

High-Throughput Screening of Metal–Organic Frameworks for Ethane–Ethylene Separation Using the Machine Learning Technique

Prosun Halder and Jayant K. Singh*

 Cite This: *Energy Fuels* 2020, 34, 14591–14597

 Read Online

ACCESS |

 Metrics & More

 Article Recommendations

 Supporting Information

ABSTRACT: A hybrid approach combining machine learning algorithms with molecular simulation is utilized to screen hypothetical metal–organic framework (h-MOF) database for the best material to separate ethane (C_2H_6) and ethylene (C_2H_4). In particular, we rationalized the relation between structural and chemical properties of h-MOF with the C_2H_6/C_2H_4 selectivity. 8% h-MOFs were chosen randomly from the h-MOF dataset as a training set. The simulations were conducted at 298 K and 1 bar using a multicomponent grand-canonical Monte Carlo method to obtain the C_2H_6/C_2H_4 selectivity. Based on the training set, the random forest (RF) model was developed to predict the selectivity of the rest of the h-MOFs. Among all the chemical and structural properties, void fraction plays a significant role in predicting the equilibrium C_2H_6/C_2H_4 selectivity. The trained machine learning model can reasonably predict the C_2H_6/C_2H_4 selectivity of the remaining h-MOF materials with an RF score of 0.89. Four h-MOFs have shown the best performance, which was compared with the previously discovered materials. The top four h-MOFs were further simulated at different pressures to obtain the adsorption isotherms. Further, the energy contribution of secondary building units and the local density profiles were analyzed to understand the enhanced interaction between h-MOF atoms and C_2H_6 .

INTRODUCTION

The worldwide production of over 150 million tons makes ethylene (C_2H_4) the leading organic compound produced industrially. Primarily, ethylene is used as a feedstock for petrochemical industries wherein it undergoes a polymerization reaction to produce polyethylene, which is widely used plastic containing polymer chains with different chain lengths. Industrially, it is produced by steam cracking and thermal decomposition of naphtha or ethane (C_2H_6), in which a certain amount of C_2H_6 residue coexists in the product that is required to separate to produce polymer-grade C_2H_4 . The traditional industrial separation process is cryogenic distillation for ethane/ethylene separation, one of the most energy-intensive chemical industry processes. This process requires large distillation columns at high pressure because of their similar volatilities and size.^{1,2} Hence, there is a need to find an alternative technique to minimize energy consumption.

Among the technologies recently developed, adsorbent-based gas separation through pressure swing adsorption, temperature swing adsorption, and membrane-based separation are promising technologies to replace the traditional industrial separation methods. Previously, numerous researchers have investigated various types of nanoporous materials for different applications.^{3–7}

Utilizing this simple approach, an ideal adsorbent should have high C_2H_4 selectivity and high C_2H_4 uptake. In this context, numerous researchers have studied the adsorption phenomena of C_2H_4 , C_2H_6 , and their mixture on various types of adsorbents using experiments and in-silico methods. Some of the adsorbents are γ - Al_2O_3 ,⁸ ITQ-59 zeolitic material,⁹ metal–organic framework (MOF) $Fe_2(O_2)(dobdc)$,¹⁰ and so forth.

The most studied nanoporous materials are MOF,¹¹ covalent organic frameworks,¹² activated carbon,¹³ zeolite,^{14,15} zeolitic imidazolate framework,¹⁶ and porous polymer network.¹⁷ Among the various adsorbents, MOF has attracted particular interest because of its easy synthesis, high surface area, and porosity. Several studies have focused on discovering the best MOF material for CO_2 ,¹¹ CH_4 ¹⁸ adsorption, and the separation of CO_2/CH_4 ,¹⁹ CO_2/N_2 .²⁰ Numerous studies were conducted on the development of potential adsorbents for the separation of C_2H_4/C_2H_6 . For example, Chen et al. reported an iron-based MOF PCN-250,²¹ which has a selectivity of 2.0 at 298 K and 1 bar pressure. Similarly, Pires et al. reported the C_2H_6/C_2H_4 selectivity of the most common MOF, that is, IRMOF-8²² at 298 K and 318 K, up to a pressure of 1000 kPa. The selectivity values of IRMOF-8 range between 1.6 and 3.4 depending on temperature and pressure. Until now, microporous MOF $Fe_2(dobdc)$ developed by Bloch et al.¹⁰ has the highest C_2H_6/C_2H_4 selectivity, which is 4.3. Discovery of new potential MOFs for a particular application is an ongoing research area. However, for a computational study, MOFs can be constructed from modular molecular building blocks, typically metal-clusters, and organic linkers. These building blocks can be assembled to form an almost unlimited number of MOFs. The hypothetical MOF (h-MOF) database generation method was inspired by the work

Received: September 10, 2020

Revised: October 16, 2020

Published: November 9, 2020



on database creation of hypothetical zeolites by Earl and Deem.²³ Using this approach, Wilmer et al.²⁴ created a database of h-MOFs. The database of experimentally discovered and computationally generated MOFs is in millions, and the best material for C₂H₆/C₂H₄ separation is still an open question. To find a suitable material for a specific application, experimentally synthesizing the vast materials database is practically not viable. Similarly, a direct molecular simulation approach for each fictitious or experimentally realized material is an inefficient way to use computational resources and time. Therefore, a cost-effective screening technique is imminent to find the best MOF for the C₂H₆/C₂H₄ separation.

In this context, a hybrid approach combining machine learning algorithms with molecular simulations could be an effective way to screen materials. It is well known that MOFs' chemical and structural properties can directly affect the gas adsorption and gas separation. The mathematical models can predict the amount of adsorbed gas or the selectivity based on the chemical and structural properties. In the literature, there have been several studies that showed the power of machine learning techniques. For example, Haranczyk et al.²⁵ used the random forest (RF) algorithm to screen the nanoporous material genome (NMG) database of about 670,000 nanoporous materials for xenon/krypton separations. Similarly, Smit et al.²⁶ used a neural network algorithm to screen the same NMG database containing over 850,000 materials to explore the limits of H₂ storage. There are several works in the literature in which machine learning-based techniques were used, viz., for the selection of compounds that bind to proteins,²⁷ MOF analogues for water adsorption,²⁸ zeolitic materials for carbon capture,²⁹ and so forth.

In this work, our focus is to identify the best MOF materials from the h-MOF database using molecular simulations coupled with a machine learning algorithm. In particular, we will rationalize the relationship between structural and chemical properties of h-MOF with the C₂H₆/C₂H₄ selectivity. This hybrid approach will further provide us with the guideline for designing a suitable tool that will fill the gap between molecular simulation and machine learning algorithms.

MODEL AND METHODOLOGY

To train and validate the machine learning algorithm, we used the h-MOFs database created by the Snurr group, which is available at <https://www.materialsproject.org/porous/www.materialsproject.org/porous/>.²⁴ The database consists of 137,953 h-MOFs. These MOFs were generated from a library of building blocks. Further, the generated MOFs were compared with the experimental and energetically optimized structures. The Forcite module of Materials Studio was used to optimize the geometry of each MOF structure. The structural and chemical descriptors are often correlated with the selectivity or adsorption capacity of any porous material. Hence, we calculated several structural and chemical descriptors to leverage the machine learning model's accuracy. The list of chemical and structural properties in this work is tabulated in Table 1 and Table 2, respectively. Zeo++,³⁰ an open-source software, was used to calculate all the structural descriptors. Also, it can provide information about the regions which are accessible to the adsorbed gas. The h-MOFs with zero accessible surface area were removed from the dataset to avoid unnecessary time and resources to test the materials that are the worst materials for gas adsorption or separation study. After removing the h-MOF

Table 1. Chemical Descriptors of MOFs

descriptor	description
hydrogen (H)	number of hydrogen atoms per unit cell
carbon (C)	number of carbon atoms per unit cell
nitrogen (N)	number of nitrogen atoms per unit cell
oxygen (O)	number of oxygen atoms per unit cell
fluorine (F)	number of fluorine atoms per unit cell
chlorine (Cl)	number of chlorine atoms per unit cell
bromine (Br)	number of bromine atoms per unit cell
vanadium (V)	number of vanadium atoms per unit cell
copper (Cu)	number of copper atoms per unit cell
zinc (Zn)	number of zinc atoms per unit cell
zirconium (Zr)	number of zirconium atoms per unit cell
total degree of unsaturation	$[(C \times 2) + 2 - H]/2$
degree of unsaturation	total degree of unsaturation/C
metallic percentage	$(\text{number of metal atoms}/C) \times 100$
oxygen-to-metal ratio	$[2 \times O]/\text{number of metal atoms}$
nitrogen-to-oxygen ratio	N/O

Table 2. Structural Descriptors of MOFs

descriptor	description
crystal density (ρ)	mass of crystalline material per volume
accessible surface area (a)	accessible surface area along pore wall
void fraction (ϵ_v)	fractional of materials that is free volume
largest included sphere diameter (D_i)	largest sphere to fit inside the material
largest free sphere diameter (D_f)	largest sphere to percolate through material

having a 100% nonaccessible surface area, the machine learning algorithm was used on the rest of the 115,302 h-MOF materials.

The equilibrium adsorption data of C₂H₆ and C₂H₄ in MOF simulations were calculated using multicomponent grand-canonical Monte Carlo (m-GCMC). 9224 h-MOF materials (8% of the h-MOF dataset) were chosen as a training set. A 50:50 molar mixture of C₂H₆ and C₂H₄ was considered at 298 K and 1 bar total pressure for the study. The Monte Carlo moves consist of particle exchange (insertion/deletion), particle translation, particle rotation, and particle identity exchange which were used with probabilities 0.6, 0.15, 0.15, and 0.1, respectively. Each Monte Carlo simulation was performed with 2000 cycles to reach the equilibrium states and another 5000 cycles to calculate the equilibrium adsorption capacity and selectivity.

The MOFs were modeled using the universal force field.³¹ On the other hand, TraPPE³² force field was used to model ethane and ethylene molecules. The Lennard-Jones (LJ) parameters between the unlike atoms were calculated using Lorentz–Berthelot mixing rules. An LJ potential cutoff distance of 12.0 Å was used for all the simulations. The box length was twice the LJ potential cutoff distance to satisfy the minimum image convention.

In this study, we used the RF regression algorithm³³ to predict the C₂H₆/C₂H₄ selectivity. The scikit-learn library³⁴ was used to utilize the RF regression algorithm. In the RF algorithm, we used 250 trees to make the forest. The permutation feature importance model inspection technique was used to calculate feature importance.³⁵ An additional cross-validation analysis was also performed using the shuffle split cross-validation strategy³⁵ with an n_{split} value of 10. The training set in each split was taken as 10% of the training dataset. In high-throughput screening, we require to perform an optimum number of

simulations to maximize machine learning accuracy as well as to limit the computational cost. Further, larger training set may not improve the machine learning model accuracy significantly. We have tested the machine learning models with a training dataset of 6, 8, and 10% of the h-MOF dataset and compared the model accuracy. As we increase the size of training set, the maximum machine learning model accuracy was achieved when the training data size is 8% of the h-MOF dataset. Further increase in the training data size does not improve model accuracy. Hence, 8% of the h-MOFs was chosen as a training set.

RESULTS AND DISCUSSION

Molecular Screening. The significant parameter to describe the efficiency of separation is the adsorption selectivity, as shown in eq 1.

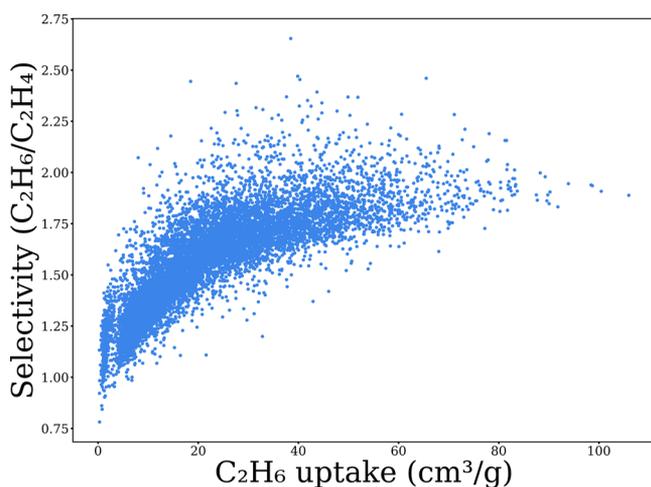


Figure 1. Performance plot of training data (9224 h-MOFs). C_2H_6/C_2H_4 selectivity against volumetric C_2H_6 uptake.

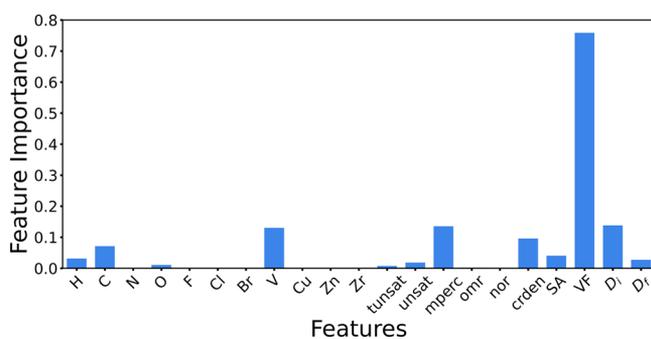


Figure 2. Average feature importance among the decision tree regressors in the RF.

$$S_{C_2H_6/C_2H_4} = \frac{x_{C_2H_6}/x_{C_2H_4}}{y_{C_2H_6}/y_{C_2H_4}} \quad (1)$$

Here, x_i ($i = C_2H_6, C_2H_4$) is the mole fraction in the adsorbed phase and y_i is the mole fraction in the bulk phase (i.e., $y_{C_2H_6}/y_{C_2H_4} = 1.0$ in our work). From the GCMC simulations, we obtained the ethane uptake and ethane selectivity. A good MOF will have high selectivity with high C_2H_6 uptake.

From Figure 1, we can see that a large amount of MOF has good selectivity, but ethane uptake is less. A handful of MOFs have a large volume of ethane uptake. The MOFs with the

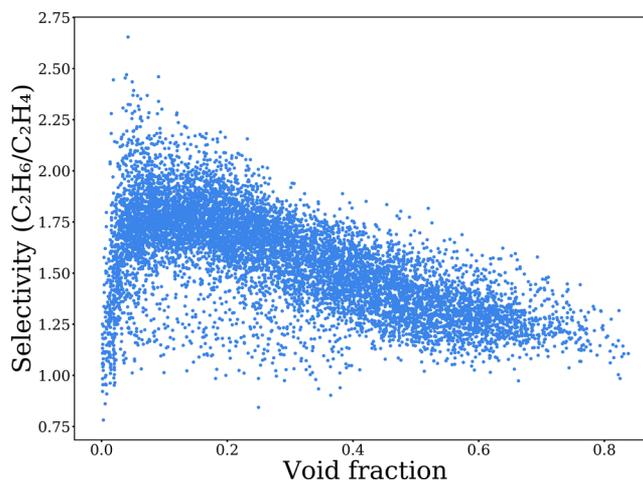


Figure 3. C_2H_6/C_2H_4 selectivity against VF.

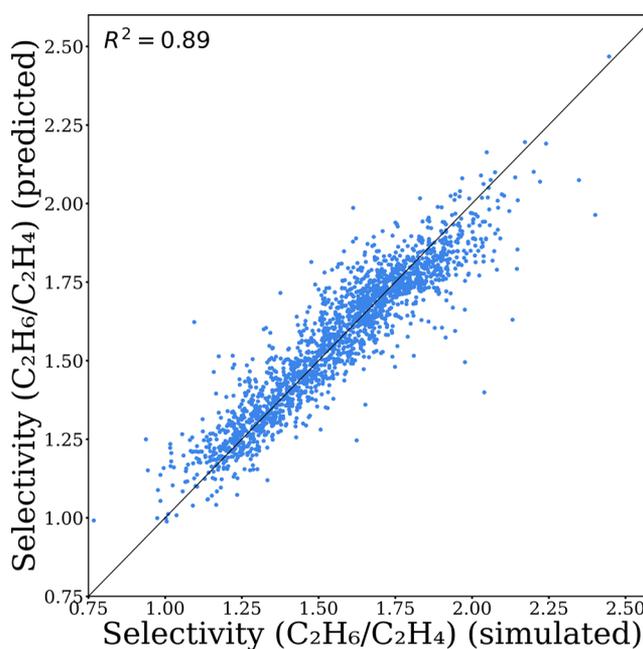


Figure 4. Parity plot of simulated C_2H_6/C_2H_4 selectivity against predicted C_2H_6/C_2H_4 selectivity.

selectivity less than 1.0 are ethylene-selective materials. As our interest is to find the best MOF, which will have the highest C_2H_6 selectivity, we should focus on the right top corner region of Figure 1.

To evaluate the performance of RF model, R^2 values were calculated, which are shown in eq 2.

$$R^2 = 1 - \frac{\sum_{MOF}^{N_{MOF}} (S_{true} - S_{predict})^2}{\sum_{MOF}^{N_{MOF}} (S_{true} - S_{mean})^2} \quad (2)$$

Here, N_{MOF} , S_{true} , $S_{predict}$ and S_{mean} represent the number of MOFs taken in the RF model, simulated selectivity, predicted selectivity, and average selectivity, respectively.

First, the RF model was used where chemical properties are taken as descriptors. In this case, the RF score was low at around 0.36. Hence, the RF model with chemical properties only as descriptors is not suitable to make predictions. Similarly, the RF model was used only with structural properties. However, in this case, the RF score improved significantly to 0.70. Furthermore,

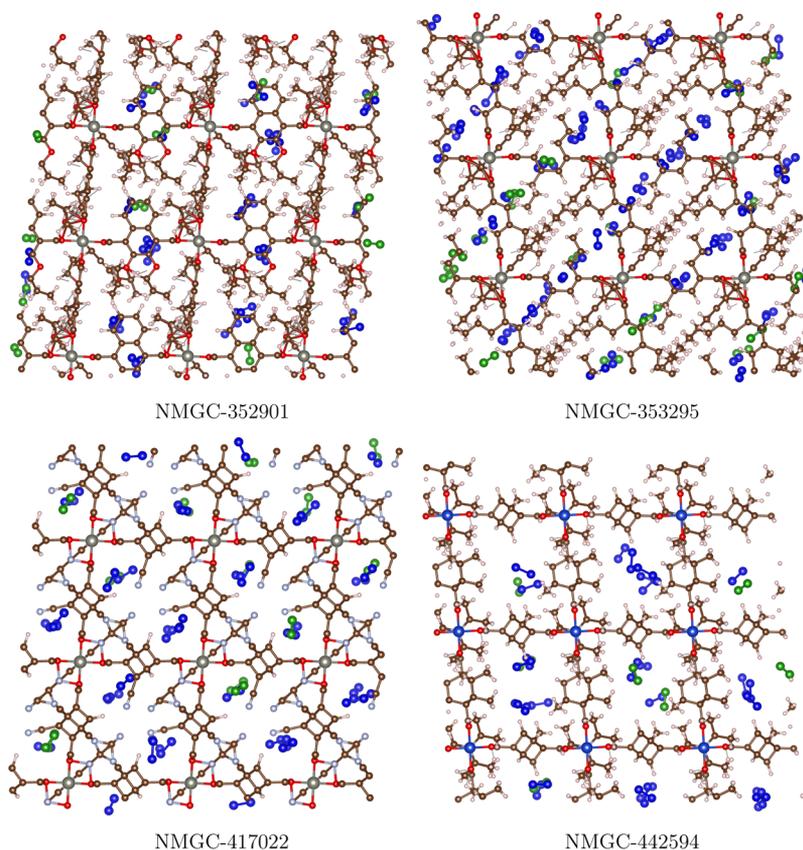


Figure 5. Top-performing h-MOFs at 298 K and 1 bar total pressure. Blue and green atoms represent ethane and ethylene, respectively.

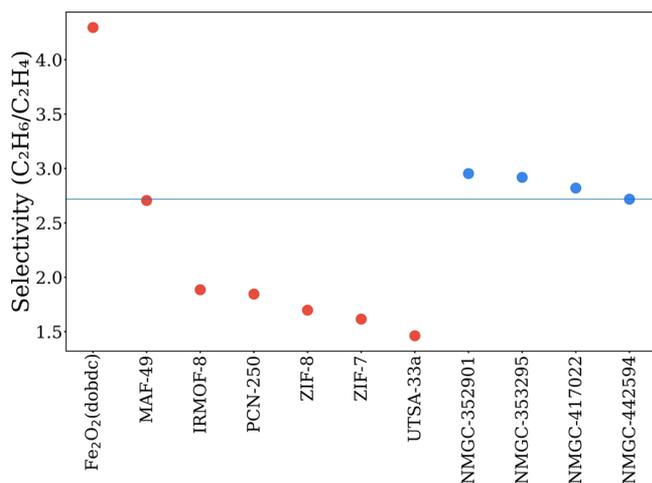


Figure 6. C₂H₆/C₂H₄ selectivity comparison with previously reported best-performing materials C₂H₆/C₂H₄ selectivity at 298 K and 1 bar total pressure. Red and blue colored filled circles represent previously reported best-performing materials and best-performing h-MOFs, respectively.

to improve the RF score, we used the combined structural and chemical properties. The RF score was improved to 0.89 using the shuffle split cross-validation approach. From the machine learning perspective, this RF model can be applied to the rest of the materials.

Also, we calculated the average feature importance among the decision tree regressors in the RF model. The descriptor having higher feature importance is more critical in improving the RF model. From Figure 2, we can clearly come to the conclusion

that void fraction (VF) is by far the most important feature. VF played a larger role in increasing the RF score in the RF model. After VF, the number of vanadium atoms per unit cell, the largest included sphere diameter (D_i), and metallic percentage (m_{perc}) played an important role in prediction as well.

To find out the physical significance of the feature importance and the model performance, we plotted the C₂H₆/C₂H₄ selectivity against the most important feature of this RF model, that is, VF. From Figure 3, we can observe that there is a strong correlation between the VF of MOF materials and C₂H₆/C₂H₄ selectivity. As the VF increases, C₂H₆/C₂H₄ selectivity increases. However, after a certain value of VF, selectivity starts decreasing gradually. Further, we tried to observe the relationship between selectivity and D_i . In this case, the relationship between selectivity and D_i is not strong enough. Similarly, other descriptors have some correlation with the ethane selectivity. Among those descriptors, some of the correlations were inconclusive by visualization means. The relationship of structural and chemical features with C₂H₆/C₂H₄ selectivity is given in the supplementary document.

Further, we tested the performance of the RF model on the test dataset. We randomly chose 2000 MOF materials from the test dataset and simulated them. Simultaneously, the RF model was applied to the randomly chosen 2000 MOF materials. The simulated C₂H₆/C₂H₄ selectivity was plotted against the predicted C₂H₆/C₂H₄ selectivity, which is shown in Figure 4. The 45° diagonal line across the plot signifies the equality between the simulated selectivity and the predicted selectivity. The root-mean-square error is 0.86.

Best Materials. The RF model was used to predict the C₂H₆/C₂H₄ selectivity of the remaining h-MOF materials.

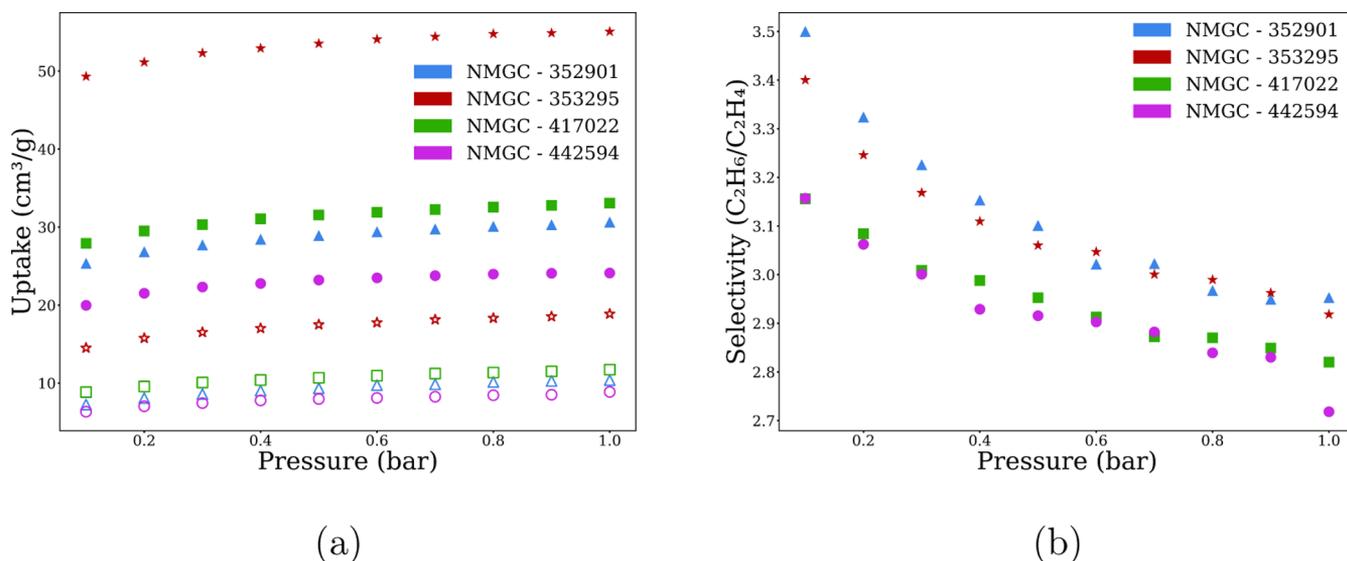
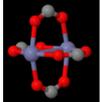
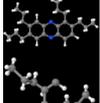


Figure 7. (a) Adsorption isotherms of C₂H₆ (solid) and C₂H₄ (open) of top-performing h-MOFs. (b) Selectivity comparison of the top-performing h-MOFs.

Table 3. Energy Contribution of the SBU with C₂H₆ and C₂H₄ of NMGC-353295 at 298 K and 1 bar Total Pressure

Nomenclature	SBU	$E_{SBU-C_2H_6}$ (kcal/mol)	$E_{SBU-C_2H_4}$ (kcal/mol)
Node		-1.179	-1.04
Linker-1		-5.502	-5.268
Linker-2		-2.217	-2.04

Subsequently, molecular simulations were performed on the predicted best-performing materials. Among those materials, only four h-MOF materials were found to have C₂H₆/C₂H₄ selectivity greater than 2.7, which are shown in Figure 5. In the figures, the blue atoms and the green atoms represent ethane and ethylene molecules, respectively. It is clearly seen that the adsorption of gases occurs near the organic linkers. We compared the selectivity of the best h-MOF materials with existing nanoporous materials at 1 bar pressure and 298 K temperature. The selectivity comparison is shown in Figure 6. It is evident from the figure that only one material discovered by Li et al.³⁶ surpasses the performance of the materials identified in this work.

The top four h-MOF systems were further simulated at different pressures ranging from 0.1 to 1.0 bar to get the adsorption isotherms of C₂H₆ and C₂H₄. From Figure 7a it is clearly seen that, among the top four materials, NMGC-353295 has the best C₂H₆ adsorption uptake, that is, 55.06 cm³/g at 298 K and 1 bar. The C₂H₆/C₂H₄ selectivity is higher at a lower pressure for each h-MOF system. However, with increasing pressure, the C₂H₆/C₂H₄ selectivity decreases for all top-performing h-MOF systems.

Table 3 shows the interaction energy of secondary building units (SBUs) of NMGC-353295 with C₂H₆ and C₂H₄ at 298 K and 1 bar total pressure. The interaction energy suggests that the C₂H₆ atoms are more favorable to achieve higher adsorption. The energy contribution between the node and C₂H₆ and C₂H₄ is the lowest because of the limited space near the metal-containing node. Among all the SBUs, linker-1 has the maximum contribution because of a higher number of binding sites. The energy interaction of SBUs with C₂H₆ and C₂H₄ of the top three MOFs indicates its affinity toward C₂H₆ as well (details are shown in the Supporting Information). Further, we calculated the local density profile of C₂H₆ and C₂H₄ from SBUs, which is shown in Figure 8. In each case, C₂H₆ density is found to be higher than C₂H₄ density. The first peak suggests the primary adsorption layer near the surface of SBUs, which is also in line with the energy interaction data of SBU with C₂H₆ and C₂H₄.

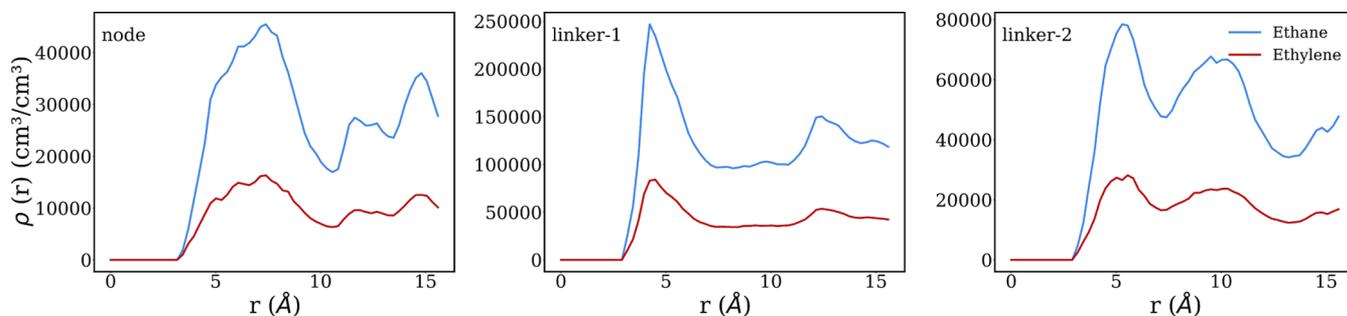


Figure 8. Local density of C₂H₆ and C₂H₄ against the distance from the SBUs of NMGC-353295 at 298 K and 1 bar total pressure.

CONCLUSIONS

In this study, we were able to correlate MOFs' chemical and structural properties with the C₂H₆ selectivity using the RF algorithm. RF was modeled by utilizing the simulated C₂H₆ selectivity values of 8% of the h-MOF database. The RF model with chemical or structural properties showed poor performance in predicting the C₂H₆/C₂H₄ selectivity. However, the R² value improves significantly to 0.89 when both the chemical and structural properties were considered as features. The most important feature to make the model successful was the VF of h-MOFs.

Further, we were able to produce the adsorption isotherms of C₂H₆ and C₂H₄ of the top four h-MOF materials. All MOFs show affinity toward C₂H₆ gas molecules. Among the SBUs, the metallic nodes show the least interaction energy with the C₂H₆ and C₂H₄. The organic linkers show the maximum affinity toward C₂H₆ gas molecules compared to C₂H₄. Hence, the systematic modification of organic linkers should be the next objective to improve further the C₂H₆/C₂H₄ selectivity.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.energyfuels.0c03063>.

Trained RF model for predicting C₂H₆/C₂H₄ selectivity (<https://github.com/gdev7/h-mof-ethane-ethylene-separation>); relationship between C₂H₆/C₂H₄ selectivity with chemical and structural descriptors of MOFs; energy contribution of SBUs with C₂H₆ and C₂H₄ of rest of the top four h-MOFs; and local density profiles of C₂H₆ and C₂H₄ against the distance from the SBUs of rest of the top four h-MOFs (PDF)

AUTHOR INFORMATION

Corresponding Author

Jayant K. Singh – Department of Chemical Engineering, Indian Institute of Technology Kanpur, Kanpur 208016, India; Prescience Insilico Private Limited, Bangalore 560049, India; orcid.org/0000-0001-8056-2115; Phone: +91 512 259 6141; Email: jayantks@iitk.ac.in; Fax: +91 512 259 0104

Author

Prosun Halder – Department of Chemical Engineering, Indian Institute of Technology Kanpur, Kanpur 208016, India; orcid.org/0000-0002-3988-2712

Complete contact information is available at: <https://pubs.acs.org/doi/10.1021/acs.energyfuels.0c03063>

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by the Science and Engineering Research Board, India, under the proposal MTR/2019/000194. The authors also thank the High Performance Computing Center of Indian Institute of Technology Kanpur for providing computational resources.

REFERENCES

(1) Lee, C. Y.; Bae, Y.-S.; Jeong, N. C.; Farha, O. K.; Sarjeant, A. A.; Stern, C. L.; Nickias, P.; Snurr, R. Q.; Hupp, J. T.; Nguyen, S. T. *J. Am. Chem. Soc.* **2011**, *133*, 5228–5231.

- (2) Li, J.-R.; Kuppler, R. J.; Zhou, H.-C. *Chem. Soc. Rev.* **2009**, *38*, 1477–1504.
- (3) Morris, R. E.; Wheatley, P. S. *Angew. Chem., Int. Ed.* **2008**, *47*, 4966–4981.
- (4) Kreno, L. E.; Leong, K.; Farha, O. K.; Allendorf, M.; Van Duyne, R. P.; Hupp, J. T. *Chem. Rev.* **2012**, *112*, 1105–1125.
- (5) Lee, J.; Farha, O. K.; Roberts, J.; Scheidt, K. A.; Nguyen, S. T.; Hupp, J. T. *Chem. Soc. Rev.* **2009**, *38*, 1450–1459.
- (6) Horcajada, P.; Serre, C.; Vallet-Regi, M.; Sebban, M.; Taulelle, F.; Férey, G. *Angew. Chem., Int. Ed.* **2006**, *45*, 5974–5978.
- (7) Halder, P.; Maurya, M.; Jain, S. K.; Singh, J. K. *Phys. Chem. Chem. Phys.* **2016**, *18*, 14007–14016.
- (8) Yang, R. T.; Kikkides, E. S. *AIChE J.* **1995**, *41*, 509–517.
- (9) Bereciartua, P. J.; et al. *Science* **2017**, *358*, 1068–1071.
- (10) Bloch, E. D.; Queen, W. L.; Krishna, R.; Zdrozny, J. M.; Brown, C. M.; Long, J. R. *Science* **2012**, *335*, 1606–1610.
- (11) Sumida, K.; Rogow, D. L.; Mason, J. A.; McDonald, T. M.; Bloch, E. D.; Herm, Z. R.; Bae, T.-H.; Long, J. R. *Chem. Rev.* **2012**, *112*, 724–781.
- (12) Cote, A. P.; Benin, A. I.; Ockwig, N. W.; O'Keeffe, M.; Matzger, A. J.; Yaghi, O. M. *Science* **2005**, *310*, 1166–1170.
- (13) Sircar, S.; Golden, T. C.; Rao, M. B. *Carbon* **1996**, *34*, 1–12.
- (14) Deem, M. W.; Pophale, R.; Cheeseman, P. A.; Earl, D. J. *J. Phys. Chem. C* **2009**, *113*, 21353–21360.
- (15) Pophale, R.; Cheeseman, P. A.; Deem, M. W. *Phys. Chem. Chem. Phys.* **2011**, *13*, 12407–12412.
- (16) Park, K. S.; Ni, Z.; Cote, A. P.; Choi, J. Y.; Huang, R.; Uribe-Romo, F. J.; Chae, H. K.; O'Keeffe, M.; Yaghi, O. M. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 10186–10191.
- (17) Lu, W.; Yuan, D.; Zhao, D.; Schilling, C. I.; Plietzsch, O.; Muller, T.; Bräse, S.; Guenther, J.; Blümel, J.; Krishna, R.; Li, Z.; Zhou, H.-C. *Chem. Mater.* **2010**, *22*, 5964–5972.
- (18) Bácia, P. S.; Bastin, L.; Hurtado, E. J.; Silva, J. A. C.; Rodrigues, A. E.; Chen, B. *Sep. Sci. Technol.* **2008**, *43*, 3494–3521.
- (19) Yan, Q.; Lin, Y.; Kong, C.; Chen, L. *Chem. Commun.* **2013**, *49*, 6873–6875.
- (20) Kochetygov, I.; Bulut, S.; Asgari, M.; Queen, W. L. *Dalton Trans.* **2018**, *47*, 10527–10535.
- (21) Chen, Y.; Qiao, Z.; Wu, H.; Lv, D.; Shi, R.; Xia, Q.; Zhou, J.; Li, Z. *Chem. Eng. Sci.* **2018**, *175*, 110–117.
- (22) Pires, J.; Pinto, M. L.; Saini, V. K. *ACS Appl. Mater. Interfaces* **2014**, *6*, 12093–12099.
- (23) Earl, D. J.; Deem, M. W. *Ind. Eng. Chem. Res.* **2006**, *45*, 5449–5454.
- (24) Wilmer, C. E.; Leaf, M.; Lee, C. Y.; Farha, O. K.; Hauser, B. G.; Hupp, J. T.; Snurr, R. Q. *Nat. Chem.* **2011**, *4*, 83–89.
- (25) Simon, C. M.; Mercado, R.; Schnell, S. K.; Smit, B.; Haranczyk, M. *Chem. Mater.* **2015**, *27*, 4459–4475.
- (26) Thornton, A. W.; Simon, C. M.; Kim, J.; Kwon, O.; Deeg, K. S.; Konstas, K.; Pas, S. J.; Hill, M. R.; Winkler, D. A.; Haranczyk, M.; Smit, B. *Chem. Mater.* **2017**, *29*, 2844–2854.
- (27) Zhang, Y.; Yang, S.; Jiao, Y.; Liu, H.; Yuan, H.; Lu, S.; Ran, T.; Yao, S.; Ke, Z.; Xu, J.; Xiong, X.; Chen, Y.; Lu, T. *J. Chem. Inf. Model.* **2013**, *53*, 3163–3177.
- (28) Canepa, P.; Arter, C. A.; Conwill, E. M.; Johnson, D. H.; Shoemaker, B. A.; Soliman, K. Z.; Thonhauser, T. *J. Mater. Chem. A* **2013**, *1*, 13597–13604.
- (29) Thornton, A. W.; Dubbeldam, D.; Liu, M. S.; Ladewig, B. P.; Hill, A. J.; Hill, M. R. *Energy Environ. Sci.* **2012**, *5*, 7637–7646.
- (30) Willems, T. F.; Rycroft, C. H.; Kazi, M.; Meza, J. C.; Haranczyk, M. *Microporous Mesoporous Mater.* **2012**, *149*, 134–141.
- (31) Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. *J. Am. Chem. Soc.* **1992**, *114*, 10024–10035.
- (32) Martin, M. G.; Siepmann, J. I. *J. Phys. Chem. B* **1998**, *102*, 2569–2577.
- (33) Breiman, L. *Mach. Learn.* **2001**, *45*, 5–32.
- (34) Pedregosa, F.; et al. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
- (35) Larose, D. T.; Larose, C. D. *Data Mining and Predictive Analytics*; John Wiley & Sons, 2015.

(36) Li, L.; Lin, R.-B.; Krishna, R.; Li, H.; Xiang, S.; Wu, H.; Li, J.; Zhou, W.; Chen, B. *Science* **2018**, *362*, 443–446.