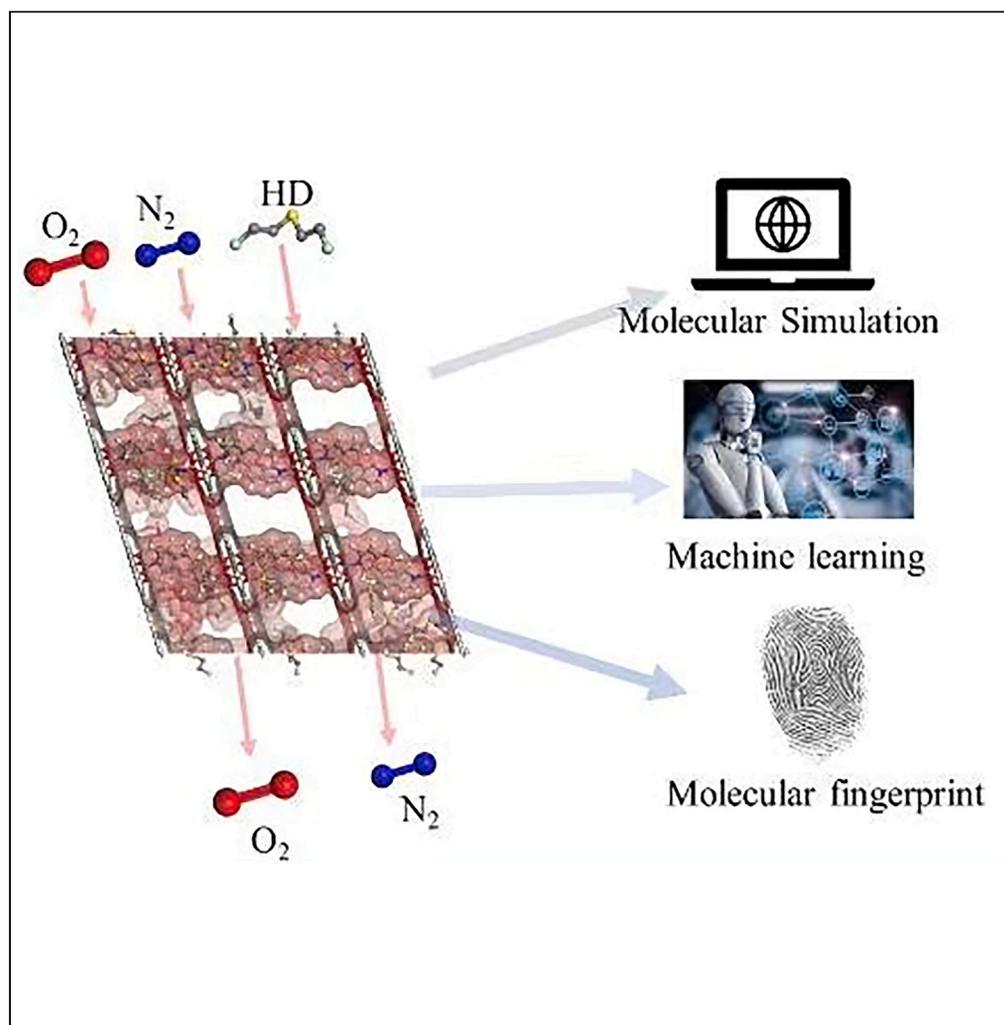


Article

Molecular fingerprint and machine learning enhance high-performance MOFs for mustard gas removal



Jing Ni, Jinfeng Li,
Shuhua Li, ..., Yue
Zhao, Hong Liang,
Zhiwei Qiao

sa11226532@mail.ustc.edu.cn
(Y.Z.)
lhong@gzhu.edu.cn (H.L.)
zqiao@gzhu.edu.cn (Z.Q.)

Highlights

Three key descriptors were found to influence the HD adsorption loading of MOFs

Ten high-performance MOFs were identified for the removal of mustard gas from air

Three crucial substructures were screened out to design the new materials

Random Forest and MACCS molecular fingerprinting can obtain a good predicted ability

Ni et al., iScience 27, 110042
June 21, 2024 © 2024 The
Authors. Published by Elsevier
Inc.
[https://doi.org/10.1016/
j.isci.2024.110042](https://doi.org/10.1016/j.isci.2024.110042)

Article

Molecular fingerprint and machine learning enhance high-performance MOFs for mustard gas removal

Jing Ni,¹ Jinfeng Li,¹ Shuhua Li,¹ He Zheng,² Zhongyuan Ming,¹ Li Li,² Heguo Li,² Shouxin Zhang,² Yue Zhao,^{2,*} Hong Liang,^{1,*} and Zhiwei Qiao^{1,3,4,*}

SUMMARY

Chemical warfare agents (CWAs), epitomized by the notoriously used mustard gas (HD), represent a class of exceptionally toxic chemicals whose airborne removal is paramount for battlefield safety. This study integrates high-throughput computational screening (HTCS) with advanced machine learning (ML) techniques to investigate the efficacy of metal-organic frameworks (MOFs) in adsorbing and capturing trace amounts of HD present in the air. Our approach commenced with a comprehensive univariate analysis, scrutinizing the impact of six distinct descriptors on the adsorption efficiency of MOFs. This analysis elucidated a pronounced correlation between MOF density and the Henry coefficient in the effective capture of HD. Then, four ML algorithms were employed to train and predict the performance of MOFs. The Random Forest (RF) algorithm demonstrates strong model learning and good generalization, achieving the best prediction result of 98.3%. In a novel exploratory stride, we incorporated a 166-bit MACCS molecular fingerprinting (MF) to identify critical functional groups within adsorbents. From the top 100 MOFs analyzed, 22 optimal functional groups were identified. Leveraging these insights, we designed three innovative substructures, grounded in these key functional groups, to enhance HD adsorption efficiency. In this work, the combination of MF and ML could provide a new direction for efficient screening of MOFs for the capture of HD in the air. The outcomes of this study offer substantial potential to revolutionize the domain of CWA capture. This represents a significant stride toward developing practical solutions that enhance both environmental protection and battlefield security.

INTRODUCTION

Chemical warfare agents (CWAs) are highly toxic chemicals that were employed during World Wars I and II, resulting in a substantial number of casualties among both combatants and civilians. CWAs are utilized in warfare to inflict irreversible and fatal injuries upon the enemy,¹ exhibiting a considerably higher lethality rate compared to thermal weapons like firearms. The category of CWAs encompasses nerve agents, blister agents, agents causing lung damage, and blood agents, all of which are relatively inexpensive to produce and transport, leading to profound and incalculable damage.² Among them, mustard gas (HD),^{3,4} as a blister agent within the realm of chemical warfare agents, stands out as one of the earliest and most widely used. HD can enter the human body through the skin, respiratory tract, and eyes, causing extensive harm, and it induces erosion of the skin and various tissue cells, resulting in skin and cell necrosis, eye damage, severe blistering of the skin in affected individuals, and in extreme cases, permanent blindness and death.⁵ During World War II, the Japanese army deployed HD in their aggressive war against China, resulting in tens of thousands of casualties. Consequently, numerous researchers have extensively investigated techniques for handling chemical warfare agents like HD, aiming to ensure effective protection for humans against these agents.^{6,7}

The pursuit of a suitable adsorbent for HD currently stands as a focal point of investigation. Traditional adsorbents, such as zeolites and activated carbon, display limited adsorption capacity, along with low productivity and high energy costs, thereby constraining their practical applications.⁸ Graphene, a flexible conductor with a super-large specific surface area and a rich pore structure, is explored as an adsorbent material and finds widespread use in various fields, including energy, biology, and medicine. However, graphene, being a two-dimensional monolayer structure, is prone to causing π - π stacking during the adsorption process due to its inherent characteristics, leading to a reduction in the adsorption efficiency of graphene.⁹ Given the potentially fatal damage caused by toxic agents, it is imperative to identify a material with robust adsorption capacity and excellent chemical stability. From an energy perspective, the cost of the adsorbent must also be considered.

¹Guangzhou Key Laboratory for New Energy and Green Catalysis, School of Chemistry and Chemical Engineering, Guangzhou University, Guangzhou 510006, China

²State Key Lab NBC Protect Civilian, Beijing 102205, P.R. China

³Joint Institute of Guangzhou University & Institute of Corrosion Science and Technology, Guangzhou University, Guangzhou 510006, China

⁴Lead contact

*Correspondence: sa11226532@mail.ustc.edu.cn (Y.Z.), lhong@gzhu.edu.cn (H.L.), zqiao@gzhu.edu.cn (Z.Q.)

<https://doi.org/10.1016/j.isci.2024.110042>



Metal-organic frameworks (MOFs), a class of one, two, or three-dimensional crystalline hybridized porous materials with a designable pore structure, are currently the focus of researchers' attention.^{10,11} MOFs exhibit highly tunable pore sizes, larger porosities, and higher selectivity than conventional porous materials,¹² meanwhile, their excellent properties can also be utilized in gas storage, catalytic hydrolysis, drug loading, etc. Lee et al.¹³ reported a MOF-fiber composite material; synthesized using a novel low-temperature hybrid solvent, and in this material, an aluminum porphyrin-based metal-organic skeleton (Al-PMOF) was attached to the surface of polymerized knitted fabrics, forming a solid membrane for the photocatalytic detoxification of HD simulants (CEES). This photocatalytic detoxification material provides a novel approach to producing protective clothing, masks, and other protective equipment. Xiao et al.¹⁴ introduced the design concept of metal-organic framework filters (MOFilters) for the adsorption of toxic gases, CWAs, volatile organic compounds (VOCs), and other environmental challenges. They also summarized the fundamental processes of filtering, adsorbing, and degrading pollutants to facilitate the better design and optimization of MOF filters. Florencia et al.¹⁵ examined the adsorption of CEES by 10 Zr-MOFs with different open metal sites, secondary building unit connectivity, and surface area/pore volume (specific surface area) and found that each of these factors affected the capture adsorption of CEES. Based on this study, researchers can choose the right topology and thus the right MOF based on the properties of the target adsorbate. Martos et al.¹⁶ selected 2,932 MOFs for the adsorption study of chemical warfare agents and their simulants, and to avoid the impact of competitive water adsorption, they chose 156 hydrophobic MOFs and identified optimal materials for adsorbing chemical warfare agents such as sarin, HD, and soman. Emelianova et al.¹⁷ conducted Monte Carlo simulations on sarin and three simulants—dimethyl methyl phosphonate (DIMP), diisopropyl fluorophosphate (DIFP), and dimethyl methyl phosphonate (DMMP), they selected DMMP as the simulant most closely resembling the adsorption characteristics of sarin. Carmen Montoro et al.¹⁸ investigated the adsorption of MOF-5 on HD, a class of hazardous volatile gases, and suggested that the pore size of the MOF and the hydrophobicity of its surface are important factors affecting adsorption. Yajiao Hao et al.¹⁹ selected MOFs and their derivatives as oxidants, employing visible or ultraviolet light to oxidize organic sulfur compounds, including simulators of HD, into less toxic sulfoxide byproducts. Moghadam et al.²⁰ performed high throughput calculations on 3,385 MOFs and established structure-mechanical stability relationships for MOF materials. Pardakhti et al.²¹ investigated the application of machine learning (ML) algorithms to forecast the methane adsorption capabilities of MOFs and believed ML is a quicker and more precise alternative solution for screening adsorbents. Borboudakis et al.²² utilized ML as a more efficient screening method to predict high-performance materials, through training on data obtained from high-throughput computational screening (HTCS). Fanourgakis et al.²³ conducted training in machine learning (ML) algorithms on MOFs and achieved successful prediction of the adsorption properties across different material families, thus substantiating the method's universality and transferability. Wang et al.^{24,25} employed interpretable ML methodologies to expedite the discovery of promising MOFs for the selective separation of ethane and ethylene. They proposed an integrated ML framework that associates MOF structures with gas adsorption capabilities through structural decomposition, feature integration, and predictive modeling. Zhang et al.²⁶ introduced a two-step integrated process design methodology involving MOF synthesis and pressure/vacuum swing adsorption (P/VSA) for gas separation purposes.

In this study, the adsorption and purification capabilities of 31,399 MOFs for HD were investigated utilizing HTCS and ML. Univariate analyses were performed to characterize the relationships between the MOF's feature descriptors and the adsorption, selectivity, and trade-off variables. Three adsorption metrics of MOFs were also analyzed using machine learning training to select the 10 best MOFs. Finally, the optimal functional group was determined using the molecular fingerprinting (MF) technique, and efforts were made to summarize the excellent structures, providing directions for the design of high-performance MOFs.

RESULTS AND DISCUSSION

Univariate analysis

Univariate analysis serves as an initial inquiry into dataset properties, constituting a fundamental aspect of data examination. It primarily focuses on elucidating the attributes and behaviors of individual variables while delineating their optimal ranges. In this study, a total of 31,399 HMOFs were selected, and the structural descriptors (VSA, LCD, ρ , and φ) as well as energetic descriptors (K and Q_{st}^0) of the MOFs were calculated through molecular dynamics simulations. Subsequently, univariate analyses were performed on the adsorption properties of MOF descriptors and HD within the MOFs. To comprehensively evaluate the performance of the MOFs, a variable, TSN (Trade-off between $S_{X/(Y1 + Y2)}$ and N_X), was introduced, defined by the formula:

$$TSN = N_X \times \ln \left(\frac{S_{X/(Y1+Y2)}}{Y1+Y2} \right) \quad (\text{Equation 1})$$

Here, X represents HD, while Y1 and Y2 respectively denote nitrogen (N_2) and oxygen (O_2). N represents adsorption capacity, and S represents selectivity.

TSN is employed to balance the relationship between selectivity and adsorption by applying a logarithmic operation to the selectivity, thereby minimizing the difference between the two orders of magnitude. Selectivity and adsorption have become common criteria, and the utilization of TSN allows for a better synthesis of the two variables. This approach was previously used to evaluate zeolite adsorption by Shah et al.²⁷

As depicted in Figures 1A and 1C, the adsorption capacity (N) and selectivity (S) of mustard gas exhibit an increasing and then decreasing trend with the increase of LCD. This pattern arises because when the LCD is very small, gas molecules cannot penetrate the interior of the MOF, resulting in nearly zero values for both N and S . As the LCD increases, gas molecules progressively enter the MOF's interior, leading to a gradual rise in both N and S . However, with further increases in LCD, the contact area between gas molecules and the MOF pore wall

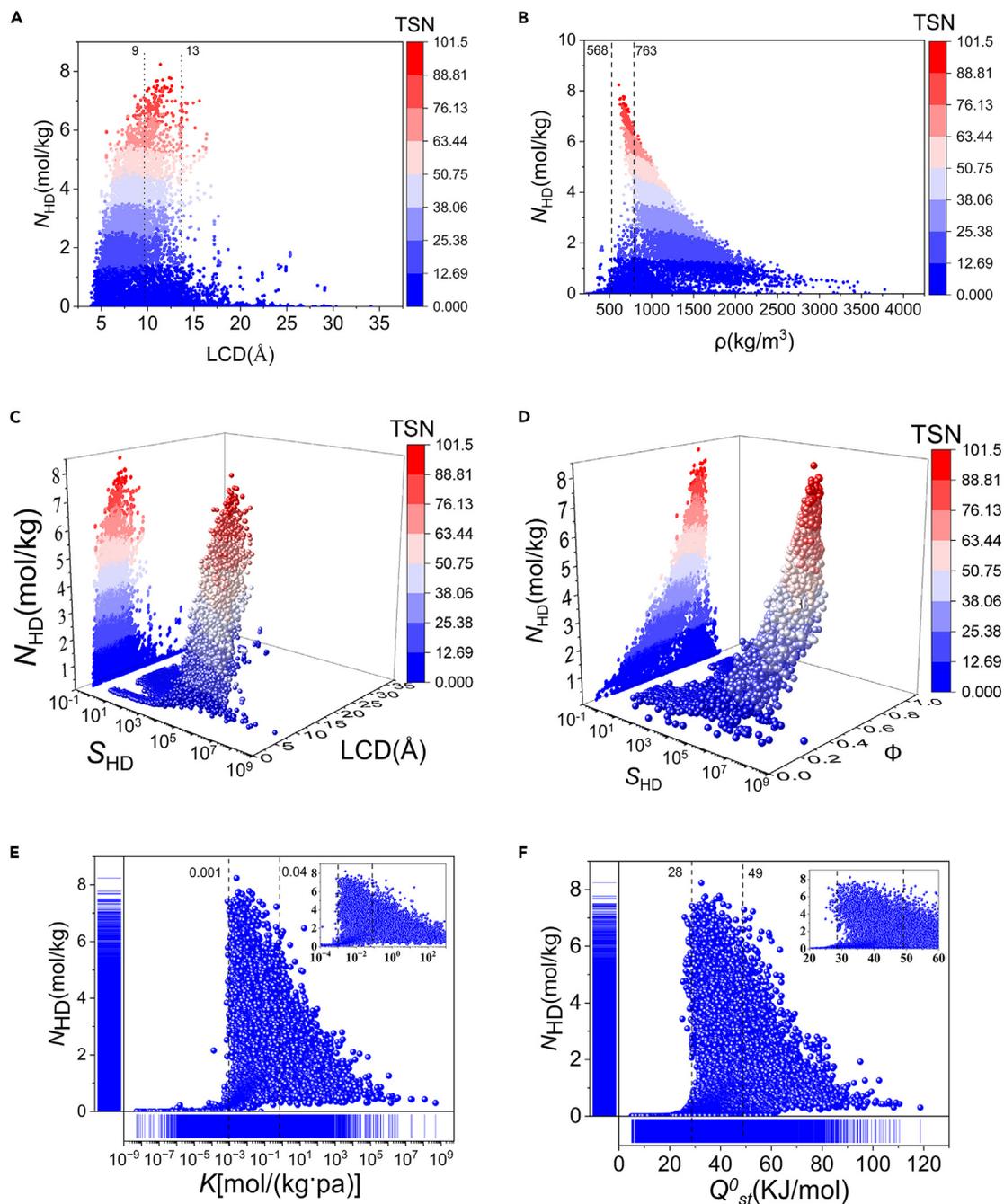


Figure 1. Univariate analysis chart

(A) The relationship between N and LCD.

(B) The relationship between ρ and N .

(C) The relationship between LCD, S , and N .

(D) The relationship between ϕ , S , and LCD.

(E) The relationship between K and N .

(F) The relationship between Q_{st}^0 and N .

The colored of the dot in A–D represent the value of the TSN.

decreases, weakening the force between them and diminishing the wrapping effect, ultimately causing a decrease in N . Concurrently, S decreases as the pore diameter becomes larger, which is unfavorable for differentiating the gas mixture.²⁸ In the LCD range of (9, 13), both N and TSN reach their peaks, representing the optimal adsorption interval.

As depicted in Figure 1B, when VSA approaches zero, it suggests a negligible surface area of the MOF. At this point, both N and S are close to zero. As VSA increases, the accessible area of gas molecules with the MOF pore wall expands, resulting in an increase in both N and S for the target gas. However, with excessively large VSA, the adsorption space and the number of adsorption sites for the main components (N_2 and O_2) in the mixed gas increase. As VSA continues to grow, the relative contact area between mustard gas and the MOF pore walls decreases, impeding mustard gas adsorption. Consequently, both N and S decrease. The optimal adsorption range for VSA is (1565, 2043), where both the adsorption quantity and the equilibrium value TSN of the MOF reach their highest values. Moreover, in Figure 1B, ρ and LCD exhibit a similar trend, with the adsorption amount increasing with ρ . When ρ becomes excessively large, the adsorption amount decreases sharply and the optimal adsorption interval for ρ is (568, 763). Compared to other descriptors, the optimal range of ρ contains the fewest MOFs and thus represents the narrowest optimal range. Therefore, it appears that ρ and the performance metric TSN exhibit a strong correlation.

In addition to the four structural descriptors, two energy descriptors (K and Q_{st}^0) are considered. Figure 1E depicts the relationship between K and N . When K is zero, the weak interaction between the MOF framework and HD molecules leads to poor MOF performance, with both N and S approaching zero. As K increases, it becomes evident that HD can be easily adsorbed, resulting in a rapid increase in the adsorption amount (N). However, when K is too large, N starts to decrease gradually, and the optimal adsorption interval for K is determined to be (0.001, 0.04). Figure 1F illustrates the trend of the N_{HD} - Q_{st}^0 relationship. When Q_{st}^0 is zero, both N and S are zero. This is because mustard gas molecules need some energy to enter the interior of the MOF. The optimal adsorption range for Q_{st}^0 is found to be (28, 49), a conclusion utilized by Qiao et al.²⁹ in the adsorption separation of CO_2 . Upon closer inspection, the optimal adsorption range of K is much smaller than that of Q_{st}^0 , implying that K is more relevant than Q_{st}^0 for the performance index TSN.

Based on the univariate analysis, it is initially concluded that there exists a strong correlation between ρ and K about the adsorption performance of the MOF. However, the univariate analysis can only identify the most important influencing factors and cannot fully elucidate the combined effects among each descriptor and the adsorption performance of the material. Therefore, the ML method is adopted to further analyze this complex adsorption system.

Machine learning

Machine learning encapsulates the interplay among multiple variables, enabling validation not only of univariate analyses but also of the intricate weighting of multiple variables, which mutually scrutinize and harmonize one another, thereby collectively influencing performance outcomes. In this study, the adsorption performance of 31,399 HMOFs concerning mustard gas was analyzed using four ML regression algorithms: Random Forest (RF), Extreme Gradient Boosting (XGB), Decision Tree (DT), and Light Gradient Boosting Regression (LGBR). For each ML algorithm, the entire dataset was divided into training and test sets, with 70% allocated for model training and the remainder for model testing. To ensure the stability and homogeneity of the model, k -fold cross-validation ($k = 5$) was employed. The principle of k -fold cross-validation involves dividing the entire dataset into k groups, with one group chosen as the test set each time and the remaining $k - 1$ groups used as the prediction set. The average of the cross-validation results was chosen as the assessment of the model's accuracy. For more information, please refer to Figure S3. Furthermore, the evaluation metrics for model prediction accuracy include the determination coefficient (R^2), mean absolute error (MAE), and root-mean-square error (RMSE).

As illustrated in Figure 2, the four ML models—RF, XGB, DT, and LGBR—are employed to predict the test set, resulting in an R^2 exceeding 0.95, signifying the robust predictive performance of these algorithms. Traditionally, individuals tend to assess prediction accuracy based on model complexity. However, this approach may not be universally correct since different algorithms are suitable for different systems. Following ML algorithm judging criteria, higher R^2 values and lower MAE or RMSE values indicate heightened prediction accuracy and more reliable results. The predictive effectiveness of these four algorithms is ranked as RF > LGBR > XGB > DT, highlighting the RF algorithm as the most suitable for this system. RF demonstrates robust model learning ability and good generalization, as evidenced by its selection as the optimal ML model in previous research.²¹ In Figure 2C, we use the DT algorithm as the relevant ML model involves DT regressor (which can fit multiple MOFs into the same leaf node based on their feature values). This is an expected result, similar to previous work by Demir et al.³⁰ Moreover, DT is not the optimal algorithm in this system we study. Furthermore, TSN is reintroduced as a performance metric in this study. This decision is motivated by the fact that choosing S as a performance metric results in lower prediction accuracy due to the uneven distribution of S data. Using TSN as a trade-off variable provides a more rationalized representation of the data distribution. Employing N and S as performance indicators for prediction and comparing the magnitudes of the three values of R^2 , MAE, and RMSE further supports the identification of RF as the algorithm most suitable for the model. The ranking of prediction accuracy for the three performance metrics is TSN > N > S , reinforcing the appropriateness of considering TSN as the primary performance metric.

In Figure 3A, the RF algorithm exhibits the highest R^2 value, as well as the lowest MAE and RMSE values in predicting the test set. Consequently, RF is identified as the optimal algorithm in conjunction with the aforementioned study. Subsequently, we employ the RF algorithm to conduct a relative importance analysis of the MOF structure, the relative importance share of the six descriptors is illustrated in Figure 3B. RF assesses descriptor importance by considering the ratio between the sum of changes in the mean square error (MAE) resulting from splits on each descriptor and the sum of branching nodes. The larger the relative importance share of a descriptor, the greater its contribution to MOF adsorption is implied. As depicted in the figure, the relative importance percentage of K is 61%, significantly surpassing other descriptors. Following closely, Q_{st}^0 holds a relative importance percentage of 25%, making it the second most crucial descriptor. Among the four structural descriptors (φ , ρ , LCD, VSA), ρ has the highest relative importance. The relative importance ranking of descriptors for MOF is $K > Q_{st}^0 > \rho > LCD > \varphi > VSA$. Notably, the importance proportions of LCD and φ are closely aligned, implying their similar impact on MOF adsorption performance. In summary, the energy descriptor K and the structural descriptor ρ play pivotal roles in the adsorption

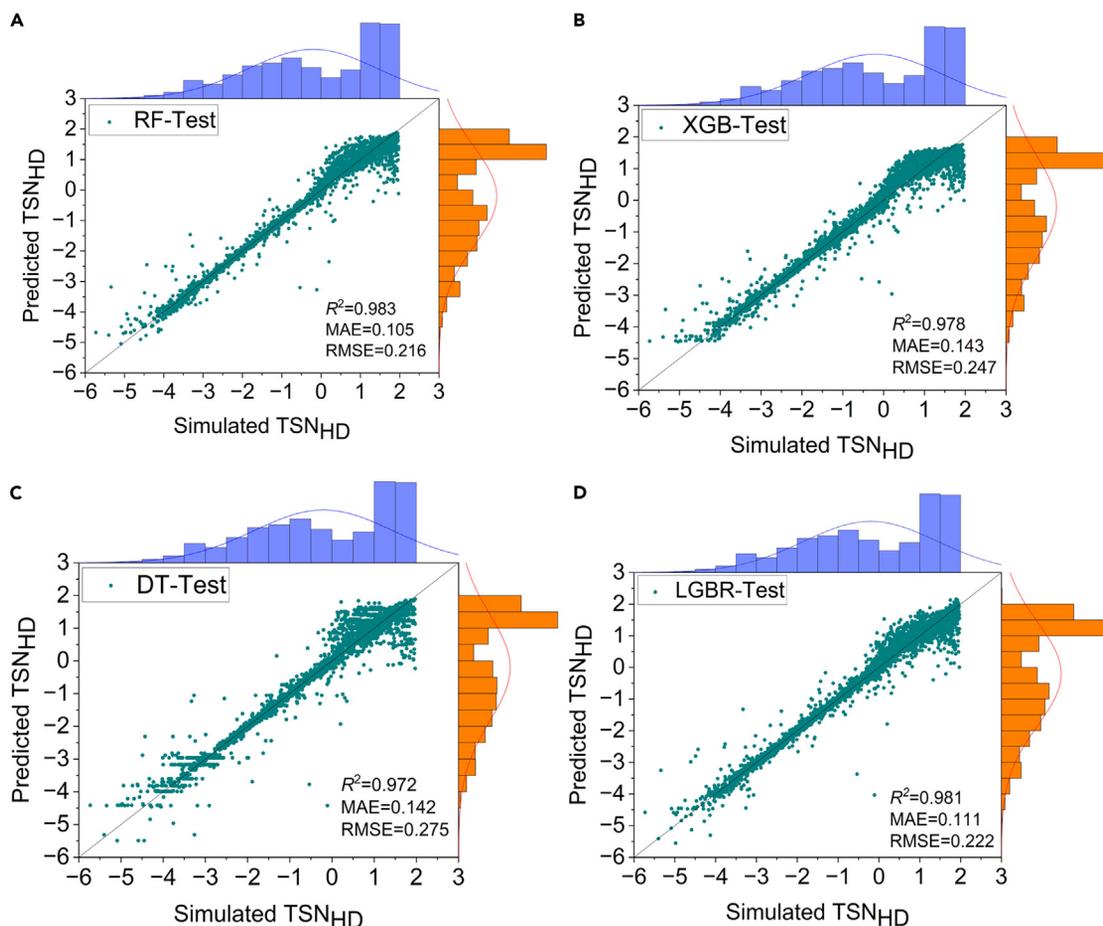


Figure 2. Machine learning graphs

Performance of four algorithms (A) RF, (B) XGB, (C) DT, and (D) LGBR for predicting TSN in test set. The bars height is used to represent the density of the points, and the smoothed curve is a normal distribution curve.

selectivity of MOFs, corroborating our earlier conclusions from univariate analysis. Moreover, the correlation between K and the adsorption of toxic agents by MOFs provides insights for the screening and design of novel MOFs.

Table 1 presents the ten optimal MOFs selected based on TSN. In the preceding univariate analysis, we outlined approximate outstanding ranges for MOF descriptors. Notably, the descriptors of these ten optimal MOFs fall within the designated ranges, further affirming the conclusions drawn from our univariate analysis.

Molecular fingerprint

MF is a technique that disassembles molecules into substructures and converts the substructures into numerical representations. We can combine ML and MF to screen for superior substructures that could potentially be the key to designing superior materials. In this study, MOF design criteria were explored by analyzing the MOF structure, and MACCS fingerprints were generated through the combination of the OpenBabel and PaDEL Descriptor packages. OpenBabel³¹ is an open-source toolkit utilized for describing chemical data across various chemical languages and processing chemical data in fields such as materials science, organic chemistry, and computer chemistry. PaDEL Descriptor³² is a dedicated software designed for calculating molecular descriptors and fingerprints, being the only one that supports over 90 different molecular file formats. The initial step involved the conversion of the "cif" file format of the HMOF to the "sdf" file format using the OpenBabel package. Subsequently, the MOF in "sdf" format was transformed into MACCS fingerprints using the PaDEL Descriptor package. MACCS is a commonly applied fingerprinting method, and its fingerprints have been utilized in prior studies to investigate the structural properties of MOFs.³³ MACCS is a two-dimensional descriptor-based fingerprint with a length of 166, covering a wide range of chemical features. MACCS fingerprints also have their limitations. They are binary codes and cannot represent specific numerical properties of molecules, such as information about the charge and bond length of atoms.

In the previous study, K and Q_{st}^0 were identified as the two most crucial descriptors. These descriptors, along with the MACCS of 166 bits, were amalgamated to predict the three properties of adsorbents (N , S , and TSN) using the RF algorithm. The primary goal was to forecast

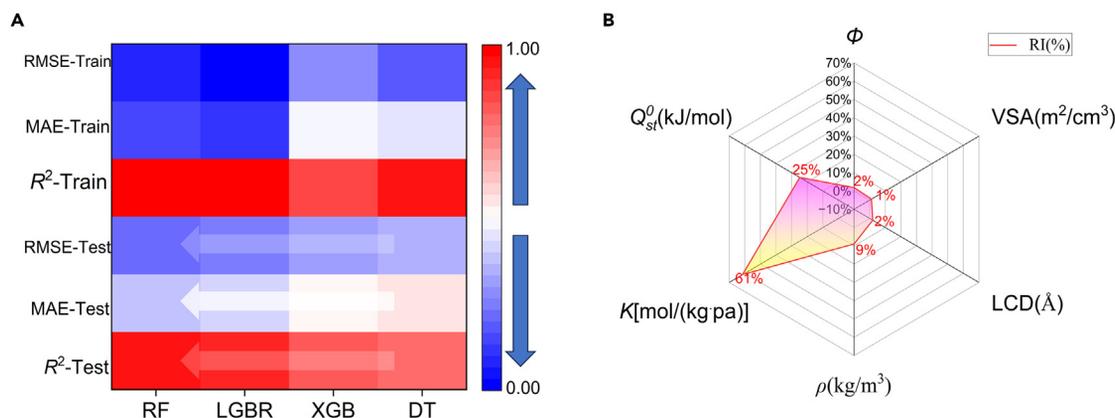


Figure 3. Algorithmic contrast heatmaps and related importance maps

(A) Comparison of algorithm performance. The color represents the value of RMER, MAE and R². (B) Relative importance of the six descriptors of MOF.

performance indexes through the MF-assisted ML method, conducting a thorough examination of the impact of each MOF structure on performances. The objective was to identify the structure exerting the most significant influence on performances, thereby providing insights into the design of novel MOFs.

As depicted in Figure 4A on the test set, when *N* and TSN were designated as performance indicators, the R² values consistently surpassed 0.8. This observation underscores the appropriateness of employing MACCS fingerprints. However, due to the dispersed nature of *S* values, predictive accuracy diminishes when *S* is chosen as the performance indicator, aligning with findings from our prior work. According to ML algorithm evaluation criteria, and in comparison, with *N* and *S*, TSN emerges as the more suitable performance indicator. Simultaneously, as illustrated in Figure 4B, a comprehensive analysis of metal content was conducted in all MOFs and the top 100 MOFs. Notably, within HMOFs, only four metals are present: Zn, Cu, V, and Zr.

Firstly, the fingerprints of 31,399 HMOFs were converted, and the frequency of each fingerprint bit was tallied. The top 10 bits with the highest frequency were identified as 164, 157, 159, 22, 124, 146, 130, 102, 140, and 16. These conventional molecular fingerprints appear both in good-performing MOFs and poor-performing MOFs. From the perspective of big data analysis, these fingerprints have no significant effect on finding MOFs with enhanced mustard gas adsorption performance. Detailed information can be found in Table S6. To mitigate interference from traditional fingerprints, MOFs were ranked based on the TSN performance metric. The top 100 MOFs were meticulously selected, and the frequency of each fingerprint bit within this subset was precisely tallied. The ten bits with the highest frequencies were 164, 159, 157, 146, 130, 124, 102, 22, 148, and 140. Significantly, these ten bits closely overlapped with our previously identified common bits, hindering a definitive assessment of their exceptional nature. In this context, a frequency difference approach was employed to define exceptional fingerprint bits. By subtracting the frequency of each fingerprint bit in the top 100 MOFs from its frequency across all 31,399 MOFs, bits with differences exceeding 10% were filtered out. As illustrated in Figure 4C, a higher difference in frequency indicates a concentration of these fingerprint bits within the top 100 MOFs, suggesting their potential significance as critical structures influencing the exceptional characteristics of MOFs. The resulting exceptional fingerprint bits selected were 103, 7, 31, 76, 75, 121, 125, 145, 99, 50, 45, 144, 162, 26, 142, 161, 122, 158, 94, 163, 117, 126, and 134.

Table 1. Ten MOFs with optimal performance for mustard gas adsorption

NO	ID ^a	ϕ	VSA/(m ² /cm ³)	LCD/(Å)	ρ /(kg/m ³)	K/(mol/kg/Pa)	Q_{st}^0 /(kJ/mol)	N/(mol/Kg)	<i>S</i>	TSN
1	6004092	0.796	1787	11.381	610.7	0.0025	32.5	8.235	222736	101.4
2	6001603	0.763	1670	10.905	622.5	0.0054	38.3	7.669	289057	96.4
3	26666	0.836	1824	12.520	679.0	0.0032	34.0	7.755	236527	96.0
4	5050062	0.821	2011	11.054	669.2	0.0196	39.8	7.680	219301	94.4
5	5049995	0.843	1903	12.341	663.0	0.0134	39.51	7.774	184916	94.2
6	6001799	0.741	1847	10.137	655.8	0.0007	28.6	7.4 14	333129	94.3
7	5062975	0.824	1828	10.891	697.3	0.0014	29.0	7.405	295457	93.3
8	22878	0.807	1937	11.892	689.0	0.0019	31.7	7.435	262443	93.0
9	22303	0.828	1947	12.101	699.6	0.0034	34.7	7.436	258154	92.7
10	27210	0.811	1904	11.063	665.1	0.0089	37.0	7.508	225861	92.6

^aStructure for HMOFs could be checked by the IDs in the HMOF database: <http://hmof.northwestern.edu>.³⁰

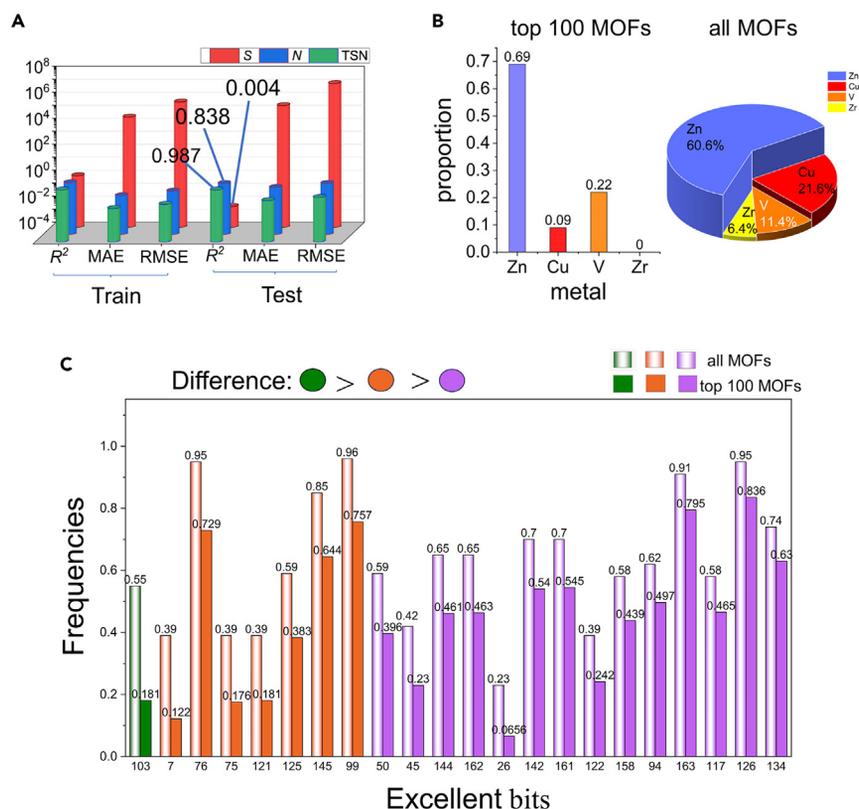


Figure 4. Molecular fingerprints and metal content maps

(A) Performance of the RF algorithm for three adsorption properties indicators.

(B) Proportion of MOFs with the four metal centers in the top 100 MOFs and in all MOFs.

(C) The fingerprint bits with difference in frequency of occurrence between top 100 MOFs dataset and all MOFs dataset exceeding 10%.

VB, VIB, and VIIB are classified as transition metal elements (Group 7). Among HMOFs, only the element V belongs to these three subgroups. Transition metals can function as binding sites for HD adsorption, and the integration of a high density of transition metals into the MOF structure can enhance the stability of MOFs.^{33,34} As indicated in Table 2, outstanding MOFs for HD adsorption should contain no fewer than two six-membered rings and aromatic rings. Hydrogen bonding in the six-membered rings improves the adsorption efficiency of MOFs,³⁵ while aromatic rings provide additional adsorption sites for MOFs, enhancing the affinity between HD and MOFs, which is favorable for MOF adsorption.³⁶ The negatively charged halogen atoms in the framework (Group 103 and Group 134) can coordinate with the metal ions in the framework, dispersing positive charges.³⁷ Additionally, the inclusion of halogen elements enhances the polarizability of the framework, promoting interaction between the framework and HD molecules.³⁸ The C=C double bond is a nonpolar bond, and HD is also a nonpolar molecule in this study. This results in an instantaneous dipole when they are nearby, and the dispersive force generated between the framework and HD molecules facilitates HD adsorption. N-containing heterocycles can directly coordinate with metals through deprotonation, yielding electron-neutral zeolite-like structures. Mulliken population analysis reveals that N is more negatively charged than pristine C atoms, and the addition of N atoms allows for stronger electrostatic interactions between the HD and the MOF.³⁹

The structure of MOFs is digitized and the public fingerprints of outstanding materials are microanalyzed by MF. The adsorption simulation diagrams of MOFs are depicted in Figure 5, with three MOFs selected from the top 10. Figures 5B and 5D represent different angles of the same MOF, with black dots indicating mustard gas. Simultaneously, their structures are decomposed one by one using the bit number of MACCS fingerprints. These four diagrams reveal a higher concentration of mustard gas near the functional groups, thereby verifying the influence of functional groups on MOF adsorption. In Figures 5A and 5B, a majority of the mustard gases cluster near chlorine (Cl) atoms, polycyclic rings, C=C bonds, and oxygen atoms. Similarly, in Figures 5C and 5D, most mustard gases cluster near polycyclic rings, N-containing heterocycles, and Cl atoms. Furthermore, in all four adsorption diagrams, certain mustard gases appear fragmented and are situated close to the center of the pore. This fragmentation is attributed to the combined influence of multiple functional groups along the surrounding pore walls. Notably, these fragmented mustard gases do not exhibit specific proximity to a particular functional group. These observations reinforce the robust correlation between the superior fingerprint positions and the adsorption properties of the MOF. For the MOF-HD system, it is suggested that pivotal information determining MOF performance may be concealed within the molecular framework (MF).

As depicted in Figure 6, three novel organic ligands have been designed, incorporating the majority of the exemplary fingerprint structures outlined in Table 2. All three organic ligands encompass (a) aromatic rings, (b) double bonds, (c) halogens, (d) nitrogen-containing

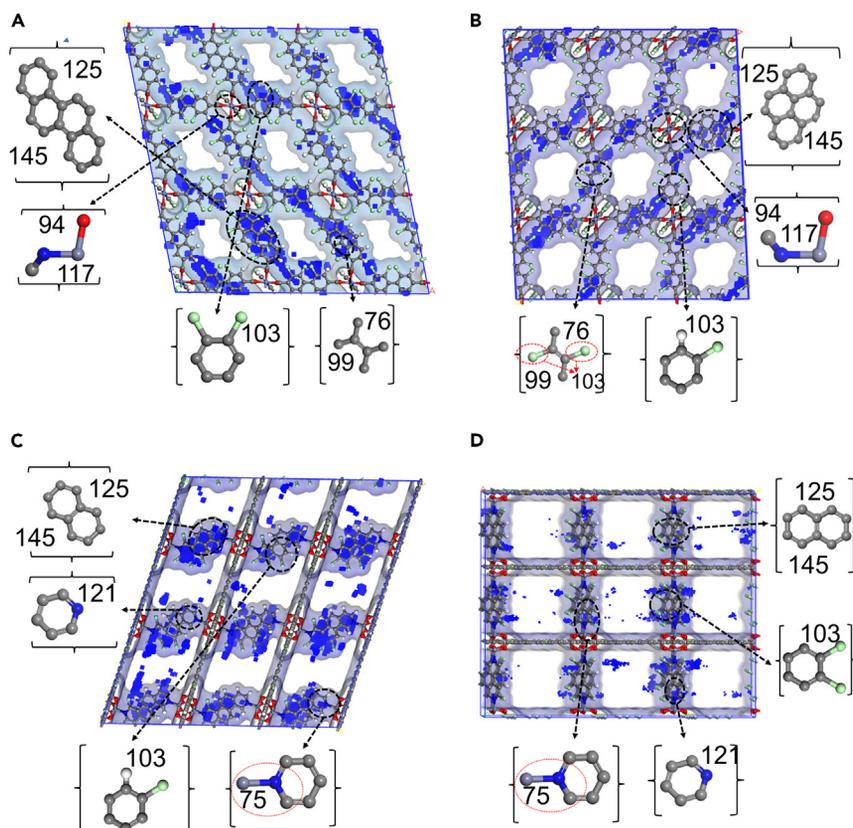


Figure 5. The adsorption simulation diagrams of three high-performance MOFs

(A) 26666, (B) and (D) 5049995, (C) 5050062. (B), and (D) are the same MOF from different views. Blue dots represent HD.

influencing MOF adsorption are the energy descriptors Q_{st}^0 and K . Subsequently, four machine learning models—RF, XGB, DT, and LGBR—were employed to predict the three performance metrics. The results indicated that the RF algorithm was deemed more suitable for the system, with TSN exhibiting superior performance compared to the other two metrics. Regarding HD adsorption, the top 10 high-performance MOFs were screened, and it was observed that the descriptors of these optimal MOFs fell within the range outlined in the univariate analysis. Finally, the molecular fingerprinting technique was employed to generate MACCS fingerprints. Utilizing a combination of ML and MF, the top 22 optimal fingerprint bits were selected. These positions corresponded to excellent structures such as -Cl, C=C, N-containing heterocycles, aromatic rings, and oxygen atoms. Three new organic ligands were designed based on the optimal fingerprint bits, providing a directional approach for designing adsorbents for HD.

Limitations of the study

The limitation of this study should also be noted. The three new organic ligands we have designed have not been subjected to theoretical simulations or experimental validation at this time.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
 - Lead contact
 - Materials availability
 - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
 - Molecular model

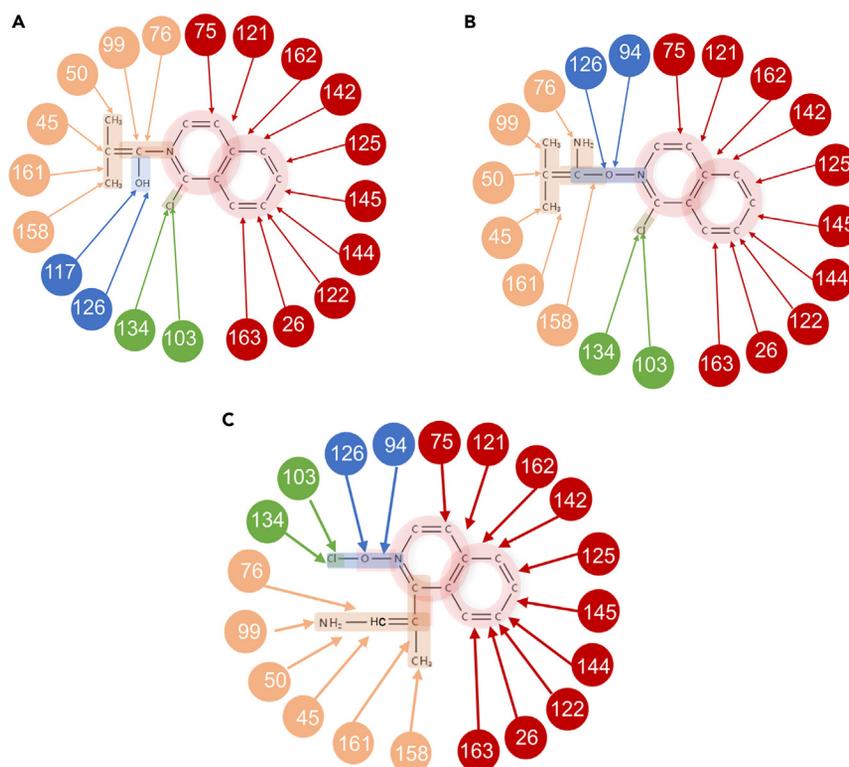


Figure 6. The three substructures employed in the design of high-performance MOFs

(A–C) Three substructures with different functional groups. The number represents the bit of MFs in Table 2. The color represents the different combinations in Table 2.

- Force field parameter
- Simulation method
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.isci.2024.110042>.

ACKNOWLEDGMENTS

The authors gratefully thank the National Natural Science Foundation of China (21978058), the Pearl River Talent Recruitment Program (2019QN01L255), the Natural Science Foundation of Guangdong Province (2022A1515011446, 2020A1515010800), the China Postdoctoral Science Foundation (2022M720868), Guangzhou Municipal Science and Technology Project (202102020875 and 202201020169), and R & D Program of Joint Institute of GZHU & ICoST (GI202102) for their financial support.

AUTHOR CONTRIBUTIONS

Conceptualization, software, validation, investigation, writing—original draft, writing—review and editing, project administration, visualization, and supervision, J.N.; formal analysis, S.L.; investigation, H.L.; methodology, validation, writing—review and editing, and funding acquisition, Z.M.; software, writing—review and editing, and funding acquisition, H.Z.; writing—review and editing and funding acquisition, L.L., H.L., S.Z., and Y.Z.; conceptualization, methodology, software, resources, data curation, writing—original draft, writing—review and editing, and funding acquisition, Z.Q.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: December 31, 2023

Revised: April 30, 2024

Accepted: May 16, 2024

Published: May 21, 2024

REFERENCES

- Zhu, B., Sheng, R., Chen, T., Rodrigues, J., Song, Q.-H., Hu, X., and Zeng, L. (2022). Molecular engineered optical probes for chemical warfare agents and their mimics: Advances, challenges and perspectives. *Coord. Chem. Rev.* 463, 214527. <https://doi.org/10.1016/j.ccr.2022.214527>.
- Chan, J., Yeung, R., and Tang, S. (2002). An overview of chemical warfare agents. *Hong Kong J. Emerg. Med.* 9, 201–205. <https://doi.org/10.1177/102490790200900404>.
- Kullander, F., Wåsterby, P., and Landström, L. (2016). Ultraviolet Raman scattering from persistent chemical warfare agents. *SPIE* 9824, 53–61. <https://doi.org/10.1117/12.2223601>.
- Paromov, V., Kumari, S., Brannon, M., Kanaparthi, N.S., Yang, H., Smith, M.G., and Stone, W.L. (2011). Protective Effect of Liposome-Encapsulated Glutathione in a Human Epidermal Model Exposed to a Mustard Gas Analog. *J. Toxicol.* 2011, 109516. <https://doi.org/10.1155/2011/109516>.
- Yang, H., Guo, Y., Han, S., Xi, H., Lin, H., Yuan, R., Ding, Z., Long, J., and Lin, Q. (2023). Synthesis of Mo-doped symbiotic mixture of Bi₂WO₆ and Bi₂O₆(OH)₃(NO₃)₃·1.5H₂O nanosheets with enhanced photocatalytic degradation for mustard gas simulator 2-CEEs. *Res. Chem. Intermed.* 49, 3381–3393. <https://doi.org/10.1007/s11164-022-04926-z>.
- Yang, X., Huang, Q., Zhang, L., Li, L., Chen, Y., Wang, W., Liang, H., Wu, Y., Zheng, H., Zhao, Y., and Qiao, Z. (2023). Computational screening and machine learning of hydrophobic metal-organic frameworks for removal of chemical warfare agents from air. *Appl. Mater. Today* 31, 101738. <https://doi.org/10.1016/j.apmt.2023.101738>.
- Wang, W., Zhang, L., Cai, C., Li, S., Liang, H., Wu, Y., Zheng, H., and Qiao, Z. (2023). Machine learning assisted high-throughput computational screening of MOFs for the capture of chemical warfare agents from the air. *Separation and Purification Technology* 325, 124546. <https://doi.org/10.1016/j.seppur.2023.124546>.
- Zhang, G., Liang, Y., Cui, G., Dou, B., Lu, W., Yang, Q., and Yan, X. (2023). Grand canonical Monte Carlo simulation of the adsorption and separation of carbon dioxide and methane using functionalized Mg-MOF-74. *Energy Rep.* 9, 2852–2860. <https://doi.org/10.1016/j.egy.2023.01.121>.
- Wang, M., Zhou, J., Zhang, M., Li, S., Li, L., Chen, G., Huang, X., Liu, G., Wang, J., and Xu, D. (2023). Facile fabrication of bioinspired hierarchical porous MOFs for selective adsorption of Congo red and Malachite green from vegetables and fruits juices. *Environmental Technology & Innovation* 30, 103132. <https://doi.org/10.1016/j.eti.2023.103132>.
- Yan, Y., Shi, Z., Li, H., Li, L., Yang, X., Li, S., Liang, H., and Qiao, Z. (2022). Machine learning and in-silico screening of metal-organic frameworks for O₂/N₂ dynamic adsorption and separation. *Chem. Eng. J.* 427, 131604. <https://doi.org/10.1016/j.cej.2021.131604>.
- Yuan, X., Deng, X., Cai, C., Shi, Z., Liang, H., Li, S., and Qiao, Z. (2021). Machine learning and high-throughput computational screening of hydrophobic metal-organic frameworks for capture of formaldehyde from air. *Green Energy Environ.* 6, 759–770. <https://doi.org/10.1016/j.gee.2020.06.024>.
- Bai, X., Shi, Z., Xia, H., Li, S., Liu, Z., Liang, H., Liu, Z., Wang, B., and Qiao, Z. (2022). Machine-Learning-Assisted High-Throughput computational screening of Metal-Organic framework membranes for hydrogen separation. *Chem. Eng. J.* 446, 136783. <https://doi.org/10.1016/j.cej.2022.136783>.
- Lee, D.T., Jamir, J.D., Peterson, G.W., and Parsons, G.N. (2020). Protective Fabrics: Metal-Organic Framework Textiles for Rapid Photocatalytic Sulfur Mustard Simulant Detoxification. *Matter* 2, 404–415. <https://doi.org/10.1016/j.matt.2019.11.005>.
- Xiao, H., Low, Z.-X., Gore, D.B., Kumar, R., Asadnia, M., and Zhong, Z. (2022). Porous metal-organic framework-based filters: Synthesis methods and applications for environmental remediation. *Chem. Eng. J.* 430, 133160. <https://doi.org/10.1016/j.cej.2021.133160>.
- Son, F.A., Wasson, M.C., Islamoglu, T., Chen, Z., Gong, X., Hanna, S.L., Lyu, J., Wang, X., Idrees, K.B., Mahle, J.J., et al. (2020). Uncovering the Role of Metal-Organic Framework Topology on the Capture and Reactivity of Chemical Warfare Agents. *Chem. Mater.* 32, 4609–4617. <https://doi.org/10.1021/acs.chemmater.0c00986>.
- Matito-Martos, I., Moghadam, P.Z., Li, A., Colombo, V., Navarro, J.A.R., Calero, S., and Fairen-Jimenez, D. (2018). Discovery of an Optimal Porous Crystalline Material for the Capture of Chemical Warfare Agents. *Chem. Mater.* 30, 4571–4579. <https://doi.org/10.1021/acs.chemmater.8b00843>.
- Emelianova, A., Reed, A., Basharova, E.A., Kolesnikov, A.L., and Gor, G.Y. (2023). Closer Look at Adsorption of Sarin and Simulants on Metal-Organic Frameworks. *ACS Appl. Mater. Interfaces* 15, 18559–18567. <https://doi.org/10.1021/acsami.3c02713>.
- Montoro, C., Linares, F., Procopio, E.Q., Senkowska, I., Kaskel, S., Galli, S., Masciocchi, N., Barea, E., and Navarro, J.A.R. (2011). Capture of Nerve Agents and Mustard Gas Analogues by Hydrophobic Robust MOF-5 Type Metal-Organic Frameworks. *J. Am. Chem. Soc.* 133, 11888–11891. <https://doi.org/10.1021/ja2042113>.
- Hao, Y., Papazyan, E.K., Ba, Y., and Liu, Y. (2022). Mechanism-Guided Design of Metal-Organic Framework Composites for Selective Photooxidation of a Mustard Gas Simulant under Solvent-Free Conditions. *ACS Catal.* 12, 363–371. <https://doi.org/10.1021/acscatal.1c04755>.
- Moghadam, P.Z., Rogge, S.M., Li, A., Chow, C.-M., Wieme, J., Moharrami, N., Aragonés-Anglada, M., Conduit, G., Gomez-Gualdrón, D.A., Van Speybroeck, V., and Fairen-Jimenez, D. (2019). Structure-Mechanical Stability Relations of Metal-Organic Frameworks via Machine Learning. *Matter* 1, 219–234. <https://doi.org/10.1016/j.matt.2019.03.002>.
- Pardakhti, M., Moharreri, E., Wanik, D., Suib, S.L., and Srivastava, R. (2017). Machine Learning Using Combined Structural and Chemical Descriptors for Prediction of Methane Adsorption Performance of Metal Organic Frameworks (MOFs). *ACS Comb. Sci.* 19, 640–645. <https://doi.org/10.1021/acscombsci.7b00056>.
- Borboudakis, G., Stergiannakos, T., Frysalis, M., Klontzas, E., Tsamardinos, I., and Froudakis, G.E. (2017). Chemically intuited, large-scale screening of MOFs by machine learning techniques. *npj Comput. Mater.* 3, 40. <https://doi.org/10.1038/s41524-017-0045-8>.
- Fanourgakis, G.S., Gkagkas, K., Tyljanakis, E., and Froudakis, G.E. (2020). A Universal Machine Learning Algorithm for Large-Scale Screening of Materials. *J. Am. Chem. Soc.* 142, 3814–3822. <https://doi.org/10.1021/jacs.9b11084>.
- Wang, Z., Zhou, T., and Sundmacher, K. (2022). Interpretable machine learning for accelerating the discovery of metal-organic frameworks for ethane/ethylene separation. *Chem. Eng. J.* 444, 136651. <https://doi.org/10.1016/j.cej.2022.136651>.
- Wang, Z., Zhou, Y., Zhou, T., and Sundmacher, K. (2022). Identification of optimal metal-organic frameworks by machine learning: Structure decomposition, feature integration, and predictive modeling. *Comput. Chem. Eng.* 160, 107739. <https://doi.org/10.1016/j.compchemeng.2022.107739>.
- Zhang, X., Zhou, T., and Sundmacher, K. (2022). Integrated metal-organic framework (MOF) and pressure/vacuum swing adsorption process design: MOF matching. *AIChE J.* 68, e17788. <https://doi.org/10.1002/aic.17788>.
- Shah, M.S., Tsapatsis, M., and Siepmann, J.I. (2016). Frontispiece: Identifying Optimal Zeolitic Sorbents for Sweetening of Highly Sour Natural Gas. *Angew. Chem. Int. Ed. Engl.* 55, 5938–5942. <https://doi.org/10.1002/anie.201600612>.
- Watanabe, T., and Sholl, D.S. (2012). Accelerating Applications of Metal-Organic Frameworks for Gas Adsorption and Separation by Computational Screening of Materials. *Langmuir* 28, 14114–14128. <https://doi.org/10.1021/la301915s>.
- Qiao, Z., Zhang, K., and Jiang, J. (2016). In silico screening of 4764 computation-ready, experimental metal-organic frameworks for CO₂ separation. *J. Mater. Chem. A Mater.* 4, 2105–2114. <https://doi.org/10.1039/C5TA08984K>.
- Demir, H., and Keskin, S. (2023). Revealing acetylene separation performances of anion-pillared MOFs by combining molecular simulations and machine learning. *Chem.*

- Eng. J. 464, 142731. <https://doi.org/10.1016/j.cej.2023.142731>.
31. O'boyle, N.M., Banck, M., James, C.A., Morley, C., Vandermeersch, T., and Hutchison, G.R. (2011). Open Babel: An open chemical toolbox. *J. Cheminform.* 3, 33. <https://doi.org/10.1186/1758-2946-3-33>.
 32. Yap, C.W. (2011). PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints. *J. Comput. Chem.* 32, 1466–1474. <https://doi.org/10.1002/jcc.21707>.
 33. Yuan, X., Li, L., Shi, Z., Liang, H., Li, S., and Qiao, Z. (2022). Molecular-fingerprint machine-learning-assisted design and prediction for high-performance MOFs for capture of NMHCs from air. *Advanced Powder Materials* 1, 100026. <https://doi.org/10.1016/j.apmate.2021.12.002>.
 34. Wen, M., Li, G., Liu, H., Chen, J., An, T., and Yamashita, H. (2019). Metal-organic framework-based nanomaterials for adsorption and photocatalytic degradation of gaseous pollutants: recent progress and challenges. *Environ. Sci.: Nano* 6, 1006–1025. <https://doi.org/10.1039/C8EN01167B>.
 35. Park, J.M., and Jung, S.H. (2020). A remarkable adsorbent for removal of bisphenol S from water: Aminated metal-organic framework, MIL-101-NH₂. *Chem. Eng. J.* 396, 125224. <https://doi.org/10.1016/j.apmate.2021.12.002>.
 36. Tang, H., and Jiang, J. (2021). In silico screening and design strategies of ethane-selective metal-organic frameworks for ethane/ethylene separation. *AIChE J.* 67, e17025. <https://doi.org/10.1002/aic.17025>.
 37. Hu, Q.-H., Wang, Y.-G., Gao, X., Shi, Y.-Z., Lin, S., Liang, R.-P., and Qiu, J.-D. (2023). Halogen microregulation in metal-organic frameworks for enhanced adsorption performance of ReO₄/TcO₄. *J. Hazard Mater.* 446, 130744. <https://doi.org/10.1016/j.jhazmat.2023.130744>.
 38. Wu, W., Su, J., Jia, M., Li, Z., Liu, G., and Li, W. (2020). Vapor-phase linker exchange of metal-organic frameworks. *Sci. Adv.* 6, eaax7270. <https://doi.org/10.1126/sciadv.aax7270>.
 39. Lu, X., Tang, Y., Yang, G., and Wang, Y.-Y. (2023). Porous functional metal-organic frameworks (MOFs) constructed from different N-heterocyclic carboxylic ligands for gas adsorption/separation. *CrystEngComm* 25, 896–908. <https://doi.org/10.1039/D2CE01667B>.
 40. Wilmer, C.E., Leaf, M., Lee, C.Y., Farha, O.K., Hauser, B.G., Hupp, J.T., and Snurr, R.Q. (2011). Large-scale screening of hypothetical metal-organic frameworks. *Nat. Chem.* 4, 83–89. <https://doi.org/10.1038/nchem.1192>.
 41. Dubbeldam, D., Calero, S., Ellis, D.E., and Snurr, R.Q. (2016). RASPA: molecular simulation software for adsorption and diffusion in flexible nanoporous materials. *Mol. Simulat.* 42, 81–101. <https://doi.org/10.1080/08927022.2015.1010082>.
 42. Willems, T.F., Rycroft, C.H., Kazi, M., Meza, J.C., and Haranczyk, M. (2012). Algorithms and tools for high-throughput geometry-based analysis of crystalline porous materials. *Microporous Mesoporous Mater.* 149, 134–141. <https://doi.org/10.1016/j.micromeso.2011.08.020>.
 43. Hantal, G., Jedlovsky, P., Hoang, P.N.M., and Picaud, S. (2007). Calculation of the Adsorption Isotherm of Formaldehyde on Ice by Grand Canonical Monte Carlo Simulation. *J. Phys. Chem. C* 111, 14170–14178. <https://doi.org/10.1021/jp0742564>.
 44. Kadantsev, E.S., Boyd, P.G., Daff, T.D., and woo, T.K. (2013). Fast and Accurate Electrostatics in Metal Organic Frameworks with a Robust Charge Equilibration Parameterization for High-Throughput Virtual Screening of Gas Adsorption. *J. Phys. Chem. Lett.* 4, 3056–3061. <https://doi.org/10.1021/jz401479k>.
 45. Qiao, Z., Xu, Q., Cheetham, A.K., and Jiang, J. (2017). High-Throughput Computational Screening of Metal-Organic Frameworks for Thiol Capture. *J. Phys. Chem. C* 121, 22208–22215. <https://doi.org/10.1021/acs.jpcc.7b07758>.
 46. Rappe, A.K., Casewit, C.J., Colwell, K.S., Goddard, W.A., and Skiff, W.M. (1992). UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *J. Am. Chem. Soc.* 114, 10024–10035. <https://doi.org/10.1021/ja00051a040>.
 47. Ewald, P.P. (1921). Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Ann. Phys.* 369, 253–287. <https://doi.org/10.1002/andp.19213690304>.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
HMOFs	http://hmofs.northwestern.edu	
Software and algorithms		
ML code(RF,XGB,DT,LGBR)	scikit-learn: machine learning in Python — scikit-learn 1.4.2 documentation	
python	Welcome to Python.org	
OpenBabel	Install Open Babel — Open Babel 3.0.1 documentation (open-babel.readthedocs.io)	
PaDEL-Descriptor	http://padel.nus.edu.sg/software/padeldescriptor	
Materials Studio	BIOVIA Materials Studio Dassault Systèmes (3ds.com)	

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and codes should be directed to and will be fulfilled by the lead contact, Zhiwei Qiao (zqiao@gzhu.edu.cn)

Materials availability

This study did not generate new unique reagents.

Data and code availability

- All data for this study were publicly available and could be found in [method details](#).
- All code could be found in the github (ML code: <https://github.com/dancingBit/MLCode/tree/MLCode02>).
- Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Our study is computational science research.

METHOD DETAILS

Molecular model

In this study, the database of 137953 hypothetical MOFs (HMOFs) designed by Snurr's group⁴⁰ was utilized. From this database, 31,399 hydrophobic MOFs were selectively chosen to mitigate the potential competitive adsorption of water vapor. Six descriptors (volumetric surface area (VSA(m²/cm³)), largest cavity diameter (LCD(Å)), density (ρ (kg/m³)), void fraction (φ), heat of adsorption (Q_{st}^0 (kJ/mol)), and Henry coefficient (K (mol/kg/Pa))) were applied to characterize the structure and energy of the MOFs. Using the RASPA⁴¹ software package, φ was estimated by detecting the framework utilizing a He atom with a 2.58 Å diameter. The VSA was computed by rolling an N₂ probe with a diameter of 3.64 Å on the framework surfaces of the MOFs. The Zeo++ software package⁴² was employed for estimating LCD and ρ . Within RASPA, the NVT-MC scheme (where N is the number of particles, V is the volume of the system, and T is the temperature of the system) was utilized to simulate K and Q_{st}^0 .

Force field parameter

Giant canonical Monte Carlo (GCMC) simulations⁴³ were employed to model the adsorption separation properties of 31,399 HMOFs in N₂ and O₂ concerning HD interactions. Given the static condition of the MOF, where its atoms remain stationary, considerations for the potential energies associated with bond angle bending, double angle bending, and bond expansion of the MOFs are omitted. However, it is crucial to account for interactions between MOFs and mustard gas, as well as the potential energy associated with mustard gas. These interactions are described by the Lennard-Jones (LJ) potential function for inter-atomic interactions:

$$u_{LJ+elec}(r) = \sum 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \sum \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \quad (\text{Equation 2})$$

We employed the MEPO-QEq⁴⁴ method to estimate the atomic charges of MOFs, and this method facilitates a swift and efficient assessment of electrostatic interactions. The accuracy of this method has been validated in a study by Qiao et al.⁴⁵ The LJ parameters for MOFs are outlined in [Table S1](#), utilizing the widely accepted universal force field (UFF).⁴⁶

Simulation method

A diverse selection of MOFs was utilized as adsorbents in our extensive computational study. In adherence to common assumptions associated with computations involving a substantial quantity of adsorbents, the MOF structures were treated as rigid. The focus of our simulations was on the adsorption of a ternary gas mixture (HD, N₂, and O₂), where N₂ and O₂ represented the primary components of air. The proportions of N₂, O₂, and HD were assumed to be 0.789:0.21:0.001, as outlined in [Table S2](#). The force field parameters for HD, N₂, and O₂ were derived from the TraPPE force field. The simulation was conducted at room temperature (298 K) and atmospheric pressure (101.325 kPa), and the conditions were maintained consistently throughout the simulation (constant chemical potential, volume, and temperature). The interactions of MOFs with HD, N₂, and O₂ were calculated using the Lorentz-Bertelot rule, and the LJ interactions were computed with a spherical truncation radius of 12 Å, including a long-range correction. Electrostatic interactions between the frameworks and the gas molecules, as well as between the gas molecules, were determined by summation via the Ewald method.⁴⁷ In each MOF, the GCMC simulation consisted of 100,000 cycles, the first 50,000 cycles were dedicated to equilibrating the simulated system, and the last 50,000 cycles were used to calculate the system average. Each cycle comprised *n* simulated motions (where *n* is the number of adsorbate molecules), including rotation, translation, exchange, and regeneration motions. The exchange motions, in turn, included insertion and deletion. Each GCMC simulation was conducted independently, and all simulations were performed using the RASPA software package.

QUANTIFICATION AND STATISTICAL ANALYSIS

Python was used for machine learning analysis, OpenBabel and PaDEL Descriptor were used to disassemble the molecular fingerprints, and Materials Studio was used to simulate the adsorption of mustard gas by MOF.