

## Selecting adsorbents to separate diverse near-azeotropic chemicals

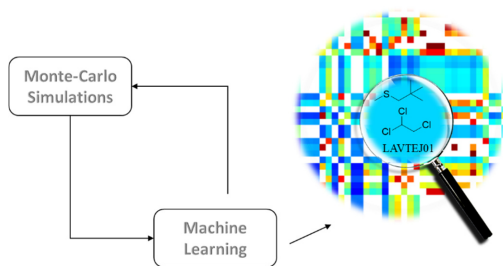
Farhad Gharagheizi , Dai Tang , David S. Sholl\* 

School of Chemical & Biomolecular Engineering, Georgia Institute of Technology, Atlanta, GA  
30332-0100, USA.

\*Correspondence to: [david.sholl@chbe.gatech.edu](mailto:david.sholl@chbe.gatech.edu)

**Abstract:** Industrial separations of near-azeotropic chemicals, species with very similar boiling points, are energy- and capital-intensive. Adsorption-based processes can energy-efficiently separate near-azeotropic mixtures provided suitable adsorbent materials can be found. Among the full diversity of industry-relevant molecules millions of these mixtures exist, meaning that discovery of mixture-specific adsorbents by direct experiment is infeasible. We show that vast numbers of adsorbents and adsorbing molecules can be explored in a powerful way by coupling atomistic simulations with machine learning. This concept is demonstrated by describing the adsorption of ~54,000 industry-relevant chemicals in an experimentally-derived set of thousands of metal-organic framework materials. Our results identify thousands of near-azeotropic mixtures that can be efficiently separated using adsorption and open possibilities for creating adsorption processes for complex mixtures with many components.

## Table of Contents Graphic



## Introduction

Chemical separations are integral to the chemical industry. Industrial practice is dominated by methods like distillation that use phase changes, so chemical separations use vast amounts of energy<sup>1</sup>. Distillation of near-azeotropic mixtures is particularly energy- and capital-intense. Adsorption-based separations are one alternative to distillation, provided that suitable adsorbent materials can be developed. Although multiple factors affect performance of adsorption processes<sup>2-3</sup>, a central quantity of interest is the adsorption isotherm, defining the equilibrium uptake of species of interest in an adsorbent. A comprehensive survey of experiments found ~10<sup>4</sup> single-component isotherms<sup>4</sup>. Mixture isotherms, however, are more important in assessing chemical separations. The number of mixture isotherms that have been experimentally measured is small; at best there are dozens of extant examples<sup>4-5</sup>. The sparsity of this data is illustrated by noting that collections of thousands of crystalline adsorbent materials<sup>6-7</sup> and billions of distinct molecules<sup>8</sup> are available.

Metal-organic frameworks (MOFs) are crystalline nanoporous materials composed of metal nodes coordinated to bi- or multi-functional organic linkers. MOFs exhibit high porosity and surface area<sup>9</sup>. More importantly, many of their properties can be varied by the diverse combinations of metal nodes and organic linkers that are available. As a result, MOFs have attracted considerable attention as sorbents for adsorption-based separation processes.

To date, two main categories of computational techniques have been used to study the adsorption in MOFs; molecular simulations and machine learning. The first category includes ab-initio quantum chemical approaches<sup>10</sup>, Monte-Carlo simulations<sup>11-12</sup>, molecular dynamics methods<sup>11</sup>, and hybrid molecular simulation methods<sup>10</sup>. Quantum chemical approaches normally lead to more accurate results but molecular simulations methods are typically far faster, making

them more suitable for high throughput studies<sup>13</sup>. Although multiple studies have used molecular simulations in high throughput computational screening of MOFs for CO<sub>2</sub> capture<sup>14</sup>, H<sub>2</sub>/N<sub>2</sub> separation<sup>15</sup>, CO<sub>2</sub>/H<sub>2</sub> separation<sup>16</sup>, H<sub>2</sub>/CH<sub>4</sub> separation<sup>17</sup>, separation of hexane isomers<sup>18</sup>, and capture of toxic chemicals<sup>19-20</sup>, there are limitations to using “computational brute-force” tool in large-scale high throughput computational screening of sorbents for adsorptive separation, particularly when millions of molecule/MOF systems are involved. A second category of computational approaches to predicting adsorption in MOFs and similar materials relies on machine learning models. Recent reports have used this approach for high-throughput screening of MOFs for H<sub>2</sub> storage<sup>21-22</sup>, CH<sub>4</sub> adsorption capacity<sup>23</sup>, CO<sub>2</sub> capture<sup>24-25</sup>, N<sub>2</sub> capture<sup>24</sup>, and methanethiol/ethanethiol adsorption<sup>26</sup>. These models are promising, but their application has been limited to a single molecule or pair of molecules. In addition, [most of these examples](#) have correlated adsorption loading with a limited number of mostly geometrical descriptors of MOFs, [although there have been several studies that have included energy-based or chemical descriptors](#) which limits their prediction quality. Here we introduce a far richer set of descriptors that should be useful in many contexts and also tackle the task of making predictions for a large and diverse collection of adsorbing molecules.

Below we consider ~24,000 molecules relevant to the chemical industry that contains ~10<sup>7</sup> near-azeotropic binary mixtures<sup>27</sup>. If a set of ~10<sup>4</sup> sorbents is considered, the complete set of possible binary separations include >10<sup>11</sup> binary adsorbed mixtures. This large set of mixtures and adsorbents can be termed “adsorption space”<sup>28</sup>. There is no feasible way to fully explore adsorption space by direct experiments. We show below, however, that this task can be accomplished by combining detailed molecular simulations of adsorption with machine learning (ML) techniques.

Commented [CHBE1]: Add refs here

## Materials and Methods

We tackled the task of efficiently predicting adsorption isotherms for a wide range of molecules in metal-organic frameworks (MOFs), a diverse class of crystalline nanoporous adsorbents. Numerous studies have shown that molecular simulations can reliably predict adsorption of small molecules (e.g., CO<sub>2</sub>, CH<sub>4</sub>) in MOFs<sup>29</sup>. Tang *et al.* recently used Grand Canonical Monte Carlo (GCMC) simulations to calculate single-component adsorption isotherms of 24 molecules in 471 MOFs at 300 K<sup>28</sup>. For the vast majority of these isotherms, no experimental data of any kind is available. We used 9,615 isotherms from 460 MOFs from Tang *et al.* as the basis for this study (see section S.1.1 for more details).

To apply ML techniques to molecular adsorption, a rich set of descriptors associated with the adsorbates and adsorbents must be available. Molecular descriptors play a fundamental role in the development of structure-property relationships in chemistry<sup>30</sup>, pharmaceutical science<sup>31</sup>, environmental science<sup>32</sup>, and materials science<sup>33</sup>. To date, however, the range of descriptors that have been explored in studies of nanoporous crystalline materials such as MOFs has been limited. Considerable attention has been paid to macroscopic geometrical descriptors for these materials, including the pore-limiting diameter (PLD)<sup>34</sup>, largest cavity diameter (LCD)<sup>34</sup>, volumetric and gravimetric surface areas (VSA and GSA)<sup>34</sup>, pore volume ( $v_p$ )<sup>34</sup>, and void fraction ( $v_f$ )<sup>12, 35</sup>. It is unlikely, however, that these quantities include enough information to describe the full diversity of molecular adsorption that can exist. To address this shortcoming, we developed a set of thousands of descriptors of crystalline porous materials by adapting methods that have been previously developed for molecular systems (see S.1.2 for details). A key step in this process was adapting molecular descriptors to extended materials such that the results are

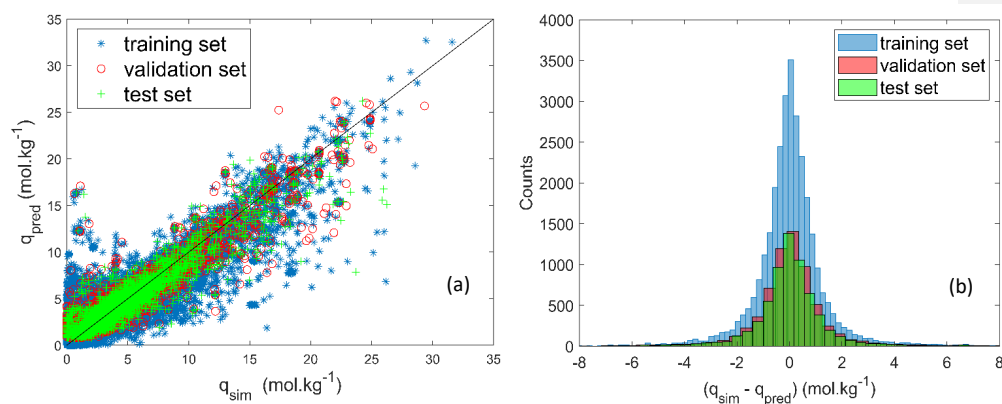
intensive with respect to periodic boundary conditions. Feature reduction methods were used to reduce this large number of descriptors to those that most strongly differentiate between MOFs, leading to a set of 25 adsorbent descriptors (see S.1.3).

Although very large numbers of molecular descriptors are available, we chose to use only a small set of descriptors for adsorbing molecules that are rooted in the law of corresponding states. Specifically, we used the critical temperature,  $T_c$ , the critical pressure  $P_c$ , and the acentric factor,  $\omega$ . The models of Gharagheizi *et al.*<sup>36</sup> were used to predict the  $T_c$ , the  $P_c$  and the  $\omega$  for each molecule. Using these models rather than experimental data has the advantage that our calculations can readily be extended to arbitrary molecules. A comparison between the predicted  $T_c$ ,  $P_c$  and  $\omega$  and the actual values reported is in Figure S1.

## Results and Discussion

The adsorption isotherms from Tang *et al.* define ~43,000 distinct data points with non-negative heats of adsorption. We used multiple regression genetic programming (MRGP)<sup>37</sup> to seek a model that uses the 28 descriptors listed in Table S1 to predict the adsorbed amount for each molecule (see section S.1.4.1 for details) after randomly splitting the underlying dataset into a training set, validation set, and test set. The resulting model, Eq. (S1), uses 14 descriptors (3 molecular descriptors, 11 MOF descriptors). This model, while simple to evaluate numerically, does not have a simple form that is identifiable in terms of “classic” isotherms derived from physical principles. The predictions of Eq. (S1) are compared to the underlying GCMC data in Figure 1. The quality of fit is very similar for the training, validation and test sets, giving a heuristic sense that the fitting is robust. Figure 2(b) shows the differences between the model predictions and GCMC data are approximately normally distributed with a standard deviation of

1.55 mol.kg<sup>-1</sup>. This difference is less than  $\pm 1$  mol.kg<sup>-1</sup> for 72 % of all data points and only 5 % of the data points show a deviation more than  $\pm 3$  mol.kg<sup>-1</sup>. Similar behavior can be seen in the histogram of percentage error (Figure S.12).



**Figs 1.** (a) A comparison between the model predictions of Eq. (S1),  $q_{\text{pred}}$ , and GCMC simulations,  $q_{\text{sim}}$ , for 9,615 adsorption isotherms in MOFs at room temperature from Tang et al. and (b) the distributions of deviations between model predictions and the GCMC simulations. The training set, validation set and test set used in developing Eq. (S1) are shown in separate colors.

It is important to compare our model with other means of efficiently predicting adsorption isotherms. Tang *et al.*<sup>28</sup> proposed a physically-motivated approximation to estimate adsorption isotherms in MOFs with significantly less computational effort than a direct GCMC simulation. Specifically, they assumed that each isotherm can be approximated by a Langmuir isotherm, an

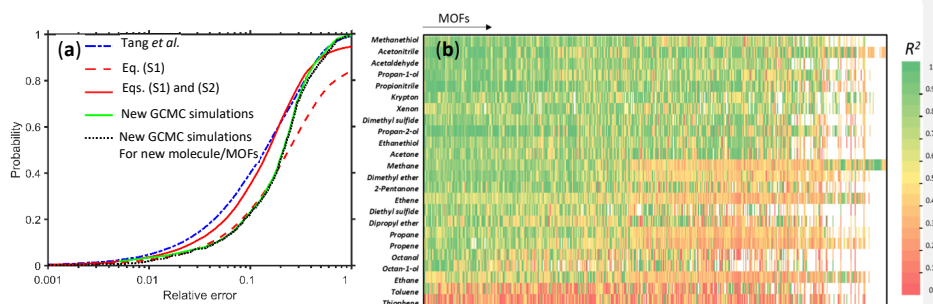
approach that only involves two parameters, the Henry's constant and the saturation loading. They then used a series of physical descriptors to estimate saturation loadings and a single GCMC simulation to directly measure the Henry's constant for each molecule/MOF pair of interest. This approach requires far more computational resources than Eq. (S1) because a molecular simulation must be performed for every adsorbate/adsorbent pair of interest.

Developing Eq. (S1) also made no a priori assumptions about the functional form of the isotherms. The predictions of Eq. (S1) are compared to the prior approach of Tang *et al.*<sup>28</sup> in Table 1 and Figure 2. In terms of  $R^2$  and  $RMSE$ , the performance of the model is better than the approach proposed by Tang *et al.*. Figure 2(a) shows a cumulative probability plot of the relative error of predictions from our ML-based model (dashed red line) and the previous model of Tang *et al.*<sup>28</sup> (blue line) compared to the original GCMC simulations. The probability of observing a relative error of <30% (an uncertainty range that is not uncommon in replicate experimental isotherm measurements of adsorption in MOFs<sup>38</sup>) using our simulation-free model and the model of Tang *et al.* (which uses one GCMC simulation per isotherm) is 0.59 and 0.75, respectively.

**Table 1-** A comparison between the model predictions and the corresponding GCMC simulations and also the estimations based on the physically-motivated model proposed by Tang *et al.*<sup>28</sup>. N is the number of independent adsorption state points considered.

Statistical parameter	training set	validation set	test set	overall	Tang <i>et al.</i> <sup>28</sup>
$R^2$	0.828	0.822	0.811	0.824	0.803

RMSE	1.536	1.594	1.557	1.548	1.752
N	30255	6482	6482	43219	43219



**Figure 2.** (a) Cumulative relative error associated with Eq. (S1) (dashed red line) compared with the original GCMC simulations of Tang *et al.* and the model of Tang *et al.*<sup>28</sup> (dashed blue line) compared with the original GCMC simulations of Tang *et al.*. The cumulative error associated with a model combining Eq. (S1) and Eq. (S2) is shown as a solid red curve. The solid green curve shows the cumulative error associated with comparing Eqs. (S1) and (S2) to the set of 1425 GCMC simulations described in the text that were not included in the work of Tang *et al.* or in the development of our ML-based model. The dotted black curve shows the cumulative error associated with comparing Eqs. (S1) and (S2) to the set of 1092 (out of 1425) GCMC simulations described in the text that neither their molecules nor their MOFs were included in the work of Tang *et al.* or in the development of our ML-based model. (b) A comparison between the predicted adsorption isotherms in 9,615 molecule-MOF pairs and the corresponding GCMC simulations. Each pixel represents a molecule/MOF system, color coded by  $R^2$ . White indicates a

system where no significant adsorption occurs in GCMC. Molecules in the figure are arranged vertically in order of decreasing average  $R^2$ , and MOFs are arranged from left to right in order of decreasing average  $R^2$ .

An interesting feature of Figure 2(a) is that for a small fraction of cases, Eq. (S1) performs very poorly. For example, 7% of the comparisons using Eq. (S1) show a relative error larger than 10, a level of error so large as to make the prediction lack even qualitative value. To explore the origins of these effects, Figure 2(b) compares all the 9,615 adsorption isotherms predicted by Eq. (S1) with their corresponding GCMC simulations. Figure 2(b) suggests there is some underlying structure to the ability of Eq. (S1) to make accurate predictions. For example, the predictions for thiophene and toluene, the only cyclic molecules in our data set, are systematically less accurate than the other molecules. The existence of systematic structure in the model performance can also be seen in Figs. S3-S6, which replot the data from Figure 2(a) with several alternative orderings. These figures hint that Eq. (S1) is less accurate for MOFs with high density (Figure S3), for MOFs with small PLDs (Figure S4) or for MOFs with low pore volumes (Figs. S5 and S6).

In order to improve our model's reliability, we sought an approach to predict the model's applicability domain (AD) without increasing numerical cost. In general terms, a model's AD is the space for which the model can make reliable predictions. Unfortunately, there is no widely-accepted mathematical algorithm for determining an AD<sup>39</sup>. To estimate the AD of Eq. (S1), we defined "reliable predictions" as the predictions within  $\pm 20\%$  of the corresponding GCMC simulation. This definition defines an integer variable that is 1 (0) for reliable (unreliable) predictions. We used MRGP to develop the classification model given by Eq. (S2) for this

variable (see S.1.4.2 for details). As with Eq. (S1), using Eq. (S2) requires no information other than the MOF and molecule descriptors; no GCMC or other molecular simulation data is needed.

Applying Eq. (S2) shows that Eq. (S1) is predicted to be reliable for 44% of the 43,219 individual state points and 33% of the 9,615 complete adsorption isotherms for which we have GCMC data. For these cases, the model predicts an adsorption loading within  $\pm 1 \text{ mol.kg}^{-1}$  71% of the time. Figure 2(a) shows a cumulative plot of relative error that indicates for cases classified as reliable by Eq. (S2), Eq. (S1) has a 76% probability of generating data with a relative error less than 30%. Figure 2(b) shows that Eqs. (S2) and (S1), which do not require any molecular simulation data, make predictions with comparable accuracy to the carefully tuned physical model of Tang *et al.* which requires at least one GCMC calculation for every adsorbate/adsorbent example.

The approach described above opens possibilities for screening adsorption-based separations on a scale that has not been previously possible. To demonstrate this concept, we considered a library of 54,000 organic molecules of interest to the chemical industry<sup>27</sup> and a set of >4,700 MOFs from the CoRE MOF database<sup>6</sup>. We restricted our attention to molecules with freezing point below 300 K (as predicted by the model of Gharagheizi *et al.*<sup>40</sup>). The resulting set of 23,923 molecules and their predicted bulk properties are given in Data S4 (see S.1.5 for more details). We computed the full set of 5,009 descriptors for 4,763 MOFs, even though only a subset of these descriptors is used in our model. This data is available in Data S5, and a list of 46 MOFs from the CoRE database that was excluded is given in Table S5. In all, this approach includes  $1.12 \times 10^8$  molecule/MOF systems. We applied Eqs. (S1) and (S2) to predict the adsorbed amount in each case at the vapor pressure of the adsorbing molecule at 300 K. Our

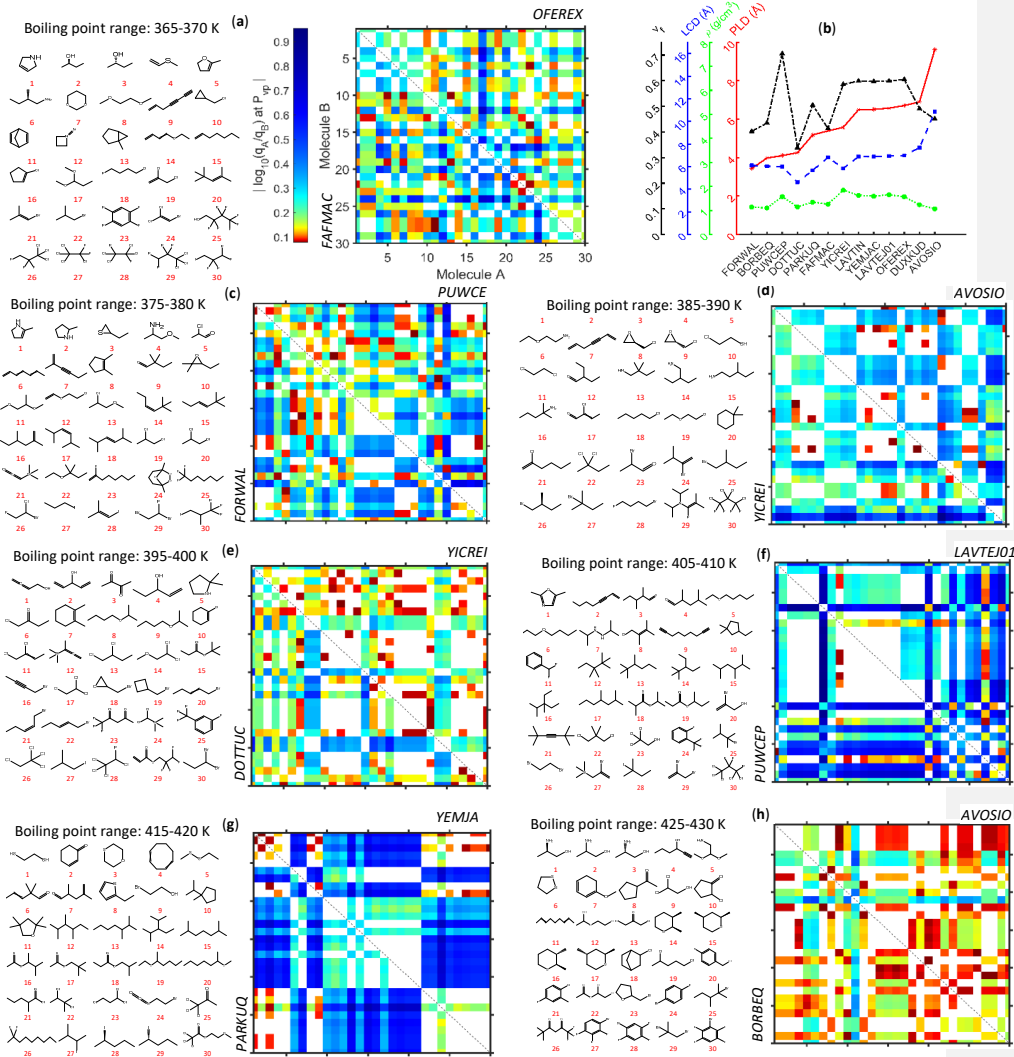
classification model predicts that Eq. (S1) is reliable in 8.46 % of these examples, that is, for  $9.54 \times 10^6$  distinct molecule/MOF systems.

It is important to evaluate our model's accuracy for examples unrelated to the data used to train and validate the model. To this end, we performed GCMC simulations for 475 molecule/MOF systems that included 37 molecules that were not included in the work of Tang *et al.* and 158 distinct MOFs, only 36 of which appeared in the work of Tang *et al.* (see S.1.6 for details). Figure 2(a) shows the relative error of the predictions from Eqs. (S1) and (S2) for these new GCMC simulations (solid green line). Our simulation-free model shows a relative error  $< 30\%$  73% of the time, compared to 76% of the time for the training data of Tang *et al.*. This is a powerful indication that our simulation-free approach can accurately predict adsorption isotherms for a diverse range of adsorbates and adsorbents.

Having established that our ML-based model gives useful predictions, we turned to finding specific MOFs to separate near-azeotropic binary mixtures of a diverse range of molecules. We define two molecules to be near-azeotropic if their boiling points differ by less than 5 K. Among the  $2.86 \times 10^8$  binary mixtures possible from the molecules listed in Data S4, there are  $1.02 \times 10^7$  near-azeotropic pairs. We analyzed  $4.8 \times 10^{10}$  near-azeotropic adsorbed mixtures/MOF combinations. In each case we calculated the single component adsorption uptake of each molecule at its bulk phase vapor pressure, then used the ratio of these uptakes as a proxy for the ability of a MOF to separate a molecular pair. We found  $6.2 \times 10^8$  examples for which our classification model (Eq. (S2)) predicts that Eq. (S1) is reliable and the predicted ratio of single component loadings exceeds 1.2. Details are given in Data S6. Of the  $1.02 \times 10^7$  near-azeotropic pairs, at least one MOF for which this loading ratio  $> 1.2$  was identified for  $8.8 \times 10^5$  pairs. Significant attention is paid to individual materials capable of separating individual near-

azeotropic pairs of commercial interest <sup>41-43</sup>. Our results greatly expand the range of examples that can be considered for these problems.

Specific examples of MOFs for separation of near-azeotropic pairs are shown in Figure 3. In each temperature range, a set of 30 molecules is shown, and the pairs for which the proxy quantity defined above is  $>1.2$  are illustrated for two specific MOFs. This proxy is clearly only approximate. More detailed information about mixture adsorption could be obtained for specific examples in a numerically efficient way by using the single component isotherms from Eq. (S1) in conjunction with Ideal Adsorbed Solution Theory <sup>44</sup>. More precise information could be found by performing mixture GCMC calculations for examples of special interest.



**Figure 3.** 14 examples of selecting MOFs for room temperature separation of near-azotropic binary mixtures. The format of (c)-(h) mimics (a), which highlights adsorption of the 30 molecules listed in two MOFs, FAFMAC and OFFEREX, which show adsorptive separation

using the metric defined in the text for 80% and 72% of the mixtures. (b) A summary of physical properties of the MOFs shown in the figure. In (c)-(h), the MOFs shown have adsorptive separations for (c) 72% (FORWAL) and 72% (PUWCEP), (d) 63% (YICREI) and 62% (AVOSIO), (e) 68% (DOTTUC) and 58% (YICREI), (f) 67% (PUWCEP) and 75% (LAVTEJ01), (g) 70% (PARKUQ) and 70% (YEMJAC), and (h) 59% (BORBEQ) and 63% (AVOSIO) of the mixtures defined by the 30 molecules shown for each temperature.

## Conclusion

We have introduced a highly numerically efficient model for single component adsorption of arbitrary molecules in MOFs at room temperature using ML methods. This model makes predictions at the same level of precision as detailed molecular simulations using generic force fields (FFs) in defect-free models of crystals that are assumed to be rigid during adsorption. Although there is considerable evidence that this level of precision yields practically useful information<sup>45</sup>, opportunities clearly exist to extend our ideas to situations where generic FFs are likely to be inadequate (e.g. open metal sites and/or materials that undergo significant adsorption-induced deformation). There is also likely to be considerable scope for improving upon the set of descriptors we have introduced for MOFs; we speculate that better descriptors exist that would improve the quality of both our adsorption isotherm model and classification model. Despite these caveats, our work demonstrates an exciting capability that extends the ability to make predictions about adsorption-based separations far beyond the scope of previous activities in this important area.

## Associated Content

The supporting information is available free of charge on the ACS Publications website.

Materials and methods, Figures S1 to S10, Tables S1 to S6, numerical values of descriptors for 460 MOFs, the descriptors suggested to MOFs, the ML model predictions, the 23,923 molecules and their predicted properties, list of 875,591 near-azeotropic pairs, new GCMC simulations, and Modified RASPA source code.

#### Author Information

\*E-mail: [david.sholl@chbe.gatech.edu](mailto:david.sholl@chbe.gatech.edu)

#### ORCID

Farhad Gharagheizi: [0000-0003-3922-5867](https://orcid.org/0000-0003-3922-5867)

Dai Tang: [0000-0002-7850-8694](https://orcid.org/0000-0002-7850-8694)

David S. Sholl: [0000-0002-2771-9168](https://orcid.org/0000-0002-2771-9168)

#### Acknowledgments

FG and DSS received funding from the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy (EERE) under the Advanced Manufacturing Office Award Number DE-EE0007888. DT and DSS received funding from the Nanoporous Materials Genome Center, funded by the U.S. Department of Energy, Office of Science, Basic Energy Sciences, under Award #DEFG02-17ER16362.

### Authors contributions

All authors conceived the study and participated in writing the manuscript. FG led development of material descriptors and machine-learning methods. DT performed the molecular simulations.

### Competing interests

The authors declare no competing financial interests.

### References

1. Sholl, D. S.; Lively, R. P., Seven Chemical Separations to Change the World. *Nature* **2016**, *532*, 435-437.
2. Bae, Y. S.; Snurr, R. Q., Development and Evaluation of Porous Materials for Carbon Dioxide Separation and Capture. *Angew Chem Int Ed Engl* **2011**, *50*, 11586-11596.
3. Keskin, S.; van Heest, T. M.; Sholl, D. S., Can Metal-Organic Framework Materials Play a Useful Role in Large-Scale Carbon Dioxide Separations? *ChemSusChem* **2010**, *3*, 879-891.
4. Siderius, D. W.; Shen, V. K.; Johnson III, R. D.; van Zee, R. D., *Nist/Arpa-E Database of Novel and Emerging Adsorbent Materials*. Eds. ed.; National Institute of Standards and Technology: Gaithersburg MD, 20899, 2018.
5. Wu, C.-W.; Sircar, S., Comments on Binary and Ternary Gas Adsorption Selectivity. *Separation and Purification Technology* **2016**, *170*, 453-461.
6. 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. *Chemistry of Materials* **2014**, *26*, 6185-6192.

7. Moghadam, P. Z.; Li, A.; Wiggin, S. B.; Tao, A.; Maloney, A. G. P.; Wood, P. A.; Ward, S. C.; Fairen-Jimenez, D., Development of a Cambridge Structural Database Subset: A Collection of Metal–Organic Frameworks for Past, Present, and Future. *Chemistry of Materials* **2017**, *29*, 2618-2625.
8. Visini, R.; Awale, M.; Reymond, J. L., Fragment Database Fdb-17. *J. Chem. Inf. Model.* **2017**, *57*, 700-709.
9. Zhou, H. C.; Long, J. R.; Yaghi, O. M., Introduction to Metal-Organic Frameworks. *Chem Rev* **2012**, *112*, 673-674.
10. Odoh, S. O.; Cramer, C. J.; Truhlar, D. G.; Gagliardi, L., Quantum-Chemical Characterization of the Properties and Reactivities of Metal-Organic Frameworks. *Chem Rev* **2015**, *115*, 6051-6111.
11. Evans, J. D.; Fraux, G.; Gaillac, R.; Kohen, D.; Trousselet, F.; Vanson, J.-M.; Coudert, F.-X., Computational Chemistry Methods for Nanoporous Materials. *Chemistry of Materials* **2016**, *29*, 199-212.
12. Dubbeldam, D.; Calero, S.; Ellis, D. E.; Snurr, R. Q., Raspa: Molecular Simulation Software for Adsorption and Diffusion in Flexible Nanoporous Materials. *Molecular Simulation* **2016**, *42*, 81-101.
13. Borboudakis, G.; Stergiannakos, T.; Frysali, M.; Klontzas, E.; Tsamardinos, I.; Froudakis, G. E., Chemically Intuited, Large-Scale Screening of Mofs by Machine Learning Techniques. *npj Computational Materials* **2017**, *3*, 1-7.
14. Li, S.; Chung, Y. G.; Simon, C. M.; Snurr, R. Q., High-Throughput Computational Screening of Multivariate Metal-Organic Frameworks (Mtv-Mofs) for Co<sub>2</sub> Capture. *J Phys Chem Lett* **2017**, *8*, 6135-6141.

15. Azar, A. N. V.; Velioglu, S.; Keskin, S., Large-Scale Computational Screening of Metal Organic Framework (Mof) Membranes and Mof-Based Polymer Membranes for H<sub>2</sub>/N<sub>2</sub> Separations. *ACS Sustain Chem Eng* **2019**, *7*, 9525-9536.
16. Avci, G.; Velioglu, S.; Keskin, S., High-Throughput Screening of Mof Adsorbents and Membranes for H<sub>2</sub> Purification and Co<sub>2</sub> Capture. *ACS Appl Mater Interfaces* **2018**, *10*, 33693-33706.
17. Altintas, C.; Avci, G.; Daglar, H.; Gulcay, E.; Erucar, I.; Keskin, S., Computer Simulations of 4240 Mof Membranes for H<sub>2</sub>/Ch<sub>4</sub> Separations: Insights into Structure-Performance Relations. *J Mater Chem A Mater* **2018**, *6*, 5836-5847.
18. Peng, L.; Zhu, Q.; Wu, P.; Wu, X.; Cai, W., High-Throughput Computational Screening of Metal-Organic Frameworks with Topological Diversity for Hexane Isomer Separations. *Phys Chem Chem Phys* **2019**, *21*, 8508-8516.
19. Moghadam, P. Z.; Fairen-Jimenez, D.; Snurr, R. Q., Efficient Identification of Hydrophobic Mofs: Application in the Capture of Toxic Industrial Chemicals. *Journal of Materials Chemistry A* **2016**, *4*, 529-536.
20. Matito-Martos, I.; Moghadam, P. Z.; Li, A.; Colombo, V.; Navarro, J. A. R.; Calero, S.; Fairen-Jimenez, D., Discovery of an Optimal Porous Crystalline Material for the Capture of Chemical Warfare Agents. *Chemistry of Materials* **2018**, *30*, 4571-4579.
21. Gopalan, A.; Bucior, B. J.; Bobbitt, N. S.; Snurr, R. Q., Prediction of Hydrogen Adsorption in Nanoporous Materials from the Energy Distribution of Adsorption Sites. *Molecular Physics* **2019**, 1-12.

22. Bobbitt, N. S.; Snurr, R. Q., Molecular Modelling and Machine Learning for High-Throughput Screening of Metal-Organic Frameworks for Hydrogen Storage. *Molecular Simulation* **2019**, *45*, 1069-1081.
23. Fanourgakis, G. S.; Gkagkas, K.; Tylanakis, E.; Klontzas, E.; Froudakis, G., A Robust Machine Learning Algorithm for the Prediction of Methane Adsorption in Nanoporous Materials. *J Phys Chem A* **2019**, *123*, 6080-6087.
24. Fernandez, M.; Barnard, A. S., Geometrical Properties Can Predict Co<sub>2</sub> and N<sub>2</sub> Adsorption Performance of Metal-Organic Frameworks (Mofs) at Low Pressure. *ACS Comb Sci* **2016**, *18*, 243-252.
25. Fernandez, M.; Boyd, P. G.; Daff, T. D.; Aghaji, M. Z.; Woo, T. K., Rapid and Accurate Machine Learning Recognition of High Performing Metal Organic Frameworks for Co<sub>2</sub> Capture. *J Phys Chem Lett* **2014**, *5*, 3056-3060.
26. Liang, H.; Yang, W.; Peng, F.; Liu, Z.; Liu, J.; Qiao, Z., Combining Large-Scale Screening and Machine Learning to Predict the Metal-Organic Frameworks for Organosulfurs Removal from High-Sour Natural Gas. *APL Materials* **2019**, *7*.
27. Yaws, C. L., *The Yaws Handbook of Physical Properties for Hydrocarbons and Chemicals : Physical Properties for More Than 54,000 Organic and Inorganic Chemical Compounds, Coverage for C1 to C100 Organics and Ac to Zr Inorganics*, Second edition. ed.; Elsevier: Oxford, UK, 2015.
28. Tang, D.; Wu, Y.; Verploegh, R. J.; Sholl, D. S., Efficiently Exploring Adsorption Space to Identify Privileged Adsorbents for Chemical Separations of a Diverse Set of Molecules. *ChemSusChem* **2018**, *11*, 1567-1575.

29. Colon, Y. J.; Snurr, R. Q., High-Throughput Computational Screening of Metal-Organic Frameworks. *Chem Soc Rev* **2014**, *43*, 5735-5749.
30. Katritzky, A. R.; Kuanar, M.; Slavov, S.; Hall, C. D.; Karelson, M.; Kahn, I.; Dobchev, D. A., Quantitative Correlation of Physical and Chemical Properties with Chemical Structure: Utility for Prediction. *Chem Rev* **2010**, *110*, 5714-5789.
31. Cherkasov, A., et al., Qsar Modeling: Where Have You Been? Where Are You Going To? *J Med Chem* **2014**, *57*, 4977-5010.
32. Mamy, L.; Patureau, D.; Barriuso, E.; Bedos, C.; Bessac, F.; Louchart, X.; Martin-Laurent, F.; Miege, C.; Benoit, P., Prediction of the Fate of Organic Compounds in the Environment from Their Molecular Properties: A Review. *Crit Rev Environ Sci Technol* **2015**, *45*, 1277-1377.
33. Puzyn, T.; Rasulev, B.; Gajewicz, A.; Hu, X.; Dasari, T. P.; Michalkova, A.; Hwang, H. M.; Toropov, A.; Leszczynska, D.; Leszczynski, J., Using Nano-Qsar to Predict the Cytotoxicity of Metal Oxide Nanoparticles. *Nat Nanotechnol* **2011**, *6*, 175-178.
34. 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. *Microporous and Mesoporous Materials* **2012**, *149*, 134-141.
35. Torres-Knoop, A.; Balaji, S. P.; Vlugt, T. J.; Dubbeldam, D., A Comparison of Advanced Monte Carlo Methods for Open Systems: Cfcmc Vs Cbmc. *J Chem Theory Comput* **2014**, *10*, 942-952.
36. Gharagheizi, F.; Eslamimanesh, A.; Mohammadi, A. H.; Richon, D., Determination of Critical Properties and Acentric Factors of Pure Compounds Using the Artificial Neural

- Network Group Contribution Algorithm. *Journal of Chemical & Engineering Data* **2011**, *56*, 2460-2476.
37. Veeramachaneni, K.; Arnaldo, I.; Derby, O.; O'Reilly, U.-M., Flexgp. *Journal of Grid Computing* **2014**, *13*, 391-407.
38. Park, J.; Howe, J. D.; Sholl, D. S., How Reproducible Are Isotherm Measurements in Metal-Organic Frameworks? *Chemistry of Materials* **2017**, *29*, 10487-10495.
39. Netzeva, T., et al., *Current Status of Methods for Defining the Applicability Domain of (Quantitative) Structure-Activity Relationships-the Report and Recommendations of Ecvam Workshop 52*, 2005; Vol. 33, p 155-73.
40. Gharagheizi, F.; Ilani-Kashkouli, P.; Kamari, A.; Mohammadi, A. H.; Ramjugernath, D., A Group Contribution Model for the Prediction of the Freezing Point of Organic Compounds. *Fluid Phase Equilibria* **2014**, *382*, 21-30.
41. Li, L.; Lin, R. B.; Krishna, R.; Li, H.; Xiang, S.; Wu, H.; Li, J.; Zhou, W.; Chen, B., Ethane/Ethylene Separation in a Metal-Organic Framework with Iron-Peroxo Sites. *Science* **2018**, *362*, 443-446.
42. Cadiau, A.; Adil, K.; Bhatt, P. M.; Belmabkhout, Y.; Eddaoudi, M., A Metal-Organic Framework-Based Splitter for Separating Propylene from Propane. *Science* **2016**, *353*, 137-140.
43. Koh, D. Y.; McCool, B. A.; Deckman, H. W.; Lively, R. P., Reverse Osmosis Molecular Differentiation of Organic Liquids Using Carbon Molecular Sieve Membranes. *Science* **2016**, *353*, 804-807.
44. Walton, K. S.; Sholl, D. S., Predicting Multicomponent Adsorption: 50 Years of the Ideal Adsorbed Solution Theory. *AIChE Journal* **2015**, *61*, 2757-2762.

45. Agrawal, M.; Sholl, D. S., Effects of Intrinsic Flexibility on Adsorption Properties of Metal-Organic Frameworks at Dilute and Nondilute Loadings. *ACS Appl Mater Interfaces* **2019**, *11*, 31060-31068.