

Large-Scale Free Energy Calculations on a Computational MOF Database: Toward Synthetic Likelihood Predictions

Ryther Anderson[†] and Diego A. Gómez-Gualdrón^{†*}

[†]Department of Chemical and Biological Engineering, Colorado School of Mines, Golden CO 80401, USA

ABSTRACT: Metal-organic frameworks (MOFs) have captivated the research community due to their modular crystal structure that can be tailored to suit diverse applications. However, identifying ideal MOFs for an application of choice is difficult due to the millions (or more) possible MOFs one could consider. Although computational screening of MOF databases has provided a fast way to evaluate MOF performance, experimental validation of the predicted “exceptional” MOFs is uncommon due to uncertainty on the synthetic likelihood of computationally constructed MOFs, hence hindering material discovery. Aiming to leverage the perspective provided by large datasets, we calculated the free energies of each MOF in a topologically diverse database of 8,500 frameworks and evaluated to what extent descriptors of MOF thermodynamic stability “discriminate” previously synthesized MOFs. Upon defining a relative free energy, $\Delta_{\text{LMF}_{\text{FL}}}$, that corrects for some force field artifacts specific to MOF nodes, we found that previously synthesized MOFs in our database clustered in a region below $\Delta_{\text{LMF}_{\text{FL}}} = 4.4$ kJ/mol per atom. This suggests that a MOF below this $\Delta_{\text{LMF}_{\text{FL}}}$ threshold may have a higher *probability* of being synthesized, although other factors may ultimately impair synthetic accessibility. For instance, when isomorphism occurs, multiple isomorphs may reside under the $\Delta_{\text{LMF}_{\text{FL}}}$ threshold and relative stability among isomorphs come into play. From 32 isomorphous MOF series we examined in detail, we found the synthesized isomorph was the one with the lowest free energy in 80% of cases, and in 20% of cases to be within 1 kJ/mol of the latter. These findings indicate that for a MOF to be synthetic accessible a “low” crystal free energy is necessary, albeit in some cases it may not be sufficient due to the role of other factors not considered here (e.g. solvents, modulators and kinetics).

1. INTRODUCTION

Nanoporous crystals, led by a class of materials known as metal-organic frameworks (MOFs), have the potential to provide technological breakthroughs in fields as diverse as catalysis,¹⁻⁴ chemical separations,⁵⁻⁸ molecular sensing,⁹⁻¹¹ drug delivery,¹²⁻¹⁵ and energy storage.¹⁶⁻¹⁹ MOFs are promising in such diverse fields due to the tunability of their pore geometry and chemistry, a property which allows researchers to precisely control how molecules adsorb, diffuse, and react within the framework. This tunability is the direct result of the combinatoric nature of MOF construction, where each combination of MOF building blocks and topology defines a unique MOF, which can be thought of as a point in “MOF-space.” Given the number of these combinations (i.e. the vast size of MOF space), synthesizing and experimentally testing every possible MOF is impossible. Therefore, discovering a high-performing MOF for a given target application is a major challenge that requires new, sophisticated methods for searching MOF space.

In recent years, high throughput computational screening (HTCS) has emerged as an efficient method (relative to experimental screening) to not only explore the MOF-space,²⁰ but also for exploring other porous-material spaces.^{21,22} By screening databases of experimentally reported structures (e.g. the CoRE MOF^{23,24} or the IZA zeolite²⁵ databases), already explored material-space can be revisited for new applications. However, screening databases of hypothetical materials

allows venturing beyond the bounds of known material-space and opens the door to new material designs that could outperform existing ones.^{20,22,26,27} A typical HTCS study of hypothetical materials consists of the computational synthesis of a database of prototypes, followed by molecular simulations to calculate performance-defining properties for a target application. An early example of such databases of material prototypes is the Predicted Crystallography Open Database (PCOD),²⁸ which was primarily populated with hypothetical zeolites.²⁹ Ideally, some of the promising material prototypes in a database would then be selected for synthesis and experimental validation of their properties, thus accelerating material discovery. In the case of MOFs, a few high-performing materials have been synthesized as the result of this type of HTCS study. For example, the chemically robust NU-800 for methane storage¹⁶ or of the small-pore SBMOF-1 for Xe/Kr separation.³⁰ However, many HTCS studies do not result in the synthesis of a high-performing MOF, which is perhaps due to the way databases of MOF prototypes are constructed.

MOFs in these databases are built by assembling building blocks primarily on the basis of geometrical rules informed by coordination chemistry.^{20,31-34} Thus, although the majority of MOF prototypes may *look* feasible, there is uncertainty as to whether a given prototype can actually be synthesized. Currently, synthesis decisions following HTCS are based on chemical intuition about MOF synthesizability. Given the

cost of exploratory experiments, synthesis is not necessarily attempted for the best MOF prototypes. Rather, synthesis is attempted for “good” prototypes that resemble previously synthesized MOFs (e.g. a prototype within a known isorecticular series), at the expense of missing potentially higher-performing MOFs. This hinders the effectiveness of simulation-guided MOF discovery, and effectively constrains the exploration of MOF-space to the vicinities of already existing MOFs. Incidentally, the tendency to constrain exploration around existing materials also pervades purely experimental work. As noted in our earlier work, despite the appeal of MOF tunability, MOF-space has arguably been underexplored, with 50% of over 40,000 articles on MOFs (published at the time) focusing on only 30 MOFs.³³

Clearly, one way to modify typical HTCS protocols to accelerate MOF discovery (and encourage wider experimental exploration of MOF-space) is to predict the “synthesizability” of MOF prototypes in a database. The development of an effective criterion for synthesizability is tied to understanding how MOF synthesis occurs. Although recent experiments have provided some insights into the fundamental processes controlling MOF nucleation and growth,³⁵ there remains a lack of consensus on exactly what factors control whether a MOF can be synthesized or not, and whether these factors are thermodynamic or kinetic in nature. Molecular simulations show potential for clarifying how and why certain MOFs are observed experimentally, and others are not. Yet, compared to the number of studies that focus on MOF property prediction, only a few simulation studies have investigated MOF properties directly related to their synthesis.²⁷⁻⁴⁷

In 2004, Mellot-Draznieks et al. used Monte Carlo simulations to anneal building blocks (starting from a disordered arrangement) into MOFs.^{36,37} Although the method was successful in predicting potential MOF structures for a given set of building blocks, few conclusions about MOF stability or formation mechanisms could be made due to the absence of dynamics and an unrealistic description of building block interactions. Since 2015, researchers have started to use molecular dynamics (MD) simulations to follow kinetic aspects of MOF self-assembly. Starting with disordered building blocks in implicit solvent, Yoneya et al. were able to form a crystalline framework over a long MD simulation by tuning the metal-linker interactions.³⁸ However, the effects that this artificial tuning of the building block interactions might have on the formation mechanism remain unclear. Biswal and Kusalik used equilibrium MD to observe the early formation stages of MOF-2/MOF-3 in explicit solvent. However, only a primarily amorphous phase formed during the simulation.^{39,40} Kollias et al.⁴¹ used metadynamics simulations to probe the free-energy landscape during the formation of crystal-like secondary building units (SBUs). Notably, Colón et al.⁴² used MD simulations and the finite string method to probe the free-energy

landscape during the assembly of 27 SBUs into a ~2.5 nm MOF-5 crystallite.⁴² Evidently, the use of biased MD to model MOF self-assembly is encouraging, but questions remain on how the choice of interaction parameters, reaction coordinates, and model simplifications could impact conclusions.

Another set of simulation studies have focused on the energetics of fully assembled MOFs, often with the goal of understanding the synthetic preference for certain isomorphs. MOF isomorphs, while all having the same composition, can have considerably different properties^{16,33,43} due to differences in how the building blocks are arranged in space. That is, due to differences in crystal topology.⁴⁴⁻⁴⁶ Accordingly, when a high-performing MOF prototype is identified during HTCS, there is a chance that a worse-performing isomorph is actually favored during synthesis. Using molecular mechanics, Amirjalayer et al. showed that the experimentally observed HKUST-1/Cu-BTC⁴⁷ (**tbo** topology) has a lower strain energies than its unobserved **pto** isomorph. Analogous findings were reported for the experimentally observed MOF-14⁴⁸ (of the catenated, **pto-c** topology) and its unobserved **tbo-c** isomorph.⁴⁹ Impeng et al. showed that three experimentally observed Cu-paddlewheel MOFs of the **nbo-b** topology had lower strain energy than four hypothesized isomorphs, respectively.⁵⁰

Using density functional theory (DFT) energies, Cai et al.⁵¹ explained the preference between the **tbo** and **fmj** topologies for four MOFs (one of them being Cu-BTC) depending on how the BTC linker was functionalized. Arhangelskis et al. used DFT energies to determine the most stable isomorph among eight zeolitic imidazole framework (ZIF) prototypes based on a CF₃IM linker, which was subsequently synthesized.⁵² Based on the similarity of DFT formation energies, Liu and Truhlar explained the often reported coexistence of the MOF isomorphs NU-901 and NU-1000 (**scu** and **csq** topology, respectively).⁵³ Curiously, an inspection of the literature⁵⁴⁻⁵⁶ seem to indicate that synthesis more readily yields the reportedly less stable isomorph, NU-1000, which could be due to a number of synthetic factors not being considered in the calculations.

Although strain/lattice energies have shown some success in explaining synthetic preference for a given MOF isomorph,⁴⁹⁻⁵² these quantities do not account for thermal effects. On the other hand, there are instances where temperature (while keeping other synthesis conditions constant) has been reported to change the topology of the synthesized MOF.⁵⁷⁻⁵⁹ Thus, in principle, a temperature-dependent quantity such as free energy should be considered in MOF synthetic accessibility predictions. Gee and Sholl⁶⁰ calculated the harmonic free energy and free energy of solvation for 25 ZIF isomorphs to assess their relative stability. A similar evolution in prediction methods can be seen in the pharmaceuticals field, where free energy calculations⁶¹⁻⁶⁵ are taking the place of traditional lattice energy calculations,⁶⁶⁻⁶⁸ as the former increasingly appear necessary to predict

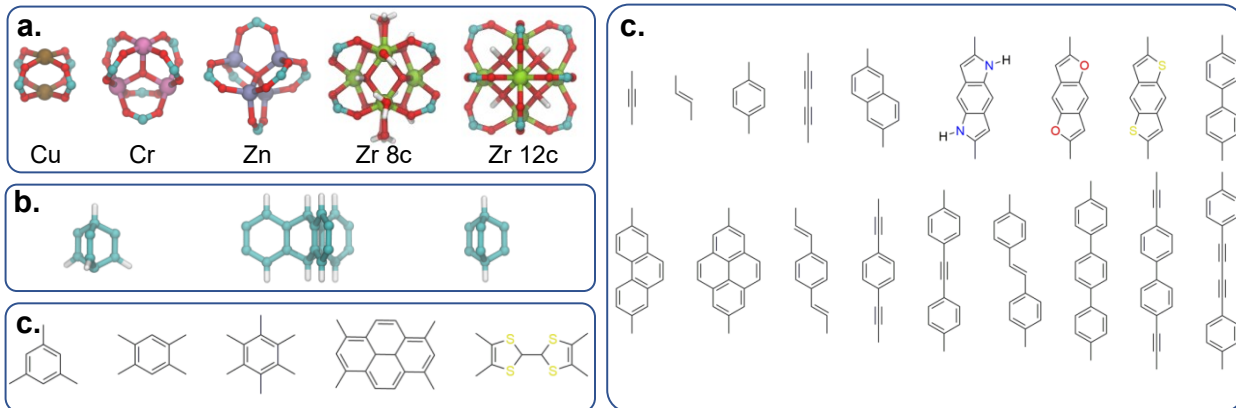


Figure 1. The computational building blocks used to construct our 8,500-MOF database. a) nodular inorganic building blocks (NIBBs). b) 3-dimensional, nodular organic building blocks (NOIBs). c) 2-dimensional nodular organic building blocks. d) connecting building blocks (CBBs). NOIBs approximately correspond to so-called “secondary building units” (SBUs) in MOF synthesis literature. NOIBs correspond to the node of multitopic organic linkers. CBBs approximately correspond to ditopic linkers and/or to the arms of multitopic organic linkers.

experimentally observed structures among energetically close organic crystal isomorphs.⁶¹

Large-scale, free energy calculations on MOF databases used for HTCS, however, have not been attempted despite the potential insights that could be gleaned from the large amount of simulation data (e.g. through the emergence of well-defined structure-property relationships relevant to MOF synthesizability). Here we decided to assess the viability of classical MD simulations to estimate MOF free energies in a way that is (sufficiently) accurate, yet still amenable for large-scale screening. Inspired by methods used in the pharmaceuticals field for crystal structure prediction of organic molecules,^{61–65} we decided to apply and evaluate free energy calculation methods in a database of 8,500 MOF prototypes: the quasi-harmonic approximation (QHA) and thermodynamic integration, specifically the Frenkel-Ladd (FL) path.^{69–71}

Our MOF prototypes were computationally constructed using ToBaCCo-3.0,^{33,34} from 34 building blocks combinatorially assembled according to 110 topological blueprints. As is typical of this procedure,^{17,20,72–74} one expects most generated MOFs to not have been experimentally reported—regardless of whether they are actually synthesizable—but one also expects to generate some experimentally reported MOFs. Additionally, with sufficient topological diversity in the database, one expects to find numerous, non-experimentally observed isomorphs of the previously synthesized MOFs. We aimed to exploit this composition of the database to assess to what extent free energies are able to “discriminate” previously synthesized MOFs. By using the free energies of previously synthesized MOFs as a point of reference, we interrogated whether *i)* previously synthesized MOFs share common thermodynamic stability characteristics despite varied composition and/or *ii)* previously synthesized MOFs present the highest thermodynamic stability among MOFs of identical composition (i.e.

isomorphs). In addressing these questions, we explored possibilities and limitations for the use of MOF free energy as a descriptor for synthetic *likelihood* of MOFs in computational databases.

2. METHODS

2.1 MOF Database Construction. Our 8,500-MOF database was created using ToBaCCo-3.0,^{33,34} with the building blocks shown in **Fig. 1**. These building blocks were selected to engender topological and chemical diversity, i.e. so our database spans a wide range of textural and energetic properties. In contrast to the databases we created for recent studies that focused on the prediction of adsorption properties,^{73,75} the current database does not include functionalized MOFs, which allowed us to explore more topologies, linker shapes, and coordination schemes, and increase the occurrence of isomorphism. The geometry of each MOF prototype was optimized in LAMMPS⁷⁶ according to the relevant force field: UFF,⁷⁷ Dreiding,⁷⁸ or UFF4MOF.^{79,80} Note, however, that since parameters for Cu in UFF are defined for tetrahedral coordination, we altered the equilibrium angle to 90° to be consistent with the Cu-paddlewheel geometry. Similarly, in Dreiding equilibrium bond lengths and angles are unavailable for some metals (Cu, Cr, Zr), so we adopted the equilibrium bonds/angles we used in our UFF implementation but otherwise retained the functional forms and force constants used for Dreiding. Note that these modifications to Dreiding and UFF have been previously validated for MOFs.⁸¹

To optimize MOF geometries, we employed an iterative approach. First, only the atom positions were relaxed using the fast inertial relaxation engine (FIRE) algorithm⁸² with a timestep of 10.0 fs. Then the atom positions and simulation cell parameters were relaxed simultaneously using a conjugate gradient algorithm. We repeated this process (relaxing only the atom positions and then the atom positions and cell parameters together) until the energy change from the previous iteration was less than 4.184×10^{-6} kJ/mol. The individual

optimizations were halted when the maximum force on any atom was less than 4.184×10^{-6} kJ/(mol Å) and the energy changed by less than 1×10^{-4} % from the previous step.

2.2 Free Energy Calculations. All free energy calculations were performed using molecular dynamics (MD) in LAMMPS⁷⁶ with a timestep of 1 fs. The thermostat of Bussi et al.⁸³ (damping parameter of 100 timesteps) was used for all simulations and a Nosé-Hoover barostat (damping parameter of 1000 timesteps) was used for NPT simulations. We calculated quasi-harmonic approximation (QHA) free energies using:

$$F_{\text{QHA}} = U + \frac{1}{2} \int_{v=0}^{v_{\text{max}}} h\nu D(\nu) d\nu + k_B T \int_{v=0}^{v_{\text{max}}} D(\nu) \ln \left[1 - \exp \left(\frac{-h\nu}{k_B T} \right) \right] d\nu \quad (1)$$

where U is the potential energy at 0 K for the equilibrium cell shape, ν are vibrational frequencies, and $D(\nu)$ is the vibrational (phonon) density of states (PDOS).^{84–86} $D(\nu)$ is normalized such that $\int D(\nu) d\nu = 3N$ where N is the number of atoms in the system.⁸⁷ To apply Eq. 1 we needed the PDOS for each MOF, which we obtained by Fourier-transforming an averaged, mass-weighted velocity trajectory (i.e. the mass-weighted velocity of each atom at each timestep).^{87,88} First, we calculated the equilibrium simulation cell shape in the NPT ensemble at zero pressure by averaging each parameter over 100 ps after 100 ps of equilibration. The velocity of each MOF atom was then output every 5 timesteps for 10 ps after 50 ps of equilibration in the NVT ensemble (at the equilibrium cell shape). These velocities were used to compute the PDOS for each MOF. More details are provided in **Section S1**.

We calculated the Frenkel-Ladd (FL) path free energies using the method of Freitas et al. which is described in detail in ref.⁷¹. Briefly, the FL path is a thermodynamic integration method in which the system is switched from an Einstein crystal (where each atom is constrained to its defined equilibrium position by a harmonic potential) to a force field representation. During thermodynamic integration, an initial system is switched to a final system according to a parametrical Hamiltonian

$$H(\lambda) = \lambda H_f + (1 - \lambda) H_i \quad (2)$$

where H_i is the initial Hamiltonian, H_f is the final Hamiltonian, and λ is the switching parameter. For $\lambda = 0$, $H(\lambda)$ is H_i and for $\lambda = 1$, $H(\lambda)$ is H_f .^{69,71,89} The difference in free energy (ΔF) between the initial and final states is equal to the work required to switch between the Hamiltonians:

$$\Delta F = \int_0^1 \left\langle \frac{\partial H}{\partial \lambda} \right\rangle_\lambda d\lambda \quad (3)$$

where angled brackets denote an ensemble average at each value of λ . Since the free energy of an Einstein crystal (F_{EC}) is known analytically, an absolute free energy of the solid is given by:

$$F_{\text{FL}} = \Delta F + F_{\text{EC}} \quad (4)$$

where ΔF is the work required to switch from the Einstein crystal to the force field representation, and where

$$F_{\text{EC}} = 3k_B T \sum_{i=1}^N \ln \left(\frac{\hbar \omega_i}{k_B T} \right) \quad (5)$$

with ω_i being the harmonic frequency of atom i in an N -atom system. One way to calculate the ΔF in Eq. 3 is by performing a series of equilibrium MD simulations for each value of λ , during which $\langle \partial H / \partial \lambda \rangle$ is calculated as an ensemble average. These averages can then be numerically integrated from $\lambda = 0$ to $\lambda = 1$ to yield ΔF . However, Freitas et al. showed, following work by de Koning,⁹⁰ that ΔF can be calculated as the average work of two *non-equilibrium* switching processes.⁷¹ First, the system is switched from the Einstein crystal to the forcefield representation (i.e. λ is changed from 0 to 1) during the course of a single non-equilibrium MD trajectory. Then, the system is switched back to the Einstein crystal after equilibration with the forcefield Hamiltonian. The forward work is denoted $W_{i \rightarrow f}^{\text{irr}}$ and the reverse work is denoted $W_{f \rightarrow i}^{\text{irr}}$, with the superscript indicating that the forward/reverse switching processes are irreversible. For sufficiently long switching trajectories, the heat dissipated during switching is small⁹¹ and the heat dissipated during the forward and reverse processes is of the same magnitude and can be cancelled (see ref.⁷¹ for a detailed justification), meaning:

$$\Delta F = \frac{1}{2} (W_{i \rightarrow f}^{\text{irr}} - W_{f \rightarrow i}^{\text{irr}}) \quad (6)$$

Accordingly, from Eq. 4-6 one can write:

$$F_{\text{FL}} = \frac{1}{2} (W_{i \rightarrow f}^{\text{irr}} - W_{f \rightarrow i}^{\text{irr}}) + 3k_B T \sum_{i=1}^N \ln \left(\frac{\hbar \omega_i}{k_B T} \right) \quad (7).$$

Typically, there are additional terms that correct for fixing the center of mass during the simulation, however we found this term to be significantly smaller than the statistical error for our systems (see **Section S1**).^{71,89} In principle, $W_{i \rightarrow f}^{\text{irr}}$ and $W_{f \rightarrow i}^{\text{irr}}$ should be averages calculated from several replicate simulations. However, we found that a single simulation with long enough switching time was sufficiently precise (see **Section S1**).

Practically, we implemented the FL path free energy calculations in a single MD simulation. First, the equilibrium simulation cell parameters were calculated with the same method used for our PDOS calculations (see above). We then switched to the NVT ensemble (as

the FL path is only valid at fixed volume) using the equilibrium cell shape. Next, we calculated the mean-squared displacement (MSD) of each atom type (according to the force field) as an ensemble average over 50 ps after 50 ps of equilibration (MSDs are used in Eq. 8, below). We then ran the switching procedure, with switching time of 400 ps, using the LAMMPS command `fix ti/spring`.⁷¹ Details on the convergence of free energy with switching time are provided **Section S1** and **Figure S1**. When using `fix ti/spring` the system is switched from the force field representation to the Einstein crystal representation, equilibrated with the Einstein crystal Hamiltonian (in our case for 30 ps), then switched back, which is the reverse of the traditional FL path (described above). However, we can simply reverse the sign of the calculated work to obtain the ΔF used in Eq. 4.

The “strength” of the springs for each atom in the Einstein crystal is defined by force constants. Here, we calculated a different force constant (α_j) for each atom type j in the simulation (according to the relevant force field), using:

$$\alpha_j = \frac{3k_B T}{\langle (\Delta r)^2 \rangle_j} \quad (8)$$

where $\langle (\Delta r)^2 \rangle_j$ is the mean squared displacement (MSD) of atom type j (calculated as described above). This ensures that the MSDs for the real and Einstein crystals are similar, which makes the switching process more efficient.⁷¹ However, we enforced a lower bound on α_j of 41.84 kJ/(mol Å²) (see detailed justification in **Section S1** and **Fig. S2**). If α_j is too small, Eq. 3 can be difficult to evaluate due to divergence in the integrand from overlapping atoms.⁸⁹ In addition, if an atom travels too far from its Einstein crystal equilibrium position LAMMPS may lose that atom when building neighbor lists.

3. RESULTS AND DISCUSSION

3.1 Comparison of MOF free energy calculations methods. The quasi-harmonic approximation (QHA) assumes that each MOF vibrational mode can be reasonably modeled with a harmonic potential. By contrast, no such assumption is made in the Frenkel-Ladd (FL) path method. In principle, then, one can take the FL free energies to be more accurate. However, should QHA and FL free energies differ by a sufficiently small amount, one would prefer to calculate QHA free energies in MOF databases as they require shorter simulations (**Fig. S3**). Thus, we start our discussion by comparing the QHA and FL path free energies (both calculated using UFF4MOF) in our 8,500-MOF database.

Fig. 2 is a parity plot comparing QHA (y-axis) and FL (x-axis) free energies. All free energies in the present work are calculated at 300 K and reported on a “per atom”

basis unless otherwise specified. If the energies calculated using the two methods were equal, all points would lie on the line $y = x$ (the parity line). Based on **Fig. 2**, there is an *apparent* systematic deviation between FL path and QHA free energies, which is consistent with observations by Pastorino and Gamba as well as Stoessel and Nowak when comparing QHA and FL free energies of sulfur crystals and biomolecules, respectively.^{85,92} Note that purely systematic deviation in QHA free energies would not matter when using free energy as a metric of relative thermodynamic stability, as free energy differences between individual structures would be maintained. Such a scenario would be reflected by a perfectly linear correlation between QHA and FL free energies.

To quantify the correlation between QHA and FL free energies, we report the Pearson (R) and Spearman rank (S) correlation coefficients. An R value of one indicates a perfectly linear correlation between QHA and FL free energies, whereas an S value of one indicates that the stability ranking of MOFs based on QHA or FL path free energies would be identical. We found both R and S to be close to one (**Fig. 2**, $R = 0.997$, $S = 0.981$). These R and S values suggest that one would observe similar free energy vs. MOF property relationships with either the

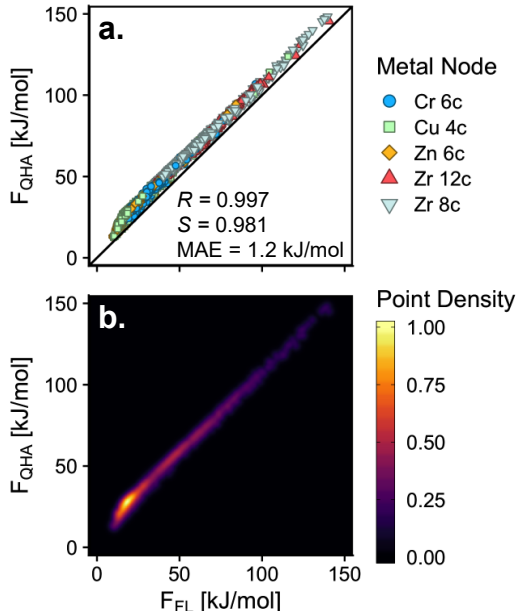


Figure 2. Comparison between free energies calculated with the QHA (F_{QHA}) versus free energies calculated via the FL path (F_{FL}). a) Parity plot with point shape and color corresponding to the type of MOF inorganic node. b) Parity plot illustrating the point density according to the color bar. Insets are the Pearson (R) and Spearman Ranking (S) correlation coefficients, and the mean absolute error (MAE) calculated for adjusted QHA free energies.

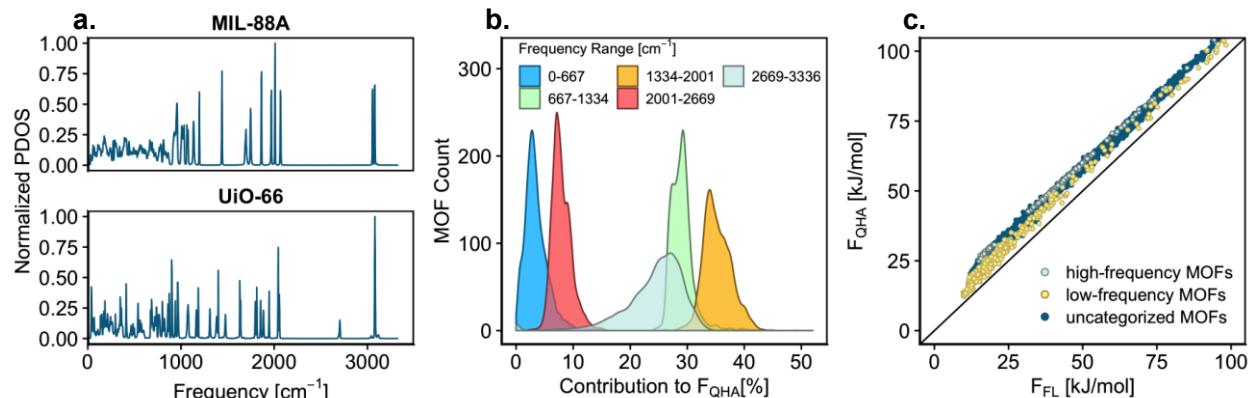


Figure 3. a) Phonon density of states (PDOS) for two previously synthesized MOFs (names are shown above each plot). b) Histograms of the contribution of frequencies in five ranges to the total QHA free energy (F_{QHA}) for each MOF (e.g. for most MOFs, frequencies in the 0-667 cm⁻¹ range contribute less than 10% to F_{QHA}). c) Parity plot comparing F_{QHA} and F_{FL} (FL free energies), with point colors indicating low-frequency, high-frequency, and uncategorized MOFs.

QHA or the FL method (i.e. broad trends would be preserved when switching between methods). However, the reproduction of broad trends has no bearing on whether QHA and FL path free energies are equally valid criteria for assessing the relative thermodynamic stability of MOFs.

Thus, we calculated the absolute deviation between the QHA and FL free energies, after removing the apparent systematic error from the QHA values, which we did by subtracting the y-intercept of a linear model fit to the data shown in **Fig. 2** (9.0 kJ/mol). Note that this subtraction does not affect differences among QHA free energies. However, it enables us to better evaluate the ability of the less expensive QHA method to correctly assess relative thermodynamic stability of MOFs. Once this adjustment was made, we found the absolute errors to span the 0-7.5 kJ/mol per atom range (**Fig. S4**), with the mean absolute error (MAE) being 1.2 kJ/mol per atom. We will see in **Sections 3.3** and **3.4** that *i*) previously synthesized MOFs encountered in our database cluster below a 4.4 kJ/mol per atom free energy threshold, and *ii*) free energy differences between MOF isomorphs are typically small (with a mean difference of 3 kJ/mol and many differences of less than 1 kJ/mol). On these grounds, one can see that the use of QHA free energies to assess thermodynamic stability (and perhaps synthetic likelihood) of MOFs is potentially problematic.

Based on the above, in subsequent sections our discussion will be based on FL free energies. Before proceeding, however, we think it valuable to briefly examine the origins of the differences between QHA and FL free energies in MOFs. MOFs present vibrations over a wide range of frequencies (0-3336 cm⁻¹)—typically with a gap around 2300 cm⁻¹—as illustrated by representative phonon density of states (PDOS) in **Fig. 3a** and **Fig. S5**. By dividing the overall frequency interval evenly into five ranges, we present in the histograms in **Fig. 3b** to what extent vibrations in each range collectively contribute to the QHA free energy of MOFs.

Fig. 3b shows that high (> 2669 cm⁻¹), mid-high (2001-2669 cm⁻¹), mid (1334-2001 cm⁻¹), mid-low (667-1334 cm⁻¹) and low (< 667 cm⁻¹) frequency ranges typically contribute ~25%, ~7%, ~35%, ~30%, and ~3%, respectively, to the QHA free energy. The exact collective contribution of vibrations in each range, however, varies with MOF composition.

Informed by the 95th percentile threshold in each histogram, we classified MOFs whose high-frequency (mid-low-frequency) vibrations contribute more than 31% (32%) to the QHA free-energy as “high-frequency” (“low-frequency”) MOFs (details in **Section S2**). As reported in earlier works,⁹³⁻⁹⁶ high-frequency vibrations correspond to bond stretching modes involving (light) H atoms. Mid-frequency vibrations mostly correspond to bond stretching modes *not* involving H or (heavy) metal atoms, and in-plane-angle-bending modes in aromatic rings. Mid-low-frequency vibrations mostly correspond to out-of-plane-angle-bending modes in aromatic rings, and stretching modes involving metal atoms (e.g. in M-O coordination bonds). Low frequencies mostly correspond to collective motions, including twisting of organic linkers and rocking of inorganic nodes. The PDOS of high-frequency MOFs in our database are dominated by C-H stretching modes, whereas low-frequency MOFs present very few (or none) of these modes (**Figure S6**).

The location of high- and low-frequency MOFs in the parity plot in **Fig. 3c** reveals that the QHA free energy error is more significant for high-frequency MOFs (albeit more “systematic”). This is perhaps surprising as high-frequency MOFs are probably more “harmonic” and thus expected to be more amenable to QHA. To understand this observation, consider that the QHA free energy (Eq. 1) can be considered as the lattice energy U “corrected” by vibrational terms: the zero-point energy (ZPE) and the temperature-dependent logarithmic term to U . At 300 K, ZPE becomes the dominant correction for frequencies above 334 cm⁻¹. Given that U is less than the

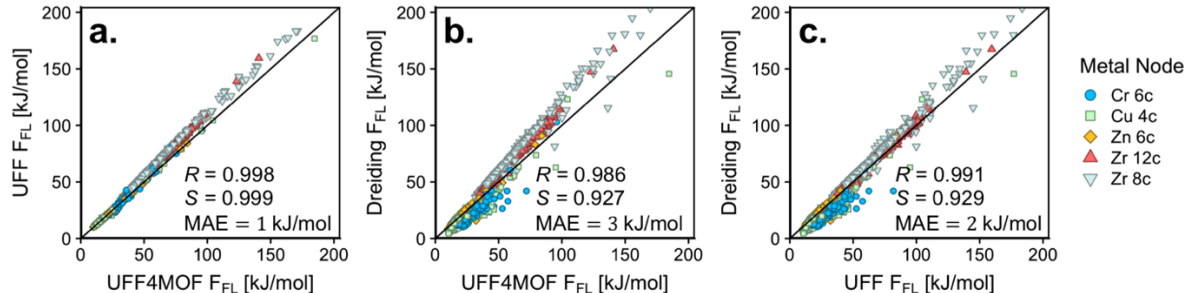


Figure 4. Parity plots comparing FL free energies (F_{FL}) calculated with different force fields. a) UFF vs. UFF4MOF. b) Dreiding vs. UFF4MOF. c) Dreiding vs. UFF. Point shape and color correspond to the type of metal node in each MOF plotted. Different point features indicate the type of metal node in the MOF. Insets are the Pearson (R) and Spearman Ranking (S) correlation coefficients, and the mean absolute error (MAE).

FL path free energy, the error between QHA and FL free energies must be primarily due to the ZPE, which is consistent with the QHA free energies of “high frequency” MOFs presenting larger errors (Fig. 3c) and the strong correlation between the error and ZPE (Fig. S6). Note, however, that the error tends to be more systematic (and hence less consequential for relative thermodynamic stability) in high-frequency MOFs than in low-frequency ones. This discussion is extended somewhat in Section S2, including details on temperature dependence for QHA and FL free energies (Figs. S8 and S9). However, a deeper discussion onto this subject is beyond the scope of this contribution.

3.2 Comparison between free energies calculated with different force fields. A question that emerges in any classical simulation study is the sensitivity of calculations to force field selection. Up to this point, we have focused on free energies calculated using UFF4MOF, because this update to UFF by Addicoat et al.^{79,80} was specifically designed for MOFs. While several MOF-specific force fields (of varied generalizability) exist,^{97–99} here we focus on a comparison of UFF4MOF with the generic force fields Dreiding⁷⁸ and UFF.¹⁰⁰ This was motivated by the tendency of HTCS studies to use these generic force fields (given their essentially complete transferability), as well as by the work of Boyd et al.⁸¹ who reported that Dreiding and UFF perform reasonably well when predicting MOF properties such as bulk modulus and thermal expansion coefficients.

Parity plots comparing FL free energies among force fields are presented in Fig. 4. The closest agreement in free energies is between UFF and UFF4MOF, especially for MOFs with Cu (paddlewheel) nodes (Fig. S12). The largest deviations are observed for MOFs with Cr, Zn, and especially Zr nodes, which can be rationalized by considering that UFF4MOF only modifies the UFF parameters that involve metals (e.g. M-O bonds, O-M-O, M-O-C angles, etc.),^{79,80} and there are more of these bond/angles in the Cr, Zn, and Zr-MOFs. The largest disagreement in free energies is between Dreiding and UFF4MOF, which is expected, given that Dreiding uses

different functional forms and force constants than UFF4MOF for all interactions. The “outliers” in Fig. 4b–c correspond to MOFs whose lattice parameters change significantly when switching from UFF4MOF/UFF to Dreiding.

Although the point spread in Fig. 4 seems large, density plots (Fig. S12) show that most points remain near the parity line in each plot. Thus, a more comprehensive comparison is given by the values of R , S and MAE shown in Fig. 4, as well as histograms for the absolute errors (Fig. S12). In all cases R and S are close to one, indicating that similar free energy vs. MOF property relationships would be obtained with UFF4MOF, UFF, or Dreiding force field. However, consistent with our discussion comparing QHA and FL free energies (MAE ~ 1 kJ/mol) we must conclude that the use of Dreiding (MAE ~ 3 kJ/mol) instead of UFF4MOF to assess the relative thermodynamic stability of MOFs may be problematic. Similarly, UFF free energies may not be sufficiently accurate (MAE ~ 1 kJ/mol) when dealing with small differences in thermodynamic stability. From this point on, then, we primarily base our discussion on results using UFF4MOF, although we will point (where appropriate) differences and similarities in results were one to use Dreiding or UFF.

3.3 Correlations between free energy and MOF properties (identifying an overall thermodynamic stability threshold). Prior studies^{101–104} postulated criteria for zeolite synthetically “feasibility,” with varied degree of success. Inspired by these works, we decided to explore whether a similar criterion can emerge for MOFs, and assess to what extent such criterion succeeds or fails. However, acknowledging the numerous factors that may facilitate or hamper the synthesis of a particular MOF, in the context of HTCS, one probably can only talk of a synthetic *likelihood* criterion. An effective criterion would enable a preliminary step in HTCS studies as to eliminate MOFs with low synthetic likelihood from a computational database. Thus, computer resources can be

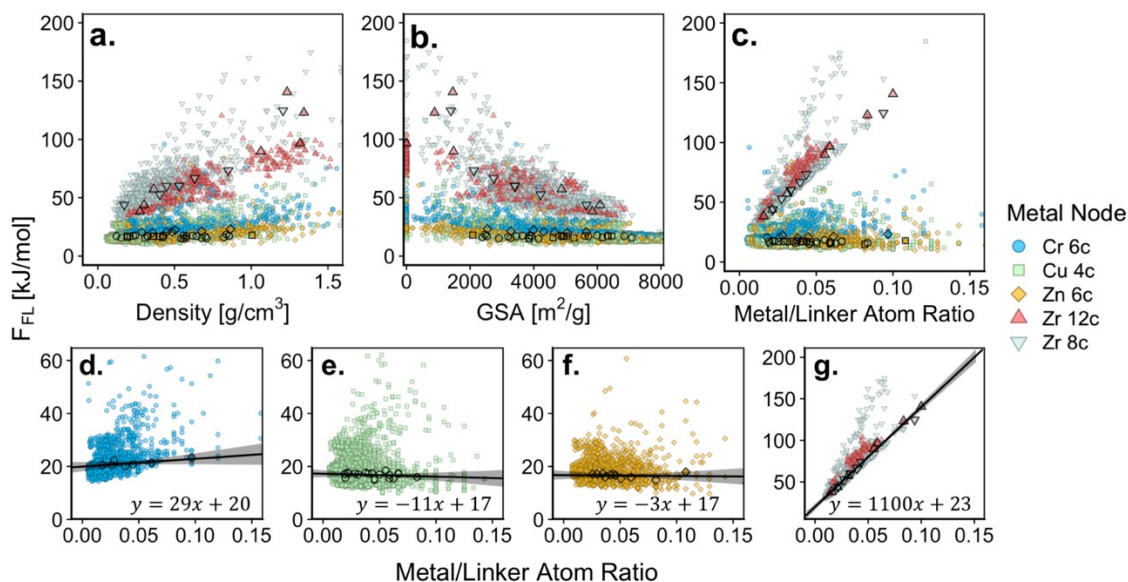


Figure 5. FL free energies (F_{FL}) versus MOF properties. a) F_{FL} vs. MOF density. b) F_{FL} vs. MOF gravimetric surface area (GSA). c) F_{FL} vs. metal to linker atoms ratio. d-f) breakdown of F_{FL} vs. metal to linker atoms ratio plot in c by metal node. d) 6c Cr MOFs. e) 4c Cu MOFs. f) 6c Zn MOFs. g) 8/12c Zr MOFs. Point features indicate the inorganic node in the MOF. Points outlined in black correspond to previously synthesized MOFs found in our 8,500-MOF database. Black lines in plots d-f are linear fits (equation shown) for the free energies of previously synthesized MOFs of each inorganic node type, and shaded regions correspond to the 95% confidence interval of each linear fit.

allocated only to MOFs that have a higher chance of being synthesized. Even in instances where calculations on non-synthesizable MOFs could still be valuable (e.g. for machine learning^{75,105}), the criterion could be used to downsize a database that is too large for costly simulations.

Given that MOFs have diverse compositions, postulating a synthetic likelihood criterion is perhaps less straightforward than for zeolites. For example, while Li et al. identified a criterion for zeolite stability based on the relationship between T-site to T-site distances and O-site to O-site distances,¹⁰¹ such a criterion is only possible because all zeolites are arrangements of TO_4 tetrahedra. An analogous criterion certainly does not exist for MOFs in general. On the other hand, Simperler et al.¹⁰⁴ identified a region in lattice energy vs. density plots where previously synthesized zeolites tended to cluster, hence postulated a “synthetic feasibility factor” based on how close (or far) a hypothetical zeolite is to (or from) this region. However, note that synthesis of one “unfeasible” zeolite has been recently accomplished,¹⁰⁶ which highlights that factors like these should be interpreted from a probabilistic perspective. Nonetheless, it is easier to imagine an extension of this lattice-stability-based approach to MOFs. Thus, we examined the correlation between MOF free energies and MOF chemical/textural properties to see whether previously synthesized MOFs exhibited a clear relationship with any of these properties.

First, we needed to identify the MOFs in our database that had been previously synthesized. To accomplish this, we calculated the MOFid, a systematic MOF identifier developed by Bucior et al.,¹⁰⁷ for each MOF in our database. We then compared the MOFids in our database to those of the CoRE MOFs (a database of previously synthesized MOFs).^{23,24} Since the MOFid for a specific MOF is unique,¹⁰⁷ if a MOFid from our database was also present in the CoRE MOFs, that means that the corresponding MOF in our database has been previously synthesized. Through this procedure, we found 54 previously synthesized MOFs in our database. For comparison, consider that Bucior et al.¹⁰⁷ found 50 and 18 previously synthesized MOFs from 10,159- and 1614-MOF subsets of earlier computationally constructed databases by Gómez-Gualdrón and coworkers^{17,34} and Wilmer and coworkers.²⁰

After identifying the existing MOFs in our database, we proceeded to plot FL free energies against MOF properties (Fig. 5a-c, additional relationships are shown in Fig. S14). Contrary to zeolites, we found that previously synthesized MOFs do not cluster in a single region. Rather, they tended to cluster in different regions depending on the composition of their metal nodes. For instance, the Zr MOFs occupy higher free energy regions than other MOFs. We found that this occupation of high free energy regions (Zr-MOFs > Cr-MOFs > Zn-MOFs ~ Cu-MOFs) correlates with the strain energy (per atom) of the corresponding isolated nodes (Table S2). The relative energies of the isolated nodes also explains the different

behavior of free energy versus the metal/linker atom ratio for different MOF compositions. To calculate this ratio, any Cu, Cr, Zn, or Zr atom was counted as a metal, and any atom bonded to a carbon was counted as a linker atom (a sufficient criteria for the linkers in our database, but not for MOFs in general). For instance, in Zr-MOFs, free energy increases rapidly with the metal/linker atom ratio as adding more Zr-nodes will always increase potential energy. (Qualitative similar trends are obtained when using UFF or Dreiding as shown in **Figs. S14** and **S15**).

When examining the different correlations separately, we can observe that the free energies of previously synthesized MOFs tend towards the lower (free energy) boundaries of the corresponding region. This indicates that among MOFs of the same type, those with the lower free energies tend to have demonstrated synthetic accessibility. However, developing separate criteria for different MOF types is inconvenient and less informative than a single criterion. The clearest pattern for previously synthesized MOFs emerges from **Fig. 5c**, where a linear relationship (for each MOF type) between the free energy of synthesized MOFs and metal/linker atom ratio is apparent for each MOF type, allowing us to fit four separate linear models to these free energies, as illustrated in **Fig. 5d-g**. Thus, other synthetic factors aside, it is reasonable to assume that MOFs on and under the corresponding fit line—which can be seen as a thermodynamic stability threshold—have higher probability of being synthesized.

The data shown in **Fig. 5c** can be transformed by subtracting the relevant linear model from each MOF free energy. We refer to the free energy transformed in this manner as $\Delta_{LM}F_{FL}$. The results of this transformation are presented in **Fig. 6**, where all the previously synthesized MOFs in our database now cluster within a small $\Delta_{LM}F_{FL}$ region. The maximum $\Delta_{LM}F_{FL}$ among experimentally observed MOFs in our database is 4.4 kJ/mol. If one were to consider all MOFs with $\Delta_{LM}F_{FL}$ higher than this value as having insufficient thermodynamic stability, one could deem them to be synthetically *unlikely*. Thus, one could eliminate 1,613 MOFs from our 8,500-MOF database (or 1,502 and 707 MOFs when postulating an analogous threshold on the basis of calculations with UFF and Dreiding, respectively, **Table S2**). We visually inspected many of the MOFs categorized as synthetically unlikely, and a large portion of these are not visibly strained (**Fig. S18**). This indicates that (i) not only is visual inspection inefficient, but also insufficient to identify MOFs with low thermodynamic stability, and (ii) a geometry-based criterion for MOF synthetic likelihood is not evident.

Note that the exact $\Delta_{LM}F_{FL}$ threshold presented here is specific to free energies calculated with UFF4MOF at 300 K. However, following the strategy outlined here, analogous criteria can emerge for other forcefields (as it occurred for UFF and Dreiding, **Fig. S17**) and temperatures. However, we do not expect pronounced changes in the general threshold across the typical temperature range for MOF synthesis (around 300-400 K). The reason is that within this range, we observed only

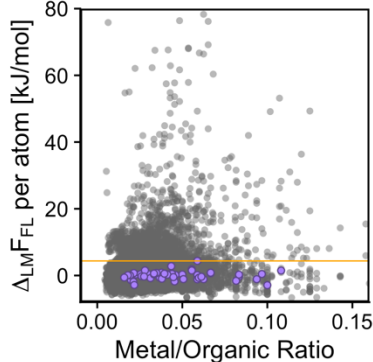


Figure 6. Relative FL free energy ($\Delta_{LM}F_{FL}$) versus metal to linker atoms ratio of each MOF. $\Delta_{LM}F_{FL}$ is obtained by subtracting the corresponding free energy from the linear fits in **Fig. 5d-g** from F_{FL} . The horizontal line shows the bound that the 54 previously synthesized MOFs (purple points) in our 8,500-MOF database are all found below.

small variations (at the database scale) of MOF free energy with temperature (**Fig. S19**)—which is not to say that the synthetic accessibility of a particular MOF is not susceptible to temperature changes. Finally, it is important to note that probably there are MOFs below the $\Delta_{LM}F_{FL}$ threshold that are synthetically inaccessible, as there will be instances where high thermodynamic stability will not be sufficient to warrant successful synthesis (as other factors come into play). However, the absence of experimentally reported MOFs above the postulated thermodynamically stability threshold suggests that “low” crystal free energy is a pre-requisite for synthetic accessibility.

3.4. Consequences of MOF isomorphism on the application of the overall thermodynamic stability criterion. Given the occurrence of isomorphism in MOFs, it is possible that a MOF that is considered synthetically likely (as per the threshold in **Section 3.3**), may not actually be synthetically accessible simply because a more stable isomorph exists that is preferentially formed during synthesis. Crucially, previous work³³ has shown that key MOF properties (e.g. mechanical stability,³³ adsorption selectivity,³³ and catalytic activity⁴³) can vary dramatically among isomorphous MOFs. Therefore, to inform synthetic attempts following HTCS, it is important to predict the most stable MOF within every isomorphous family in a database. Additionally, by only retaining the most stable isomorphs (by eliminating less stable isomorphs, even if they fall below the $\Delta_{LM}F_{FL}$ threshold) one could further reduce the size of a computational MOF database. Thus, we now proceed to assess thermodynamic stability within various MOFs isomorphous series to determine (i) how close in stability isomorphs typically are, (ii) how often multiple isomorphs fall below the $\Delta_{LM}F_{FL}$ threshold, and (iii) how often the experimentally observed isomorph corresponds to the most stable isomorph within its series.

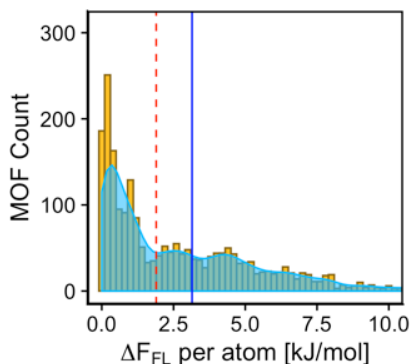


Figure 7. A histogram of the free energy differences ΔF_{FL} per atom observed between isomorphs in our 8,500 MOF database. The solid, blue line indicates the mean difference, and the dashed, red line indicates the median difference.

To this end, we first found which MOFs in our databases had isomorphs. To accomplish this, we identified all the unique linkers in our database by their canonical SMILES string, and then considered any MOFs with the same node and linker(s) to be isomorphs. We found that the free energy differences among isomorphs can frequently be well under 1 kJ/mol per atom (**Fig. 7**), which is less than the width of the stability region in **Fig. 6**, indicating that when a MOF is found to lie below the $\Delta_{LM}F_{FL} = 4.4$ kJ/mol threshold, it is extremely common to find many of its isomorphs (when existing) to also lie below the aforementioned threshold (**Fig. S20**). Thus, it follows that to examine synthetic likelihood for “low” free-energy MOFs belonging to isomorphous series, one may need to also consider their stability relative to their isomorphs.

To directly assess whether free energies can accurately differentiate between an experimentally observed MOF and its non-observed isomorphs, we identified all the isomorphs for each previously synthesized MOF in our database. This yielded 32 isomorphous series in our database that also included at least one existing MOF (not all previously synthesized MOFs in our database had isomorphs). The linkers observed among the MOFs in these series are shown in **Fig. S21**. Two of the 32 series contained multiple (two) experimentally observed isomorphs, as discussed in **Section S5** along with other case-studies. For each of the 32 series, we determined whether the experimentally observed isomorph had the lowest free energy at 300 K. This was the case for 26 (81%) of the isomorphous series (**Table S4**), or 24 (80%) if we do not consider the series with two experimental MOFs. Changes in isomorph prediction accuracies with temperature and force field are

given in **Tables S6** and **S7** (e.g. when using UFF and Dreiding, respectively, 23 and 18 experimentally observed isomorphs had the lowest free energy within their series).

In **Fig. 8**, we show one of the 32 studied isomorphous series, where the free energy of the experimentally observed NOTT-100 (**fof** topology)¹⁰⁸ is less than 0.5 kJ/mol per atom lower than three other possible isomorphs. Note that in contrast to the whole database, by definition we can say that all MOFs (experimentally observed and non-observed) in the subset of these 32 isomorphous series have had their synthesis attempted. Therefore, the abovementioned accuracy to predict the experimentally observed MOF can be used as a preliminary estimate to gauge the effectiveness of thermodynamic stability (as noted by free energies) as a criterion for MOF synthetic likelihood.

Underscoring the importance of “low” free energy for a MOF to be synthetically accessible, note that although for six isomorphous series the experimentally observed member did not have the lowest free energy, the free energy of the experimentally isomorph was (at most) within 2.5% of the lowest one in the series (within 2.0 and 2.1% when using UFF and Dreiding, respectively). The infrequent discrepancies between free energy rankings and the experimentally reported isomorph may be due to *i*) synthetic factors (e.g. solvents, modulators, and different synthesis temperatures for different MOFs) not accounted for by our free energy calculation, *ii*) experimental oversight where the lowest free energy isomorph has formed during synthesis but not reported— isomorphs that are close in free energy may coexist such as the NU-1000 and NU-901 cases discussed in **Section S5**—or *iii*) simply a prediction error inherent to either the forcefield or our free energy calculation method. Based on *ii*) and *iii*), one may consider the use of a tolerance before deciding to discard a MOF isomorph as synthetically unlikely.

Although the effect of solvation on absolute free energies have been previously noted to be small,⁸⁶ given the small free energy difference among isomorphs, it seems that examining the effect of solvents may be important to modulate the order of stability of isomorphous MOFs under synthesis conditions. However, considering the diversity of solvents used in MOF synthesis, and the computational cost associated with including solvent and calculating free energies of solvation, such task may be impractical for HTCS at this point (although it merits study in smaller, rationally-selected MOF subsets).

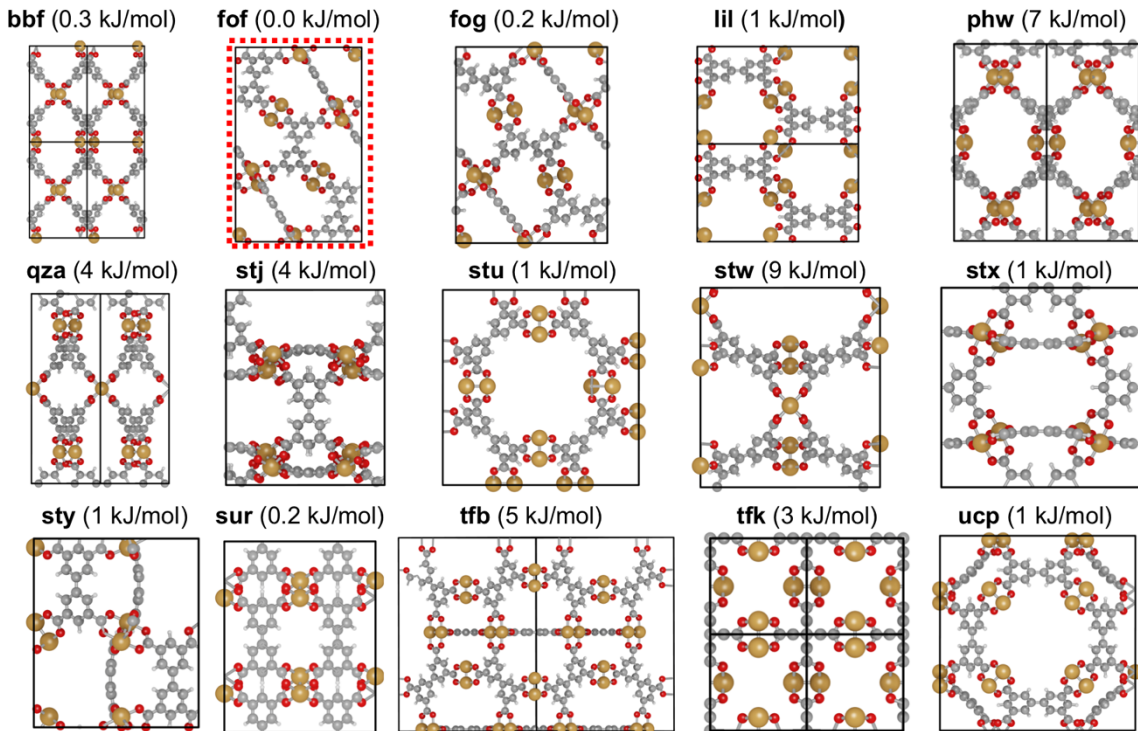


Figure 8. The experimentally observed NOTT-100 (outlined by the dashed line) along with all their isomorphs found in our database. Isomorphs are labeled by their topology, with their free energy per atom relative to the NOTT-100 (**fof** topology) presented in parenthesis.

3.5. Can relative MOF thermodynamic stability be assessed with alternative metrics? While not 100% effective, the FL free energies of “activated” (i.e. solvent-free) MOFs have shown considerable potential to inform the synthetic likelihood of MOF prototypes. Thus, free energies could be used to eliminate some synthetically unlikely MOFs from databases (reducing their size and increasing the computational efficiency of HTCS), and/or to inform synthesis attempts (e.g. when presented with two “promising” MOF prototypes, an experimentalist may decide to attempt synthesis for the one with the lowest free energy, provided it does not have a isomorph with lower free-energy). However, for this approach to work, free energies would need to be calculated for all MOFs in a database. Hence, an important question to ask is whether more affordable metrics could be used as effectively. Although we have not envisioned a geometry-based metric at this point, we decided to briefly inspect whether metrics such as potential energy at 0 K (strain energy, U_{0K}) and at a finite temperature (here, U_{300K}) could be as effective as the FL free energy as a stability metric. This was partly motivated by the strong correlation between free, potential and strain energy (**Table 1** and **Fig. S22**), which presents a MAE smaller than 0.8 kJ/mol.

For the 300 K temperature considered here, it turns out that using potential or strain energy in an analogous manner to how we used free energy data in **Fig. 5** to obtain **Fig. 6**, yields similar results. That is, stability

Table 1. Correlation metrics between different energies calculated for all MOFs in our database. R and S are the Pearson and Spearman Ranking correlation coefficients, respectively, and (MAE) is the mean absolute error.

Correlation	R	S	MAE (kJ/mol)
F_{FL} vs. $(F_{QHA})_{adjusted}$	0.997	0.981	1.2
F_{FL} vs. U_{300K}	0.999	0.992	0.5
F_{FL} vs. U_{0K}	0.998	0.984	0.8

regions for MOFs could also be identified on the basis of strain or potential energy (**Fig. S23**), and which could be used to discard a roughly similar number of MOFs deemed synthetically unlikely. However, as noted earlier, a more challenging task (since smaller free energy differences are involved) is to be able to discard low-energy MOFs that have a more stable isomorph.

From **Section 3.4** we already know that on the basis of having the lowest free energy, we would correctly predict the experimentally observed isomorph for 26 (out of 32 isomorph series) to be synthetically accessible, and would need a 2.5% tolerance to classify all experimentally observed isomorphs as synthetically accessible (at the expense of including false positives). If we were to use potential or strain energy as a proxy for free energy, we would correctly predict the experimentally observed isomorph for 25 and 22 isomorph series, respectively. Thus, using potential

(strain) energy results in one (four) more false negatives than free energy. A tolerance of 3.3% would be needed to classify all the experimentally observed isomorphs in our database as synthetically accessible using potential energy, slightly higher than the tolerance required using free energy. However, a dramatically higher tolerance (93.0%) would be required when using strain energy. Therefore, using strain energy to identify synthetically accessible isomorphs would certainly result in significantly more false positives.

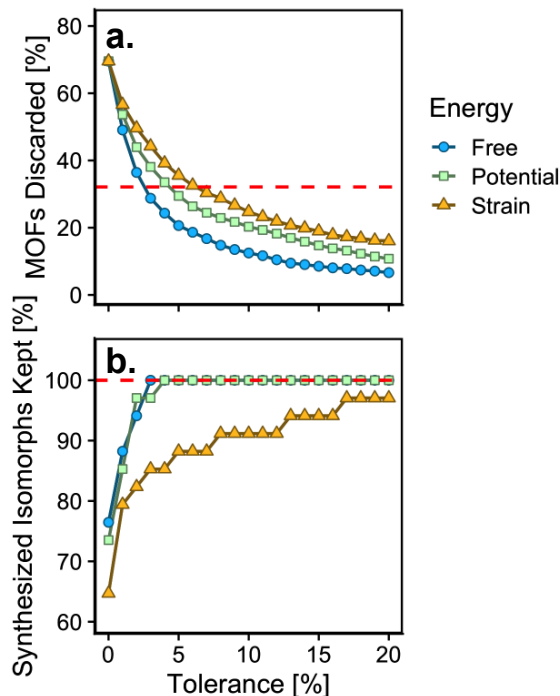


Figure 9. a) Percentage of MOFs that can be discarded because they are synthetically inaccessible as a function of tolerance, the dashed line corresponds to the percent of MOFs discarded at the tolerance for free energy (2.5%) that retains all of the experimentally observed isomorphs in our database. b) Percentage of synthetically accessible isomorphs that are kept as a function of tolerance, the dashed line corresponds to all 34 isomorphs. The tolerance is defined as being within a certain percent of the lowest free, potential or strain energy predicted within an isomorph series.

Of course, it would be possible to lower the tolerance when using strain or potential energy as a proxy for free energy, if one were to accept the occurrence of false negatives. As the balance between false negatives and false positives is a recurring subject in efforts to accelerate HTCS, we calculated the percentage of synthetically accessible isomorphs that would be kept in our database subset as a function of tolerance (Fig. 9b), as well as the percentage of MOFs that can be discarded as they have a more stable isomorph (Fig. 9a). A loose tolerance would result in fewer MOFs being discarded, fewer false negatives, and more false positives. A tight

tolerance allows for more MOFs to be discarded, more false negatives, and fewer false positives. At the tolerance required to retain all the synthetically accessible isomorphs in our database (i.e. an estimate of the tolerance required to eliminate the false negatives), the free energy metric enables the elimination of the largest number of MOFs (i.e. minimizes the number of false positives).

Based on the above observations, we conclude that free and potential energy would definitively outperform strain energy for finding stable isomorphs. Both free and potential energy require low tolerances (2.5 and 3.3%, respectively) to remove false negatives in our data, and would likely perform similarly in the vast majority of cases. Note that the discussion here has focused on calculations at 300 K. However, if one were to calculate free and potential energies at different temperatures, the number of correctly identified experimentally observed isomorphs may change (see Table S6). Interestingly, for the 32 isomorph series studied here, calculations at 300 K performed best at predicting the experimentally observed isomorphs among tested temperatures (100 K, 300 K, 350 K, 400 K).

To complete our discussion regarding the effectiveness of free, potential and strain energies to characterize isomorph stability, we identified the lowest free energy isomorph within each of the 975 isomorph series we found in our 8,500-MOF database. Only 797 (82%) and 759 (78%) were also the lowest potential and strain energy isomorph, respectively. Consequently, if we consider free energy to be the most accurate identifier of a synthetically accessible isomorphs, using potential or strain energy as a proxy for free energy may lead to some incorrect classifications.

CONCLUSIONS

In this contribution, we demonstrated the first large-scale calculation of MOF free energies, which we performed on a topologically diverse, computationally constructed 8500-MOF database. Based on our results, we recommended using the Frenkel-Ladd (FL) method over the quasi-harmonic approximation for molecular dynamics free energy calculations in diverse MOFs. While we discussed our results based on calculations with the (MOF-specific) UFF4MOF forcefield, generic forcefields UFF and Dreiding performed surprisingly well at replicating trends in free energy observed with UFF4MOF. Thermodynamic stability of the “activated” (i.e. solvent-free) MOF crystals at 300 K was found to correlate strongly (albeit not perfectly) with MOF synthesizability based on the following key observations: (i) the 54 previously synthesized MOFs identified in our database all clustered below a free energy threshold of 4.4 kJ/mol, and (ii) in 80% of 32 studied isomorph series, the experimentally observed isomorph had the lowest free energy within its series. As no MOFs with “high” free energy were found to be experimentally reported, “low” free energy seems a necessary, albeit sometimes

insufficient, condition for a MOF to be synthetically accessible. Instances where low free energy is insufficient to warrant synthetic accessibility may be due to solvents, modulators, and kinetic considerations that were not inspected here, but that merit closer inspection in the future. Nonetheless, high throughput calculation of MOF thermodynamic stability seems plausible and a key first step towards gauging the synthetic likelihood of MOFs in computational databases.

ASSOCIATED CONTENT

Supporting Information.

Example LAMMPS inputs for calculating the FL path free energy of a MOF from our database.
Additional figures and details on simulation methodology.
Additional figures and discussion relating to the comparison of the QHA and FL path methods.
Additional figures relating to the sensitivity of free energy to force field selection
Additional free energy versus MOF property relationships.
Additional discussion, data, and figures relating to the composition of the experimentally observed isomorphs in our database and isomorph stability calculations.

AUTHOR INFORMATION

Corresponding Author

* dgomezgualdron@mines.edu

Funding Sources

NSF CAREER

ACKNOWLEDGMENTS

D.A.G.-G. acknowledges funding from NSF CAREER (CBET 1846707). Simulations were made possible by the Mio supercomputer cluster at Colorado School of Mines

REFERENCES

- (1) Wei, X.; Wang, S.; Hua, Z.; Chen, L.; Shi, J. Metal–Organic Framework Nanosheet Electrocatalysts for Efficient H₂ Production from Methanol Solution: Methanol-Assisted Water Splitting or Methanol Reforming? *ACS Appl. Mater. Interfaces* **2018**, *10*, 25422–25428.
- (2) Schweitzer, B.; Archuleta, C.; Seong, B.; Anderson, R.; Gómez-Gualdrón, D. A. Electronic Effects Due to Organic Linker-Metal Surface Interactions: Implications on Screening of MOF-Encapsulated Catalysts. *Phys. Chem. Chem. Phys.* **2020**, *22*, 2475–2487.
- (3) Duan, J.; Chen, S.; Zhao, C. Ultrathin Metal–Organic Framework Array for Efficient Electrocatalytic Water Splitting. *Nat. Commun.* **2017**, *8*, 15341.
- (4) Yang, D.; Ortuño, M. A.; Bernales, V.; Cramer, C. J.; Gagliardi, L.; Gates, B. C. Structure and Dynamics of Zr₆O₈ Metal–Organic Framework Node Surfaces Probed with Ethanol Dehydration as a Catalytic Test Reaction. *J. Am. Chem. Soc.* **2018**, *140*, 3751–3759.
- (5) Herm, Z. R.; Bloch, E. D.; Long, J. R. Hydrocarbon Separations in Metal–Organic Frameworks. *Chem. Mater.* **2014**, *26*, 323–338.
- (6) Herm, Z. R.; Wiers, B. M.; Mason, J. A.; van Baten, J. M.; Hudson, M. R.; Zajdel, P.; Brown, C. M.; Masciocchi, N.; Krishna, R.; Long, J. R. Separation of Hexane Isomers in a Metal–Organic Framework with Triangular Channels. *Science* **2013**, *340*, 960–964.
- (7) Venna, S. R.; Carreon, M. A. Highly Permeable Zeolite Imidazolate Framework-8 Membranes for CO₂/CH₄ Separation. *J. Am. Chem. Soc.* **2010**, *132*, 76–78.
- (8) Chung, Y. G.; Gómez-Gualdrón, D. A.; Li, P.; Leperi, K. T.; Deria, P.; Zhang, H.; Vermeulen, N. A.; Stoddart, J. F.; You, F.; Hupp, J. T.; et al. In Silico Discovery of Metal–Organic Frameworks for Precombustion CO₂ Using a Genetic Algorithm. *Sci. Adv.* **2016**, *2*, e1600909.
- (9) Xie, Z.; Ma, L.; deKrafft, K. E.; Jin, A.; Lin, W. Porous Phosphorescent Coordination Polymers for Oxygen Sensing. *J. Am. Chem. Soc.* **2010**, *132*, 922–923.
- (10) Lan, A.; Li, K.; Wu, H.; Olson, D. H.; Emge, T. J.; Ki, W.; Hong, M.; Li, J. A Luminescent Microporous Metal–Organic Framework for the Fast and Reversible Detection of High Explosives. *Angew. Chemie Int. Ed.* **2009**, *48*, 2334–2338.
- (11) Gustafson, J. A.; Wilmer, C. E. Computational Design of Metal–Organic Framework Arrays for Gas Sensing: Influence of Array Size and Composition on Sensor Performance. *J. Phys. Chem. C* **2017**, *121*, 6033–6038.
- (12) Li, Q.-L.; Wang, J.-P.; Liu, W.-C.; Zhuang, X.-Y.; Liu, J.-Q.; Fan, G.-L.; Li, B.-H.; Lin, W.-N.; Man, J.-H. A New (4,8)-Connected Topological MOF as Potential Drug Delivery. *Inorg. Chem. Commun.* **2015**, *55*, 8–10.
- (13) Roth Stefaniak, K.; Epley, C. C.; Novak, J. J.; McAndrew, M. L.; Cornell, H. D.; Zhu, J.; McDaniel, D. K.; Davis, J. L.; Allen, I. C.; Morris, A. J.; et al. Photo-Triggered Release of 5-Fluorouracil from a MOF Drug Delivery Vehicle. *Chem. Commun.* **2018**, *54*, 7617–7620.
- (14) Hartlieb, K. J.; Ferris, D. P.; Holcroft, J. M.; Kandela, I.; Stern, C. L.; Nassar, M. S.; Botros, Y. Y.; Stoddart, J. F. Encapsulation of Ibuprofen in CD-MOF and Related Bioavailability Studies. *Mol. Pharm.* **2017**, *14*, 1831–1839.
- (15) Chen, X.; Tong, R.; Shi, Z.; Yang, B.; Liu, H.;

- Ding, S.; Wang, X.; Lei, Q.; Wu, J.; Fang, W. MOF Nanoparticles with Encapsulated Autophagy Inhibitor in Controlled Drug Delivery System for Antitumor. *ACS Appl. Mater. Interfaces* **2018**, *10*, 2328–2337.
- (16) Gomez-Gualdrón, D. A.; Gutov, O. V.; Krungleviciute, V.; Borah, B.; Mondloch, J. E.; Hupp, J. T.; Yildirim, T.; Farha, O. K.; Snurr, R. Q. Computational Design of Metal–Organic Frameworks Based on Stable Zirconium Building Units for Storage and Delivery of Methane. *Chem. Mater.* **2014**, *26*, 5632–5639.
- (17) Gómez-Gualdrón, D. A.; Colón, Y. J.; Zhang, X.; Wang, T. C.; Chen, Y.-S.; Hupp, J. T.; Yildirim, T.; Farha, O. K.; Zhang, J.; Snurr, R. Q. Evaluating Topologically Diverse Metal–Organic Frameworks for Cryo-Adsorbed Hydrogen Storage. *Energy Environ. Sci.* **2016**, *9*, 3279–3289.
- (18) Zheng, J.; Tian, J.; Wu, D.; Gu, M.; Xu, W.; Wang, C.; Gao, F.; Engelhard, M. H.; Zhang, J.-G.; Liu, J.; et al. Lewis Acid–Base Interactions between Polysulfides and Metal Organic Framework in Lithium Sulfur Batteries. *Nano Lett.* **2014**, *14*, 2345–2352.
- (19) de Combarieu, G.; Morcrette, M.; Millange, F.; Guillou, N.; Cabana, J.; Grey, C. P.; Margiolaki, I.; Férey, G.; Tarascon, J.-M. Influence of the Benzoquinone Sorption on the Structure and Electrochemical Performance of the MIL-53(Fe) Hybrid Porous Material in a Lithium-Ion Battery. *Chem. Mater.* **2009**, *21*, 1602–1611.
- (20) Wilmer, C. E.; Leaf, M.; Lee, C. Y.; Farha, O. K.; Hauser, B. G.; Hupp, J. T.; Snurr, R. Q. Large-Scale Screening of Hypothetical Metal–Organic Frameworks. *Nat. Chem.* **2011**, *4*, 83–89.
- (21) Yan, T.; Lan, Y.; Tong, M.; Zhong, C. Screening and Design of Covalent Organic Framework Membranes for CO₂/CH₄ Separation. *ACS Sustain. Chem. Eng.* **2019**, *7*, 1220–1227.
- (22) Kim, J.; Abouelnasr, M.; Lin, L. C.; Smit, B. Large-Scale Screening of Zeolite Structures for CO₂ Membrane Separations. *J. Am. Chem. Soc.* **2013**, *135*, 7545–7552.
- (23) Chung, Y. G.; Camp, J.; Haranczyk, M.; Sikora, B. J.; Bury, W.; Krungleviciute, V.; Yildirim, T.; Farha, O. K.; Sholl, D. S.; Snurr, R. Q. Computation-Ready, Experimental Metal–Organic Frameworks: A Tool to Enable High-Throughput Screening of Nanoporous Crystals. *Chem. Mater.* **2014**, *26*, 6185–6192.
- (24) Chung, Y. G.; Haldoupis, E.; Bucior, B. J.; Haranczyk, M.; Lee, S.; Zhang, H.; Vogiatzis, K. D.; Milisavljevic, M.; Ling, S.; Camp, J. S.; et al. Advances, Updates, and Analytics for the Computation-Ready, Experimental Metal–Organic Framework Database: CoRE MOF 2019. *J. Chem. Eng. Data* **2019**, *64*, 5985–5998.
- (25) Baerlocher, C.; McCusker, L. B. Database of Zeolite Structures. <http://www.iza-structure.org/databases/> (accessed Mar. 2020)
- (26) Martin, R. L.; Willems, T. F.; Lin, L. C.; Kim, J.; Swisher, J. A.; Smit, B.; Haranczyk, M. Similarity-Driven Discovery of Zeolite Materials for Adsorption-Based Separations. *ChemPhysChem* **2012**, *13*, 3595–3597.
- (27) Sikora, B. J.; Wilmer, C. E.; Greenfield, M. L.; Snurr, R. Q. Thermodynamic Analysis of Xe/Kr Selectivity in over 137 000 Hypothetical Metal–Organic Frameworks. *Chem. Sci.* **2012**, *3*, 2217–2223.
- (28) Grazulis, S.; Chateigner, D.; Downs, R. T.; Yokochi, A. F. T.; Quiros, M.; Lutterotti, L.; Manakova, E.; Butkus, J.; Moeck, P.; Le Bail, A. Crystallography Open Database - an Open-Access Collection of Crystal Structures. *J. Appl. Crystallogr.* **2009**, *42*, 726–729.
- (29) Deem, M. W.; Pophale, R.; Cheeseman, P. A.; Earl, D. J. Computational Discovery of New Zeolite-Like Materials. *J. Phys. Chem. C* **2009**, *113*, 21353–21360.
- (30) Banerjee, D.; Simon, C. M.; Plonka, A. M.; Motkuri, R. K.; Liu, J.; Chen, X.; Smit, B.; Parise, J. B.; Haranczyk, M.; Thallapally, P. K. Metal–Organic Framework with Optimally Selective Xenon Adsorption and Separation. *Nat. Commun.* **2016**, *7*, ncomms11831.
- (31) Addicoat, M. A.; Couprie, D. E.; Heine, T. AuToGraFS: Automatic Topological Generator for Framework Structures. *J. Phys. Chem. A* **2014**, *118*, 9607–9614.
- (32) Boyd, P. G.; Woo, T. K. A Generalized Method for Constructing Hypothetical Nanoporous Materials of Any Net Topology from Graph Theory. *CrystEngComm* **2016**, *18*, 3777–3792.
- (33) Anderson, R.; Gómez-Gualdrón, D. A. Increasing Topological Diversity during Computational “Synthesis” of Porous Crystals: How and Why. *CrystEngComm* **2019**, *21*, 1653–1665.
- (34) Colón, Y. J.; Gómez-Gualdrón, D. A.; Snurr, R. Q. Topologically Guided, Automated Construction of Metal–Organic Frameworks and Their Evaluation for Energy-Related Applications. *Cryst. Growth Des.* **2017**, *17*, 5801–5810.
- (35) Van Vleet, M. J.; Weng, T.; Li, X.; Schmidt, J. R. In Situ, Time-Resolved, and Mechanistic Studies of Metal–Organic Framework Nucleation and Growth. *Chem. Rev.* **2018**, *118*, 3681–3721.
- (36) Draznieks, C. M.; Newsam, J. M.; Gorman, A. M.; Freeman, C. M.; Férey, G. De Novo Prediction of Inorganic Structures Developed

- through Automated Assembly of Secondary Building Units (AASBU Method). *Angew. Chem. Int. Ed. Engl.* **2000**, *39*, 2270–2275.
- (37) Mellot-Draznieks, C.; Dutour, J.; Férey, G. Hybrid Organic–Inorganic Frameworks: Routes for Computational Design and Structure Prediction. *Angew. Chemie Int. Ed.* **2004**, *43*, 6290–6296.
- (38) Yoneya, M.; Tsuzuki, S.; Aoyagi, M. Simulation of Metal–Organic Framework Self-Assembly. *Phys. Chem. Chem. Phys.* **2015**, *17*, 8649–8652.
- (39) Biswal, D.; Kusalik, P. G. Molecular Simulations of Self-Assembly Processes in Metal–Organic Frameworks: Model Dependence. *J. Chem. Phys.* **2017**, *147*, 44702.
- (40) Biswal, D.; Kusalik, P. G. Probing Molecular Mechanisms of Self-Assembly in Metal–Organic Frameworks. *ACS Nano* **2017**, *11*, 258–268.
- (41) Kollias, L.; Cantu, D. C.; Tubbs, M. A.; Rousseau, R.; Glezakou, V.-A.; Salvalaglio, M. Molecular Level Understanding of the Free Energy Landscape in Early Stages of Metal–Organic Framework Nucleation. *J. Am. Chem. Soc.* **2019**, *141*, 6073–6081.
- (42) Colón, Y. J.; Guo, A. Z.; Antony, L. W.; Hoffmann, K. Q.; de Pablo, J. J. Free Energy of Metal–Organic Framework Self-Assembly. *J. Chem. Phys.* **2019**, *150*, 104502.
- (43) Deria, P.; Gómez-Gualdrón, D. A.; Hod, I.; Snurr, R. Q.; Hupp, J. T.; Farha, O. K. Framework-Topology-Dependent Catalytic Activity of Zirconium-Based (Porphinato)Zinc(II) MOFs. *J. Am. Chem. Soc.* **2016**, *138*, 14449–14457.
- (44) Li, M.; Li, D.; O’Keeffe, M.; Yaghi, O. M. Topological Analysis of Metal–Organic Frameworks with Polytopic Linkers and/or Multiple Building Units and the Minimal Transitivity Principle. *Chem. Rev.* **2014**, *114*, 1343–1370.
- (45) O’Keeffe, M.; Yaghi, O. M. Deconstructing the Crystal Structures of Metal–Organic Frameworks and Related Materials into Their Underlying Nets. *Chem. Rev.* **2012**, *112*, 675–702.
- (46) Ockwig, N. W.; Delgado-friedrichs, O. Reticular Chemistry : Occurrence and Taxonomy of Nets and Grammar for the Design of Frameworks. **2005**, *38*, 176–182.
- (47) Chui, S. S.-Y.; Lo, S. M.-F.; Charmant, J. P. H.; Orpen, A. G.; Williams, I. D. A Chemically Functionalizable Nanoporous Material $[\text{Cu}_3(\text{TMA})_2(\text{H}_2\text{O})_3]_n$. *Science* **1999**, *283*, 1148–1150.
- (48) Chen, B.; Eddaoudi, M.; Hyde, S. T.; O’Keeffe, M.; Yaghi, O. M. Interwoven Metal–Organic Framework on a Periodic Minimal Surface with Extra-Large Pores. *Science* **2001**, *291*, 1021–1023.
- (49) Amirjalayer, S.; Tafipolsky, M.; Schmid, R. Exploring Network Topologies of Copper Paddle Wheel Based Metal–Organic Frameworks with a First-Principles Derived Force Field. *J. Phys. Chem. C* **2011**, *115*, 15133–15139.
- (50) Impeng, S.; Cedeno, R.; Dürholt, J. P.; Schmid, R.; Bureekaew, S. 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.
- (51) Cai, Y.; Kulkarni, A. R.; Huang, Y.-G.; Sholl, D. S.; Walton, K. S. Control of Metal–Organic Framework Crystal Topology by Ligand Functionalization: Functionalized HKUST-1 Derivatives. *Cryst. Growth Des.* **2014**, *14*, 6122–6128.
- (52) Arhangelskis, M.; Katsenis, A. D.; Novendra, N.; Akimbekov, Z.; Gandrath, D.; Marrett, J. M.; Ayoub, G.; Morris, A. J.; Farha, O. K.; Frišić, T.; et al. Theoretical Prediction and Experimental Evaluation of Topological Landscape and Thermodynamic Stability of a Fluorinated Zeolitic Imidazolate Framework. *Chem. Mater.* **2019**, *31*, 3777–3783.
- (53) Chen, Z.; Hanna, S. L.; Redfern, L. R.; Alezi, D.; Islamoglu, T.; Farha, O. K. Reticular Chemistry in the Rational Synthesis of Functional Zirconium Cluster-Based MOFs. *Coord. Chem. Rev.* **2019**, *386*, 32–49.
- (54) Islamoglu, T.; Otake, K.; Li, P.; Buru, C. T.; Peters, A. W.; Akpınar, I.; Garibay, S. J.; Farha, O. K. Revisiting the Structural Homogeneity of NU-1000, a Zr-Based Metal–Organic Framework. *CrystEngComm* **2018**, *20*, 5913–5918.
- (55) Webber, T. E.; Liu, W.-G.; Desai, S. P.; Lu, C. C.; Truhlar, D. G.; Penn, R. L. Role of a Modulator in the Synthesis of Phase-Pure NU-1000. *ACS Appl. Mater. Interfaces* **2017**, *9*, 39342–39346.
- (56) Garibay, S. J.; Iordanov, I.; Islamoglu, T.; DeCoste, J. B.; Farha, O. K. Synthesis and Functionalization of Phase-Pure NU-901 for Enhanced CO₂ Adsorption: The Influence of a Zirconium Salt and Modulator on the Topology and Phase Purity. *CrystEngComm* **2018**, *20*, 7066–7070.
- (57) Wang, P.; Luo, L.; Fan, J.; Lv, G.-C.; Song, Y.; Sun, W.-Y. Syntheses, Structures, Sorption and Magnetic Properties of Copper(II) Frameworks with Varied Topologies. *Microporous Mesoporous Mater.* **2013**, *175*, 116–124.
- (58) Forster, P. M.; Burbank, A. R.; Livage, C.; Férey, G.; Cheetham, A. K. The Role of

- Temperature in the Synthesis of Hybrid Inorganic–Organic Materials: The Example of Cobalt Succinates. *Chem. Commun.* **2004**, *4*, 368–369.
- (59) Sun, Y.-X.; Sun, W.-Y. Influence of Temperature on Metal–Organic Frameworks. *Chinese Chem. Lett.* **2014**, *25*, 823–828.
- (60) Gee, J. A.; Sholl, D. S. Characterization of the Thermodynamic Stability of Solvated Metal–Organic Framework Polymorphs Using Molecular Simulations. *J. Phys. Chem. C* **2013**, *117*, 20636–20642.
- (61) Hoja, J.; Reilly, A. M.; Tkatchenko, A. First-Principles Modeling of Molecular Crystals: Structures and Stabilities, Temperature and Pressure. *WIREs Comput. Mol. Sci.* **2017**, *7*, e1294.
- (62) Reilly, A. M.; Tkatchenko, A. Role of Dispersion Interactions in the Polymorphism and Entropic Stabilization of the Aspirin Crystal. *Phys. Rev. Lett.* **2014**, *113*, 55701.
- (63) Nyman, J.; Day, G. M. Static and Lattice Vibrational Energy Differences between Polymorphs. *CrystEngComm* **2015**, *17*, 5154–5165.
- (64) Rossi, M.; Gasparotto, P.; Ceriotti, M. Anharmonic and Quantum Fluctuations in Molecular Crystals: A First-Principles Study of the Stability of Paracetamol. *Phys. Rev. Lett.* **2016**, *117*, 115702.
- (65) Schneider, E.; Vogt, L.; Tuckerman, M. E. Exploring Polymorphism of Benzene and Naphthalene with Free Energy Based Enhanced Molecular Dynamics. *Acta Crystallogr. Sect. B* **2016**, *72*, 542–550.
- (66) Reilly, A. M.; Tkatchenko, A. Seamless and Accurate Modeling of Organic Molecular Materials. *J. Phys. Chem. Lett.* **2013**, *4*, 1028–1033.
- (67) Case, D. H.; Campbell, J. E.; Bygrave, P. J.; Day, G. M. Convergence Properties of Crystal Structure Prediction by Quasi-Random Sampling. *J. Chem. Theory Comput.* **2016**, *12*, 910–924.
- (68) Curtis, F.; Li, X.; Rose, T.; Vázquez-Mayagoitia, Á.; Bhattacharya, S.; Ghiringhelli, L. M.; Marom, N. GAtor: A First-Principles Genetic Algorithm for Molecular Crystal Structure Prediction. *J. Chem. Theory Comput.* **2018**, *14*, 2246–2264.
- (69) Frenkel, D.; Ladd, A. J. C. 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.
- (70) Aragones, J. L.; Valeriani, C.; Vega, C. Note: Free Energy Calculations for Atomic Solids through the Einstein Crystal/Molecule Methodology Using GROMACS and LAMMPS. *J. Chem. Phys.* **2012**, *137*, 146101.
- (71) Freitas, R.; Asta, M.; de Koning, M. Nonequilibrium Free-Energy Calculation of Solids Using LAMMPS. *Comput. Mater. Sci.* **2016**, *112*, 333–341.
- (72) Dureckova, H.; Krykunov, M.; Aghaji, M. Z.; Woo, T. K. Robust Machine Learning Models for Predicting High CO₂ Working Capacity and CO₂/H₂ Selectivity of Gas Adsorption in Metal Organic Frameworks for Precombustion Carbon Capture. *J. Phys. Chem. C* **2019**, *123*, 4133–4139.
- (73) Anderson, R.; Seong, B.; Peterson, Z.; Stevanak, M.; Carreon, M. A.; Gómez-Gualdrón, D. A. Exploiting Hydrophobicity and Hydrophilicity in Nanopores as a Design Principle for “Smart” MOF Microtanks for Methane Storage. *Mol. Syst. Des. Eng.* **2020**, *5*, 166–176.
- (74) Ahmed, A.; Seth, S.; Purewal, J.; Wong-Foy, A. G.; Veenstra, M.; Matzger, A. J.; Siegel, D. J. Exceptional Hydrogen Storage Achieved by Screening Nearly Half a Million Metal–Organic Frameworks. *Nat. Commun.* **2019**, *10*, 1568.
- (75) Anderson, R.; Biong, A.; Gómez-Gualdrón, D. A. Adsorption Isotherm Predictions for Multiple Molecules in MOFs Using the Same Deep Learning Model. *J. Chem. Theory Comput.* **2020**, *16*, 1271–1283.
- (76) Plimpton, S. Fast Parallel Algorithms for Short-Range Molecular Dynamics. *J. Comput. Phys.* **1995**, *117*, 1–19.
- (77) Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. UFF, a Full Periodic Table Force Field for Molecular Mechanics and Molecular Dynamics Simulations. *J. Am. Chem. Soc.* **1992**, *114*, 10024–10035.
- (78) Mayo, S. L.; Olafson, B. D.; Goddard, W. A. DREIDING: A Generic Force Field for Molecular Simulations. *J. Phys. Chem.* **1990**, *94*, 8897–8909.
- (79) Addicoat, M. A.; Vankova, N.; Akter, I. F.; Heine, T. Extension of the Universal Force Field to Metal–Organic Frameworks. *J. Chem. Theory Comput.* **2014**, *10*, 880–891.
- (80) Coupry, D. E.; Addicoat, M. A.; Heine, T. Extension of the Universal Force Field for Metal–Organic Frameworks. *J. Chem. Theory Comput.* **2016**, *12*, 5215–5225.
- (81) Boyd, P. G.; Moosavi, S. M.; Witman, M.; Smit, B. Force-Field Prediction of Materials Properties in Metal–Organic Frameworks. *J. Phys. Chem. Lett.* **2017**, *8*, 357–363.
- (82) Bitzek, E.; Koskinen, P.; Gähler, F.; Moseler, M.; Gumbusch, P. Structural Relaxation Made Simple. *Phys. Rev. Lett.* **2006**, *97*, 170201.
- (83) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling through Velocity

- Rescaling. *J. Chem. Phys.* **2007**, *126*, 14101.
- (84) Vasileiadis, M. *Calculation of the Free Energy of Crystalline Solids*. Imperial College London: London, UK. **2013**.
- (85) Pastorino, C.; Gamba, Z. Free-Energy Calculations of Elemental Sulphur Crystals via Molecular Dynamics Simulations. *J. Chem. Phys.* **2003**, *119*, 2147–2154.
- (86) Gee, J. A.; Sholl, D. S. Characterization of the Thermodynamic Stability of Solvated Metal-Organic Framework Polymorphs Using Molecular Simulations. *J. Phys. Chem. C* **2013**, *117*, 20636–20642.
- (87) Lin, S. T.; Blanco, M.; Goddard, W. A. The Two-Phase Model for Calculating Thermodynamic Properties of Liquids from Molecular Dynamics: Validation for the Phase Diagram of Lennard-Jones Fluids. *J. Chem. Phys.* **2003**, *119*, 11792–11805.
- (88) Camiola, V. D.; Tozzini, V. Collective Mode Mining from Molecular Dynamics Simulations: A Comparative Approach. *Int. J. Comput. Methods* **2018**, *15*, 1–18.
- (89) Frenkel, Daan; Smith, B. *Understanding Molecular Simulations From Algorithms to Applications*; Academic Press, **2002**.
- (90) de Koning, M. Optimizing the Driving Function for Nonequilibrium Free-Energy Calculations in the Linear Regime: A Variational Approach. *J. Chem. Phys.* **2005**, *122*, 104106.
- (91) Watanabe, M.; Reinhardt, W. P. Direct Dynamical Calculation of Entropy and Free Energy by Adiabatic Switching. *Phys. Rev. Lett.* **1990**, *65*, 3301–3304.
- (92) Nowak, P.; Stoessel, J. P. Absolute Free Energies in Biomolecular Systems. *Macromolecules* **1990**, *23*, 1961–1965.
- (93) Hoffman, A. E. J.; Vanduyfhuys, L.; Nevjestic, I.; Wieme, J.; Rogge, S. M. J.; Depauw, H.; Van Der Voort, P.; Vrielinck, H.; Van Speybroeck, V. Elucidating the Vibrational Fingerprint of the Flexible Metal–Organic Framework MIL-53(Al) Using a Combined Experimental/Computational Approach. *J. Phys. Chem. C* **2018**, *122*, 2734–2746.
- (94) Krylov, A.; Vtyurin, A.; Petkov, P.; Senkovska, I.; Maliuta, M.; Bon, V.; Heine, T.; Kaskel, S.; Slyusareva, E. Raman Spectroscopy Studies of the Terahertz Vibrational Modes of a DUT-8 (Ni) Metal–Organic Framework. *Phys. Chem. Chem. Phys.* **2017**, *19*, 32099–32104.
- (95) Kamencek, T.; Bedoya-Martínez, N.; Zojer, E. Understanding Phonon Properties in Isoreticular Metal-Organic Frameworks from First Principles. *Phys. Rev. Mater.* **2019**, *3*, 1–16.
- (96) Civalieri, B.; Napoli, F.; Noël, Y.; Roetti, C.; Dovesi, R. Ab-Initio Prediction of Materials Properties with CRYSTAL: MOF-5 as a Case Study. *CrystEngComm* **2006**, *8*, 364–371.
- (97) Dürholt, J. P.; Fraux, G.; Coudert, F.-X.; Schmid, R. Ab Initio Derived Force Fields for Zeolitic Imidazolate Frameworks: MOF-FF for ZIFs. *J. Chem. Theory Comput.* **2019**, *15*, 2420–2432.
- (98) Bristow, J. K.; Tiana, D.; Walsh, A. Transferable Force Field for Metal–Organic Frameworks from First-Principles: BTW-FF. *J. Chem. Theory Comput.* **2014**, *10*, 4644–4652.
- (99) Bureekaew, S.; Amirjalayer, S.; Tafipolsky, M.; Spickermann, C.; Roy, T. K.; Schmid, R. MOF-FF - A Flexible First-Principles Derived Force Field for Metal-Organic Frameworks. *Phys. Status Solidi Basic Res.* **2013**, *250*, 1128–1141.
- (100) Rappé, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. UFF, a Full Periodic Table Force Field for Molecular Mechanics and Molecular Dynamics Simulations. *J. Am. Chem. Soc.* **1992**, *114*, 10024–10035.
- (101) Li, Y.; Yu, J.; Xu, R. Criteria for Zeolite Frameworks Realizable for Target Synthesis. *Angew. Chemie Int. Ed.* **2013**, *52*, 1673–1677.
- (102) Dawson, C. J.; Kapko, V.; Thorpe, M. F.; Foster, M. D.; Treacy, M. M. J. Flexibility As an Indicator of Feasibility of Zeolite Frameworks. *J. Phys. Chem. C* **2012**, *116*, 16175–16181.
- (103) Bushuev, Y. G.; Sastre, G. Feasibility of Pure Silica Zeolites. *J. Phys. Chem. C* **2010**, *114*, 19157–19168.
- (104) Simperler, A.; Foster, M. D.; Friedrichs, O. D.; Bell, R. G.; Almeida Paz, F. A.; Klinowski, J. Hypothetical Binodal Zeolitic Frameworks. *Acta Crystallogr. Sect. B Struct. Sci.* **2005**, *61*, 263–279.
- (105) Jablonka, K. M.; Ongari, D.; Moosavi, S. M.; Smit, B. Big-Data Science in Porous Materials: Materials Genomics and Machine Learning. **2020**, arXiv:2001.06728.arXiv.org e-Print archive, <https://arxiv.org/abs/2001.06728?context=cond-mat> (accessed Feb. 2020).
- (106) Mazur, M.; Wheatley, P. S.; Navarro, M.; Roth, W. J.; Polojij, M.; Mayoral, A.; Eliášová, P.; Nachtigall, P.; Čejka, J.; Morris, R. E. Synthesis of ‘Unfeasible’ Zeolites. *Nat. Chem.* **2016**, *8*, 58–62.
- (107) Bucior, B. J.; Rosen, A. S.; Haranczyk, M.; Yao, Z.; Ziebel, M. E.; Farha, O. K.; Hupp, J. T.; Siepmann, J. I.; Aspuru-Guzik, A.; Snurr, R. Q. Identification Schemes for Metal–Organic Frameworks To Enable Rapid Search and Cheminformatics Analysis. *Cryst. Growth Des.* **2019**, *19*, 6682–6697.
- (108) Lin, X.; Telepeni, I.; Blake, A. J.; Dailly, A.; Brown, C. M.; Simmons, J. M.; Zoppi, M.; Walker, G. S.; Thomas, K. M.; Mays, T. J.; et al. High Capacity Hydrogen Adsorption in Cu(II) Tetracarboxylate Framework Materials: The

Role of Pore Size, Ligand Functionalization, and Exposed Metal Sites. *J. Am. Chem. Soc.* **2009**, *131*, 2159–2171.

TOC Figure

