I&EC research
Industrial & Engineering Chemistry Research

pubs.acs.org/IECR

Article

Thermal Stability of Metal–Organic Frameworks (MOFs): Concept, Determination, and Model Prediction Using Computational Chemistry and Machine Learning

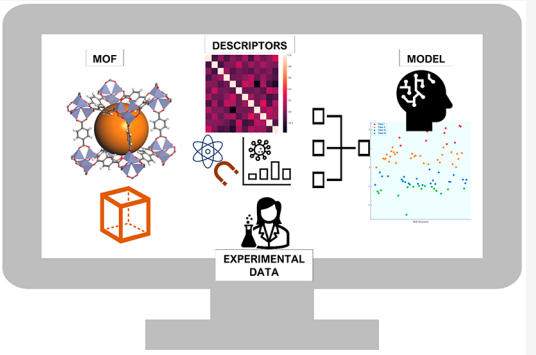
Harold U. Escobar-Hernandez, Lisa M. Pérez, Pingfan Hu, Fernando A. Soto, Maria I. Papadaki, Hong-Cai Zhou, and Qingsheng Wang*

Cite This: *Ind. Eng. Chem. Res.* 2022, 61, 5853–5862

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: The indubitable rise of metal–organic framework (MOF) technology has opened the potential for commercialization as alternative materials with a versatile number of applications that range from catalysis to greenhouse gas capture. However, there are several factors that constrain the direct scale-up of MOFs from laboratory to industrial plant given the insufficient knowledge about the overall safety in synthesis processes. This article focuses on the study of MOF thermal stability, from concept to prediction, and the factors that influence such stability. The core of this work is a thermal stability prediction model for MOFs. This model can be applied to existing and new MOF structures, and it will allow for an estimation of the thermal stability temperature range of MOFs. This work contributes to the overall advancement of MOF technology and the efforts for its commercial use at industrial scale, combining both experimental data and computational techniques.



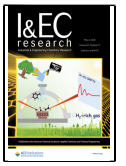
1. INTRODUCTION

The increasing interest in metal–organic framework (MOF) research is closely related to many applications because of their unique properties.[1−3] Due to their significantly large internal surface area derived from their morphology, MOFs have historically shown a direct use in adsorption-driven applications. This includes gas storage, separation, catalysis, water remediation, and energy storage. Altogether, MOFs have potential for everyday life applications.[4−6] MOFs are polycrystalline porous frameworks formed by inorganic metal clusters or secondary building units, joined by organic linkers.[1,7] The exchangeable character of the secondary building units and organic linkers allows the existence of a large number of stable MOF structures, by combining different transition metals with different topologies. Different synthesis techniques and applications have been the focus of mainstream MOF research. This research has rapidly reached industrial and academic interest from various companies like BASF to automotive companies and start-ups.[1] The research, development, scale-up, and commercialization of MOFs is undeniably an emerging and rapidly growing field with potential impact.

The importance of thermal stability in MOFs has been stated on different fronts. MOF functionality and structural integrity can be significantly compromised under different thermal conditions. This makes some MOFs unfeasible for applications. According to Devic and Serre, there is a clear gap for studies to look specifically into the thermal stability of MOFs and the different temperature ranges.[8] Besides, the understanding in terms of the toxicity, stability, and overall safety of MOF technology is limited.[9−11] Furthermore, it is crucial to understand and assess any inherent risk related to the use of MOFs in practical applications associated with their thermal stability. Therefore, this article aims to use a comprehensive approach to systematically explain the fundamental factors influencing MOF stability by using computational chemistry tools. This will contribute to a better understanding of MOF technology.

Yaghi et al. pointed out that the knowledge of precise crystal structures was key for MOF development as porous structures, including modeling and quantum chemical calculations for material discovery and to study properties.[1] Consequently, different computational chemistry techniques and software were used to study MOFs since 2010. This includes the use of density functional theory (DFT) calculations, *ab initio*, and grand canonical Monte Carlo simulations to better understand CO_2 sequestration, gas absorption, and MOF bulk properties

Received: February 17, 2022
Revised: April 4, 2022
Accepted: April 7, 2022
Published: April 21, 2022



https://doi.org/10.1021/acs.iecr.2c00561
Ind. Eng. Chem. Res. 2022, 61, 5853−5862

prediction.[12−16] Meanwhile, several studies have explored the application of machine learning and quantitative structure−property relationships (QSPR) in the realm of MOFs, specifically to predict their gas-sorption-related properties and behavior.[17−20] It is worth noting that these studies have focused on MOF properties for specific applications, but thermal stability has not been thoroughly studied using these methods. Rosen et al.[21,22] made a significant contribution to MOFs computational studies by developing the Quantum MOF database (QMOF) with information obtained through computational chemistry. Their work is key to accelerating MOF discovery since it uses methods that are more robust.

This article focuses on the prediction of such stability property, including the factors that influence it. The core of this work is a thermal stability prediction model that can be applied to existing and new MOF structures. This model will serve to estimate the stability temperature range of MOFs. The method of choice for this predictive model is QSPR, a well-known predictive method based on information encoded in the compound chemical structure.[23] The aspects explored will contribute to the advancement of MOF technology and the efforts for its commercial use at the industrial scale.

2. THERMAL STABILITY OF MOFS

MOF stability can be defined as the ability of MOFs to maintain structural integrity under different conditions (e.g., pressure, temperature, solvent, water presence). This is also defined as MOF resilience; once MOFs undergo abnormal conditions and even after structural changes, they show restoration of the secondary building unit and organic linker bond. The concept of MOF thermal stability does not have a consensus definition in literature. Yet, thermal stability tests upon synthesis have become standard to verify and report the structural integrity of MOFs under certain temperature conditions. Thermogravimetry analysis (TGA) is often used to test for thermal stability. However, these data are not always thoroughly analyzed or used, but they are still presented as part of supporting information in MOF synthesis focused research.[24] The impact of thermal stability highly depends on the MOF application. For instance, in applications of liquid or gas phase adsorption, thermal stability plays an important role in ensuring the integrity of the MOF during the interaction with the fluid, as well as high-pressure conditions. Yuan et al. point out that MOF stability is highly dependent on several factors, including framework structure, particle size, crystal defects, and operation conditions. They also highlight the importance of addressing MOF stability to open the possibilities for industrialization of the materials and exploitation of theirfull potential.[25]

There are two factors identified to contribute to the structural integrity of MOFs: linker characteristics and metal−linker bonds. Linker characteristics influence the mechanical stability of the pores formed and therefore contribute to the overall MOF stability. Studies on linker thermolysis suggest that the linker characteristics dominate the stable character of the MOF. Kinetic factors also influence MOF stability. For instance, linker length decreases MOF stability opposite to what is thermodynamically expected. This is explained by the rigidity of linkers directly creating more stable MOFs. Regarding metal−linker bonds, Zhou et al. demonstrate that for carboxylate-based MOFs, stability is closely related to the metal−oxygen bond. In the case of the coordination bonds constructed by M^{3+} ions oxides, stability changes with the cation in the form V < Fe < Cr < Al. On the other hand, ion M^{4+}−carboxylate interaction is strong and contributes to the stability of the MOF by requiring more ligands and higher connection cluster numbers. This prevents guest molecules' intervention. Additionally, specific linkers are associated with higher thermal stabilities (e.g., $[M(OH)(COO)_2]n$ chains).[25,26] Stability has also been linked to the porosity of the framework given their relationship to the structure.[27−29]

Gas storage is perhaps the most relevant MOF application related to thermal stability. This is due to MOFs' high internal surface area and incredibly fast adsorption and desorption kinetics. It is crucial to consider how field conditions differ from those of synthesis, which poses a question regarding the scale-up capabilities of MOFs. The current experimental methods that simulate gas adsorption do not reflect the exact conditions in the field.[1,30] In field conditions, system components such as valves, piping, and sensors need to be considered when calculating the real equipment capacity. HKUST-1, a Cu-based MOF, is considered feasible for methane storage in natural gas-driven vehicles. Using the basics of gas sorption, it is straightforward to see that the heat of absorption Q_{st} generated to store 20 kg of methane can be estimated to be 25000 kJ. This significant amount of thermal energy would be released upon tank fueling. Considering the heat capacity of the MOF, an estimated temperature difference of 120 °C is expected after each fueling event.[1] HKUST-1 has a reported stability temperature of ∼336 °C.[31] This temperature increase in the system would result in a significant nearness to the thermal point required to compromise the MOF structural integrity, a lower life cycle, and compromised mechanical integrity of the tank. These effects have a direct impact on the safety of the vehicle and are linked to tank overpressure. This is yet another proof that further analysis into scale-up capabilities of MOFs is needed. Since it has been shown that temperature is not the only factor to consider in reaction safety, in fact, overpressure accounts for high consequence scenarios.[32]

Thermal stability of MOFs is commonly measured using TGA or differential-scanning calorimetry (DSC). In both methods, the sample is heated at controlled rates while monitoring the apparent mass change and temperature profile. The changes observed in the different stages of both TGA and DSC results have served as a clear indication of the overall stability of MOFs. Different mass and structure stages can be identified, and these findings can be extended to estimate the stability of new MOFs. The time to gather experimental TGA data, including equipment calibration, makes data acquisition and analysis of MOFs time-consuming.[33] MOF TGA result interpretation is done around the drastic changes of slope in the data. It can be generalized that the early stages of mass loss correspond to desolvation or dehydration of the remaining additional molecules in the MOF pores. This is followed by a more significant noticeable mass loss corresponding to the point of stability. This change is considered a decarboxylation reaction of the organic linker, and it contributes to the MOF's loss of structural integrity.[34] Although this extends to MOFs across the board, other mechanisms should be taken into account depending on the organic linker structure, such as double-bond breakage in the internal structure of the linker as the first step of loss of structural integrity, followed by decarboxylation.

https://doi.org/10.1021/acs.iecr.2c00561
Ind. Eng. Chem. Res. 2022, 61, 5853−5862

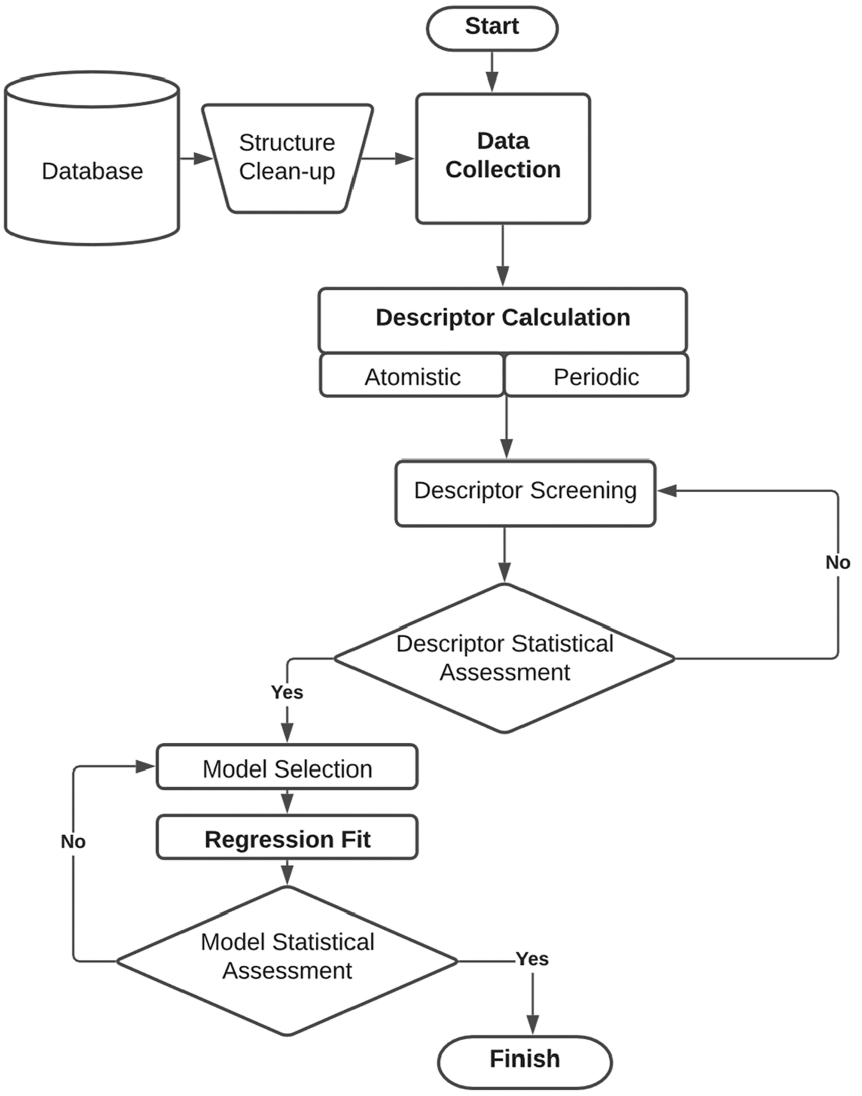


Figure 1. Summary of QSPR methodology.

3. METHODOLOGY

QSPR models determine a mathematical relationship between a physicochemical property and a series of descriptive parameters (descriptors) inherent to the chemical structure of the substance. Quintero et al. thoroughly summarize safety-related properties with QSPR studies, including but not limited to toxicity, explosibility, flammability, and self-reactive properties.[35] QSPR have been especially highlighted for flammability properties such as flash point, minimum ignition energy, autoignition temperature, and flammability limits.[36−39] For this research, the property of interest is the thermal stability point as defined in the following sections.

In summary, the QSPR process comprises five stages. First is the collection of experimental data with the target property information. Additionally, collection of the clean and optimized molecular structures is needed for the second step, descriptor calculation. Third is descriptor screening followed by application of a regression method of choice to fit the calculated descriptor and target data to generate the QSPR model. Last, the evaluation and validation of the model are performed.[38] Figure 1 illustrates the summary of the QSPR technique used in this work.

3.1. Data Collection. This stage involved the collection of known MOF structures. Part of these data were available from previous work and other data were extracted from the literature, commonly reported in the supporting information. The data extracted followed the procedure described in the methodology section. One challenge to build the model was to include structures with known experimental data and workable structure files available. The main source of MOF structures was the Cambridge Crystallographic Database Centre (CCDC).[40] However, many researchers have pointed out that the structure files do not reflect the accurate molecular structure of the MOF due to the disorder, solvent, water, or other guest molecules as part of the deposited file.[22,41] Although this can be overcome by manual cleanup of the structures, in most cases, this process will affect the MOF structure. Even after geometry optimization, there is no certainty of the equivalency between experimental and optimized structures.

The CoRE MOF database was used to find the structures, which had TGA experimental data.[42,43] This database is built from structures in the CCDC and includes structures where most of these discrepancies have been cleaned up using a

https://doi.org/10.1021/acs.iecr.2c00561
Ind. Eng. Chem. Res. 2022, 61, 5853−5862

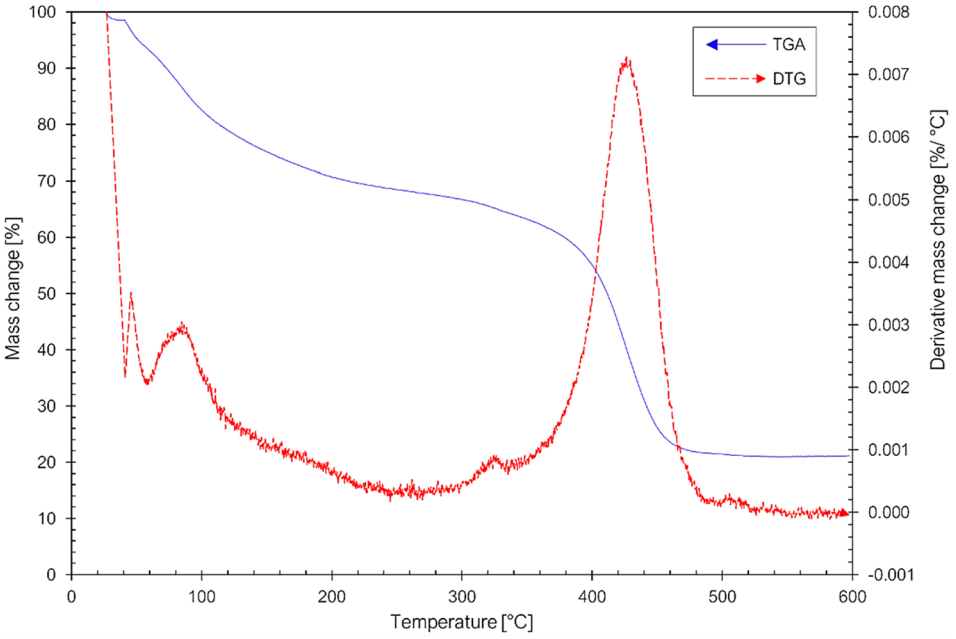


Figure 2. Example of TGA sample and DTG curves for PCN-224(Zr).

topology-based crystal generator. This procedure is simulation-wise equivalent to MOF experimental activation.[42] The dataset used in this research was built with a subset of structures present in the CoRE database. Minor structure cleanup such as removal of duplicate atoms or occupancy correction created by symmetry was needed. It was important to only use accurate and clean structures. This would ensure discarding or repairing of problematic MOFs with omitted, unbounded, overlapping, or extra atoms, and other structural issues.[22] Therefore, the subset of MOFs used was smaller than originally expected, with a final dataset size of 80.

3.1.1. Target Property. As mentioned, the target property for this work is the thermal stability point. Thermal stability point is analogous to onset temperature. These two concepts allow describing physical phenomena punctually, even though these concepts might not have a physical meaning in and of themselves. One concrete example in runaway reaction research is the term "onset" temperature, defined as the temperature point where reactions stop being dominated by kinetics and lose control, resulting in a notable change in the temperature rate or derivative.[44−46] This concept is highly subjective and dependent on the equipment used. Different scientists determine different points of inflections depending on their data using expert judgment and subject knowledge. In addition, different equipment data provide different precision in the measures, therefore determining different "onset" temperatures.[47] Overall, the concept aims to explain the same phenomenon at the start of untempered reactions. This explanation serves to present our understanding of the MOF thermal stability point. We define the thermal stability point as the TGA data point where the rate of mass loss changes drastically according to the slope of the mass loss curve. From a chemical perspective, this is the point where the MOF has ceased to exist as intended and lost its structural integrity. Since the phenomenon of the stability loss does not necessarily mean decomposition of the MOF, this point is not referred to as decomposition or onset temperature but as thermal stability point. The target property of thermal stability point is calculated from experimental TGA results.

In Figure 2, the TGA profile for the PCN-224(Zr) MOF is depicted. The typical behavior of the weight change upon heat exposure in TGA is represented by the sudden changes in the slope of the curve. Significant changes or peaks in the derivative thermogravimetric curve (DTG) represent changes in the structural integrity of the MOF. On the basis of expert knowledge in the MOF synthesis field, the first peak corresponds to desolvation and dehydration of the structure. Moreover, the most significant peak is associated with the overall loss of structural integrity of the MOF itself, and it corresponds to the maximum DTG peak temperature value.

The procedure for data treatment to find thermal stability points consisted of capturing DTG plots for each MOF structure using the Lagrange derivative method to calculate the slope at each TGA point using the data extraction tool WebPlot digitizer.[48] The maximum value from the DTG profiles was correlated with the temperature to establish the thermal stability point for each MOF. These results are the target data used in the QSPR analysis and can be found in Table S1. We used the *k*-means clustering method to gain intuition about the target data and split the data into four groups as seen in Table 2. This defined the ranges of temperature of each group while avoiding randomness and closeness of the temperature points.

3.1.2. Descriptor Calculation. Descriptors are molecular structure properties with hidden relationships with the target property.[23] Some descriptors are calculated using simulation tools and others gathered from the summary of the structure information. For this step, given the complexity of the crystalline character of MOFs, different resources were used to calculate MOF descriptors. Materials Studio[49] was first used for final structure cleanup and to calculate some of the descriptors included in the final set. All the cells were regularized or normalized according to the lattice lengths of the unit cells, as reported in the CCDC. Given some cells in the dataset exhibit significantly large lattices, supercells were

https://doi.org/10.1021/acs.iecr.2c00561
Ind. Eng. Chem. Res. 2022, 61, 5853−5862

Table 1. Descriptors Selected for QSPR Model after Screening

descriptor	definition	type	reference[17,50,55,57,58]	software used for calculation
Fraction of p valence electrons	Composition-weighted fraction of p fraction of valence electrons: valence electrons in the structure that represent p states	Atomic	Meredig et al.	Python
Electron affinity mean	Composition-weighted arithmetic mean of the electron affinity of the elements in the structure	Atomic	He et al.	Python
Ionization energy geometric mean	Composition-weighted geometric mean of the ionization energy of the elements in the structure	Atomic	He et al.	Python
Period number standard deviation	Composition-weighted standard deviation of the period of the elements in the structure	Atomic	He et al.	Python
Melting temperature standard deviation	Composition-weighted standard deviation of the melting temperature point of the elements in the structure	Atomic	He et al.	Python
Metal electronegativity	Electronegativity of the MOF metal	Atomic	Yaghi et al.	Materials Studio
Density	Structure density [g/cm^3]	Periodic	Sarkisov et al.	PoreBlazer
Molecular weight	Molecular weight of the unit cell [g/mol]	Periodic	Sarkisov et al.	PoreBlazer
Accessible surface area S_{AC}	Mass-specific surface area formed by the path of the probe particle center over the surface of the atoms at the boundary of the structure [m^2/cm^3]	Periodic	Sarkisov et al.	PoreBlazer
Volume fraction $V_{(PO,T)}/V$	Volume ratio between the total probe-occupiable volume and the total framework volume	Periodic	Sarkisov et al.	PoreBlazer
Cell volume	Total volume of the framework cell [Å^3]	Periodic	Materials Studio	Materials Studio
Connolly surface area	Mass-specific surface area formed by the path of the probe particle limit over the surface of the atoms at the boundary of the structure	Periodic	Connolly et al.	Materials Studio

created to simulate cubic geometries with an approximate volume that was within 5 standard deviations of the maximum original unit cell volume (7×10^5 Å^3) while also still in the same order of magnitude of volume (10^5 Å^3). Cells preserved their original geometry so only integer multipliers were applied to the original lattice lengths to get cubic-resembling cells. These supercells were used for both atomic and periodic descriptor calculation. The structures were obtained directly from the CoRE database.[43]

3.1.2.1. Atomic Descriptors. On the basis of the work by Rosen et al. and using information on the QMOF database, a set of 165 atomic descriptors was calculated.[22] Within the atomic set, two subsets of atomic descriptor calculation methods were chosen: 45 descriptors were based on the work by Meredig et al.[50] and 120 descriptors based on the work by He et al.[17] These descriptors describe the composition of MOFs in terms of the properties of the atoms in the cell, such as average molecular weight, electronegativity, and atom number. These descriptors were chosen to include the cell characteristics and the metal atoms' presence in the cell, as well as to summarize the composition of the MOF. These descriptors were calculated for the set of normalized structures using Python with the featurization included in the Atomic Simulation Environment (ASE 3.21.1[51]) building on the available code by Rosen et al.[22] The complete list of descriptors including atomic descriptors before descriptor selection is part of the Supporting Information in Table S2.

Meredig et al. developed an element composition-based materials screening method for inorganic solid formation energy predictions based on coupling physically motivated heuristics and a machine learning model to a quantum mechanics calculations database, including DFT calculations data-fed models.[50,52] The featurization includes 103 percentage elemental composition descriptors and 17 statistical measures of elemental descriptors within the crystal structure including atomic weight, group number, electronegativity, and fraction of valence electrons per type. This set of descriptors was calculated using the Python libraries Matminer v0.5.9 and Pymatgen 2020.3.13.

On the other hand, He et al. developed a screening method to identify metallic MOF structures at the level of semilocal DFT band theory combining machine learning techniques, statistical multivoting, and *ab initio* calculations performed in VASP.[17] The model includes composition-based descriptors used to characterize the metallic or nonmetallic character of inorganic solids in the Open Quantum Materials Database.[53,54] The model used MOF structures in the CoRE database.[42,43] There are 45 statistical descriptors of elemental properties used in this model. This includes nine elemental properties such as atomic number, period number, and electron affinity. Five statistical quantities were calculated on those elemental properties over the number of atoms in the chemical formula. This set of descriptors was calculated using the Python libraries ASE 3.21.1 and Scipy 1.3.1.

3.1.2.2. Periodic Descriptors. The PoreBlazer tool was used to calculate descriptors with periodic character. PoreBlazer uses the MOF structure from the file containing the atom coordinates, as well as the cell dimensions and the angles of the unit cell.[55] The force field used is the universal force field (UFF),[56] a force field with parameters for the full range of elements in the periodic table including MOF transition metals and actinoids. The simulation conditions are standard temperature (298 K) and cutoff distance of 12.8 Å and cubelet size of 0.2 Å with 500 trials per atom for the surface area calculation.

PoreBlazer v04 was compiled and run for the set of structures normalized. These descriptors describe the capability of MOFs to adsorb guest molecules and are defined according to the geometries of the MOF which determine the properties of the pores. The calculation of pore properties on the regularized cells is key since it normalizes the properties across the MOF spectrum. Additionally, only intensive properties were included by using specific properties by mass or volume unit. These descriptors are also included in Table S2. The full list of descriptors used for the model after descriptor selection is included in Table 1.

The challenge of this work was to apply similar methodologies, including the calculation of descriptors for MOFs as highly crystalline and by definition, periodic structures. This contrasts with other works with smaller molecules and descriptor values widely available in databases. Instead, descriptor calculation complexity in this work includes both

https://doi.org/10.1021/acs.iecr.2c00561
Ind. Eng. Chem. Res. 2022, 61, 5853−5862

the atomic and periodic character of MOF and therefore aims to accurately correlate these with the concept of thermal stability.

3.1.2.3. Descriptor Selection. The total number of descriptors prior to selection was 159, including nonzero atomic and periodic descriptors. To include nonlinear relationships between the descriptors and target property, a series of nine transformations were applied to the descriptor data (i.e., polynomial, trigonometric, logarithmic, exponential, radical, and rational). This added an extra layer of possible hidden correlations to the model at the cost of significantly increasing the total number of descriptors in the model.

The number of MOF descriptors in the data set was large compared to the data set size. Therefore, a reduction in the number of descriptors is needed, since the accuracy of the model in machine learning depends on the quality of the domain of descriptors data.[59,60] As depicted in Figure 1, descriptor screening and statistical significance determination are crucial to ensure that the model includes only descriptors highly correlated to the target property. It is standard to have 10−15% ratio of descriptors per total number of molecules in the data set. The process for descriptor screening is to significantly reduce the number of descriptors used in the model.[41] An initial screening excluded zero value columns, descriptors with low data variability, and descriptors providing similar information.

The first technique for descriptor reduction to avoid overfitting due to multicollinearity was reduction by correlation. We identified highly correlated independent variables according to the Pearson correlation coefficient matrix for effective feature reduction.[25,59] We assumed coefficients higher than 0.70 indicated high correlation, and thus we selected variables to discard based on subject matter expertise. Given the significantly large number of descriptors for this model, the original Pearson matrix is not shown but can be found in Supporting Information. The number of descriptors after both the initial screening and the correlation reduction was reduced by ∼93%. The second technique used was the direct feature versus target importance. This shows the level at which descriptors are contributing to the target property.[59] As with the correlation matrix, only the descriptors with the highest importance index were included. Based on the rule of thumb for data set to descriptor ratio, only 12 descriptors were selected.

3.1.3. Machine Learning Model Building. With the set of selected descriptors and target property set, the next step was to build the QSPR classification model. The model would allow us to predict with enough level of accuracy the cluster group of new MOF structures based on the initial data set clustering and the descriptor calculation for these. As part of the data preprocessing, an important step was the scalar standardization to transform data to dimensionless form and provide the model with the same distribution.[61] The method used the sample standard score, which center-scales the descriptor variables using the average value of each descriptor, resulting in zero mean descriptor data.[62] The models were built using the Python library Scikitlearn 0.24.2.

4. RESULTS AND DISCUSSION

4.1. Data Treatment. The distribution of the TGA data upon clustering is shown in Table 2. Based on the clustering, four thermal stability classes are established for the MOFs in the dataset. The distribution of the classes is close to homogeneous, which ensures a good class data input for the model. However, class I had a slightly less representation in the data set, which, although not critical for the model validity, is important to consider when analyzing the model results. The detailed results of the TGA data clustering is part of the Supporting Information. After descriptor calculations, selections are followed, and the feature importance of the remaining descriptors is calculated. An average of three runs using mean decrease impurity are carried out. These results are shown in Figure 3. The top 12 independent descriptors as seen in Figure 4 are selected for the final model.

Table 2. Distribution of TGA Data Clustering of MOF Structures

TGA stability class	TGA temperature range mean [°C]	percentage of MOFs in the class [%]
class I	569	12.5
class II	454	37.5
class III	335	31.0
class IV	275	19.0

4.2. Machine Learning. The results of the classification model are depicted in Figure 5. The model used a ratio of 85/15% training to testing sets. A classification report shows prediction statistics per class using the accuracy formula in eq 1. The precision column represents the model exactness per class calculated based on the successes of the model per class or percent of correct classification for the testing set. The recall is the measure of the model completeness; it describes the percentage of instances classified correctly. The F_1 score is the weighted harmonic mean of precision and recall.

$$\text{accuracy}(y, \hat{y}) = \frac{1}{n_{\text{samples}}} \sum_{i=0}^{n_{\text{samples}}-1} 1(\hat{y}_i = y_i) \quad (1)$$

Five QSPR models with different predictive machine-learning models were run using the same target and descriptor data. This produced the selected final classification model that, based on the intrinsic molecular structure of a MOF, will place such MOF within an acceptable thermal stability point range. Thus, this will give an initial assessment of the stability of the MOF itself. Values for accuracy and area under the receiver operating characteristic curve (ROC AUC) were computed and presented in Table 3. As evidenced, the *k*-nearest neighbor model showed the highest balanced precision, resulting in the best classification model for MOF thermal stability classes.

The model with the highest accuracy in classifying the thermal stability class for the test set results in an accuracy score of 0.86 after cross-validation (value of balanced accuracy taking into account the imbalance nature of the dataset). The resulting ROC AUC score is 0.82. For both measures, the closeness to perfect prediction for the test set shows the robustness of the thermal stability classification model. The parameters used for the model development are reported in Table 4. Smaller test set sizes result in higher accuracy results (the Supporting Information includes the hyperparameter tuning and results for several tests to train size ratio). However, given the general rules of thumb for test-to-train ratio, the higher test set size is reported to reflect higher applicability of the classification model.

https://doi.org/10.1021/acs.iecr.2c00561
Ind. Eng. Chem. Res. 2022, 61, 5853−5862

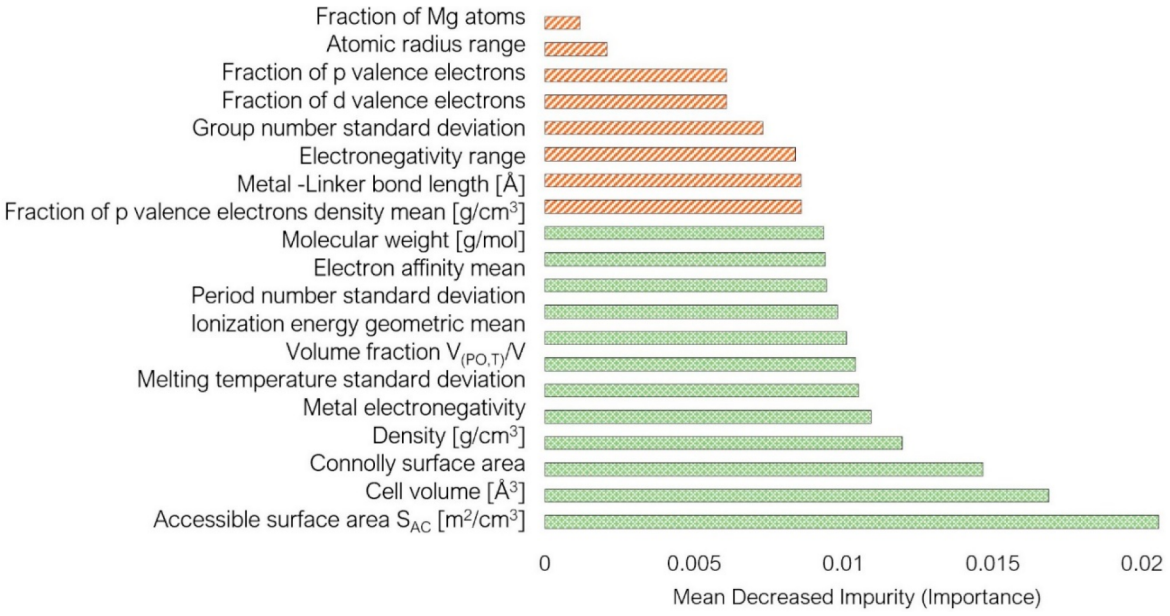


Figure 3. Descriptor importance for the target property.

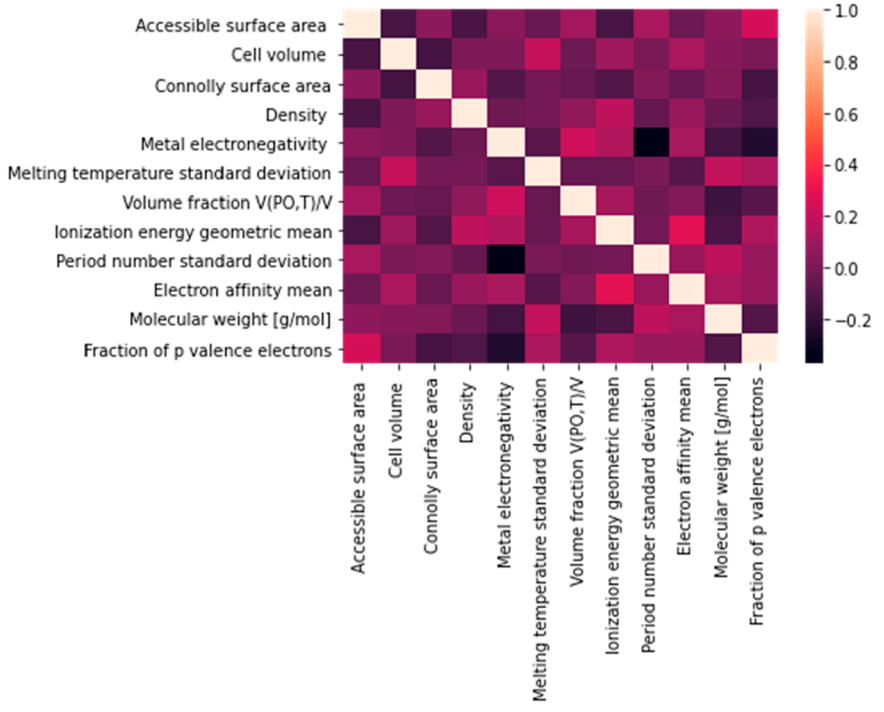


Figure 4. Descriptor correlation matrix.

5. CONCLUSIONS

This study presents an advanced predictive model for determining the thermal stability of MOFs by combining knowledge from computational chemistry and machine learning algorithms. The approach of machine learning in this study aligns with the state of the art in terms of materials research.[63] The *k*-nearest neighbor model has the highest prediction accuracy of 86.1%. Although this model considers different attempts to predict the MOF characteristics, it is important to recognize other factors that could potentially have a significant impact on the thermal stability of each MOF. For example, different synthesis conditions and techniques have resulted in different stability.[47] Therefore, some factors such as solvent type, solvent properties, pressure, temperature, and time-to-synthesize should be studied in the future. Additionally, some properties such as pK_a, which could be important in the prediction of thermal stability, have not been included in this model. Different theories point out the importance of the hard−soft acid−base (HSAB) concept for MOF thermal stability while other theories point out the significance of the radius and precise electronic structure of the metal on the bond strength. Altogether, characteristics of the electronic structure of the metal have an influence on the bond strength and therefore on the MOF thermal stability.

https://doi.org/10.1021/acs.iecr.2c00561
Ind. Eng. Chem. Res. 2022, 61, 5853−5862

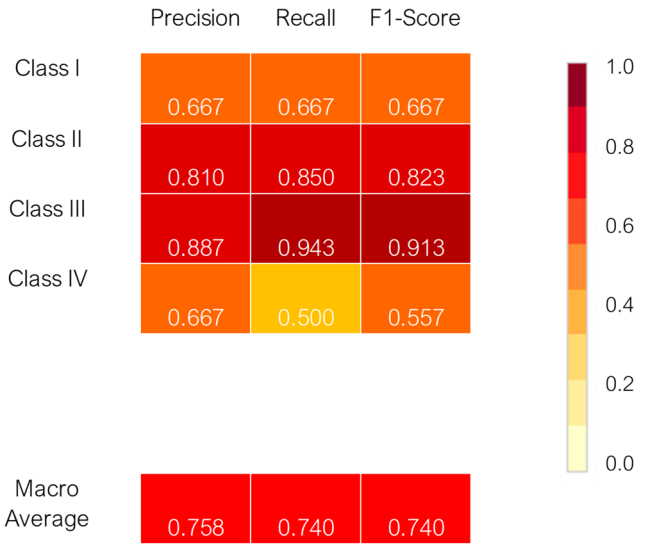


Figure 5. Classification report of the QSPR model by MOF class.

Table 3. Classification Results by Various Models

model	accuracy	ROC AUC score
k-nearest neighbors	0.861	0.820
support vector machine	0.665	0.553
random forest	0.610	0.735
quadratic discriminant analysis	0.583	0.681
multilayer perceptron	0.637	0.772

Table 4. Classification Model Parameters

parameter	value utilized
testing set size	15%
number of *k* neighbors	8
weight	uniform
algorithm	*k*-dimensional tree
leaf size	30
p-value	2
metric	Manhattan

On the basis of the important analysis using random forest classifier, the descriptors selected from the overall large pool of calculated descriptors are the ones that contribute more significantly to the MOF thermal stability. A next step for this work is to improve the model by integrating the use of graph neural networks to increase the capability of interpretation of the model. Additionally, by expanding the amount of experimental data, the sensitivity and robustness of the model will further increase.[44,45]

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.iecr.2c00561.

Table S1 listing TGA stability points from experimental data extracted using WebPlotDigitizer; Figure S1 showing TGA stability points clustered in different classes; Figure S2 showing application of the elbow method to determine the number of *k* clusters based on the distance between data points and their assigned clusters; Table S2 of complete list of all calculated descriptors including atomic and periodic descriptors (PDF)

Pearson correlation matrix of raw descriptor data prior to descriptor selection (XLSX)

Hyperparameter tuning and results reporting (XLSX)

■ AUTHOR INFORMATION

Corresponding Author

Qingsheng Wang − *Artie McFerrin Department of Chemical Engineering, Texas A&M University, College Station, Texas 77843-3122, United States;* orcid.org/0000-0002-6411-984X; Email: qwang@tamu.edu

Authors

Harold U. Escobar-Hernandez − *Artie McFerrin Department of Chemical Engineering, Texas A&M University, College Station, Texas 77843-3122, United States;* orcid.org/0000-0002-4289-1214

Lisa M. Pérez − *Division of Research, High Performance Research Computing, Texas A&M University, College Station, Texas 77843-3361, United States;* orcid.org/0000-0003-1176-1027

Pingfan Hu − *Artie McFerrin Department of Chemical Engineering, Texas A&M University, College Station, Texas 77843-3122, United States*

Fernando A. Soto − *Energy Engineering, Penn State Greater Allegheny, McKeesport, Pennsylvania 15132, United States*

Maria I. Papadaki − *Department of Environmental & Natural Resources Management, University of Patras, Agrinio GR30100, Greece*

Hong-Cai Zhou − *Department of Chemistry, Texas A&M University, College Station, Texas 77843-3255, United States;* orcid.org/0000-0002-9029-3788

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.iecr.2c00561

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work receivedi financial support from the Mary Kay O'Connor Process Safety Center at Texas A&M University. Portions of this research were conducted with the advanced computing resources and consultation provided by Texas A&M High Performance Research Computing.

■ REFERENCES

(1) Yaghi, O. M.; Kalmutzki, M. J.; Diercks, C. S. *Introduction to Reticular Chemistry: Metal-Organic Frameworks and Covalent Organic Frameworks*; Wiley, 2019.

(2) Zhou, H.-C.; Long, J. R.; Yaghi, O. M. Introduction to Metal-Organic Frameworks. *Chem. Rev.* **2012**, *112* (2), 673−674.

(3) García, H.; Navalón, S. *Metal-Organic Frameworks: Applications in Separations and Catalysis*; Wiley, 2018.

(4) Fu, J.; Tian, Y.; Wu, J. Seeking metal-organic frameworks for methane storage in natural gas vehicles. *Adsorption* **2015**, *21* (6), 499−507.

(5) Safaei, M.; Foroughi, M. M.; Ebrahimpoor, N.; Jahani, S.; Omidi, A.; Khatami, M. A review on metal-organic frameworks: Synthesis and applications. *Trac-Trend. Anal. Chem.* **2019**, *118*, 401−425.

(6) Czaja, A. U.; Trukhan, N.; Müller, U. Industrial applications of metal-organic frameworks. *Chem. Soc. Rev.* **2009**, *38* (5), 1284−1293.

(7) Li, H.; Eddaoudi, M.; O'Keeffe, M.; Yaghi, O. M. Design and synthesis of an exceptionally stable and highly porous metal-organic framework. *Nature* **1999**, *402* (6759), 276−279.
(8) Devic, T.; Serre, C. High valence 3p and transition metal based MOFs. *Chem. Soc. Rev.* **2014**, *43* (16), 6097−6115.
(9) News Sina, An explosion occurred in the Institute of Chemistry of the Chinese Academy of Sciences News Sina (China), April 1, 2021. https://news.sina.cn/gn/2021-04-01/detail-ikmxzfmk0500184.d.html?from=wap (accessed May 1, 2021).
(10) Friščić, T.; Julien, P. A.; Mottillo, C. Environmentally-Friendly Designs and Syntheses of Metal-Organic Frameworks (MOFs). *ACS Symp. Ser.* **2014**, *1186*, 161−183.
(11) Julien, P. A.; Mottillo, C.; Friščić, T. Metal-organic frameworks meet scalable and sustainable synthesis. *Green Chem.* **2017**, *19* (12), 2729−2747.
(12) McDonald, T. M.; Mason, J. A.; Kong, X.; Bloch, E. D.; Gygi, D.; Dani, A.; Crocellà, V.; Giordanino, F.; Odoh, S. O.; Drisdell, W. S.; Vlaisavljevich, B.; Dzubak, A. L.; Poloni, R.; Schnell, S. K.; Planas, N.; Lee, K.; Pascal, T.; Wan, L. F.; Prendergast, D.; Neaton, J. B.; Smit, B.; Kortright, J. B.; Gagliardi, L.; Bordiga, S.; Reimer, J. A.; Long, J. R. Cooperative insertion of CO_2 in diamine-appended metal-organic frameworks. *Nature* **2015**, *519* (7543), 303−308.
(13) Tsivion, E.; Long, J. R.; Head-Gordon, M. Hydrogen Physisorption on Metal-Organic Framework Linkers and Metalated Linkers: A Computational Study of the Factors That Control Binding Strength. *J. Am. Chem. Soc.* **2014**, *136* (51), 17827−17835.
(14) Dzubak, A. L.; Lin, L.-C.; Kim, J.; Swisher, J. A.; Poloni, R.; Maximoff, S. N.; Smit, B.; Gagliardi, L. *Ab initio* carbon capture in open-site metal-organic frameworks. *Nat. Chem.* **2012**, *4* (10), 810−816.
(15) Fairen-Jimenez, D.; Moggach, S. A.; Wharmby, M. T.; Wright, P. A.; Parsons, S.; Düren, T. Opening the Gate: Framework Flexibility in ZIF-8 Explored by Experiments and Simulations. *J. Am. Chem. Soc.* **2011**, *133* (23), 8900−8902.
(16) Boyd, P. G.; Moosavi, S. M.; Witman, M.; Smit, B. Force-Field Prediction of Materials Properties in Metal-Organic Frameworks. *J. Phys. Chem. Lett.* **2017**, *8* (2), 357−363.
(17) He, Y.; Cubuk, E. D.; Allendorf, M. D.; Reed, E. J. Metallic Metal-Organic Frameworks Predicted by the Combination of Machine Learning Methods and *Ab Initio* Calculations. *J. Phys. Chem. Lett.* **2018**, *9* (16), 4562−4569.
(18) Dureckova, H.; Krykunov, M.; Aghaji, M. Z.; Woo, T. K. Robust Machine Learning Models for Predicting High CO_2 Working Capacity and CO_2/H_2 Selectivity of Gas Adsorption in Metal Organic Frameworks for Precombustion Carbon Capture. *J. Phys. Chem. C* **2019**, *123* (7), 4133−4139.
(19) Fernandez, M.; Woo, T. K.; Wilmer, C. E.; Snurr, R. Q. Large-Scale Quantitative Structure-Property Relationship (QSPR) Analysis of Methane Storage in Metal-Organic Frameworks. *J. Phys. Chem. C* **2013**, *117* (15), 7681−7689.
(20) Mukherjee, K.; Colón, Y. J. Machine learning and descriptor selection for the computational discovery of metal-organic frameworks. *Mol. Simul.* **2021**, *47* (10−11), 857−877.
(21) Rosen, A. S.; Fung, V.; Huck, C.; O'Donnell, C. T.; Horton, M. K.; Truhlar, D. G.; Persson, K. A.; Notestein, J. M.; Snurr, R. Q. High-Throughput Predictions of Metal-Organic Framework Electronic Properties: Theoretical Challenges, Graph Neural Networks, and Data Exploration. *ChemRxiv* **2021**, DOI: 10.26434/chemrxiv-2021-6cs91.
(22) Rosen, A. S.; Iyer, S. M.; Ray, D.; Yao, Z.; Aspuru-Guzik, A.; Gagliardi, L.; Notestein, J. M.; Snurr, R. Q. Machine learning the quantum-chemic al properties of metal-organic frameworks for accelerated materials discovery. *Matter* **2021**, *4* (5), 1578−1597.
(23) Roy, K.; Kar, S.; Das, R. N. *A Primer on QSAR/QSPR Modeling: Fundamental Concepts*; Springer International Publishing, 2015.
(24) Escobar-Hernandez, H. U.; Shen, R.; Papadaki, M. I.; Powell, J. A.; Zhou, H.-C.; Wang, Q. Hazard Evaluation of Metal-Organic Framework Synthesis and Scale-up: A Laboratory Safety Perspective. *ACS Chem. Health Saf.* **2021**, *28* (5), 358−368.
(25) Yuan, S.; Feng, L.; Wang, K.; Pang, J.; Bosch, M.; Lollar, C.; Sun, Y.; Qin, J.; Yang, X.; Zhang, P.; Wang, Q.; Zou, L.; Zhang, Y.; Zhang, L.; Fang, Y.; Li, J.; Zhou, H.-C. Stable Metal-Organic Frameworks: Design, Synthesis, and Applications. *Adv. Mater.* **2018**, *30* (37), 1704303.
(26) Yuan, S.; Zou, L.; Qin, J.-S.; Li, J.; Huang, L.; Feng, L.; Wang, X.; Bosch, M.; Alsalme, A.; Cagin, T.; Zhou, H.-C. Construction of hierarchically porous metal-organic frameworks through linker labilization. *Nat. Commun.* **2017**, *8* (1), 15356.
(27) Lv, X.-L.; Yuan, S.; Xie, L.-H.; Darke, H. F.; Chen, Y.; He, T.; Dong, C.; Wang, B.; Zhang, Y.-Z.; Li, J.-R.; Zhou, H.-C. Ligand Rigidification for Enhancing the Stability of Metal-Organic Frameworks. *J. Am. Chem. Soc.* **2019**, *141* (26), 10283−10293.
(28) Wang, S.-L.; Hu, F.-L.; Zhou, J.-Y.; Zhou, Y.; Huang, Q.; Lang, J.-P. Rigidity versus Flexibility of Ligands in the Assembly of Entangled Coordination Polymers Based on Bi- and Tetra Carboxylates and N-Donor Ligands. *Cryst. Growth Des.* **2015**, *15* (8), 4087−4097.
(29) Cavka, J. H.; Jakobsen, S.; Olsbye, U.; Guillou, N.; Lamberti, C.; Bordiga, S.; Lillerud, K. P. A New Zirconium Inorganic Building Brick Forming Metal Organic Frameworks with Exceptional Stability. *J. Am. Chem. Soc.* **2008**, *130* (42), 13850−13851.
(30) Dincă, M.; Dailly, A.; Liu, Y.; Brown, C. M.; Neumann, D. A.; Long, J. R. Hydrogen Storage in a Microporous Metal-Organic Framework with Exposed Mn^{2+} Coordination Sites. *J. Am. Chem. Soc.* **2006**, *128* (51), 16876−16883.
(31) Chui, S. S.-Y.; Lo, S. M.-F.; Charmant, J. P. H.; Orpen, A. G.; Williams, I. D. A Chemically Functionalizable Nanoporous Material $[Cu_3(TMA)_2(H_2O)_3]_n$. *Science* **1999**, *283* (5405), 1148−1150.
(32) Valdes, O. R.; Moreno, V. C.; Waldram, S.; Véchot, L.; Mannan, M. S. Runaway decomposition of dicumyl peroxide by open cell adiabatic testing at different initial conditions. *Process Saf. Environ. Prot.* **2016**, *102*, 251−262.
(33) Dirion, J.-L.; Reverte, C.; Cabassud, M. Kinetic parameter estimation from TGA: Optimal design of TGA experiments. *Chem. Eng. Res. Des.* **2008**, *86* (6), 618−625.
(34) Gadipelli, S.; Guo, Z. Postsynthesis Annealing of MOF-5 Remarkably Enhances the Framework Structural Stability and CO_2 Uptake. *Chem. Mater.* **2014**, *26* (22), 6333−6338.
(35) Quintero, F. A.; Patel, S. J.; Muñoz, F.; Mannan, M. S. Review of Existing QSAR/QSPR Models Developed for Properties Used in Hazardous Chemicals Classification System. *Ind. Eng. Chem. Res.* **2012**, *51* (49), 16101−16115.
(36) Wang, B.; Zhou, L.; Xu, K.; Wang, Q. Prediction of Minimum Ignition Energy from Molecular Structure Using Quantitative Structure-Property Relationship (QSPR) Models. *Ind. Eng. Chem. Res.* **2017**, *56* (1), 47−51.
(37) Park, S.; Bailey, J. P.; Pasman, H. J.; Wang, Q.; El-Halwagi, M. M. Fast, easy-to-use, machine learning-developed models of prediction of flash point, heat of combustion, and lower and upper flammability limits for inherently safer design. *Comput. Chem. Eng.* **2021**, *155*, 107524.
(38) Jiao, Z.; Escobar-Hernandez, H. U.; Parker, T.; Wang, Q. Review of recent developments of quantitative structure-property relationship models on fire and explosion-related properties. *Process Saf. Environ. Prot.* **2019**, *129*, 280−290.
(39) Jiao, Z.; Hu, P.; Xu, H.; Wang, Q. Machine Learning and Deep Learning in Chemical Health and Safety: A Systematic Review of Techniques and Applications. *ACS Chem. Health Saf.* **2020**, *27* (6), 316−334.
(40) Groom, C. R.; Bruno, I. J.; Lightfoot, M. P.; Ward, S. C. The Cambridge structural database. *Acta Crystallogr. B* **2016**, *72* (2), 171−179.
(41) Chen, T.; Manz, T. A. Identifying misbonded atoms in the **2019** CoRE metal-organic framework database. *RSC Adv.* **2020**, *10* (45), 26944−26951.
(42) Chung, Y. G.; Camp, J.; Haranczyk, M.; Sikora, B. J.; Bury, W.; Krungleviciute, V.; Yildirim, T.; Farha, O. K.; Sholl, D. S.; Snurr, R. Q. Computation-Ready, Experimental Metal-Organic Frameworks: A

https://doi.org/10.1021/acs.iecr.2c00561
Ind. Eng. Chem. Res. 2022, 61, 5853−5862

Tool To Enable High-Throughput Screening of Nanoporous Crystals. *Chem. Mater.* **2014**, *26* (21), 6185−6192.
(43) Chung, Y. G.; Haldoupis, E.; Bucior, B. J.; Haranczyk, M.; Lee, S.; Zhang, H.; Vogiatzis, K. D.; Milisavljevic, M.; Ling, S.; Camp, J. S.; Slater, B.; Siepmann, J. I.; Sholl, D. S.; Snurr, R. Q. Advances, Updates, and Analytics for the Computation-Ready, Experimental Metal-Organic Framework Database: CoRE MOF 2019. *J. Chem. Eng. Data* **2019**, *64* (12), 5985−5998.
(44) Wang, Q. Theoretical and Experimental Evaluation of Chemical Reactivity. Ph.D. Dissertation, Texas A&M University, College Station, Texas, USA, 2010.
(45) Narayan, T. C.; Miyakai, T.; Seki, S.; Dincă, M. High Charge Mobility in a Tetrathiafulvalene-Based Microporous Metal-Organic Framework. *J. Am. Chem. Soc.* **2012**, *134* (31), 12932−12935.
(46) Farha, O. K.; Özgür Yazaydın, A.; Eryazici, I.; Malliakas, C. D.; Hauser, B. G.; Kanatzidis, M. G.; Nguyen, S. T.; Snurr, R. Q.; Hupp, J. T. De novo synthesis of a metal-organic framework material featuring ultrahigh surface area and gas storage capacities. *Nat. Chem.* **2010**, *2* (11), 944−948.
(47) Healy, C.; Patil, K. M.; Wilson, B. H.; Hermanspahn, L.; Harvey-Reid, N. C.; Howard, B. I.; Kleinjan, C.; Kolien, J.; Payet, F.; Telfer, S. G.; Kruger, P. E.; Bennett, T. D. The thermal stability of metal-organic frameworks. *Coordin. Chem. Rev.* **2020**, *419*, 213388.
(48) Marin, F.; Rohatgi, A.; Charlot, S. WebPlotDigitizer, a polyvalent and free software to extract spectra from old astronomical publications: application to ultraviolet spectropolarimetry. *arXiv* **2017**, 1708.02025.
(49) BIOVIA Dassault Systèmes Materials Studio 2018, 2018; BIOVIA Dassault Systèmes: San Diego, CA, USA, 2018. https://www.3ds.com/products-services/biovia/products/molecular-modeling-simulation/biovia-materials-studio (accessed Mar 1, 2020).
(50) Meredig, B.; Agrawal, A.; Kirklin, S.; Saal, J. E.; Doak, J. W.; Thompson, A.; Zhang, K.; Choudhary, A.; Wolverton, C. Combinatorial screening for new materials in unconstrained composition space with machine learning. *Phys. Rev. B* **2014**, *89* (9), 094104.
(51) Hjorth Larsen, A.; Jørgen Mortensen, J.; Blomqvist, J.; Castelli, I. E.; Christensen, R.; Dułak, M.; Friis, J.; Groves, M. N.; Hammer, B.; Hargus, C.; Hermes, E. D.; Jennings, P. C.; Bjerre Jensen, P.; Kermode, J.; Kitchin, J. R.; Leonhard Kolsbjerg, E.; Kubal, J.; Kaasbjerg, K.; Lysgaard, S.; Bergmann Maronsson, J.; Maxson, T.; Olsen, T.; Pastewka, L.; Peterson, A.; Rostgaard, C.; Schiøtz, J.; Schütt, O.; Strange, M.; Thygesen, K. S.; Vegge, T.; Vilhelmsen, L.; Walter, M.; Zeng, Z.; Jacobsen, K. W. The atomic simulation environment—a Python library for working with atoms. *J. Phys-Condens. Mater.* **2017**, *29* (27), 273002.
(52) Zagorac, D.; Müller, H.; Ruehl, S.; Zagorac, J.; Rehme, S. Recent developments in the Inorganic Crystal Structure Database: theoretical crystal structure data and related features. *J. Appl. Crystallogr.* **2019**, *52* (5), 918−925.
(53) Kirklin, S.; Saal, J. E.; Meredig, B.; Thompson, A.; Doak, J. W.; Aykol, M.; Rühl, S.; Wolverton, C. The Open Quantum Materials Database (OQMD): assessing the accuracy of DFT formation energies. *npj Comput. Mater.* **2015**, *1* (1), 15010.
(54) Saal, J. E.; Kirklin, S.; Aykol, M.; Meredig, B.; Wolverton, C. Materials Design and Discovery with High-Throughput Density Functional Theory: The Open Quantum Materials Database (OQMD). *JOM* **2013**, *65* (11), 1501−1509.
(55) Sarkisov, L.; Bueno-Perez, R.; Sutharson, M.; Fairen-Jimenez, D. Materials Informatics with PoreBlazer v4.0 and the CSD MOF Database. *Chem. Mater.* **2020**, *32* (23), 9849−9867.
(56) Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *J. Am. Chem. Soc.* **1992**, *114* (25), 10024−10035.
(57) Connolly, M. L. Solvent-Accessible Surfaces of Proteins and Nucleic Acids. *Science* **1983**, *221* (4612), 709−713.
(58) Connolly, M. L. Analytical molecular surface calculation. *J. Appl. Crystallogr.* **1983**, *16* (5), 548−558.
(59) Jiao, Z.; Sun, Y.; Hong, Y.; Parker, T.; Hu, P.; Mannan, M. S.; Wang, Q. Development of Flammable Dispersion Quantitative Property-Consequence Relationship Models Using Extreme Gradient Boosting. *Ind. Eng. Chem. Res.* **2020**, *59* (33), 15109−15118.
(60) Hu, P.; Jiao, Z.; Zhang, Z.; Wang, Q. Development of Solubility Prediction Models with Ensemble Learning. *Ind. Eng. Chem. Res.* **2021**, *60* (30), 11627−11635.
(61) Shanker, M.; Hu, M. Y.; Hung, M. S. Effect of data standardization on neural network training. *Omega-Int. J. Manage S Eng.* **1996**, *24* (4), 385−397.
(62) Kuhn, M.; Johnson, K. *Applied Predictive Modeling*; Springer, 2013; Vol. *26*.
(63) Keith, J. A.; Vassilev-Galindo, V.; Cheng, B.; Chmiela, S.; Gastegger, M.; Müller, K.-R.; Tkatchenko, A. Combining Machine Learning and Computational Chemistry for Predictive Insights Into Chemical Systems. *Chem. Rev.* **2021**, *121* (16), 9816−9872.

https://doi.org/10.1021/acs.iecr.2c00561
Ind. Eng. Chem. Res. 2022, 61, 5853−5862