Combining Automated Machine Learning and Molecular Simulation to Advance the Discovery of COF-based Membranes for Acid Gas Separation

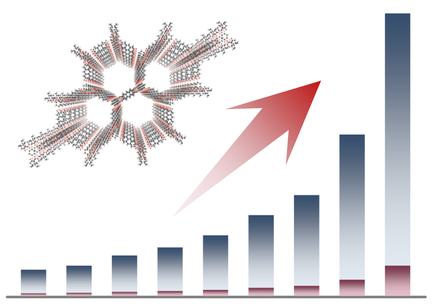
Bingru Xin[1], Minggao Feng[2], Min Cheng[1], Zhongde Dai[3], Li Zhou[1], Yiyang Dai[1]*, Xu Ji[1]

[1] School of Chemical Engineering, Sichuan University, China

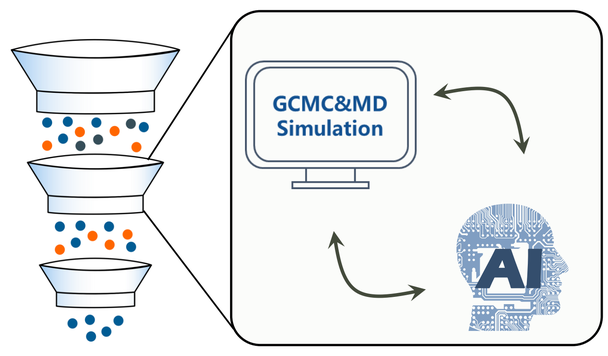
[2] Department of Chemistry and Materials Innovation Factory, University of Liverpool, UK

[3] College of Carbon Neutrality Future Technology, Sichuan University, China

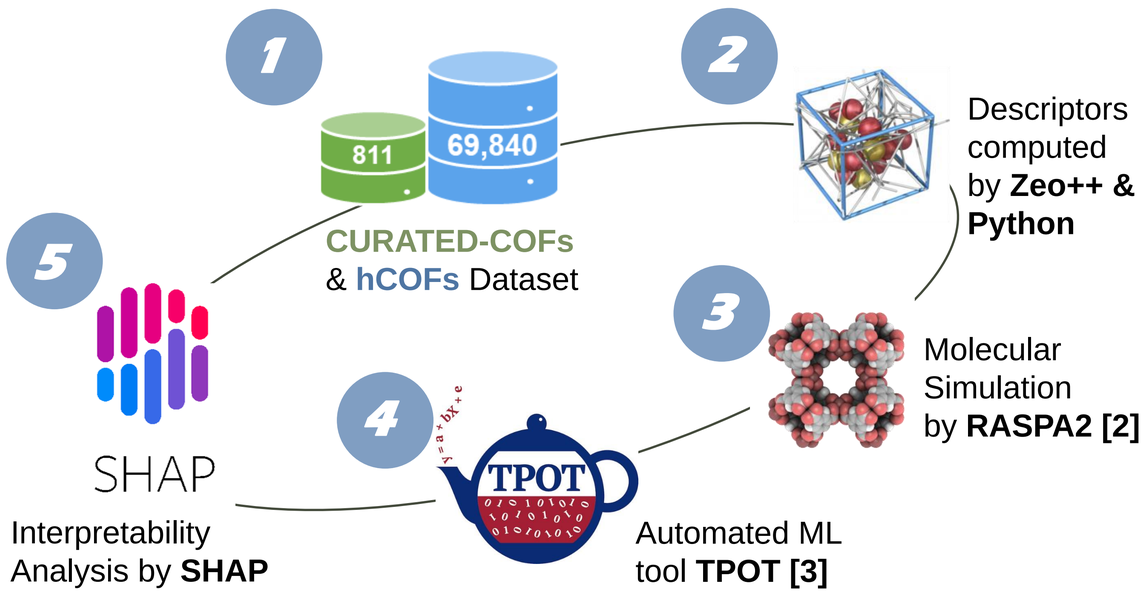
Introduction



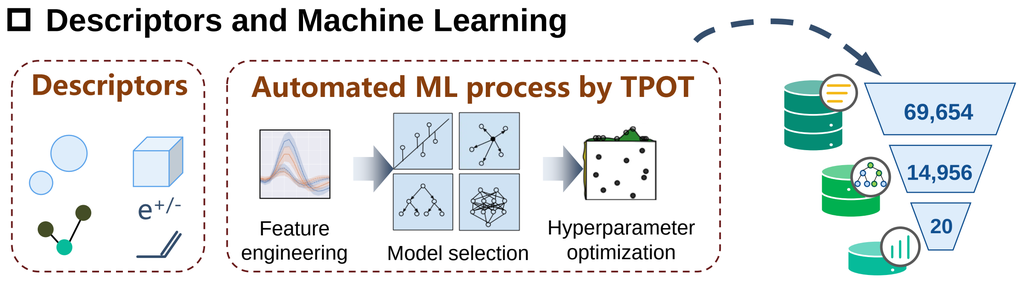
- **Membrane separation [1] ,** which has the advantages of small footprint, low energy consumption, and so on, is a promising technology in gas separation.
- **Covalent organic frameworks (COFs)** are an emerging class of crystal network porous membrane materials for gas separation.
- **Growing in number:** At present, there are over 10^3 COFs experimentally synthesized (e.g. CURATED-COFs) and over 10^5 computer-synthesized COFs (e.g. hCOFs). A method is urgently needed to accelerate material evaluation and discovery.



Methods



Descriptors and Machine Learning



Molecular Simulation and Membrane Calculation

$$P_i = \frac{[\phi \times N_i \times D_i \times \rho_i]}{f_i}$$

$$S_{i/j} = P_i / P_j$$

i & j: *gas species (H_2S, CO_2 or CH_4);*
ϕ: *void Fraction of COF;*
N_i: *gas uptake(mol/kg) calculated by **GCMC;***
D_i: *gas self-diffusivity (m^2/s) calculated by **MD;***
f_i: *partial fugacity of gas (Pa).*
ρ_i: *density of COF (g/cm^3)*

Results & Discussion

1. The automated machine learning process TPOT builds models with high prediction accuracy

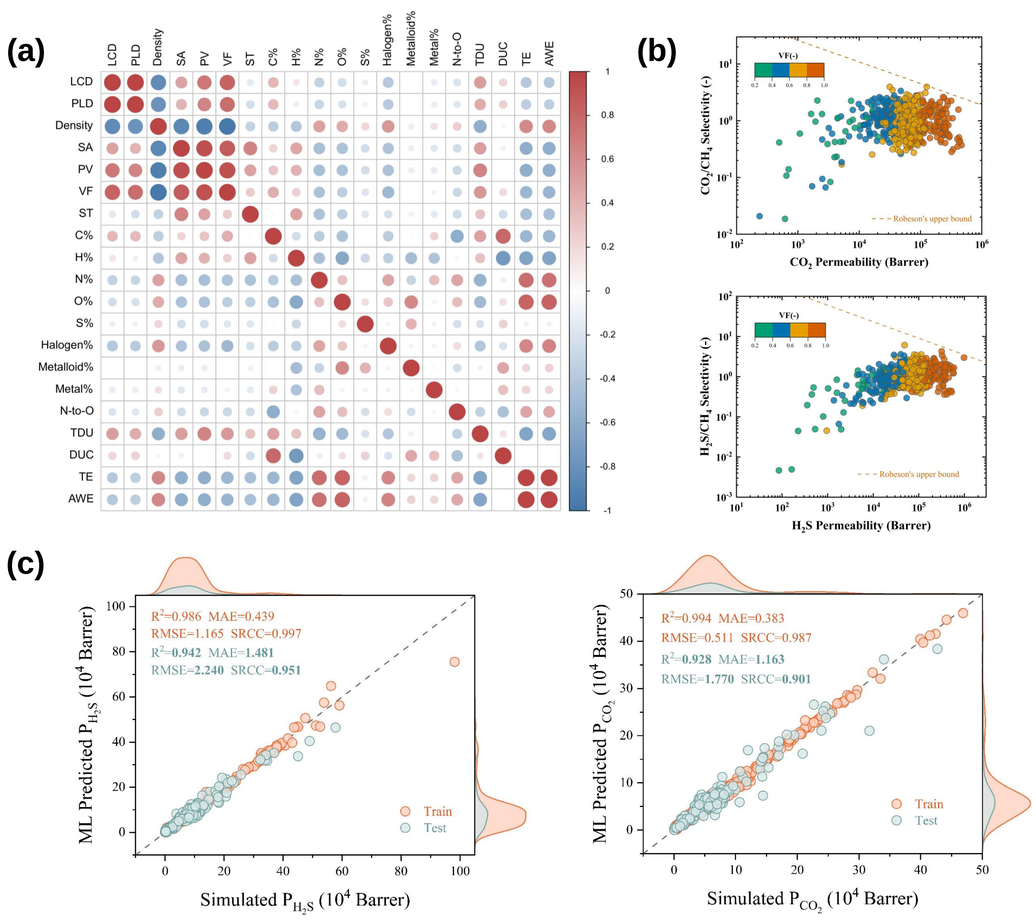


Fig. 1. The heatmap of feature correlation(a) and the gas permeability of the COF-based membranes at 10 bar, 298 K. The VF indicates void fraction(b). Marginal distribution scatter plots of gas permeability data predicted by the ML model compared with MS results.(c)

3. Rapid Screening 69,654 hCOFs

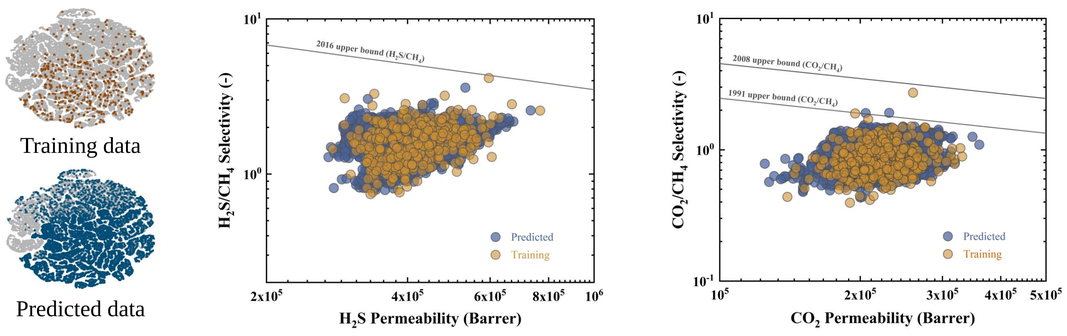


Fig. 3. H_2S/CH_4 & CO_2/CH_4 permeability-selectivity distributions obtained from hCOFs after ML projection.

2. Interpretability analysis: void fraction (VF) is all we need

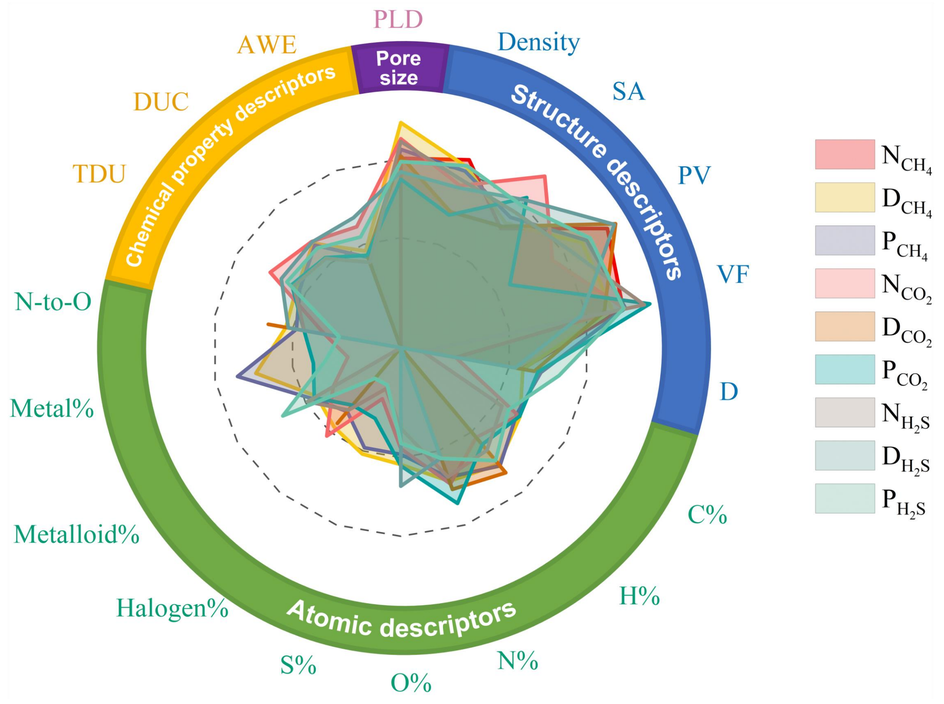


Fig. 2. Radar plots of feature importance for the gas adsorption, diffusion, and permeability of COFs. The SHAP values for each type of explainer have been normalized and the scale range shows a Log distribution from 0.01 to 1.

Conclusions

- **A New Workflow:** Combining ML and MS to predict the separation performance of 70,000+ COF-based membranes for H_2S/CH_4 & CO_2/CH_4.
- **High Accuracy:** Most ML models predict that R^2 is greater than 0.9.
- **Interpretability ML:** Structure descriptors (especially void fraction) contributed the most to the membrane gas permeability.
- **TOP High-Performance COFs:** large pore sizes and high void fraction.
- **Future work:** Consider using Active Learning (AL) and Deep Learning (DL) methods for material simulation and discovery.

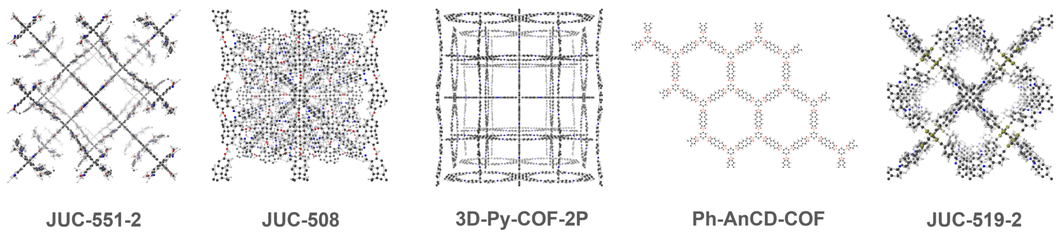


Fig. 4. Structural visualization of top 5 COFs in CURATED-COFs. All snapshots are taken looking down the crystallographic c-axis of each structure.

References

[1] Yuan, Shushan, et al. "Covalent organic frameworks for membrane separation." Chemical Society Reviews 48.10 (2019): 2665-2681.

[2] Dubbeldam, David, et al. "RASPA: molecular simulation software for adsorption and diffusion in flexible nanoporous materials." Molecular Simulation 42.2 (2016): 81-101.

[3] Olson, Randal S., and Jason H. Moore. "TPOT: A tree-based pipeline optimization tool for automating machine learning." Workshop on automatic machine learning. PMLR, 2016.

Acknowledgements

The financial support provided by the Young Scientists Fund of the National Natural Science Foundation of China (No. 52170112) and the National Key Research and Development Program of China (No. 2023YFF0720102)

About the author

Master Bingru Xin's research interests include AI for material and process in chemical engineering, you can view my research works in ***https://xinbingru.github.io/***