

耦合机器学习与高通量计算研究疏水 MOFs 在 CO₂/C₂H₂ 膜分离中的构效关系

韩荣美, 韩 琪, 张政清*, 孙玉绣*, 乔志华

(天津工业大学 先进分离膜材料全国重点实验室 化学工程与技术学院, 天津 300387)

摘要: 二氧化碳/乙炔(CO₂/C₂H₂)分离在工业过程中的重要性推动了高效膜材料的开发,但膜材料在潮湿环境下的分离性能不稳定,因此开发能够在潮湿条件下保持优异性能的分膜成为关键。本研究首先以 ZIF-8 的亨利系数为标准,建立了疏水性 MOF 数据库。通过巨正则蒙特卡罗(grand canonical Monte Carlo, GCMC)模拟和分子动力学(MD)模拟分别计算了用于 CO₂/C₂H₂ 分离的 MOF 的热力学性质和扩散性质,筛选出 Top 20 个高性能疏水 MOF 膜。随后,基于 MOFs 的特征描述符和模拟结果,训练和测试了 5 种不同的机器学习(ML)模型,并采用最优的 ML 模型揭示了影响 CO₂/C₂H₂ 分离性能的关键因素。最后,通过建立构效关系分析,确定了 CO₂/C₂H₂ 分离的最佳 MOF 结构参数范围,为 MOF 膜材料的开发与优化提供了理论依据。

关键词: 金属-有机骨架; 分子模拟; 机器学习; CO₂/C₂H₂ 分离

中图分类号: TQ051.8 **文献标志码:** A **文章编号:** 1007-8924(2025)03-0136-09

doi: 10.16159/j.cnki.issn1007-8924.2025.03.014

乙炔(C₂H₂)是一种重要的化工原料,用于生产塑料、橡胶、合成纤维等,但工业乙炔中通常含有一定量的二氧化碳(CO₂)^[1-2]。为了确保下游工艺的效率 and 产品的纯度,高效分离 CO₂ 是乙炔纯化过程中必不可少的环节。然而,由于 CO₂ 和 C₂H₂ 的分子尺寸(CO₂: 0.33 nm × 0.33 nm × 0.57 nm, C₂H₂: 0.32 nm × 0.33 nm × 0.54 nm)和物理化学性质(如极性)非常接近^[3],这使得 CO₂/C₂H₂ 分离成为工业规模生产中具有挑战性的化学分离之一。与传统的分离技术(如深冷分离和溶剂吸收)相比,膜分离技术因其高效性、低能耗和经济性,在实现绿

色化工和环境友好型生产中具有重要意义。然而,实现膜分离技术的关键在于开发具有高渗透性和高选择性的新型膜材料^[4-6]。

金属-有机框架(MOFs)作为一种新型功能化纳米多孔材料,因其具有规则的化学结构、高孔隙率、高比表面积以及可调的孔结构和化学环境等诸多特点,在气体储存和气体膜分离领域展现出了广泛的应用前景^[7-8]。例如,Zhang 等^[9]通过晶种诱导界面生长策略构建了无缺陷的 MOF-802 膜,这种膜对 H₂/CO₂ 具有良好的分离性能,H₂ 的渗透率约为 3540 GPU,选择性约为 23.1,并且发现 MOF-

收稿日期: 2024-12-31; 修改稿收到日期: 2025-02-16

基金项目: 国家自然科学基金项目(22108202)

第一作者简介: 韩荣美(2000-),女,山东德州人,硕士研究生,研究方向为 MOF 膜材料筛选与设计。* 通讯作者,张政清, E-mail: zhangzhengqing@tiangong.edu.cn; 孙玉绣, E-mail: yuxiusun@tiangong.edu.cn

引用本文: 韩荣美,韩 琪,张政清,等. 耦合机器学习与高通量计算研究疏水 MOFs 在 CO₂/C₂H₂ 膜分离中的构效关系[J]. 膜科学与技术,2025,45(3):136-144.

Citation: Han R M, Han Q, Zhang Z Q, *et al.* Coupling machine learning with high-throughput computing study the structure-performance correlation of hydrophobic MOFs for membrane-based CO₂/C₂H₂ separation[J]. Membrane Science and Technology(Chinese), 2025, 45(3): 136-144.

802 膜在水蒸气存在下也表现出很高的分离性能。Zhou 等^[10]采用快速电流驱动合成的方法,在多孔阳极氧化铝载体上制备了 ZIF-8 膜,对丙烯/丙烷具有较好的分离能力,在 0.1 MPa 和室温下,膜的 C₃H₆ 渗透率为 52.0 GPU, C₃H₆/C₃H₈ 的选择性为 304.8。Liu 等^[11]制备了蜂窝状霍夫曼型 ZJU-74a 膜,可用于分离 C₂H₂ 与 CO₂ 混合气体,膜的 C₂H₂/CO₂ 分离选择性约为 2.58,而且膜的耐用性(实际应用中的关键参数)也非常出色。然而,当前绝大多数 MOF 材料对水或水蒸气敏感^[12],易引起晶体骨架变形甚至坍塌,导致其降低或失去孔道结构和高比表面积,进而影响 MOF 材料的各种性能及其实际应用^[12-13]。因此,疏水性 MOF 的筛选研究在推动 MOF 的实际应用中具有重要的现实意义。

随着金属-有机框架实验合成和修饰技术的不断进步^[14-16],理论上,借助这些手段,研究人员有望设计并合成出种类繁多的 MOF 材料。迄今为止,已有超过 10 万种 MOFs 通过实验合成^[17],而且通过化学理论和计算机模拟技术预测的潜在可合成 MOFs 数量更是庞大。例如, G3mez-Gualdr3n 等^[18]通过“自上而下”的方法构建了包含 1.4 万种 MOFs 的数据库; Snurr 课题组^[19]和 Li 课题组^[20]则采用“自下而上”的方法,利用高通量计算成功构建了包含 13 万余种和 1.1 万种 MOFs。如此庞大的 MOFs 数量无疑为其在气体分离膜领域的应用提供了巨大潜力。然而,仅依赖实验方法来筛选和制备高性能的特定气体分离 MOFs,犹如大海捞针,

这在一定程度上限制了研发效率。因此,针对特定应用场景寻找合适的 MOFs 成为一项极具挑战性的任务。在此情形下,高通量计算筛选成为从海量材料中筛选出高性能目标 MOFs 最为有效的研究方法。进一步通过与机器学习相结合,不仅可以高效预测数以万计 MOF 材料的分离性能,减少实验成本和时间,还能够深入挖掘其微观结构与宏观性能之间的复杂关联,为气体分离材料的理性设计提供有力的理论支持,从而加速材料的筛选和开发^[21-22]。

本文提出了一种基于机器学习辅助的高通量 MOF 材料筛选策略,用于研究气体膜分离构效关系。首先建立疏水性 MOF 数据库,然后利用巨正则蒙特卡罗(GCMC)模拟和分子动力学(MD)模拟分别计算出用于 CO₂/C₂H₂ 分离的 MOF 膜分离性能,并筛选出性能优异的 MOF 膜材料。使用 MOFs 的特征描述符和模拟结果对 5 种不同的机器学习算法进行训练和测试,选择最佳的预测模型,揭示了影响 CO₂/C₂H₂ 膜分离性能的关键因素并获得用于 CO₂/C₂H₂ 分离的最佳参数范围,为 MOF 膜材料的高效开发与优化提供了重要的理论和实际意义。

1 方法

本研究提出了一种耦合机器学习与分子模拟的方法,用于研究 MOFs 膜在 CO₂/C₂H₂ 分离中的构效关系。工作流程如图 1 所示。

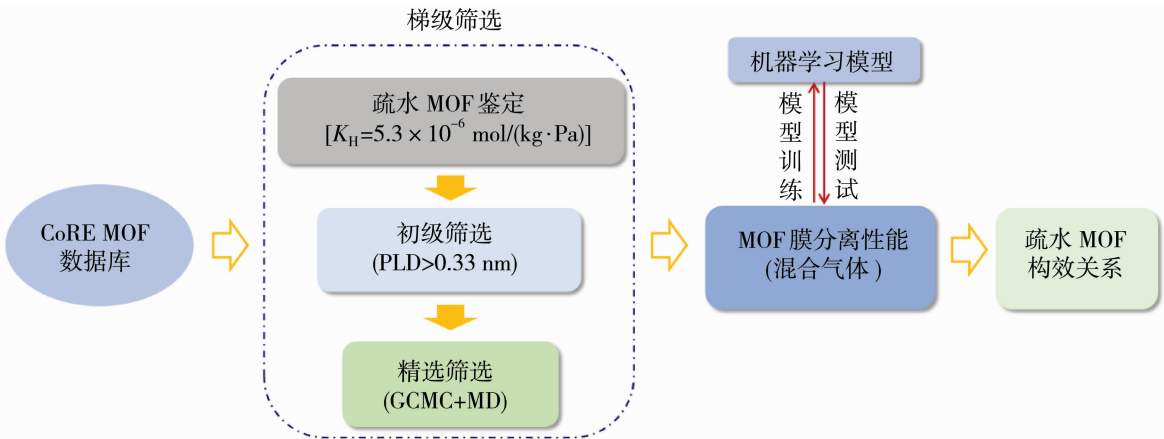


图 1 用于 CO₂/C₂H₂ 分离的 MOFs 膜研究工作流程图

Fig. 1 The workflow illustration of MOFs for membrane-based CO₂/C₂H₂ separation

1.1 疏水 MOFs 数据库预筛选

在执行分子模拟计算之前,需要对 CoRE -

MOF 数据库中的 10 143 个有序且无溶剂的 MOFs 进行预筛选处理^[23]。由于大多数 MOFs 在潮湿环

境中不稳定,为了减小水汽对 MOF 分离性能的影响,首先通过水在 MOFs 中的亨利常数(K_H)来判断 MOFs 的亲疏水性。由于 ZIF-8 是疏水性 MOF,因此利用 Widom 插入法计算水分子在 ZIF-8 中 K_H 值,并作为鉴定 MOF 亲疏水性的标准。如果 MOF 的 K_H 值低于 ZIF-8 的 K_H 值 [$K_H=5.3\times10^{-6}$ mol/(kg·Pa)],则将其归类为疏水性 MOF。计算的 K_H 值与先前报道的计算结果一致 [$K_H=5.0\times10^{-6}$ mol/(kg·Pa)]^[24] 验证了计算的准确性。为了减少计算资源和时间,同时保证气体分子能够通过 MOF 孔隙进行膜分离,进一步过滤掉孔径过小的 MOFs,以 MOF 的孔限直径 (PLD) 大于气体分子的最大动力学尺寸 (CO_2 为 0.33 nm, C_2H_2 为 0.33 nm)^[25] 和质量比表面积 (gASA) 大于 0 m²/g 为标准,最终确定了包含 1 733 个疏水性 MOF 的数据库,用于研究中的分子模拟和机器学习部分。对于机器学习中使用的物理特征,包括 PLD,最大空腔直径 (LCD)、LCD/PLD,密度(ρ)、孔隙度(φ)、可及孔体积 (AV) 和质量比表面积 (gASA) 都是通过执行 Zeo++ 软件包^[26] 计算得到的。对于 gASA,探针分子的球半径设置为 N_2 分子动力学直径的一半 (0.184 nm)。

1.2 分子模拟与性能指标计算

巨正则系综蒙特卡罗 (GCMC) 模拟已经被广泛用于计算多孔材料中气体的吸附性质。因此,在 298.0 K 和 0.1 MPa 下进行 GCMC 模拟,以获得 $\text{CO}_2/\text{C}_2\text{H}_2$ 二元混合物在 1 733 种疏水 MOFs 中的热力学性质。混合气体中 CO_2 和 C_2H_2 的摩尔比设

置为 50 : 50,以模拟石化行业中生产的 CO_2 和 C_2H_2 的典型混合比例。在 GCMC 模拟过程中,MOF 采用刚性骨架模型,Lennard-Jones (LJ) 势参数采用 UFF 模型中的参数^[27],MOF 中各原子的电荷分配采用 EEq 理论^[28]。对于 CO_2 分子采用 EPM2 模型^[29],其中 C—O 键长为 0.116 nm,键角为 180°。 C_2H_2 则采用全原子模型^[30],气体分子力场参数见表 1。每个 MOF 都进行了单独的 GCMC 模拟,每次模拟包括: 1×10^5 次循环用于系统平衡, 2×10^6 次循环用于热力学性质计算。在 GCMC 模拟过程中,考虑了多种分子运动类型,包括分子的平移 (0.15%)、旋转 (0.15%)、插入 (0.15%)、交换 (0.5%) 和删除 (0.05%),以模拟气体分子的吸附过程。最终的 GCMC 模拟结果体现了热力学平衡构型,作为分子动力学 (MD) 模拟的初始构型,用于进一步计算 CO_2 和 C_2H_2 气体分子在 MOF 中的扩散性质。在 MD 模拟中,使用 Nosé-Hoover 恒温器,在 NVT 系综下进行模拟。分子运动轨迹采用 Velocity Verlet 算法进行积分,时间步长为 1.0 fs。对每个 MOF 材料运行 5.0 ns 的 MD 模拟,其中最后 3.0 ns 用于计算 CO_2 和 C_2H_2 分子的扩散性质。

所有分子模拟 (GCMC 和 MD) 均使用 HT-CADSS 代码进行^[31]。对于气体分子间的相互作用,采用 Lorentz-Berthelot 混合规则计算不同原子之间的交叉相互作用。长程静电相互作用,LJ 势和库仑相互作用采用 Ewald 求和方法处理,LJ 势和库仑相互作用的截断半径为 1.2 nm。周期性边界条件 (PBC) 应用于 x,y,z 3 个方向。

表 1 气体分子的力场参数

Table 1 Force field parameters of gas molecules

分子	$\epsilon/k_b/K$	σ/nm	电荷/e	键长/nm
CO_2_C	28.129 0	0.275 7	0.651 2	0.116 0 (C—O)
CO_2_O	80.507 0	0.303 3	−0.325 6	—
C_2H_2_C	57.876 0	0.380 0	−0.278 0	0.121 1 (C—C)
C_2H_2_H	0.000 0	0.000 0	0.278 0	0.107 1 (C—H)

为了评价 MOFs 的膜分离性能,研究中计算了 3 个常用的膜性能评价指标^[17, 32-33]: 气体渗透性 (permeability, P_i)、膜选择性 (membrane selectivity, $S_{\text{mem}}^{\text{mix}}$) 以及气体渗透性和膜选择性之间的权衡 (membrane performance score, MPS)。此外,还引入吸附选择性 (adsorption selectivity, $S_{\text{ads}}^{\text{mix}}$) 和扩散选择性 (diffusion selectivity, $S_{\text{diff}}^{\text{mix}}$) 计算评价标准的公式,主要因为膜的选择性与其吸附和扩散性能密

切相关。各项性能的计算公式如式 (1)~(5):

$$P_i=\frac{D_{s,i}^{\text{mix}}\times C_i}{f_i} \tag{1}$$

$$S_{\text{mem},\text{CO}_2/\text{C}_2\text{H}_2}^{\text{mix}}=S_{\text{ads},\text{CO}_2/\text{C}_2\text{H}_2}^{\text{mix}}\times S_{\text{diff},\text{CO}_2/\text{C}_2\text{H}_2}^{\text{mix}} \tag{2}$$

$$\text{MPS}=S_{\text{mem},\text{CO}_2/\text{C}_2\text{H}_2}^{\text{mix}}\times P_i \tag{3}$$

$$S_{\text{ads},\text{CO}_2/\text{C}_2\text{H}_2}^{\text{mix}}=\frac{N_{\text{ads},\text{CO}_2}/N_{\text{ads},\text{C}_2\text{H}_2}}{y_{\text{CO}_2}/y_{\text{C}_2\text{H}_2}} \tag{4}$$

$$S_{\text{diff}, \text{CO}_2/\text{C}_2\text{H}_2}^{\text{mix}} = \frac{D_{s, \text{CO}_2}^{\text{mix}}}{D_{s, \text{C}_2\text{H}_2}^{\text{mix}}} \quad (5)$$

式中: $N_{\text{ads}, i}$ 表示吸附条件下的气体吸收量 (cm^3/g); y_i 表示气体混合物的体积成分; D_s 表示混合气中气体的自扩散系数 (m^2/s), 通过 Einstein 方程计算得到^[17, 21]; C_i 表示吸附气体的摩尔浓度 (mol/m^3); f_i 表示混合物中气体种类的分压 (Pa); i 表示气体种类, CO_2 或 C_2H_2 。

1.3 机器学习

收集 MOF 的物理和化学特征参数以及分子模拟结果作为机器学习 (ML) 的输入文件。为了避免机器学习模型因强相关性导致的多重共线性问题, 在训练模型之前, 需要计算每个描述符之间的 Pearson 相关系数 r , 见方程 (6)。图 2 中显示的是由 6 个物理特征 (PLD、LCD/PLD、 ρ 、 φ 、AV 和 gASA) 和 13 个化学特征 {碳原子分数 [$n(\text{Carbon})\%$]、氢原子分数 [$n(\text{Hydrogen})\%$]、氧原子分数 [$n(\text{Oxygen})\%$]、氮原子分数 [$n(\text{Nitrogen})\%$]、金属原子分数 [$n(\text{Metal})\%$]、卤素原子分数 [$n(\text{Halogen})\%$]、金属离子(簇)种类 (MT)、有机连接体种类 (LT)、氧-金属比 (omr)、不饱和度 (unsat)、总不饱和度 (tunsat)、金属百分比 (mperc), 和拓扑结构类型 (TT)} 描述符组成的皮尔逊相关矩阵^[21, 34]。并未发现相关系数大于 0.9 的特征^[22], 因此选择所有的物理特征和化学特征作为输入数据来训练和测试 ML 模型。为了寻找具有较高预测精度的 ML 算法, 采用 5 种不同的 ML 模型进行训练和预测, 其模型包括随机森林 (RF)、极限梯度提升 (XGBoost)、轻量的梯度提升机 (LightGBM)、CatBoost 和直方图梯度提升 (Hist-GradientBoosting)。在机器学习过程中, 随机选择了 90% 的 MOFs 数据集作为训练集, 10% 的数据作为测试集。为了避免过拟合, 采用了 5 折交叉验证, 所有模型均使用 Python 编写的 scikit-learn 模块进行实现。最终, 基于较高精度的 ML 模型揭示了影响气体分离性能的关键因素。为了评价 ML 模型预测性, 使用决定系数 (R^2)、平均绝对误差 (MAE) 和均方根误差 (RMSE) 这 3 个评价指标。计算公式如式 (6)~(9):

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}} \quad (6)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (X_i - \hat{X}_i)^2}{\sum_{i=1}^n (X_i - \bar{X})^2} \quad (7)$$

$$\text{MAE} = \frac{\sum_{i=1}^n |X_i - \hat{X}_i|}{n} \quad (8)$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (X_i - \hat{X}_i)^2}{n}} \quad (9)$$

式中: X_i 和 Y_i 表示模拟值; $\hat{X}$ 表示预测值; $\bar{X}$ 和 $\bar{Y}$ 分别表示 X_i 和 Y_i 的平均值; n 表示 MOF 膜的总个数; i 表示第 i 个 MOF 膜。 R^2 越高, MAE 值和 RMSE 值越低, 代表 ML 模型预测性能越好。

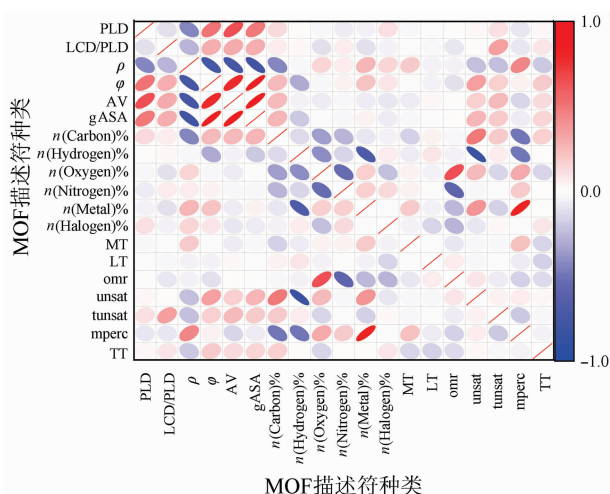


图 2 19 个特征描述符组成的 Pearson 相关矩阵
Fig. 2 Pearson correlation matrix of the 19 descriptors

2 结果与讨论

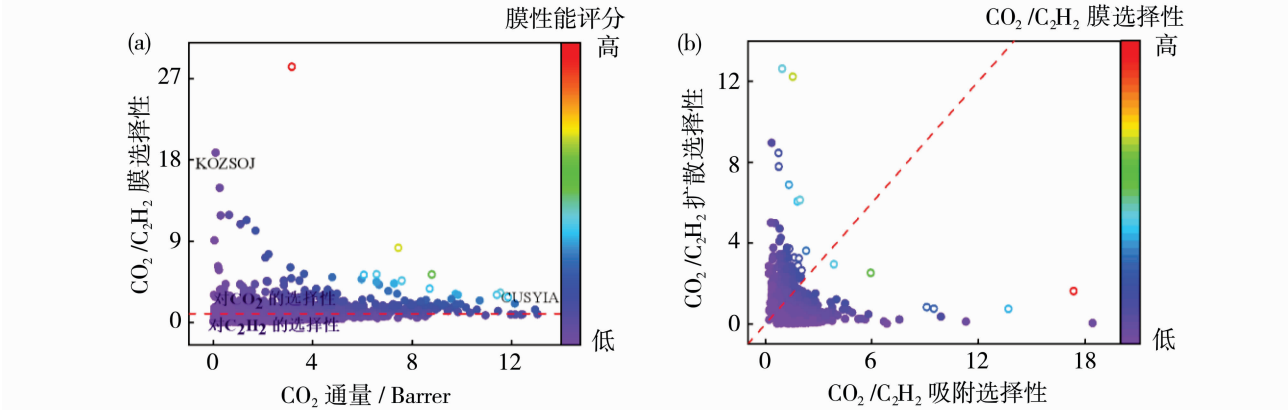
2.1 高性能疏水 MOF 膜筛选

为了筛选出用于 $\text{CO}_2/\text{C}_2\text{H}_2$ 分离的高性能疏水 MOF 膜, 采用分子模拟方法, 在 298.0 K 和 0.1 MPa 的条件下计算了 MOF 膜对 $\text{CO}_2/\text{C}_2\text{H}_2$ 二元混合物的热力学性质和扩散系数, 并评估了各 MOF 的膜分离性能。

图 3 (a) 绘制了疏水 MOF 膜的 CO_2 通量 (P_{CO_2})、选择性 ($S_{\text{mem}, \text{CO}_2/\text{C}_2\text{H}_2}$) 和膜性能评分 (MPS) 之间的关系。图中黑色虚线划定了 2 个不同的区域, 大多数 MOFs 位于黑色虚线的上方, 表明它们在 CO_2 选择性方面优于 C_2H_2 选择性。此外, 还发现, MOF 膜的 CO_2 渗透性与 $\text{CO}_2/\text{C}_2\text{H}_2$ 膜选择性之间存在 trade-off 效应。例如, KOZSOJ 具有较高的膜选择性 (18.81), 但其 CO_2 通量较差 ($8.37 \times$

10^3 Barrer)。同样,CUSYIA 具有较高的 CO_2 通量 (1.30×10^6 Barrer),但其膜选择性(1.45)远低于 KOZSOJ。因此,在计算筛选中,采用了更全面的膜分离性能评价指标(MPS),具有高 MPS 值的疏水 MOF 膜可被视为 $\text{CO}_2/\text{C}_2\text{H}_2$ 分离的潜在候选材料。为了更直观地展示 MOF 膜性能之间的差距,数据点根据 MPS 值的大小进行着色,颜色较深的数据点表示 MOFs 的膜性能评分较高,而颜色较浅的数据点表示膜性能评分较低。同时,使用空心圆

圈标记出 20 个膜分离性能最好的疏水 MOF 膜 ($8.93 \times 10^3 \geq \text{MPS} \geq 2.48 \times 10^3$)。同时,通过疏水 MOF 膜的吸附选择性($S_{\text{ads},\text{CO}_2/\text{C}_2\text{H}_2}$)、扩散选择性($S_{\text{diff},\text{CO}_2/\text{C}_2\text{H}_2}$)和膜选择性($S_{\text{mem},\text{CO}_2/\text{C}_2\text{H}_2}$)的关系[图 3 (b)]可以观察到,位于虚线下方的 MOF 数量略多,表明膜分离机制主要由吸附过程主导。另一方面,对于 Top 20 的 MOFs,大多数疏水 MOF 位于红色虚线之上,暗示着膜分离机制主要由扩散为过程主导。



数据点分别根据 MPS 和膜选择性($S_{\text{mem},\text{CO}_2/\text{C}_2\text{H}_2}$)值的大小进行着色,空心圆圈代表基于 MPS 值选择的最佳疏水 MOF 膜候选。虚线表示 $S_{\text{mem},\text{CO}_2/\text{C}_2\text{H}_2} = 1$ 。虚线上的点代表 MOF 膜的 $S_{\text{ads},\text{CO}_2/\text{C}_2\text{H}_2}$ 与 $S_{\text{diff},\text{CO}_2/\text{C}_2\text{H}_2}$ 相等。

图 3 疏水 MOF 膜 CO_2 通量(P_{CO_2})、选择性($S_{\text{mem},\text{CO}_2/\text{C}_2\text{H}_2}$)和膜性能评分(MPS) (a)以及吸附选择性($S_{\text{ads},\text{CO}_2/\text{C}_2\text{H}_2}$)、扩散选择性($S_{\text{diff},\text{CO}_2/\text{C}_2\text{H}_2}$)和膜选择性($S_{\text{mem},\text{CO}_2/\text{C}_2\text{H}_2}$)之间的关系(b)

Fig. 3 Relationship among P_{CO_2} , $S_{\text{mem},\text{CO}_2/\text{C}_2\text{H}_2}$ and MPS (a); adsorption ($S_{\text{ads},\text{CO}_2/\text{C}_2\text{H}_2}$), diffusion ($S_{\text{diff},\text{CO}_2/\text{C}_2\text{H}_2}$) and membrane selectivity ($S_{\text{mem},\text{CO}_2/\text{C}_2\text{H}_2}$) of hydrophobic MOFs (b)

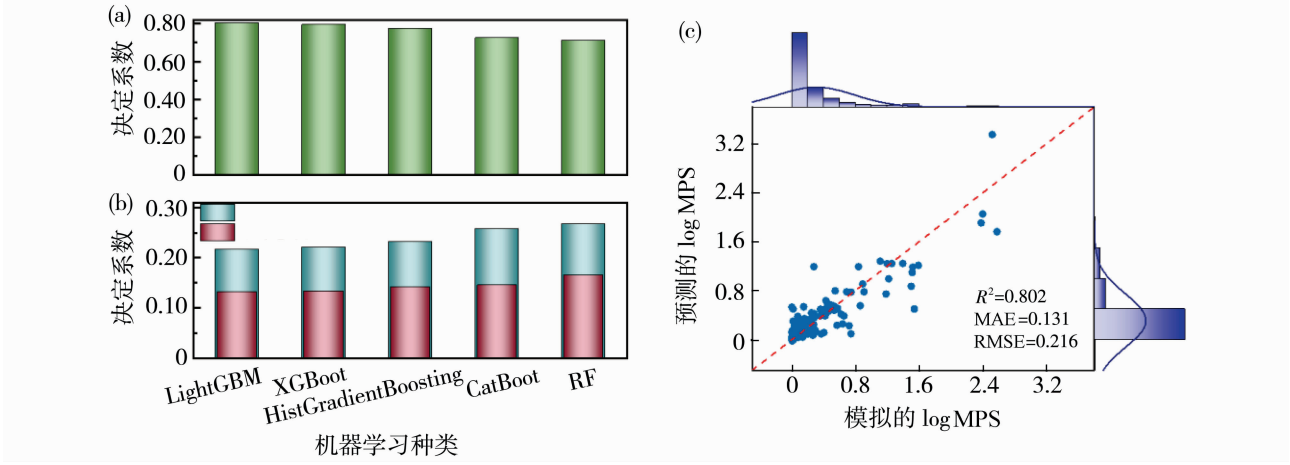
2.2 机器学习模型的性能评价

众所周知,MOFs 的物理和化学特征可以直接影响气体分离性能,MPS 是综合评价多孔 MOF 膜气体分离性能指标,因此,利用机器学习技术来揭示控制膜性能评分(MPS)的关键因素。为了寻找一种适合预测 MPS 的机器学习算法,分别对 RF、XGBoost、LightGBM、CatBoost 和 HistGradientBoosting 5 种机器学习算法进行了训练和测试。图 4(a)和 4(b)绘制了 5 种机器学习算法预测膜性能评分(MPS)的评估结果。可以看出,LightGBM 表现为最准确的模型,具有最高的 R^2 值(0.802),以及最低的 RMSE(0.216)和 MAE(0.131)。XGBoost 模型紧随其后, R^2 为 0.794, RMSE 为 0.221, MAE 为 0.133。相比之下,随机森林(RF)在五种机器学习算法中表现最差, R^2 为 0.711, RMSE 为

0.267, MAE 为 0.165。因此,预测膜性能评分(MPS)的模型准确性排序为: LightGBM > XGBoost > HistGradientBoosting > CatBoost > RF。图 4(c)进一步展示了预测值与计算值之间的良好一致性,因为大多数数据点集中在对角线附近。所有预测数据点呈向上倾斜的直线分布,且大部分数据点位于对角线上。

2.3 多变量分析

如上所述,LightGBM 模型在 5 种不同的 ML 模型中对预测 MPS 展现了较高的预测精度。因此,采用 LightGBM 模型进行多变量分析,以挖掘潜在的结构-性能关系。图 5 展示了 Top 20 的 MOF 膜分离性能的特征重要性分析。从整体来看,疏水 MOFs 的物理特征的重要性约为 52.14%,高于化学特征(47.86%)的重要性,表明物理特征对



边缘直方图的高度表示点的密度;预测结果和计算结果均以对数标度绘制

图 4 基于决定系数(R^2)(a), 均方根误差(RMSE)和平均绝对误差(MAE)(b)评估机器学习模型预测膜性能评分(MPS)的准确性;LightGBM 预测的 MPS 与分子模拟计算的结果比较(c)

Fig. 4 Evaluation of five ML algorithms based on R^2 (a) and RMSE and MAE values (b) for the MPS; comparison of the LightGBM-predicted MPS values of the hydrophobic MOFs with simulated MPS values on the testing set (c)

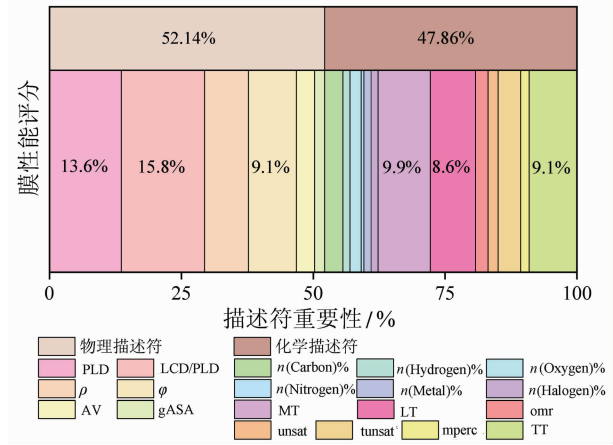


图 5 疏水 MOF 特征的相对重要性分析

Fig. 5 Relative importance analysis of hydrophobic MOF features

CO₂/C₂H₂ 分离性能具有显著影响。为了进一步挖掘有效信息并为实验合成 MOFs 提供理论指导,在图 5 中列出了所有物理描述符和化学描述符的重要性分析。MOFs 的特征描述符对 MPS 的贡献存在显著差异,其中 LCD/PLD 是所有描述符中最重要特征(15.8%),其次是 PLD(13.6%)和 MT(9.9%)。相比之下, $n(\text{Halogen})\%$ 和 $n(\text{Nitrogen})\%$ 的相对重要性较低,分别为 1.34% 和 0.53%,这表明这两种特征对 MOF 膜分离性能(MPS)的影响较弱。总之,影响 MOF 膜分离性能的关键是 LCD/PLD 和 PLD,这意味着可以通过调

节孔径大小来获得性能优异的 MOF 膜用于 CO₂/C₂H₂ 分离。

2.4 单变量分析

了解疏水 MOF 膜分离性能与其结构特征之间的关系,对于有效指导高分离性能新型材料的设计与合成,特别是在特定气体分离领域,具有重要意义。为了获得更优的 MOF 膜结构参数,绘制了 MPS 与机器学习模型预测的 2 个最重要结构参数(LCD/PLD 和 PLD)之间的构效关系图,见图 6。从图 6(a)可以看出,MOF 膜性能评分(MPS)随着 LCD/PLD 的增加而减小,表明当 MOF 结构的 PLD 和 LCD 值相近时,MOF 膜可获得较高的 MPS 值。最优的 MPS 值范围对应的 LCD/PLD 值为 1.11~2.12。同样,图 6(b)展示了 MPS 随着 PLD 的增加而减小,最佳的 PLD 范围为 0.373~0.605 nm,表明相对较小的孔隙有利于增强气体分子与 MOF 膜之间的相互作用,从而实现 CO₂/C₂H₂ 的高效分离。这些特征可作为设计具有优异膜分离性能 MOF 材料的重要指标。

3 结论

提出了一种耦合机器学习与分子模拟的方法,探索 MOFs 膜在 CO₂/C₂H₂ 分离中的构效关系。首先,以 ZIF-8 的亨利系数为标准,鉴定了 10 143 个 CoRE-MOF 数据库中的 MOFs 的疏水性,建立

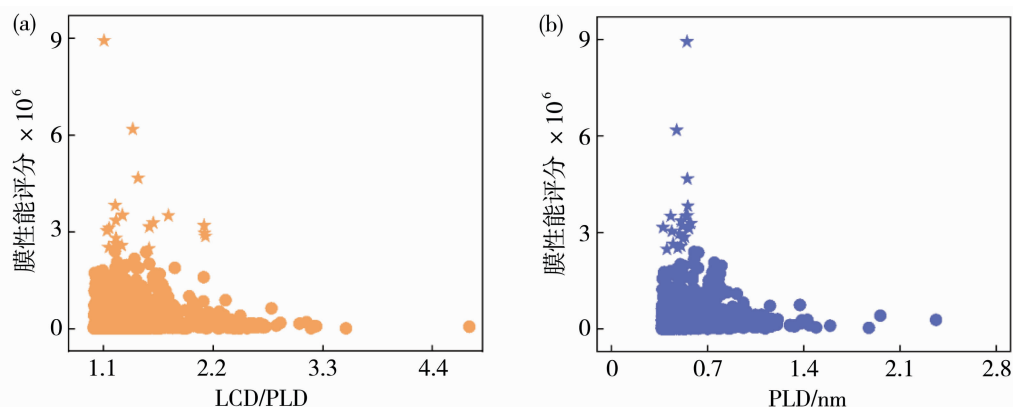


图 6 疏水 MOF 膜分离 $\text{CO}_2/\text{C}_2\text{H}_2$ 的膜性能评分(MPS)与 LCD/PLD(a)和 PLD(b)的关系

Fig. 6 The MPS of hydrophobic MOF membrane for $\text{CO}_2/\text{C}_2\text{H}_2$ separation as a function of LCD/PLD (a) and PLD (b)

了疏水 MOF 数据库。接着,筛选出孔限直径 (PLD) ≥ 0.33 nm 且 $\text{gASA} > 0$ m^2/g 的 1 733 个疏水 MOFs 用于后续分子模拟和机器学习。在分子模拟中,采用巨正则蒙特卡罗 (GCMC) 模拟和分子动力学 (MD) 模拟,分别计算了 $\text{CO}_2/\text{C}_2\text{H}_2$ 在 MOF 中的热力学性质和扩散系数,并评估了膜分离性能 [如 CO_2 通量, $\text{CO}_2/\text{C}_2\text{H}_2$ 膜选择性和膜性能评分 (MPS)]。最终,筛选出 Top 20 个性能优异的疏水 MOF 膜。随后,基于 MOFs 的特征描述符和模拟结果,训练和测试了包括随机森林 (RF)、极限梯度提升 (XGBoost)、轻量的梯度提升机 (LightGBM)、CatBoost 和直方图梯度提升 (HistGradientBoosting) 在内的 5 种不同的 ML 模型。结果显示,LightGBM 模型具有最高的预测准确性 ($R^2 = 0.802$)。通过多变量分析,发现 LCD/PLD 和 PLD 是影响 MOF 膜分离性能 (MPS) 的关键因素。最终,建立了构效关系,指出 LCD/PLD 和 PLD 值过大将不利于 $\text{CO}_2/\text{C}_2\text{H}_2$ 膜分离性能。最佳 MOF 结构参数范围为 LCD/PLD 值在 1.11~2.12 之间,PLD 值范围为 0.373 nm 至 0.605 nm。通过筛选 CoRE-MOF 数据库并构建构效关系,为后续 MOF 膜的开发和优化提供了理论指导。未来,将引入逆向设计策略,以指导实验合成高性能的 $\text{CO}_2/\text{C}_2\text{H}_2$ 分离膜候选结构。

参考文献:

- [1] Li X Y, Song Y, Zhang C X, *et al.* Inverse $\text{CO}_2/\text{C}_2\text{H}_2$ separation in a pillared-layer framework featuring a chlorine-modified channel by quadrupole-moment sieving [J]. *Sep Purif Technol*, 2021, 279: 119608.
- [2] Meng L K, Yang L X, Chen C L, *et al.* Selective acetylene adsorption within an imino-functionalized nanocage-based metal-organic framework [J]. *ACS Appl Mater Interfaces*, 2020, 12(5): 5999-6006.
- [3] Yang S Q, Xing B, Wang L L, *et al.* Boosting acetylene packing density within an isorecticular metal-organic framework for efficient $\text{C}_2\text{H}_2/\text{CO}_2$ separation [J]. *Chem & Bio Engineering*, 2024, 1(3): 245-251.
- [4] Park H B, Kamcev J, Robeson L M, *et al.* Maximizing the right stuff: The trade-off between membrane permeability and selectivity [J]. *Science*, 2017, 356(6343): eaab0530.
- [5] Li W B. Metal-organic framework membranes: Production, modification, and applications [J]. *Prog Mater Sci*, 2019, 100: 21-63.
- [6] Kang Z X, Fan L L, Sun D F. Recent advances and challenges of metal-organic framework membranes for gas separation [J]. *J Mater Chem A*, 2017, 5(21): 10073-10091.
- [7] Furukawa H, Ko N, Go Y B, *et al.* Ultrahigh porosity in metal-organic frameworks [J]. *Science*, 2010, 329(5990): 424-428.
- [8] Farha O K, Eryazici I, Jeong N C, *et al.* Metal-organic framework materials with ultrahigh surface areas: Is the sky the limit? [J]. *J Am Chem Soc*, 2012, 134(36): 15016-15021.
- [9] Zhang Y X, Liu Y, Hang Y T, *et al.* Crystal seeds induced interfacial growth of zirconium metal-organic framework membranes towards efficient hydrogen purification [J]. *J Membr Sci*, 2024, 689: 122171.
- [10] Zhou S, Wei Y Y, Li L B, *et al.* Paralyzed membrane: Current-driven synthesis of a metal-organic framework with sharpened propene/propane separation

- [J]. *Sci Adv*, 2018, 4(10): eaau1393.
- [11] Liu H, Cong S, Yan X, *et al.* Honeycomb-like hofmann-type metal-organic framework membranes for C₂H₂/CO₂ and H₂/CO₂ separation[J]. *J Membr Sci*, 2023, 669:121282.
- [12] Xie L H, Xu M M, Liu X M, *et al.* Hydrophobic metal-organic frameworks: Assessment, construction, and diverse applications[J]. *Adv Sci*, 2020, 7(4): 1901758.
- [13] Huang H L, Zhang W J, Liu D H, *et al.* Understanding the effect of trace amount of water on CO₂ capture in natural gas upgrading in metal-organic frameworks: A molecular simulation study[J]. *Ind Eng Chem Res*, 2012, 51(30): 10031-10038.
- [14] Li H L, Eddaoudi M, O'Keeffe M, *et al.* Design and synthesis of an exceptionally stable and highly porous metal-organic framework [J]. *Nature*, 1999, 402(6759): 276-279.
- [15] Chae H K, Siberio-Pérez D Y, Kim J, *et al.* A route to high surface area, porosity and inclusion of large molecules in crystals[J]. *Nature*, 2004, 427(6974): 523-527.
- [16] Shafiei M, Alivand M S, Rashidi A, *et al.* Synthesis and adsorption performance of a modified micro-mesoporous MIL-101 (Cr) for VOCs removal at ambient conditions[J]. *Chem Eng J*, 2018, 341: 164-174.
- [17] Daglar H, Keskin S. Recent advances, opportunities, and challenges in high-throughput computational screening of MOFs for gas separations[J]. *Coord Chem Rev*, 2020, 422: 213470.
- [18] Gómez-Gualdrón D A, Colón Y J, Zhang X, *et al.* Evaluating topologically diverse metal-organic frameworks for cryo-adsorbed hydrogen storage[J]. *Energ Environ Sci*, 2016, 9(10): 3279-3289.
- [19] Wilmer C E, Leaf M, Lee C Y, *et al.* Large-scale screening of hypothetical metal-organic frameworks [J]. *Nat Chem*, 2012, 4(2): 83-89.
- [20] Li S, Chung Y G, Simon C M, *et al.* High-throughput computational screening of multivariate metal-organic frameworks (MTV-MOFs) for CO₂ capture[J]. *J Phys Chem Lett*, 2017, 8(24): 6135-6141.
- [21] Zhang Z Q, Cao X H, Geng C X, *et al.* Machine learning aided high-throughput prediction of ionic liquid@ MOF composites for membrane-based CO₂ capture[J]. *J Membr Sci*, 2022, 650: 120399.
- [22] He Y J, Cao X H, Zhang Z Q, *et al.* Discovery of high-performing metal-organic frameworks for efficient SF₆/N₂ separation: A combined computational screening, machine learning, and experimental study[J]. *Ind Eng Chem Res*, 2023, 62(19): 7642-7649.
- [23] Chung Y G, Haldoupis E, Bucior B J, *et al.* Advances, updates, and analytics for the computation-ready, experimental metal-organic framework database: CoRE MOF 2019[J]. *J Chem Eng Data*, 2019, 64(12): 5985-5998.
- [24] Moghadam P Z, Fairen-Jimenez D, Snurr R Q. Efficient identification of hydrophobic MOFs: Application in the capture of toxic industrial chemicals[J]. *J Mater Chem A*, 2016, 4(2): 529-536.
- [25] Li J R, Kuppler R J, Zhou H C. Selective gas adsorption and separation in metal-organic frameworks[J]. *Chem Soc Rev*, 2009, 38(5): 1477-1504.
- [26] Willems T F, Rycroft C H, Kazi M, *et al.* Algorithms and tools for high-throughput geometry-based analysis of crystalline porous materials[J]. *Micropor Mesopor Mater*, 2012, 149(1): 134-141.
- [27] Rappé A K, Casewit C J, Colwell K S, *et al.* UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations[J]. *J Am Chem Soc*, 1992, 114(25): 10024-10035.
- [28] Wilmer C E, Kim K C, Snurr R Q. An extended charge equilibration method[J]. *J Phys Chem Lett*, 2012, 3(17): 2506-2511.
- [29] Harris J G, Yung K H. Carbon dioxide's liquid-vapor coexistence curve and critical properties as predicted by a simple molecular model[J]. *J Phys Chem*, 1995, 99(31): 12021-12024.
- [30] Fischer M, Hoffmann F, Fröba M. New microporous materials for acetylene storage and C₂H₂/CO₂ separation: Insights from molecular simulations[J]. *Chem Phys Chem*, 2010, 11(10): 2220-2229.
- [31] Yang Q Y, Liu D H, Zhong C L, *et al.* Development of computational methodologies for metal-organic frameworks and their application in gas separations [J]. *Chem Rev*, 2013, 113(10): 8261-8323.
- [32] Avci G, Velioglu S, Keskin S. High-throughput screening of MOF adsorbents and membranes for H₂ purification and CO₂ capture[J]. *ACS Appl Mater*, 2018, 10(39): 33693-33706.
- [33] Daglar H, Keskin S. High-Throughput screening of metal organic frameworks as fillers in mixed matrix membranes for flue gas separation[J]. *Adv Theory Simul*, 2019, 2(11): 1900109.

[34] Halder P, Singh J K. High-throughput screening of metal-organic frameworks for ethane-ethylene separation using the machine learning technique[J]. *Energy Fuel*, 2020, 34(11): 14591-14597.

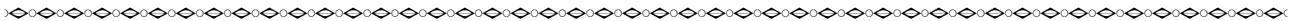
Coupling machine learning with high-throughput computing study
the structure-performance correlation of hydrophobic MOFs for
membrane-based CO₂/C₂H₂ separation

HAN Rongmei, HAN Qi, ZHANG Zhengqing,
SUN Yuxiu, QIAO Zhihua

(State Key Laboratory of Advanced Separation Membrane Materials, School of Chemical
Engineering and Technology, Tiangong University, Tianjin 300387, China)

Abstract: The CO₂/C₂H₂ separation in industrial processes has driven the development of highly efficient membrane materials. However, the presence of water under humidity environments can significantly impact the separation performance of these membranes. Thus, the development of separation membranes capable of maintaining superior performance under humidity conditions has become a critical area of focus. In this study, the hydrophobic MOF database was established, with the Henry coefficient of ZIF-8 serving as the reference. Then, the thermodynamic properties and diffusion coefficients of MOFs used for CO₂/C₂H₂ separation were calculated using grand canonical Monte Carlo (GCMC) and molecular dynamics (MD) simulations. Based on these calculations, the membrane separation performance was evaluated, and the top 20 hydrophobic MOF membranes with high separation performance were identified. The MOF descriptors and simulation results were utilized as inputs to train and test five different machine learning (ML) models and the key factors influencing CO₂/C₂H₂ separation performance were uncovered with the optimized ML model. Finally, the optimal range of MOF structural parameters for CO₂/C₂H₂ separation was determined. This study provides theoretical foundations for the development and optimization of MOF membrane materials.

Key words: metal-organic frameworks; molecular simulation; machine learning; CO₂/C₂H₂ separation



广 告 目 次

中科瑞阳膜技术(北京)有限公司	(封 面)	利得膜(北京)新材料技术有限公司	前插 5
沃顿科技股份有限公司	(封 二)	上海湛信科技发展有限公司	前插 6
北京碧水源膜科技有限公司	(封 三)	大连卡瑞博纳科技有限公司	前插 7
常州冀德环保科技有限公司	(封 四)	海水资源高效利用化工技术教育部工程研究中心	前插 8
海南立昇净水科技实业有限公司	前插 1	蓝星工程有限公司	前插 9
绵阳市聚合新材料有限公司	前插 2	天邦膜技术国家工程研究中心有限责任公司	前插 10
自然资源部天津海水淡化与综合利用研究所	前插 3		
山东招金膜天股份有限公司	前插 4		