JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

pubs.acs.org/jcim Article

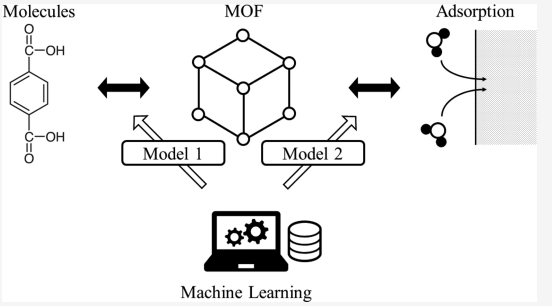
Correlation between the Metal and Organic Components, Structure Property, and Gas-Adsorption Capacity of Metal−Organic Frameworks

Shunsuke Yuyama and Hiromasa Kaneko*

Cite This: J. Chem. Inf. Model. 2021, 61, 5785−5792 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Metal−organic frameworks (MOFs) are materials in which metals and organic compounds form crystalline and porous structures. Previous studies have investigated the relationships between the structure properties and physical properties of MOFs through molecular simulations, but the overall relationships in MOFs, including the relationships between the metals and organic components and the experimentally measured physical properties, have not been clarified. In this study, we developed two regression models between three elements in MOFs: the components, structure properties, and gas-adsorption capacities as physical properties. Using a nonlinear regression analysis method, we succeeded in predicting the structure properties from the components and the physical properties from the structure properties.



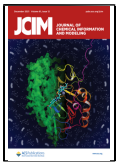
1. INTRODUCTION

Metal−organic frameworks (MOFs) are crystalline materials composed of metal ions and organic ligands, and they have attracted much attention in recent years. Kitagawa et al.[1] summarized the properties of MOFs as (1) the large surface area and porosity, (2) free and highly designable, and (3) soft crystalline materials. Yaghi et al.[2,3] formed multinuclear metal clusters with multidentate ligands around a metal ion. This cluster is called a secondary building unit (SBU)[4] and can form a stronger structure than a metal ion alone. The initially synthesized MOF-5 was found to combine extraordinary architectural stability, porosity, and an ultrahigh surface area.[3] This stability is largely attributed to $Zn_4O(COO)_6$, which is the SBU of MOF-5. The combination of the SBU, which has a well-defined shape, and an organic linker is found to play an important role in the MOF structure.[5,6] The function of the SBU is to form an open metal site (OMS), which has empty orbitals and induces Lewis acid−base reactions, highly active catalysis, and interactions with guest molecules.[7−9]

MOFs have attracted attention as gas adsorbents, gas separators,[10] and catalysts owing to the properties of the SBU and OMS. In gas adsorption, various interactions occur between the adsorbed guest molecules and MOFs, such as hydrogen bonding and $\pi-\pi$ interactions. In particular, the OMS has been found to exhibit strong affinity for certain molecules through Lewis acid−base interactions.[11,12] In addition, MOFs are ideal homogeneous catalysts because they have both high selectivity as molecular catalysts and high reactivity recycling as heterogeneous catalysts.[13] For example, it has been reported that Ti-based MOFs can combine photocatalytic activity and the light-absorbing properties of the organic linkers.[14] In addition, a Pd catalyst covered with HKUST-1[15] or Ni-MOF-74[16] can induce electron transfer, leading to active and robust catalysts. Other applications include modifying MOF surfaces[17] to form network structures for photon excitation[18] and constructing MOF films with complex structures by controlling crystal growth.[19]

Many studies and technological developments have been performed and made to reproduce MOFs using computers.[20] Calculations using ab initio methods and classical force fields,[21,22] which describe the structures of MOFs and the potential energies of guest molecules, are highly accurate but require significant computational resources.[23] Therefore, Monte Carlo (MC) methods and molecular dynamics simulations have been used to describe MOFs.[24−28] Grand canonical MC simulations enable high-throughput computational screening of MOFs on a large scale. Although these simulations allow the entire large set of material candidates to be explored, calculating more than a few thousand structures can become computationally unfeasible because of the need to deal with complex guest molecules, intramolecular polarization, and interactions with aqueous media.[29]

Received: October 2, 2021
Published: December 13, 2021



https://doi.org/10.1021/acs.jcim.1c01205
J. Chem. Inf. Model. 2021, 61, 5785−5792

To overcome the above problems, machine-learning approaches have been introduced. In regression analysis, the model $y = f(\boldsymbol{X})$ is constructed with an objective variable y, such as a physical property or activity, and explanatory variables $\boldsymbol{X}$, such as molecular descriptors and experimental conditions. The correlation between the molecular structure and physical properties is called the quantitative structure−property relationship.[30] y can be generated by experiments, which are expensive and time-consuming, or molecular simulations, which are computationally expensive. Conversely, $\boldsymbol{X}$ can be calculated from the molecular structure and the experimental conditions can be set, which means that $\boldsymbol{X}$ can be generated at a low cost. Thornton et al.[31] used structural descriptors to perform high-throughput screening calculations of nanoporous materials, including MOFs. Pardakhti et al.[32] combined structural and chemical descriptors to construct regression models. Fernandez et al.[33] developed new descriptors weighted by the atomic properties. Therefore, it is possible to perform large-scale computational screening at a low cost by applying machine learning to MOFs. In addition, the studies of quantitative structure−property relationships were conducted to efficiently design MOFs.[20,34−37]

After the target structure property of a MOF is designed, the MOF is then synthesized. The actual structure is determined by X-ray diffraction or thermogravimetric analysis, and the physical properties, such as the gas-adsorption capacity, are measured. Although many simulations and machine-learning studies of the structure properties and physical properties of MOFs have been performed, it is important to obtain the components of the MOFs, that is, the metals and organic compounds, to synthesize the target structure properties of MOFs and determine the target values of the physical properties. Mathematical models are required for the relationships between the components, structure properties, and physical properties of MOFs.

The objective of this study is to construct models between the components of MOFs (the SBU and organic linker) and the structure properties, and between the structure properties and the physical properties. Model 1 was constructed between the components and the structure properties to predict the structure properties from the components. Model 2 was constructed between the structure properties and the physical properties to predict the physical properties from the structure properties. To construct these models, we used three types of data of MOFs: the molecular information about the components, the structure properties, and the physical properties. For the components, each MOF was decomposed into metals and organic compounds, and molecular descriptors were calculated. The structure properties were calculated from the structure properties optimized by molecular simulation. We focused on the gas-adsorption amount as a physical property. There are two types of gas-adsorption amounts: the gas-adsorption amount measured in actual experiments and the gas-adsorption amount calculated by molecular simulation. The effectiveness of the proposed method was tested using a dataset from a public database and an experimental dataset collected from published studies.

2. METHOD

In this study, we constructed two models: one between the components that comprise the MOF and the crystal properties, and another between the crystal properties and the physical properties.

2.1. Components.

The components that comprise MOFs are classified as SBUs, which are mainly metallic elements, and organic compounds, which are connected to the SBUs. Because the properties of SBUs and organic compounds greatly differ, it is necessary to identify and classify each component. In the field of chemoinformatics, simplified molecular input line entry system (SMILES),[38] fingerprints, and chemical table files (MOL files)[39] are mainly used to represent molecules. We used MOFid,[40] in which SMILES is used to represent inorganic and organic molecules, to accurately represent the MOFs.

Next, we calculated the molecular descriptors for each molecule. The molecular descriptors summarize the structural features of molecules, which can be calculated from SMILES. After obtaining SMILES of the components from MOFid, we classified them into inorganic and organic molecules. The descriptors, including the molecular weight, maximum partial charge, and log P, were then calculated for each molecule using RDKit.[41,42] For MOFs containing multiple organic molecules and inorganic molecules, the average of the descriptors was used. After deleting the variables whose values were zero for all of the samples, the total number of descriptors that could be calculated was 367, including 200 for the inorganic molecules and 167 for the organic molecules. For the central metal of the SBU, 27 descriptors, such as the atomic weight, melting point, and electron orbitals, could be obtained using Pymatgen.[43,44] After the descriptor of the atomic orbitals, which includes string data, was changed to dummy variables, 44 variables were obtained. The total number of explanatory variables was 411.

2.2. Structure Properties.

Because most MOFs are porous materials, the structure properties, including the internal pores, porosity, and density, are required to represent MOFs. We used the Zeo++ application[45−47] to calculate the descriptors of the structure properties through simulation. The eight properties given in Table 1, which are considered to be important to represent structure properties, were used in this study.

Table 1. Variables of the Structural Characteristics

name	meaning
LCD	largest cavity diameter
PLD	pore-limiting diameter
LFPD	largest free pore diameter
density	density
ASA	accessible surface area
NASA	nonaccessible surface area
AV VF	accessible volume void fraction
AV	accessible volume density

2.3. Physical Properties.

Although MOFs have a variety of physical properties, we focused on the gas-adsorption capacity in this study. For the gas-adsorption capacity, we collected previously reported experimental data (Table 2) and the results of molecular simulations from MOF database.[40,48] Because the collected adsorption dataset contained missing values, we used the Gaussian mixture model to interpolate the missing values.[49] Although the gas-adsorption capacity was predicted at a certain temperature and pressure in previous studies, our dataset included a variety of temperatures and pressures to predict the gas-adsorption capacities for wide ranges of temperature and pressure, which led to the representation of the "breathing behavior" of MOFs.

https://doi.org/10.1021/acs.jcim.1c01205
J. Chem. Inf. Model. 2021, 61, 5785−5792

Table 2. CO_2 Dataset with the MOF Refcode, Name, and Reference

refcode	name	ref.	refcode	name	ref.
CIXZEP	Tb16(TATB)16(DMA)24	50	MAVRAE	NJU-Bai3	61
	Co(BDP)	51	SAHYIK	MOF-5	52
CUSXOE	MOF-200	52			62
EDUSIF	IRMOF-1	53		BeBTB	63
EDUSUR	IRMOF-3	53	SUKXUS01	MOF-205	52
EPISOM	NOTT-140	54	TODYUJ	CPO-27-Mg	59
ERIRIG	MOF-177	53	UXISAW	Cu3(TPBTM)	64
		55	VAMYOA	DUT-117Pd	65
		52	VAMZAN	DUT-117Ni	65
FIJDOS	MOF-74	53	VAMZER	DUT-117Cu	65
FIQCEN	HKUST-1,MOF-199,CuBTC	53	VIXPUP	HNUST-3	66
GECXUH	MOF-2	53	VIXPUP01	NOTT-125	67
GEYFIB	HNUST-1	56	VUJBIM	PCN-61	64
HABRAF	PCN-68	57			57
HAFGON	MFM-136	58	VUJBOS	PCN-66	57
LASYOU	MOF-505	53	XOZXOA	ZnO(FMA)	68
LECQEQ	CPO-27-Ni	59	XUDQEU	DUT-95	65
LEHXUT	CuBTB	60			

2.4. Modeling. *2.4.1. Model 1: Relationship between the Components and the Structure Properties.* The aim of this model is to represent the relationship between the descriptors of the components, such as the SBU and organic linkers, and the structure properties. The model $y = f(\boldsymbol{X})$ was constructed between the components $\boldsymbol{X}$ and the structure property y. Because multiple y variables needed to be estimated, as shown in Table 1, a model was constructed for each y variable so that all of the structure properties could be predicted.

2.4.2. Model 2: Relationship between the Structure Properties and the Physical Properties. The aim of this model is to represent the relationship between the structure properties and the gas-adsorption capacity as a physical property. The model $y = f(\boldsymbol{X})$ was constructed between the structure property descriptors $\boldsymbol{X}$ and the gas-adsorption capacity $\boldsymbol{y}$. Because the gas-adsorption capacity included values at different pressures in our dataset, multiple y variables needed to be estimated. Therefore, a model was constructed for each pressure so that the change in the gas-adsorption capacity could be expressed when the pressure changed.

3. RESULTS AND DISCUSSION

3.1. Model 1. The CoRE-MOF dataset[69] containing MOFs extracted from the Cambridge Crystallographic Data Centre[70] was used as the dataset. The solvent was removed from the MOF files as preprocessing. We used 9318 samples. $\boldsymbol{X}$ contained a total of 411 structural descriptors calculated from the molecules of the components, and y was set to the eight variables shown in Table 1.

LightGBM,[71] which is a gradient boosting decision tree, was used as the regression method. The dataset was divided into training data and test data at a ratio of 8:2. The training data were used to construct the model, and the test data were used to verify its predictive ability. Optuna[72] was used to determine the hyperparameters.

The prediction results of the test data are shown in Figures 1 and S1. The horizontal axis represents the actual y value and the vertical axis represents the estimated y value. The coefficient of determination (r^2), the mean absolute error (MAE), and the root-mean-square error (RMSE) of the test data are given in Tables 3 and S1. The r^2 values showed that

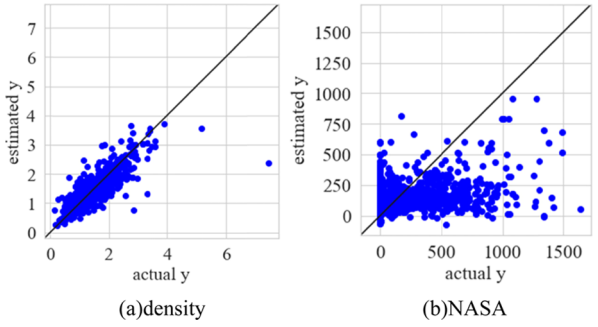


Figure 1. Plots of the actual y values against the estimated y values for structure property prediction with model 1.

Table 3. r^2, MAE, and RMSE for the Test Data with Model 1

name	r^2	MAE	RMSE
density	0.711	0.167	0.274
NASA	0.161	149.0	234.0

each model was able to represent more than 45% of the variance of each y, except for NASA. To clarify the difference between r^2 of the density and NASA, we determined the relationships between the descriptors in $\boldsymbol{X}$ and the density or NASA. The degree to which $\boldsymbol{X}$ contributes to y can be expressed as the variable importance, which can be calculated using the Gini impurity of random forests. The four variables in $\boldsymbol{X}$ with the highest importance for the density and NASA are shown in Table 4. For the density, SMR_VSA and VSA_Estate are descriptors related to the surface area of the molecule, which indicates that the surface area is important information for estimating the density. However, while NASA is structural information about the crystal surface area, the most important variable, BalabanJ, is a descriptor of the molecular graph and does not represent the surface area. Therefore, there were no structural descriptors in $\boldsymbol{X}$ that could explain NASA. Because the pore diameter and volume are very important for gas adsorption, the usefulness of this model was confirmed because the variables LCD and AV_VF were predicted.

3.2. Model 2. We used the amount of the gas adsorption of CO_2, N_2, and Ar as physical properties. For CO_2 adsorption,

https://doi.org/10.1021/acs.jcim.1c01205
J. Chem. Inf. Model. 2021, 61, 5785−5792

Table 4. Feature Importance Calculated with Random Forests

	variable name	importance
density	SMR_VSA7 of the organic linker	0.143
	VSA_Estate6 of the organic linker	0.071
	ExactMolwt of SBU	0.064
	Kappa3 of SBU	0.050
NASA	BalabanJ of organic linker	0.025
	PEOE_VSA3 of the organic linker	0.023
	MinAbsEStateIndex of the organic linker	0.017
	Kappa3 of SBU	0.016

we used the 50 samples collected from the research specified in Table 2, and for N_2 and Ar, we used 7121 and 7786 samples included in MOFDB.

As shown in model 1, LightGBM with hyperparameter selection by Optuna was used as the regression analysis method. The N_2 and Ar datasets were split into training and test data at a ratio of 8:2, with the training data used to construct the model and the test data used to verify its predictive ability. For the CO_2 dataset, owing to the small number of samples, leave-one-out cross-validation was used to validate the model.

The prediction results for CO_2 adsorption are shown in Figures 2 and 3. Figure 2 shows the plot of the actual y values

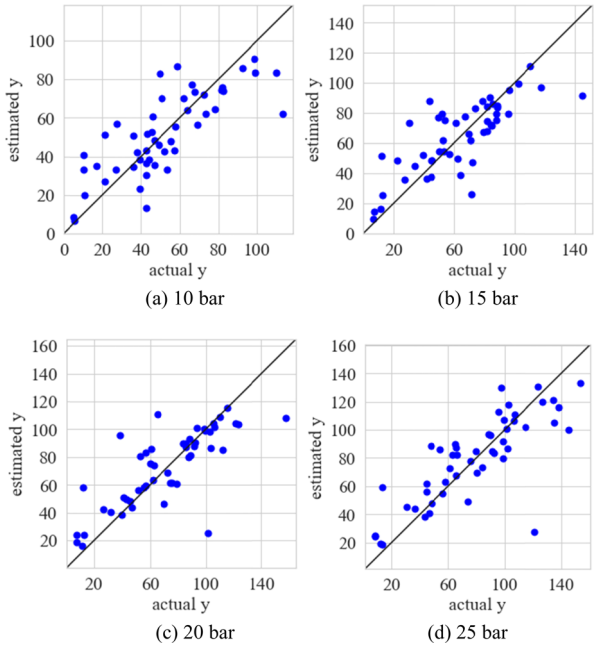


Figure 2. Plots of the actual y values against the estimated y values for CO_2 adsorption capacity at different pressures with model 2.

against the estimated y values. Figure 3 shows the actual y values against the estimated y values at each pressure for the MOF with refcode VAMZAN[60] as an example. It was confirmed that the model can accurately predict the y values even with a small dataset. In addition, the adsorption isotherms showed that the models can predict the y values at both low and high pressures. The MAE scores at each pressure are shown in Figure 4, and these indicated that the model showed highly predictive ability over the entire pressure range.

The prediction results for N_2 and Ar are shown in Figures 5 and 6, respectively, and examples of the adsorption isotherms

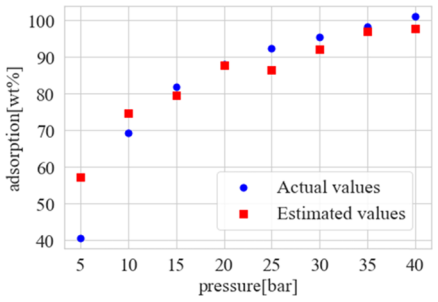


Figure 3. Plots of pressures vs gas-adsorption capacity for the CO_2 adsorption isotherm for VAMZAN with model 2.

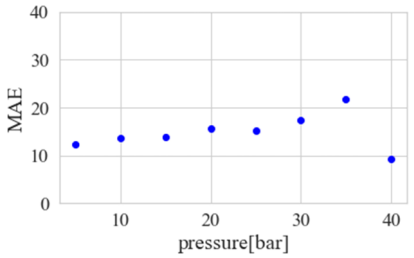


Figure 4. MAE scores for the predicted CO_2 adsorption capacity at each pressure with model 2.

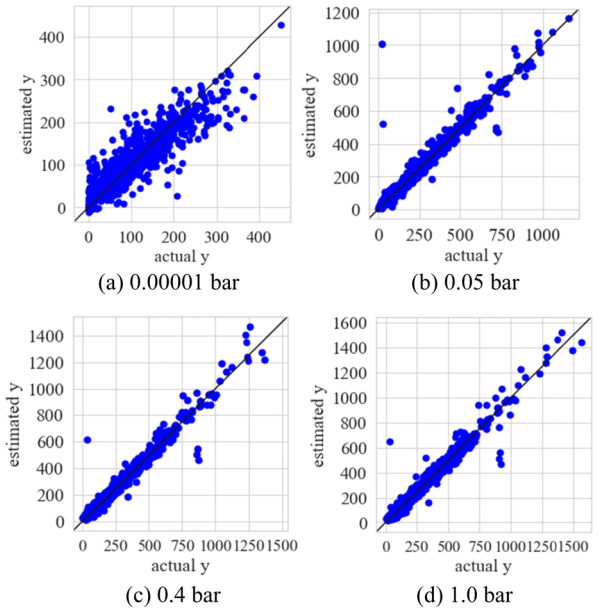


Figure 5. Plots of actual y values vs estimated y values at different pressures for N_2 adsorption capacity with model 2.

for the MOFs with refcode RUKHAI[73] and COXHON[74] are shown in Figures 7 and 8, respectively. From Figures 5 and 6, it can be seen that there were a few samples with large prediction errors, but overall, the y values were estimated with high accuracy. From Figures 7 and 8, gas adsorption was estimated with high accuracy for each gas in a wide pressure range. There were some outliers. Therefore, to further improve the prediction accuracy, it is necessary to increase the number of samples at a variety of pressures.

The MAE scores for N_2 and Ar are shown in Figure 9. The error was slightly larger in the low-pressure range, but the overall prediction was highly accurate. As shown in Figures 4 and 9, the results differed depending on the gas. The MAE of

https://doi.org/10.1021/acs.jcim.1c01205
J. Chem. Inf. Model. 2021, 61, 5785−5792

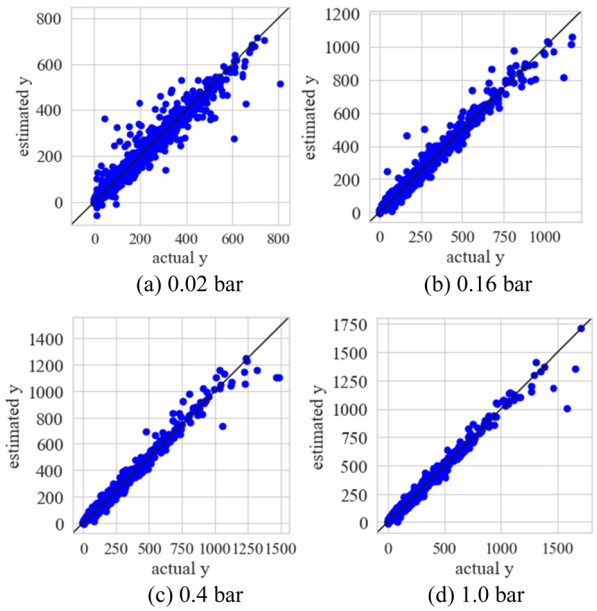


Figure 6. Plots of the actual *y* values against the estimated *y* values for Ar adsorption at different pressures with model 2.

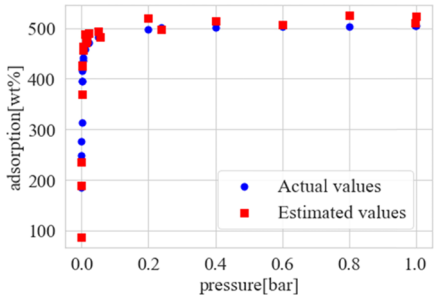


Figure 7. Plots of pressures vs gas-adsorption capacity for the N_2 adsorption isotherm for RUKHAI with model 2.

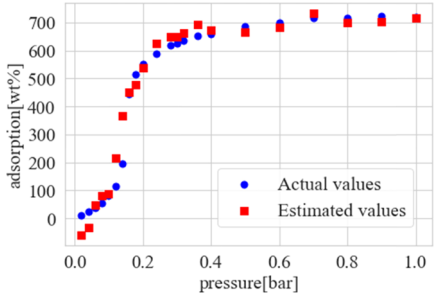


Figure 8. Plots of pressures vs gas-adsorption capacity for the Ar adsorption isotherm for COXHON with model 2.

CO_2 became larger as the pressure increased, but the MAE values of the other two gases were larger in the low-pressure region. One of the possible reasons for this is the interaction induced by the OMS. As mentioned above, it is known that there are electrostatic interactions between guest molecules and the OMS. Because CO_2 adsorption under low-pressure conditions is largely due to this interaction, the variable of OMS in ***X*** may have contributed to the prediction. However, as the pressure increases, the influence of the interaction becomes relatively small, and the volume and the surface area of the pores have a large effect. Therefore, it is desirable to

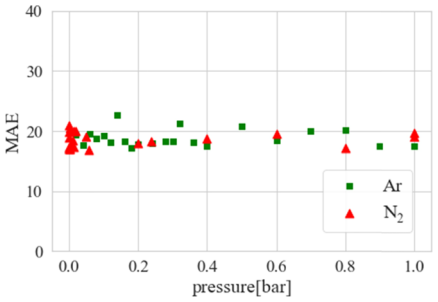


Figure 9. MAE scores for the predicted N_2 and Ar adsorption capacity at each pressure with model 2.

change the *x* variables used on the basis of the pressure to improve the predictive ability of the model. Regarding the difference in the results depending on the type of gas, the major difference between CO_2 and the other two gases is the molecular polarity. The CO_2 molecule more strongly interacts with the OMS than the nonpolar molecules N_2 and Ar. It is also known that Ar adsorbs on the pore walls at low pressures.[75] Therefore, the physical properties of the organic linker, which is the pore wall, play a major role in adsorption, and variables describing the physical properties of the organic linker will be required. From the adsorption isotherms, the amount of gas adsorption in each type of MOF differs greatly, but the differences can be recognized and predicted from the structure property alone. It is confirmed that the amount of gas adsorption has a strong relationship with the structure property.

4. CONCLUSIONS

We have developed two regression models to clarify the overall relationship between the components, structure properties, and physical properties of MOFs. To represent MOFs, molecular descriptors were calculated as components, which were used to decompose the MOFs into SBUs and organic compounds. The structure properties were calculated using the Zeo++ simulation. Model 1 can predict the structure properties from the components and shows the relationship between the components and the structure properties. For the physical properties, we focused on the amount of the gas adsorption of CO_2, N_2, and Ar and predicted the amounts of gas adsorption for wide ranges of temperature and pressure. Model 2 can predict the amount of gas adsorption from the structure properties. It was confirmed that the types of adsorption isotherms differed for the MOFs, but they were dependent on the structure properties. As a result, the two regression models were able to predict the *y* values with high accuracy. Based on the obtained results, when we want to obtain a MOF with a desired amount of gas adsorption, we can predict the structure properties and physical properties. In addition, when we know the components of the MOF before performing experiments, we can design a MOF with desired gas adsorption using the two models.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jcim.1c01205.

Details about the plots and scores for structure property prediction with model 1 (PDF)

https://doi.org/10.1021/acs.jcim.1c01205
J. Chem. Inf. Model. 2021, 61, 5785−5792

AUTHOR INFORMATION

Corresponding Author

Hiromasa Kaneko − *Department of Applied Chemistry, School of Science and Technology, Meiji University, Kawasaki, Kanagawa 214-8571, Japan;* orcid.org/0000-0001-8367-6476; Phone: +81-44-934-7197; Email: hkaneko@meiji.ac.jp

Author

Shunsuke Yuyama − *Department of Applied Chemistry, School of Science and Technology, Meiji University, Kawasaki, Kanagawa 214-8571, Japan;* orcid.org/0000-0002-6695-3061

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jcim.1c01205

Notes

The authors declare no competing financial interest.

The MOF crystal structure data are available in the CoRE-MOF dataset, and Ar and N_2 gas adsorption data are available in MOFDB. The CO_2 gas adsorption data that support the findings of this study were obtained from the references.

ACKNOWLEDGMENTS

This work was supported by a Grant-in-Aid for Scientific Research (KAKENHI) (Grant Number 19 K15352) from the Japan Society for the Promotion of Science. We thank Tim Cooper, PhD, from Edanz (https://jp.edanz.com/ac) for editing a draft of this manuscript.

REFERENCES

(1) Kitagawa, S. Studies on Synthesis and Properties of Porous Coordination Polymers. *Bull. J. Soc. Coord. Chem.* **2008**, *51*, 13−19.

(2) 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*, 276−279.

(3) Kalmutzki, M. J.; Hanikel, N.; Yaghi, O. M. Secondary building units as the turning point in the development of the reticular chemistry of MOFs. *Sci. Adv.* **2018**, *4*, No. eaat9180.

(4) Morris, R. E. Modular materials from zeolite-like building blocks. *J. Mater. Chem.* **2005**, *15*, 931−938.

(5) Tranchemontagne, D. J.; Mendoza-Cortés, J. L.; O'Keeffe, M.; Yaghi, O. M. Secondary building units, nets and bonding in the chemistry of metal−organic frameworks. *Chem. Soc. Rev.* **2009**, *38*, 1257−1283.

(6) Serre, C.; Millange, F.; Surblé, S.; Férey, G. A Route to the Synthesis of Trivalent Transition-Metal Porous Carboxylates with Trimeric Secondary Building Units. *Angew. Chem., Int. Ed. Engl.* **2004**, *43*, 6285−6289.

(7) Dincă, M.; Long, J. R. Hydrogen Storage in Microporous Metal−Organic Frameworks with Exposed Metal Sites. *Angew. Chem., Int. Ed. Engl.* **2008**, *47*, 6766−6779.

(8) Wiersum, A. D.; Chang, J. S.; Serre, C.; Llewellyn, P. L. An Adsorbent Performance Indicator as a First Step Evaluation of Novel Sorbents for Gas Separations: Application to Metal−Organic Frameworks. *Langues* **2013**, *29*, 3301−3309.

(9) Dietzel, P. D. C.; Johnsen, R. E.; Fjellvåg, H.; Bordiga, S.; Groppo, E.; Chavan, S.; Blom, R. Adsorption properties and structure of CO_2 adsorbed on open coordination sites of metal−organic framework Ni_2(dhtp) from gas adsorption, IR spectroscopy and X-ray diffraction. *Chem. Commun.* **2008**, *41*, 5125−5127.

(10) Li, H.; Wang, K.; Sun, Y.; Lollar, C. T.; Li, J.; Zhou, H. -C. Recent advances in gas storage and separation using metal−organic frameworks. *Mater. Today* **2018**, *21*, 108−121.

(11) Mukherjee, S.; Manna, B.; Desai, A. V.; Yin, Y.; Krishna, R.; Babarao, R.; Ghosh, S. K. Harnessing Lewis acidic open metal sites of metal−organic frameworks: the foremost route to achieve highly selective benzene sorption over cyclohexane. *Chem. Commun.* **2016**, *52*, 8215−8218.

(12) Bloch, E. D.; Queen, W. L.; Krishna, R.; Zadrozny, J. M.; Brown, C. M.; Long, J. R. Hydrocarbon Separations in a Metal−Organic Framework with Open Iron(II) Coordination Sites. *Science* **2012**, *335*, 1606−1610.

(13) Diercks, C. S.; Liu, Y.; Cordova, K. E.; Yaghi, O. M. The role of reticular chemistry in the design of CO_2 reduction catalysts. *Nat. Mater.* **2018**, *17*, 301−307.

(14) Dan-Hardi, M.; Serre, C.; Frot, T.; Rozes, L.; Maurin, G.; Sanchez, C.; Férey, G. A New Photoactive Crystalline Highly Porous Titanium(IV) Dicarboxylate. *J. Am. Chem. Soc.* **2009**, *131*, 10857−10859.

(15) Chen, Y.; Sakata, O.; Nanba, Y.; Kumara, L. S. R.; Yang, A.; Song, C.; Koyama, M.; Li, G.; Kobayashi, H.; Kitagawa, H. Electronic origin of hydrogen storage in MOF-covered palladium nanocubes investigated by synchrotron X-rays. *Commun. Chem.* **2018**, *1*, 61.

(16) Li, Z.; Yu, R.; Huang, J.; Shi, Y.; Zhang, D.; Zhong, X.; Wang, D.; Wu, Y.; Li, Y. Platinum−nickel frame within metal−organic framework fabricated *in situ* for hydrogen enrichment and molecular sieving. *Nat. Commun.* **2015**, *6*, 8248.

(17) Hirai, K.; Furukawa, S. Surface Chemistry of Porous Coordination Polymers (PCPs) or Metal−Organic Frameworks (MOFs). *Hyomen Kagaku* **2012**, *33*, 519−523.

(18) Kouno, H.; Ogawa, T.; Amemori, S.; Mahato, P.; Yanai, N.; Kimizuka, N. Triplet energy migration-based photon upconversion by amphiphilic molecular assemblies in aerated water. *Chem. Sci.* **2016**, *7*, 5224−5229.

(19) Huang, C.; Liu, C.; Chen, X.; Xue, Z.; Liu, K.; Qiao, X.; Li, X.; Lu, Z.; Zhang, L.; Lin, Z.; Wang, T. A Metal−Organic Framework Nanosheet-Assembled Frame Film with High Permeability and Stability. *Adv. Sci.* **2020**, *7*, No. 193180.

(20) Sturluson, A.; Huynh, M. T.; Kaija, A. R.; Laird, C.; Yoon, S.; Hou, F.; Feng, Z.; Wilmer, C. E.; Colón, Y. J.; Chung, Y. G.; Siderius, D. W.; Simon, C. M. The role of molecular modelling and simulation in the discovery and deployment of metal−organic frameworks for gas storage and separation. *Mol. Simul.* **2019**, *45*, 1082−1121.

(21) Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A., III; Skiff, W. M. UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *J. Am. Chem. Soc.* **1992**, *114*, 10024−10035.

(22) Mayo, S. L.; Olafson, B. D.; Goddard, W. A. DREIDING: a generic force field for molecular simulations. *J. Phys. Chem.* **1990**, *94*, 8897−8909.

(23) Becker, T. M.; Lin, L. -C.; Dubbeldam, D.; Vlugt, T. J. H. Polarizable Force Field for CO_2 in M-MOF-74 Derived from Quantum Mechanics. *J. Phys. Chem. C* **2018**, *122*, 24488−24498.

(24) Férey, G.; Serre, C. Large breathing effects in three-dimensional porous hybrid matter: facts, analyses, rules and consequences. *Chem. Soc. Rev.* **2009**, *38*, 1380−1399.

(25) Férey, G.; Serre, C.; Devic, T.; Maurin, G.; Jobic, H.; Llewellyn, P. L.; Weireld, G. D.; Vimont, A.; Daturif, M.; Chang, J.-S. Why hybrid porous solids capture greenhouse gases? *Chem. Soc. Rev.* **2011**, *40*, 550−562.

(26) Schneemann, A.; Bon, V.; Schwedler, I.; Senkovska, I.; Kaskel, S.; Fischer, R. A. Flexible metal−organic frameworks. *Chem. Soc. Rev.* **2014**, *43*, 6062−6096.

(27) Fanourgakis, G. S.; Gkagkas, K.; Tylianakis, E.; Froudakis, G. A Generic Machine Learning Algorithm for the Prediction of Gas Adsorption in Nanoporous Materials. *J. Phys. Chem. C* **2020**, *124*, 7117−7126.

(28) Basdogana, Y.; Keskin, S. Simulation and modelling of MOFs for hydrogen storage. *CrystEngComm* **2015**, *17*, 261−275.

(29) Chen, L.; Morrison, C. A.; Düren, T. Improving Predictions of Gas Adsorption in Metal−Organic Frameworks with Coordinatively

Unsaturated Metal Sites: Model Potentials, ab initio Parameterization, and GCMC Simulations. *J. Phys. Chem. C* **2012**, *116*, 18899−18909.
(30) Katritzky, A. R.; Sild, S.; Karelson, M. Correlation and Prediction of the Refractive Indices of Polymers by QSPR. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 1171−1176.
(31) 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.; Smit, B. Correction to Materials Genome in Action: Identifying the Performance Limits of Physical Hydrogen Storage. *Chem. Mater.* **2017**, *29*, 10243.
(32) Pardakhti, M.; Moharreri, E.; Wanik, D.; Suib, S. L.; Srivastava, R. Machine Learning Using Combined Structural and Chemical Descriptors for Prediction of Methane Adsorption Performance of Metal Organic Frameworks (MOFs). *ACS Comb. Sci.* **2017**, *19*, 640−645.
(33) Fernandez, M.; Trefiak, N. R.; Woo, T. K. Atomic Property Weighted Radial Distribution Functions Descriptors of Metal−Organic Frameworks for the Prediction of Gas Uptake Capacity. *J. Phys. Chem. C* **2013**, *117*, 14095−14105.
(34) Rosen, A. S.; Notestein, J. M.; Snurr, R. Q. Structure−Activity Relationships That Identify Metal−Organic Framework Catalysts for Methane Activation. *ACS Catal.* **2019**, *9*, 3576−3587.
(35) 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-chemical properties of metal−organic frameworks for accelerated materials discovery. *Materials* **2021**, *4*, 1578−1597.
(36) Kancharlapalli, S.; Gopalan, A.; Haranczyk, M.; Snurr, R. Q. Fast and Accurate Machine Learning Strategy for Calculating Partial Atomic Charges in Metal−Organic Frameworks. *J. Chem. Theory Comput.* **2021**, *17*, 3052−3064.
(37) Bobbitt, N. S.; Snurr, R. Q. Molecular modelling and machine learning for high-throughput screening of metal−organic frameworks for hydrogen storage. *Mol. Simul.* **2019**, *45*, 1069−1081.
(38) Weininger, D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 31−36.
(39) Dalby, A.; Nourse, J. G.; Hounshell, W. D.; Gushurst, A. K. I.; Grier, D. L.; Leland, B. A.; Laufer, J. Description of several chemical structure file formats used by computer programs developed at Molecular Design Limited. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 244−255.
(40) Bucior, B. J.; Rosen, A. S.; Haranczyk, M.; Yao, Z.; Ziebel, M. E.; Farha, O. K.; Hupp, J. T.; Siepmann, J. I.; Aspuru-Guzik, A.; Snurr, R. Q. Identification Schemes for Metal−Organic Frameworks To Enable Rapid Search and Cheminformatics Analysis. *Cryst. Growth Des.* **2019**, *19*, 6682−6697.
(41) https://rdkit.org/docs_jp/index.html (accessed April 18, 2021).
(42) https://www.rdkit.org/docs/source/rdkit.Chem.Descriptors.html#rdkit.Chem.Descriptors.ExactMolWt (accessed April 18, 2021).
(43) https://pymatgen.org/introduction.html (accessed April 18, 2021).
(44) Ong, S. P.; Richards, W. D.; Jain, A.; Hautier, G.; Kocher, M.; Cholia, S.; Gunter, D.; Chevrier, V. L.; Persson, K. A.; Ceder, G. Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis. *Comput. Mater. Sci.* **2013**, *68*, 314−319.
(45) http://zeoplusplus.org/ (accessed April 18, 2021).
(46) Willems, T. F.; Rycroft, C. H.; Kazi, M.; Meza, J. C.; Haranczyk, M. Algorithms and tools for high-throughput geometry-based analysis of crystalline porous materials. *Micro Mes. Mater.* **2012**, *149*, 134−141.
(47) Martin, R. L.; Smit, B.; Haranczyk, M. Addressing challenges of identifying geometrically diverse sets of crystalline porous materials. *J. Chem. Info Model.* **2012**, *52*, 308−318.
(48) https://mof.tech.northwestern.edu/ (accessed April 18, 2021).
(49) https://datachemeng.com/iterative_gaussian_mixture_regression/ (accessed April 18, 2021).
(50) Park, Y. K.; Choi, S. B.; Kim, H.; Kim, K.; Won, B. H.; Choi, K.; Choi, J. S.; Ahn, W. S.; Won, N.; Kim, S.; Jung, D. H.; Choi, S. H.; Kim, G. H.; Cha, S. S.; Jhon, Y. H.; Yang, J. K.; Kim, J. Crystal Structure and Guest Uptake of a Mesoporous Metal−Organic Framework Containing Cages of 3.9 and 4.7 nm in Diameter. *Ang. Chem., Int. Ed.* **2007**, *46*, 8230−8233.
(51) Taylor, M. K.; Runčevski, T.; Oktawiec, J.; Gonzalez, M. I.; Siegelman, R. L.; Mason, J. A.; Ye, J.; Brown, C. M.; Long, J. R. Tuning the Adsorption-Induced Phase Change in the Flexible Metal−Organic Framework Co(bdp). *J. Am. Chem. Soc.* **2016**, *138*, 15019−15026.
(52) Furukawa, H.; Ko, N.; Go, Y. B.; Aratani, N.; Choi, S. B.; Choi, E.; Yazaydin, A. Ö.; Snurr, R. Q.; O'Keeffe, M.; Kim, J.; Yaghi, O. M. Ultrahigh Porosity in Metal−Organic Frameworks. *Science* **2010**, *329*, 424−428.
(53) Millward, A. R.; Yaghi, O. M. Metal−Organic Frameworks with Exceptionally High Capacity for Storage of Carbon Dioxide at Room Temperature. *J. Am. Chem. Soc.* **2005**, *127*, 17998−17999.
(54) Tan, C.; Yang, S.; Champness, N. R.; Lin, X.; Blake, A. J.; Lewisa, W.; Schröder, M. High capacity gas storage by a 4,8-connected metal−organic polyhedral framework. *Chem. Commun.* **2011**, *47*, 4487−4489.
(55) Herm, Z. R.; Swisher, J. A.; Smit, B.; Krishna, R.; Long, J. R. Metal−Organic Frameworks as Adsorbents for Hydrogen Purification and Precombustion Carbon Dioxide Capture. *J. Am. Chem. Soc.* **2011**, *133*, 5664−5667.
(56) Zheng, B.; Liu, H.; Wang, Z.; Yu, X.; Yia, P.; Bai, J. Porous NbO-type metal−organic framework with inserted acylamide groups exhibiting highly selective CO_2 capture. *CrystEngComm* **2013**, *15*, 3517−3520.
(57) Yuan, D.; Zhao, D.; Sun, D.; Zhou, H. C. An Isoreticular Series of Metal−Organic Frameworks with Dendritic Hexacarboxylate Ligands and Exceptionally High Gas-Uptake Capacity. *Ang. Chem., Int. Ed.* **2010**, *49*, 5357−5361.
(58) Benson, O.; Silva, I. D.; Argent, S. P.; Cabot, R.; Savage, M.; Godfrey, H. G. W.; Yan, Y.; Parker, S. F.; Manuel, P.; Lennox, M. J.; Mitra, T.; Easun, T. L.; Lewis, W.; Blake, A. J.; Besley, E.; Yang, S.; Schröder, M. Amides Do Not Always Work: Observation of Guest Binding in an Amide-Functionalized Porous Metal−Organic Framework. *J. Am. Chem. Soc.* **2016**, *138*, 14828−14831.
(59) Dietzel, P. D. C.; Besikiotisa, V.; Blom, R. Application of metal−organic frameworks with coordinatively unsaturated metal sites in storage and separation of methane and carbon dioxide. *J. Mater. Chem.* **2009**, *19*, 7362−7370.
(60) Zheng, B.; Yang, Z.; Bai, J.; Lia, Y.; Li, S. High and selective CO_2 capture by two mesoporous acylamide-functionalized rht-type metal−organic frameworks. *Chem. Commun.* **2012**, *48*, 7025−7027.
(61) Duan, J.; Yang, Z.; Bai, J.; Zheng, B.; Lia, Y.; Li, S. Highly selective CO_2 capture of an agw-type metal−organic framework with inserted amides: experimental and theoretical studies. *Chem. Commun.* **2012**, *48*, 3058−3060.
(62) Kukulka, W.; Cendrowski, K.; Michalkiewicz, B.; Mijowska, E. MOF-5 derived carbon as material for CO_2 absorption. *RSC Adv.* **2019**, *9*, 18527−18537.
(63) Sumida, K.; Hill, M. R.; Horike, S.; Dailly, A.; Long, J. R. Synthesis and Hydrogen Storage Properties of $Be_{12}(OH)_{12}$(1,3,5-benzenetribenzoate)$_4$. *J. Am. Chem. Soc.* **2009**, *131*, 15120−15121.
(64) Zheng, B.; Bai, J.; Duan, J.; Wojtas, L.; Zaworotko, M. J. Enhanced CO_2 Binding Affinity of a High-Uptake rht-Type Metal−Organic Framework Decorated with Acylamide Groups. *J. Am. Chem. Soc.* **2011**, *133*, 748−751.
(65) Müller, P.; Bon, V.; Senkovska, I.; Getzschmann, J.; Weiss, M. S.; Kaskel, S. Crystal Engineering of Phenylenebis(azanetriyl)-tetrabenzoate Based Metal−Organic Frameworks for Gas Storage Applications. *Cryst. Growth Des.* **2017**, *17*, 3221−3228.
(66) Wang, Z.; Zheng, B.; Liu, H.; Lin, X.; Yu, X.; Yi, P.; Yun, R. High-Capacity Gas Storage by a Microporous Oxalamide-Functionalized NbO-Type Metal−Organic Framework. *Cryst. Growth Des.* **2013**, *13*, 5001−5006.

https://doi.org/10.1021/acs.jcim.1c01205
J. Chem. Inf. Model. 2021, 61, 5785−5792

(67) Alsmail, N. H.; Suyetin, M.; Yan, Y.; Cabot, R.; Krap, C. P.; Lü, J.; Easun, T. L.; Bichoutskaia, E.; Lewis, W.; Blake, A. J.; Schröder, M. Analysis of High and Selective Uptake of CO_2 in an Oxamide-Containing $\{Cu_2(OOCR)_4\}$-Based Metal−Organic Framework. *Chem. − Eur. J.* **2014**, *20*, 7317−7324.
(68) Xue, M.; Liu, Y.; Schaffino, R. M.; Xiang, S.; Zhao, X.; Zhu, G. S.; Qiu, S. L.; Chen, B. New Prototype Isoreticular Metal−Organic Framework $Zn_4O(FMA)_3$ for Gas Storage. *Inorg. Chem.* **2009**, *48*, 4649−4651.
(69) Chung, Y. G.; Camp, J.; Haranczyk, M.; Sikora, B. J.; Bury, W.; Krungleviciute, V.; Yildirim, T.; Farha, O. K.; Sholl, D. S.; Snurr, R. Q. Computation-Ready, Experimental Metal−Organic Frameworks: A Tool To Enable High-Throughput Screening of Nanoporous Crystals. *Chem. Mater.* **2014**, *26*, 6185−6192.
(70) https://www.ccdc.cam.ac.uk/ (accessed April 18, 2021).
(71) https://lightgbm.readthedocs.io/en/latest/index.html (accessed April 18, 2021).
(72) https://optuna.readthedocs.io/en/stable/index.html (accessed April 18, 2021).
(73) Vukotic, V. N.; O'Keefe, C. A.; Zhu, K.; Harris, K. J.; To, C.; Schurko, R. W.; Loeb, S. J. Mechanically Interlocked Linkers inside Metal−Organic Frameworks: Effect of Ring Size on Rotational Dynamics. *J. Am. Chem. Soc.* **2015**, *137*, 9643−9651.
(74) Li, C. R.; Li, S. L.; Zhang, X. M. D_3-Symmetric Supramolecular Cation $\{(Me_2NH_2)_6(SO_4)\}^{4+}$ As a New Template for 3D Homochiral (10,3)-a Metal Oxalates. *Cryst. Growth Des.* **2009**, *9*, 1702−1707.
(75) Cho, H. S.; Deng, H.; Miyasaka, K.; Dong, Z.; Cho, M.; Neimark, A. V.; Kang, J. K.; Yaghi, O. M.; Terasaki, O. Extra adsorption and adsorbate superlattice formation in metal−organic frameworks. *Nature* **2015**, *527*, 503−507.

https://doi.org/10.1021/acs.jcim.1c01205
J. Chem. Inf. Model. 2021, 61, 5785−5792