Chemical
Science

ROYAL SOCIETY
OF CHEMISTRY

PERSPECTIVE

View Article Online
View Journal

Check for updates

Cite this: DOI: 10.1039/d6sc02765b

All publication charges for this article have been paid for by the Royal Society of Chemistry

Received 3rd April 2026
Accepted 4th June 2026

DOI: 10.1039/d6sc02765b

rsc.li/chemical-science

Interrogating the synthetic likelihood of metal–organic frameworks: a digital discovery perspective

Xiaoyu Wu and Jianwen Jiang *

Digital discovery of metal–organic frameworks (MOFs) has advanced rapidly, driven by the tremendously large number of experimentally synthesized and computationally designed structures, high-throughput screening, and artificial intelligence. Yet a fundamental bottleneck remains: many hypothetical MOFs (hMOFs) may never reach a chemical laboratory. This gap has rendered the synthetic likelihood of MOFs a central challenge in translating digital MOF discovery into experimental synthesis and test. In this perspective, we provide an overview of recent progress in interrogating the synthetic likelihood of MOFs. First, thermodynamic analysis, focusing on free energy as a physically grounded metric for assessing synthesizability, is presented. Then, emerging data-driven heuristics, such as synthetic scores, classification models for synthesizability prediction, and machine-learning methods for predicting synthesis conditions directly from atomic structures, are discussed. Finally, we offer an outlook on future directions, including scalable free-energy calculations, synthesis-aware inverse design, and unified databases that incorporate both successful and failed synthesis attempts. It is highly anticipated that integrating these advances will transform MOF discovery from performance-driven screening into synthesis-informed design, thereby accelerating the realization of computationally designed structures in experiments.

1 Introduction

The 2025 Nobel Prize in Chemistry recognized Susumu Kitagawa, Richard Robson, and Omar Yaghi for their pioneering development of metal–organic frameworks (MOFs).[1] As crystalline nanoporous materials with remarkable tunability in topology and functionality, MOFs have shown great potential in a wide variety of applications. Along with over 120 000 experimentally synthesized MOFs and MOF-like structures,[2,3] millions of hypothetical MOFs (hMOFs) have been computationally designed.[4–7] More recently, generative artificial intelligence has attracted substantial attention as a powerful tool for generating hMOF structures, either conditionally or unconditionally, significantly expanding the accessible design space of MOFs.[8–10] These hMOFs were computationally screened for specific applications (*e.g.*, CO_2 capture, H_2 storage and H_2 production),[11–13] and top-performing hMOFs were identified with the implicit assumption that they would be synthesized.[14] Without interrogating synthetic likelihood, however, such computational screening risks to produce synthetically implausible MOFs that are excellent on a plot but not in a chemical laboratory.[15]

It is of central importance to develop reliable measures for the synthetic likelihood of MOFs, thereby bridging the divide between digital discovery and experimental realization. In this perspective, we present the recent efforts on reliable and actionable approaches for evaluating the synthesizability of MOFs. Our objective is to address the key translation gap in digital discovery of MOFs, turn the design from "performance optimization" into "performance optimization subject to synthesizability", and promote closer integration between *in silico* design and actual experimental synthesis.

We adopt synthetic likelihood here as an operational metric to describe the propensity of a MOF that can be synthesized experimentally. This notion is intentionally broader than a purely thermodynamic definition of stability. While thermodynamic competitiveness provides a necessary physical foundation, the synthetic likelihood of a MOF is also governed by chemical validity, pathway dependence of crystallization, and kinetic accessibility.[16] In particular, crystallization can be controlled by both thermodynamic and kinetic factors, implying that the lowest free-energy polymorph is not necessarily an experimentally producible structure under a given synthetic route.[17] Consequently, synthetic likelihood should not be treated as a standalone quantitative descriptor, but rather as a set of progressively enforceable constraints that can be incorporated into a digital discovery workflow.

At the most fundamental level, a MOF must satisfy chemical validity, including reasonable oxidation states and coordination environments, as well as overall charge balance.[18] A structure that violates these basic chemical rules undermines the

---

Department of Chemical and Biomolecular Engineering, National University of Singapore, 117576, Singapore. E-mail: chejj@nus.edu.sg

reliability of downstream simulation or prediction.[15,19,20] In addition, database-level inconsistency and differences in structural curation may propagate into systematic uncertainties in large-scale simulation results.[21] Once chemical plausibility is ensured, thermodynamic factors, particularly the free energy of the MOF compared with that of experimentally realized structures, provide physically grounded criteria for evaluating thermodynamic competitiveness.[22–27] Beyond physics-based criteria, data-driven heuristics offer complementary routes for estimating synthetic likelihood. In this context, machine-learning models trained on known experimental MOFs can capture statistical patterns in the chemical space of MOFs that correlate with experimental realizability and synthesis conditions.[28,29]

Importantly, synthetic likelihood is multidimensional. A number of factors such as activation stability, hydrolytic or thermal robustness, and mechanical stability come into a complex interplay in determining whether computationally designed MOFs can be synthesized and practically utilized.[6,30] In this perspective, we discuss two major approaches for quantifying synthetic likelihood in MOF discovery: thermodynamic stability analysis and data-driven heuristics.

2 Thermodynamic stability analysis

As a comparative route for assessing synthetic likelihood, thermodynamic stability determines whether a structure is energetically competitive relative to experimentally realizable structures. A key step in the thermodynamics-informed synthetic likelihood is based on free energy. Compared with inorganic solids, however, evaluating the free energies of MOFs presents additional challenges. This is because MOFs exhibit enormous diversity in metal nodes, coordination environments, linker chemistries, and structural topologies, making direct comparison of free energies across chemically distinct MOFs less meaningful. Instead, synthetic likelihood is often evaluated through relative free energies, which compare the energetic competitiveness of a structure within chemically comparable structures (*i.e.*, in similar metal–organic coordination environments).[22] Another important consideration arises from isomorphism, where multiple MOFs may share identical chemical composition but differ in structural topology.[31–33] In such cases, several structures may satisfy general thermodynamic criteria, yet synthesis typically favors the isomorph with the lowest free energy. Nevertheless, it should be noted that thermodynamic stability alone does not fully determine synthesizability, because crystallization can also be governed by kinetic effects, that is, metastable MOFs may be experimentally accessible under kinetic control.[34]

2.1 Free energy calculation

Among available approaches, the Frenkel–Ladd (FL) thermodynamic integration method[35] provides a formally exact route for calculating the Helmholtz free energy of a crystalline solid at a specified cell volume (Fig. 1). In this method, the free energy difference between a target crystal (*e.g.*, a MOF) and a reference Einstein crystal is obtained by integrating along a coupling parameter λ that continuously transforms one Hamiltonian into the other. This λ-dependent Hamiltonian $H(\lambda)$ is defined as:

$$H(\lambda) = (1 - \lambda)H_\mathrm{i} + \lambda H_\mathrm{f} \tag{1}$$

where H_i is the Hamiltonian of the Einstein crystal and H_f is the Hamiltonian of a MOF crystal described by an interatomic force field or electronic-structure method. The free energy difference between the two crystals ΔF is obtained through thermodynamic integration over λ between 0 and 1. The absolute free energy (F_FL) of the MOF can be evaluated from the following equation:

$$F_\mathrm{FL} = F_\mathrm{EC} + \Delta F = 3k_\mathrm{B}T\sum_{i=1}^{N}\ln\left(\frac{\hbar\omega_i}{k_\mathrm{B}T}\right) + \Delta F \tag{2}$$

where N is the number of framework atoms, k_B is the Boltzmann constant, $\hbar$ is the reduced Planck constant, T is the temperature under consideration, and ω_i is the harmonic frequency of atom i determined from its mass and spring constant. The spring constant can be computed from the mean-squared displacement of atoms *via* the equipartition theorem.

The FL method is most commonly implemented using classical molecular dynamics (MD) simulation with an empirical force field, because extensive samplings along the coupling pathway are required. In principle, *ab initio* molecular dynamics (AIMD) simulation can also be applied in the FL method, but the computational cost typically restricts such applications to relatively small systems (*e.g.*, <100 atoms). As an alternative, the Quasi-Harmonic Approximation (QHA) estimates the vibrational contribution to the Helmholtz free energy of a crystal from phonon spectra at a given volume and, when evaluated over different volumes, can be used to estimate the Gibbs free energy under a constant-pressure condition,[36]

$$F_\mathrm{QHA} = U + \frac{1}{2}\int_0^{\nu_\mathrm{max}} h\nu D(\nu)\mathrm{d}\nu + k_\mathrm{B}T\int_0^{\nu_\mathrm{max}} D(\nu)\ln\left[1 - \exp\left(-\frac{h\nu}{k_\mathrm{B}T}\right)\right]\mathrm{d}\nu \tag{3}$$

where U is the potential energy of an equilibrium structure, $D(\nu)$ is the phonon density of states, and ν denotes the vibrational frequency. The phonon density of states may be obtained from either lattice dynamics calculations based on density-functional theory (DFT) or velocity autocorrelation functions derived from an MD trajectory. Compared with the FL method, the QHA neglects strong anharmonic effects but can be evaluated using significantly shorter simulations or harmonic phonon calculations. As a result, it provides a practical approximation for

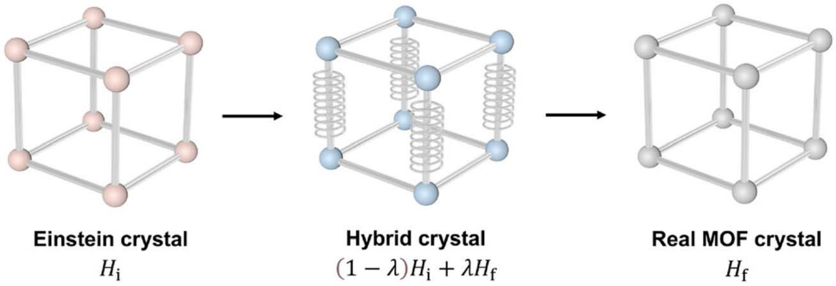

Fig. 1 Schematic illustration of the Frenkel–Ladd thermodynamic integration method. Pink system: Einstein crystal with H_i; blue: a hybrid crystal with $H(\lambda) = (1 - \lambda)H_i + \lambda H_f$; gray: a real MOF crystal with H_f. (Reproduced with permission from ref. 26, John Wiley and Sons, copyright 2026.)

estimating vibrational free energies across a large collection of MOFs.

An example based on thermodynamic stability to predict synthetic likelihood was illustrated for iron–sulfur MOFs assembled from Fe_4S_4 cluster and 1,4-benzenedithiolate (BDT) linker.[24] Specifically, a library of hundreds of hypothetical Fe_4S_4-BDT MOFs with different charge-balancing cations was constructed. Based on the FL method, their free energies were calculated, revealing a clear thermodynamic hierarchy among competing structures. For all five examined cations (TEA, TPA, MPA, MNP, and TPP), 1D structures were found to consistently possess lower free energies than their 2D and 3D counterparts. This study provides a direct thermodynamic basis for synthesizability, *i.e.*, an experimentally realizable structure occupies the lowest point in the free-energy landscape under a given chemical composition and environmental condition. Recently, we also applied the FL method to investigate how F_{FL} correlates with reticular chemistry in 10 556 computational-ready experimental (CoRE) MOFs,[37] thereby linking thermodynamic stability to synthetic likelihood.[26] As presented in Fig. 2, MOFs with more constricted cavities display both a higher median F_{FL} and a broader distribution of F_{FL}; however, those with larger pore volumes generally fall within a lower F_{FL} regime. In terms of topology, MOFs with **srs** topology tend to show a lower median F_{FL}. Due to a well-matched metal–linker coordination environment and minimal lattice strain, Zn- and Mg-MOFs are likewise skewed toward lower F_{FL}. Nevertheless, Eu- and In-MOFs often exhibit higher F_{FL} because of greater structural distortion and less favorable packing. Additionally, linker chemistry also contributes, with nitrogen-containing moieties exhibiting the strongest negative correlation with free energy. Our results suggest that synthetic likelihood is governed synergistically by the pore size and volume, topology, metal-node chemistry, as well as linker environment, rather than by a single factor alone.

2.2 Relative free energy

The chemical space of MOFs is extremely vast, comprising a great variety of metal nodes with different coordination environments, as well as numerous types of organic linkers. Consequently, absolute free energies cannot be directly compared across diverse frameworks. As shown in Fig. 3, MOFs with different metal nodes (*e.g.*, Cu, Zn, Cr, and Zr) occupy distinct regions in the free-energy landscape.[22] This separation arises primarily from differences in the intrinsic strain energies of metal nodes and the ways these environments are described (*e.g.*, by interatomic force fields). A metal node with a higher intrinsic strain or a coordination complexity tends to produce a systematically higher free energy.

To circumvent this issue, Anderson and Gómez-Gualdrón introduced a relative free energy $\Delta_{LM}F$.[22] This metric can be obtained by subtracting a metal node-specific linear model of free energy f from F_{FL}:

$$\Delta_{LM}F = F_{FL} - f = F_{FL} - a \times \left(\frac{N_m}{N_{m+n}}\right) - b \qquad (4)$$

where N_m and N_{m+n} are the number of metal and total atoms, respectively, in a MOF unit cell, and the coefficients a and b are taken from the specific confidence interval of the linear model f for each metal-node type. By removing the systematic energetic contribution associated with each metal-node type, $\Delta_{LM}F$ effectively normalizes the free-energy landscape, mitigating force-field artifacts and enabling consistent synthetic likelihood comparison across chemically diverse MOFs.

The above idea was adopted in our recent study on high-throughput screening of hMOFs for CO_2 capture.[23] Fig. 4a shows the free energies *versus* the percentage of metal atoms for 148 hMOFs, which were extracted from the *ab initio* REPEAT charge (ARC) MOF database[38] and predicted with high CO_2 capture performance. To determine the synthetic likelihood of these hMOFs, the free energy (F) values of 79 CoRE MOFs[37] with different metal-node types (pillared Cu-paddlewheel, pillared Zn-paddlewheel, Zn_4O, and V_3O_3) were calculated and used as benchmarks (Fig. 4b). For each metal-node type, the F values of CoRE MOFs were fitted against the percentage of metal atoms, acting as a reference line. By subtracting the reference line (F_{LM}) from the absolute F, the relative free energies $\Delta_{LM}F$ were obtained. As shown in Fig. 4c, all the 79 CoRE MOFs were capped within $\Delta_{LM}F$ of ~4.2 kJ mol^{-1}, which was taken as the upper bound for thermodynamic stability. Thereafter, 41 hMOFs among 148 were identified to possess $\Delta_{LM}F$ exceeding the upper bound, being considered thermodynamically unstable and unlikely to be synthesized.

A complementary thermodynamic metric was recently introduced by Rosen and co-workers, who constructed convex-hull phase diagrams for over 20 000 MOFs and coordination

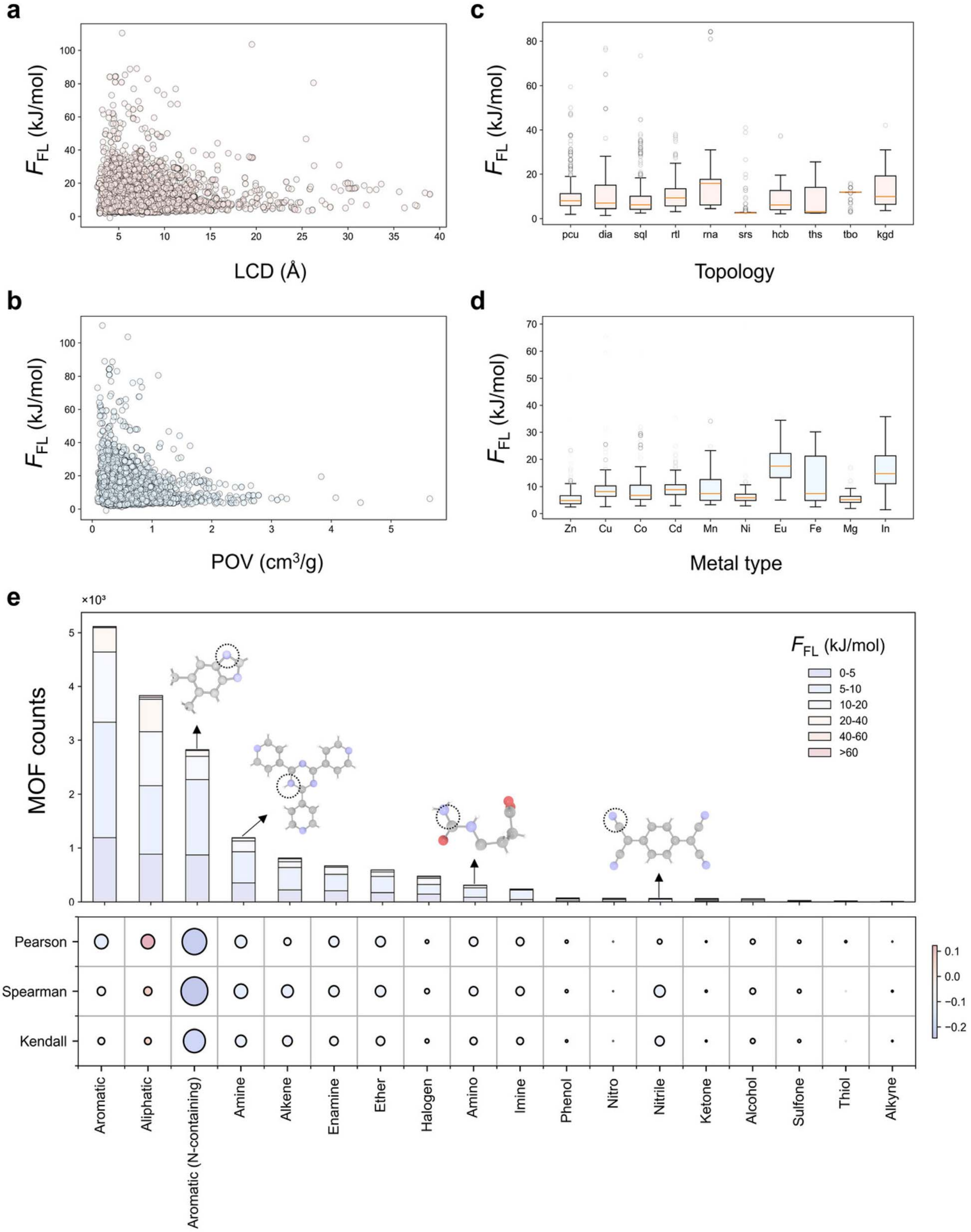

Fig. 2 Relationships between F_{FL} and (a) largest cavity diameter (LCD), (b) probe-centred pore volume (POV), (c) topology, and (d) metal type in 10 556 CoRE MOFs. (e) Linker-dependence and correlation coefficients (Kendall, Spearman, and Pearson) for F_{FL}. Representative linkers are shown for substructures with the most negative correlations. (Reproduced with permission from ref. 26, John Wiley and Sons, copyright 2026.)

polymers in the QMOF-Thermo Database. The energy above hull was used to quantify the metastability of MOFs with respect to phase transition or decomposition into competing materials, and most MOFs were predicted to be thermodynamically metastable.[39] While the energy above hull can serve as a useful synthesizability metric for filtering newly designed MOFs, one

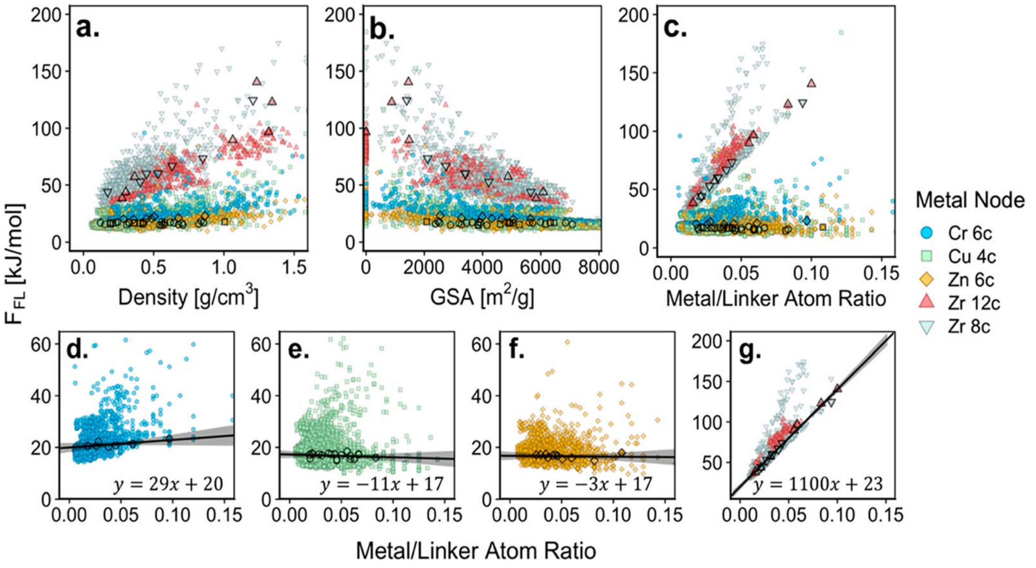

Fig. 3 F_{FL} *versus* (a) density, (b) gravimetric surface area (GSA), and (c) metal/linker atom ratio for MOFs with different metal-node types; (d–g) breakdown of F_{FL} *versus* metal/linker atom ratio. Black points denote synthesized MOFs, black lines are linear fits for synthesized MOFs with different metal-node types, and shaded regions correspond to 95% confidence interval of linear fit. (Reproduced with permission from ref. 22, American Chemical Society, copyright 2020.)

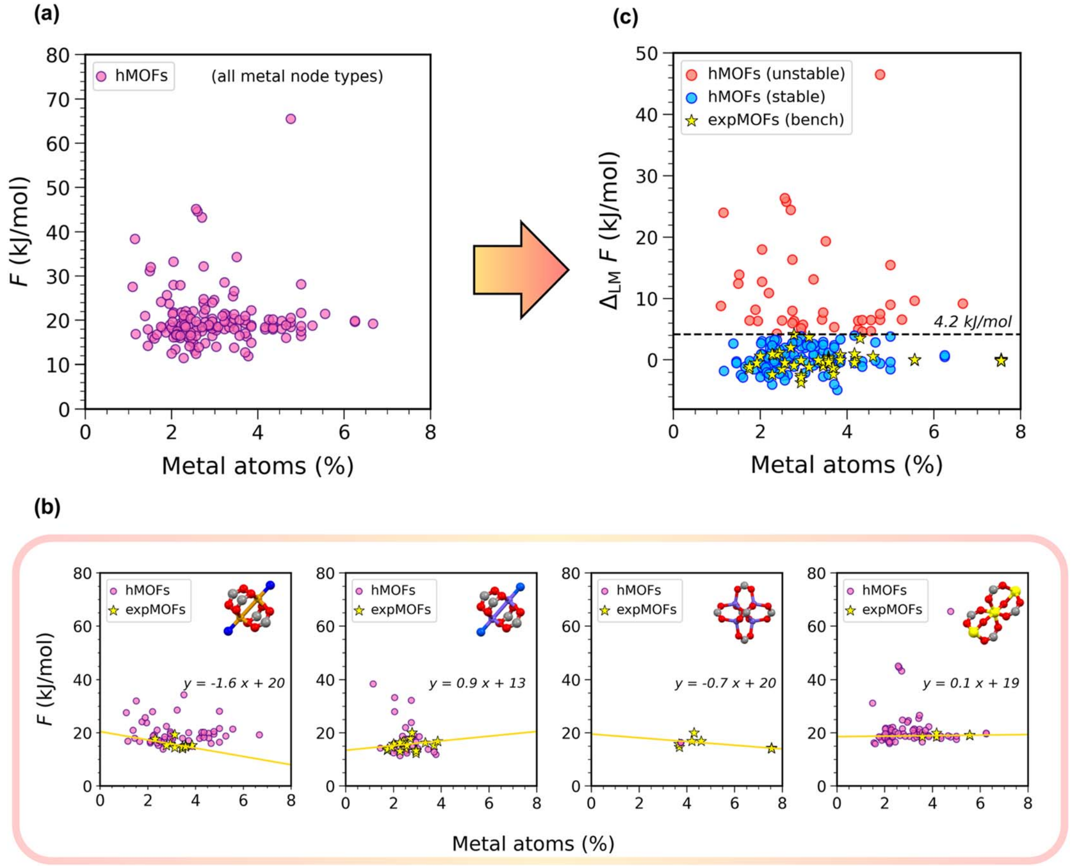

Fig. 4 (a) Free energies of 148 hMOFs with different metal-node types *versus* the percentage of metal atoms. (b) Free energies of hMOFs and experimental MOFs with different metal-node types. (c) Relative free energies $\Delta_{LM}F$ of 148 hMOFs and selected benchmarking experimental MOFs, where hMOFs with $\Delta_{LM}F > 4.2$ kJ mol^{-1} were considered not synthesizable. (Reproduced with permission from ref. 23, Springer Nature, copyright 2023.)

should be cautious because it may not explicitly capture kinetic factors or synthesis conditions.

2.3 Isomorph curation

A further point in thermodynamic stability analysis for synthetic likelihood is the presence of isomorphic MOFs, where identical building blocks (metal nodes and organic linkers) assemble into multiple structural topologies.[31–33] Large hMOF databases frequently contain such isomorphic structures and several of them may also satisfy the thermodynamic stability criterion based on relative free energies. Consequently, evaluating thermodynamic competition among isomorphs is important for refining the set of candidate structures. By grouping structures with identical chemical compositions and comparing their relative free energies, unstable isomorphs are systematically removed. Such isomorph curation can effectively reduce redundancy and help produce a more confident list of synthesizable hMOF candidates, ensuring that digital discovery prioritizes the most thermodynamically plausible structures for experimental synthesis.

Crystal structure prediction (CSP) provides a closely related route for connecting building block chemistry with synthetic likelihood. Instead of only filtering pre-existing MOFs, CSP aims to predict which structure is most likely to assemble from given metal nodes and organic linkers.[40] *Ab initio* CSP studies have demonstrated that experimentally realized MOFs could be recovered from underlying building blocks and that predicted low-energy structures could guide the discovery of new MOFs.[33] In this sense, CSP complements the relative free-energy analysis and isomorph curation by explicitly addressing which structure is most likely to be realized for a given chemical composition.

3 Data-driven heuristics

Beyond physics-based stability analysis, data-driven heuristics have emerged as a powerful route to evaluate the synthetic likelihood of MOFs by learning directly from accumulated literature records of experimentally synthesized MOFs. In principle, data-driven heuristics can capture the statistical synthesizability information that is difficult to encode explicitly by stability analysis, especially as the chemical space of synthesized MOFs is biased.[41] Recent progress in predictive models has further expanded this paradigm, enabling not only the classification of likely synthesizable MOFs, but also the prediction of synthesis conditions.[28,42–44] These advances suggest that synthesizability prediction of MOFs is evolving from a purely stability-centered question into a broader informatics problem. We will discuss three representative data-driven heuristics: synthetic scores, direct prediction of synthesizability, and prediction of synthesis conditions.

3.1 Synthetic scores

The concept of synthetic accessibility has long been explored in drug discovery, where prioritizing compounds that are both functional and synthetically tractable is essential for efficient chemical development.[45] Among the most widely used metrics are the synthetic accessibility (SA) score[46] and synthetic complexity (SC) score,[47] both proposed to estimate the difficulty of synthesizing organic molecules. The SA score evaluates synthetic feasibility by combining fragment contributions derived from large databases of known molecules with penalties associated with structural complexity such as rings, stereochemistry, and molecular size. In contrast, the SC score adopts a data-driven strategy, learning synthetic complexity directly from a large reaction dataset based on the principle that reaction products tend to be more complex than their precursors. Both SA and SC scores provide rapid heuristics for filtering MOFs that are likely to be experimentally accessible. However, these scores remain inherently molecule-centered and therefore capture only part of the synthesizability information for MOFs. In particular, they primarily evaluate linker chemistry while neglecting the influence of metal coordination environment and structural topology, which also play critical roles in determining whether a MOF can be realized experimentally.[48]

As exemplified in an inverse design study (Fig. 5) aimed at shifting the chemical space of MOFs toward improved property distributions, the linkers optimized for maximum heat capacity ($c_p^{\max}$) exhibit a shift toward lower SA scores, typically suggesting favorable synthetic accessibility.[49] However, many of these linkers display significantly higher SC scores, indicating large, highly elaborate, and often unrealistic linker architectures. In practice, several linkers with lower SA but higher SC values are unlikely to be suitable for MOF construction. To address this limitation, a constraint of SC < 4 was imposed during the inverse design process and this constrained optimization produced linkers that more closely resemble experimental structures in CoRE-MOFs while maintaining a similarly improved heat capacity distribution.[49] This study illustrates that, in the design of linkers for MOFs, the SC score can complement the SA score to provide a more discriminating indicator of "linker-like" realism, and incorporating such a constraint may guide generative models toward candidates that are not only high-performing but also more chemically synthesizable.

3.2 Prediction of synthesizability

While machine learning (ML) has been widely applied to predict materials properties, direct prediction of MOF synthesizability is limited, largely because reliable negative data are scarce. In the field of MOFs, unsuccessful syntheses are rarely reported, and the absence of reported structures does not rigorously prove that targeted MOFs are unsynthesizable.[50–53] Such scarcity of reliable negative data limits conventional supervised models for predicting synthesizability.

Fast, non-ML algorithmic methods have therefore recently emerged as practical tools for evaluating the synthetic likelihood of MOFs. A representative example is MOFSynth,[54] which assesses synthesizability by comparing the energy and geometry of linker conformation in a MOF with those of the corresponding isolated linkers. This strategy provides a physically interpretable and computationally efficient measure of whether the linker conformation required by a target MOF is chemically

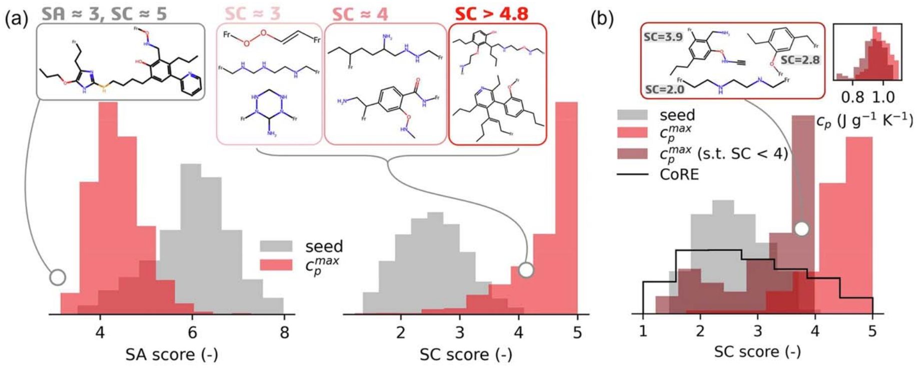

Fig. 5 Synthetic accessibility and synthetic complexity of MOFs optimized for maximum specific heat capacity. (a) Distributions of SA and SC scores for the inverse designed and seed sets. While the dreamed set shifts toward lower SA scores, its SC scores are shifted higher, indicating increased linker complexity. (b) Constrained (with SC < 4) and unconstrained optimization from 100 seed MOFs. (Reproduced with permission from ref. 49, Springer Nature, copyright 2025.)

reasonable. More recently, MOFSynth-ADV[55] extended this concept into an open-source engine for automated evaluation of MOF synthesizability. Such rule- or descriptor-based algorithmic methods are particularly valuable because they do not require a large set of failed synthesis data, which remain scarce in the literature.

The scarcity of reliable negative data has motivated the use of one-class classification (OCC),[56] in which an ML model is trained only on known positive examples and then used to identify candidates that are most consistent with the distribution of experimentally realized materials. Toward this end, a complementary evaluation strategy termed the maximum fraction difference (MFD) method was proposed.[57] It quantifies how effectively a model separates the score distribution of known positive samples from that of an unlabeled query dataset. Rather than relying on standard classification metrics that require both positive and negative labels, the MFD identifies a threshold at which the difference between the positive fraction of a ground-truth dataset and that of a query dataset is maximized, thereby providing both an assessment criterion and a practical decision threshold for candidate prioritization. Based on the MFD, a DeepSVDD model was shown to outperform several traditional OCC approaches in predicting MOF synthesizability from metal–linker combinations (Fig. 6).[57] More broadly, this work highlights that the prediction of synthesizability can serve as an important filter alongside property screening, enabling the prioritization of MOFs that are not only computationally promising but also more likely to be experimentally synthesized.

3.3 Prediction of synthesis conditions

While thermodynamic metrics and data-driven classifiers act as important indicators of whether a MOF is likely to be experimentally realizable, they do not directly answer a practical question faced by experimental chemists: how should the MOF be actually synthesized?[58] In other words, synthetic likelihood ultimately depends not only on whether a structure is stable or statistically plausible, but also on whether suitable synthesis conditions can be identified. Traditionally, these conditions are determined through trial-and-error guided by chemical intuition and analogy to previously reported systems. However, with the rapidly growing availability of structural databases and synthesis reports, recent studies have started to explore whether ML can directly infer synthesis conditions from the structure of a targeted MOF.[28,29,59–61] Such exploration represents a natural extension from the prediction of synthetic likelihood, moving beyond assessing whether the targeted MOF is synthesizable toward predicting how it can be synthesized.

A representative example is the development of the SynMOF database, where synthesis conditions were curated from the literature by automatic data mining and linked with crystallographic information files.[28] As illustrated in Fig. 7, the SynMOF database integrates key synthesis conditions such as solvent, temperature, additive, and reaction time with building blocks (metal type and organic linker). ML models trained on this database were able to learn correlations between synthesis conditions and framework chemistry, enabling the prediction of plausible experimental settings for previously unseen MOFs. Interestingly, the trained ML models were shown to outperform human experts in MOF synthesis, highlighting the complexity of such synthesis and the potential of ML in facilitating experimental realization of MOFs. Nevertheless, the transferability of such ML models should be interpreted with caution. Because literature-derived databases primarily encode known chemical and topological patterns, their predictions are expected to be most reliable for MOFs within or close to the training domain. For genuinely first-of-its-kind topologies or unusual metal–linker combinations without close experimental analogues, predicted conditions should be regarded as plausible starting points rather than definitive synthesis recipes.

4 Outlook

We have discussed the recent progress in interrogating the synthetic likelihood of MOFs through thermodynamic stability analysis and data-driven heuristics. Despite these advances, several key challenges remain before digital MOF discovery can

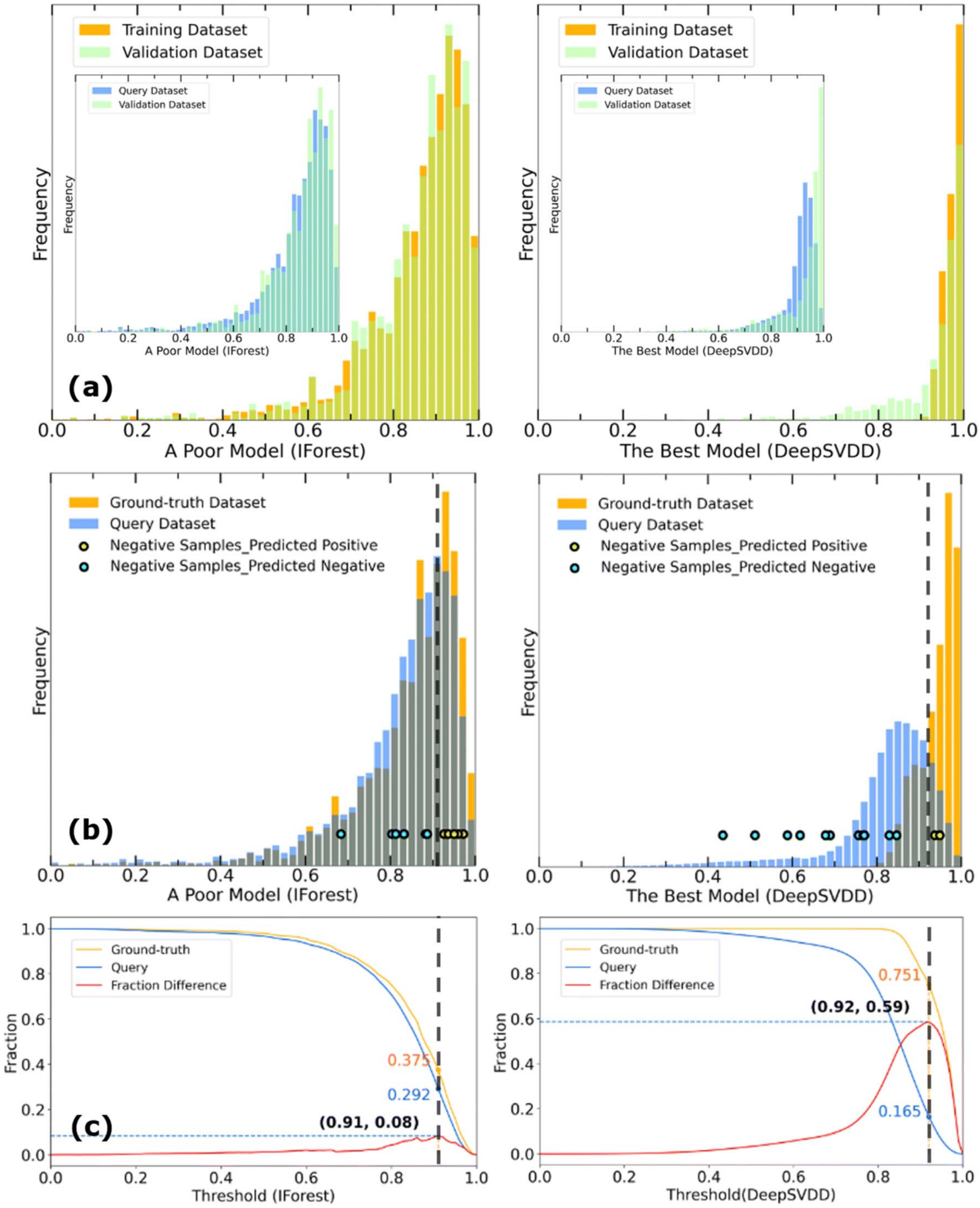

Fig. 6 Comparison of a poor model (isolation forest with 6 metal features and 1613-dimensional Mordred linker descriptors, MFD = 0.08) and the best model (DeepSVDD with 205 metal features and 2048-dimensional ECFP linker features, MFD = 0.59). (a) Score distributions of the positive validation dataset against the training dataset (main panel) and query dataset (inset). (b) Score distributions of the ground-truth and query datasets, with predictions for 14 true negatives; blue and yellow dots denote negative and positive predictions, respectively. (c) Positive fraction distributions of the ground-truth and query datasets, and their difference, *versus* the normalized score. The dashed line marks the optimal threshold. (Reproduced with permission from ref. 57, Royal Chemistry Society, copyright 2024.)

reliably translate computational design into experimental realization. These challenges also provide new and exciting opportunities for further advancement in this vibrant field.

First, an important direction lies in balancing accuracy and computational efficiency in free-energy calculations. Methods such as FL thermodynamic integration provide rigorous estimates of free energies but are computationally expensive for large-scale screening of MOFs. By contrast, approximate approaches such as the QHA offer improved efficiency but remain based primarily on vibrational degrees of freedom;

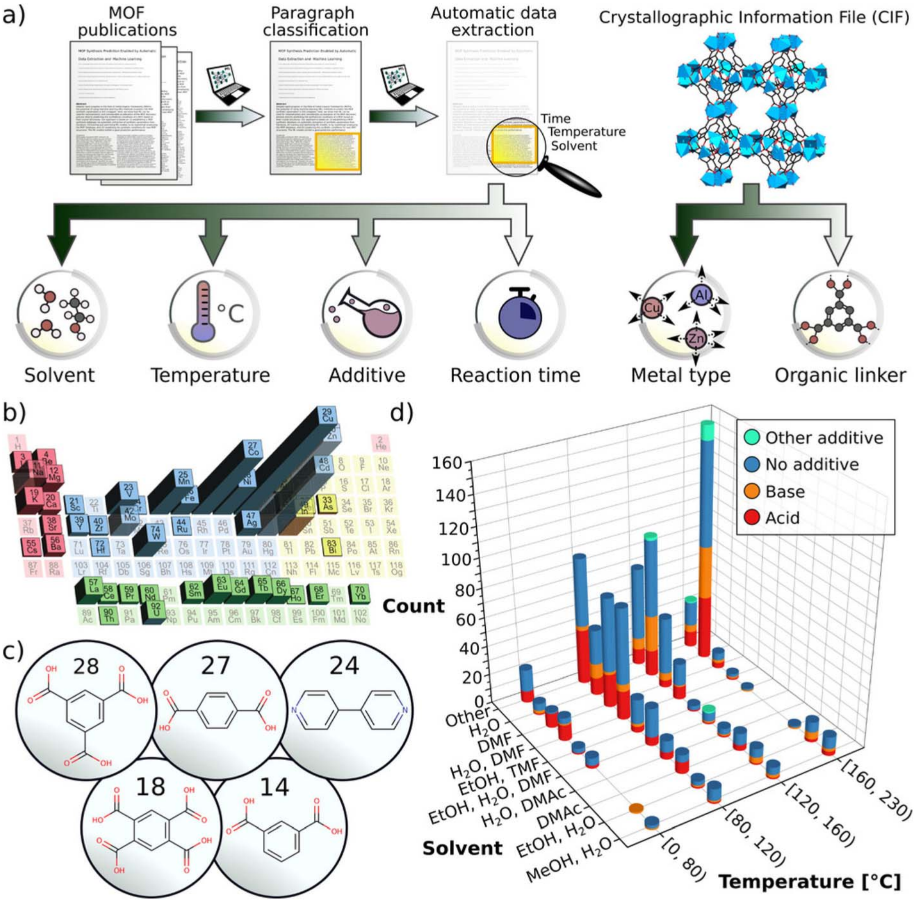

Fig. 7 (a) Data mining pipeline and content of the SynMOF database. (b) Statistics on the most common metals. (c) Structures and occurrences of the most common organic linkers. (d) 3D graph exhibiting correlation between solvent type, additive, and temperature. (Reproduced with permission from ref. 28, John Wiley and Sons, copyright 2022.)

therefore, they may neglect strong anharmonic effects, rotational or torsional motions, and large-amplitude structural flexibility that are common in porous materials. Force-field artifacts are usually removed through human interpretation in relative free energies on a node-wise basis for predicting synthetic likelihood. A promising path forward is the development of ML potentials (MLPs), which are capable of reproducing DFT accuracy while maintaining computational costs comparable to classical force fields. By enabling long-time scale MD simulations with near *ab initio* fidelity, MLPs can significantly expand the scope of free-energy-based prediction of synthetic likelihood across the large chemical space of MOFs.

Second, synthetic likelihood should be integrated directly into inverse design workflows. Current generative models and bottom-up assembly approaches can efficiently construct vast libraries of MOFs, yet many generated structures remain synthetically unrealistic due to impractical linkers, unstable coordination environments, or incompatible topologies. Incorporating synthetic accessibility constraints, such as linker synthetic complexity through SC scores or thermodynamic stability criteria, during a generative process can help steer inverse design toward an experimentally feasible chemical space of MOFs. Furthermore, emerging large language models trained on chemical annotations provide a complementary strategy for estimating synthetic plausibility based on partial structural information like metal nodes and organic linkers, even before a full crystal structure is constructed.

Finally, progress in data-driven synthesizability prediction is currently limited by the lack of negative or failed synthesis data. Most published datasets contain only successfully synthesized MOFs, making it difficult for ML models to distinguish between genuinely unsynthesizable structures and those that have not yet been attempted. Therefore, future endeavors should aim to develop unified synthesis databases that record both successful and failed experiments, including detailed synthesis conditions, precursor choices, and other underlying information. Such databases would enable more robust learning of key factors governing MOF synthesis and provide a more realistic foundation for predicting synthetic likelihood.

Author contributions

XW conceptualized the project and wrote the manuscript. All authors reviewed and edited the manuscript. JJ acquired the funding and supervised the project.

Conflicts of interest

There are no conflicts of interest to declare.

Data availability

As this is a perspective article, no original data are associated.

Acknowledgements

We gratefully acknowledge the National Research Foundation Singapore (NRF-CRP26-2021RS-0002) and A*STAR LCER-FI (LCERFI01-0015 U2102d2004 and LCERFI01-0033 U2102d2006) for financial support, and the National University of Singapore (CFP01-CF-077 and CFP03-CF-016) and National Supercomputing Centre Singapore for computational resources.

References

1 Press Release, *Nobel Prize in Chemistry 2025, NobelPrize.org*, https://www.nobelprize.org/prizes/chemistry/2025/press-release/, accessed 10 Mar. 2026.

2 L. T. Glasby, J. L. Cordiner, J. C. Cole and P. Z. Moghadam, Topological Characterization of Metal–Organic Frameworks: A Perspective, *Chem. Mater.*, 2024, **36**, 9013–9030.

3 G. Zhao, L. M. Brabson, S. Chheda, J. Huang, H. Kim, K. Liu, K. Mochida, T. D. Pham, Prerna, G. G. Terrones, S. Yoon, L. Zoubritzky, F.-X. Coudert, M. Haranczyk, H. J. Kulik, S. M. Moosavi, D. S. Sholl, J. I. Siepmann, R. Q. Snurr and Y. G. Chung, CoRE MOF DB: A curated experimental metal-organic framework database with machine-learned properties for integrated material-process screening, *Matter*, 2025, **8**, 102140.

4 C. E. Wilmer, M. Leaf, C. Y. Lee, O. K. Farha, B. G. Hauser, J. T. Hupp and R. Q. Snurr, Large-scale screening of hypothetical metal–organic frameworks, *Nat. Chem.*, 2011, **4**, 83–89.

5 P. G. Boyd, A. Chidambaram, E. García-Díez, C. P. Ireland, T. D. Daff, R. Bounds, A. Gładysiak, P. Schouwink, S. M. Moosavi, M. M. Maroto-Valer, J. A. Reimer, J. A. R. Navarro, T. K. Woo, S. Garcia, K. C. Stylianou and B. Smit, Data-driven design of metal–organic frameworks for wet flue gas CO_2 capture, *Nature*, 2019, **576**, 253–256.

6 A. Nandy, S. Yue, C. Oh, C. Duan, G. G. Terrones, Y. G. Chung and H. J. Kulik, A database of ultrastable MOFs reassembled from stable fragments with machine learning models, *Matter*, 2023, **6**, 1585–1603.

7 X. Wu and J. Jiang, Precision-engineered metal–organic frameworks: fine-tuning reverse topological structure prediction and design, *Chem. Sci.*, 2024, **15**, 16467–16479.

8 Z. Yao, B. Sánchez-Lengeling, N. S. Bobbitt, B. J. Bucior, S. G. H. Kumar, S. P. Collins, T. Burns, T. K. Woo, O. K. Farha, R. Q. Snurr and A. Aspuru-Guzik, Inverse design of nanoporous crystalline reticular materials with deep generative models, *Nat. Mach. Intell.*, 2021, **3**, 76–86.

9 H. Park, S. Majumdar, X. Zhang, J. Kim and B. Smit, Inverse design of metal–organic frameworks for direct air capture of CO_2 *via* deep reinforcement learning, *Digital Discovery*, 2024, **3**, 728–741.

10 H. Park, X. Yan, R. Zhu, E. A. Huerta, S. Chaudhuri, D. Cooper, I. Foster and E. Tajkhorshid, A generative artificial intelligence framework based on a molecular diffusion model for the design of metal-organic frameworks for carbon capture, *Commun. Chem.*, 2024, **7**, 1–18.

11 X. Wu, Q. Liu and J. Jiang, Discovering robust metal-organic frameworks with open copper sites for precombustion CO_2 capture: Data-efficient exploration and exploitation by active learning, *Chem. Eng. J.*, 2025, **521**, 167021.

12 W.-T. Kim, W.-G. Lee, H.-E. An, H. Furukawa, W. Jeong, S.-C. Kim, J. R. Long, S. Jeong and J.-H. Lee, Machine learning-assisted design of metal–organic frameworks for hydrogen storage: A high-throughput screening and experimental approach, *Chem. Eng. J.*, 2025, **507**, 160766.

13 X. Niu, Z. Zhang, X. Wu, Y. Liu, Y. Cui and J. Jiang, Machine learning guided discovery of water stable metal–organic frameworks for photocatalytic hydrogen production, *Chem. Sci.*, 2026, **17**, 5376–5386.

14 P. Z. Moghadam, Y. G. Chung and R. Q. Snurr, Progress toward the computational discovery of new metal–organic framework adsorbents for energy applications, *Nat. Energy*, 2024, **9**, 121–133.

15 A. J. White, M. Gibaldi, J. Burner, R. A. Mayo and T. K. Woo, High Structural Error Rates in "Computation-Ready" MOF Databases Discovered by Checking Metal Oxidation States, *J. Am. Chem. Soc.*, 2025, **147**, 17579–17583.

16 A. Li, R. Bueno-Perez, D. Madden and D. Fairen-Jimenez, From computational high-throughput screenings to the lab: taking metal–organic frameworks out of the computer, *Chem. Sci.*, 2022, **13**, 7990–8002.

17 A. K. Cheetham, G. Kieslich and H. H. M. Yeung, Thermodynamic and Kinetic Effects in the Crystallization of Metal–Organic Frameworks, *Acc. Chem. Res.*, 2018, **51**, 659–667.

18 M. Gibaldi, A. Kapeliukha, A. White, J. Luo, R. A. Mayo, J. Burner and T. K. Woo, MOSAEC-DB: a comprehensive database of experimental metal–organic frameworks with verified chemical accuracy suitable for molecular simulations, *Chem. Sci.*, 2025, **16**, 4085–4100.

19 X. Jin, K. M. Jablonka, E. Moubarak, Y. Li and B. Smit, MOFChecker: a package for validating and correcting metal–organic framework structures, *Digital Discovery*, 2025, **4**, 1560–1569.

20 G. Zhao, P. Zhao and Y. G. Chung, MOFClassifier: A Machine Learning Approach for Validating Computation-Ready Metal–Organic Frameworks, *J. Am. Chem. Soc.*, 2025, **147**, 33343–33349.

21 H. Daglar, H. C. Gulbalkan, G. Avci, G. O. Aksu, O. F. Altundal, C. Altintas, I. Erucar and S. Keskin, Effect of Metal–Organic Framework Database Selection on the Assessment of Gas Storage and Separation Potentials of MOFs, *Angew. Chem., Int. Ed.*, 2021, **60**, 7828–7837.

22 R. Anderson and D. A. Gómez-Gualdrón, Large-Scale Free Energy Calculations on a Computational Metal–Organic Frameworks Database: Toward Synthetic Likelihood Predictions, *Chem. Mater.*, 2020, **32**, 8106–8119.

23 S. A. Mohamed, D. Zhao and J. Jiang, Integrating stability metrics with high-throughput computational screening of metal–organic frameworks for CO_2 capture, *Commun. Mater.*, 2023, **4**, 1–10.

24 J. Mao, N. Jiang, A. Darù, A. S. Filatov, J. E. Burch, J. Hofmann, S. M. Vornholt, K. W. Chapman, J. S. Anderson and A. L. Ferguson, Structure and Synthesizability of Iron–Sulfur Metal–Organic Frameworks, *J. Am. Chem. Soc.*, 2025, **147**, 17651–17667.

25 A. Niyongabo Rubungo, F. Fajardo-Rojas, D. A. Gómez-Gualdrón and A. B. Dieng, Highly Accurate and Fast Prediction of MOF Free Energy *via* Machine Learning, *J. Am. Chem. Soc.*, 2025, **147**, 48035–48045.

26 X. Wu, R. Zheng, Q. Liu and J. Jiang, Digital Discovery of Synthesizable Metal–Organic Frameworks *via* Molecular Dynamics-Informed, High-Fidelity Deep Learning, *Adv. Funct. Mater.*, 2026, **36**, e74277.

27 K. Tolborg, J. Klarbring, A. M. Ganose and A. Walsh, Free energy predictions for crystal stability and synthesisability, *Digital Discovery*, 2022, **1**, 586–595.

28 Y. Luo, S. Bag, O. Zaremba, A. Cierpka, J. Andreo, S. Wuttke, P. Friederich and M. Tsotsalas, MOF Synthesis Prediction Enabled by Automatic Data Mining and Machine Learning, *Angew. Chem., Int. Ed.*, 2022, **61**, e202200242.

29 Y. Kang and J. Kim, ChatMOF: an artificial intelligence system for predicting and generating metal-organic frameworks using large language models, *Nat. Commun.*, 2024, **15**, 4705.

30 Z. Zhang, A. S. Palakkal, X. Wu, J. Jiang and Z. Jiang, Discovering Ultra-Stable Metal–Organic Frameworks for CO_2 Capture from A Wet Flue Gas: Integrating Machine Learning and Molecular Simulation, *Environ. Sci. Technol.*, 2025, **59**, 9123–9133.

31 S. Impeng, R. Cedeno, J. P. Dürholt, R. Schmid and S. Bureekaew, Computational Structure Prediction of (4,4)-Connected Copper Paddle-wheel-based MOFs: Influence of Ligand Functionalization on the Topological Preference, *Cryst. Growth Des.*, 2018, **18**, 2699–2706.

32 M. Arhangelskis, A. D. Katsenis, N. Novendra, Z. Akimbekov, D. Gandrath, J. M. Marrett, G. Ayoub, A. J. Morris, O. K. Farha, T. Friščić and A. Navrotsky, Theoretical Prediction and Experimental Evaluation of Topological Landscape and Thermodynamic Stability of a Fluorinated Zeolitic Imidazolate Framework, *Chem. Mater.*, 2019, **31**, 3777–3783.

33 Y. Xu, J. M. Marrett, H. M. Titi, J. P. Darby, A. J. Morris, T. Friščić and M. Arhangelskis, Experimentally Validated *Ab Initio* Crystal Structure Prediction of Novel Metal–Organic Framework Materials, *J. Am. Chem. Soc.*, 2023, **145**, 3515–3525.

34 A. J. Rieth, A. M. Wright and M. Dincă, Kinetic stability of metal–organic frameworks for corrosive and coordinating gas capture, *Nat. Rev. Mater.*, 2019, **4**, 708–725.

35 D. Frenkel and A. J. C. Ladd, New Monte Carlo method to compute the free energy of arbitrary solids. Application to the fcc and hcp phases of hard spheres, *J. Chem. Phys.*, 1984, **81**, 3188–3193.

36 M. Vasileiadis, *Calculation of the Free Energy of Crystalline Solids*, PhD thesis, Imperial College London, London, UK, 2013.

37 Y. G. Chung, E. Haldoupis, B. J. Bucior, M. Haranczyk, S. Lee, H. Zhang, K. D. Vogiatzis, M. Milisavljevic, S. Ling, J. S. Camp, B. Slater, J. I. Siepmann, D. S. Sholl and R. Q. Snurr, Advances, Updates, and Analytics for the Computation-Ready, Experimental Metal–Organic Framework Database: CoRE MOF 2019, *J. Chem. Eng. Data*, 2019, **64**, 5985–5998.

38 J. Burner, J. Luo, A. White, A. Mirmiran, O. Kwon, P. G. Boyd, S. Maley, M. Gibaldi, S. Simrod, V. Ogden and T. K. Woo, ARC-MOF: A Diverse Database of Metal-Organic Frameworks with DFT-Derived Partial Atomic Charges and Descriptors for Machine Learning, *Chem. Mater.*, 2023, **35**, 900–916.

39 B. Dallmann, A. Saha and A. S. Rosen, Predicting the Thermodynamic Limits of Metal–Organic Framework Metastability, *J. Am. Chem. Soc.*, 2026, **148**, 19487–19501.

40 J. P. Darby, M. Arhangelskis, A. D. Katsenis, J. M. Marrett, T. Friščić and A. J. Morris, Ab Initio Prediction of Metal-Organic Framework Structures, *Chem. Mater.*, 2020, **32**, 5835–5844.

41 Z. Song, S. Lu, M. Ju, Q. Zhou and J. Wang, Accurate prediction of synthesizability and precursors of 3D crystal structures *via* large language models, *Nat. Commun.*, 2025, **16**, 6530.

42 Z. Zheng, O. Zhang, C. Borgs, J. T. Chayes and O. M. Yaghi, ChatGPT Chemistry Assistant for Text Mining and the Prediction of MOF Synthesis, *J. Am. Chem. Soc.*, 2023, **145**, 18048–18062.

43 Z. Han, Y. Yang, J. Rushlow, J. Huo, Z. Liu, Y.-C. Hsu, R. Yin, M. Wang, R. Liang, K.-Y. Wang and H.-C. Zhou, Development of the design and synthesis of metal–organic frameworks (MOFs) – from large scale attempts, functional oriented modifications, to artificial intelligence (AI) predictions, *Chem. Soc. Rev.*, 2025, **54**, 367–395.

44 J. Zhang, J. Li, G. Zhao, Q. Wang, Y.-G. Guo and C. Yang, Mining Solid-State Electrolytes from Metal–Organic Framework Databases through Large Language Models and Representation Clustering, *J. Am. Chem. Soc.*, 2025, **147**, 40496–40506.

45 J. C. Baber and M. Feher, Predicting Synthetic Accessibility: Application in Drug Discovery and Development, *Mini-Rev. Med. Chem.*, 2004, **4**, 681–692.

46 P. Ertl and A. Schuffenhauer, Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions, *J. Cheminf.*, 2009, **1**, 8.

47 C. W. Coley, L. Rogers, W. H. Green and K. F. Jensen, SCScore: Synthetic Complexity Learned from a Reaction Corpus, *J. Chem. Inf. Model.*, 2018, **58**, 252–261.

48 S. M. Moosavi, A. Nandy, K. M. Jablonka, D. Ongari, J. P. Janet, P. G. Boyd, Y. Lee, B. Smit and H. J. Kulik, Understanding the diversity of the metal-organic framework ecosystem, *Nat. Commun.*, 2020, **11**, 1–10.

49 C. Cleeton and L. Sarkisov, Inverse design of metal-organic frameworks using deep dreaming approaches, *Nat. Commun.*, 2025, **16**, 4806.

50 D. Gleaves, N. Fu, E. M. Dilanga Siriwardane, Y. Zhao and J. Hu, Materials synthesizability and stability prediction using a semi-supervised teacher-student dual neural network, *Digital Discovery*, 2023, **2**, 377–391.

51 N. C. Frey, J. Wang, G. I. Vega Bellido, B. Anasori, Y. Gogotsi and V. B. Shenoy, Prediction of Synthesis of 2D Metal Carbides and Nitrides (MXenes) and Their Precursors with Positive and Unlabeled Machine Learning, *ACS Nano*, 2019, **13**, 3031–3041.

52 E. R. Antoniuk, G. Cheon, G. Wang, D. Bernstein, W. Cai and E. J. Reed, Predicting the synthesizability of crystalline inorganic materials from the data of known material compositions, *npj Comput. Mater.*, 2023, **9**, 155.

53 J. Jang, J. Noh, L. Zhou, G. H. Gu, J. M. Gregoire and Y. Jung, Synthesizability of materials stoichiometry using semi-supervised learning, *Matter*, 2024, **7**, 2294–2312.

54 C. G. Livas, P. N. Trikalitis and G. E. Froudakis, MOFSynth: A Computational Tool toward Synthetic Likelihood Predictions of MOFs, *J. Chem. Inf. Model.*, 2024, **64**, 8193–8200.

55 C. G. Livas, E. Klontzas, P. N. Trikalitis and G. E. Froudakis, MOFSynth-ADV: An Open-Source Engine for Synthesizability Evaluation of Metal–Organic Frameworks, *J. Chem. Inf. Model.*, 2026, **66**, 5040–5045.

56 A. Vriza, A. B. Canaj, R. Vismara, L. J. Kershaw Cook, T. D. Manning, M. W. Gaultois, P. A. Wood, V. Kurlin, N. Berry, M. S. Dyer and M. J. Rosseinsky, One class classification as a practical approach for accelerating π–π co-crystal discovery, *Chem. Sci.*, 2021, **12**, 1702–1719.

57 C. Zhang, D. Antypov, M. J. Rosseinsky and M. S. Dyer, Accelerating metal–organic framework discovery *via* synthesisability prediction: the MFD evaluation method for one-class classification models, *Digital Discovery*, 2024, **3**, 2509–2522.

58 D. Wonanke, A. Longa, A. Pankajakshan, L. Himanen, A. N. Ladines, J. A. Márquez, M. A. Addicoat, D. Crittenden, M. Scheidgen, P. Lio, S. Dehnen, C. Wöll and T. Heine, FAIR-MOFs: A Comprehensive Database for Accelerating the Discovery and Synthesis of Metal-Organic Frameworks, *ChemRxiv*, preprint, 2025, DOI: 10.26434/chemrxiv-2025-zjjdc.

59 T. M. Pruyn, A. Aswad, S. T. Khan, J. Huang, R. Black and S. M. Moosavi, MOF-ChemUnity: Literature-Informed Large Language Models for Metal–Organic Framework Research, *J. Am. Chem. Soc.*, 2025, **147**, 43474–43486.

60 L. T. Glasby, K. Gubsch, R. Bence, R. Oktavian, K. Isoko, S. M. Moosavi, J. L. Cordiner, J. C. Cole and P. Z. Moghadam, DigiMOF: A Database of Metal–Organic Framework Synthesis Information Generated *via* Text Mining, *Chem. Mater.*, 2023, **35**, 4510–4524.

61 L. Pilz, M. Koenig, M. Schwotzer, H. Gliemann, C. Wöll and M. Tsotsalas, Enhancing the Quality of MOF Thin Films for Device Integration Through Machine Learning: A Case Study on HKUST-1 SURMOF Optimization, *Adv. Funct. Mater.*, 2024, **34**, 2404631.