

文章编号: 1671-4229(2024)05-0025-12

# 面向手性药物布洛芬拆分和运载的高能金属有机框架的机器学习 - 大数据研究

乔智威, 柯诗晴, 关珂昕, 许方怡, 黄晓珊, 关雅芳

(广州大学 化学化工学院, 广东 广州 510006)

**摘要:** 药物分离/药物负载材料已成为药物可控释放及药物制备技术中的重要研究对象之一。为了从大量的已有材料中选取可高效负载药物的候选者并探索其负载机理, 该研究从 CoRE-MOF 2019 数据库中抽取了 1 000 种 MOFs 材料, 通过高通量计算探讨它们对布洛芬药物的吸附负载性能。首先, 将 MOFs 的 8 种结构/能量描述符分别与 MOFs 对布洛芬药物分子的选择性 ( $S_{S-IBU/N_2}$ )、吸附量 ( $N_{S-IBU}$ ) 和两者权衡值 (Trade-off between  $S_{S-IBU/N_2}$  and  $N_{S-IBU}$ ) 进行单变量分析, 初步探索了不同描述符和 3 种性能指标之间的关系趋势。其次, 采用随机森林、极限梯度提升 (Extreme Gradient Boosting, XGB)、梯度提升树、轻量梯度提升树、反向传播神经网络和支持向量回归 6 种机器学习算法, 分别对 8 种描述符和 3 种性能评价标准 (吸附选择性  $S_{S-IBU/N_2}$ 、吸附量  $N_{S-IBU}$  和权衡值 TSN) 进行大数据训练和挖掘, 并建立定量关系。结果表明, 6 种 ML 算法的预测精度都是  $N_{S-IBU} > TSN > S_{S-IBU/N_2}$ 。对于  $S_{S-IBU/N_2}$  来说, XGB 表现出最佳的预测效果 ( $R^2 = 0.83$ )。接着, 基于 XGB 模型使用形状添加解释器 (SHapley Additive explanation, SHAP) 方法来解释和分析 MOF 描述符对性能指标的重要程度。MOF 吸附过程中产生的总能量被认为是关键的影响因素, 它与 TSN 和  $N_{S-IBU}$  都呈现正相关的趋势。最终, 结合毒理学分析, 推荐和设计了一系列高性能 MOF 材料。文章从分子层面、高通量计算到大数据挖掘, 系统地研究了布洛芬药物分子在 MOF 中的吸附运载机理, 为药物运载材料提供了理论指导。

**关键词:** 布洛芬; 金属有机框架; 机器学习

**中图分类号:** R979.9 **文献标志码:** A

## A machine Learning-Big Data Study on High-Energy Metal-Organic frameworks for chiral drug Ibuprofen resolution and transportation

QIAO Zhi-wei\*, KE Shi-qing, GUAN Ke-xin, XU Fang-yi, HUANG Xiao-shan, GUAN Ya-fang

(School of Chemistry and Chemical Engineering, Guangzhou University, Guangzhou 510006, China)

**Abstract:** Drug separation/drug loading materials have become one of the significant research objects in controlled drug release and drug preparation technology. In order to select candidates for efficient drug loading from a large number of existing materials and explore their loading mechanisms, 1 000 MOFs materials were extracted from the CoRE-MOF 2019 database for this study, and their adsorptive loading performance the drug ibuprofen was explored by high-throughput calculations. Firstly, the eight structure/energy descriptors of MOFs were analyzed by univariate analysis with the adsorption selectivity ( $S_{S-IBU/N_2}$ ), adsorption capacity ( $N_{S-IBU}$ ) and trade-off value (TSN) of MOFs for ibuprofen

收稿日期: 2024-08-08; 修回日期: 2024-09-10

基金项目: 国家自然科学基金项目(22478085, 21978058); 广东省自然科学基金项目(2023A1515240076, 2022A1515011446)

作者简介: 乔智威(1986—), 男, 教授. E-mail: zqiao@gzhu.edu.cn

引文格式: 乔智威, 柯诗晴, 关珂昕, 等. 面向手性药物布洛芬拆分和运载的高能金属有机框架的机器学习 - 大数据研究[J]. 广州大学学报(自然科学版), 2024, 23(5): 25-36.

drug molecules, and the relationship trend between different descriptors and the three performance indicators were preliminarily explored. Secondly, six machine learning models including Random Forest, Extreme Gradient Boosting (XGB), Gradient Boosting Decision Tree, Light Gradient Boosting Machine, Backpropagation Neural Network, and Support Vector Regression six machine learning algorithms with eight descriptors and three performance evaluation criteria (Adsorption selectivity  $S_{S-IBU/N_2}$ , Adsorption capacity  $N_{S-IBU/N_2}$  and Trade-off value TSN) for big data training and mining, were used to establish quantitative relationships. The results show that the prediction accuracy of the six ML algorithms is  $N_{S-IBU} > TSN > S_{S-IBU/N_2}$ . For  $S_{S-IBU/N_2}$ , XGB showed the best prediction ( $R^2 = 0.83$ ). Subsequently, based on the XGB model, the SHaple Additive explanation (SHAP) method was used to explain and analyze the importance of MOF descriptors to performance indicators. The total energy generated during MOF adsorption is considered to be the key influencing factor, and it shows a positive correlation trend with both TSN and  $N_{S-IBU}$ . Finally, combined with toxicological analysis, a series of high-performance MOF materials were recommended and designed. This work, from molecular level, high-throughput computing to big data mining, systematically studied the adsorption and delivery mechanism of ibuprofen drug molecules in MOF, which provides theoretical guidance for drug delivery materials.

**Key words:** Ibuprofen; Metal Organic Framework; machine learning

金属有机框架 (Metal-Organic Frameworks, MOFs) 是一类具有比表面积大、孔结构可调以及孔界限清晰的新型纳米材料,其结构由金属配体和有机连接体组成<sup>[1]</sup>。相较于传统的纳米多孔材料,MOF 的优势在于合成过程中能够调节材料的化学和结构特性。这也使 MOFs 在吸附分离、催化作用、传感器、质子传导等各种应用中体现出奇妙的功能<sup>[2-3]</sup>。除此之外,研究人员还关注到 MOFs 在药物储存、运输以及成像的药物化学领域的用途,其中,非处方药物布洛芬 (ibuprofen, IBU) 在医学和生物学中被广泛应用导致其需求量日益增加,它于二十世纪六十年代后期首次推出,用于治疗风湿性关节炎。如今,它已成为一种核心的非甾体抗炎药 (Non-steroidal Anti-Inflammatory Drugs, NSAID), 常用于缓解各种关节炎、痛经、痤疮和发烧等。IBU 存在 S-IBU 和 R-IBU 两种对映异构体,其中, S-IBU 在体外和体内均具有活性,因此,大多数研究首选考虑 S-IBU<sup>[4]</sup>。目前常用于 IBU 分离的方法有色谱法和萃取法等,但这些方法都存在着原料消耗大、分离纯度低、污染环境等不足,其中,脂质、聚合物纳米粒子、金属团簇和碳结构作为储存材料更是具有载药量低、药物释放快和含毒性等局限性<sup>[5]</sup>。近年来,MOFs 的多功能结构、改性后仍可控制的尺寸,以及形状和高度的均匀性

在多功能药物释放和刺激响应方面表现出色,且 MOFs 的弱配位键也保证了良好的生物降解性,这为药物释放和生物相容性提供了新的可能性。首次研究发现 MOFs 可提供较高的药物容量 (每克多孔固体可吸附 1.4 克 IBU, 比介孔二氧化硅材料的吸附量高 4 倍) 和较长的释放时间 (21 天)<sup>[6-7]</sup>。Srarker 等<sup>[8]</sup> 合成了一种稳定的多孔金属有机框架 Zr-DASDCA, 用草酰氯或对苯二甲酰氯进行改性,从而将各种官能团引入 Zr-DASDCA 上。研究表明,官能化后的 MOFs 不仅能够降低 IBU 分子的释放率,还能诱导它在长达 10 天的时间内持续释放。Horcajada 等<sup>[7]</sup> 在 MIL-101 (Cr) 上浸渍了高达 50 wt% 的 IBU 分子,并在 24 小时内观察到 70% 的总释放量。这一结果归因于 MIL-101 (Cr) 的高表面积 (4 500 m<sup>2</sup>/g) 以及 IBU 分子与路易斯酸金属位点和 MOF 上的有机分子之间形成的特殊相互作用。然而,由于框架中存在有毒的金属铬, MIL-101 (Cr) 在药物储存和输送方面的实际应用受到了限制。因而 MOFs 的生物相容性也至关重要,它代表着 MOF 材料在生物系统中的耐受性和相互作用,包括 MOF 与生物体内的细胞、组织或器官之间的相容性,以及其在体内的生物降解和代谢行为。通常为了降低有机连接剂和金属的毒性作用,使用氨基酸、核碱基和糖等分子作为连接分子,并且

选择生物相容性金属阳离子来构建生物 MOFs。Lawson 等<sup>[9]</sup>建立了两种用 IBU 给药的生物相容性介孔 MOFs (Zn MOF-74 和 UTSA-74), 发现 Zn MOF-74 是一种更优越的药物载体, 因为它能有效地释放更多的 IBU 分子, 并能在药物释放过程中溶解, 从而减少对人体的侵袭。Hartlieb 等<sup>[10]</sup>介绍了一种 IBU 分子与生物相容性 CD-MOF 的药用共晶体, IBU/CD-MOF-1 共晶体与 IBU 钾盐对照品一样能够迅速吸 IBU 气体, 并在 20 分钟内达到血浆浓度峰值。Miller 等<sup>[11]</sup>合成了由金属铁和烟酸组成的 Bio-MIL-1, 并证明了将治疗剂作为 MOF 框架成分的重要性。

目前, 计算就绪的实验金属有机框架 (Computation-Ready Experimental MOFs, CoRE-MOFs) 2019 数据库和假设的金属有机框架 (hypothetical MOFs, hMOFs) 数据库仍在更新中<sup>[12-13]</sup>, 能够从庞大的数据集中快速定位在药物吸收、储存、稳定释放以及生物相容性方面表现最优异的候选材料无疑是面临挑战的。实验研究虽能提供准确的数据和深入的材料理解, 但在材料的初步筛选和优化阶段往往耗时长、成本高, 且难以覆盖广泛的材料空间。相比之下, 机器学习 (Machine Learning, ML) 算法结合高通量计算筛选 (High-Throughput Computational Screening, HTCS) 方法可以通过模拟成千上万种不同的 MOF 结构和组合, 迅速识别出具有理想特性的材料, 为实验研究提供有力的指导。这不仅加快了研究进程, 也大幅降低了研究成本, 使得在短时间内探索和发现新材料成为可能。目前, 针对 MOFs 开发的大数据驱动方法主要集中在气体存储和分离特性等方面。Daglar 等<sup>[14]</sup>通过结合分子模拟和 ML 技术, 评估了 MOFs 膜和 MOF 聚合物混合基质膜在 6 种不同气体 (He/H<sub>2</sub>、He/N<sub>2</sub>、He/CH<sub>4</sub>、H<sub>2</sub>/N<sub>2</sub>、H<sub>2</sub>/CH<sub>4</sub> 和 N<sub>2</sub>/CH<sub>4</sub>) 上的分离性能。Wilmer 等<sup>[15]</sup>从 102 个构建块的化学库中生成了 137 953 种 hMOFs 并快速筛选出特定应用的最佳候选材料。Gao 等<sup>[16]</sup>结合 ML 辅助的 HTCS 和分子指纹技术, 加速发现适用于 H<sub>2</sub>S/CO<sub>2</sub>/CH<sub>4</sub> 分离的高性能 MOFs, 旨在提高天然气净化效率。但 MOFs 中药物传递的基本机制仍然难以捉摸, 因此, 原子和分子计算非常有助于在微观水平上提供基础物理学, 弥补实验上的局限, 并最终协助合理设计出新型纳米材料以更

好地给药。Babarao 等<sup>[4]</sup>采用蒙特卡罗 (Monte Carlo, MC) 模拟和分子动力学 (Molecular Dynamics, MD) 模拟来计算 298 K 下 MOF 中的负载和迁移率, 他们还采用模拟退火技术和第一性原理 (Density Functional Theory, DFT) 来研究单个 IBU 分子, 以鉴定 IBU 的优先吸附位点。Liu 等<sup>[17]</sup>对 28 种 MOFs 进行巨正则蒙特卡罗 (Grand Canonical Monte Carlo, GCMC) 模拟以评估它们对降压药物氨氯地平的吸附性能, 并通过 MD 模拟筛选出作为氨氯地平药物载体的有希望的候选者。Erucar 等<sup>[18]</sup>使用 MC 模拟来研究 IBU 和两种化妆品分子在生物相容性 MOFs 中的释放和储存, 并将模拟方法扩展到另外不同的 MOFs 中用于预测它们的储存性能。Liu 等<sup>[19]</sup>收集 100 多种出版物中的数据并结合 ML 算法评估了不同 MOFs 的 IBU 分子负载能力。Berinini 等<sup>[20]</sup>首先使用 IBU 分子在 MIL-53、MIL-100 和 MIL-101 中的吸附和释放的科实验数据验证模型, 使用 GCMC 来预测新型多孔材料在药物传递应用中的宏观性能。这种智能筛选和设计的方法能够帮助人们深入了解药物分子在多孔材料中的行为, 这也在药物的负载和释放中起着关键作用<sup>[21]</sup>。尽管如此, 利用计算模拟结合 ML 算法针对大量数据集来研究 MOFs 对药物分子的吸附机制的工作并不多, 分析 MOFs 的结构特性对 MOFs 药物吸附-传递现象的影响也尤为关键。

本工作将结合 6 种 ML 算法和 HTCS 方法筛选出生物相容性较好的 MOFs 材料, 作为 S-IBU 分子的载体系统。首先, 随机从 CoRE-MOFs 2019 数据库中挑选出 1 000 个小样本数据集, 对其建立结构-性能关系; 其次, ML 算法对多元描述符-性能建立回归预测模型, 并与模拟方法计算出的目标值进行拟合; 接着, 使用 Shapely 加法解释器从 S-IBU 负载能力的角度分析了特征效应的重要性, 以量化影响负载 S-IBU 分子的关键因子; 最后, 从金属中心类型、孔隙结构相似性以及 S-IBU 分子的吸附位点等方面, 探索最优 MOF 材料的共性。本工作中的方法、分析和结论为加速 S-IBU 吸附剂的发现提供了一种有效的途径, 对关键因素的评估能够更好地理解 MOF-IBU 的工作机制, 为设计出更具 S-IBU 靶向性的吸附剂提供指导方针, 最终为开发更安全、更高效的药物传输系统奠定坚实的基础。

## 1 方 法

### 1.1 分子模拟

从 CoRE-MOFs 2019 数据库<sup>[22]</sup> 中随机抽取 1 000 个 MOFs 进行重新编号,在固定压力下,使用 Materials Studio 软件中的吸附模块批量模拟 S-IBU 分子和 N<sub>2</sub> 分子(摩尔分数比 0.001:0.999)在 MOFs 中的吸附行为。所有 MOFs 材料和气体分子在结构优化的基础上,使用传统的 Metropolis 蒙特卡罗方法模拟吸附剂结构,模拟过程中,MOF 被视为是刚性结构,周期性边界应用三维系统,并且模拟的每个 MOF 晶胞沿 xyz 3 个维度扩展到至少 24 Å。在计算 Lennard-Jones(LJ)相互作用时,采取一种简化的处理方法,12 Å 以内的 MOF 框架上的原子或分子与气体分子之间的相互作用将被精确计算,超出此距离的粒子相互作用将被忽略。此外,使用 Ewald 求和用于计算静电相互作用。模拟总共执行 100 000 个周期,其中,前 10 000 个周期用于系统的平衡过程,余下的 90 000 个周期用于计算统计平均值。每个周期内试验  $n$  次操作( $n$  为吸附分子数量),随机执行 5 种不同类型的操作:平移、旋转、扭曲、生成与消除以及交换,各操作被赋予相同的发生概率。所有分子模拟均在 Materials studio 软件中进行,LJ 参数被用于描述 MOF 框架和吸附分子的非键相互作用力:

$$u_{LJ+elec}(r) = \sum 4\varepsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \sum \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \quad (1)$$

$u_{LJ+elec}(r)$  表示两原子之间的势能; $r_{ij}$  代表两个原子中心之间的距离; $\varepsilon_{ij}$  代表势能井的深度; $\sigma_{ij}$  是两个原子的势能恰好为零,从相互吸引转变为相互排斥的临界点之间的距离。其中, $\sum \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}$  是 Coulomb 相互作用项, $q_i$  和  $q_j$  是两个粒子的电荷; $\varepsilon_0 = 8.8542 \times 10^{-12} \text{ C}^2 \text{ N}^{-1} \text{ m}^{-2}$  代表真空中的介电常数。MOF 框架的 LJ 参数来源于通用力场(Universal Force Field, UFF)<sup>[23]</sup>,见表 S1(附录)。MOF 的原子电荷同样使用 MEPO-Qeq 方法<sup>[24]</sup> 进行估算。仿真研究表明<sup>[25]</sup>,UFF 可以准确预测 MOFs 的吸附和扩散。5 种结构描述符包括体积比表面积(Void Surface Area, VSA) ( $\text{m}^2/\text{cm}^3$ )、孔径

极限直径(Pore Limiting Diameter, PLD) (Å)、最大孔径(Largest Cavity Diameter, LCD) (Å)、密度  $\rho$  ( $\text{kg}/\text{m}^3$ )和孔隙率  $\phi$ 。借助直径为 2.58 Å 的 He 和直径为 3.64 Å 的 N<sub>2</sub> 作为探针,在 RASPA 软件包中计算  $\phi$  和 VSA。PLD、LCD 和  $\rho$  通过 Zeo++ 程序包估算。3 种能量描述符系统吸附热 Total Energy(kcal/mol)、布洛芬分子的等位热  $Q_{st,S-IBU}$  和氮气分子的等位热  $Q_{st,N_2}$  (kJ/mol)通过 Metropolis MC 方法在无限稀释状态下计算。在模拟吸附的过程中,固定总压力设置为 100 kPa,温度为 298 K,S-IBU 和 N<sub>2</sub> 的逸度比为 99.9:0.1。N<sub>2</sub> 的两个 N 原子通过一个非常强的三键(一条  $\sigma$  键和两条  $\pi$  键)连接在一起,N-N 键长为 1.10 Å。S-IBU 的分子构型包括一个苯环,它与一个异丙基(一个中心碳原子连接两个甲基)相连,并且这个中心碳原子还带有一个羧酸基团(-COOH)。苯环上的 C-C 键长约为 1.39 Å,C-H 键长约为 1.09 Å;羧酸基团中的 C-O 双键约为 1.22 Å,C-O 单键约为 1.39 Å;异丙基中的 C-C 双键约为 1.52 Å。两种分子的三维结构如图 1 所示。

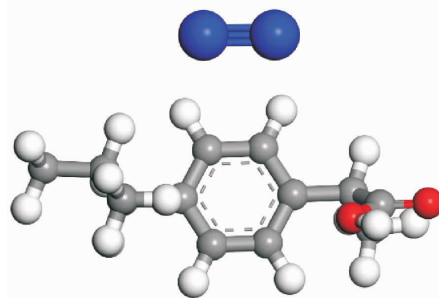


图 1 N<sub>2</sub> 和 S-Ibuprofen 的模型  
Fig. 1 Models of N<sub>2</sub> and S-Ibuprofen

### 1.2 方法

根据模拟出的两种气体的吸附量  $N_{S-IBU}$  和  $N_{N_2}$  (molecular/u·c),可以计算出另外两种吸附剂评价标准吸附选择性  $S_{S-IBU/N_2}$  和权衡值(Trade-off between  $S_{S-IBU/N_2}$  and  $N_{S-IBU}$ , TSN)。气体混合物在特定材料上的吸附选择性,即材料吸附一种成分优于另一种成分的能力,也是决定吸附分离性能可行性的主要参数之一<sup>[26]</sup>。然而,吸附剂的吸附容量仅代表它捕获足量目标分子的能力,不能用于选择性区分目标分子和其他分子。为了确保分离过程的高效性和经济性,因此,引入一个权衡值来评估和优化这两个参数之间的平衡。 $S_{S-IBU/N_2}$  和

TSN 的定义公式如下:

$$x'_i = \frac{x_i - \min(x)}{\max(x) - \min(x)}, \quad (2)$$

$$S_{S-IBU/N_2} = \frac{N_{S-IBU} \cdot y_{N_2}}{N_{N_2} \cdot y_{S-IBU}}, \quad (3)$$

$$TSN = N_{S-IBU} \cdot \ln(S_{S-IBU/N_2}). \quad (4)$$

本工作中采用了6种ML算法对于上述3种评价指标进行回归预测,所有模型均取自Scikit-learn软件包,并在Python3.8的环境下运行。6种ML算法包括随机森林(Random Forest, RF)、极限梯度提升(Extreme Gradient Boosting, XGB)、梯度提升树(Gradient Boosting Decision Tree, GBDT)、轻量梯度提升树(Light Gradient Boosting Machine, LGBM)、反向传播神经网络(Backpropagation Neural Network, BPNN)和支持向量回归(Support Vector Regression, SVR),设置训练集和测试集的比例为7:3。在使用ML算法进行拟合前,数据预处理是一个重要且必不可少的阶段,其主要目标是获得正确且对进一步数据挖掘算法有效的最终数据集<sup>[27]</sup>。首先,将投入的所有特征都进行标准化处理,方法见公式(2)。此外,由于 $S_{S-IBU/N_2}$ 的数值跨度非常大,预测时也对其进行一定的缩放。所有的ML算法都交叉运行5次, $R^2$ 、MAE和RMSE为评价指标。在ML算法预测的可解释性背景下,SHAP和轮廓线有助于权衡每个描述符因子的贡献度。SHAP通过遵循相加性、对称性和空效原则,为每个独立特征提供独特的解决方案<sup>[28]</sup>。SHAP值依赖于应用的ML算法( $f$ )和用于预测的输入数据( $x$ )。通过一个解释器( $g$ )的可解释函数和解释方法来解释模型做出的预测 $f(x)$ <sup>[29]</sup>。

## 2 结果

### 2.1 结构-性能关系

对Materials Studio模拟出的1000个数据进行筛选,去除对S-IBU分子吸附量为0的材料,最后得到391个MOFs用于进一步分析。首先,对结构-性能进行最基础的单变量分析,寻找单一材料属性与其在298 K和1 bar下的S-IBU负载能力之间的相关性。图2(a)和图2(b)显示了所有MOFs的VSA和LCD的统计直方图。以VSA和LCD为代表,6种结构描述符的分布存在两种类型。VSA的频数范围符合正态分布,因此,大多数的

MOF结构的比表面积都接近于平均值。而LCD的直方图符合偏态分布,较大一部分MOF的孔径都小于15 Å。4种描述符( $\rho$ 、PLD、LCD和 $\phi$ )与 $S_{S-IBU/N_2}$ 的趋势关系被建立在图2(c)中,对于3种描述符PLD、LCD和 $\phi$ 相同的是, $S_{S-IBU/N_2}$ 都在它们范围较大的区域内有较高值。这是因为 $N_2$ 的分子动力学直径(3.64 Å)明显小于S-IBU分子( $5 \times 10$  Å),其中,PLD和LCD被已知是一对相关性较高的描述符,当孔径较小时, $N_2$ 分子率先进入到MOF结构中,此时的 $S_{S-IBU/N_2}$ 几乎为零。而随着孔径的不断接近到5 Å直至不断增大, $S_{S-IBU/N_2}$ 也随着S-IBU分子开始进入MOFs而不断上升。当 $\phi$ 处于0.8~1的范围内,MOFs对于S-IBU分子的吸附选择性较高。这可能是因为在孔隙率较高时,MOFs材料的孔容积越大,此时能够容纳更多的S-IBU分子。而 $\rho$ 与其他3种特征呈现的趋势则相反,当 $\rho$ 逐渐增大时, $S_{S-IBU/N_2}$ 逐渐降低。MOF结构的密度相对较大意味着它们的内部结构更加紧凑,这可能会限制直径较大的S-IBU分子进入到孔隙中,较小的 $N_2$ 则容易进入而导致吸附选择性降低。除此之外,分析了能量描述符和性能指标之间的关系。放出的能量越多,MOF表面或孔隙与分子之间的相互作用强度更高,因而吸附量和吸附选择性都更高。

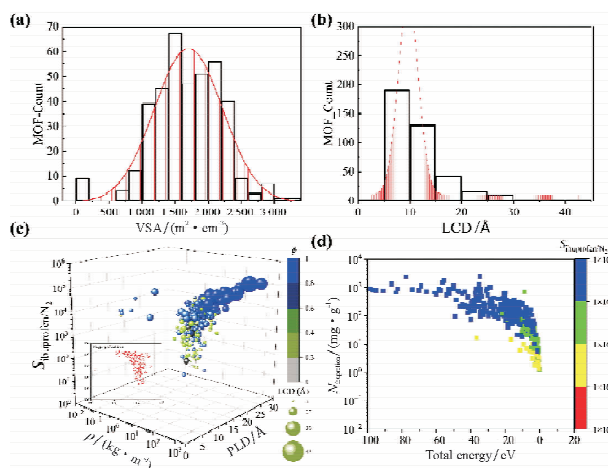


图2 特征分布及单变量关系图

Fig. 2 Feature distribution and univariate relationship diagram  
注:(a)VSA和(b)LCD的数据直方图;(c) $\rho$ 、PLD、LCD和 $\phi$ 的三维关系;(d)Total Energy和 $N_{S-IBU}$ 、 $S_{S-IBU/N_2}$ 的映射关系。

### 2.2 ML多元分析

本工作采用6种ML算法对3种性能指标( $N_{S-IBU}$ 、 $S_{S-IBU/N_2}$ 和TSN)进行回归预测。首先,对8

种描述符的数据进行预处理,将它们投放在(0,1)的范围内。图 3(a) 显示的是 8 种描述符进行预处理后的数值范围。可以观察到它们中位数相差不大,这能够消除数据间的差异性,以此来提高模型预测的收敛速度和泛化能力。总体而言,图 3(b) 中 6 种 ML 算法的预测精度都是  $N_{S-IBU} > TSN > S_{S-IBU/N_2}$ 。RF、GBRT、XGB、LGBM 和对测试集上的  $N_{S-IBU}$  和 TSN 都表现出  $R^2 > 0.9$  的预测效果,特别是 GBRT 和 XGB 的  $R^2$  几乎接近 1 的超高准确度。对于  $S_{S-IBU/N_2}$  来说,XGB 表现出最佳的预测效果 ( $R^2 = 0.83$ )。BPNN 和 SVR 模型仅仅对  $N_{S-IBU}$  的拟合程度较好,对 TSN 和  $S_{S-IBU/N_2}$  的预测明显低于其他模型。图 3(c) 是表现最佳的 XGB 模型对测试集上 TSN 的拟合效果,预测值与模拟值之间存在很强的线性关系。图 3(d) 比较了 6 种 ML 模型的运行时间,可以观察到 LGBM 和 SVR 运行速度都是最快的,而 XGB 和 GBRT 的速度相对较慢。这是因为 XGB 和 GBRT 在处理数据或者调优参数需要更多的计算资源和时间,因此,它们表现出相对较长的运行时间。综合考虑模型的执行时间和准确性,对于需要快速响应的大数据集,LGBM 可能更为适合;而对于需要高准确度的应用,XGB 模型是更好的选择。

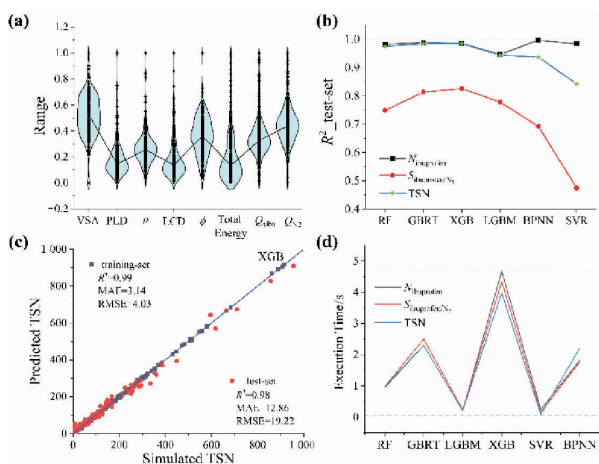


图 3 6 种 ML 算法预测效果及运行时间的对比

Fig. 3 Comparison of the prediction performance and running time of six ML algorithms

注:(a)8 种描述符归一化后的数值分布;(b)6 种回归算法 RF, GBRT, LGBM, XGB, BPNN 和 SVR 对  $N_{S-IBU}$ ,  $S_{S-IBU/N_2}$  和 TSN 的预测精度;(c)在测试集上使用 XGB 算法预测的 TSN 和模拟 TSN 值的拟合效果;(d) RF, GBRT, LGBM, XGB, SVR 和 BPNN 预测  $N_{S-IBU}$ ,  $S_{S-IBU/N_2}$  和 TSN 所需的运行时间。

## 2.3 基于 SHAP 值的重要性分析计算

模型的可解释性正在成为机器学习的重要研究趋势。为了进一步探讨这些变量如何影响 S-IBU 分子的负载性能,使用 SHAP 中的树解释器解释和挖掘 MOF 描述符对性能指标的重要程度。针对 TSN 和  $N_{S-IBU}$  两种性能指标,SHAP 首先计算每个样本中各个描述符对目标值的贡献度 (SHAP value),然后将每一列描述符所有样本中 SHAP value 的绝对值相加,最后计算它们的平均贡献值。图 4(a) ~ 图 4(b) 左侧的柱状图和右侧的束图都是基于 XGB 模型进行训练得到的。系统中产生的总能量 Total Energy 被认为是关键的影响因素,它与 TSN 和  $N_{S-IBU}$  都呈现正相关的趋势。位于较大的  $Q_{st,S-IBU}$  与两种指标都负相关。图 4(c) 显示的是  $Q_{st,S-IBU}$  和 LCD 基于 XGB 模型中的特征交互图,划定  $Q_{st,S-IBU}$  的 SHAP\_value 值小于 0 是对最终目标值有负影响的区域。[60, 100] 范围内的  $Q_{st,S-IBU}$  存在正的 shap value 值,此时,布洛芬与 MOF 间的作用力较强而产生更高的吸附能力。如果  $Q_{st,S-IBU}$  值过高 ( $Q_{st,S-IBU} > 100$ ),可能意味着 MOF 孔道过于紧密,S-IBU 分子难以进入,反而会降低吸附量。因此,适中的  $Q_{st,S-IBU}$  能够提供足够的相互作用力的同时,保持适当的孔道可进入性,使布洛芬分子能够更多地被 MOF 材料吸附。尽管较高的等位吸附热代表 MOF 与吸附质之间及较强的作用力,但对 S-IBU 分子产生较高吸附能的 MOFs 普遍具有较小的孔径,这使得 S-IBU 分子很难进入到材料孔道中,从而影响 MOFs 对 S-IBU 分子的负载量和吸附选择性。除此之外,与 MOFs 孔径大小相关的两种描述符 PLD 和 LCD 的重要性分别居于第三、五位,当它们的值较大时,贡献度也就越大,这说明孔径大小对提高 MOFs 对 S-IBU 分子的吸附能力也十分重要。LCD 与  $N_{S-IBU}$  和 TSN 呈现正相关的关系,这是因为当 MOF 的 LCD 小于布洛芬的动力学直径 ( $5 \times 10 \text{ \AA}$ ) 时,布洛芬分子无法进入 MOF 内部,从而导致吸附量较低或基本不发生吸附。当 LCD 逐渐增大到大于布洛芬的动力学直径时,布洛芬分子逐渐进入 MOF 孔道中发生吸附作用。这也与刘蓓等<sup>[30]</sup>的研究一致。

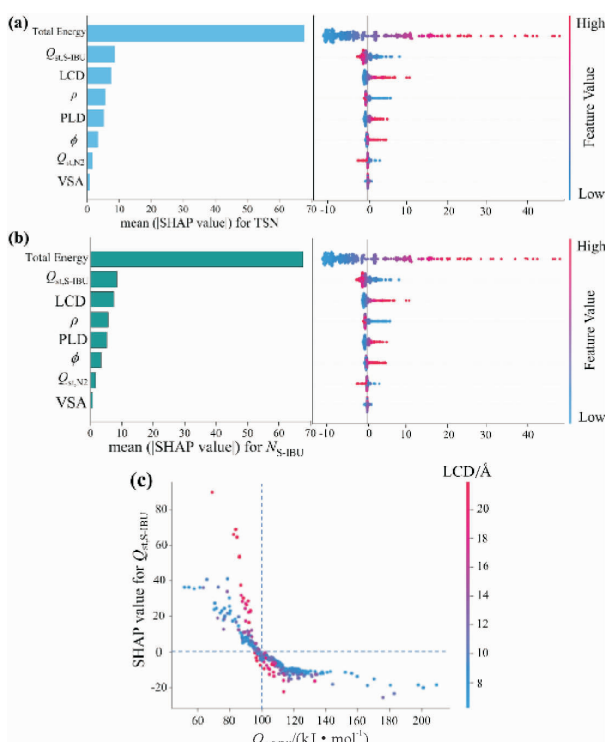


图4 每种描述符的 SHAP 贡献度

Fig. 4 SHAP contribution of each descriptor

注:SHAP 解释器中 8 种描述符对 (a) TSN, (b)  $N_{S-IBU}$  的重要性; (c)  $Q_{st,S-IBU}$  和 LCD 在 XGB 模型中对 TSN 进行预测的形状依赖图。图(a-c)中每一个点都代表单个数据点。颜色代表特征值,红色点数值较高,蓝色点数值较低;正的 SHAP 值表示对预测值有正相关影响,负的 SHAP 值则负相关。

## 2.4 优异 MOFs 的共性

大量的研究表明<sup>[31-32]</sup>, MOFs 中的金属中心类型和开放金属位点 (Open Metal Site, OMS) 可能存在于气体的最佳吸附点位。因此,探究金属中心对于 S-IBU 分子负载可能存在的作用机理。首先,统计出 391 个 MOFs 中的金属中心类型,图 5 (a) 仅显示出出现频数大于 5 次的金属中心。这些金属中心被划分为三大类,分别统计了它们在所有金属中心中的占比,发现过渡金属元素的占比 (约 78%) 远远高于另外两种金属类型,其中, Cu、Zn 和 Co 的数量较多。尽管如此,这并不能代表过渡金属元素是影响 S-IBU 分子负载量的关键,因而对根据 TSN 排序前 100 位 MOFs 同样统计金属元素,结果显示在图 5 (b)。同样的,出现频次大于 5 以上的排名前 6 位的 Cu、Zn、Ag、Mn、Cd 和 Cr 都属于过渡金属。随着向较优子集的逐渐移动,过渡金属的占比从约 78% 增长到约 89%。这说

明含有过渡金属元素的 MOF 结构很有希望成为负载 S-IBU 分子的高性能材料。这可能是由于过渡金属离子具有空的 d 轨道,可以与布洛芬分子中的羧基形成稳定的配位键<sup>[33]</sup>。这种配位作用可以增强布洛芬分子在 MOF 表面的吸附强度,提高负载效率。此外,过渡金属的配位环境可以引入极性基团或产生局部电荷,这可能促进与布洛芬分子的羧基或其他官能团形成氢键,提高吸附稳定性<sup>[34]</sup>。

此外,作为具有药物负载功能的 MOF 材料,良好的生物相容性也必须考虑。就人类的毒理学风险而言,金属元素可以区分成两大类:一类是低生物利用度使它们对生命不是必需的;另一类是迄今为止只发现过负面影响的元素。在后一组中,包括具有软“酸”化学特性的金属,如 Hg、Pb 和 Cd 等<sup>[35]</sup>。当今被认为是人类必不可少的金属元素包括 Na、K、Mg、Ca、Mn、Fe、Co、Ni、Cu 和 Zn 等<sup>[36]</sup>。在表 S2 (附录) 列出了排名前 15 位的 MOF 材料,发现它们大部分都以人体所需的微量元素为金属中心。在去除含有 Ag、In 为金属中心的 MOF 材料,使用 Yang 等<sup>[37]</sup> 开发的拓扑数据方法来探索剩余 10 种 MOFs 的孔隙结构是否具有相同点,并量化了 S-IBU 分子化学环境的相似性。Lee 等<sup>[38]</sup> 利用一种孔隙识别方法来量化孔隙结构的相似性,并使用拓扑数据分析对用于甲烷储存的材料进行分类。因此,这种方法不仅能够找到与负载 S-IBU 性能最佳结构在几何相似的结构,还可以根据其孔形的相似性设计和组织新的用于捕获 S-IBU 分子的结构。其中,两种材料间的相似性被转化为根据持久同源性计算出的归一化距离,距离越小即代表材料之间越相似<sup>[39]</sup>。图 5 (c) 显示的是 10 种 MOFs 材料的两两条形码转化的距离,右边的颜色带映射距离大小。首先,以排名第一的 MOF 材料为中心,发现排名分别为第 2、3、6、9 和 14 的 MOF 材料与其的距离都依次较近,且第 2、3 和 6 的 MOF 材料的距离基本接近于 0,数据表明,它们之间具有相似的孔隙结构。基于以上得出的结论,使用 Materials Studio 观察这 6 种 MOFs 的结构特征,它们的孔道结构如图 6 (a) ~ 图 6 (f) 所示。

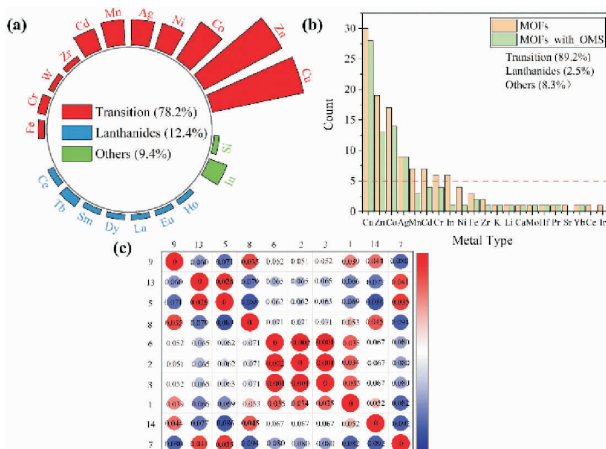


图5 MOFs 金属中心类型分布及优异 MOFs 间的相关性分析

Fig. 5 Distribution of metal centers in MOFs and correlation analysis between superior MOFs

注:(a)在 391 种 MOFs 中出现频次大于 5 以上的金属中心类型,圆圈中心的占比表示在所有 MOFs 中的金属中心类型占比;(b) TOP 100 MOFs 中金属中心和 OMS 的个数,占比统计的是 TOP 100 MOFs 中的金属中心类型;(c) 10 种生物相容性较好的优异 MOFs 之间的 Landscape 距离。

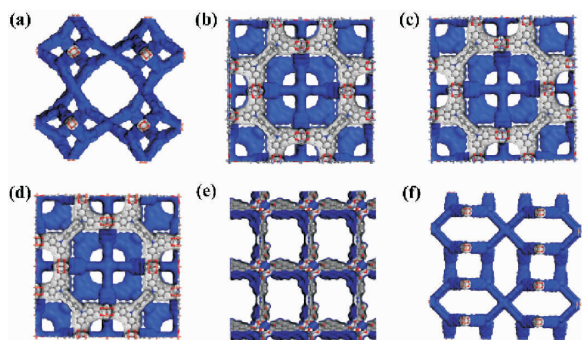


图6 6 种具有孔道相似性的优异 MOFs

Fig. 6 Six exceptional MOFs with pore similarities

注:(a)为 AVAJUE;(b)为 ACOCUS;(c)为 ACODAZ;(d)为 ACO-COM;(e)为 5024922;(f)为 AVAKAL

发现它们都具有较大的 PLD,范围在 12 ~ 19 Å 之间,这说明它们的孔道尺寸相对适中,S-IBU 分

子能够在进入后被有效吸附。此外,这 6 种结构中菱形和方形的孔道结构居多,相比于圆形孔道,菱形和方形孔道的空间利用率更高,并且均匀分布的孔道连通性能减少扩散阻力,提高吸附能力<sup>[40-41]</sup>。这为指导新材料的合成在孔道形状上提供一定的见解。

### 3 结 论

本研究旨在从小数据样本集上筛选出新型 MOFs 药物载体用于吸附和传递 S-IBU 分子,并深入分析优异 MOFs 的共性。首先,初步统计描述符的分布区间并建立它们与 3 种性能指标 ( $N_{S-IBU}$ 、 $S_{S-IBU}$  和 TSN) 间的关系,发现特征描述符的分布有两种类型:正态和偏态分布,PLD、LCD 和  $\phi$  与  $S_{S-IBU}/N_2$  存在正相关的趋势,且  $\phi$  在 0.8 ~ 1 的范围内存在较好的  $S_{S-IBU}/N_2$ 。其次,比较 6 种 ML 模型 (RF、XGB、GBRT、LGBM、BRNN 和 SVR) 对 3 种指标的预测效果和运行时间,XGB 的预测精度最佳,LGBM 和 SVR 的运行时间最短。紧接着,对最佳模型 XGB 使用 SHAP 解释器来计算每种描述符对输出样本的权重值,以量化它们的特征贡献度。系统产出的总能量 Total Energy、 $Q_{st,S-IBU}$  和 LCD 是影响 S-IBU 分子负载的关键因素。统计了所有 MOFs 和 TOP 100 MOFs 的金属中心类型,发现过渡金属的占比最高,且占比随着向最优子集的逐渐移动从 78% 上升到 89%。最后,根据 TSN 筛选出了前 15 个优异 MOFs 并考虑它们的生物相容性,挑选出 10 种生物相容性较好的 MOFs 并量化它们的拓扑相似性,其中,有 6 种 MOFs 的孔道存在相似性,以较为规则的菱形和方形孔道结构为主。

### 参考文献:

- [1] 杨文远,梁红,乔智威.高通量筛选金属-有机框架:分离天然气中的硫化氢和二氧化碳[J].化学学报,2018,76(10):785-792.
- [2] 蔡铖智,李丽凤,邓小梅,等.基于机器学习和高通量计算筛选金属有机框架的甲烷/乙烷/丙烷分离性能[J].化学学报,2020,78(5):427-436.
- [3] 刘克清,崔燕,于琼,等.氨基酸修饰的 MOF-808 催化降解神经毒剂的模拟酶性质研究[J].化工学报,2024,75(2):616-625.
- [4] Babarao R, Jiang J. Unraveling the energetics and dynamics of ibuprofen in mesoporous metal-organic frameworks[J]. The

- Journal of Physical Chemistry C, 2009, 113(42): 18287-18291.
- [5] 裴震, 郭建栋, 张倩, 等. 金属-有机骨架抗菌复合材料与纤维的研究进展及应用[J]. 复合材料学报, 2021, 38(8): 2396-2403.
- [6] Horcajada P, Serre C, Vallet-Regí M, et al. Metal-organic frameworks as efficient materials for drug delivery[J]. *Angewandte Chemie International Edition*, 2006, 45(36): 5974-5978.
- [7] Horcajada P, Serre C, Maurin G, et al. Flexible porous metal-organic frameworks for a controlled drug delivery[J]. *Journal of the American Chemical Society*, 2008, 130(21): 6774-6780.
- [8] Sarker M, Shin S, Jhung S H. Synthesis and functionalization of porous Zr-diaminostillbenedicarboxylate metal-organic framework for storage and stable delivery of ibuprofen[J]. *ACS Omega*, 2019, 4(6): 9860-9867.
- [9] Lawson S, Newport K, Schueddig K, et al. Optimizing ibuprofen concentration for rapid pharmacokinetics on biocompatible zinc-based MOF-74 and UTSA-74[J]. *Materials Science and Engineering: C*, 2020, 117: 111336.
- [10] Hartlieb K J, Ferris D P, Holcroft J M, et al. Encapsulation of ibuprofen in CD-MOF and related bioavailability studies[J]. *Molecular Pharmaceutics*, 2017, 14(5): 1831-1839.
- [11] Miller S R, Heurtaux D, Baati T, et al. Biodegradable therapeutic MOFs for the delivery of bioactive molecules[J]. *Chemical Communications*, 2010, 46(25): 4526-4528.
- [12] Burner J, Luo J, White A, et al. ARC-MOF: A diverse database of metal-organic frameworks with DFT-derived partial atomic charges and descriptors for machine learning[J]. *Chemistry of Materials*, 2023, 35(3): 900-916.
- [13] Bobbitt N S, Shi K, Bucior B J, et al. MOFX-DB: An online database of computational adsorption data for nanoporous materials[J]. *Journal of Chemical & Engineering Data*, 2023, 68(2): 483-498.
- [14] Daglar H, Erucar I, Keskin S. Recent advances in simulating gas permeation through MOF membranes[J]. *Materials Advances*, 2021, 2(16): 5300-5317.
- [15] Wilmer C E, Leaf M, Lee C Y, et al. Large-scale screening of hypothetical metal-organic frameworks[J]. *Nature Chemistry*, 2012, 4(2): 83-89.
- [16] Gao W, Zheng W, Yan K, et al. Accelerating the discovery of acid gas-selective MOFs for natural gas purification: A combination of machine learning and molecular fingerprint[J]. *Fuel*, 2023, 350: 128757.
- [17] Liu J, Yang Z, Che Y, et al. Computational investigation of metal organic frameworks as potential drug carriers for antihypertensive amlodipine[J]. *AIChE Journal*, 2022, 68(1): e17474.
- [18] Erucar I, Keskin S. Efficient storage of drug and cosmetic molecules in biocompatible metal organic frameworks: A molecular simulation study[J]. *Industrial & Engineering Chemistry Research*, 2016, 55(7): 1929-1939.
- [19] Liu X, Wang Y, Yuan J, et al. Prediction of the ibuprofen loading capacity of MOFs by machine learning[J]. *Bioengineering*, 2022, 9(10): 517.
- [20] Bernini M C, Fairen-Jimenez D, Pasinetti M, et al. Screening of bio-compatible metal-organic frameworks as potential drug carriers using monte carlo simulations[J]. *Journal of Materials Chemistry B*, 2014, 2(7): 766-774.
- [21] Kotzabasaki M, Froudakis G E. Review of computer simulations on anti-cancer drug delivery in MOFs[J]. *Inorganic Chemistry Frontiers*, 2018, 5(6): 1255-1272.
- [22] 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]. *Journal of Chemical & Engineering Data*, 2019, 64(12): 5985-5998.
- [23] Rappe 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]. *Journal of the American Chemical Society*, 1992, 114(25): 10024-10035.
- [24] Kadantsev E S, Boyd P G, Daff T D, et al. Fast and accurate electrostatics in metal organic frameworks with a robust charge equilibration parameterization for high-throughput virtual screening of gas adsorption[J]. *Journal of Physical Chemistry Letters*, 2013, 4: 3056-3061.
- [25] Garberoglio G, Skoulidas A I, Johnson J K. Adsorption of gases in metal organic materials: Comparison of simulations and

- experiments[J]. *The Journal of Physical Chemistry B*, 2005, 109(27): 13094-13103.
- [26] Pillai R S, Pinto M L, Pires J, et al. Understanding gas adsorption selectivity in IRMOF-8 using molecular simulation[J]. *ACS Applied Materials & Interfaces*, 2015, 7(1): 624-637.
- [27] Garcia S, Luengo J, Herrera F. Tutorial on practical tips of the most influential data preprocessing algorithms in data mining [J]. *Knowledge-Based Systems*, 2016, 98: 1-29.
- [28] Shapley L S. A value for n-person games [EB/OL]. (1988-10-01) [2024-07-08]. <https://www.semanticscholar.org/paper/A-Value-for-n-person-Games-Shapley/19088a582f2eb657ac1803f1ea1b79058d5c3dc7>.
- [29] Lundberg S M, Lee S-I. A unified approach to interpreting model predictions, December 03, 2017 [EB/OL]. (2017-12-03) [2024-03-15]. 2017; arXiv:1705.07874v2. <https://www.xueshufan.com/publication/2618851150>.
- [30] 刘蓓, 廉源会, 李智, 等. 生物金属-有机骨架材料中药物吸附及扩散的分子模拟研究[J]. *化学学报*, 2014, 72(8): 942-948.
- [31] Li Y, Situ Y, Guan K, et al. High dynamic separation performance of metal-organic frameworks for D<sub>2</sub>/H<sub>2</sub>: Independent or competitive adsorption? [J]. *AIChE Journal*, 2024, 70(1): e18283.
- [32] Guan K, Xu F, Huang X, et al. Deep learning and big data mining for metal-organic frameworks with high performance for simultaneous desulfurization and carbon capture[J]. *Journal of Colloid and Interface Science*, 2024, 662: 941-952.
- [33] Tella A C, Olatunji S J, Ajibade P A. Functionalization of a porous copper (ii) metal-organic framework and its capacity for loading and delivery of ibuprofen[J]. *RSC Advances*, 2024, 14(35): 25759-25770.
- [34] Seo P W, Bhadra B N, et al. Adsorptive removal of pharmaceuticals and personal care products from water with functionalized metal-organic frameworks: remarkable adsorbents with hydrogen-bonding abilities[J]. *Scientific Reports*, 2016, 6(1): 34462.
- [35] Jan A T, Azam M, Siddiqui K, et al. Heavy metals and human health: Mechanistic insight into toxicity and counter defense system of antioxidants[J]. *International Journal of Molecular Sciences*, 2015, 16(12): 29592-29630.
- [36] Maret W. The metals in the biological periodic system of the elements: Concepts and conjectures[J]. *International Journal of Molecular Sciences*, 2016, 17(1): 66.
- [37] Yang Y, Guo S, Li S, et al. Topological data analysis combined with high-throughput computational screening of hydrophobic metal-organic frameworks: Application to the adsorptive separation of C<sub>3</sub> components[J]. *Nanomaterials*, 2024, 14(3): 298.
- [38] Lee Y, Barthel S D, Dłotko P, et al. Quantifying similarity of pore-geometry in nanoporous materials[J]. *Nature Communications*, 2017, 8(1): 1-8.
- [39] 程洪, 葛美伶, 司天宇, 等. 机器学习辅助金属材料力学性能预测[J]. *材料研究与应用*, 2023, 17(6): 1070-1077.
- [40] Wang J, Chen J, Xu T. The effect of a novel D-camphoric acid-based MOF on chiral separation[J]. *Solid State Sciences*, 2019, 98: 106032.
- [41] Das S, Sekine T, Mabuchi H, et al. Three-dimensional covalent organic framework with scu-c topology for drug delivery [J]. *ACS Applied Materials & Interfaces*, 2022, 14(42): 48045-48051.

【责任编辑: 陈 钢】

附录:

表 S1 UFF 中 MOFs 的 LJ 参数

Table S1 Lennard-Jones parameters of MOFs from UFF

Atom	$\epsilon/\text{kJ/K}$	$\sigma/\text{\AA}$	Atom	$\epsilon/\text{kJ/K}$	$\sigma/\text{\AA}$	Atom	$\epsilon/\text{kJ/K}$	$\sigma/\text{\AA}$
Ac	16.60	3.10	Ge	190.69	3.81	Po	163.52	4.20
Ag	18.11	2.80	Gd	4.53	3.00	Pr	5.03	3.21
Al	254.09	4.01	H	22.14	2.57	Pt	40.25	2.45
Am	7.04	3.01	Hf	36.23	2.80	Pu	8.05	3.05
Ar	93.08	3.45	Hg	193.71	2.41	Ra	203.27	3.28
As	155.47	3.77	Ho	3.52	3.04	Rb	20.13	3.67
At	142.89	4.23	I	170.57	4.01	Re	33.21	2.63
Au	19.62	2.93	In	301.39	3.98	Rh	26.67	2.61
B	90.57	3.64	Ir	36.73	2.53	Rn	124.78	4.25
Ba	183.15	3.30	K	17.61	3.40	Ru	28.18	2.64
Be	42.77	2.45	Kr	110.69	3.69	S	137.86	3.59
Bi	260.63	3.89	La	8.55	3.14	Sb	225.91	3.94
Bk	6.54	2.97	Li	12.58	2.18	Sc	9.56	2.94
Br	126.29	3.73	Lu	20.63	3.24	Se	146.42	3.75
C	52.83	3.43	Lr	5.53	2.88	Si	202.27	3.83
Ca	119.75	3.03	Md	5.53	2.92	Sm	4.03	3.14
Cd	114.72	2.54	Mg	55.85	2.69	Sn	285.28	3.91
Ce	6.54	3.17	Mn	6.54	2.64	Sr	118.24	3.24
Cf	6.54	2.95	Mo	28.18	2.72	Ta	40.75	2.82
Cl	114.21	3.52	N	34.72	3.26	Tb	3.52	3.07
Cm	6.54	2.96	Na	15.09	2.66	Tc	24.15	2.67
Co	7.04	2.56	Ne	21.13	2.66	Te	200.25	3.98
Cr	7.55	2.69	Nb	29.69	2.82	Th	13.08	3.03
Cu	2.52	3.11	Nd	5.03	3.18	Ti	8.55	2.83
Cs	22.64	4.02	No	5.53	2.89	Tl	342.14	3.87
Dy	3.52	3.05	Ni	7.55	2.52	Tm	3.02	3.01
Eu	4.03	3.11	Np	9.56	3.05	U	11.07	3.02
Er	3.52	3.02	O	30.19	3.12	V	8.05	2.80
Es	6.04	2.94	Os	18.62	2.78	W	33.71	2.73
F	25.16	3.00	P	153.46	3.69	Xe	167.04	3.92
Fe	6.54	2.59	Pa	11.07	3.05	Y	36.23	2.98
Fm	6.04	2.93	Pb	333.59	3.83	Yb	114.72	2.99
Fr	25.16	4.37	Pd	24.15	2.58	Zn	62.39	2.46
Ga	208.81	3.90	Pm	4.53	3.16	Zr	34.72	2.78

表 S2 TSN 排名前 15 位 MOFs 材料

Table S2 Top 15 MOFs Material based on TSN

No	CSD	Metal	VSA $/(\text{m}^2 \cdot \text{cm}^{-3})$	PLD $/\text{\AA}$	$\rho$ $/(\text{kg} \cdot \text{m}^{-3})$	LCD $/\text{\AA}$	$\phi$	$Q_{\text{st},\text{S-IBU}}$ $/(\text{kJ} \cdot \text{mol}^{-1})$	$Q_{\text{N}_2}$ $/(\text{kJ} \cdot \text{mol}^{-1})$	Total Energy $/(\text{kJ} \cdot \text{mol}^{-1})$	$N_{\text{S-IBU}}$ $/(\text{mol} \cdot \text{u}^{-1} \cdot \text{c}^{-1})$	$S_{\text{S-IBU}/\text{N}_2}$	TSN
1	AVAJUE	Cu	1 016.69	19.54	164.12	37.53	0.95	69.29	7.70	2 271.27	97.583	80 786.79	1 102.65
2	ACOCUS	Zn	15 23.24	13.22	346.49	24.38	0.93	83.80	11.82	2 234.51	85.232	74 292.62	955.94
3	ACODAZ	Co	1 539.04	13.12	343.14	23.29	0.93	88.77	11.36	2 192.10	81.053	78 347.31	913.38
4	BIGZUO	Cr, Ag	1 409.6	20.99	308.30	21.72	0.90	87.17	10.90	1 905.25	77.23	127 504.2	907.91
5	ADATEG	Cu	1 492.37	13.34	238.78	27.34	0.92	84.16	11.33	1 981.44	78.491	105 265.8	907.69
6	ACOCOM	Cu	1 542.98	13.03	349.93	24.61	0.93	84.81	11.55	2 099.35	79.646	70 719.36	889.37
7	BAZGAM	Cu	826.332	24.24	126.53	42.80	0.96	86.31	10.20	1 789.28	72.982	133 974.7	861.58
8	ADUROI	Zn	1 275.01	17.12	402.57	25.69	0.85	86.23	11.27	1 928.36	74.513	99 317.53	857.35
9	5 024 922	Zn	1 437.25	15.86	295.00	20.52	0.89	91.79	10.96	1 587.45	60.812	118 193	710.29
10	AWUPAL	ln	1 726.10	9.24	348.55	18.42	0.91	93.03	12.03	1 705.51	60.829	76 045.77	683.66
11	AWUPOZ	Ln, Mn	1 742.35	9.26	353.42	18.42	0.91	88.13	11.88	1 618.00	59.264	74 275.17	664.68
12	ADASAB	Ln, Fe	1 752.40	9.28	362.43	18.45	0.89	90.22	11.83	1 517.56	55.201	72 503.02	617.78
13	ADASAB	Cu	1 659.76	13.49	343.94	22.82	0.88	82.67	10.90	1 327.33	52.165	92 727.46	596.63
14	AVAKAL	Cu	1 282.08	12.58	207.42	26.25	0.94	93.21	11.49	1 286.11	49.602	121 273.6	580.63
15	BIHBAX	Cr, Ag	1 476.98	17.56	321.68	18.57	0.90	87.41	10.63	1 222.74	49.385	126 404.3	580.14

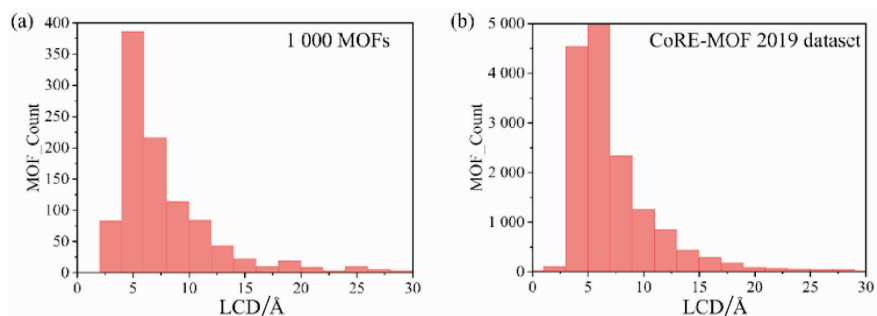


图 S1 (a) 1 000 种 MOFs 和 (b) CoRE-MOF 2019 数据库的 LCD 分布直方图

Figure S1 Histogram of LCD distribution of (a) 1 000 MOFs and (b) CoRE-MOF 2019 dataset