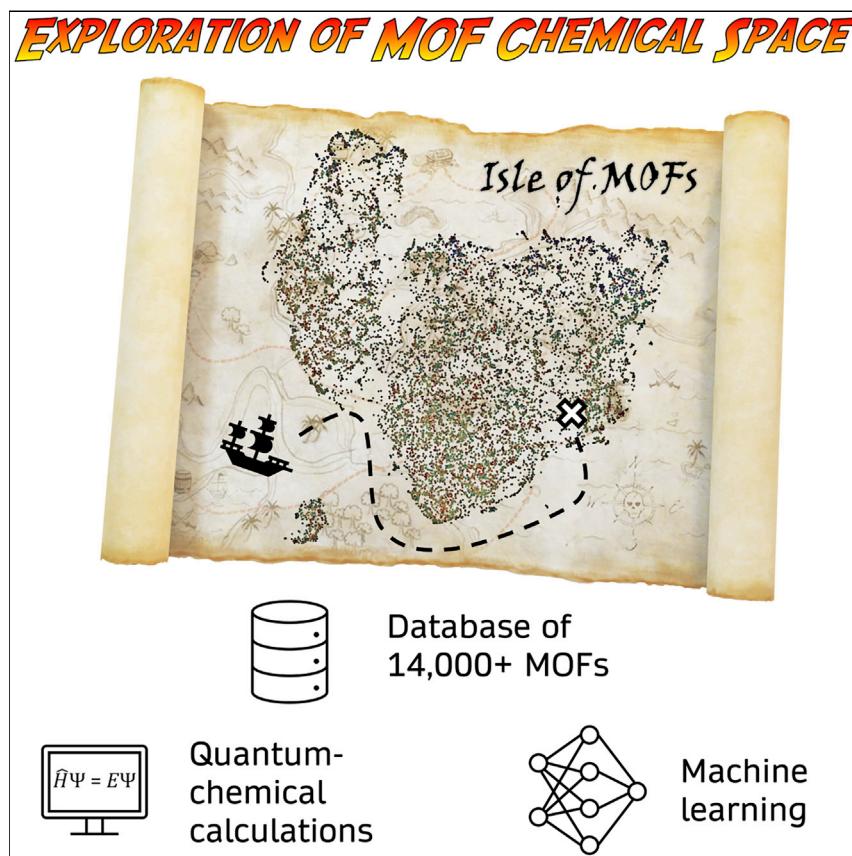


Article

Machine learning the quantum-chemical properties of metal–organic frameworks for accelerated materials discovery



Using a new database of electronic structure properties derived from high-throughput density functional theory calculations for thousands of metal–organic frameworks (MOFs), we benchmark a variety of machine learning models to accurately and rapidly predict MOF band gaps. Unsupervised dimensionality reduction techniques are also used to map the MOF feature space and identify otherwise subtle structure–property relationships. We anticipate that machine learning models derived from this new database will accelerate the discovery of promising MOFs with targeted quantum-chemical properties.



Understanding

Dependency and conditional studies on material behavior

Andrew S. Rosen, Shaelyn M. Iyer, Debmalya Ray, ..., Laura Gagliardi, Justin M. Notestein, Randall Q. Snurr

rosen@u.northwestern.edu

HIGHLIGHTS

New publicly available database of electronic structure properties for 14,000+ MOFs

Machine learning models can rapidly and accurately predict computed MOF band gaps

A crystal graph convolutional neural network achieves high predictive performance

Several MOFs with low band gaps are computationally identified

Rosen et al., Matter 4, 1578–1597
May 5, 2021 © 2021 The Authors. Published by Elsevier Inc.
<https://doi.org/10.1016/j.matt.2021.02.015>



Article

Machine learning the quantum-chemical properties of metal–organic frameworks for accelerated materials discovery

Andrew S. Rosen,^{1,7,8,*} Shaelyn M. Iyer,¹ Debmalya Ray,² Zhenpeng Yao,³ Alán Aspuru-Guzik,^{3,4,5} Laura Gagliardi,⁶ Justin M. Notestein,¹ and Randall Q. Snurr¹

SUMMARY

The modular nature of metal–organic frameworks (MOFs) enables synthetic control over their physical and chemical properties, but it can be difficult to know which MOFs would be optimal for a given application. High-throughput computational screening and machine learning are promising routes to efficiently navigate the vast chemical space of MOFs but have rarely been used for the prediction of properties that need to be calculated by quantum mechanical methods. Here, we introduce the Quantum MOF (QMOF) database, a publicly available database of computed quantum-chemical properties for more than 14,000 experimentally synthesized MOFs. Throughout this study, we demonstrate how machine learning models trained on the QMOF database can be used to rapidly discover MOFs with targeted electronic structure properties, using the prediction of theoretically computed band gaps as a representative example. We conclude by highlighting several MOFs predicted to have low band gaps, a challenging task given the electronically insulating nature of most MOFs.

INTRODUCTION

Over the last several years, significant attention has been focused on the design of novel metal–organic frameworks (MOFs), a class of materials composed of discrete inorganic nodes connected to one another via organic linkers. One of the main advantages of MOFs is that they often have predictable and atomically defined structures with properties that are directly related to the choice of underlying metal and organic building blocks.¹ In this way, it becomes possible to impart physical and chemical functionality specifically tailored for a given application.² To date, tens of thousands of MOFs have been synthesized,^{3,4} and a nearly unlimited number can be proposed^{5–7} by considering different combinations of constituent building blocks. Due to the enormous set of possible framework compositions, structures, and resulting properties,⁸ it remains difficult to discover truly top-performing MOFs for a particular application based solely on chemical intuition, conventional trial-and-error experimental testing, or serendipity alone.

High-throughput computational screening approaches based on classical simulations have proved extremely useful for more efficiently exploring the vast combinatorial space of MOF structures.^{9,10} Recently, the large quantities of data generated during these computational screening studies have led to the development of machine learning (ML) models¹¹ that can accelerate the MOF design and discovery process even further. ML-assisted screening studies have been successfully applied to

Progress and potential

Metal–organic frameworks (MOFs) are a class of crystalline solids with tunable structures that make it possible to impart chemical functionality tailored for a given application. A virtually unlimited number of MOFs can be realized, making it difficult to know which would be top-performing candidates. This is especially true for the many applications governed by quantum mechanical properties, which are computationally demanding to simulate and time-consuming to probe experimentally. Here, we present a database of quantum mechanical properties for thousands of MOFs, the result of over 170 years of computing time. With this new database, we developed machine learning models that can predict the electronic properties of MOFs in mere seconds without the need for expensive quantum mechanical simulations and powerful supercomputers. We anticipate that machine learning models constructed using this database will accelerate the discovery of MOFs for a variety of long-standing challenges.



the discovery of MOFs suitable for H₂ storage,^{12–14} CO₂ separation/capture,^{15–17} and numerous other applications predominantly (though not exclusively)^{18,19} in the area of gas storage and separations.^{10,20,21} Nonetheless, similar efforts remain almost entirely unexplored for the many applications in which the properties of interest are best described by quantum mechanical models,²² such as those based on the electronic, optical, magnetic, and/or catalytic properties of MOFs. Beyond the sheer number of possible MOFs that can be realized, the large number of atoms in MOF crystal structures often makes it computationally demanding to carry out even moderate-scale quantum-chemical screening studies, further magnifying the need for ML approaches in this area.

To date, the most relevant studies focused on training ML models to predict the quantum-chemical properties of MOFs are those of Raza et al.²³ and Korolev et al.,²⁴ who independently developed ML models that can predict the partial atomic charges of MOFs in the Computation-Ready, Experimental (CoRE) MOF database.^{25,26} Beyond these fundamental studies on partial charge prediction, however, there remains a significant gap in the literature, particularly for the discovery of MOFs with desired electronic structure properties. To the best of our knowledge, the only prior work in this area is that of He et al.,²⁷ who trained binary classification models to predict whether inorganic solids in the Open Quantum Materials Database (OQMD)^{28,29} are metallic or nonmetallic. Without retraining on MOF data, a multimodel voting procedure was then used to predict the metallic or nonmetallic behavior of 2,932 MOFs in the CoRE MOF database,²⁵ which do not have computed band gaps. Of the six identified materials with near-zero band gap at the PBE level of theory,³⁰ all are best described as metal-cyanide/thiocyanate cluster complexes, and none have H atoms in the structure. This is likely due in large part to the extreme differences between the OQMD, which consists almost entirely of inorganic compounds, and the CoRE MOF database. Furthermore, the fidelity of the metallic materials was not considered, leading to highlighted structures such as [CdC₄]_n that should actually be [Cd(CN)₂]_n.³¹

In stark contrast to the existing literature on MOFs, significant progress has been made in the development of ML models that can accelerate the quantum-chemical screening process for a wide range of inorganic and molecular compounds.^{32–39} One of the fundamental features underlying much of this work has been the use of high-throughput density functional theory⁴⁰ (DFT) workflows to construct large-scale electronic structure–property databases, such as those developed for inorganic solids^{29,41–47} and molecular systems.^{48–52} The synergistic combination of high-throughput DFT databases and ML has led to the discovery of a diverse range of materials with sought-after properties, including efficient organic light-emitting diodes,⁵³ superhard inorganic materials,⁵⁴ and thermally conductive polymers,⁵⁵ among many others.³⁸ With this in mind, there is a significant need for an analogous database of DFT-computed material properties for MOFs so that new ML models can be developed for the rapid prediction of MOF electronic structure properties. High-throughput screening, database generation, and subsequent ML model development are crucial components for realizing the full potential of reticular chemistry⁵⁶ and accelerating materials discovery in general.^{57–60}

In the present study, we leverage a recently developed high-throughput periodic DFT workflow tailored for MOF structures⁶¹ to construct a large-scale database of MOF quantum mechanical properties. This publicly available dataset—the Quantum MOF (QMOF) database⁶²—contains computed properties for 15,713 experimentally characterized MOFs after structure relaxation via DFT, including but not limited

¹Department of Chemical and Biological Engineering, Northwestern University, 2145 Sheridan Road, Evanston, IL 60208, USA

²Department of Chemistry, Chemical Theory Center, and Minnesota Supercomputing Institute, University of Minnesota, 207 Pleasant Street SE, Minneapolis, MN 55455, USA

³Chemical Physics Theory Group, Department of Chemistry and Department of Computer Science, University of Toronto, Toronto, ON M5S 3H6, Canada

⁴Vector Institute for Artificial Intelligence, Toronto, ON M5S 1M1, Canada

⁵Lebovic Fellow, Canadian Institute for Advanced Research (CIFAR) Senior Fellow, Toronto, ON M5S 1M1, Canada

⁶Department of Chemistry, Pritzker School of Molecular Engineering, James Franck Institute, Chicago Center for Theoretical Chemistry, The University of Chicago, 5735 South Ellis Avenue, Chicago, IL 60637, USA

⁷Lead contact

⁸Twitter: @Andrew_S_Rosen

*Correspondence: rosen@u.northwestern.edu
<https://doi.org/10.1016/j.matt.2021.02.015>

to optimized geometries, energies, band gaps, charge densities, density of states, partial charges, spin densities, and bond orders. We anticipate that the QMOF database will serve two primary purposes: (1) materials discovery using the as-deposited data; and (2) the evaluation and development of novel ML algorithms to reduce, or circumvent altogether, the need for otherwise expensive DFT calculations.

To demonstrate the utility of the data generated via the high-throughput DFT workflow, we use the QMOF database to develop several ML models for the prediction of MOF band gaps from nothing more than an encoding of the experimental (i.e., unrelaxed) crystal structures, drastically decreasing the number of computationally demanding quantum mechanical simulations that would need to be carried out in future screening studies. Beyond serving as a proof of concept, an ML model that can predict MOF band gaps is particularly desirable, as most MOFs are known to be electronically insulating,⁶³ which limits their potential use in electrocatalysis, sensing, energy storage, and other applications for which some degree of electrical conductivity is necessary.^{63–67} We identify a top-performing band gap regression model based on a crystal graph convolutional neural network⁶⁸ and show how dimensionality reduction techniques can be used to discover overarching structure–property relationships for the identification of MOFs with targeted electronic structure properties. We conclude by highlighting several Fe MOFs with low band gaps identified for the first time in this work.

RESULTS AND DISCUSSION

Generation and overview of the QMOF database

Prior to carrying out any periodic DFT calculations, a dataset of starting structures must be assembled. There are several databases of MOF structures that have been published to date.^{3–6,25,69} However, it is imperative to note that existing databases of synthesized MOFs cannot be used as-is for quantum-chemical screening purposes. If even a single atom is missing or duplicated in a MOF crystal structure, the resulting DFT calculations are unlikely to be physically meaningful. Put another way, the simulation unit cell is expected to be charge-neutral unless otherwise specified; any additional or missing electron in the system ruins the integrity of the resulting charge density and, therefore, all the quantum-chemical properties derived from it. These situations can arise as a result of deficiencies in the deposited experimental crystal structure and/or in the dataset curation process when generating a database of MOF crystal structures. Therefore, in this work we aim to start with a comparatively “clean” dataset of crystal structures for high-throughput computational investigation, one we will refer to as a suitably “DFT-ready” dataset of MOFs.

We considered the list of materials identified as MOFs from both the Cambridge Structural Database (CSD) MOF subset³ and the 2019 Computation-Ready, Experimental (CoRE) MOF database,⁴ the latter of which contains a relatively small number of MOFs not present in the former. All starting structures were taken directly from the CSD by querying the corresponding CSD reference code (“refcode”), and free (i.e., unbound) solvents were automatically removed from the frameworks. We chose to take the initial structures directly from the CSD as a matter of consistency and so that we could make use of valuable CSD metadata⁷⁰ (e.g., unresolved atoms, charged structures) associated with each deposited crystal structure. From this set of experimental crystal structures, we constructed a smaller DFT-ready subset of 42,349 nondisordered MOF structures (“QMOF-42349”) after an extensive suite of automated fidelity checks, as summarized in [Figure S1](#). This process serves to filter out many problematic MOFs with omitted H atoms, fractional occupancies, deleted

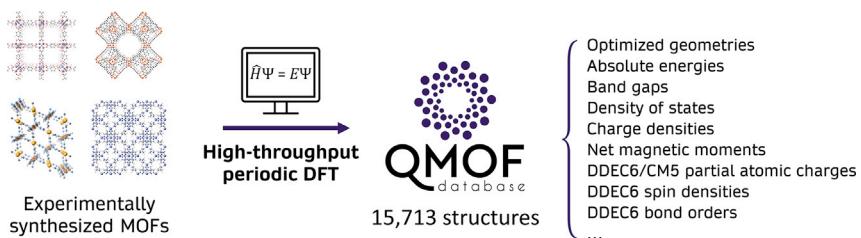


Figure 1. Overview of the QMOF database

Selected DFT-computed properties for the structurally relaxed MOFs made available in the QMOF database.

framework atoms, lone (i.e., unbonded) atoms, overlapping atoms, an improper number of charge-balancing ions, and other structural issues that have been discussed in several recent studies.^{70–76} Of these 42,349 experimental crystal structures, a subset of materials with 300 atoms or fewer per primitive cell was considered such that high-throughput DFT calculations could be carried out in an efficient manner. Full structure relaxations (including cell volume and atomic positions) were carried out via a multistage workflow⁶¹ (Table S2) at the PBE-D3(BJ)^{30,77,78} level of theory with the Vienna *ab initio* Simulation Package.^{79,80} Additional methodological details regarding the dataset construction, DFT calculations, and ML methods can be found in the [supplemental information](#).

The high-throughput periodic DFT workflow was successfully completed for 15,713 MOFs, and several DFT-computed properties were tabulated following the structure relaxation process, a selection of which are listed in Figure 1. Of these, band gaps are likely to be of interest for electronic and optical properties, especially in the search for (semi)conducting MOFs^{63,65,81,82} or screening for photocatalytic materials.⁸³ Electronic energies, particularly if converted to formation energies, may provide insight into the relative stability of MOFs.⁸⁴ Machine learning the charge density⁸⁵ is a potential way to bypass a large portion of the calculations performed with Kohn-Sham DFT.^{86–88} Both the charge density and density of states can provide insight into the electronic structure in addition to serving as promising features to predict a variety of other quantum-chemical properties.^{89,90} Partial atomic charges,^{91–94} bond orders,⁹⁵ and spin densities^{91,92} have a wide range of potential use cases, from describing electrostatic interactions in classical simulations of MOFs^{26,96} to serving as descriptors to better understand trends in catalytic reactions^{97–99} and small-molecule binding.¹⁰⁰ Furthermore, the DFT-optimized structures can be used as starting points for further quantum-chemical calculations and for analyzing geometric properties of MOFs. In addition to the curated data mentioned in Figure 1, all output data from the DFT calculations are made publicly available so that other properties of interest can be readily investigated.

Prior to highlighting how these data can be used in practice, we first investigated several properties of the QMOF database. As shown in Figure 2A, the QMOF database contains MOFs with chemical elements that span nearly the entire periodic table, which is beneficial for the development of transferable ML models. As anticipated, there is also a large number of MOFs in the QMOF database containing Cu, Zn, and Cd, which constitute the three most common types of inorganic nodes in the MOF literature.⁷⁵ Nonetheless, we note that some types of MOFs are currently under-represented in the QMOF database due in part to the dataset curation process, which filters out any MOFs with missing atomic coordinates or partial occupancies. These situations are likely to arise in MOFs with complex proton topologies

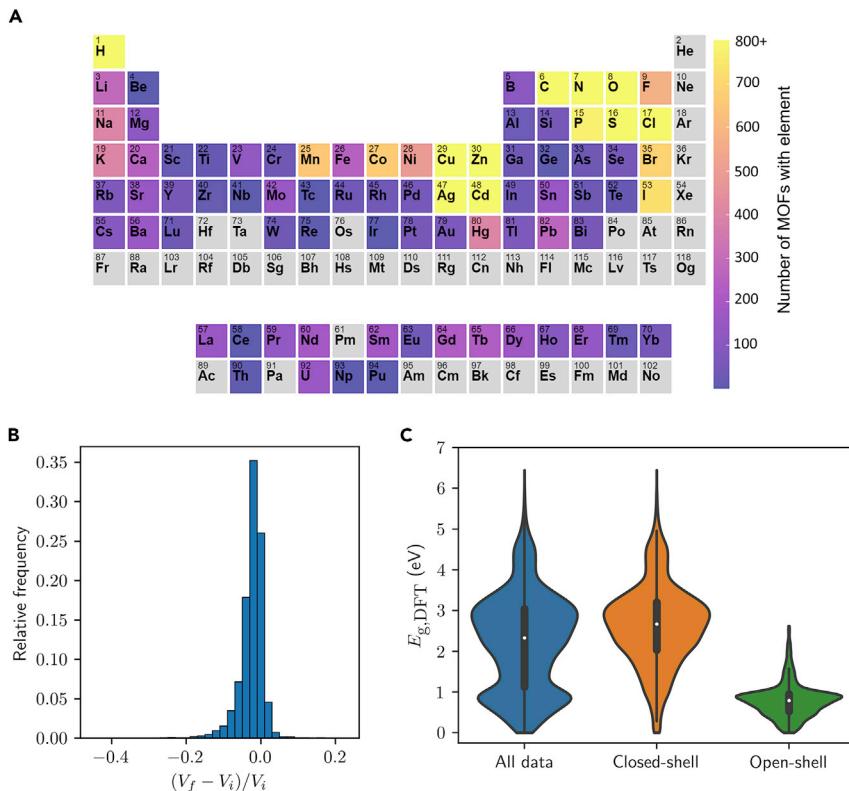


Figure 2. Distribution of properties within the QMOF database

(A) Number of MOFs in the QMOF database containing a given element. All elements that occur in greater than 800 structures are capped at 800 for ease of visualization. These include: C (15,713), H (15,713), N (12,892), O (12,821), Cu (2,882), S (2,684), Zn (2,665), Cd (2,538), Cl (1,687), and Ag (1,213). Elements in gray are not present in any structure.

(B) Histogram of the fractional change in cell volume before (V_i) and after (V_f) structure relaxation at the PBE-D3(BJ) level of theory for the MOFs in the QMOF database.

(C) Violin plots of the DFT-computed band gaps, $E_{g,DFT}$, at the PBE-D3(BJ) level of theory for the MOFs in the QMOF database. Separate distributions are shown for the entire dataset (15,713 entries), the closed-shell MOFs (12,169 entries), and the open-shell MOFs (3,544 entries). Open-shell character is defined here as having a DDEC6 atomic spin density with a magnitude greater than 0.1. A box plot, showing the extrema and interquartile range, is included in each violin with the median marked by a white dot.

that cannot be easily resolved from X-ray diffraction alone (e.g., Zr- and Hf-based MOFs),^{101,102} MOFs with defects in the crystal structure,¹⁰³ and MOFs that have undergone postsynthetic functionalization.^{104,105}

When looking at the geometries before and after structure relaxation, we find that 96.6% of the DFT-optimized MOFs had a change in cell volume of less than 10% (Figure 2B), suggesting that the removal of free solvent does not drastically alter the structural properties for most of the MOFs in this work. In the case of flexible MOFs, multiple conformations are often included in the QMOF database, which is important since they may exhibit different electronic structure properties.¹⁰⁶ As depicted in Figure S3, three distinct conformations of the flexible MOF Fe(bdp) (H_2 bdp = 1,4-benzenedipyrrolopyrazole)¹⁰⁷ are included in the QMOF database (refcodes: QUPZIM, QUPZIM01, QUPZIM02), one of which is on the extreme end of the distribution shown in Figure 2A (see also Table S5). This is not surprising given that high pressures of CH_4 are needed to stabilize the given open-pore configuration of Fe(bdp).¹⁰⁷

The distribution of DFT-computed band gaps for the fully optimized structures at the PBE-D3(BJ) level of theory is shown in [Figure 2C](#) and indicates that there is a wide spread of values from nearly 0 eV to 6.45 eV. The band gaps are not normally distributed and instead are bimodal, with peaks centered around 0.9 eV and 2.9 eV. This can be attributed to different distributions associated with closed- and open-shell materials in the QMOF database ([Figure 2C](#)), the latter of which have significantly lower band gaps at the PBE-D3(BJ) level of theory on average. With regard to partial atomic charges, a wide spread of values is also obtained ([Figure S5A](#)). On comparing the partial atomic charges before and after structure relaxation, we find that 92.4% of the ~1.2 million data points have an absolute difference of less than 0.05 q_e , and 98.9% of the points have an absolute difference of less than 0.1 q_e ([Figure S5B](#)). As has been observed on a smaller scale in prior work,^{26,69} it can be safely assumed that the partial charges remain essentially unchanged upon structure relaxation in most cases.

As a brief demonstration for how the data generated via the high-throughput DFT workflow could be used directly, we identified any porous framework materials with high-spin Fe species following the high-throughput DFT workflow. High-spin Fe complexes are known to be promising for oxidation catalysis, in particular for the activation of strong C–H bonds, and recent work has focused on stabilizing such motifs in MOFs for this purpose.^{108–110} This query of the QMOF database resulted in six unique MOFs, as shown in [Figure S6](#). Providing validation of this screening approach, two of the six MOFs—Fe₂(dobdc) (H₄dobdc = 2,5-dihydroxybenzene-1,4-dicarboxylic acid) (refcode: COKNOH)¹¹¹ and Fe₂(dobpdc) (H₄dobpdc = 4,4'-dihydroxy-(1,1'-biphenyl)-3,3'-dicarboxylic acid) (refcode: MAL-SIE)¹¹²—have already been shown to oxidize strong C–H bonds.^{108,113,114} Another two of the six MOFs—Fe₂Cl₂(bbta) (H₂bbta = 1H,5H-benzo(1,2-d:4,5-d')bistriazole) (refcode: HAYYUE)¹¹⁵ and Fe₂Cl₂(btdd) (H₂btdd = bis(1H-1,2,3-triazolo[4,5-b],[4',5'-i]dibenz[1,4]dioxin) (refcode: HAYZAL)¹¹⁵—have been computationally investigated for their use in oxidation reactions.^{97,116,117} Prior experimental studies suggest that the aforementioned MOFs exhibit high-spin Fe sites.^{108,114,115}

Machine learning models for band gap prediction

Beyond analyzing the DFT-computed properties directly, the QMOF database now makes it possible to train a wide range of ML models specifically tailored for MOFs, which are likely to have their own distinct feature space compared with isolated molecules and inorganic solids. This serves two primary purposes. The first is more theoretical: featurization methods (i.e., how each MOF structure is numerically encoded) and ML algorithms that are well suited for other materials may not be equally suitable for MOFs, so this database of quantum-chemical properties can serve as a testing ground to benchmark new ML methods. The Materials Project⁴¹ and OQMD^{28,29}, in particular have accelerated this research direction for inorganic solids, and the QM9 dataset^{48,118} (as one example) has done the same for small-molecule chemistry. The second purpose of this new database is to apply these rapid yet accurate ML models to accelerate the materials discovery process, now with the ability to train these models directly on properties computed for MOFs.

In this work, we have chosen to develop an ML regression model that can rapidly predict the DFT-computed band gaps of MOFs. Specifically, we aim to predict the computed band gaps of the DFT-optimized structures from the unoptimized, experimentally resolved MOF crystal structures such that no quantum-chemical calculations need to be carried out. To achieve this, all ML models are trained on the band gaps of the DFT-optimized structures but take representations of the

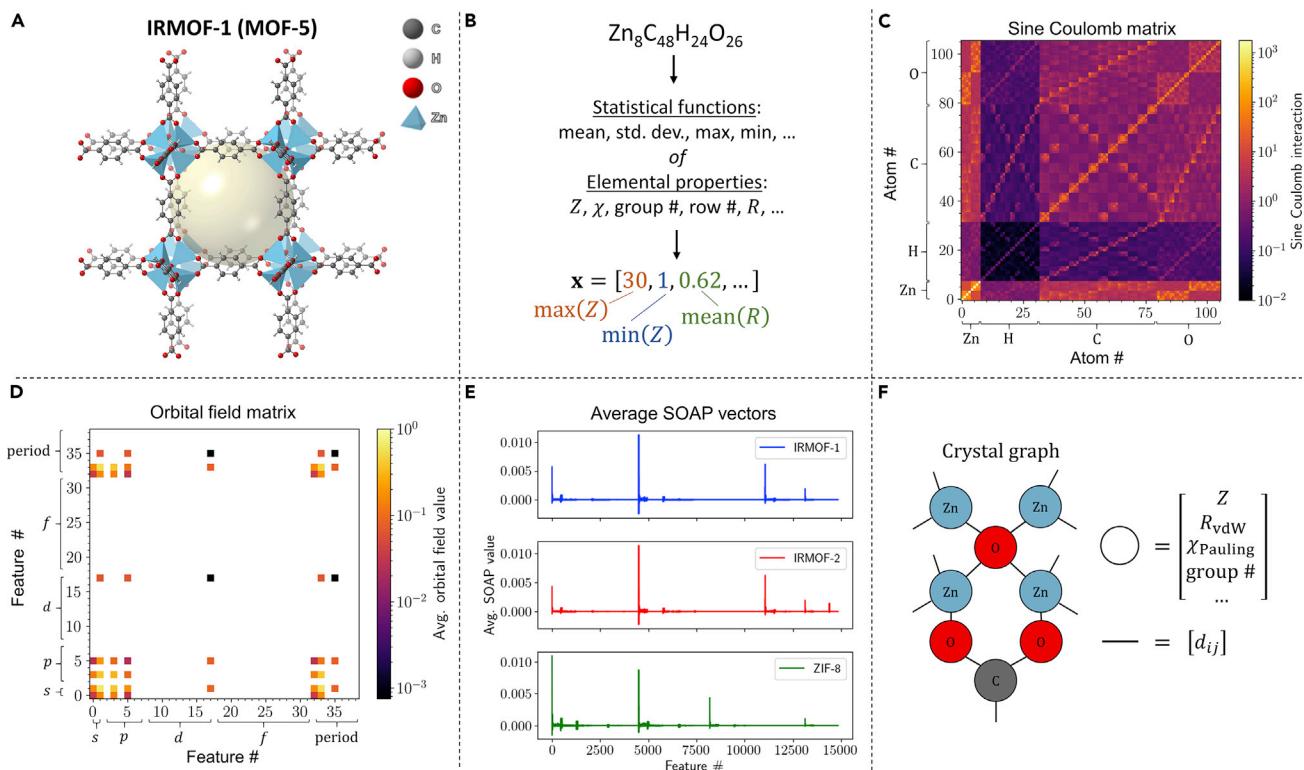


Figure 3. Visualization of various featurization methods applied to IRMOF-1

- (A) IRMOF-1 structure.
- (B) Examples of composition-based features.
- (C) Sine Coulomb matrix showing the interaction values between each pair of atoms.
- (D) Orbital field matrix showing the average interaction value between each pair of orbital- or period-based features. Only nonzero values are shown.
- (E) Averaged SOAP fingerprint of IRMOF-1 compared with IRMOF-2 and ZIF-8. Taking the dot product of any two vectors yields an unnormalized similarity score.
- (F) Schematic of a crystal graph with example node (circle) and edge (line) embeddings (only a representative portion is shown for clarity).

corresponding unrelaxed experimental structures as the input. Since the development of an ML regression model that can predict the band gaps of MOF crystal structures has not been achieved previously, we trained several ML models using a variety of common featurization methods to benchmark each approach. These featurization methods are graphically summarized in Figure 3 for a representative material IRMOF-1 (IRMOF = isoreticular MOF),¹¹⁹ also known as MOF-5 (Figure 3A). For the purposes of training ML models throughout this work, we specifically focus on a de-duplicated subset of 14,482 materials in the QM0F database ("QM0F-14482") that have gone through the full periodic DFT volume-relaxation process.

The simplest featurization methods considered in this work are the feature sets of He et al.²⁷ (with 45 statistical attributes of elemental properties, denoted "Stoichiometric-45") and Meredig and Agrawal et al.¹²⁰ (with 103 attributes describing the elemental fractions from H–Lr and 17 statistical attributes of elemental properties, denoted "Stoichiometric-120"), which rely solely on the chemical composition of each material (Figure 3B). In addition, we consider several structure-sensitive featurization approaches, including the sine Coulomb matrix¹²¹ that encodes pairwise electrostatic interactions between nuclei in a material (Figure 3C and Equation S4) and the orbital field matrix¹²² that encodes the distribution of valence electrons in each coordination environment of a material (Figure 3D). The smooth overlap of

Table 1. Benchmarking the performance of different machine learning models

ML method	MAE (eV)	R ²	<i>p</i>
Constant mean model	0.973	–	–
Sine Coulomb matrix	0.529 ± 0.008	0.643 ± 0.012	0.787 ± 0.008
Stoichiometric-45	0.437 ± 0.004	0.743 ± 0.006	0.842 ± 0.004
Stoichiometric-120	0.433 ± 0.010	0.750 ± 0.009	0.847 ± 0.005
Orbital field matrix	0.417 ± 0.008	0.763 ± 0.010	0.863 ± 0.003
SOAP	0.357 ± 0.008	0.822 ± 0.010	0.910 ± 0.003
CGCNN	0.274 ± 0.008	0.876 ± 0.011	0.932 ± 0.005

Summary of the testing-set mean absolute error (MAE), coefficient of determination (R^2), and Spearman rank-order correlation coefficient (*p*) for several machine learning methods to predict the computed band gaps of MOFs from their deposited crystal structures with free solvent removed. Kernel ridge regression was used for all featurization methods except for the crystal graphs of CGCNN, for which a convolutional neural network was constructed. The testing-set statistics are shown, averaged over five runs (using different random seeds for data splitting) with ±1 standard deviation shown. For all models, 80% of the QMOF-14482 dataset was used for training. The MAE for a dummy model that predicts the mean band gap (2.220 eV) for all the MOFs is shown for reference.

atomic positions (SOAP)^{123,124} is another structure-sensitive descriptor considered in this work, which can be used to compute the similarity between a pair of local atomic environments—and, by extension, a pair of structures—by representing the atoms as Gaussians (i.e., “smoothed positions”) and comparing the spatial overlap in the resulting atomic density fields (Figure S2 and Equations S5–S9). In all of the aforementioned examples, these features are used to develop a kernel ridge regression¹²⁵ (KRR) model (Equations S1–S3). Motivated by prior work on inorganic solids, we also investigated the use of a crystal graph convolutional neural network (CGCNN),⁶⁸ wherein an approximate crystal graph is generated for each MOF, with each node in the graph representing an atom and each edge representing the bonds that connect the atoms (Figure 3F). More detailed descriptions and full methodological details for each featurization method and ML model architecture can be found in the [supplemental information](#).

As shown in Table 1, the KRR models trained on composition-based features (i.e., Stoichiometric-45 and Stoichiometric-120) are able to capture some of the band gap trends with mean absolute errors (MAEs) of 0.43–0.44 eV (with respect to the DFT-computed values) on the out-of-sample testing set. Nonetheless, these methods are still quite limited for regression purposes given that they do not encode any information about the structural properties of the MOF. In terms of structure-sensitive methods, taking an eigenvalue spectrum of the sine Coulomb matrix fares worse than the stoichiometry-based features, yielding a testing-set MAE of 0.53 eV (Table 1). This can likely be traced back to the required use of zero-padding in the sine Coulomb matrix to ensure constant-length feature vectors between MOFs with different numbers of atoms per unit cell. The KRR model using a flattened orbital field matrix as the feature set is more accurate than the model based on the sine Coulomb matrix but shows only a minor improvement over the stoichiometry-based features. Overall, SOAP performs the best of all tested KRR descriptor sets, with an MAE of 0.36 eV and $R^2 = 0.82$ on the testing set. The marked improvement in performance with SOAP is especially clear when comparing the parity plots of the different KRR models (Figure S7).

Notably, CGCNN significantly outperforms all the aforementioned KRR models, achieving an MAE of 0.27 eV and $R^2 = 0.88$ (Table 1). As a point of reference, a trivial model that simply predicts the mean band gap for every MOF would have an MAE of 0.97 eV, suggesting that CGCNN captures much of the underlying chemistry. The

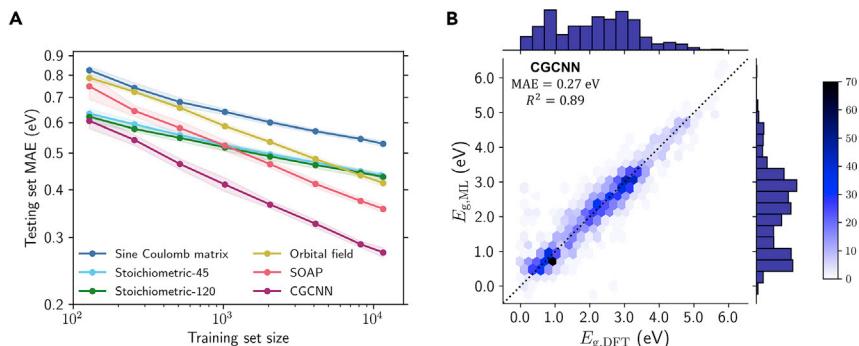


Figure 4. Learning curves and CGCNN parity plot

(A) Mean absolute error (MAE) for band gap predictions on the testing set as a function of training-set size for various machine learning methods. Each point represents the average MAE over five runs with different random seeds for data splitting, and the shaded region represents ± 1 standard deviation. The data are shown on log-log axes.

(B) Testing-set parity plot for the CGCNN model with hexagonal binning, comparing the machine learning band gaps, $E_{g,ML}$, with the PBE-D3(BJ) band gaps of the DFT-optimized structures, $E_{g,DFT}$. The color bar indicates the number of MOFs in each bin, and the line of parity is shown as a dashed line. Histograms summarizing the distribution of $E_{g,DFT}$ and $E_{g,ML}$ data are displayed parallel to the x and y axes, respectively.

performance of the CGCNN model for MOF band gaps is comparable, if not slightly better, than state-of-the-art ML band gap models trained on inorganic solids from the OQMD and Materials Project as well as the organic crystals from the Organic Materials Database (OMDB).^{44,68,126,127} It is also worth noting that the experimentally measured band gaps of MOFs can vary by several tenths of an electronvolt depending on the synthesis and post-treatment conditions.¹²⁸ As such, an MAE less than 0.3 eV is promising for the identification of structure–property trends and for sorting material candidates by band gap, the latter of which is further justified by the CGCNN’s high Spearman rank-order correlation coefficient of $\rho = 0.93$. For context, it took ~8 min (7 min for a one-time encoding of the crystal graphs and 1 min to evaluate the neural network) on a modern laptop computer to predict the band gaps of all 14,482 MOFs in the QMOF-14482 set using the CGCNN model. In stark contrast, it took over 1.5 million hours (~170 years) of computing time on the Stampede2 supercomputer^{129,130} to carry out the structure relaxations and compute the band gaps via DFT.

The learning curves for each of the six models are shown in Figure 4A, highlighting the testing-set MAE as a function of the training-set size. Of all the individual models, CGCNN has the lowest MAE regardless of training-set size. While SOAP has a worse testing-set MAE than simpler stoichiometric models when trained on fewer than 1,000 MOFs, SOAP has a significantly higher learning rate such that it performs much better for larger training-set sizes (although it still underperforms compared with CGCNN). Reassuringly, the MAEs of the top-performing CGCNN and SOAP methods have not plateaued with respect to the training-set size over the range of values considered in this work (i.e., up to ~10⁴ training points). This indicates that both CGCNN and SOAP are capable of encoding the MOF crystal structures with sufficient uniqueness between structures and that the performance of the ML algorithms could be further improved if a greater number of training examples were provided. The testing-set parity plot for the CGCNN trained on 80% of the QMOF-14482 MOF dataset is shown in Figure 4B. As one would expect based on the relatively low MAE and high R^2 , the agreement with the DFT predictions is quite strong, and this is generally true across the full range of band gap values.

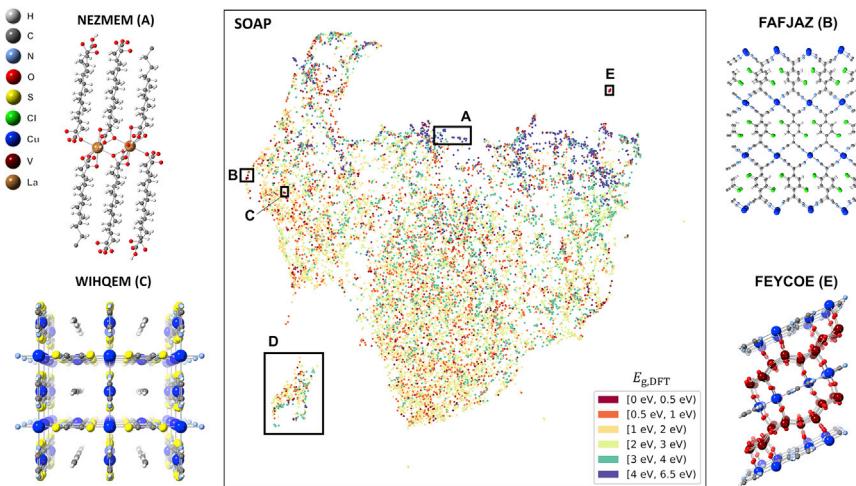


Figure 5. UMAP based on the SOAP average similarity kernel

Unsupervised structural dimensionality reduction performed using UMAP, with a distance matrix obtained from the SOAP average similarity kernel of the unrelaxed structures in the QMOF-14482 dataset. The PBE-D3(BJ) band gaps of the DFT-optimized structures, $E_{g,\text{DFT}}$, are overlaid on the UMAP. Selected MOFs in the projection are highlighted.

Dimensionality reduction for structure–property analysis

While the kernel-based methods have a higher MAE than CGCNN when predicting MOF band gaps, the underlying descriptors can still be used for dimensionality reduction, an unsupervised learning task that can cluster structurally similar MOFs in feature space for the purposes of identifying interpretable structure–property relationships. Using the uniform manifold approximation and projection (UMAP) algorithm to carry out the dimensionality reduction,^{131,132} the distance between each MOF in the reduced space can be related to the distance in feature space, such that clusters of points tend to have similar structures (Equation S10). By overlaying the DFT-computed band gaps over the UMAP, regions of low and high band gaps can emerge, making it possible to identify otherwise subtle structure–property trends.

As an example, selecting several MOFs in region A of the SOAP-based UMAP (Figure 5) yields materials with long, linear alkane-based linkers (e.g., refcode: NEZMEM),¹³³ which consistently have high band gaps regardless of the coordinating metal. The low band gap MOFs are more scattered throughout the reduced feature space, but as one example, region B of Figure 5 contains framework materials with linkers consisting of various TCNQ (TCNQ = 7,7,8,8-tetracyano-quinodimethane) derivatives, with several of these materials previously shown to have high electrical conductivities (e.g., refcodes: BISVUW, FAFJAZ¹³⁵). The projection in Figure 5 can be used to find MOFs that are structurally similar to a given material of interest as well. For instance, $\text{Cu}[\text{Ni}(\text{pdt})_2] \cdot \text{C}_2\text{H}_2$ (pdt^{2-} = 2,3-pyrazinedithiolate) (refcode: HIVPOU)¹³⁶ is in the QMOF-14482 dataset, and it is known to be one of the rare examples of a three-dimensional, porous framework that exhibits room-temperature electrical conductivity.¹³⁶ Perhaps unsurprisingly, one of the closest points to $\text{Cu}[\text{Ni}(\text{pdt})_2] \cdot \text{C}_2\text{H}_2$ is the isostructural framework $\text{Cu}[\text{Cu}(\text{pdt})_2] \cdot \text{C}_2\text{H}_2$ (refcode: WIHQEM)¹³⁷ (region C), which has also been studied for its conductive properties.^{138,139} In general, we find that the SOAP-based UMAP places greater emphasis on the similarity of the organic linkers rather than the metal identity, likely due to the averaging scheme used in the generation of the similarity kernel (Table S6). Modifications to the SOAP encoding that better account for the discrete building-block nature of MOFs, such as variations

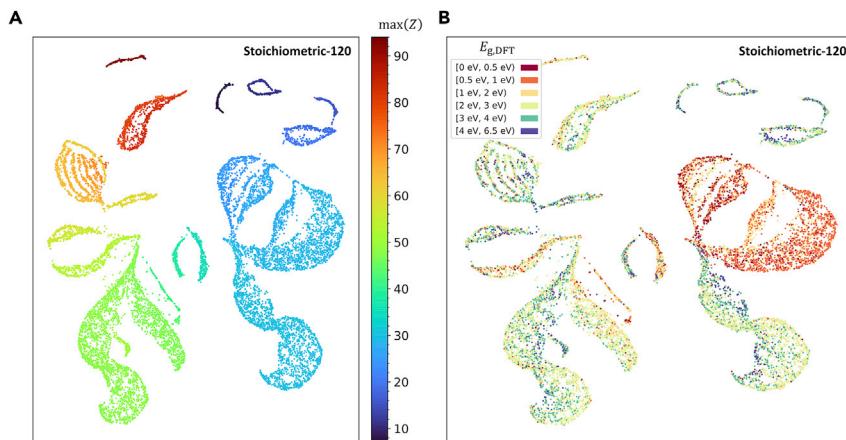


Figure 6. UMAP based on the Stoichiometric-120 features

Unsupervised dimensionality reduction performed using UMAP, with a distance matrix obtained using a Euclidean distance metric of the Stoichiometric-120 encodings for the structures in the QMOF-14482 dataset. The (A) maximum atomic number in each structure, $\text{max}(Z)$, and (B) PBE-D3(BJ) band gaps for the corresponding DFT-optimized structures, $E_{g,\text{DFT}}$, are overlaid on the UMAPs.

on the recently developed coarse-grained SOAP (cg-SOAP) method,¹⁴⁰ may yield improvements in the future.

Similar to what has been done in prior work with revised autocorrelation functions,¹⁴¹ we can use the SOAP similarity kernel to understand the diversity of structures in the QMOF-14482 dataset and identify structural outliers. The most apparent example is the isolated cluster of points in region D of Figure 5. Investigation of these crystal structures indicates that they are predominantly frameworks with high fluorine content, such as MOFs with fluorinated linkers (e.g., refcodes: MUQCEH,¹⁴² HADMOR¹⁴³) or metal-fluoride species (e.g., refcode: EMEJAJ¹⁴⁴), which leads to a large difference in the average SOAP fingerprint compared with most other MOFs in the dataset. The isolated region E of Figure 5 where there is a low-band gap cluster contains polyoxovanadate-based MOFs, some of which have already been investigated for their conductive and electrocatalytic properties (e.g., refcodes: FEYCOE,¹⁴⁵ XEHYEP¹⁴⁶).

While the SOAP-based UMAP is useful for identifying local trends in feature space, significantly greater clustering is observed when using the Stoichiometric-120 encoding. As is evident in Figure 6A, the UMAP based on the Stoichiometric-120 encoding largely partitions the MOF chemical space by the maximum atomic number in each chemical formula. The variations within a given cluster are due to more subtle differences in the elemental fractions and compositional features that compose the Stoichiometric-120 descriptor. Notably, the band gaps are well separated between and within each cluster in the reduced space (Figure 6B). For these reasons, the Stoichiometric-120 UMAP is one useful way to obtain a global view of the QMOF database. For instance, we find that the QMOF-14482 dataset closely overlaps with both the larger QMOF-42349 dataset it was drawn from and the separate CoRE MOF 2019 database⁴ based on the reduced space of Stoichiometric-120 features (Figures S8 and S9). The data in Figure 6B also emphasize how the Zn-containing MOFs (east cluster, $\text{max}(Z) = 30$) tend to have lower band gaps than MOFs with first-row transition metals (south-east cluster, $\text{max}(Z) = 23\text{--}29$) at the PBE-D3(BJ) level of theory. To

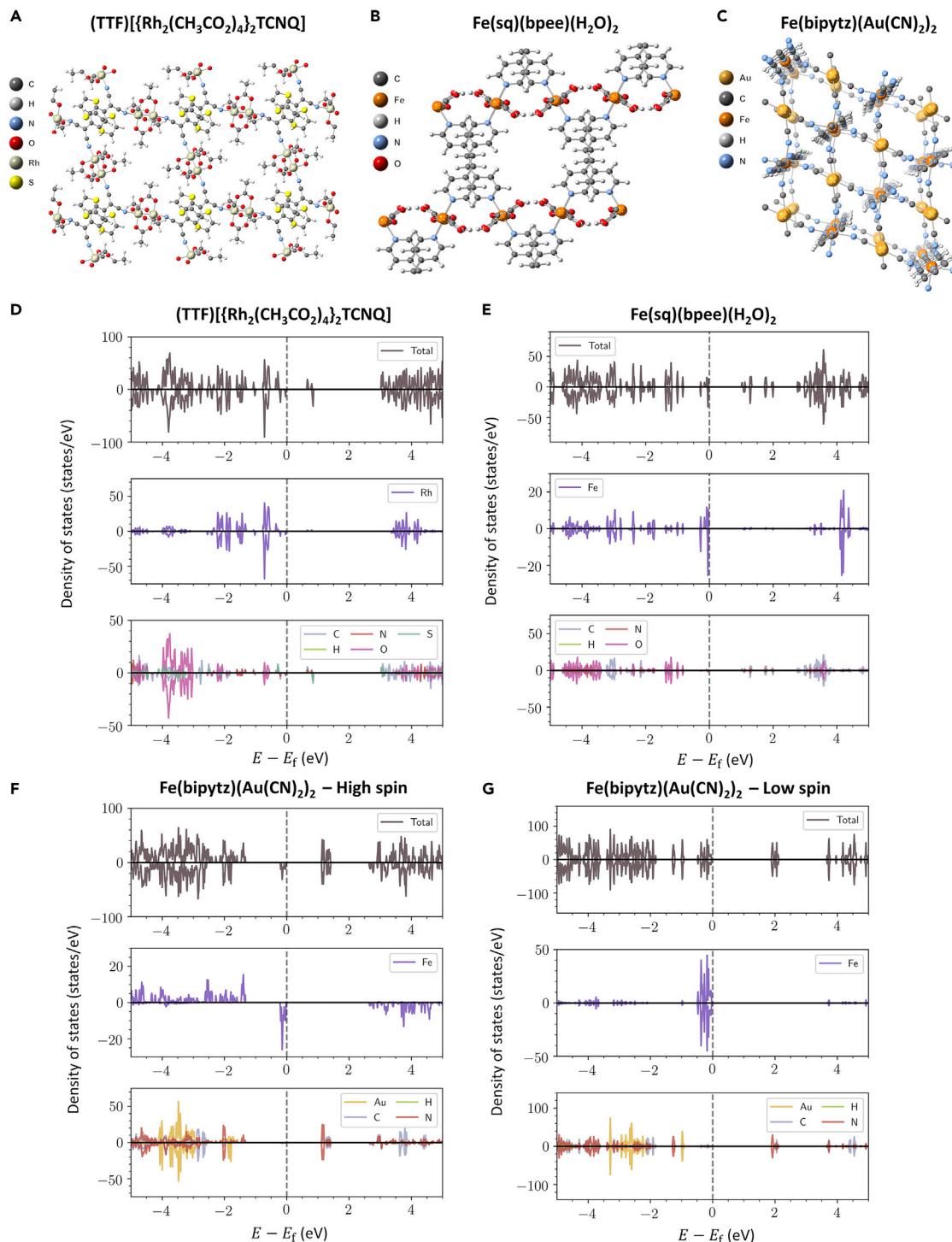


Figure 7. HSE06-D3(BJ) density of states for selected MOFs

Structures of (A) (TTF)[{Rh₂(CH₃CO₂)₄}₂TCNQ], (B) Fe(sq)(bpee)(H₂O)₂, and (C) Fe(bipytz)(Au(CN)₂)₂. Total and projected density of states (DOS) at the HSE06-D3(BJ) level of theory for (D) (TTF)[{Rh₂(CH₃CO₂)₄}₂TCNQ], (E) Fe(sq)(bpee)(H₂O)₂, (F) Fe(bipytz)(Au(CN)₂)₂ (high spin), and (G) Fe(bipytz)(Au(CN)₂)₂ (low spin). The energy, E , in eV is shown with respect to the Fermi level, E_F . DOS values above and below zero refer to the spin-up and spin-down channels, respectively.

enable additional data exploration, interactive versions of the UMAPs are available in the supporting dataset.⁶²

Highlighting notable low band gap MOFs

We conclude by highlighting several framework materials identified in this work that have low band gaps, motivated in part by the search for a greater number of (semi)conducting MOFs. It should be noted that while the PBE-D3(BJ) level of theory makes it possible to generate a sufficiently large database for the purposes of ML model development and to identify structure–property relationships, it is known to underestimate band gaps like essentially all generalized gradient approximation functionals.^{147,148} As such, we carried out full structure relaxations and corresponding band gap calculations using the hybrid-level HSE06-D3(BJ) functional^{149–151} on select materials to generate more accurate band gap predictions. As a point of reference, materials with band gaps in excess of ~4 eV are often classified as electronic insulators, including many of the most commonly studied MOFs (e.g., MOF-5,¹²⁸ UiO-66 [UiO = Universitetet i Oslo],¹⁵² ZIF-8 [ZIF = zeolitic imidazolate framework¹⁵³]).^{63,147} Generally, lower band gaps are necessary to support electrical conductivity (although it is not the sole factor required for achieving high electrical conductivities).⁶³

When the CGCNN model is used to predict the band gaps of all 42,349 structures that compose the QMOF-42349 dataset, one of the lowest band gap materials is predicted to be Ag(DCl)₂ (DCl = 2,5-Cl,Cl-N,N'-dicyanoquinone diamine) (refcode: OTARUX),¹⁵⁴ which is known from experiments to exhibit metallic character via organic radicals that connect the Ag(I) cations.¹⁵⁴ The introduction of radical or redox-active linking units is a well-established strategy to increase the electrical conductivity of framework materials.⁶³ Although Ag(DCl)₂ is arguably best described as a coordination polymer, one notable MOF in the QMOF-42349 dataset with a low predicted band gap and a radical-containing linker is (TTF)[{Rh₂(CH₃CO₂)₄}₂TCNQ] (TTF = tetrathiafulvalene) (refcode: WAQMEJ)¹⁵⁵—a pillared layer framework material built from Rh(II) paddlewheels and a TTF-TCNQ charge-transfer salt (Figure 7A). The HSE06-D3(BJ) band gap for this material is found to be particularly small with a value of 0.71 eV, which can be directly attributed to a reduced conduction-band minimum (CBM) from the TTF and TCNQ components (Figure 7D). Furthermore, the valence band maximum (VBM) also exhibits hybridization between the 4d orbitals of Rh and 2p orbitals of C and N atoms belonging to the radical TCNQ linker, which is important for applications involving electron transport. In contrast, one of the most electronically insulating structures in the QMOF-42349 dataset based on CGCNN-predicted band gap is the nonporous coordination polymer Sr[C₂H₄(SO₃)₂] (refcode: GUTYAW),¹⁵⁶ which has an HSE06-D3(BJ) band gap of 8.36 eV (Figure S12).

Consistent with prior experimental work,¹⁵⁷ we also find several Fe-containing materials in the QMOF-42349 dataset with low band gaps, many of which have not yet been studied for their electronic properties. One representative example is Fe(s-q)(bpee)(H₂O)₂ (bpee = 1,2-bis(4-pyridyl)ethylene; sq = squareate) (refcode: RAX-NEK),¹⁵⁸ shown in Figure 7B, which has a band gap of 1.06 eV at the HSE06-D3(BJ) level of theory. The high-spin Fe(II) species in an octahedral crystal field with t_{2g}⁴e_g² electron configuration dominate the VBM in this material, whereas the bpee linker (as opposed to the bridging sq species or inorganic node) make up the conduction band edge (Figure 7E).

Another noteworthy example is the three-dimensional porous framework material Fe(bipytz)(Au(CN)₂)₂ (bipytz = 3,6-bis(4-pyridyl)-1,2,4,5-tetrazine)

(refcode: LOJLAZ),¹⁵⁹ shown in Figure 7C. At the HSE06-D3(BJ) level of theory, we find that the high spin state exhibits a band gap of 1.17 eV (Figure 7F), similar to that of Fe(sq)(bpee)(H₂O)₂. The projected density of states indicates that the Au(I) species are unrelated to the relatively low band gap; instead, the low band gap can be attributed to the combination of Fe(II) and bipytz linker. Fe(bipytz)(Au(CN)₂)₂ is known to be a spin-crossover framework (with a sharp spin transition around 290 K),¹⁵⁹ and we find the low-spin HSE06-D3(BJ) band gap to be 1.95 eV (Figure 7G), suggesting that the material may have tunable electronic properties as a function of temperature. For the low-spin case, the VBM is composed of Fe 3d orbitals and the CBM is composed of N 2p orbitals. The reduction in band gap from low to high spin state can be rationalized on the basis of crystal field theory. In the high spin state, the Fe(II) centers have a t_{2g}⁴e_g² electronic configuration, whereas in the low spin state they have a t_{2g}⁶e_g⁰ electron configuration. This occupation of the e_g orbitals in the high spin state is directly related to the predicted ~0.8 eV reduction in the band gap compared with the low spin state. For both highlighted Fe-containing frameworks, the band gaps are lower—or comparable in the low spin state for Fe(bipytz)(Au(CN)₂)₂—than those of several Fe-containing MOFs that have been studied for their conductive properties, such as Fe₂(dobdc), Fe₂(dsbdc) (H₄dsbdc = 2,5-disulphydrylbenzene-1,4-dicarboxylic acid), and Fe(bpz).^{157,160} Collectively, these findings demonstrate the practical utility of the QMOF database for identifying MOFs with targeted quantum-chemical properties.

Conclusion

In this work, we have developed a database of quantum-chemical properties for over 14,000 MOF structures (the QMOF database)⁶² via a high-throughput periodic DFT workflow.⁶¹ DFT-computed geometries, energies, band gaps, densities of states, partial charges, spin densities, bond orders, and related electronic structure properties are made publicly available.⁶² We highlight how this database can be used to identify MOFs with targeted electronic structure properties and then develop several ML models to predict the DFT-computed band gaps using descriptors derived from the unoptimized MOF crystal structures. A CGCNN⁶⁸ is found to achieve high predictive performance for this task, making it possible to circumvent large numbers of computationally expensive DFT calculations in future studies. While not as accurate as CGCNN for regression purposes, we show that both the SOAP^{123,124} and composition-based features¹²⁰ can be used to discover otherwise subtle structure–property relationships in the QMOF database via unsupervised dimensionality reduction techniques. Finally, we show how top-performing ML models generated from the database of DFT-computed properties can be used to aid in the discovery of MOFs with desired quantum-chemical properties—in this case, discovering MOFs with low band gaps that could be suitable candidates to consider further for applications in which electrical conductivity is necessary.

Importantly, the QMOF database now makes it possible to pursue several important research directions that are reliant on a large database of quantum-chemical properties for MOFs beyond those directly discussed in this work. For instance, with the success of transfer learning,^{126,161} multitask learning,¹⁶² and Δ-ML¹⁶³ methods in materials research, the QMOF database can serve as a valuable resource to increase the accuracy—and reduce the required training-set size—for ML models tasked with the prediction of new MOF properties not present in the QMOF database. Since the output of any ML models will depend on the chosen density functional approximation, related transfer learning approaches may also prove useful in generalizing ML model predictions to other levels of theory using the PBE-D3(BJ) data as a starting point. Instead of relying on representation approaches that were originally designed

for inorganic solids or small molecules, the QMOF database can also be used to develop better methods for the encoding of MOF structures in ML models. Even outside the areas of high-throughput DFT screening, data mining, and ML, there are countless possible use cases for the QMOF database. As just one example, the DFT-generated properties in the QMOF database could be used to develop and/or benchmark (semi)empirical methods (e.g., tight binding approaches¹⁶⁴ or molecular mechanics force fields¹⁶⁵) with the hopes of achieving high accuracies for MOF structures.

We conclude by noting that the QMOF database should be considered a living resource; several updates to the QMOF database are planned for the future, and we welcome the development of subsets, modifications, and supplements to the database that suit the diverse needs of the MOF community. With all this in mind, we anticipate that the QMOF database will accelerate the materials design and discovery process while being specifically tailored for the chemical space of experimentally realized MOF structures.

EXPERIMENTAL PROCEDURES

Resource availability

Lead contact

Further information and requests for resources and materials should be directed to and will be fulfilled by the lead contact, Andrew S. Rosen (rosen@u.northwestern.edu).

Materials availability

This study did not generate new, unique reagents.

Data and code availability

The landing page for the QMOF database can be found at the following GitHub repository: <https://github.com/ahrens93/QMOF>. Data associated with the QMOF database is hosted via Figshare and has the following permanent DOI: [10.6084/m9.figshare.13147324](https://doi.org/10.6084/m9.figshare.13147324). All data associated with this work is made publicly available, including results from the DFT calculations, Python scripts to reproduce the ML analyses, code to reproduce the automated DFT screening process, and other related resources.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.matt.2021.02.015>.

ACKNOWLEDGMENTS

A.S.R. was supported by a fellowship award through the National Defense Science and Engineering Graduate Fellowship Program, sponsored by the Air Force Research Laboratory (AFRL), the Office of Naval Research, and the Army Research Office. A.S.R. also acknowledges support by a Ryan Fellowship from the International Institute for Nanotechnology and a Presidential Fellowship from The Graduate School at Northwestern University. This work was further supported by the US Department of Energy, Office of Basic Energy Sciences, Division of Chemical Sciences, Geosciences and Biosciences through the Nanoporous Materials Genome Center under award number DE-FG02-17ER16362. The authors acknowledge computing support from the Extreme Science and Engineering Discovery Environment (XSEDE) Stampede2 through allocation CTS180057 supported by National

Science Foundation grant number ACI-1548562 (A.S.R., R.Q.S.), the Quest High Performance Computing (HPC) facility at Northwestern University (A.S.R., S.M.I., R.Q.S.), the Mustang HPC environment via the Department of Defense High Performance Computing Modernization Program at the AFRL (A.S.R.), and the Minnesota Supercomputing Institute at the University of Minnesota (D.R., L.G.). A.A.-G. thanks Dr. Anders G. Frøseth for his generous support.

AUTHOR CONTRIBUTIONS

A.S.R.: conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, writing – original draft, writing – review & editing, visualization, supervision, project administration, funding acquisition. S.M.I.: methodology, software, formal analysis, investigation, data curation, writing – review & editing. D.R.: formal analysis, writing – review & editing. Z.Y.: writing – review & editing. A.A.-G.: writing – review & editing, supervision, funding acquisition. L.G.: writing – review & editing, supervision, funding acquisition. J.M.N.: writing – review & editing, supervision. R.Q.S.: writing – review & editing, supervision, funding acquisition.

DECLARATION OF INTERESTS

R.Q.S. has a financial interest in the start-up company NuMat Technologies, which is seeking to commercialize MOFs. A.A.-G. is a member of the advisory board for Matter.

Received: November 11, 2020

Revised: January 16, 2021

Accepted: February 17, 2021

Published: April 5, 2021

REFERENCES

- Yaghi, O.M., Kalmutzki, M.J., and Diercks, C.S. (2020). Introduction to Reticular Chemistry: Metal-Organic Frameworks and Covalent Organic Frameworks (John Wiley & Sons).
- Kalmutzki, M.J., Hanikel, N., and Yaghi, O.M. (2018). Secondary building units as the turning point in the development of the reticular chemistry of MOFs. *Sci. Adv.* 4, eaat9180.
- Moghadam, P.Z., Li, A., Wiggin, S.B., Tao, A., Maloney, A.G.P., Wood, P.A., et al. (2017). Development of a Cambridge Structural Database subset: a collection of metal-organic frameworks for past, present, and future. *Chem. Mater.* 29, 2618–2625.
- 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. (2019). Advances, updates, and analytics for the computation-ready, experimental metal-organic framework database: CoRE MOF 2019. *J. Chem. Eng. Data* 64, 5985–5998.
- Wilmer, C.E., Leaf, M., Lee, C.Y., Farha, O.K., Hauser, B.G., Hupp, J.T., and Snurr, R.Q. (2012). Large-scale screening of hypothetical metal-organic frameworks. *Nat. Chem.* 4, 83–89.
- Colón, Y.J., Gómez-Gualdrón, D.A., and Snurr, R.Q. (2017). Topologically guided, automated construction of metal-organic frameworks and their evaluation for energy-related applications. *Cryst. Growth Des.* 17, 5801–5810.
- Boyd, P.G., and Woo, T.K. (2016). A generalized method for constructing hypothetical nanoporous materials of any net topology from graph theory. *CrystEngComm* 18, 3777–3792.
- Ejsmont, A., Andreo, J., Lanza, A., Galarda, A., Macreadie, L., Wuttke, S., Canossa, S., Ploetz, E., and Goscianska, J. (2020). Applications of reticular diversity in metal-organic frameworks: an ever-evolving state of the art. *Coord. Chem. Rev.* 430, 213655.
- Colón, Y.J., and Snurr, R.Q. (2014). High-throughput computational screening of metal-organic frameworks. *Chem. Soc. Rev.* 43, 5735–5749.
- Jablonka, K.M., Ongari, D., Moosavi, S.M., and Smit, B. (2020). Big-data science in porous materials: materials genomics and machine learning. *Chem. Rev.* 120, 8066–8129.
- Chibani, S., and Coudert, F.-X. (2020). Machine learning approaches for the prediction of materials properties. *APL Mater.* 8, 80701.
- Anderson, G., Schweitzer, B., Anderson, R., and Gomez-Gualdrón, D.A. (2018). Attainable volumetric targets for adsorption-based hydrogen storage in porous crystals: molecular simulation and machine learning. *J. Phys. Chem. C* 123, 120–130.
- Bucior, B.J., Bobbitt, N.S., Islamoglu, T., Goswami, S., Gopalan, A., Yildirim, T., Farha, O.K., Bagheri, N., and Snurr, R.Q. (2019). Energy-based descriptors to rapidly predict hydrogen storage in metal-organic frameworks. *Mol. Syst. Des. Eng.* 4, 162–174.
- Thornton, A.W., Simon, C.M., Kim, J., Kwon, O., Deeg, K.S., Konstas, K., Pas, S.J., Hill, M.R., Winkler, D.A., Haranczyk, M., and Smit, B. (2017). Materials Genome in action: identifying the performance limits of physical hydrogen storage. *Chem. Mater.* 29, 2844–2854.
- Anderson, R., Rodgers, J., Argueta, E., Biong, A., and Gomez-Gualdrón, D.A. (2018). Role of pore chemistry and topology in the CO₂ capture capabilities of MOFs: from molecular simulation to machine learning. *Chem. Mater.* 30, 6325–6337.
- Dureckova, H., Krykunov, M., Aghajani, M.Z., and Woo, T.K. (2019). 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* 123, 4133–4139.
- Yao, Z., Sanchez-Lengeling, B., Bobbitt, N.S., Bucior, B.J., Kumar, S.G.H., Collins, S.P., et al. (2021). Inverse design of nanoporous

- crystalline reticular materials with deep generative models. *Nat. Mach. Intell.* 76–86.
18. Moghadam, P.Z., Rogge, S.M.J., Li, A., Chow, C.-M., Wieme, J., Moharrami, N., Aragones-Anglada, M., Conduit, G., Gomez-Gualdon, D.A., Van Speybroeck, V., and Fairen-Jimenez, D. (2019). Structure-mechanical stability relations of metal-organic frameworks via machine learning. *Matter* 1, 219–234.
 19. Moosavi, S.M., Chidambaram, A., Talirz, L., Haranczyk, M., Stylianou, K.C., and Smit, B. (2019). Capturing chemical intuition in synthesis of metal-organic frameworks. *Nat. Commun.* 10, 539.
 20. Shi, Z., Yang, W., Deng, X., Cai, C., Yan, Y., Liang, H., Liu, Z., and Qiao, Z. (2020). Machine-learning-assisted high-throughput computational screening of high performance metal-organic frameworks. *Mol. Syst. Des. Eng.* 5, 725–742.
 21. Chong, S., Lee, S., Kim, B., and Kim, J. (2020). Applications of machine learning in metal-organic frameworks. *Coord. Chem. Rev.* 423, 213487.
 22. Mancuso, J.L., Mroz, A.M., Le, K.N., and Hendon, C.H. (2020). Electronic structure modeling of metal-organic frameworks. *Chem. Rev.* 120, 8641–8715.
 23. Raza, A., Sturluson, A., Simon, C., and Fern, X. (2020). Message passing neural networks for partial charge assignment to metal-organic frameworks. *J. Phys. Chem. C* 124, 19070–19082.
 24. Korolev, V.V., Mitrofanov, A., Marchenko, E.I., Eremin, N.N., Tkachenko, V., and Kalmykov, S.N. (2020). Transferable and extensible machine learning derived atomic charges for modeling hybrid nanoporous materials. *Chem. Mater.* 32, 7822–7831.
 25. Chung, Y.G., Camp, J., Haranczyk, M., Sikora, B.J., Bury, W., Krungleviciute, V., Yildirim, T., Farha, O.K., Sholl, D.S., and Snurr, R.Q. (2014). Computation-ready, experimental metal-organic frameworks: a tool to enable high-throughput screening of nanoporous crystals. *Chem. Mater.* 26, 6185–6192.
 26. Nazarian, D., Camp, J.S., and Sholl, D.S. (2016). A comprehensive set of high-quality point charges for simulations of metal-organic frameworks. *Chem. Mater.* 28, 785–793.
 27. He, Y., Cubuk, E.D., Allendorf, M.D., and Reed, E.J. (2018). Metallic metal-organic frameworks predicted by the combination of machine learning methods and ab initio calculations. *J. Phys. Chem. Lett.* 9, 4562–4569.
 28. Saal, J.E., Kirklin, S., Aykol, M., Meredig, B., and Wolverton, C. (2013). Materials design and discovery with high-throughput density functional theory: The Open Quantum Materials Database (OQMD). *JOM* 65, 1501–1509.
 29. Kirklin, S., Saal, J.E., Meredig, B., Thompson, A., Doak, J.W., Aykol, M., et al. (2015). The Open Quantum Materials Database (OQMD): assessing the accuracy of DFT formation energies. *npj Comput. Mater.* 1, 15010.
 30. Perdew, J.P., Burke, K., and Ernzerhof, M. (1996). Generalized gradient approximation made simple. *Phys. Rev. Lett.* 77, 3865–3868.
 31. Abrahams, B.F., Hardie, M.J., Hoskins, B.F., Robson, R., and Williams, G.A. (1992). Topological rearrangement within a single crystal from a honeycomb cadmium cyanide $[Cd(CN)_2]$, 3D net to a diamond net. *J. Am. Chem. Soc.* 114, 10641–10643.
 32. von Lilienfeld, O.A., Müller, K.-R., and Tkatchenko, A. (2020). Exploring chemical compound space with quantum-based machine learning. *Nat. Rev. Chem.* 4, 347–358.
 33. Butler, K.T., Davies, D.W., Cartwright, H., Isayev, O., and Walsh, A. (2018). Machine learning for molecular and materials science. *Nature* 559, 547–555.
 34. Ramprasad, R., Batra, R., Pilania, G., Mannodi-Kanakkithodi, A., and Kim, C. (2017). Machine learning in materials informatics: recent applications and prospects. *npj Comput. Mater.* 3, 54.
 35. Ward, L., and Wolverton, C. (2017). Atomistic calculations and materials informatics: a review. *Curr. Opin. Solid State Mater. Sci.* 21, 167–176.
 36. Chen, C., Zuo, Y., Ye, W., Li, X., Deng, Z., and Ong, S.P. (2020). A critical review of machine learning of energy materials. *Adv. Energy Mater.* 10, 1903242.
 37. Morgan, D., and Jacobs, R. (2020). Opportunities and challenges for machine learning in materials science. *Annu. Rev. Mater. Res.* 50, 71–103.
 38. Saal, J.E., Oliynyk, A.O., and Meredig, B. (2020). Machine learning in materials discovery: confirmed predictions and their underlying approaches. *Annu. Rev. Mater. Res.* 50, 49–69.
 39. Suh, C., Fare, C., Warren, J.A., and Pyzer-Knapp, E.O. (2020). Evolving the materials Genome: how machine learning is fueling the next generation of materials discovery. *Annu. Rev. Mater. Res.* 50, 1–25.
 40. Jain, A., Hautier, G., Moore, C.J., Ping Ong, S., Fischer, C.C., Mueller, T., Persson, K.A., and Ceder, G. (2011). A high-throughput infrastructure for density functional theory calculations. *Comput. Mater. Sci.* 50, 2295–2310.
 41. Jain, A., Ong, S.P., Hautier, G., Chen, W., Richards, W.D., Dacek, S., Cholia, S., Gunter, D., Skinner, D., Ceder, G., and Persson, K.A. (2013). The Materials Project: a materials Genome approach to accelerating materials innovation. *APL Mater.* 1, 11002.
 42. Curtarolo, S., Setyawan, W., Wang, S., Xue, J., Yang, K., Taylor, R.H., Nelson, L.J., Hart, G.L.W., Sanvitto, S., Buongiorno-Nardelli, M., et al. (2012). AFLOWLIB.ORG: a distributed materials properties repository from high-throughput ab initio calculations. *Comput. Mater. Sci.* 58, 227–235.
 43. Winther, K.T., Hoffmann, M.J., Boes, J.R., Mamun, O., Bajdich, M., and Bligaard, T. (2019). Catalysis-hub.org, an open electronic structure database for surface reactions. *Sci. Data* 6, 75.
 44. Borysov, S.S., Geilhufe, R.M., and Balatsky, A.V. (2017). Organic Materials Database: an open-access online database for data mining. *PLoS One* 12, e0171501.
 45. Landis, D.D., Hummelshoj, J.S., Nestorov, S., Greeley, J., Dulak, M., Bligaard, T., Nørskov, J.K., and Jacobsen, K.W. (2012). The computational materials repository. *Comput. Sci. Eng.* 14, 51–57.
 46. Draxl, C., and Scheffler, M. (2018). NOMAD: the FAIR concept for big data-driven materials science. *MRS Bull.* 43, 676–682.
 47. Chanussot, L., Das, A., Goyal, S., Lavril, T., Shuaibi, M., Riviere, M., et al. (2020). The Open Catalyst 2020 (OC20) dataset and community challenges. *arXiv*, 2010.09990.
 48. Ramakrishnan, R., Dral, P.O., Rupp, M., and Von Lilienfeld, O.A. (2014). Quantum chemistry structures and properties of 134 kilo molecules. *Sci. Data* 1, 140022.
 49. Montavon, G., Rupp, M., Gobre, V., Vazquez-Mayagoitia, A., Hansen, K., Tkatchenko, A., Müller, K.-R., and Von Lilienfeld, O.A. (2013). Machine learning of molecular electronic properties in chemical compound space. *N. J. Phys.* 15, 95003.
 50. Smith, J.S., Isayev, O., and Roitberg, A.E. (2017). ANI-1, A data set of 20 million calculated off-equilibrium conformations for organic molecules. *Sci. Data* 4, 170193.
 51. Blau, S., Spotte-Smith, E., Wood, B., Dwarknath, S., and Persson, K. (2020). Accurate, automated density functional theory for complex molecules using on-the-fly error correction. *ChemRxiv*. <https://doi.org/10.26434/chemrxiv.13076030>.
 52. Balcells, D., and Skjelstad, B.B. (2020). The tmQM dataset—quantum geometries and properties of 86k transition metal complexes. *J. Chem. Inf. Model.* 60, 6135–6146.
 53. Gómez-Bombarelli, R., Aguilera-Iparraguirre, J., Hirzel, T.D., Duvenaud, D., MacLaurin, D., Blood-Forsythe, M.A., Chae, H.S., Einzinger, M., Ha, D.-G., Wu, T., et al. (2016). Design of efficient molecular organic light-emitting diodes by a high-throughput virtual screening and experimental approach. *Nat. Mater.* 15, 1120–1127.
 54. Mansouri Tehrani, A., Oliynyk, A.O., Parry, M., Rizvi, Z., Couper, S., Lin, F., Miyagi, L., Sparks, T.D., and Brogch, J. (2018). Machine learning directed search for ultraincompressible, superhard materials. *J. Am. Chem. Soc.* 140, 9844–9853.
 55. Wu, S., Kondo, Y., Kakimoto, M., Yang, B., Yamada, H., Kuwajima, I., et al. (2019). Machine-learning-assisted discovery of polymers with high thermal conductivity using a molecular design algorithm. *npj Comput. Mater.* 5, 66.
 56. Lyu, H., Ji, Z., Wuttke, S., and Yaghie, O.M. (2020). Digital reticular chemistry. *Chem* 6, 2219–2241.
 57. Flores-Leonar, M.M., Mejia-Mendoza, L.M., Aguilar-Granda, A., Sanchez-Lengeling, B., Tribukait, H., Amador-Bedolla, C., and Aspuru-Guzik, A. (2020). Materials acceleration platforms: on the way to

- autonomous experimentation. *Curr. Opin. Green. Sustain. Chem.* 25, 100370.
58. Tabor, D.P., Roch, L.C., Saikin, S.K., Kreisbeck, C., Sheberla, D., Montoya, J.H., Dwarakanath, S., Aykol, M., Ortiz, C., Tribukait, H., et al. (2018). Accelerating the discovery of materials for clean energy in the era of smart automation. *Nat. Rev. Mater.* 3, 5–20.
 59. Coley, C.W., Eyke, N.S., and Jensen, K.F. (2020). Autonomous discovery in the chemical sciences part I: progress. *Angew. Chem. Int. Ed.* 59, 22858–22893.
 60. Coley, C.W., Eyke, N.S., and Jensen, K.F. (2020). Autonomous discovery in the chemical sciences part II: outlook. *Angew. Chem. Int. Ed.* 59, 23414–23436.
 61. Rosen, A.S., Notestein, J.M., and Snurr, R.Q. (2019). Identifying promising metal-organic frameworks for heterogeneous catalysis via high-throughput periodic density functional theory. *J. Comput. Chem.* 40, 1305–1318.
 62. QMof Database. <https://dx.doi.org/10.6084/m9.figshare.13147324>.
 63. Xie, L.S., Skorupskii, G., and Dinca, M. (2020). Electrically conductive metal-organic frameworks. *Chem. Rev.* 120, 8536–8580.
 64. Baumann, A.E., Burns, D.A., Liu, B., and Thoi, V.S. (2019). Metal-organic framework functionalization and design strategies for advanced electrochemical energy storage devices. *Commun. Chem.* 86, <https://doi.org/10.1038/s42004-019-0184-6>.
 65. D'Alessandro, D.M. (2016). Exploiting redox activity in metal-organic frameworks: concepts, trends and perspectives. *Chem. Commun.* 52, 8957–8971.
 66. Downes, C.A., and Marinescu, S.C. (2017). Electrocatalytic metal-organic frameworks for energy applications. *ChemSusChem* 10, 4374–4392.
 67. Allendorf, M.D., Dong, R., Feng, X., Kaskel, S., Matoga, D., and Stavila, V. (2020). Electronic devices using open framework materials. *Chem. Rev.* 120, 8581–8640.
 68. Xie, T., and Grossman, J.C. (2018). Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Phys. Rev. Lett.* 120, 145301.
 69. Nazarian, D., Camp, J.S., Chung, Y.G., Snurr, R.Q., and Sholl, D.S. (2017). Large-scale refinement of metal-organic framework structures using density functional theory. *Chem. Mater.* 29, 2521–2528.
 70. Li, A., Bueno-Perez, R., Wiggin, S., and Fairen-Jimenez, D. (2020). Enabling efficient exploration of metal-organic frameworks in the Cambridge Structural Database. *CrystEngComm* 22, 7152–7161.
 71. Zarabadi-Poor, P., and Marek, R. (2019). Comment on "Database for CO₂ separation performances of MOFs based on computational materials screening.". *ACS Appl. Mater. Interfaces* 11, 16261–16265.
 72. Barthel, S., Alexandrov, E.V., Proserpio, D.M., and Smit, B. (2018). Distinguishing metal-organic frameworks. *Cryst. Growth Des.* 18, 1738–1747.
 73. Altintas, C., Avci, G., Daglar, H., Azar, A.N.V., Erucar, I., Velioglu, S., and Keskin, S. (2019). An extensive comparative analysis of two MOF databases: high-throughput screening of computation-ready MOFs for CH₄ and H₂ adsorption. *J. Mater. Chem. A* 7, 9593–9608.
 74. Velioglu, S., and Keskin, S. (2020). Revealing the effect of structure curations on the simulated CO₂ separation performances of MOFs. *Mater. Adv.* 1, 341–353.
 75. 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., and Snurr, R.Q. (2019). Identification schemes for metal-organic frameworks to enable rapid search and cheminformatics analysis. *Cryst. Growth Des.* 19, 6682–6697.
 76. Chen, T., and Manz, T.A. (2020). Identifying misbonded atoms in the 2019 CoRE metal-organic framework database. *RSC Adv.* 10, 26944–26951.
 77. Grimme, S., Antony, J., Ehrlich, S., and Krieg, H. (2010). A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H–Pu. *J. Chem. Phys.* 132, 154104.
 78. Grimme, S., Ehrlich, S., and Goerigk, L. (2011). Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* 32, 1456–1465.
 79. Kresse, G., and Furthmüller, J. (1996). Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B* 54, 11169–11186.
 80. Kresse, G., and Joubert, D. (1999). From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B* 59, 1758–1775.
 81. Hendon, C.H., Tiana, D., and Walsh, A. (2012). Conductive metal-organic frameworks and networks: fact or fantasy? *Phys. Chem. Chem. Phys.* 14, 13120–13132.
 82. Sun, L., Campbell, M.G., and Dinca, M. (2016). Electrically conductive porous metal-organic frameworks. *Angew. Chem. Int. Ed.* 55, 3566–3579.
 83. Singh, A.K., Montoya, J.H., Gregoire, J.M., and Persson, K.A. (2019). Robust and synthesizable photocatalysts for CO₂ reduction: a data-driven materials discovery. *Nat. Commun.* 10, 443.
 84. Yang, L.-M., Ravindran, P., Vajeeston, P., Svelle, S., and Tilset, M. (2013). A quantum mechanically guided view of Cd-MOF-5 from formation energy, chemical bonding, electronic structure, and optical properties. *Microporous Mesoporous Mater.* 175, 50–58.
 85. Gong, S., Xie, T., Zhu, T., Wang, S., Fadel, E.R., Li, Y., and Grossman, J.C. (2019). Predicting charge density distribution of materials using a local-environment-based graph convolutional network. *Phys. Rev. B* 100, 184103.
 86. Grisafi, A., Fabrizio, A., Meyer, B., Wilkins, D.M., Corminboeuf, C., and Ceriotti, M. (2018). Transferable machine-learning model of the electron density. *ACS Cent. Sci.* 5, 57–64.
 87. Chandrasekaran, A., Kamal, D., Batra, R., Kim, C., Chen, L., and Ramprasad, R. (2019). Solving the electronic structure problem with machine learning. *npj Comput. Mater.* 22, <https://doi.org/10.1038/s41524-019-0162-7>.
 88. Kamal, D., Chandrasekaran, A., Batra, R., and Ramprasad, R. (2020). A charge density prediction model for hydrocarbons using deep neural networks. *Mach. Learn. Sci. Technol.* 1, 25003.
 89. Kolb, B., Lentz, L.C., and Kolpak, A.M. (2017). Discovering charge density functionals and structure-property relationships with PROPHet: a general framework for coupling machine learning and first-principles methods. *Sci. Rep.* 7, 1192.
 90. Pilania, G., Wang, C., Jiang, X., Rajasekaran, S., and Ramprasad, R. (2013). Accelerating materials property predictions using machine learning. *Sci. Rep.* 3, 2810.
 91. Manz, T.A., and Limas, N.G. (2016). Introducing DDEC6 atomic population analysis: part 1. Charge partitioning theory and methodology. *RSC Adv.* 6, 47771–47801.
 92. Limas, N.G., and Manz, T.A. (2016). Introducing DDEC6 atomic population analysis: part 2. Computed results for a wide range of periodic and nonperiodic materials. *RSC Adv.* 6, 45727–45747.
 93. Limas, N.G., and Manz, T.A. (2018). Introducing DDEC6 atomic population analysis: part 4. Efficient parallel computation of net atomic charges, atomic spin moments, bond orders, and more. *RSC Adv.* 8, 2678–2707.
 94. Marenich, A.V., Jerome, S.V., Cramer, C.J., and Truhlar, D.G. (2012). Charge Model 5: an extension of hirshfeld population analysis for the accurate description of molecular interactions in gaseous and condensed phases. *J. Chem. Theor. Comput.* 8, 527–541.
 95. Manz, T.A. (2017). Introducing DDEC6 atomic population analysis: part 3. Comprehensive method to compute bond orders. *RSC Adv.* 7, 45552–45581.
 96. Haldoupis, E., Nair, S., and Sholl, D.S. (2012). Finding MOFs for highly selective CO₂/N₂ adsorption using materials screening based on efficient assignment of atomic point charges. *J. Am. Chem. Soc.* 134, 4313–4323.
 97. Rosen, A.S., Notestein, J.M., and Snurr, R.Q. (2019). Structure-activity relationships that identify Metal-organic framework catalysts for methane activation. *ACS Catal.* 9, 3576–3587.
 98. Yang, B., Wu, X.-P., Gagliardi, L., and Truhlar, D.G. (2019). Methane functionalization by an Ir(III) catalyst supported on a metal-organic framework: an alternative explanation of steric confinement effects. *Theor. Chem. Acc.* 138, 107.
 99. Sours, T., Patel, A., Nørskov, J., Siahrostami, S., and Kulkarni, A. (2020). Circumventing scaling relations in oxygen electrochemistry using metal-organic frameworks. *J. Phys. Chem. Lett.* 11, 10029–10036.

100. Rosen, A.S., Mian, M.R., Islamoglu, T., Chen, H., Farha, O.K., Notestein, J.M., and Snurr, R.Q. (2020). Tuning the redox activity of metal-organic frameworks for enhanced, selective O₂ binding: design rules and ambient temperature O₂ chemisorption in a cobalt-triazolate framework. *J. Am. Chem. Soc.* 142, 4317–4328.
101. Planas, N., Mondloch, J.E., Tussupbayev, S., Borycz, J., Gagliardi, L., Hupp, J.T., Farha, O.K., and Cramer, C.J. (2014). Defining the proton topology of the Zr₆-based metal-organic framework NU-1000. *J. Phys. Chem. Lett.* 5, 3716–3723.
102. Klet, R.C., Liu, Y., Wang, T.C., Hupp, J.T., and Farha, O.K. (2016). Evaluation of Brønsted acidity and proton topology in Zr- and Hf-based metal-organic frameworks using potentiometric acid-base titration. *J. Mater. Chem. A* 4, 1479–1485.
103. Ren, J., Ledwaba, M., Musyoka, N.M., Langmi, H.W., Mathe, M., Liao, S., and Pang, W. (2017). Structural defects in metal-organic frameworks (MOFs): formation, detection and control towards practices of interests. *Coord. Chem. Rev.* 349, 169–197.
104. Deria, P., Mondloch, J.E., Karagiardidi, O., Bury, W., Hupp, J.T., and Farha, O.K. (2014). Beyond post-synthesis modification: evolution of metal-organic frameworks via building block replacement. *Chem. Soc. Rev.* 43, 5896–5912.
105. Syed, Z.H., Sha, F., Zhang, X., Kaphan, D.M., Delferro, M., and Farha, O.K. (2020). Metal-organic framework nodes as a supporting platform for tailoring the activity of metal catalysts. *ACS Catal.* 10, 11556–11566.
106. Ling, S., and Slater, B. (2015). Unusually large band gap changes in breathing metal-organic framework materials. *J. Phys. Chem. C* 119, 16667–16677.
107. Mason, J.A., Oktawiec, J., Taylor, M.K., Hudson, M.R., Rodriguez, J., Bachman, J.E., Gonzalez, M.I., Cervellino, A., Guagliardi, A., Brown, C.M., et al. (2015). Methane storage in flexible metal-organic frameworks with intrinsic thermal management. *Nature* 527, 357–361.
108. Xiao, D.J., Bloch, E.D., Mason, J.A., Queen, W.L., Hudson, M.R., Planas, N., Borycz, J., Dzubak, A.L., Verma, P., Lee, K., et al. (2014). Oxidation of ethane to ethanol by N₂O in a metal-organic framework with coordinatively unsaturated iron(II) sites. *Nat. Chem.* 6, 590–595.
109. Vogiatzis, K.D., Haldoupis, E., Xiao, D.J., Long, J.R., Siepmann, J.I., and Gagliardi, L. (2016). Accelerated computational analysis of metal-organic frameworks for oxidation catalysis. *J. Phys. Chem. C* 120, 18707–18712.
110. Osadchii, D., Olivos Suarez, A.I., Szécsényi, Á., Li, G., Nasalevich, M.A., Dugulan, A.I., Serra-Crespo, P., Hensen, E.J.M., Veber, S.L., Fedin, M.V., et al. (2018). Isolated Fe sites in metal-organic framework catalyze the direct conversion of methane to methanol. *ACS Catal.* 8, 5542–5548.
111. Queen, W.L., Hudson, M.R., Bloch, E.D., Mason, J.A., Gonzalez, M.I., Lee, J.S., Gygi, D., Howe, J.D., Lee, K., Darwish, T.A., et al. (2014). Comprehensive study of carbon dioxide adsorption in the metal-organic frameworks M₂(dobdc) (M = Mg, Mn, Fe, Co, Ni, Cu, Zn). *Chem. Sci.* 5, 4569–4581.
112. Gygi, D., Bloch, E.D., Mason, J.A., Hudson, M.R., Gonzalez, M.I., Siegelman, R.L., et al. (2016). Hydrogen storage in the expanded pore metal-organic frameworks M₂(dobpd) (M = Mg, Mn, Fe, Co, Ni, Zn). *Chem. Mater.* 28, 1128–1138.
113. Verma, P., Vogiatzis, K.D., Planas, N., Borycz, J., Xiao, D.J., Long, J.R., et al. (2015). Mechanism of oxidation of ethane to ethanol at iron(IV)-Oxo sites in magnesium-diluted Fe₂(dobdc). *J. Am. Chem. Soc.* 137, 5770–5781.
114. Xiao, D.J., Oktawiec, J., Milner, P.J., and Long, J.R. (2016). Pore environment effects on catalytic cyclohexane oxidation in expanded Fe₂(dobdc) analogues. *J. Am. Chem. Soc.* 138, 14371–14379.
115. Reed, D.A., Keitz, B.K., Oktawiec, J., Mason, J.A., Runčevski, T., Xiao, D.J., Darago, L.E., Crocellà, V., Bordiga, S., and Long, J.R. (2017). A spin transition mechanism for cooperative adsorption in metal-organic frameworks. *Nature* 550, 96–100.
116. Rosen, A.S., Notestein, J.M., and Snurr, R.Q. (2020). High-valent metal-oxo species at the nodes of metal-triazolate frameworks: the effects of ligand-exchange and two-state reactivity for C-H bond activation. *Angew. Chem. Int. Ed.* 132, 19662–19670.
117. Rosen, A.S., Notestein, J.M., and Snurr, R.Q. (2020). Comparing GGA, GGA+U, and meta-GGA functionals for redox-dependent binding at open metal sites in Metal-organic frameworks. *J. Chem. Phys.* 152, 224101.
118. Ruddigkeit, L., Van Deursen, R., Blum, L.C., and Reymond, J.-L. (2012). Enumeration of 166 billion organic small molecules in the chemical universe database GDB-17. *J. Chem. Inf. Model.* 52, 2864–2875.
119. Li, H., Eddaoudi, M., O'Keeffe, M., and Yaghi, O.M. (1999). Design and synthesis of an exceptionally stable and highly porous metal-organic framework. *Nature* 402, 276–279.
120. Meredig, B., Agrawal, A., Kirklin, S., Saal, J.E., Doak, J.W., Thompson, A., Zhang, K., Choudhary, A., and Wolverton, C. (2014). Combinatorial screening for new materials in unconstrained composition space with machine learning. *Phys. Rev. B* 89, 94104.
121. Faber, F., Lindmaa, A., von Lilienfeld, O.A., and Armiento, R. (2015). Crystal structure representations for machine learning models of formation energies. *Int. J. Quant. Chem.* 115, 1094–1101.
122. Lam Pham, T., Kino, H., Terakura, K., Miyake, T., Tsuda, K., Takigawa, I., and Chi Dam, H. (2017). Machine learning reveals orbital interaction in materials. *Sci. Technol. Adv. Mater.* 18, 756–765.
123. Bartók, A.P., Kondor, R., and Csányi, G. (2013). On representing chemical environments. *Phys. Rev. B* 87, 184115.
124. De, S., Bartók, A.P., Csányi, G., and Ceriotti, M. (2016). Comparing molecules and solids across structural and alchemical space. *Phys. Chem. Chem. Phys.* 18, 13754–13769.
125. Pronobis, W., and Müller, K.-R.. Kernel methods for quantum chemistry. In *Machine Learning Meets Quantum Physics*; Schütt K.T., Chmiela S., von Lilienfeld O.A., Tkatchenko A., Tsuda K., Müller K.-R., Springer; pp 25–36.
126. Chen, C., Ye, W., Zuo, Y., Zheng, C., and Ong, S.P. (2019). Graph networks as a universal machine learning framework for molecules and crystals. *Chem. Mater.* 31, 3564–3572.
127. Olsthoorn, B., Geihufe, R.M., Borysov, S.S., and Balatsky, A.V. (2019). Band gap prediction for large organic crystal structures with machine learning. *Adv. Quan. Technol.* 2, 1900023.
128. Gascon, J., Hernández-Alonso, M.D., Almeida, A.R., van Klink, G.P.M., Kapteijn, F., and Mul, G. (2008). Isoreticular MOFs as efficient photocatalysts with tunable band gap: an operando FTIR study of the photoinduced oxidation of propylene. *ChemSusChem* 1, 981–983.
129. Stanzione, D., Barth, B., Gaffney, N., Gaither, K., Hempel, C., Minyard, T., Mehringer, S., Werner, E., Tufo, H., Panda, D., and Teller, P. (2017). Stampede 2: the evolution of an XSEDE supercomputer. In *Proceedings of the Practice and Experience in Advanced Research Computing 2017 on Sustainability, Success and Impact*; D. Hart, ed. (Association for Computing Machinery). <https://doi.org/10.1145/3093338.3093385>.
130. Towns, J., Cockerill, T., Dahan, M., Foster, I., Gaither, K., Grimshaw, A., Hazlewood, V., Lathrop, S., Lifka, D., Peterson, G.D., et al. (2014). XSEDE: accelerating scientific discovery. *Comput. Sci. Eng.* 16, 62–74.
131. McInnes, L., Healy, J., and Melville, J. (2018). UMAP: uniform manifold approximation and projection for dimension reduction. *arXiv*, 1802.03426.
132. Leland, M., John, H., Nathaniel, S., and Lukas, G. (2018). UMAP: uniform manifold approximation and projection. *J. Open Source Softw.* 3, 861.
133. Zeng, X.-Z., Zhang, A.-Y., Bu, D., and Li, Y.-W. (2013). Hydrothermal synthesis, structure and thermal properties of a novel three-dimensional La(III)-Sebacate framework. *Chin. J. Struct. Chem.* 32, 120–124.
134. Zhang, Z., Zhao, H., Matsushita, M.M., Awaga, K., and Dunbar, K.R. (2014). A new metal-organic hybrid material with intrinsic resistance-based bistability: monitoring in situ room temperature switching behavior. *J. Mater. Chem. C* 2, 399–404.
135. Lopez, N., Zhao, H., Ota, A., Prosvirin, A.V., Reinheimer, E.W., and Dunbar, K.R. (2010). Unprecedented binary semiconductors based on TCNQ: single-crystal X-ray studies and physical properties of Cu(TCNQ_x) X = Cl, Br. *Adv. Mater.* 22, 986–989.
136. Aubrey, M.L., Kapelewski, M.T., Melville, J.F., Oktawiec, J., Presti, D., Gagliardi, L., and Long, J.R. (2019). Chemiresistive detection of gaseous hydrocarbons and interrogation of charge transport in Cu[Ni(2,3-pyrazinedithiolate)₂] by gas adsorption. *J. Am. Chem. Soc.* 141, 5005–5013.

137. Peng, Y.-L., Pham, T., Li, P., Wang, T., Chen, Y., Chen, K.-J., Forrest, K.A., Space, B., Cheng, P., Zaworotko, M.J., and Zhang, Z. (2018). Robust ultramicroporous metal-organic frameworks with benchmark affinity for acetylene. *Angew. Chem. Int. Ed.* 57, 10971–10975.
138. Takaishi, S., Hosoda, M., Kajiwara, T., Miyasaka, H., Yamashita, M., Nakanishi, Y., et al. (2009). Electroconductive porous coordination polymer Cu[Cu(pdtt)₂] composed of donor and acceptor building units. *Inorg. Chem.* 48, 9048–9050.
139. Kobayashi, Y., Jacobs, B., Allendorf, M.D., and Long, J.R. (2010). Conductivity, doping, and redox chemistry of a microporous dithiolene-based metal-organic framework. *Chem. Mater.* 22, 4120–4122.
140. Nicholas, T.C., Goodwin, A., and Deringer, V.L. (2020). Understanding the geometric diversity of inorganic and hybrid frameworks through structural coarse-graining. *Chem. Sci.* 11, 12580–12587.
141. Moosavi, S.M., Nandy, A., Jablonka, K.M., Ongari, D., Janet, J.P., Boyd, P.G., Lee, Y., Smit, B., and Kulik, H.J. (2020). Understanding the diversity of the metal-organic framework ecosystem. *Nat. Commun.* 11, 4068.
142. Hulvey, Z., Furman, J.D., Turner, S.A., Tang, M., and Cheetham, A.K. (2010). Dimensionality trends in metal-organic frameworks containing perfluorinated or nonfluorinated benzenedicarboxylates. *Cryst. Growth Des.* 10, 2041–2043.
143. Taylor, M.K., Runcevski, T., Oktawiec, J., Gonzalez, M.I., Siegelman, R.L., Mason, J.A., Ye, J., Brown, C.M., and Long, J.R. (2016). Tuning the adsorption-induced phase change in the flexible metal-organic framework Co(bdp). *J. Am. Chem. Soc.* 138, 15019–15026.
144. Cui, X., Chen, K., Xing, H., Yang, Q., Krishna, R., Bao, Z., Wu, H., Zhou, W., Dong, X., Han, Y., et al. (2016). Pore chemistry and size control in hybrid porous materials for acetylene capture from ethylene. *Science* 353, 141–144.
145. Li, S., Zhang, L., Lu, B., Yan, E., Wang, T., Li, L., Wang, J., Yu, Y., and Mu, Q. (2018). A new polyoxovanadate-based metal-organic framework: synthesis, structure and photo-/electro-catalytic properties. *New J. Chem.* 42, 7247–7253.
146. Yan, B., Luo, J., Dube, P., Sefat, A.S., Greedan, J.E., and Maggard, P.A. (2006). Spin-gap formation and thermal structural studies in reduced hybrid layered vanadates. *Inorg. Chem.* 45, 5109–5118.
147. Choudhuri, I., and Truhlar, D.G. (2019). HLE17: an efficient way to predict band gaps of complex materials. *J. Phys. Chem. C* 123, 17416–17424.
148. Borlido, P., Aull, T., Huran, A.W., Tran, F., Marques, M.A.L., and Botti, S. (2019). Large-scale benchmark of exchange-correlation functionals for the determination of electronic band gaps of solids. *J. Chem. Theor. Comput.* 15, 5069–5079.
149. Heyd, J., Scuseria, G.E., and Ernzerhof, M. (2003). Hybrid functionals based on a screened Coulomb potential. *J. Chem. Phys.* 118, 8207–8215.
150. Krka, A.V., Vydrov, O.A., Izmaylov, A.F., and Scuseria, G.E. (2006). Influence of the exchange screening parameter on the performance of screened hybrid functionals. *J. Chem. Phys.* 125, 224106.
151. Moellmann, J., and Grimme, S. (2014). DFT-D3 study of some molecular crystals. *J. Phys. Chem. C* 118, 7615–7621.
152. Valenzano, L., Civalleri, B., Chavan, S., Bordiga, S., Nilsen, M.H., Jakobsen, S., Lillerud, K.P., and Lamberti, C. (2011). Disclosing the complex structure of UiO-66 metal organic framework: a synergic combination of experiment and theory. *Chem. Mater.* 23, 1700–1718.
153. Saliba, D., Ammar, M., Rammal, M., Al-Ghoul, M., and Hmaded, M. (2018). Crystal growth of ZIF-8, ZIF-67, and their mixed-metal derivatives. *J. Am. Chem. Soc.* 140, 1812–1823.
154. Naito, T., Kakizaki, A., Inabe, T., Sakai, R., Nishibori, E., and Sawa, H. (2011). Growth of nanocrystals in a single crystal of different materials: a way of giving function to molecular crystals. *Cryst. Growth Des.* 11, 501–506.
155. Sekine, Y., Tonouchi, M., Yokoyama, T., Kosaka, W., and Miyasaka, H. (2017). Built-in TTF-TCNO charge-transfer salts in π -stacked pillared layer frameworks. *CrystEngComm* 19, 2300–2304.
156. Salami, T.O., Patterson, S.N., Jones, V.D., Masello, A., and Abboud, K.A. (2009). Synthesis, characterization, thermal study, and crystal structure of a new layered alkaline earth metal sulfonate: Sr[C₂H₄(SO₃)₂]. *Inorg. Chem. Commun.* 12, 1150–1153.
157. Sun, L., Hendon, C.H., Park, S.S., Tulchinsky, Y., Wan, R., Wang, F., Walsh, A., and Dincă, M. (2017). Is iron unique in promoting electrical conductivity in MOFs? *Chem. Sci.* 8, 4450–4457.
158. Manna, S.C., Zangrandi, E., Ribas, J., and Chaudhuri, N.R. (2005). Squarato-bridged polymeric networks of iron(II) with N-donor coligands: syntheses, crystal structures and magnetic properties. *Inorgan. Chim. Acta* 358, 4497–4504.
159. Clements, J.E., Price, J.R., Neville, S.M., and Kepert, C.J. (2014). Perturbation of spin crossover behavior by covalent post-synthetic modification of a porous metal-organic framework. *Angew. Chem. Int. Ed.* 126, 10328–10332.
160. Spirk, S., Grzywa, M., Reschke, S., Fischer, J.K.H., Sippel, P., Demeshko, S., von Nidda, H.-A., and Volkmer, D. (2017). Single-crystal to single-crystal transformation of a nonporous Fe(II) Metal-organic framework into a porous metal-organic framework via a solid-state reaction. *Inorg. Chem.* 56, 12337–12347.
161. Lee, J., and Asahi, R. (2020). Transfer learning for materials informatics using crystal graph convolutional neural network. *arXiv*, 2007.09932.
162. Sanyal, S., Balachandran, J., Yadati, N., Kumar, A., Rajagopalan, P., Sanyal, S., and Talukdar, P. (2018). MT-CGCNN: integrating crystal graph convolutional neural network with multitask learning for material property prediction. *arXiv*, 1811.05660.
163. Ramakrishnan, R., Dral, P.O., Rupp, M., and von Lilienfeld, O.A. (2015). Big data meets quantum chemistry approximations: the Δ -machine learning approach. *J. Chem. Theor. Comput.* 11, 2087–2096.
164. Bannwarth, C., Ehlert, S., and Grimme, S. (2019). GFN2-xTB—an accurate and broadly parametrized self-consistent tight-binding quantum chemical method with multipole electrostatics and density-dependent dispersion contributions. *J. Chem. Theor. Comput.* 15, 1652–1671.
165. Spicher, S., and Grimme, S. (2020). Robust atomistic modeling of materials, organometallic, and biochemical systems. *Angew. Chem. Int. Ed.* 59, 15665–15673.