Chemical Science

ROYAL SOCIETY OF CHEMISTRY

EDGE ARTICLE

View Article Online
View Journal | View Issue

Cite this: *Chem. Sci.*, 2025, **16**, 17450

All publication charges for this article have been paid for by the Royal Society of Chemistry

A mechanism-guided inverse engineering framework to unlock design principles of H-bonded organic frameworks for gas separation

Yong Qiu, [a] Lei Wang,[a] Letian Chen,[b] Yun Tian, *[a] Zhen Zhou*[ab] and Jianzhong Wu [c]

Received 13th June 2025
Accepted 26th August 2025

DOI: 10.1039/d5sc04332h

rsc.li/chemical-science

The diverse combinations of novel building blocks offer a vast design space for hydrogen-bonded organic frameworks (HOFs), rendering them highly promising for gas separation and purification. However, the underlying separation mechanism facilitated by their unique hydrogen-bond networks has not yet been fully understood. In this work, a comprehensive understanding of the separation mechanisms was achieved through an iterative data-driven inverse engineering approach established upon a hypothetical HOF database possessing nearly 110 000 structures created by a materials genomics method. Leveraging a simple yet universal feature extracted from hydrogen bonding information with unambiguous physical meanings, the entire design space was exploited to rapidly identify the optimization route towards novel HOF structures with superior Xe/Kr separation performance (selectivity > 10^3). This work not only provides the first large-scale HOF database, but also demonstrates the enhanced machine learning interpretability of our model-driven iterative inverse design framework, offering new insights into the rational design of nanoporous materials for gas separation.

Introduction

The efficient adsorption and separation of xenon (Xe) and krypton (Kr) hold substantial industrial significance due to their specific and extensive applications in the nuclear industry, space exploration, commercial lighting and medical devices.[1–4] However, due to the nearly identical kinetic diameters (Kr = 3.66 Å; Xe = 4.05 Å), ultra-low concentrations (Kr = 1.14 ppmv; Xe = 0.09 ppmv), and extremely low polarizabilities, the major challenges associated with the separation of Xe and Kr remain significant.[5] Traditionally, cryogenic distillation is the most mature technology to separate pure Xe and Kr (20/80 (v/v)) from air, but these processes are energy- and capital-intensive.[6] Metal–organic frameworks (MOFs) and covalent organic frameworks (COFs) are the most developed porous materials for Xe/Kr separation, yet their commercialization is significantly hindered by issues related to material stability, high costs, and environmental pollution. Therefore, exploring stable, environmentally friendly porous adsorbents with controllable structures to balance capacity and selectivity remains both appealing and challenging.[7]

Compared with other classes of nanoporous materials dominated by relatively stronger intra-molecular interactions like MOFs and COFs, HOFs possess unique advantages, such as milder synthesis conditions, facile solution processability and recyclability, which are usually attributed to their weaker intermolecular interactions and higher reversibility of hydrogen bonding.[5] In addition, previous studies have demonstrated that HOFs exhibit significant potential for Xe/Kr separation, with excellent thermal and chemical stability, and can be efficiently recovered and reused through simple recrystallization, thereby greatly reducing the operational cost of the adsorbents.[1,7]

To date, the total number of experimentally reported HOF structures is still limited to 10^2, primarily owing to the difficulty in materials synthesis.[8,9] Although the separation of Xe/Kr in MOFs[10,11] was attributed to the modified polarizability of guest molecules induced by designed functional groups, the separation mechanism for Xe/Kr in HOF materials remains unknown, which severely hinders the rational design of high-performance HOF materials. With a large unexplored design space, the recent application of modular design in HOF synthesis,[12,13] and the emergence of automated laboratories,[14–16] *in silico* design of HOFs is highly attractive as a viable alternative to advance the fundamental understanding of the separation mechanism and accelerate the discovery of targeted HOFs.

As a well-established high-throughput computational method for predicting gas adsorption and separation, Classical

[a]*Interdisciplinary Research Center for Sustainable Energy Science and Engineering (IRC4SE^2), School of Chemical Engineering, Zhengzhou University, Zhengzhou 450001, Henan, China. E-mail: ytian009@zzu.edu.cn*

[b]*School of Materials Science and Engineering, Institute of New Energy Material Chemistry, Renewable Energy Conversion and Storage Center (ReCast), Key Laboratory of Advanced Energy Materials Chemistry (Ministry of Education), Nankai University, Tianjin 300350, China. E-mail: zhouzhen@nankai.edu.cn*

[c]*Department of Chemical and Environmental Engineering, University of California, Riverside, CA, 92521, USA*

Density Functional Theory (CDFT) has demonstrated its outstanding efficiency and accuracy under various scenarios, including thermodynamic phenomena involving diverse molecular and ionic systems.[17–21] The integration of machine learning techniques with CDFT has further improved the computational efficiency and revolutionized the design strategy in recent years.[22–25] In our previous work, the forward design strategy was employed to identify high-performance gas separation materials from an existing COF database.[26] Given the vast chemical space with an exponentially increasing number of building units, it is essential to incorporate inverse design approaches to uncover the optimization route for novel HOF structures.[27–29] Regrettably, the applicability of current inverse design methods is usually limited to the initial dataset or their analogues, which severely hinders the iteration process of new structures.[30–32]

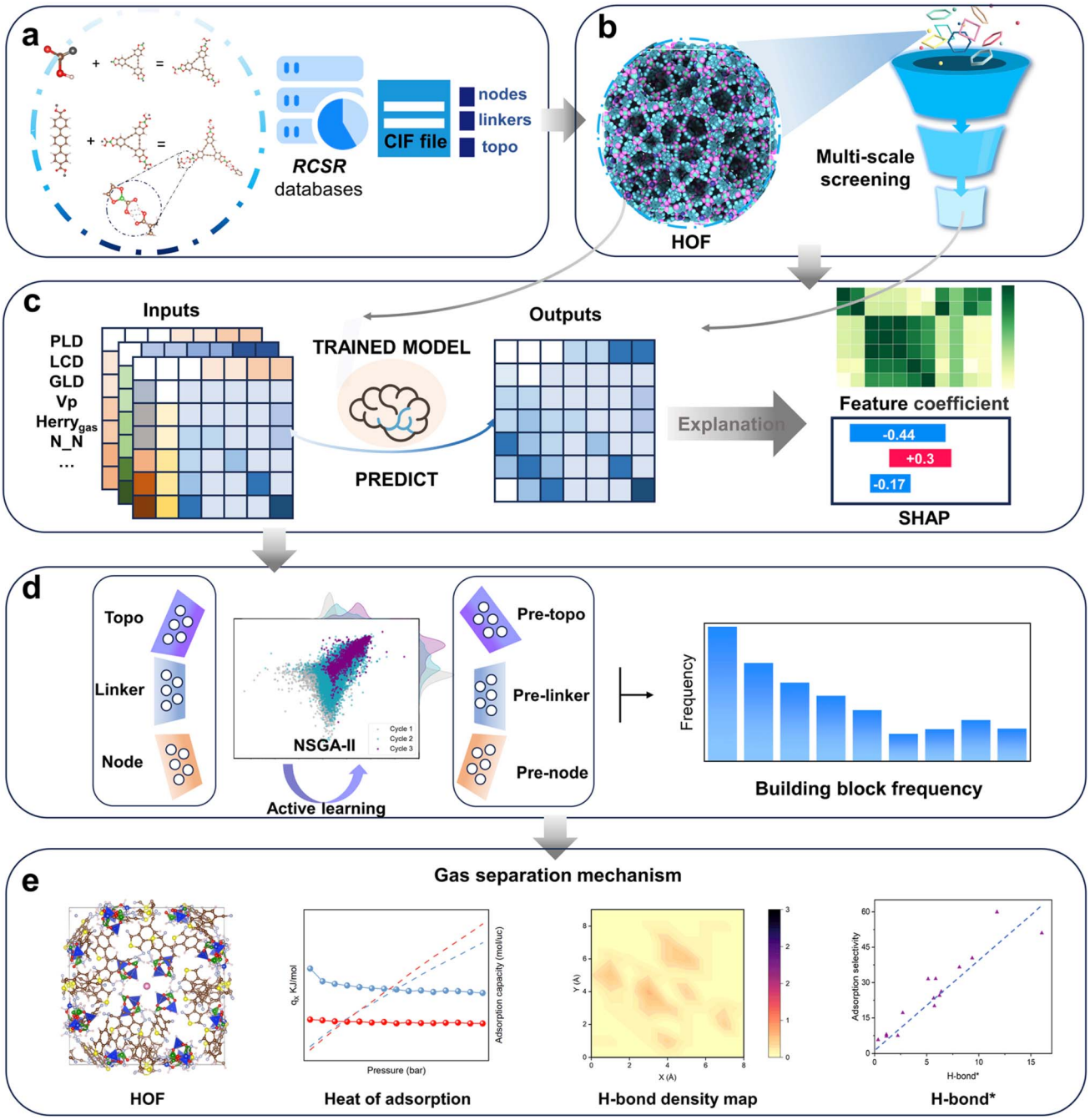

Fig. 1 Workflow of mechanism exploration based on a self-built database and the iterative inverse engineering framework. (a) *In silico* construction of a HOF database. (b) Feature extraction and high-throughput prediction of separation performance using CDFT. (c) Machine learning model training and interpretability analysis. (d) Active learning promoted iterative inverse design of HOFs. (e) Mechanism revelation of Xe/Kr separation in HOFs.

In this work, we developed a hypothetical framework to construct a large-scale HOF database with a genomics-based method as shown in Fig. 1, enabling the rapid generation of HOF structures and high-throughput predictions of the gas separation performance of Xe/Kr using CDFT. Through incorporation of a genetic algorithm with active learning, for the first time, we established an iterative inverse design framework to efficiently optimize the structure of novel HOF materials to identify the target ones. As an illustration, we successfully designed a novel HOF structure with Xe/Kr selectivity over 10^3, exceeding the best existing adsorbent for Xe/Kr separation.[1,33,34] Furthermore, after comprehensive analysis with our framework, it is conclusively demonstrated that the Xe/Kr separation mechanism in HOFs is not primarily governed by polarizability differences. Instead, the separation process is driven by a synergistic dual mechanism involving pore sieving effects and variable hydrogen bonding intensity. More excitingly, the fundamental mechanism driving Xe/Kr separation was utilized to establish a novel descriptor with enhanced interpretability and universality, which is constructed from both the hydrogen-bond networks and key structural features of the materials. With the iterative inverse engineering framework, we expect to provide a new paradigm for developing nanoporous materials using materials genomics strategies, offering guiding principles for the experimental design of HOFs.

Results

Materials genomics-based algorithms for HOF construction

The primary distinction between HOF, COF, and MOF materials lies in the bonding interactions between the building blocks (BBs).[35,36] Experiments have shown that COFs and HOFs can be interconverted through the modification of functional groups and by adjusting experimental conditions to control the material docking process.[36,37] Specifically, when the building units possess functional groups capable of acting as hydrogen-bond donors or acceptors, thereby facilitating the formation of stable hydrogen-bond networks, it becomes theoretically possible to construct novel HOF structures based on existing topological frameworks. Taking the building units and topological structures of COF materials collected in the Reticular Chemistry Structure Resource (RCSR)[38] database as a template, we propose a methodology to construct BBs for hypothetical HOFs by functionalizing the building units (as shown in Fig. 1a). Functional groups (–OH, –COOH, $–NH_2$, –F) serve as hydrogen-bond donors/acceptors, enabling diverse hydrogen-bond networks. These modified BBs constitute novel HOF structures, realizing COF-to-HOF transformation as experimentally demonstrated. Through the incorporation of these functional groups, the BBs were not only saturated but also capable of forming hydrogen-bond networks, thus satisfying the preliminary criteria for the design and synthesis of HOFs. New BBs are named by appending the functional group to the original BB name (*e.g.*, 'C6' with –COOH becomes 'C6_COOH').

The hierarchical assembly of HOFs initiates with the rational design of ligand building blocks, where ligand units (Fig. S1) can be functionalized through strategic substitution of hydrogen-bonding motifs (*e.g.*, –COOH, $–NH_2$, –OH, –F) or alternatively developed *via de novo* synthesis to create novel architectures. These designed ligands are then systematically paired with compatible topological units (*e.g.*, sql, dia, hcb) to construct a comprehensive gene library, with each entry explicitly encoding the topology–ligand relationship (topology + node + edge ligand) After assessing the degree of compatibility, BBs and topological units are assembled using pormake[39] to construct complete new HOFs. Through a parallel design approach, we significantly accelerated the materials design process and established the first theoretical HOF database, comprising nearly 110 000 distinct HOF structures. Moreover, this approach offers sufficient versatility and flexibility for users to perform targeted materials design, enabling facile modification of functional groups and integration of diverse BBs to construct materials tailored to specific applications.

Theoretical HOF materials database

Under typical industrial separation conditions (296 K, 1 bar, 20/80 molar ratio for Xe and Kr),[1,40] we performed high-throughput CDFT calculations for Xe/Kr gas separation on the constructed HOF database after excluding those with inaccessible pore geometries for gas molecules, and detailed comparison and validation with available experimental data are shown in Fig. S2. Moreover, we tested materials for Xe/Kr separation under different conditions including air (Xe/Kr = 0.086 ppm : 1.14 ppm (ref. 41)) and nuclear waste gas (Xe/Kr = 500 ppm : 50 ppm (ref. 42)). Table S1 shows consistent adsorption selectivity across these scenarios. Key structural characteristics including the largest included sphere along the free sphere path diameter (PLD), largest free sphere diameter (LCD), pore volume, framework density, pore characteristics, specific surface area, *etc.*, and chemical features including element ratios, unsaturation, electronegativity ratio, *etc.*, were calculated for subsequent machine learning implementations.

The selectivity values in our database range from 0 to 10^4, reflecting the diversity of our comprehensive dataset, with the top-performing materials significantly surpassing the separation capabilities of existing experimental HOF materials. We provide the representative structure of the high-performance HOFs (selectivity > 10^3), including its monomer composition and hydrogen-bonding patterns, as illustrated in Fig. S3. Detailed data distributions of the entire database can be found in Fig. S4 to S13. For analytical clarity, we categorized the HOFs in the database into three groups based on their Xe/Kr adsorption selectivity: Class 1 (top 10%), Class 2 (top 10–30%), and Class 3 (the remaining materials). As illustrated in Fig. 2, an apparent correlation can be drawn between the Xe/Kr separation performance and the physical structure of these materials. Materials exhibiting superior separation performance typically feature relatively smaller pore sizes with PLD around 15 Å (Fig. 2a), higher framework densities (Fig. 2b), and smaller void fractions (Fig. 2c), which can be attributed to the inert nature of Xe and Kr and the dominating size-exclusion effects. In terms of these geometric features, the void fraction plays a vital role in measuring the upper limit of capacity for gas

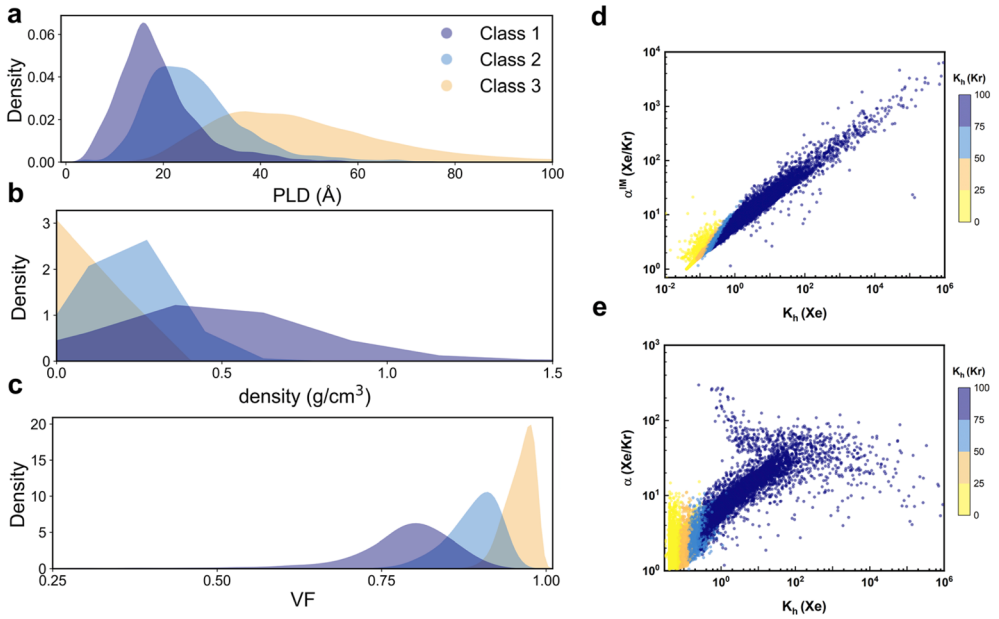

Fig. 2 Relationship between HOF framework characteristics and Xe/Kr adsorption selectivity. Density plots of the framework's physical structure and adsorption selectivity: PLD (a), density of frameworks (b), and void fraction (c), as well as the relationship between Henry's constant and ideal adsorption selectivity (d) and practical adsorption selectivity (e). In panels (d) and (e), the color map represents the variation in Henry's constant for Kr, with darker colors indicating higher Henry's coefficients.

molecule accommodation, and with well-defined pore size and density of frameworks, the optimal HOF is capable of effectively differentiating guest molecules through non-uniform intermolecular interactions with its hydrogen-bond networks, leading to enhanced separation efficiency. Meanwhile, Henry's coefficient, which accounts for the thermodynamic interactions between the guest and host molecules, also determines the relative adsorption strength and separation performance of porous materials. As shown in Fig. 2d, Henry's coefficients for Xe/Kr in HOFs exhibit a consistent upward trend, regardless of whether they are evaluated for ideal adsorption selectivity (Fig. 2d) or selectivity under practical operating conditions (Fig. 2e), indicating that separation performance might be primarily driven by the difference in interacting affinity with frameworks. However, depending solely on any single physicochemical character is inadequate for effectively distinguishing the performance of HOFs.

While experimental results show the influence of measurable physicochemical properties on performance, high-throughput calculations and machine learning models enable rapid prediction of materials performance and offer deeper insights into the underlying mechanisms. In this regard, we conducted feature selection using the Pearson correlation coefficient (Fig. 3a) integrated with unsupervised learning methods, including t-SNE and PCA (Fig. S14–S16). Our analysis revealed that the physical structural features predominantly influence the target property prediction, while the chemical characteristics exhibit a relatively weaker correlation with the separation performance. Owing to the inert nature of noble gases, Kr and Xe molecules are less prone to chemically react with the framework and are predominantly adsorbed through physical interactions. To more specifically quantify the relationship between features and performances, we applied a variety of machine learning models (listed in Table S2) to predict the actual Xe/Kr adsorption selectivity of the HOFs (Fig. S17). Through hyperparameter optimization for each model, with an 80 : 20 training-to-test set ratio and 5-fold cross-validation, we identified the model with the best performance. Our results show that tree-based models, such as Random Forest (RF) and Gradient Boosting Regression (GBR), outperformed linear regression algorithms including Least Absolute Shrinkage and Selection Operator (LASSO), Linear Support Vector Machine (Linear SVM), and Partial Least Squares (PLS) regarding data regression. Neural-network-based models like Artificial Neural Networks (ANN) and K-Nearest Neighbors (KNN) were prone to overfitting, indicating excessive model complexity. Among all of them, the GBR model delivered the best performance, achieving an R^2 of 0.905 on the test set as illustrated in Fig. 3b. In spite of the minor divergency in regions of high selectivity due to sparse data distribution in these areas, the machine learning model generally provides satisfactory predictions for separation performance based on physical properties. To further investigate the structure–property relationships of the materials, we conducted SHapley Additive exPlanations (SHAP) analysis on the trained GBR model to assess the significance of each feature, as depicted in Fig. 3c and d. The top nine most important features, all of which are physical properties, are listed, based on both the average SHAP values and individual feature contributions, which further substantiate our earlier feature selection analysis, confirming that physical properties are most critical in determining separation performance, with framework density ranking among the top three most significant features. However, our analysis of individual feature weights (Fig. 3d) reveals the coexistence of

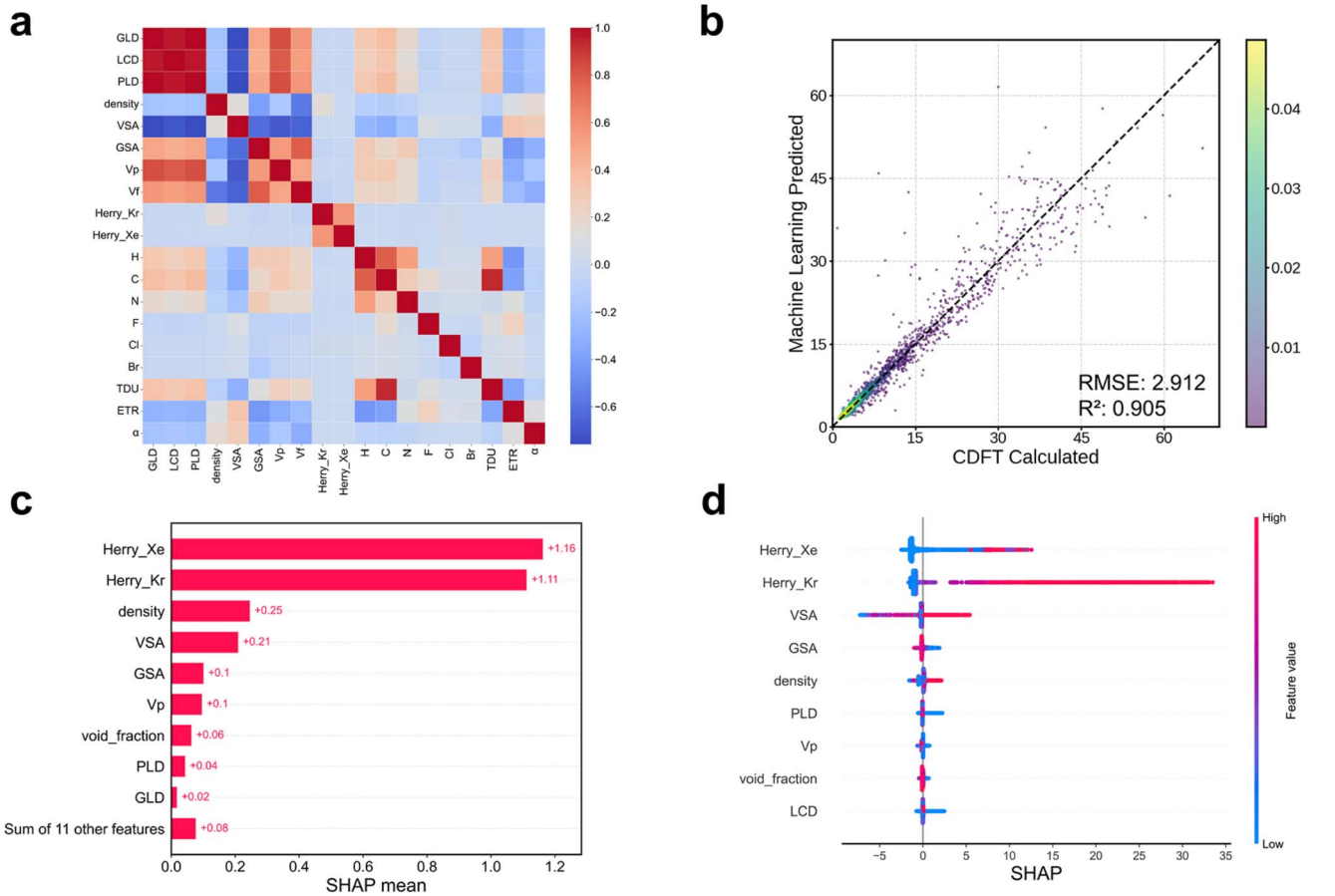

Fig. 3 Correlation of different features, importance analysis, and prediction performance. (a) Pearson correlation coefficients, with color mapping representing both the direction (positive/negative) and magnitude of correlations (the top 10 are physical features, while the others are chemical features. The full names are provided in Table S1); (b) GBR machine learning model performance on the test set; (c) average SHAP values, indicating the mean feature importance across all samples; (d) SHAP beeswarm plot visualizing the distribution of SHAP values, with higher values represented by redder colors.

both positive and negative correlations among features across different frameworks. This observation indicates that these descriptors cannot be used in isolation to reliably assess materials performance, highlighting the need to uncover the more complex underlying mechanisms that govern the intercorrelation of these individual features.

Performance-driven inverse design of HOFs

Inverse design represents a paradigm shift in materials discovery, wherein the process initiates from the desired materials performance and systematically traces back to the requisite structural properties. This approach facilitates more efficient and targeted materials development. However, materials properties are typically governed by multiple performance factors rather than a single determinant. An exclusive focus on a singular objective during the design process may lead to impractical outcomes, such as the materials with extremely high selectivity but very low adsorption capacity. To address this issue, we employed the NSGA[43] (Non-dominated Sorting Genetic Algorithm) for multi-objective optimization, incorporating both the adsorption capacity and selectivity for the target gas species of Xe. The simultaneous achievement of high adsorption density and high selectivity is set as a key indicator of superior materials performance, demonstrating both efficient molecular uptake capacity and precise target-specific recognition capabilities. Given that inverse design is intrinsically associated with the BBs and topology of the material, we incorporate these critical factors as high-dimensional input features. Leveraging the pre-trained MOF-NET[39] model, we predict Xe/Kr adsorption selectivity and Xe adsorption capacity of HOFs, enabling a more targeted and efficient design process. The model employs a resampling strategy to address data imbalance by selectively augmenting underrepresented regions within the dataset, thereby enhancing the overall balance and representativeness of the database. As shown in Fig. S18, the R^2 values for the test set are consistently approximately 0.99, demonstrating significantly improved predictive accuracy compared with models trained exclusively on physical and chemical features as seen in Fig. S17. The enhanced performance can be exclusively attributed to the model's implicit incorporation of critical structural factors, particularly the coordination environment of diverse BBs, which encompasses the synergistic effects of hydrogen-bond networks and their surrounding chemical environments.

Compared with Deng and Sarkisov's multi-objective optimization model,[44] our approach represents a substantial advancement through integration of an active learning process into the NSGA-II framework, thereby enhancing the model's adaptive capabilities and optimization efficiency. By selecting the top 10% materials predicted with the most promising performance in each generation for further materials design and high-throughput computation, we continuously assess the model's accuracy and facilitate iterative updates, which are subsequently applied to the next cycle of structure predictions, thereby creating a self-improving optimization loop. This enhanced sampling methodology effectively mitigates decision-making errors arising from materials performance deviations from the pre-trained database, while simultaneously accelerating the discovery and optimization of next-generation materials. Additionally, NSGA-II optimizes the selection process by incorporating crowding distance and elitism, which enhances multi-objective optimization, maintains population diversity, and reduces premature convergence. In this work, as a proof of concept, we implemented three genetic cycles, with each cycle generating approximately 100 000 potential structures (note that duplicate HOFs are systematically eliminated in subsequent generations to maintain structural diversity). As shown in Fig. 4a, the materials generation and performance prediction for the three cycles are depicted. Through comprehensive analysis of the performance density distribution across successive generations, it is evident that the inverse design process effectively directs the transition from performance-oriented to structure-based materials design. In the newly generated third-generation materials, the majority exhibit superior Xe gas density and adsorption selectivity. Notably, when compared with the randomly generated $\sim 10^5$ structures, the proportion of materials with excellent adsorption performance increased substantially. These results underscore the effectiveness of our inverse design approach in generating a higher proportion of candidates with optimized materials performance. Building upon these findings, we conducted a detailed frequency analysis of the BBs present in third-generation materials to identify key structural motifs contributing to these performance enhancements. Fig. 4b and c present the top-ranked node and edge building blocks, respectively. They reveals that –F, –O, and –N functional groups, prone to form strong hydrogen bonds, dominate the BB composition, which is crucial for facilitating efficient Xe/Kr gas separation. Furthermore, we observed a significantly higher frequency of the C36_COOH building block compared with other node types, which can be attributed to C36_COOH's unique structural characteristics in matching more edge building blocks, consequently promoting the formation of additional hydrogen bonds with increased framework density, as illustrated in Fig. S19. To quantitatively evaluate this structural distinction, we conducted a comparative analysis of hydrogen-bond networks across nearly 50 structural variants, in which only the node type was displaced (replacing typical high-frequency-node C36_COOH with low-frequency node C36_NH_3). As demonstrated in Fig. S20, the HOF structures with the C36_COOH node consistently exhibit more hydrogen bonds in virtually all cases, which can be attributed to the presence of additional '–O' groups in C36_COOH, which serve as effective hydrogen-bond acceptors and intensify the hydrogen-bond networks with edge components, thereby enhancing adsorption and separation capabilities.

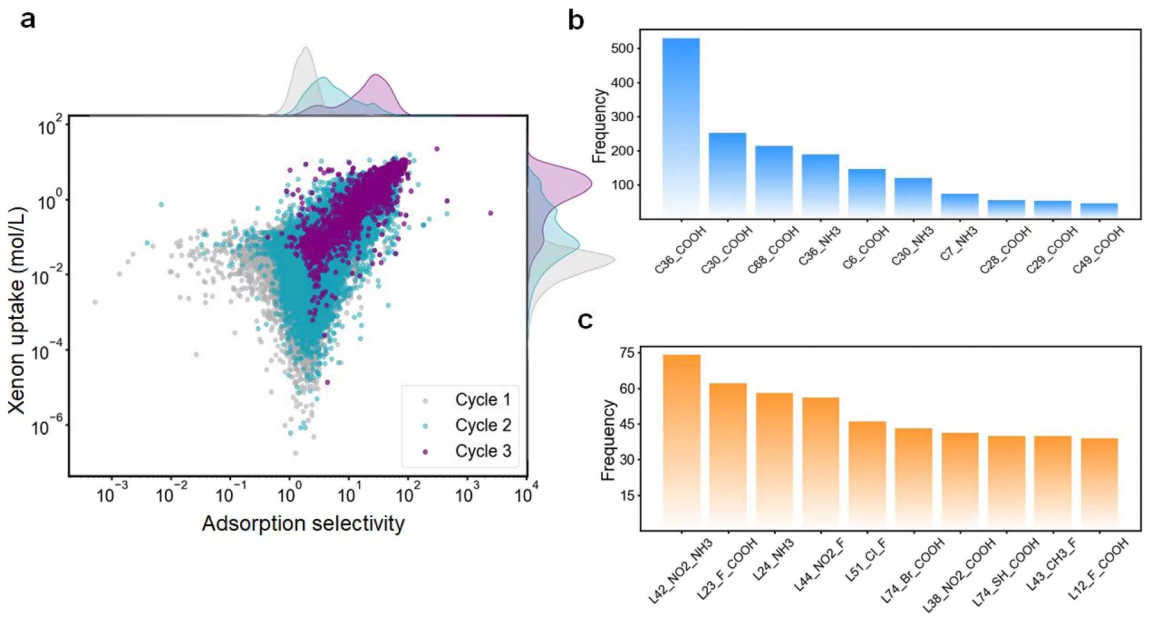

Fig. 4 Performance analysis of inverse-designed materials. (a) Inverse design performance over three cycles, with shading from light to dark indicating an increase in iteration cycles. Higher values correspond to better performance. The materials density analysis for the two performances metrics is shown at the top and right sides. Frequency analysis of the top-ranked node (b) and edge (c) building blocks for materials in cycle 3.

The separation mechanism of Xe/Kr in HOFs

To gain deeper insights into the adsorption mechanisms of guest molecules within the framework, we take 'sxt+C6_NH_3+-L39_CN_F′ as a representative study (Fig. S21). As shown in Fig. 5a and b, we first determined the adsorption sites of Xe/Kr gas molecules in the HOF based on the adsorption density map (the *XY* direction is shown in Fig. S23a). The pore size distribution (PSD) is an important structural characteristic that correlates with adsorption capacity,[45] as shown in Fig. S22. Specifically, the PSD of sxt+C6_NH3+L39_CN_F exhibits a certain correlation with adsorption density (Fig. S23a). The adsorption sites for Xe and Kr are essentially the same, while the adsorption intensity of Xe is stronger than that of Kr, as evidenced by the more concentrated adsorption pattern. As shown in the adsorption density map, the central cage-like region serves as an important adsorption site within the framework. In addition, heat of adsorption exhibited a consistent decreasing trend with increasing loading over a wide pressure range shown in Fig. 5e, reflecting that the dominating molecular interactions transit from guest–host to guest–guest, and larger initial adsorption heat for Xe indicates stronger intermolecular interaction between Xe and the framework.

To understand the adsorption separation mechanism facilitated by the unique hydrogen-bond networks, we also investigated the adsorption sites within corresponding COF structures with the same topology for direct comparison. It was observed that the corresponding COF framework adsorbs guest molecules preferentially at the edge sites (Fig. S23b), instead of the central cage-like region as promoted by the hydrogen-bond networks in HOFs (Fig. S23a). From the adsorption energy analysis (Fig. S24) and the differential charge distribution (Fig. S25), the adsorption strength for Xe at this site is higher than that for Kr and the electron transfer between Xe/Kr and the surrounding BBs is nearly zero, indicating negligible polarizability difference, consistent with the non-polar nature of inert gases. Rather than the polarizable effect,[46] the primary factor driving Xe/Kr separation is probably the difference in the intermolecular interactions between the hydrogen-bond acceptors and gas molecules, which is closely related to the intensity of hydrogen-bond networks within the system. Therefore, we statistically analyzed the distribution of hydrogen bonds within the framework, as shown in Fig. 5c. It was found that this distribution aligns well with the adsorbed density distribution of guest molecules, suggesting that the arrangement of hydrogen-bond networks reflects the relative strength of guest molecule adsorption within the framework. In contrast, there is no such apparent correlation in corresponding COF structures (Fig. 5d), which demonstrates the fundamental difference in the adsorption separation mechanisms between these two types of framework materials.

Is there a universal correlation between the hydrogen-bond networks and the gas separation performance of HOF materials? To address this question, we randomly selected approximately 500 HOF structures and examined the direct

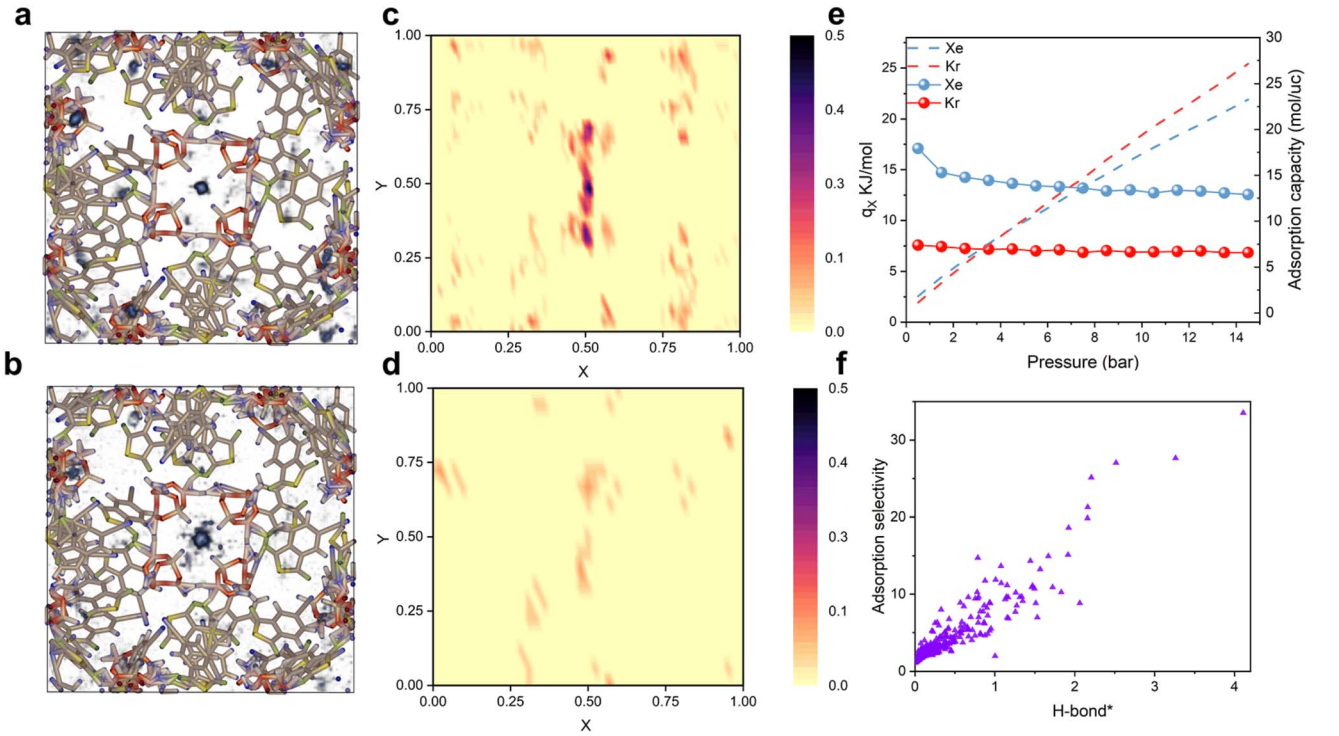

Fig. 5 Mechanistic insights into adsorption, site analysis, and descriptor optimization for selectivity prediction. sxt+C6_NH_3+L39_CN_F material: (a) Xe adsorption density distribution map (the isosurface is set at 10^{-5} mol L^{-1}); (b) Kr adsorption density distribution map; (c) hydrogen bonding distribution in the HOF material; (d) hydrogen bonding distribution density in the corresponding COF framework; (e) heat of adsorption (sphere) and adsorption density (dashed line) with pressure variation (Xe : Kr = 20 : 80, T = 296 K) in the HOF; (f) correlation between adsorption selectivity and H-bond*.

relationship between the number of hydrogen bonds and adsorption selectivity. Through Lasso regression, it was found that the number of hydrogen bonds consistently ranked among the top four features with the highest weight (Fig. S26). However, the number of hydrogen bonds itself does not exhibit a strong correlation with adsorption selectivity as seen in Fig. S27. Taking the top-ranking structural features into consideration, the density of the framework was further identified to possess a stronger correlation with adsorption selectivity as shown in Fig. S28. Besides the void fraction and the simplest yet straightforward feature, it does not involve any complex calculations like Henry's constants. While PSD is an essential characterization tool in separation systems, obtaining PSD data is time-consuming and also cannot reveal the complete structural feature. Through integration of key structural factors with the information on hydrogen-bond networks, we propose a novel descriptor, H-bond*, to represent the effective hydrogen bonding density, which effectively captures the linear correlation with adsorption selectivity without any input from molecular simulation data as illustrated in Fig. 5f. It is worth noting that the slight deviation from linearity is probably due to the varied interaction strength between guest molecules and hydrogen-bond acceptors, which is acceptable as a universal screening tool for rapid performance assessment that bypasses cumbersome molecular simulations. Additionally, the enhanced linear correlation was also confirmed with symbolic regression using SISSO as seen in Fig. S29.

$$\text{H-bond*} = [\alpha + \log(\rho_{\text{H-bond}})] \times \rho_{\text{f}}/V_{\text{f}}$$

where α is a universal parameter to avoid nonphysical results (*e.g.*, negative values for H-bond*), and $\rho_{\text{H-bond}}$ represents the averaged number density of hydrogen bonds within the framework. ρ_{f} represents the gravimetric density of the framework, and V_{f} is the void fraction of the material.

Mechanism validation with experimental data and universal applicability

In addition to the theoretical database, we have also comprehensively evaluated HOF materials reported in recent

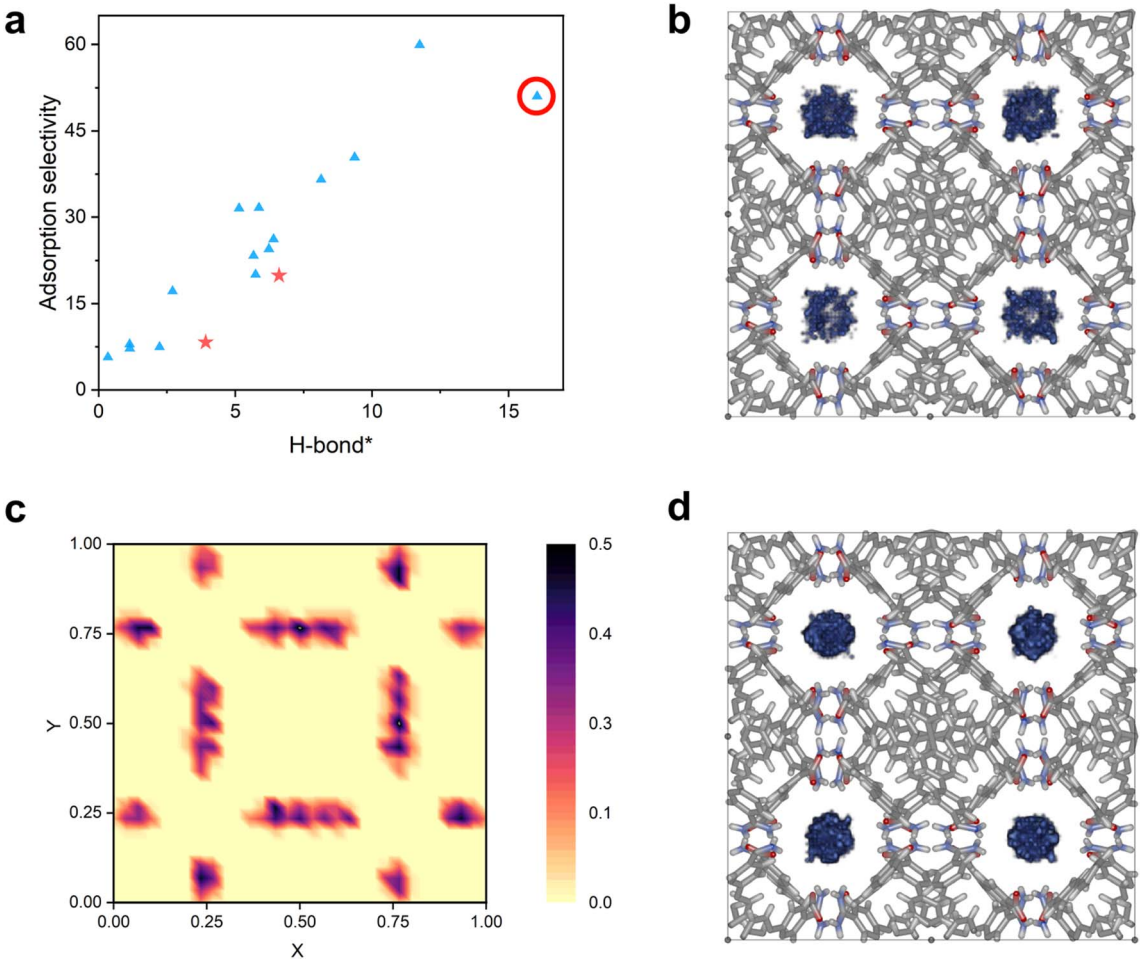

Fig. 6 Experimental validation of the separation mechanism. (a) Relationship between the proposed descriptor and adsorption selectivity from experimentally synthesized HOFs, where stars represent experimentally measured selectivity data (HIAM-103 (ref. 47) and HOF-FJU-46a (ref. 1)), and the triangles are computed selectivity with experimentally synthesized HOFs (Xe/Kr selectivity has not been measured). (b) Adsorption density map of Kr gas in the structure of Bbiphen (2D).[12] (c) Hydrogen bonding arrangement in the framework. (d) Adsorption density map of Xe gas in the structure of Bbiphen (2D),[12] where Bbiphen (2D) corresponds to the circled structure in Fig. 6(a).

experimental literature. As anticipated, through feature analysis with available experimental data, we validated the universality of our proposed descriptor, which exhibits a strong linear correlation with adsorption selectivity as shown in Fig. 6a. The separation mechanism was also verified with a typical experimental HOF material, as highlighted in Fig. 6a. Fig. 6c and d present the adsorption density maps for Xe and Kr, respectively. The hydrogen bonding distribution map in Fig. 6c revealed that the adsorbed gas molecules are predominantly located within the central-cage region, surrounded by hydrogen-bond networks, which is consistent with the mechanism proposed in this work. Guided by the universal descriptor with enhanced mechanism understanding, we suggest that experimentalists increase framework density while maintaining porosity and incorporating functional groups to enhance effective hydrogen bonding density, to realize rational design of high-performance HOF materials for gas separation.

Considering that cyano, pyrazole, and formyl functional groups have not been previously employed in HOF construction, we systematically incorporated these motifs into our design. Fig. S30 presents representative building blocks featuring these newly introduced groups. Fig. S31 displays a representative structure assembled from these building units, showing both its molecular configuration and hydrogen-bonding patterns. Through our computational framework, we generated and evaluated nearly 800 new HOF structures, analyzing their adsorption selectivity and structural characteristics. Remarkably, comparative analysis demonstrates that although these functional groups were absent in earlier models, their resulting adsorption selectivity maintains a strong correlation with our proposed H-bond* descriptor (Fig. S32). This correlation is also corroborated by experimental evidence, as several reported HOF structures containing these groups exhibit consistent performance trends.

Discussion

In this work, we have successfully demonstrated a hypothetical framework to construct the first large-scale HOF database with a materials genomics method and integrate active learning into the inverse design workflow to achieve efficient iteration and optimization towards target materials. More importantly, we have leveraged inverse design tools to elucidate the underlying design principles of high-performance HOF materials, thereby revealing key mechanisms governing gas separation performance. Utilizing this tool, we have successfully designed a novel HOF structure with superior Xe/Kr separation selectivity > 10^3, while simultaneously maintaining a high Xe adsorption capacity. After comprehensive evaluation of the separation mechanism of Xe/Kr in both theoretical and experimental databases, a novel universal descriptor was established to enable rapid performance prediction bypassing cumbersome molecular simulations, with the screening speed enhanced by at least two orders of magnitude. With continuous advancement in the modular manipulation strategies of HOFs, the material design concept we proposed in this work holds great promise for application in automated experimental laboratories. While recognizing the inherent limitations of theoretical predictions (particularly in anticipating synthetic stability and other experimental challenges[48]), we emphasize that mechanistic understanding for guiding structural modifications holds greater significance than material screening *per se*. Furthermore, the iterative inverse engineering framework can be generalized to various scenarios in promoting systematic mechanism investigation and achieving rational design of novel materials.

Conflicts of interest

There are no conflicts to declare.

Data availability

The raw data for Fig. 2–6 and the legends in the SI can be found at **https://github.com/Yong-Q/Reverse-design-of-porous-materials.git**. The database constructed in the article is available at **https://doi.org/10.5281/zenodo.14910258**. The parameters for dataset computation are described in the SI, and the machine learning code is implemented using Scikit-Learn and TensorFlow.

Detailed computational methods and parameters, as well as supporting figures and tables. See DOI: **https://doi.org/10.1039/d5sc04332h**.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (No. 22108256 and 22478361) and Natural Science Foundation of Henan Province (No. 252300421176)). The computations were performed at the National Supercomputing Center in Zhengzhou, China.

References

1 J. L. Huang, Y. B. Li, H. Zhang, Z. Yuan, S. C. Xiang, B. L. Chen and Z. J. Zhang, A Microporous Hydrogen-Bonded Organic Framework Based on Hydrogen-Bonding Tetramers for Efficient Xe/Kr Separation, *Angew. Chem., Int. Ed.*, 2023, **62**, e202315987.

2 X. M. Liu, G. L. Liu, T. Fu, K. R. Ding, J. R. Guo, Z. R. Wang, W. Xia and H. Y. Shangguan, Structural Design and Energy and Environmental Applications of Hydrogen-Bonded Organic Frameworks: A Systematic Review, *Adv. Sci.*, 2024, **11**, 2400101.

3 W. Wang, Y. Shi, W. Chai, K. W. K. Tang, I. Pyatnitskiy, Y. Xie, X. Liu, W. He, J. Jeong, J.-C. Hsieh, A. R. Lozano, B. Artman, X. Shi, N. Hoefer, B. Shrestha, N. B. Stern, W. Zhou, D. W. McComb, T. Porter, G. Henkelman, B. Chen and H. Wang, H-bonded organic frameworks as ultrasound-programmable delivery platform, *Nature*, 2025, **638**, 401–410.

4 N. P. Franks, R. Dickinson, S. L. M. de Sousa, A. C. Hall and W. R. Lieb, How does xenon produce anaesthesia?, *Nature*, 1998, **396**, 324.

5 R. B. Lin, Y. B. He, P. Li, H. L. Wang, W. Zhou and B. L. Chen, Multifunctional porous hydrogen-bonded organic framework materials, *Chem. Soc. Rev.*, 2019, **48**, 1362–1389.
6 P. K. Thallapally, J. W. Grate and R. K. Motkuri, Facile xenon capture and release at room temperature using a metal–organic framework: a comparison with activated charcoal, *Chem. Commun.*, 2012, **48**, 347–349.
7 Z. Yuan, L. Chen, X. Zhou, L. Li, Y. Li, Y. Yang, Z. Zhou, Y. Chen, S. Xiang, B. Chen and Z. Zhang, Hydrogen-bonded organic framework with tailored pores prepared by enlarging the core size for high-performance Xe/Kr separation, *J. Mater. Chem. A*, 2023, **11**, 21857–21863.
8 J. Y. Yang, J. K. Wang, B. H. Hou, X. Huang, T. Wang, Y. Bao and H. X. Hao, Porous hydrogen-bonded organic frameworks (HOFs): from design to potential applications, *Chem. Eng. J.*, 2020, **399**, 125873.
9 Y. Liu, G. G. Chang, F. Zheng, L. H. Chen, Q. W. Yang, Q. L. Ren and Z. B. Bao, Hybrid Hydrogen-Bonded Organic Frameworks: Structures and Functional Applications, *Chem.–Eur. J.*, 2023, **29**, e202202655.
10 L. Li, L. Guo, Z. Zhang, Q. Yang, Y. Yang, Z. Bao, Q. Ren and J. Li, A Robust Squarate-Based Metal–Organic Framework Demonstrates Record-High Affinity and Selectivity for Xenon over Krypton, *J. Am. Chem. Soc.*, 2019, **141**, 9358–9364.
11 J. Francis Kurisingal, D. Won Kim and C. S. Hong, Effective approaches to boost Xe/Kr separation in Metal-Organic Frameworks: a review, *Coord. Chem. Rev.*, 2024, **507**, 215731.
12 S. A. Boer, M. Morshedi, A. Tarzia, C. J. Doonan and N. G. White, Molecular Tectonics: A Node-and-Linker Building Block Approach to a Family of Hydrogen-Bonded Frameworks, *Chem.–Eur. J.*, 2019, **25**, 10006–10012.
13 Z. Fan, Y. Zou, C. Liu, S. Xiang and Z. Zhang, Hydrogen-Bonded Organic Frameworks: Functionalized Construction Strategy by Nitrogen-Containing Functional Group, *Chem.–Eur. J.*, 2022, **28**, e202200422.
14 Z. Zheng, N. Rampal, T. J. Inizan, C. Borgs, J. T. Chayes and O. M. Yaghi, Large language models for reticular chemistry, *Nat. Rev. Mater.*, 2025, **10**, 1–13.
15 T. Song, M. Luo, X. Zhang, L. Chen, Y. Huang, J. Cao, Q. Zhu, D. Liu, B. Zhang, G. Zou, G. Zhang, F. Zhang, W. Shang, Y. Fu, J. Jiang and Y. Luo, A Multiagent-Driven Robotic AI Chemist Enabling Autonomous Chemical Research On Demand, *J. Am. Chem. Soc.*, 2025, **147**, 12534–12545.
16 G. Tom, S. P. Schmid, S. G. Baird, Y. Cao, K. Darvish, H. Hao, S. Lo, S. Pablo-García, E. M. Rajaonson, M. Skreta, N. Yoshikawa, S. Corapi, G. D. Akkoc, F. Strieth-Kalthoff, M. Seifrid and A. Aspuru-Guzik, Self-Driving Laboratories for Chemistry and Materials Science, *Chem. Rev.*, 2024, **124**, 9633–9732.
17 Y. Tian, X. F. Xu and J. Z. Wu, Thermodynamic Route to Efficient Prediction of Gas Diffusivity in Nanoporous Materials, *Langmuir*, 2017, **33**, 11797–11803.
18 J. Fu, Y. Tian and J. Z. Wu, Classical density functional theory for methane adsorption in metal-organic framework materials, *AlChE J.*, 2015, **61**, 3012–3021.
19 Q. Y. Yang, D. H. Liu, C. L. Zhong and J. R. Li, Development of Computational Methodologies for Metal-Organic Frameworks and Their Application in Gas Separations, *Chem. Rev.*, 2013, **113**, 8261–8323.
20 C. Lian, M. Janssen, H. Liu and R. van Roij, Blessing and Curse: How a Supercapacitor's Large Capacitance Causes its Slow Charging, *Phys. Rev. Lett.*, 2020, **124**, 076001.
21 Y. Lin, C. Lian, M. U. Berrueta, H. Liu and R. van Roij, Microscopic Model for Cyclic Voltammetry of Porous Electrodes, *Phys. Rev. Lett.*, 2022, **128**, 206001.
22 Y. Qiu, X. Zhang, Y. Tian and Z. Zhou, Machine learning promotes the development of all-solid-state batteries, *Chin. J. Struct. Chem.*, 2023, **42**, 100118.
23 G. Zhao, Y. Chen and Y. G. Chung, High-Throughput, Multiscale Computational Screening of Metal–Organic Frameworks for Xe/Kr Separation with Machine-Learned Parameters, *Ind. Eng. Chem. Res.*, 2023, **62**, 15176–15189.
24 P. Huang, Y. Leng, C. Lian and H. Liu, Porous-DeepONet: Learning the Solution Operators of Parametric Reactive Transport Equations in Porous Media, *Engineering*, 2024, **39**, 94–103.
25 H. Tao, S. Wang, H. Liu and C. Lian, Deep Neural Network Enhanced Mesoscopic Thermodynamic Model for Unlocking the Electrode/Electrolyte Interface, *Angew. Chem., Int. Ed.*, 2025, **64**, e202418447.
26 Y. Qiu, L. T. Chen, X. Zhang, D. H. Ping, Y. Tian and Z. Zhou, A universal machine learning framework to automatically identify high-performance covalent organic framework membranes for CH_4/H_2 separation, *AlChE J.*, 2024, **70**, 18575.
27 H. Park, S. Majumdar, X. Zhang, J. Kim and B. Smit, Inverse design of metal–organic frameworks for direct air capture of CO_2 via deep reinforcement learning, *Digital Discov.*, 2024, **3**, 728–741.
28 Y. Lim, J. Park, S. Lee and J. Kim, Finely tuned inverse design of metal-organic frameworks with user-desired Xe/Kr selectivity, *J. Mater. Chem. A*, 2021, **9**, 21175–21183.
29 Z. P. Yao, B. Sánchez-Lengeling, N. S. Bobbitt, B. J. Bucior, S. G. H. Kumar, S. P. Collins, T. Burns, T. K. Woo, O. K. Farha, R. Q. Snurr and A. Aspuru-Guzik, Inverse design of nanoporous crystalline reticular materials with deep generative models, *Nat. Mach. Intell.*, 2021, **3**, 76–86.
30 M. S. Zhou, A. Vassallo and J. Z. Wu, Toward the inverse design of MOF membranes for efficient D_2/H_2 separation by combination of physics-based and data-driven modeling, *J. Membr. Sci.*, 2020, **598**, 117675.
31 M. Zhou and J. Wu, Inverse design of metal–organic frameworks for C_2H_4/C_2H_6 separation, *npj Comput. Mater.*, 2022, **8**, 256.
32 B. Kim, S. Lee and J. Kim, Inverse design of porous materials using artificial neural networks, *Sci. Adv.*, 2020, **6**, eaax9324.
33 S. K. Elsaidi, M. H. Mohamed, A. S. Helal, M. Galanek, T. Pham, S. Suepaul, B. Space, D. Hopkinson, P. K. Thallapally and J. Li, Radiation-resistant metal-organic framework enables efficient separation of krypton fission gas from spent nuclear fuel, *Nat. Commun.*, 2020, **11**, 3103.
34 D. Banerjee, C. M. Simon, A. M. Plonka, R. K. Motkuri, J. Liu, X. Chen, B. Smit, J. B. Parise, M. Haranczyk and P. K. Thallapally, Metal–organic framework with optimally

selective xenon adsorption and separation, *Nat. Commun.*, 2016, **7**, ncomms11831.

35 X. Y. Song, Y. Wang, C. Wang, D. Wang, G. W. Zhuang, K. O. Kirlikovali, P. Li and O. K. Farha, Design Rules of Hydrogen-Bonded Organic Frameworks with High Chemical and Thermal Stabilities, *J. Am. Chem. Soc.*, 2022, **144**, 10663–10687.

36 Y. Hu, W.-K. Han, Y. Liu, R.-M. Zhu, X. Yan, H. Pang and Z.-G. Gu, Mechanochemical Transition from a Hydrogen-Bonded Organic Framework to Covalent Organic Frameworks, *ACS Mater. Lett.*, 2023, **5**, 2534–2541.

37 X. Song, Y. Wang, C. Wang, D. Wang, G. Zhuang, K. O. Kirlikovali, P. Li and O. K. Farha, Design Rules of Hydrogen-Bonded Organic Frameworks with High Chemical and Thermal Stabilities, *J. Am. Chem. Soc.*, 2022, **144**, 10663–10687.

38 M. O'Keeffe, M. A. Peskov, S. J. Ramsden and O. M. Yaghi, The Reticular Chemistry Structure Resource (RCSR) Database of, and Symbols for, Crystal Nets, *Acc. Chem. Res.*, 2008, **41**, 1782–1789.

39 S. Lee, B. Kim, H. Cho, H. Lee, S. Y. Lee, E. S. Cho and J. Kim, Computational Screening of Trillions of Metal-Organic Frameworks for High-Performance Methane Storage, *ACS Appl. Mater. Interfaces*, 2021, **13**, 23647–23654.

40 M. Aneke and M. Wang, Improving the Energy Efficiency of Cryogenic Air Separation Units (ASU) through Compressor Waste Heat Recovery using Direct Binary Heat Engine Cycle, *Comput. Chem. Eng.*, 2015, **37**, 2375–2380.

41 M. T. Kapelewski, J. Oktawiec, T. Runčevski, M. I. Gonzalez and J. R. Long, Separation of Xenon and Krypton in the Metal–Organic Frameworks M2(m-dobdc) (M=Co, Ni), *Israel J. Chem.*, 2018, **58**, 1138–1143.

42 S. U. Nandanwar, K. Coldsnow, V. Utgikar, P. Sabharwall and D. Eric Aston, Capture of harmful radioactive contaminants from off-gas stream using porous solid sorbents for clean environment – a review, *Chem. Eng. J.*, 2016, **306**, 369–381.

43 K. Deb, A. Pratap, S. Agarwal and T. Meyarivan, A fast and elitist multiobjective genetic algorithm: NSGA-II, *IEEE T. Evolut. Comput.*, 2002, **6**, 182–197.

44 Z. J. Deng and L. Sarkisov, Multi-Scale Computational Design of Metal-Organic Frameworks for Carbon Capture Using Machine Learning and Multi-Objective Optimization, *Chem. Mater.*, 2024, **36**, 9806–9821.

45 X. Wu and Y. Liu, Predicting Gas Adsorption without the Knowledge of Pore Structures: A Machine Learning Method Based on Classical Density Functional Theory, *J. Phys. Chem. Lett.*, 2023, **14**, 10094–10102.

46 L. He, Y. Li, L. Li, Z. Wang, Y. Chen, F. Yuan, G. Lan, C. Chen, S. Xiang, B. Chen and Z. Zhang, A Microporous Hydrogen-Bonded Organic Framework with Open Pyrene Sites Isolated by Hydrogen-Bonded Helical Chains for Efficient Separation of Xenon and Krypton, *Angew. Chem., Int. Ed.*, 2024, **7**, e202418917.

47 F.-A. Guo, K. Zhou, J. Liu, H. Wang and J. Li, Robust Hydrogen-Bonded Organic Framework with Four-Fold Interpenetration for Adsorptive Separation of C_2H_6/C_2H_4 and Xe/Kr, *Precis. Chem.*, 2023, **1**, 524–529.

48 Z. Zhang, F. Pan, S. A. Mohamed, C. Ji, K. Zhang, J. Jiang and Z. Jiang, Accelerating Discovery of Water Stable Metal−Organic Frameworks by Machine Learning, *Small*, 2024, **20**, 2405087.