. Long, *J. Am. Chem. Soc.*, 2016, **138**, 15019–15026.
54 J. A. Mason, J. Oktawiec, M. K. Taylor, M. R. Hudson, J. Rodriguez, J. E. Bachman, M. I. Gonzalez, A. Cervellino, A. Guagliardi, C. M. Brown, P. L. Llewellyn, N. Masciocchi and J. R. Long, *Nature*, 2015, **527**, 357–361.
55 R. Wang, Q. Gao, Y. Zhong, X. Wang and D. Xu, *Sep. Purif. Technol.*, 2023, **304**, 122332.
56 Y. Hu, L. Wang, R. Nan, N. Xu, Y. Jiang, D. Wang, T. Yan, D. Liu, Y. Zhang and B. Chen, *Chem. Eng. J.*, 2023, **471**, 144851.
57 Y.-P. Li, J.-J. Ni, X.-J. Zhang, X.-L. Zhang, W. Wen, Z.-Y. Sui and X.-F. Xu, *Inorg. Chem.*, 2024, **63**, 13568–13575.
58 T. D. Pham, F. Joodaki, F. Formalik and R. Q. Snurr, *J. Phys. Chem. C*, 2024, **128**, 17165–17174.
59 S. Thaler, F. Mayr, S. Thomas, A. Gagliardi and J. Zavadlav, *npj Comput. Mater.*, 2024, **10**, 86.
60 J. Luo, O. B. Said, P. Xie, M. Gibaldi, J. Burner, C. Pereira and T. K. Woo, *npj Comput. Mater.*, 2024, **10**, 224.
61 G. Zhao and Y. G. Chung, *J. Chem. Theory Comput.*, 2024, **20**, 5368–5380.
62 C. E. Wilmer, O. K. Farha, Y.-S. Bae, J. T. Hupp and R. Q. Snurr, *Energy Environ. Sci.*, 2012, **5**, 9849.
63 Y. Jiang, L. Sun, J. Du, Y. Liu, H. Shi, Z. Liang and J. Li, *Cryst. Growth Des.*, 2017, **17**, 2090–2096.