JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
pubs.acs.org/JACS
Article

Analogy Powered by Prediction and Structural Invariants: Computationally Led Discovery of a Mesoporous Hydrogen-Bonded Organic Cage Crystal

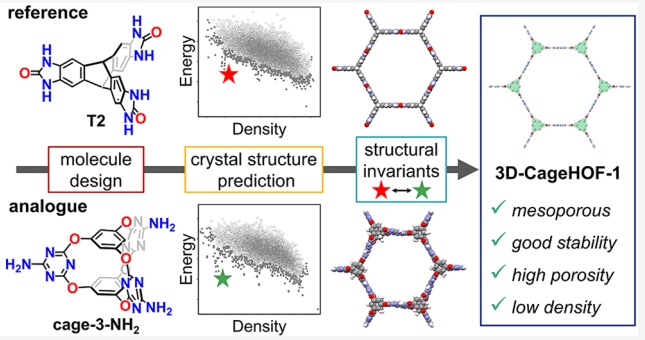
Qiang Zhu, Jay Johal, Daniel E. Widdowson, Zhongfu Pang, Boyu Li, Christopher M. Kane, Vitaliy Kurlin,* Graeme M. Day,* Marc A. Little,* and Andrew I. Cooper*

Cite This: *J. Am. Chem. Soc.* 2022, 144, 9893−9901

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Mesoporous molecular crystals have potential applications in separation and catalysis, but they are rare and hard to design because many weak interactions compete during crystallization, and most molecules have an energetic preference for close packing. Here, we combine crystal structure prediction (CSP) with structural invariants to continuously qualify the similarity between predicted crystal structures for related molecules. This allows isomorphous substitution strategies, which can be unreliable for molecular crystals, to be augmented by *a priori* prediction, thus leveraging the power of both approaches. We used this combined approach to discover a rare example of a low-density (0.54 g cm^{-3}) mesoporous hydrogen-bonded framework (HOF), **3D-CageHOF-1**. This structure comprises an organic cage (**Cage-3-NH$_2$**) that was predicted to form kinetically trapped, low-density polymorphs *via* CSP. Pointwise distance distribution structural invariants revealed five predicted forms of **Cage-3-NH$_2$** that are analogous to experimentally realized porous crystals of a chemically different but geometrically similar molecule, **T2**. More broadly, this approach overcomes the difficulties in comparing predicted molecular crystals with varying lattice parameters, thus allowing for the systematic comparison of energy−structure landscapes for chemically dissimilar molecules.

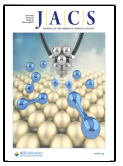


INTRODUCTION

The development of reliable methods for crystal structure prediction (CSP)[1,2] provides a powerful tool for the *ab initio* discovery of porous molecular crystals[3,4] and other functional organic solids. The lowest energy structures resulting from a computational search are assumed to be the most likely to be observed experimentally, and the probability of observing a particular structure can be related to its energy.[5] In the context of functional material discovery, CSP allows an assessment of a molecule's tendency to form crystal structures with an arrangement of molecules that provides the desired property, therefore guiding experimental workflows. Such guides can be expressed graphically through energy−structure−function maps that show the relationship between lattice energy and computed function for a molecule.[3,6]

However, predicting whether a molecule can form a stable porous crystal poses a challenge because such crystal packings often correspond to kinetically trapped, high-energy structures.[3,7−9] One approach is to apply *a priori* CSP coupled with computational methods for assessing solvent templating[8] and kinetic stability,[10] followed by the experimental screening of crystallization conditions for the molecules that are computed to have likely porous structures. An alternative, much more established approach makes use of analogy, substituting the molecule in a known porous structure with a related molecule that is expected, in principle, to be capable of adopting the same crystal packing arrangement. Such "isomorphous substitution" strategies are the basis for the reticular chemistry methods that have proved highly successful for designing porous bonded frameworks, such as metal−organic frameworks (MOFs)[11] and covalent organic frameworks (COFs).[12,13] By contrast, the crystallization of organic molecules is dependent on weaker and less directional intermolecular interactions, which makes crystallization outcomes much more difficult to predict, even for structurally similar molecular building blocks. Hence, there is an opportunity to augment isomorphous substitution methods

Received: March 10, 2022
Published: May 29, 2022



https://doi.org/10.1021/jacs.2c02653
J. Am. Chem. Soc. 2022, 144, 9893−9901

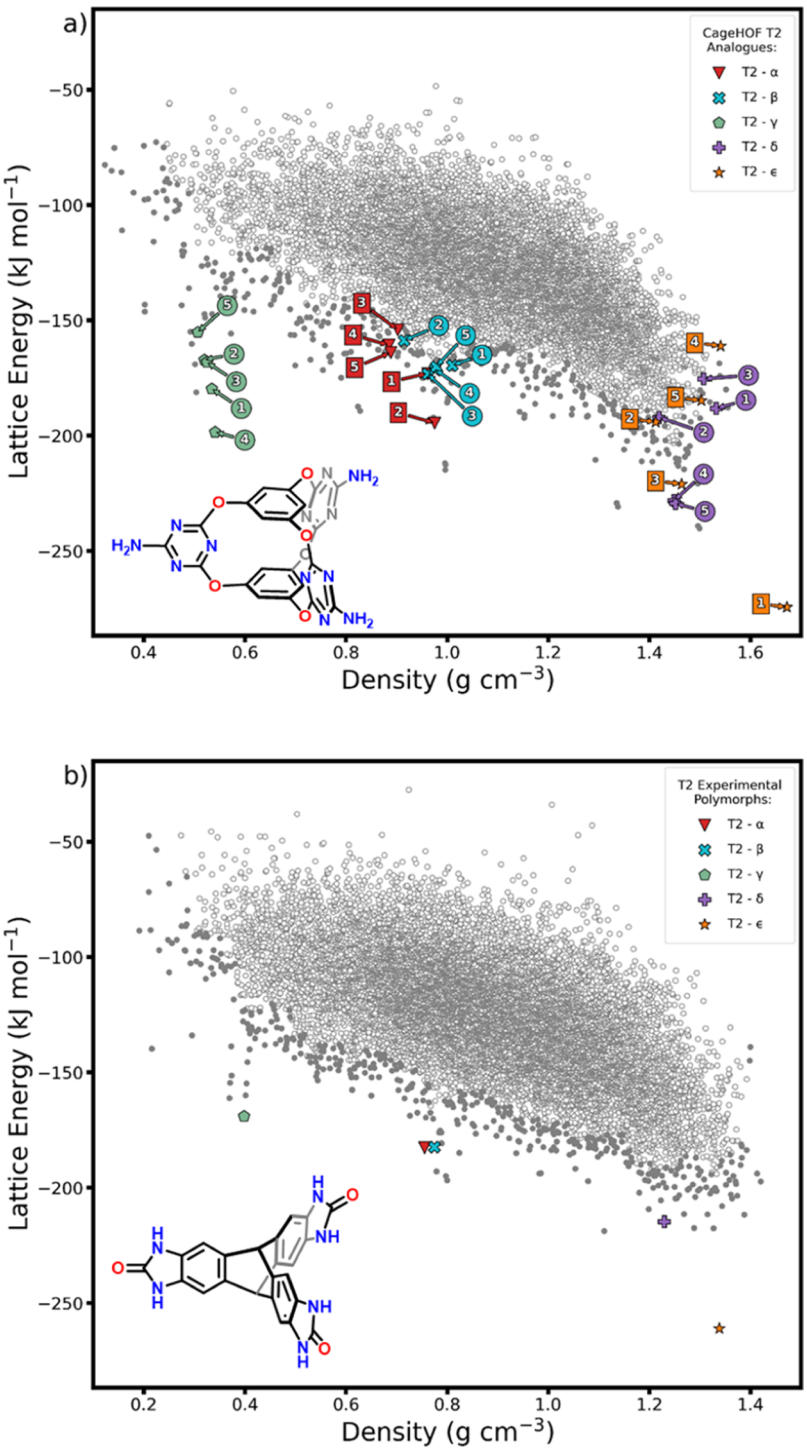


Figure 1. Energy−density distributions of the CSP structures of (a) **Cage-3-NH$_2$** and (b) **T2**. In both cases, the "leading edge" structures are shown as filled gray points to highlight those that are most likely to be found by experiment. The best matches to the five experimentally observed polymorphs of **T2** are indicated on the **T2** landscape, based on comparisons made using the COMPACK algorithm. On the **Cage-3-NH$_2$** landscape, we highlight the five closest analogues of each **T2** polymorph, based on isometry invariants (discussed below). These comparisons were restricted to the leading edge of the **Cage-3-NH$_2$** energy−density distribution, and their crystal structures are shown in Figure 2.

for organic molecules beyond simple analogy by applying CSP strategies.

Porous hydrogen-bonded frameworks (HOFs) have potential applications, such as gas storage, molecular separations, catalysis, sensing, solid electrolytes, and enzyme encapsulation.[14−17] Typically, HOFs are designed by appending hydrogen-bonding units to organic scaffolds to control their assembly.[14−16] An advantage of HOFs is that they are often highly crystalline. However, most hydrogen-bonding molecules have an energetic preference for close packing, and HOFs that do crystallize with low framework densities are frequently unstable to desolvation. More recently, some HOFs with excellent chemical and physical stabilities have been discovered,[9,18−21] although the isostructural series of porous HOFs

https://doi.org/10.1021/jacs.2c02653
J. Am. Chem. Soc. 2022, 144, 9893−9901

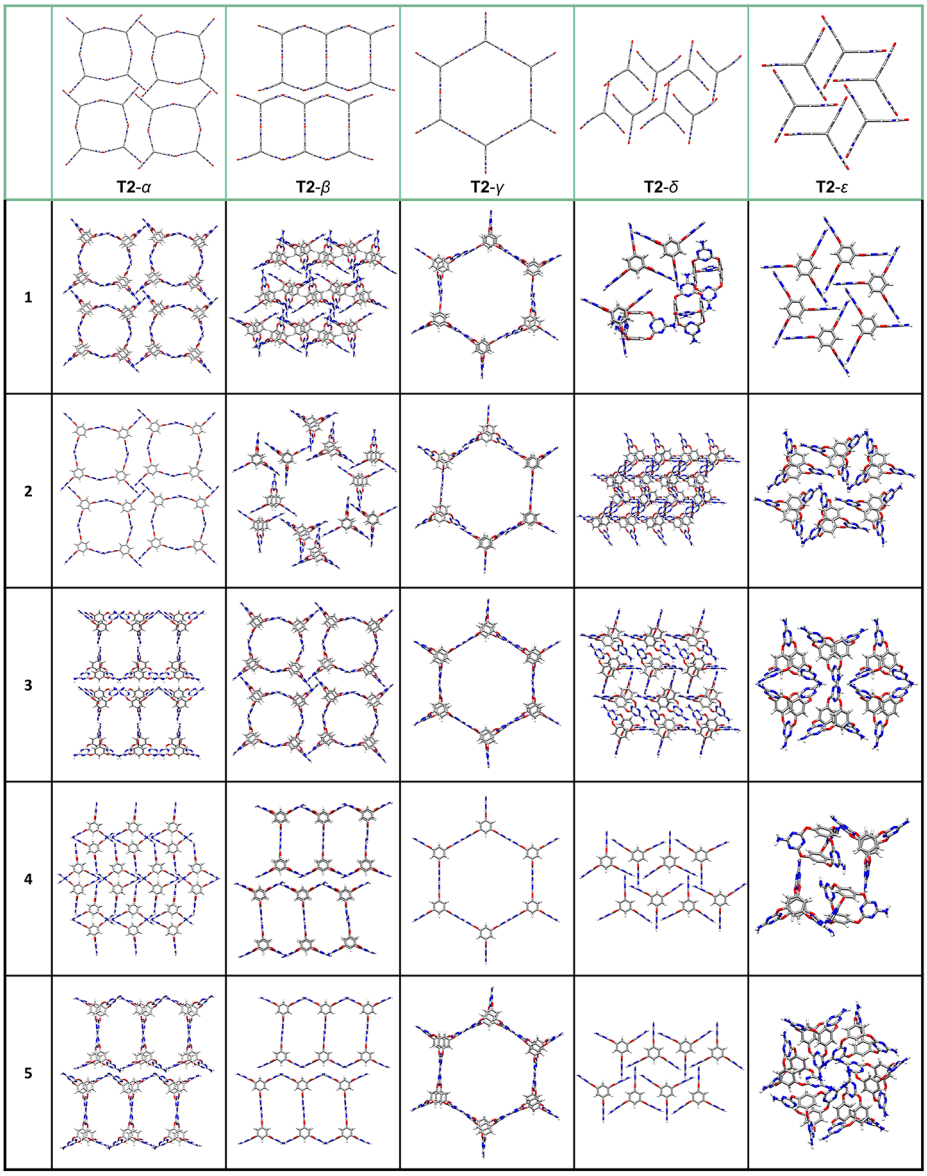


Figure 2. Crystal packing diagrams of predicted **Cage-3-NH$_2$** structures from the pre-DFTB-optimized CSP data set that were identified *via* structural invariants as the nearest neighbors to the **T2**-*α*, **T2**-*β*, **T2**-*γ*, **T2**-*δ*, and **T2**-*ε* polymorphs (from **1** to **5**, **1** being the nearest neighbor).

remain much rarer[14−16] than isoreticular MOFs[11,22] and COFs[12,13,23] for the reasons outlined above. A related challenge for HOFs with large mesopores is the tendency to form interpenetrated structures,[7,24−32] which have lower porosity levels.

A challenge for the coupled use of CSP with isomorphous substitution strategies is to define whether two phases of related but structurally dissimilar molecules are analogous or isostructural. Such comparisons are often simple to make by eye—for example, to say that two dissimilar molecules both pack as hexagonal, hydrogen-bonded nets—but it is harder to define these similarities in a formal, quantitative way. Common methods for comparing crystal structure similarity, such as root-mean-square deviation (RMSD) of atomic positions, break down when applied to dissimilar molecules. It is therefore challenging to provide a formal metric to define when the crystal packing of molecule A is "like" the crystal packing of dissimilar molecule B. This is a general challenge in supramolecular chemistry that goes beyond comparing crystal packings on CSP landscapes.

In this study, we use structural invariants—that is, geometry-based crystal descriptors that can continuously assess similarity—to compare predicted crystal structures of a trigonal cage molecule, **Cage-3-NH$_2$** (Figure 1a),[33] with five known polymorphs of a trigonal triptycene benzimidazolone molecule (**T2**, Figure 1b).[3,34] We chose **T2** because of its rich polymorphic behavior and its geometric similarity with **Cage-3-NH$_2$**. Specifically, we were interested in searching for a structural analogue of the lowest density (0.42 g cm^{-3}), highly porous polymorph, **T2**-*γ*, which has a Brunauer−Emmett−Teller (BET) surface area of 3425 m^2 g^{-1}.[3] Initially, we used CSP to predict the energy landscape of **Cage-3-NH$_2$**. We then used a new structural invariant[35,36] to continuously quantify the similarity between five polymorphs of **T2**[3,34] and the predicted crystal structures of **Cage-3-NH$_2$**. This new approach analyzes CSP results by comparison with structural analogues, and this helped us to identify a rare example of a

https://doi.org/10.1021/jacs.2c02653
J. Am. Chem. Soc. 2022, 144, 9893−9901

mesoporous HOF (**3D-CageHOF-1**) that was kinetically trapped on the **Cage-3-NH$_2$** energy landscape. This predicted HOF structure was subsequently prepared in the laboratory and has a low density, with 2.3 nm sized one-dimensional (1D) pores and an experimental BET surface area of 1750 m^2 g^{-1}. Our design strategy mimics the isoreticular approaches developed for MOFs[11,22] and COFs[13,23] by identifying chemically different building blocks that can form kinetically trapped materials with related porous structures.

Organic cage molecules with various topologies and cavity sizes are now synthetically accessible,[37,38] and their one-pot self-assembly can be screened using computational methods.[39] Here, we demonstrate a new strategy to predict the crystal structures of cage molecules and to design stable, porous HOFs, complementing isoreticular approaches used to design cage-based MOFs[40,41] and COFs.[42−44] Thinking beyond porous solids, the comparison of related CSP landscapes offers a formal method for the *a priori* design of functional materials by analogy, using known functional solids as the starting point.

RESULTS AND DISCUSSION

Solution processible organic cages are versatile three-dimensional (3D) building blocks for constructing extended materials[45,46] and close-packed porous crystals.[4] Appending organic cages with hydrogen-bonding units to modulate their 3D packing has only been investigated recently: the first cage-based HOF was reported by Han *et al.*,[33] who found that the triangular prism-shaped cage, **Cage-3-NH$_2$** (Figure 1a), formed HOF-19, which has a two-dimensional (2D) ladder-like structure (Figure S13). By extension, we were interested in exploring the crystallization behavior of **Cage-3-NH$_2$** to determine whether it could form other porous crystal structures, assisted by computational prediction.

CSP Calculations. The CSP energy−density landscape for **Cage-3-NH$_2$** is shown in Figure 1a. These initial calculations used quasi-random structure generation,[47] a rigid-molecule approximation, and an empirically parametrized intermolecular force field with atom-centered electrostatic multipoles[48] (see the Supporting Information, Section S1, Figures S1−S5, and Tables S1−S3).

As usual, density is correlated with energy: the lowest energy structures are among the densest. However, multiple spikes are apparent on the leading edge of the landscape. These spikes correspond to unusually low energy structures being predicted in several narrow density ranges. The presence of these spikes on CSP energy−density plots has, in several cases, anticipated the experimental discovery of kinetically trapped porous molecular crystals.[3,7,9] Recent computational work has shown that these spikes correspond to deep energy basins, separated from dense crystal packings by high energy barriers.[10]

The close-packed global energy minimum predicted structure of **Cage-3-NH$_2$** has a density of 1.67 g cm^{-3}, but prominent spikes on the CSP landscape are apparent at densities close to 1.34, 1.00, 0.83, and 0.54 g cm^{-3}. The overall structure of the crystal energy landscape, including the low-density spikes, is strongly reminiscent of that reported for **T2** (Figure 1b),[34] which has the same trigonal D_{3h} symmetry as **Cage-3-NH$_2$** and an analogous arrangement of hydrogen bond donors and acceptors at the ends of the three "arms" of the molecule. The spike at 1 g cm^{-3} on the **Cage-3-NH$_2$** energy−density plot was found to contain predicted structures with high similarity to the HOF-19 structure reported by Han *et al.*,[33] which is closely related to the β polymorph of **T2** (Figure 2, top row).[3] The visualization of the predicted structures along the leading edge of the CSP landscape of **Cage-3-NH$_2$** finds structures with similar crystal packing to those predicted for **T2** (Figures S4 and S5 and Table S3).[3]

When comparing CSP structures to the reported experimental crystal structure of HOF-19, we noted distortion of the molecular geometry away from the ideal D_{3h} geometry. Therefore, to improve the quality of the most important CSP structures, we re-optimized the 386 structures on the leading edge of the energy−density distribution using density functional theory-based tight binding (DFTB), which allowed for molecular flexibility within each crystal structure (Table S2, Figure S3). After re-optimization, one of the CSP structures provides a good match to the HOF-19 structure reported by Han *et al.*[33] For comparison, the **T2** leading edge structures were also re-optimized with DFTB. Energy−density distributions after re-optimization are shown in Figures S1 and S2 (Supporting Information).

Structural Invariants. We used a new isometry invariant to continuously quantify the similarity of the two crystal energy landscapes for **T2** and **Cage-3-NH$_2$**. Our aim was to determine whether structural invariants could identify analogous structures on the **Cage-3-NH$_2$** landscape that matched the four reported **T2** polymorphs, **T2**-α, **T2**-β, **T2**-γ, and **T2**-δ, plus a new densely packed **T2** polymorph, reported here for the first time, **T2**-ε, which was grown by sublimation at *ca.* 800−850 °C using a tube furnace under reduced pressure (~3.5 × 10^{-2} mbar) (see the Supporting Information, Section S3.1 for refinement and structural details). We note that such analogous crystal structures cannot be identified routinely in molecular crystal data sets because searches using unit cell dimensions or crystal packing similarities otherwise fail; for example, while both **Cage-3-NH$_2$** and **T2** are trigonal, **Cage-3-NH$_2$** has a different size and a different aspect ratio.

The structural invariant used here was the pointwise distance distribution (PDD) defined for a periodic set of points. Here, we used the center of mass of the molecules as the periodic points (see the Supporting Information, Section S2). The PDD of a periodic set (S) is obtained from the matrix $m \times k$ in which each row consists of ordered distances from a point (p) in a unit cell of S to a number (k) of its nearest neighbors in S. In the $m \times k$ matrix, the distance rows are lexicographically ordered. If any rows are identical, they collapse into a single row in the matrix, and a weighting is applied (Figure S8). The resulting PDD(S; k) is a weighted distribution of distance rows, independent of a unit cell,[36] invariant under isometry (composition of translations, rotations, and reflections). Here, we use the 100 nearest neighbors as the k value. The earth mover's distance was previously used to compare crystal compositions[49] and is now adapted to a continuous metric between PDDs (Tables S4 and S5), which is easier to compute than between complete isoset invariants.[50] The PDD is more robust and quicker to compute than past invariants,[51−53] which allowed it to be used to distinguish all the periodic crystals in the Cambridge Structural Database.[36]

In Figure 2, we highlight the five predicted **Cage-3-NH$_2$** crystal structures with the highest similarity from the leading edge of the CSP energy−density distribution, as measured by PDD invariants, to each of the five known polymorphs of **T2**. Using these geometrical comparisons, we find a strong correspondence between the two energy landscapes (Figure

https://doi.org/10.1021/jacs.2c02653
J. Am. Chem. Soc. 2022, 144, 9893−9901

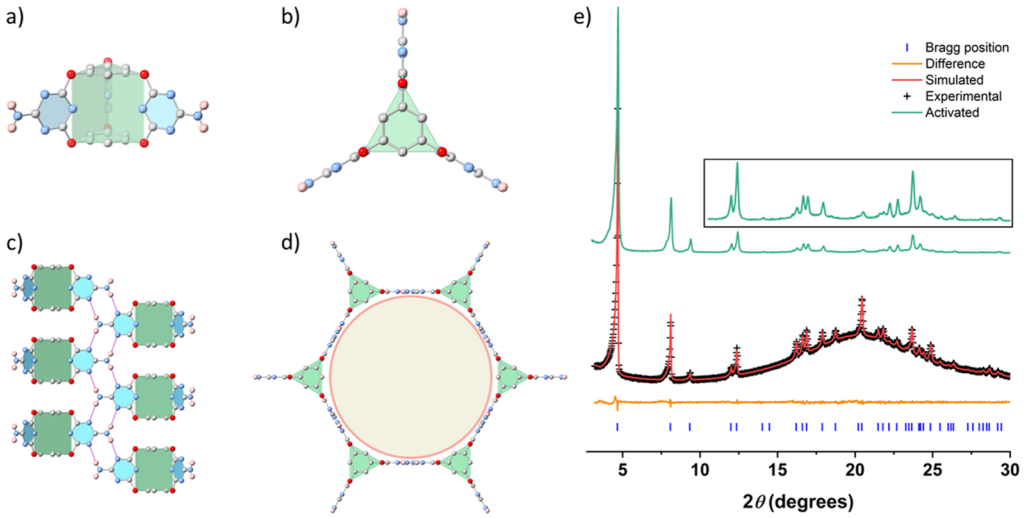


Figure 3. Crystal structures and stability of **3D-CageHOF-1**. Front view (a) and top view (b) of **Cage-3-NH$_2$** in the sc-XRD structure of **3D-CageHOF-1**; front view (c) and top view (d) of **3D-CageHOF-1**; single-crystal atom colors: C, gray; N, blue; O, red; and H, pink. H atoms on the phenyl rings are omitted for clarity. (e) PXRD pattern fitting of solvated **3D-CageHOF-1** with Pawley refinement (Cu-Kα) and the activated PXRD pattern of **3D-CageHOF-1** (green). In the insert, the peak intensities are multiplied by 10 from 2θ = 10.

1). The **Cage-3-NH$_2$** global energy minimum has a packing that is analogous to the close-packed global energy minimum **T2**-ε polymorph, and this is identified as the nearest neighbor (**1**, Figure 2) using the PDD invariant. In addition, the PDD invariant identified that the leading edge of the **Cage-3-NH$_2$** landscape is populated by structures that are analogous to the four known porous **T2** polymorphs: **T2**-α (structures **1**−**2**, Figure 2), **T2**-β (**4**−**5**, Figure 2), **T2**-γ (**1**−**5**, Figure 2), and **T2**-δ (**4**−**5**, Figure 2), albeit with more pronounced differences in crystal packing noted for **T2**-δ. There are 386 predicted structures on the leading edge of the **Cage-3-NH$_2$** energy landscape, and the PDD invariants locate packings that are isostructural to each of the five known **T2** polymorphs within the top five closest **Cage-3-NH$_2$** neighbors (Figure 2). This illustrates the power of this computationally inexpensive metric to automate structure comparisons. In the future, this method should allow the cross comparison of entire structure−energy landscapes.

The prediction of **T2** analogous structures on the CSP landscape of **Cage-3-NH$_2$** prompted us to pursue these structures experimentally. We were particularly interested in finding conditions that lead to the low-energy structure in the spike at 0.54 g cm^{-3}, which was identified as the nearest neighbor of **T2**-γ, and which is a non-interpenetrated 3D HOF with a 2.2 nm sized pore (Figure 2).

Crystallization Studies. CSP does not tell us how to access a particular predicted structure. Computational methods have been used to predict solvent effects on the crystallite size dependence of polymorph stability[54] and to screen solvent stabilization effects on CSP predicted crystal structures.[3,8] Although methods such as these could be used to guide the choice of crystallization conditions, such computational methods are expensive when applied to large CSP structure sets and were not used in this study. Instead, we studied the crystallization behavior of **Cage-3-NH$_2$** experimentally (see the Supporting Information, Section S5). Han *et al.*[33] reported that crystallizing **Cage-3-NH$_2$** from formic acid afforded HOF-19[33] (Figure S13). Here, we found that slowly diffusing diethyl ether into a solution of **Cage-3-NH$_2$** dissolved in formic acid afforded small needle-shaped crystals (Figure S14). Aniline (1 M equiv per **Cage-3-NH$_2$**) was used as a modulator to slow down the crystallization process and enable us to grow large enough crystals for single-crystal X-ray diffraction (sc-XRD) of what appeared to be the same material by powder XRD (PXRD, Figures S14 and S15). sc-XRD analysis revealed that the needle-shaped crystals had $P6_3/mmc$ symmetry in which the **Cage-3-NH$_2$** cage has a near-perfect triangular prismatic topology, and the −NH$_2$ groups hydrogen bond to triazine N atoms in neighboring cages. In total, each **Cage-3-NH$_2$** hydrogen bonds to six adjacent neighbors *via* this motif to form a 3D hydrogen-bonded network (**3D-CageHOF-1**, Figure 3, see the Supporting Information, Section S3.1 for refinement and structural details). The networks have a honeycomb shape with 2.2 nm wide hexagonal 1D channels directed along the crystallographic *c*-axis. The hydrogen bond distance was calculated to be 2.98 Å, and the shortest distance on the *c*-axis between adjacent cages was 3.44 Å, indicating that there are additional $\pi-\pi$ stacking interactions between the cages. Remarkably, the accessible void volume of **3D-CageHOF-1** is 72.5% of the unit cell volume, which is among the highest values reported for HOFs (Table S6).[14]

The CSP structure that was the target for experimental searches showed a strong match with the experimental structure, with 30 out of 30 molecules in common and an RMSD of atomic positions of 0.267 Å using CSD Mercury software[55] (Tables S1 and S2). As such, CSP led us to a new mesoporous polymorph of **Cage-3-NH$_2$** that has a much lower framework density than the previously reported HOF-19.

Characterization of 3D-CageHOF-1. HOFs with mesopores remain rare but are desirable as highly porous molecular crystals.[20,29,56] To measure the experimental porosity of **3D-CageHOF-1**, we scaled up the crystallization containing 1 M equiv of aniline per **Cage-3-NH$_2$**, and characterized the bulk material by PXRD (Figures S16). A Pawley refinement of the solvated material matched well with the simulated PXRD pattern of **3D-CageHOF-1** and indicated the sample was phase pure (Figure 3e, $P6_3/mmc$, $a = b = 21.87$ Å, $c = 8.01$ Å, $V = 3316$ Å^3, $R_{wp} = 1.41\%$, $R_p = 1.06\%$, $\chi^2 = 1.22$). After confirming the phase purity of the sample, we used supercritical CO$_2$ (scCO$_2$) to activate the crystal pores after

https://doi.org/10.1021/jacs.2c02653
J. Am. Chem. Soc. 2022, 144, 9893−9901

exchanging the crystallization with acetone. PXRD was again used to confirm the phase purity of **3D-CageHOF-1** after $scCO_2$ activation (Figures S17–S19). After degassing the $scCO_2$ activated **3D-CageHOF-1** under a dynamic vacuum at 25 °C for 15 h, we confirmed that crystallization solvents were removed from the pores by thermogravimetric analysis (TGA) (Figure S20) and NMR spectroscopy (Figure S21). Variable temperature PXRD was also used to analyze **3D-CageHOF-1**, which showed that the material remained highly crystalline and stable up to 120 °C (Figures S22 and S23), in agreement with the differential scanning calorimetry (DSC) result that no visible phase changes were observed before 165 °C (Figure S24). We attribute the stability to a combination of the 3D hydrogen-bonded network and the additional $\pi-\pi$ interactions between aromatic caps of the cages. In contrast to other HOFs, including triptycene-based HOFs with the same network topology,[16,26,27] the bulky cage cores in **3D-CageHOF-1** play an essential role in obtaining a non-interpenetrated 3D structure.

To gain further information about the crystallization behavior of **Cage-3-NH$_2$**, we carried out an *in situ* variable temperature PXRD experiment using the as-synthesized sample. We found that heating solvated crystals of **3D-CageHOF-1** from 298 to 363 K caused the structure to transform into the denser HOF-19 structure, with the PXRD patterns remaining the same after re-cooling the same to room temperature (Figure S25). However, activated crystals of **3D-CageHOF-1** were stable over the same temperature range (Figure S23), highlighting the importance of the crystallization solvent in facilitating the transformation of **3D-CageHOF-1** to HOF-19.

Gas Sorption Analysis. Encouraged by the apparent stability of the non-interpenetrated structure of **3D-CageHOF-1**, N_2 sorption at 77 K was used to measure its porosity. The sorption isotherms of **3D-CageHOF-1** had a type-IV shape with a sharp uptake at low relative pressure (<0.01), followed by a step at P/P_0 = 0.01–0.1 (Figures 4, S26–S29). The

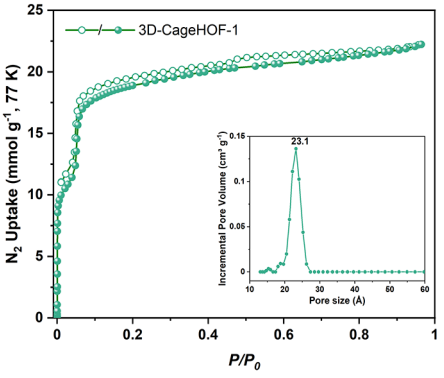


Figure 4. N_2 sorption isotherms of **3D-CageHOF-1** at 77 K and pore size distribution (inset).

desorption hysteresis loop indicates that **3D-CageHOF-1** is mesoporous. This conclusion is supported by the derived pore size distribution plot that displays a narrow peak at 23 Å, close to the value of 22 Å based on the sc-XRD structure (Figure 4, inset).

Although it is challenging to find HOF building blocks that form structures with large pores without interpenetration, this shows that the use of organic cages as building blocks is one promising strategy. To our knowledge, **3D-CageHOF-1** is the first non-interpenetrating mesoporous 3D HOF with a stable skeleton (Table S6). **3D-CageHOF-1** exhibits a BET surface area of 1750 $m^2\ g^{-1}$, which is about 2.5 times higher than HOF-19 (Figure S30).[33] Furthermore, **3D-CageHOF-1** has a much lower structure density (0.54 g cm^{-3}) than HOF-19 (1.00 g cm^{-3}) (Figure S30).

CONCLUSIONS

A cage-based HOF (**3D-CageHOF-1**) was discovered using computational CSP calculations and structural invariants to allow systematic comparisons between two CSP landscapes. In the mesoporous HOF phase, the bulky cage cores prevent network interpenetration, leading to the first example of a mesoporous non-interpenetrated 3D cage-HOF structure. **3D-CageHOF-1** has good structure stability after removing guest molecules from its pores, and its activated structure has a high BET surface area. The mesoporous structure **3D-CageHOF-1** could lead to new applications of HOFs in host–guest chemistry by enabling larger guests to occupy the crystal pores, such as enzymes.[56]

This study is also the first example of a molecular crystal that was computationally identified using a combination of CSP and structural invariants. Here, this led to the experimental discovery of a rare example of a mesoporous HOF. More broadly, we anticipate that this strategy will help to identify other functional molecular crystals by leveraging the proven power of structural analogy, supported by *a priori* lattice energy calculations and a formal metric for structural similarity between energy landscapes.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/jacs.2c02653.

Computational methodology, structural invariants, synthetic procedures, PXRD, sc-XRD, DSC, TGA, and gas sorption analysis (PDF)

Accession Codes

CCDC 2157172–2157173 contain the supplementary crystallographic data for this paper. These data can be obtained free of charge via www.ccdc.cam.ac.uk/data_request/cif, or by emailing data_request@ccdc.cam.ac.uk, or by contacting The Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge CB2 1EZ, UK; fax: +44 1223 336033.

AUTHOR INFORMATION

Corresponding Authors

Vitaliy Kurlin − *Computer Science, University of Liverpool, Liverpool L69 3BX, U.K.*; Email: vitaliy.kurlin@liverpool.ac.uk

Graeme M. Day − *Computational Systems Chemistry, School of Chemistry, University of Southampton, Southampton SO17 1BJ, U.K.*; orcid.org/0000-0001-8396-2771; Email: g.m.day@soton.ac.uk

Marc A. Little − *Materials Innovation Factory and Department of Chemistry, University of Liverpool, Liverpool L7 3NY, U.K.*; orcid.org/0000-0002-1994-0591; Email: malittle@liverpool.ac.uk

Andrew I. Cooper − *Materials Innovation Factory and Department of Chemistry and Leverhulme Research Centre for Functional Materials Design, University of Liverpool,*

Liverpool L7 3NY, U.K.; orcid.org/0000-0003-0201-1021; Email: aicooper@liverpool.ac.uk

Authors

Qiang Zhu − *Materials Innovation Factory and Department of Chemistry and Leverhulme Research Centre for Functional Materials Design, University of Liverpool, Liverpool L7 3NY, U.K.*

Jay Johal − *Computational Systems Chemistry, School of Chemistry, University of Southampton, Southampton SO17 1BJ, U.K.*

Daniel E. Widdowson − *Computer Science, University of Liverpool, Liverpool L69 3BX, U.K.;* orcid.org/0000-0002-5958-0703

Zhongfu Pang − *Materials Innovation Factory and Department of Chemistry and Leverhulme Research Centre for Functional Materials Design, University of Liverpool, Liverpool L7 3NY, U.K.*

Boyu Li − *Materials Innovation Factory and Department of Chemistry, University of Liverpool, Liverpool L7 3NY, U.K.*

Christopher M. Kane − *Materials Innovation Factory and Department of Chemistry, University of Liverpool, Liverpool L7 3NY, U.K.*

Complete contact information is available at: https://pubs.acs.org/10.1021/jacs.2c02653

Author Contributions

The manuscript was written through the contributions of all the authors.

Funding

The authors received funding from the Engineering and Physical Sciences Research Council (EPSRC) (EP/V026887/1, EP/R029431, EP/T022213, EP/N004884/1, and EP/R018472/1) and the Leverhulme Trust *via* the Leverhulme Research Centre for Functional Materials Design. This project has received funding from the European Research Council under the European Union's Horizon 2020 research and innovation program (grant agreement no. 856405). V.K. received an Industrial Fellowship from the Royal Academy of Engineering (IF2122/186). Q.Z. received a University of Liverpool Graduate Association (Hong Kong) Postgraduate Scholarship.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We acknowledge the use of the IRIDIS High-Performance Computing Facility and associated support services at the University of Southampton. *Via* our membership of the UK's HEC Materials Chemistry Consortium, which is funded by the EPSRC (EP/R029431), this work used the UK Materials and Molecular Modelling Hub for computational resources, the MMM Hub, which is partially funded by the EPSRC (EP/P020194/1 and EP/T022213/1). V.K. thanks the Royal Academy of Engineering for an Industry Fellowship (IF2122/186). Q.Z. thanks the University of Liverpool Graduate Association (Hong Kong) for a Postgraduate Scholarship. The authors thank Yu Che for helping to prepare the figures.

■ REFERENCES

(1) Price, S. L. Predicting Crystal Structures of Organic Compounds. *Chem. Soc. Rev.* **2014**, *43*, 2098−2111.

(2) Woodley, S. M.; Day, G. M.; Catlow, R. Structure Prediction of Crystals, Surfaces and Nanoparticles. *Philos. Trans. R. Soc., A* **2020**, *378*, 20190600.

(3) Pulido, A.; Chen, L.; Kaczorowski, T.; Holden, D.; Little, M. A.; Chong, S. Y.; Slater, B. J.; McMahon, D. P.; Bonillo, B.; Stackhouse, C. J.; Stephenson, A.; Kane, C. M.; Clowes, R.; Hasell, T.; Cooper, A. I.; Day, G. M. Functional Materials Discovery Using Energy-Structure-Function Maps. *Nature* **2017**, *543*, 657−664.

(4) Jones, J. T. A.; Hasell, T.; Wu, X.; Bacsa, J.; Jelfs, K. E.; Schmidtmann, M.; Chong, S. Y.; Adams, D. J.; Trewin, A.; Schiffman, F.; Cora, F.; Slater, B.; Steiner, A.; Day, G. M.; Cooper, A. I. Modular and Predictable Assembly of Porous Organic Molecular Crystals. *Nature* **2011**, *474*, 367−371.

(5) Campbell, J. E.; Yang, J.; Day, G. M. Predicted Energy-Structure-Function Maps for the Evaluation of Small Molecule Organic Semiconductors. *J. Mater. Chem. C* **2017**, *5*, 7574−7584.

(6) Day, G. M.; Cooper, A. I. Energy-Structure-Function Maps: Cartography for Materials Discovery. *Adv. Mater.* **2018**, *30*, 1704944.

(7) Cui, P.; Mcmahon, D. P.; Spackman, P. R.; Alston, B. M.; Little, M. A.; Day, G. M.; Cooper, A. I. Mining Predicted Crystal Structure Landscapes with High Throughput Crystallisation: Old Molecules, New Insights. *Chem. Sci.* **2019**, *10*, 9988−9997.

(8) McMahon, D. P.; Stephenson, A.; Chong, S. Y.; Little, M. A.; Jones, J. T. A.; Cooper, A. I.; Day, G. M. Computational Modelling of Solvent Effects in a Prolific Solvatomorphic Porous Organic Cage. *Faraday Discuss.* **2018**, *211*, 383−399.

(9) Aitchison, C. M.; Kane, C. M.; Mcmahon, D. P.; Spackman, P. R.; Pulido, A.; Wang, X.; Wilbraham, L.; Chen, L.; Clowes, R.; Zwijnenburg, M. A.; Sprick, R. S.; Little, M. A.; Day, G. M.; Cooper, A. I. Photocatalytic Proton Reduction by a Computationally Identified, Molecular Hydrogen-Bonded Framework. *J. Mater. Chem. A* **2020**, *8*, 7158−7170.

(10) Yang, S.; Day, G. M. Global Analysis of Crystal Energy Landscapes: Applying the Thresh-Old Algorithm to Molecular Crystal Structures. Nov 17, 2021, ChemRxiv preprint. https://chemrxiv.org/engage/chemrxiv/article-details/6193e70a78db4e47f810560a (accessed April 21, 2022).

(11) Horike, S.; Shimomura, S.; Kitagawa, S. Soft Porous Crystals. *Nat. Chem.* **2009**, *1*, 695−704.

(12) Diercks, C. S.; Yaghi, O. M. The Atom, the Molecule, and the Covalent Organic Framework. *Science* **2017**, *355*, No. eaal1585.

(13) Biswal, B. P.; Chandra, S.; Kandambeth, S.; Lukose, B.; Heine, T.; Banerjee, R. Mechanochemical Synthesis of Chemically Stable Isoreticular Covalent Organic Frameworks. *J. Am. Chem. Soc.* **2013**, *135*, 5328−5331.

(14) Lin, R.-B.; He, Y.; Li, P.; Wang, H.; Zhou, W.; Chen, B. Multifunctional Porous Hydrogen-Bonded Organic Framework Materials. *Chem. Soc. Rev.* **2019**, *48*, 1362−1389.

(15) Wang, B.; Lin, R.-B.; Zhang, Z.; Xiang, S.; Chen, B. Hydrogen-Bonded Organic Frameworks as a Tunable Platform for Functional Materials. *J. Am. Chem. Soc.* **2020**, *142*, 14399−14416.

(16) Li, P.; Ryder, M. R.; Stoddart, J. F. Hydrogen-Bonded Organic Frameworks: A Rising Class of Porous Molecular Materials. *Acc. Mater. Res.* **2020**, *1*, 77−87.

(17) Hisaki, I.; Xin, C.; Takahashi, K.; Nakamura, T. Designing Hydrogen-Bonded Organic Frameworks (HOFs) with Permanent Porosity. *Angew. Chem., Int. Ed.* **2019**, *58*, 11160−11170.

(18) Hu, F.; Liu, C.; Wu, M.; Pang, J.; Jiang, F.; Yuan, D.; Hong, M. An Ultrastable and Easily Regenerated Hydrogen-Bonded Organic Molecular Framework with Permanent Porosity. *Angew. Chem., Int. Ed.* **2017**, *56*, 2101−2104.

(19) Hisaki, I.; Suzuki, Y.; Gomez, E.; Cohen, B.; Tohnai, N.; Douhal, A. Docking Strategy To Construct Thermostable, Single-Crystalline, Hydrogen-Bonded Organic Framework with High Surface Area. *Angew. Chem., Int. Ed.* **2018**, *57*, 12650−12655.

(20) Ma, K.; Li, P.; Xin, J. H.; Chen, Y.; Chen, Z.; Goswami, S.; Liu, X.; Kato, S.; Chen, H.; Zhang, X.; Bai, J.; Wasson, M. C.; Maldonado, R. R.; Snurr, R. Q.; Farha, O. K. Ultrastable Mesoporous Hydrogen-Bonded Organic Framework-Based Fiber Composites toward Mustard Gas Detoxification. *Cell Rep. Phys. Sci.* **2020**, *1*, 100024.
(21) Yin, Q.; Zhao, P.; Sa, R.-J.; Chen, G.-C.; Lü, J.; Liu, T.-F.; Cao, R. An Ultra-Robust and Crystalline Redeemable Hydrogen-Bonded Organic Framework for Synergistic Chemo-Photodynamic Therapy. *Angew. Chem., Int. Ed.* **2018**, *57*, 7691−7696.
(22) Eddaoudi, M.; Kim, J.; Rosi, N.; Vodak, D.; Wachter, J.; O'Keeffe, M.; Yaghi, O. M. Systematic Design of Pore Size and Functionality in Isoreticular MOFs and Their Application in Methane Storage. *Science* **2002**, *295*, 469−472.
(23) Gropp, C.; Ma, T.; Hanikel, N.; Yaghi, O. M. Design of Higher Valency in Covalent Organic Frameworks. *Science* **2020**, *370*, No. eabd6406.
(24) Zhang, X.; Wang, J. X.; Li, L.; Pei, J.; Krishna, R.; Wu, H.; Zhou, W.; Qian, G.; Chen, B.; Li, B. A Rod-Packing Hydrogen-Bonded Organic Framework with Suitable Pore Confinement for Benchmark Ethane/Ethylene Separation. *Angew. Chem., Int. Ed.* **2021**, *60*, 10304−10310.
(25) Zhou, Y.; Kan, L.; Eubank, J. F.; Li, G.; Zhang, L.; Liu, Y. Self-Assembly of Two Robust 3D Supramolecular Organic Frameworks from a Geometrically Non-Planar Molecule for High Gas Selectivity Performance. *Chem. Sci.* **2019**, *10*, 6565−6571.
(26) Li, P.; Li, P.; Ryder, M. R.; Liu, Z.; Stern, C. L.; Farha, O. K.; Stoddart, J. F. Interpenetration Isomerism in Triptycene-Based Hydrogen-Bonded Organic Frameworks. *Angew. Chem., Int. Ed.* **2019**, *58*, 1664−1669.
(27) Li, P.; Chen, Z.; Ryder, M. R.; Stern, C. L.; Guo, Q.-H.; Wang, X.; Farha, O. K.; Stoddart, J. F. Assembly of a Porous Supramolecular Polyknot from Rigid Trigonal Prismatic Building Blocks. *J. Am. Chem. Soc.* **2019**, *141*, 12998−13002.
(28) Zentner, C. A.; Lai, H. W. H.; Greenfield, J. T.; Wiscons, R. A.; Zeller, M.; Campana, C. F.; Talu, O.; Fitzgerald, S. A.; Rowsell, J. L. C. High Surface Area and Z′ in a Thermally Stable 8-Fold Polycatenated Hydrogen-Bonded Framework. *Chem. Commun.* **2015**, *51*, 11642−11645.
(29) Yin, Q.; Li, Y.-L.; Li, L.; Lü, J.; Liu, T.-F.; Cao, R. Novel Hierarchical Meso-Microporous Hydrogen-Bonded Organic Framework for Selective Separation of Acetylene and Ethylene versus Methane. *ACS Appl. Mater. Interfaces* **2019**, *11*, 17823−17827.
(30) Yang, W.; Wang, J.; Wang, H.; Bao, Z.; Zhao, J. C.-G.; Chen, B. Highly Interpenetrated Robust Microporous Hydrogen-Bonded Organic Framework for Gas Separation. *Cryst. Growth Des.* **2017**, *17*, 6132−6137.
(31) Li, Y.-L.; Alexandrov, E. V.; Yin, Q.; Li, L.; Fang, Z.-B.; Yuan, W.; Proserpio, D. M.; Liu, T.-F. Record Complexity in the Polycatenation of Three Porous Hydrogen-Bonded Organic Frameworks with Stepwise Adsorption Behaviors. *J. Am. Chem. Soc.* **2020**, *142*, 7218−7224.
(32) Wang, B.; He, R.; Xie, L.-H.; Lin, Z.-J.; Zhang, X.; Wang, J.; Huang, H.; Zhang, Z.; Schanze, K. S.; Zhang, J.; Xiang, S.; Chen, B. Microporous Hydrogen-Bonded Organic Framework for Highly Efficient Turn-Up Fluorescent Sensing of Aniline. *J. Am. Chem. Soc.* **2020**, *142*, 12478−12485.
(33) Han, B.; Wang, H.; Wang, C.; Wu, H.; Zhou, W.; Chen, B.; Jiang, J. Postsynthetic Metalation of a Robust Hydrogen-Bonded Organic Framework for Heterogeneous Catalysis. *J. Am. Chem. Soc.* **2019**, *141*, 8737−8740.
(34) Mastalerz, M.; Oppel, I. M. Rational Construction of an Extrinsic Porous Molecular Crystal with an Extraordinary High Specific Surface Area. *Angew. Chem., Int. Ed.* **2012**, *51*, 5252−5255.
(35) Widdowson, D.; Mosca, M. M.; Pulido, A.; Cooper, A. I.; Kurlin, V. Average Minimum Distances of Periodic Point Sets − Foundational Invariants for Mapping Periodic Crystals. *Match* **2022**, *87*, 529−559.
(36) Widdowson, D.; Kurlin, V. Pointwise Distance Distributions of Periodic Sets. Aug 10, 2021, arXiv:2108.04798. arXiv preprint. http://kurlin.org/projects/periodic-geometry-topology/PDD.pdf (accessed April 14, 2022).
(37) Zhang, G.; Mastalerz, M. Organic Cage Compounds-from Shape-Persistency to Function. *Chem. Soc. Rev.* **2014**, *43*, 1934−1947.
(38) Klotzbach, S.; Beuerle, F. Shape-Controlled Synthesis and Self-Sorting of Covalent Organic Cage Compounds. *Angew. Chem., Int. Ed.* **2015**, *54*, 10356−10360.
(39) Berardo, E.; Turcani, L.; Miklitz, M.; Jelfs, K. E. An Evolutionary Algorithm for the Discovery of Porous Organic Cages. *Chem. Sci.* **2018**, *9*, 8513−8527.
(40) Swamy, S. I.; Bacsa, J.; Jones, J. T. A.; Stylianou, K. C.; Steiner, A.; Ritchie, L. K.; Hasell, T.; Gould, J. A.; Laybourn, A.; Khimyak, Y. Z.; Adams, D. J.; Rosseinsky, M. J.; Cooper, A. I. A Metal-Organic Framework with a Covalently Prefabricated Porous Organic Linker. *J. Am. Chem. Soc.* **2010**, *132*, 12773−12775.
(41) Zhang, L.; Xiang, L.; Hang, C.; Liu, W.; Huang, W.; Pan, Y. From Discrete Molecular Cages to a Network of Cages Exhibiting Enhanced CO_2 Adsorption Capacity. *Angew. Chem., Int. Ed.* **2017**, *56*, 7787−7791.
(42) Zhu, Q.; Wang, X.; Clowes, R.; Cui, P.; Chen, L.; Little, M. A.; Cooper, A. I. 3D Cage COFs: A Dynamic Three-Dimensional Covalent Organic Framework with High-Connectivity Organic Cage Nodes. *J. Am. Chem. Soc.* **2020**, *142*, 16842−16848.
(43) Ma, J.-X.; Li, J.; Chen, Y.-F.; Ning, R.; Ao, Y.-F.; Liu, J.-M.; Sun, J.; Wang, D.-X.; Wang, Q.-Q. Cage Based Crystalline Covalent Organic Frameworks. *J. Am. Chem. Soc.* **2019**, *141*, 3843−3848.
(44) Ji, C.; Su, K.; Wang, W.; Chang, J.; El-Sayed, E.-S. M.; Zhang, L.; Yuan, D. Tunable Cage-Based Three-Dimensional Covalent Organic Frameworks. *CCS Chem.* **2021**, *3*, 3094−3104.
(45) Giri, A.; Sahoo, A.; Dutta, T. K.; Patra, A. Cavitand and Molecular Cage-Based Porous Organic Polymers. *ACS Omega* **2020**, *5*, 28413−28424.
(46) Beuerle, F.; Gole, B. Covalent Organic Frameworks and Cage Compounds: Design and Applications of Polymeric and Discrete Organic Scaffolds. *Angew. Chem., Int. Ed.* **2018**, *57*, 4850−4878.
(47) 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.
(48) Price, S. L.; Leslie, M.; Welch, G. W. A.; Habgood, M.; Price, L. S.; Karamertzanis, P. G.; Day, G. M. Modelling Organic Crystal Structures Using Distributed Multipole and Polarizability-Based Model Intermolecular Potentials. *Phys. Chem. Chem. Phys.* **2010**, *12*, 8478−8490.
(49) Hargreaves, C. J.; Dyer, M. S.; Gaultois, M. W.; Kurlin, V. A.; Rosseinsky, M. J. The Earth Mover's Distance as a Metric for the Space of Inorganic Compositions. *Chem. Mater.* **2020**, *32*, 10610−10620.
(50) Anosova, O.; Kurlin, V. An Isometry Classification of Periodic Point Sets. In *International Conference on Discrete Geometry and Mathematical Morphology, DGMM 2021: Lecture Notes in Computer Science, Lindblad, May 24, 2021*; Malmberg, F. J., Sladoje, N., Eds.; Springer: Cham, 2021; pp 229−241.
(51) Himanen, L.; Jäger, M. O. J.; Morooka, E. V.; Federici Canova, F.; Ranawat, Y. S.; Gao, D. Z.; Rinke, P.; Foster, A. S. DScribe: Library of Descriptors for Machine Learning in Materials Science. *Comput. Phys. Commun.* **2020**, *247*, 106949.
(52) Pietrucci, F. Novel Enhanced Sampling Strategies for Transitions Between Ordered and Disordered Structures. In *Handbook of Materials Modeling*; Springer International Publishing: Cham, 2020; pp 597−619.
(53) Edelsbrunner, H.; Heiss, T.; Kurlin, V.; Smith, P.; Wintraecken, M. The Density Fingerprint of a Periodic Point Set. April 22, 2021, arXiv:2104.11046. arXiv preprint. https://drops.dagstuhl.de/opus/volltexte/2021/13831/ (accessed April 22, 2022).
(54) Belenguer, A. M.; Lampronti, G. I.; Cruz-Cabeza, A. J.; Hunter, C. A.; Sanders, J. K. M. Solvation and Surface Effects on Polymorph Stabilities at the Nanoscale. *Chem. Sci.* **2016**, *7*, 6617−6627.
(55) Macrae, C. F.; Sovago, I.; Cottrell, S. J.; Galek, P. T. A.; McCabe, P.; Pidcock, E.; Platings, M.; Shields, G. P.; Stevens, J. S.;

Towler, M.; Wood, P. A. Mercury 4.0 : From Visualization to Analysis, Design and Prediction. *J. Appl. Crystallogr.* **2020**, *53*, 226−235.

(56) Tang, Z.; Li, X.; Tong, L.; Yang, H.; Wu, J.; Zhang, X.; Song, T.; Huang, S.; Zhu, F.; Chen, G.; Ouyang, G. A Biocatalytic Cascade in an Ultrastable Mesoporous Hydrogen-Bonded Organic Framework for Point-of-Care Biosensing. *Angew. Chem., Int. Ed.* **2021**, *60*, 23608−23613.

https://doi.org/10.1021/jacs.2c02653
J. Am. Chem. Soc. 2022, 144, 9893−9901