Electronic Supplementary Material (ESI) for Materials Advances.

Electronic Supplementary Information (ESI)

Deep Learning Neural Network Potential for Simulating Gaseous Adsorption in Metal-Organic Frameworks

Chi-Ta Yang[a], Ishan Pandey[b], Dan Trinh[b], Chau-Chyun Chen[b], Joshua D. Howe[b], and Li-Chiang Lin[a,c,*]

[a]William G. Lowrie Department of Chemical and Biomolecular Engineering, The Ohio State University, Columbus, Ohio 43210, USA

[b]Department of Chemical Engineering, Texas Tech University, Lubbock, TX 79409, USA

[c]Department of Chemical Engineering, National Taiwan University, Taipei 10617, Taiwan

*Email: lclin@ntu.edu.tw

Table of Content:

1. Machine Learning

1.1 Learning Algorithms

Details of the machine learning (ML) algorithms[1,2] adopted in this work are described in the following along with their associated parameters. In this work, the building, learning, and evaluation of random forests (RF), support vector machines (SVM), k-nearest neighbors (KNN) were performed using Python with Scikit-learn module,[1] while those of artificial neural network (ANN) were performed using TensorFlow 2 (version 2.4.1)[3] library with Keras (version 2.4.0)[4] API. To quantify the performance of the trained ML models, mean absolute error (MAE) and the root mean square error (RMSE) were calculated. A ratio of 0.8 was used to split the datasets into training and test sets.

Random Forests. The random forests[5,6] (RF) algorithm is a powerful algorithm for classification and regression and is an ensemble of decision Trees[7]. The decision tree is developed by asking series of questions regarding features, which results in decision nodes in the tree. A leaf node is the decision node that does not have any child nodes, and the prediction is made based on this node. The classification and regression tree[8] (CART) algorithm embedded in Scikit-Learn[1] was used to develop the nodes, where Gini impurity and mean square error (MSE) were used as the cost function, respectively, in classification and regression tasks. CART searches for the feature and the corresponding threshold to produce pure subsets by minimizing impurity (in classification) and MSE (in regression). RF can introduce extra randomness since the best split is determined either from all input features or a random subset of features during the construction of a tree. In this work, the best split was based on the total number of features, and the number of trees in the forest was 100.

Support Vector Machines. Support vector machines[9,10] (SVM) is a supervised learning algorithm that is capable of performing linear/nonlinear classification and regression tasks. The working principle is that upon the classification of two classes, SVM fits a widest street (margins) to separate the two classes. One major hyperparameter is the C value that serves as the weighting for the margin violation; hence, SVM with a larger C value would possibly lead to overfitting. Through the use of kernels[11], SVM is equipped with the capability to perform nonlinear classification (and regression). In this scenario, more features in higher dimensions are considered, resulting in linear classification in the

higher dimensional space. Polynomial and Gaussian radial basis function (RBF) are two common kernels in practice. SVM regression[10] (SVR) is built based on the concept of SVM classification, but in a reverse manner. Instead of fitting the widest street with a limited margin violation in classification, SVR fits the widest street to incorporate as many data as possible. SVR introduces another hyperparameter, epsilon, that determines the width of the street. In this work, a C value of 1 and an epsilon value of 0.1 were used along with linear, polynomial, and RBF kernels. Note the results from the linear kernel was used because it resulted in the better performance.

k-Nearest Neighbors. k-nearest neighbors (KNN) algorithm is a supervised learning algorithm that can perform classification and regression tasks. KNN assumes that similar things are near to each other, so the prediction of a query point is made based on the k nearest neighbors. In our settings, 5 nearest neighbors were used for prediction and each point in the local neighborhood contributed uniformly to prediction of a query point (uniform weights).

Artificial Neural Networks. Artificial neural network (ANN) is a machine learning model inspired by the biological neural networks that power animal brains, firstly introduced by Warren McCulloch and Walter Pitts back in 1943.[12] ANN is a versatile learning method that can tackle highly complex classification and regression tasks. A standard artificial neural network consists of an input layer, hidden layers, and an output layer. ANN is called a deep neural network (DNN) when a deep stack of hidden layers is packed. In an artificial neutral network for regression, the predicted value is calculated based on weights, biases and an activation function associated with each neuron and its' inputs, as illustrated in the equation below:

$$a^{(h)} = \emptyset(a^{(in)}W^{(h)} + b)$$

, where $a^{(in)}$ is the feature/input vector of a sample and b is the vector of the bias terms, $W^{(h)}$ is the weighting matrix, and $\emptyset$ is the activation function to output the activated value $a^{(h)}$ for the next layer. Therefore, artificial neutral network is drastically different from classic force fields such as Lennard-Jones potential in which energy is calculated through a direct summation of contributions from pair-wise interactions between atoms.

In this work, five hidden layers were used except as noted in the study of ANN architecture (i.e., specifically denoted as DNN in the main text). In neural network models, the number of neurons in the first input layer was the number of input features, and one neuron was used in the output layer for the prediction. The number of neurons per hidden layer used in this work was 50

(resulting in a total of more than 30,000 parameters (e.g., weighting and biasing parameters) in our DNN), except in the section regarding the selection of features. The nonlinear or complexity nature of ANN is introduced by the activation function, and rectified linear activation function (ReLU) was used. The connections between layers are through the connection weights that provide a certain representation of the data. The learning process is to train ANN; backpropagation[13] training algorithm was used along with the Adam[14] algorithm (learning rate of 0.001), implemented in TensorFlow; 1,000 epoch was used in the training process.

1.2 Features

The feature or descriptor that encodes the atomic structure or configuration can either be from the local scope (ex: atom centered symmetry functions[15], and smooth overlap of atomic positions[16]) or global scope (ex: Coulomb matrix[17]), and the pair distance is the basic element. In adsorption applications using molecular simulations, the accurate description of short-range repulsion, dispersive interaction, and Coulombic interaction is important. To this end, we design features aiming to provide these pieces of information during the training process.

Specifically, the dispersion multipole expansion for two ground-state and spherically symmetric atoms or molecules, are considered as below.

$$U^{vdw} = -[\frac{C_4}{{r_{ij}}^4} + \frac{C_6}{{r_{ij}}^6} + \frac{C_8}{{r_{ij}}^8} + \frac{C_{10}}{{r_{ij}}^{10}} \ldots] \quad (1)$$

, where C_4 is associated with interactions of the back-to-back quadrupoles, C_6 is related with the interaction between two instantaneous dipoles, C_8 is regarding the interaction between a quadrupole and a dipole, and C_{10} includes the interactions between an octopole and a dipole and between two quadrupoles;[18,19] r_{ij} is the distance between atoms i and j. Regarding the Coulombic interaction, it is considered as below

$$U^{Coul} = \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \quad (2)$$

, where q_i and q_j are the atomic charges of atoms i and j, and ε_o is the permittivity of free space. In terms of repulsion interaction where the overlap of electronic wavefunction becomes significant, the valance repulsion dominates. Since electronic wavefunctions

decay exponentially as a function of distance, such repulsion interactions are well represented by the exponential form.[18,20] Therefore, the Born-Mayer[21] expression as below was used to provide ML algorithms with repulsion information.

$$U^{rep} = \sum A\, e^{-br_{ij}} \quad (3)$$

, where coefficient A and b are repulsive parameters.

Overall, the aforementioned terms (i.e., $e^{-r_{ij}}, r_{ij}^{-1}, r_{ij}^{-4}, r_{ij}^{-6}, r_{ij}^{-8}, r_{ij}^{-10}$) were adopted to provide ML algorithms with insights into adsorbate-adsorbent interactions. Also, N pairs of the smallest distances in each adsorbate-adsorbent atom pair were used to represent the environment of each adsorption configuration as depicted in Fig. 1 of the main text; N=4 was used and the influence of N was also discussed in the main text. Note that although the incorporation of more terms and more pair distances could possibly offer a more accurate description, it is prone to overfitting given the rather limited size of DFT training data. The impact of the feature selection was also studied and discussed in the main text. The min-max normalization was used for better training.

For the adsorbate-adsorbent atomic pairs, we considered the pairs between each element type of the adsorbate and each atom type of the adsorbent; for the latter, we differentiated Oa, Ob, Oc from O atom and Ca, Cb, Cc, Cd from C atoms of Mg-MOF-74 due to different chemistry exhibited.[20,22] This leads to a total of 9 different atom types for Mg-MOF-74 (i.e., Mg, Oa, Ob, Oc, Ca, Cb, Cc, Cd, H). Using H_2O as an example for our implementation, there are two distinct types of elements (i.e., H atoms (H1 and H2) and one O atom), so pairs of H (H_2O) and different atom types of adsorbent (H-adsorbent, i.e., H1 and H2 are not distinguished) and pairs of O(H_2O) and different atom types of adsorbent (O-adsorbent) were adopted as features. This leads to a total of 2x9=18 pairs. For each pair, as noted above, per the adsorption configuration, we determined the N shortest distances for each pair, and N is generally assigned as 4 unless otherwise noted.

Note that we found such an approach (i.e., considering the element type for the adsorbate) can yield better results as compared to treating two hydrogens separately. Specifically, testing models were built using four kinds of feature designs. The first feature set includes all the atomic pairs, i.e., H1- adsorbent, H2-adsorbent, and O-adsorbent. The second feature set considers the element pairs (O-adsorbent and H-adsorbent as

mentioned above). The third feature set specifies the H atom that is closest to a Mg of MOF as H1 and the other as H2, and then considers all the atomic pairs. The fourth feature set is similar to the third one except H1 is the closest H atom to MOF. The four feature designs are summarized in Table S1 as well as the corresponding MAE values of the test set; feature 2 (the adopted approach) exhibits a smaller value.

Table S1. Test of feature adoption upon H_2O adsorption in Mg-MOF-74 with the corresponding MAE of the test set. Note H1, H2, and O are the two H atoms and the O atom of H_2O, respectively.

	adsorbate-adsorbent atomic pairs	MAE of test set (kJ/mol)
Feature 1	H1-adsorbent, H2-adsorbent, O-adsorbent	5.27
Feature 2	H-adsorbent, O-adsorbent	4.37
Feature 3	H1 (closest to Mg) -adsorbent, H2-adsorbent, O-adsorbent	5.01
Feature 4	H1 (closest to framework) -adsorbent, H2-adsorbent, O-adsorbent	5.54

2 Monte Carlo Simulations

To generate training data points for the evaluation of potential ML algorithms as well as to generate results for the comparison of ML predicted energy with existing force filed approaches (UFF[23], Lin *et al.*[20], and scaling factor[24]), Monte Carlo (MC) simulations were performed using the open-source RASPA code[25], via various MC attempts of random translation, rotation, and reinsertion moves. The dimension of the simulation boxes in all the MC simulations was at least twice the cutoff radius (12Å) along each direction. The TraPPE force field was adopted for CO_2[26], while H_2O was modelled using the four-site TIP4P-Ew model[27]. CO molecule was described using models proposed by Lin and co-worker (i.e., the so-called ESP-MMs).[28] Regarding using MC to generate the training data set for evaluation upon CO_2 and H_2O adsorption, Buckingham potential was used with force field parameters taken from the work of Lin *et al.*[20] to describe the dispersive interactions between adsorbates and the framework; these results were also used for

comparison purpose. In comparison with results from UFF force field[23] and the scaling factor approach by Sholl and co-workers[24,29,30], 6-12 Lennard-Jones (L-J) potential was used. The Lorentz-Berthelot mixing rule was used. Coulombic potential was used to describe electrostatic interactions and the Ewald approach[31] was used to compute the long-range electrostatic interactions. The structure, force-fields parameters, and atomic charges of the framework atoms were taken from a previous work[20] except that the framework of Mg-MOF74 in CO adsorption was relaxed by the DFT method in this work. Note that even though the relaxed structure of Mg-MOF74 in CO adsorption is slightly different from that in the other two cases, it should not affect the corresponding ML results in each CO_2, H_2O, and CO adsorption case, since the comparison is among different force fields methods (i.e. DNN, UFF[23], Lin *et al.*[20], and scaling factor[24]) in each case.

The Widom particle insertion method[32] was used to compute the Henry coefficients of adsorption (i.e., initial isotherm slope at an infinite dilution condition). The Henry coefficient of adsorption (K_H) was derived based on the ratio of the average Rosenbluth weights of the adsorbate in the framework, $\langle W \rangle$, and that in the gas phase, $\langle W_{ig} \rangle$, as

$$K_H = \frac{1}{\rho_f RT} \frac{\langle W \rangle}{\langle W_{ig} \rangle} \quad (3)$$

, where ρ_f the mass density of the framework, R is the gas constant, and T is the temperature. $\langle W \rangle$ can further reduce to the average of the Boltzmann factor $\langle \exp\left(^{-\Delta U}/_{RT}\right) \rangle$, where ΔU is the predicted adsorbate-adsorbent interaction of the inserted ghost molecule. In calculating K_H through ML force fields, interaction energies for millions of randomly generated configurations were calculated using equation (3) for the estimation of K_H values.

3 Density functional theory calculations for CO adsorption

We performed fully periodic density functional theory (DFT) calculations to study the adsorption of carbon monoxide (CO) in Mg-MOF-74 with a planewave basis set, using the Vienna Ab initio Simulation Package (VASP)[33–36]. All calculations were performed using Projector augmented wave pseudopotentials[37,38] and Perdew, Burke, and Ernzerhof (PBE) generalized gradient approximation (GGA)[39]. To account for the dispersion interactions which are important during the adsorption, the third version of

dispersion correction of Grimme et al. (D3)[40] was used in all calculations. The calculations were done using the planewave basis set energy cutoff of 1,000 eV with Brillouin zone sampling done at the Γ-point, with structures relaxed until the interionic forces were less than 0.01 eV/Å.

Beginning from the previously published structures[41] we first optimized a trigonal unit cell of 162 atoms using the PBE-D3+U approach. The computed lattice volumes were within 1.5% of the neutron power diffraction data published previously[42]. We then performed static calculations to compute the binding energy of randomly generated CO molecule configurations in the Mg-MOF-74 framework. These calculations were performed for a single molecule adsorption of CO at 2,000 randomly generated coordinates at the distance of 2.0-5.5 Å from the MOF. The interaction energy at each randomly generated CO location was calculated as

$$\Delta E = E_{CO+MOF} - E_{MOF} - E_{CO} \quad (4)$$

Where E_{CO+MOF}, E_{MOF}, and E_{CO} are the energy of the adsorbed phase, the energy of the bare MOF (adsorbent), and the energy of the adsorbate (CO) in gas phase, respectively. Here, a negative ΔE corresponds to an exothermic adsorption.

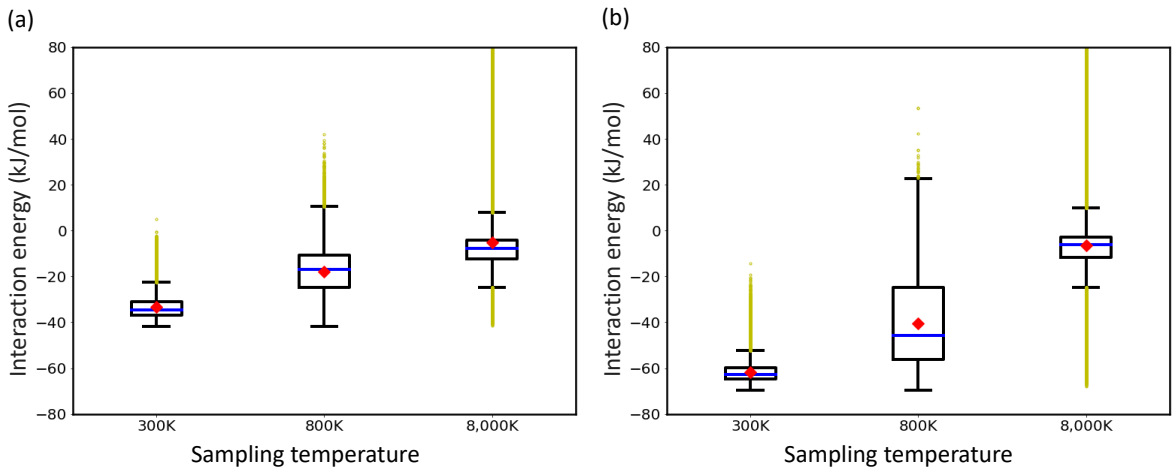

Fig. S1 Boxplots to show the energy distribution of datasets generated from NVT MC at temperatures of 300, 800, and 8,000 K for (a) CO_2 and (b) H_2O adsorption in Mg-MOF-74. Red diamond and blue line indicate the mean and median values, respectively

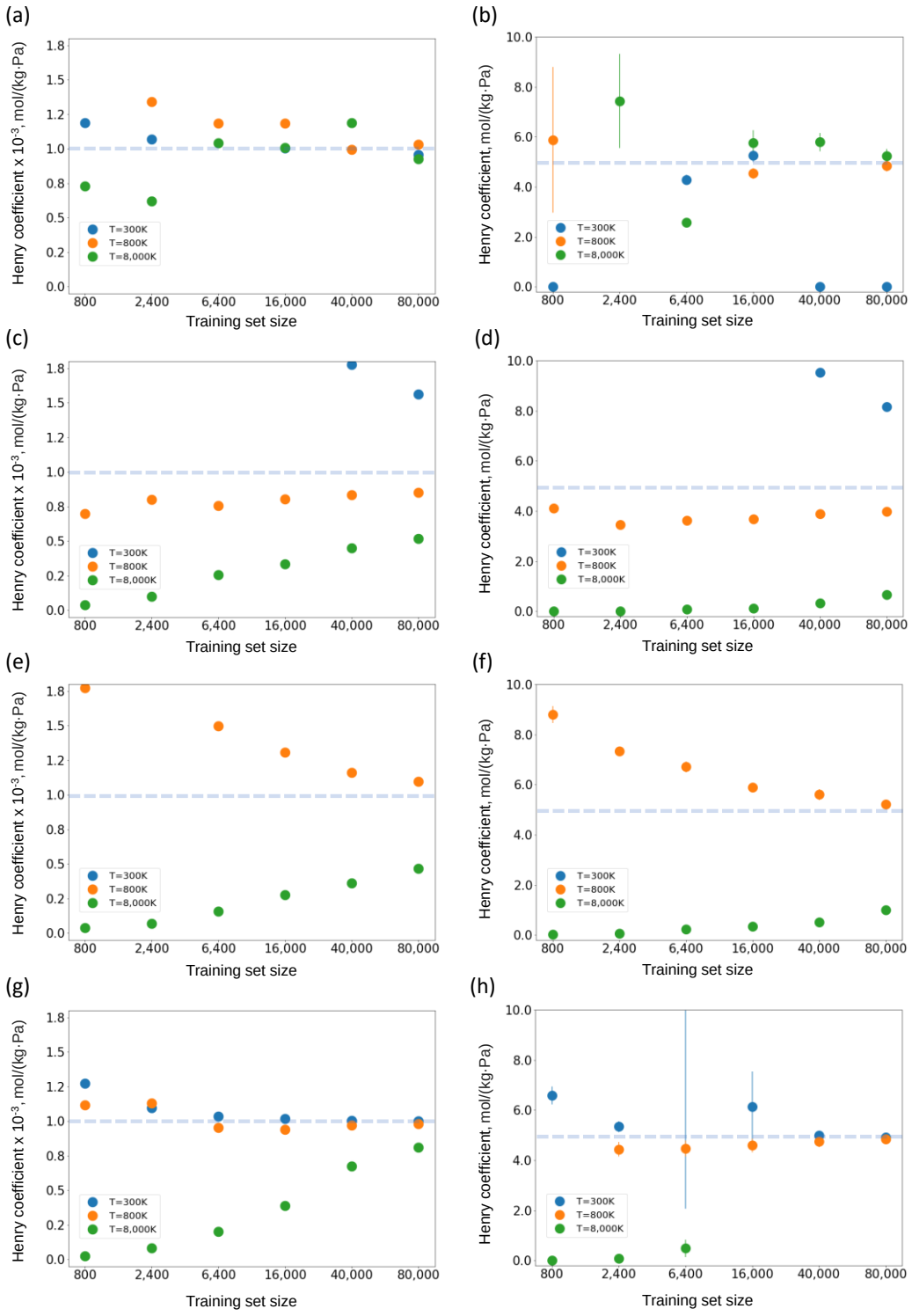

Fig. S2 Calculated Henry coefficients of CO_2 adsorption in Mg-MOF-74 using (a) DNN, (c) RF, (e) KNN, and (g) SVM algorithms. Likewise, (b, d, f, h) results of employing the respective algorithms for H_2O adsorption. The datasets from NVT MC at temperatures of 300, 800, and 8,000 K were used. Note values that are absent in the figures are poor predictions with values that are outside of the displayed range.

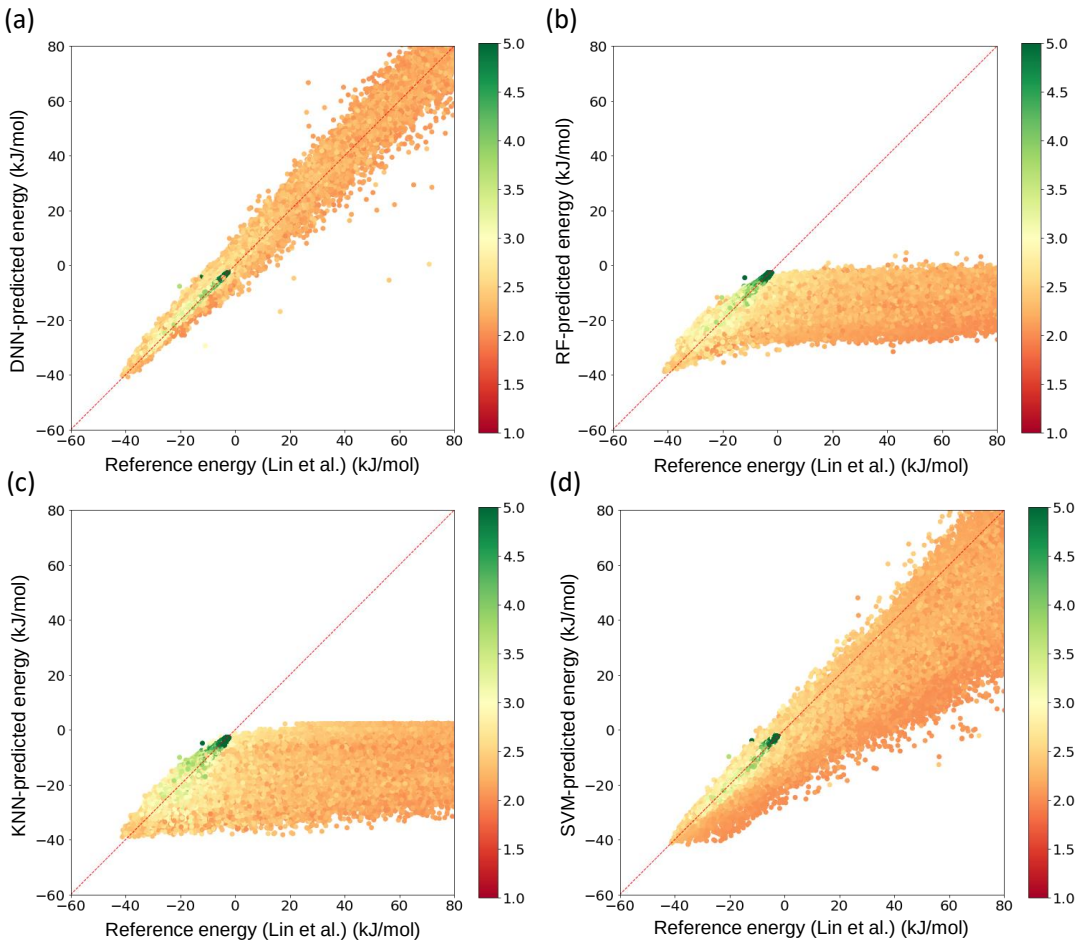

Fig. S3 Parity plots of the interaction energy computed by models employing (a) DNN (9.21, 2.04), (b) RF (52.38, 19.15), (c) KNN (53.65, 20.26), and (d) SVM (16.70, 5.50), trained with 2,400 points and that computed by the Lin *et al.*[20] force field for 0.4 M unseen random configurations for CO_2 adsorbed in Mg-MOF-74. The energies computed by the aforementioned force field were used for training and testing. The values in the parenthesis are the RMSE and MAE, respectively, of the predicted energies w.r.t. to the reference energies. Note the dataset for DNN is from NVT MC at temperatures of 8,000 K, and those for RF, SVM, and KNN are from NVT MC at temperatures of 800 K based on the best fitting of each algorithm shown in Fig. S2. The color bar represents the shortest distance (Å) between CO_2 and the framework.

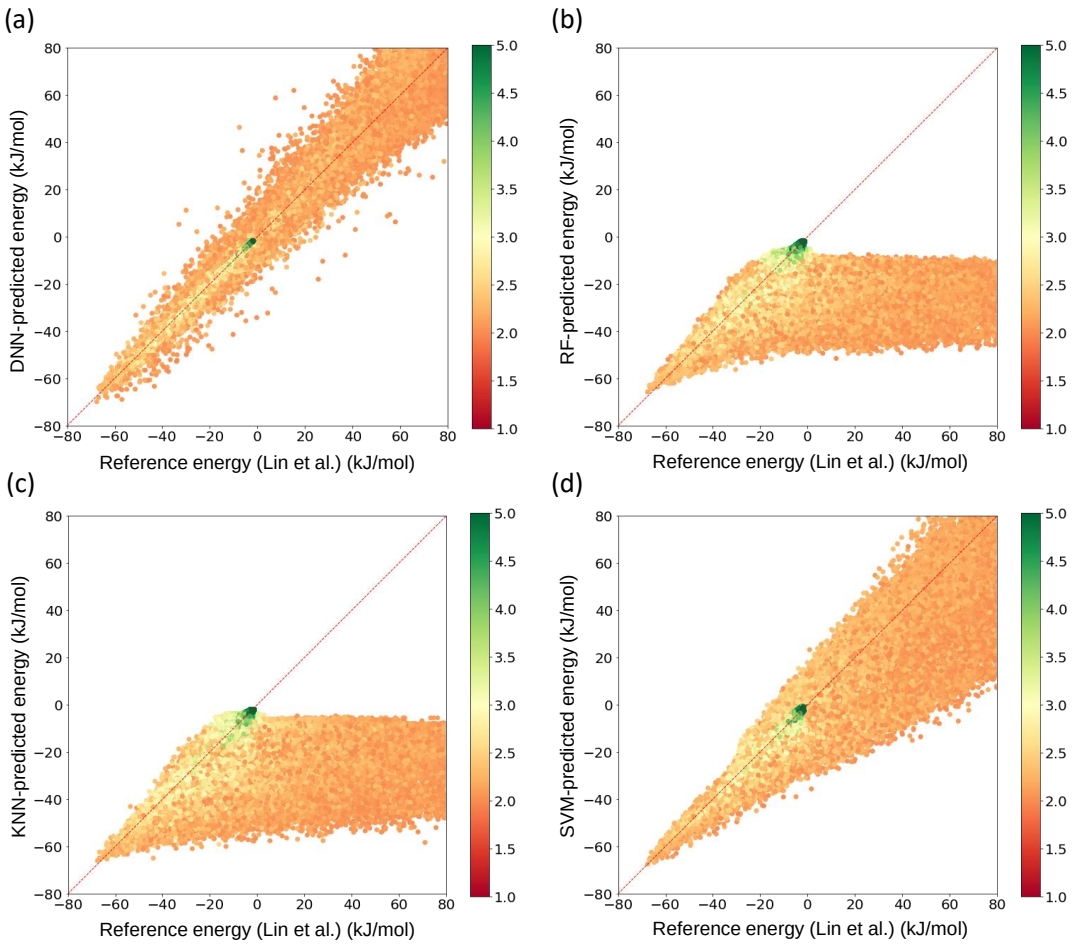

Fig. S4 Parity plots of the interaction energy computed by models employing (a) DNN (4.97, 1.32), (b) RF (28.33, 10.00), (c) KNN (28.52, 10.45), and (d) SVM (10.21, 4.09), trained with 2,400 points and that computed by the Lin *et al.*[20] force field for 0.4M unseen random configurations for H_2O adsorbed in Mg-MOF-74. The energies computed by the aforementioned force field were used for training and testing. The values in the parenthesis are the RMSE and MAE, respective, of the predicted energies w.r.t. to the reference energies. Note the dataset for DNN is from NVT MC at temperatures of 8,000 K, and those for RF, SVM, and KNN are from NVT MC at temperatures of 800 K based on the best fitting of each algorithm shown in Fig. S2. The color bar represents the distance (Å) between H_2O and the framework.

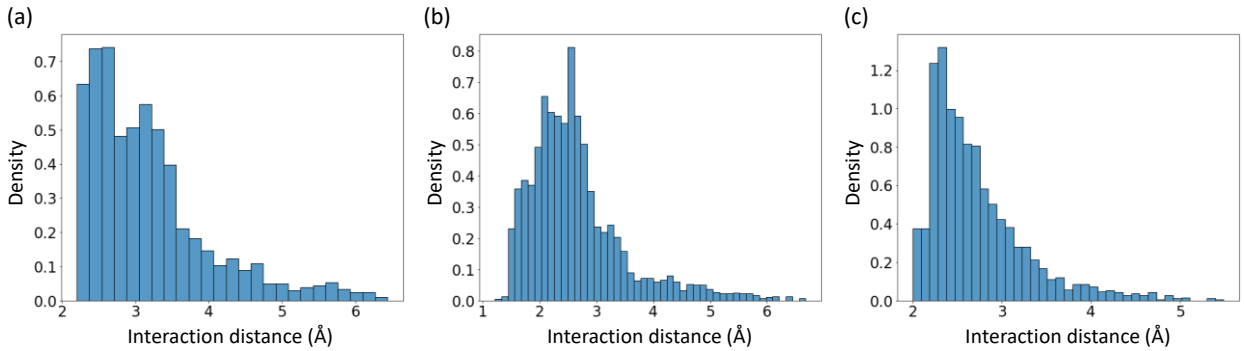

Fig. S5 Histogram of the shortest distance between the adsorbate and the MOF for configurations in the whole data sets of (a)CO_2, (b) H_2O, and (c) CO. Note that the configurations of CO_2 and H_2O adsorption used for DFT-derived DNN models were taken from the work of Lin *et al.*[20] In H_2O adsorption, there are 1,000 H_2O configurations located near the open metal sites and 2,000 H_2O configurations in the accessible pore volume. In CO_2 adsorption, there are 1,200 CO_2 random configurations inside the accessible pore volume of Mg-MOF-74. Regarding CO adsorption, a total of 2,000 CO adsorption configurations are in the distance of 2.0 ~ 5.5 Å from the framework.

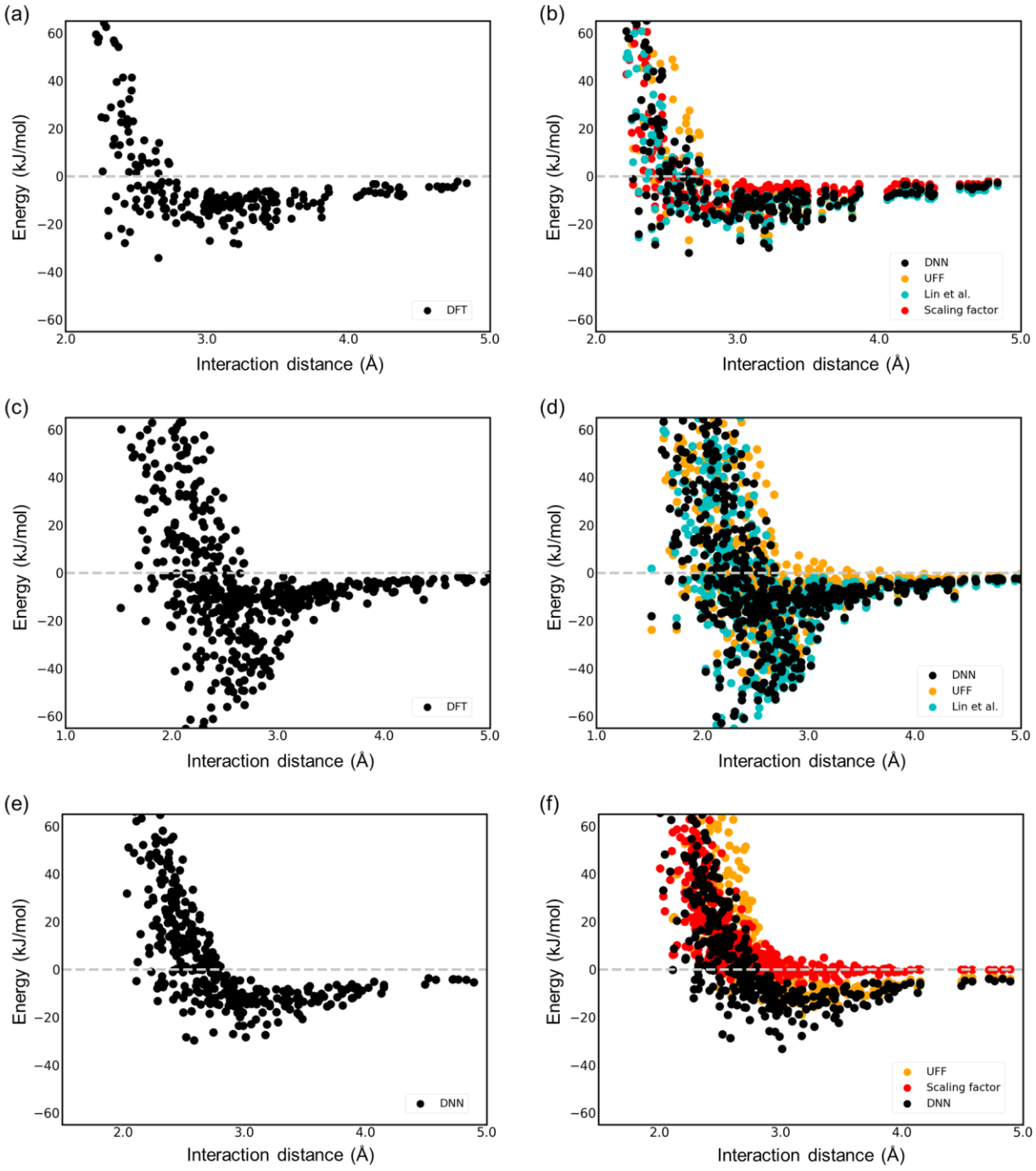

Fig. S6 The interaction energies for adsorption configurations (test sets) of (a, b) CO_2, (c, d) H_2O, and (e, f) CO in Mg-MOF-74 computed by (left) DFT and (right) the developed DNN force field of this study, UFF, the force field reported by Lin et al.[20] or the parameterized potential using the scaling factor as a function of the shortest distance between the adsorbate and the adsorbent.

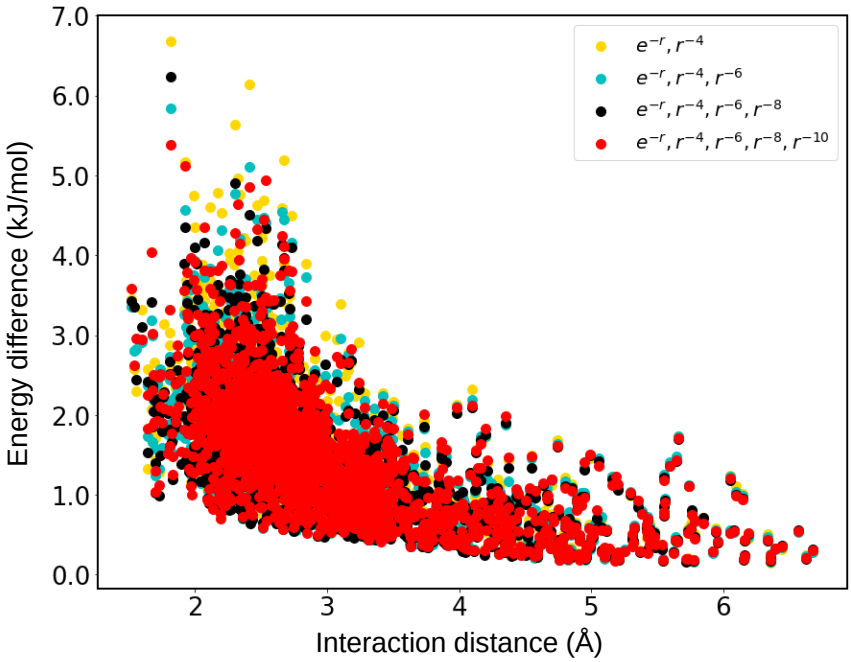

Fig. S7 Absolute energy difference between the DNN-predicted energies and the DFT reference data as a function of the shortest distance between H_2O and MOF upon different adopted features. Note the DNN of 5HL50 with the smallest four pair-distances are used in each feature entity. Results are based on 100 different trained NN models for each feature set. Configurations of interaction energies less than 10 kJ/mol were included in the calculations of RMSE.

REFERENCES

1 F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot and É. Duchesnay, *J. Mach. Learn. Res.*, 2011, **12**, 2825–2830.

2 A. Geron, *Hands-On Machine Learning with Scikit-Learn, Keras & TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*, O'Reilly Media, Inc., California, 2nd edn., 2019.

3 M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, R. Jozefowicz, Y. Jia, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, M. Schuster, R. Monga, S. Moore, D. Murray, C. Olah, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*, 2015, https://www.tensorflow.org/, (accessed 2021-12-01).

4 F. Chollet, *Keras*, https://github.com/fchollet/keras, (accessed 2021-12-01).

5 T. K. Ho, *IEEE Trans. Pattern Anal. Mach. Intell.*, 1998, **20**, 832–844.

6 L. Breiman, *Mach. Learn.*, 2001, **45**, 5–32.

7 J. R. Quinlan, *Mach. Learn.*, 1986, **1**, 81–106.

8 L. Breiman, J. H. Friedman, R. A. Olshen and C. J. Stone, *Classification and Regression Trees*, Chapman and Hall/CRC, Boca Raton, 1984

9 C. Cortes and V. Vapnik, *Mach. Learn.*, 1995, **20**, 273–297.

10 A. J. Smola and B. Schölkopf, *Stat. Comput.*, 2004, **14**, 199-222

11 B. Schölkopf, K. Tsuda and J.-P. Vert, *Kernel Methods in Computational Biology*, The MIT Press, Massachusetts, 2004.

12 W. S. McCulloch and W. Pitts, *Bull. Math. Biophys.*, 1943, **5**, 115–133.

13 D. E. Rumelhart, G. E. Hinton and R. J. Williams, *Nature*, 1986, **323**, 533–536.

14 D. P. Kingma and J. Ba, *Int. Conf. Learn. Represent*, 2015.

15 J. Behler, *J. Chem. Phys.*, 2011, **134**, 074106.

16 A. P. Bartók, R. Kondor and G. Csányi, *Phys. Rev. B*, 2013, **87**, 184115.

17 M. Rupp, A. Tkatchenko, K.-R. Müller and O. A. von Lilienfeld, *Phys. Rev. Lett.*, 2012, **108**, 058301.

18 R. J.-M. Pellenq and D. Nicholson, *J. Phys. Chem.*, 1994, **98**, 13339–13349.

19 D. J. Griffiths, *Introduction to Electrodynamics*, Prentice-Hall, Inc., New Jersey, 2nd edn., 1989.

20 L.-C. Lin, K. Lee, L. Gagliardi, J. B. Neaton and B. Smit, *J. Chem. Theory Comput.*, 2014, **10**, 1477–1488.

21 A. A. Abrahamson, *Phys. Rev.*, 1969, **178**, 76–79.

22 L.-C. Lin, A. H. Berger, R. L. Martin, J. Kim, J. A. Swisher, K. Jariwala, C. H. Rycroft, A. S. Bhown, M. W. Deem, M. Haranczyk and B. Smit, *Nat Mater*, 2012, **11**, 633–641.

23 A. K. Rappe, C. J. Casewit, K. S. Colwell, W. A. Goddard III and W. M. Skiff, *J. Am. Chem. Soc.*, 1992, **114**, 10024–10035.

24 H. Fang, P. Kamakoti, J. Zang, S. Cundy, C. Paur, P. I. Ravikovitch and D. S. Sholl, *J. Phys. Chem. C*, 2012, **116**, 10692–10701.

25 D. Dubbeldam, S. Calero, D. E. Ellis and R. Q. Snurr, *Mol. Simul.*, 2016, **42**, 81-101
26 M. G. Martin and J. I. Siepmann, *J. Phys. Chem. B*, 1998, **102**, 2569–2577.
27 H. W. Horn, W. C. Swope, J. W. Pitera, J. D. Madura, T. J. Dick, G. L. Hura and T. Head-Gordon, *J. Chem. Phys.*, 2004, **120**, 9665–9678.
28 E. H. Cho and L.-C. Lin, *J. Chem. Theory Comput.*, 2019, **15**, 6323–6332.
29 J. Zang, S. Nair and D. S. Sholl, *J. Phys. Chem. C*, 2013, **117**, 7519–7525.
30 H. Fang, R. Awati, S. E. Boulfelfel, P. I. Ravikovitch and D. S. Sholl, *J. Phys. Chem. C*, 2018, **122**, 12880–12891.
31 U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, *J. Chem. Phys.*, 1995, **103**, 8577–8593.
32 D. Frenkel and B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications*, Academic Press, California, 2nd edn., 2001.
33 G. Kresse and J. Hafner, *Phys. Rev. B Condens. Matter*, 1993, **47**, 558–561.
34 G. Kresse and J. Hafner, *Phys. Rev. B Condens. Matter*, 1994, **49**, 14251–14269.
35 G. Kresse and J. Furthmüller, *Comput. Mater. Sci.*, 1996, **6**, 15-50.
36 G. Kresse and J. Furthmüller, *Phys. Rev. B Condens. Matter*, 1996, **54**, 11169-11186.
37 P. E. Blöchl, *Phys. Rev. B Condens. Matter*, 1994, **50**, 17953–17979.
38 G. Kresse and D. Joubert, *Phys. Rev. B Condens. Matter*, 1999, **59**, 1758–1775.
39 J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, 1996, **77**, 3865–3868.
40 S. Grimme, S. Ehrlich and L. Goerigk, *J. Comput. Chem.*, 2011, **32**, 1456–1465.
41 K. Lee, J. D. Howe, L.-C. Lin, B. Smit and J. B. Neaton, *Chem. Mater.*, 2015, **27**, 668–678.
42 E. D. Bloch, M. R. Hudson, J. A. Mason, S. Chavan, V. Crocellà, J. D. Howe, K. Lee, A. L. Dzubak, W. L. Queen, J. M. Zadrozny, S. J. Geier, L.-C. Lin, L. Gagliardi, B. Smit, J. B. Neaton, S. Bordiga, C. M. Brown and J. R. Long, *J. Am. Chem. Soc.*, 2014, **136**, 10752–10761.