S.T. Yau High School Science Award

Research Report

The Team

Name of team member: Aiden Chen

School: Old Scona

City, Country: Edmonton, Canada

Name of team member: Xi Nan Jiao

School: Old Scona

City, Country: Edmonton, Canada

Name of supervising teacher: James Sochan

Job Title: School Teacher

School/Institution: Old Scona

City, Country: Edmonton, Canada

Title of Research Report

Computational Screening and Design of Metal-Organic Frameworks for CO2 Separation from Flue Gas

Date

August 22, 2024

Computational Screening and Design of Metal-Organic Frameworks for CO_2 Separation from Flue Gas

Xi Nan Jiao[1], Aiden Chen[1]

Abstract

Rising CO_2 levels, largely from flue gas emissions, are a significant contributor to global climate change. Adsorption using Metal-Organic Frameworks (MOFs) offers a promising solution for CO_2 capture due to their high surface area, tunable porosity, and selectivity. To streamline the discovery of efficient MOFs, we developed a high-throughput virtual screening (HTVS) pipeline by integrating Grand-Canonical Monte Carlo (GCMC) simulations, molecular modeling and machine learning. We screened a filtered subset of the CoREMOF database and a user-defined hypothetical MOF bank to identify candidates with high CO_2 adsorption capacity and selectivity. This approach yielded several high-performing MOFs, including five from the CoREMOF dataset and a new structure from the hypothetical bank exceeding the performance of existing MOFs. Our findings highlight the complex relationship between MOF geometries and CO_2 capture performance, emphasizing the importance of features like open metal sites and pore geometry. This computational framework accelerates MOF discovery and provides valuable insights for experimental synthesis. Future work will focus on expanding the hypothetical MOF dataset, improving simulation accuracy, and employing advanced optimization techniques to enhance the screening process.

Keywords: Metal-organic frameworks, CO_2 capture, molecule simulations, machine learning, multi-fidelity virtual screening

[1] These authors contributed equally to this work.

Acknowledgement

We are indebted to our instructor, James Sochan, who has provided valuable guidance and help in completing the project.

Declaration of Academic Integrity

The participating team declares that the paper submitted is comprised of original research and results obtained under the guidance of the instructor. To the team's best knowledge, the paper does not contain research results, published or not, from a person who is not a team member, except for the content listed in the references and the acknowledgment. If there is any misinformation, we are willing to take all the related responsibilities.

Names of team members XiNan Jiao Aiden Chen Xi Nan Jiao, Aiden Chen

Signatures of team members

Name of the instructor James Sochan

Signature of the instructor James Sochan James Sochan

Date Aug 11, 2024

Table of Contents

1 Introduction

Flue gas is a major contributor to the rising CO_2 level in today's atmosphere, consisting of approximately 80% of the CO_2 emission each year[1, 2].These high CO_2 levels trap heat from the sun, forcing temperatures to rise on a global scale. To mitigate this new phenomenon, one strategy is to remove CO_2 from the atmosphere, allowing more heat to radiate into space. Several strategies have been considered to remove CO_2 from the atmosphere, some being reducing energy needs overall, increasing efficiency of energy use and production, transitioning to low or non-carbon energy sources, or capturing carbon from the atmosphere[3, 4]. While reducing energy needs, increasing efficiency, and transitioning to green energy sources are not realistic, both in terms of economy and timescale, capturing carbon from the atmosphere offers a both cheap and immediate solution. Furthermore, the captured carbon can be repurposed as feedstock for industrial chemical production[5].

There are a wide range of CO_2 capture techniques, including absorption, adsorption, membrane, biological capture, and cryogenic separation[6]. They have been exploited in different technological pathways, such as post combustion capture, pre-combustion capture, ambient air capture, and the relatively new oxy-combustion and chemical looping combustion (CLC) processes[6]. The CO_2 concentrations from different processes vary drastically and thus require different techniques to capture the CO_2. For example, the flue gas resulting from pre-combustion contains high partial pressure of CO_2 and is suitable for chemical absorption using liquid amine-based approaches, such as the commercially available amine scrubbing[7]. In contrast, the flue gas from post-combustion of fossil fuels has low partial pressure of CO_2 that ranges from 0.03 ~ 0.19 atm, around 10-fold lower than that in the flue gas from pre-combustion process or oxy-combustion, making it suitable for adsorption-based techniques[8]. This low partial pressure usually results in low efficiency in capture and high cost from widespread deployment. Therefore, there is an urgent need to develop cost-effective materials that have high CO_2 capacity, selectivity and are easily deployable in different plants.

Adsorption is widely regarded as a promising technology for CO_2 capture as it can be retrofitted to any power plant, operated under various conditions with relatively high adsorption capacity, high CO_2 selectivity, and low energy requirement for regeneration[9]. Metal-organic frameworks are a relatively new class of materials that can be used in adsorption technologies[4,10]. They are

synthesized by self-assembly of metal nodes and organic linkers[11,12]. Due to its reticular structure, MOFs have exceptionally high surface area and pore volume, and by varying the nodes and linkers, nearly infinite variations of MOFs can be produced[13]. Therefore, MOFs can readily be tuned to maximize CO_2 capacity, selectivity, and efficiency, which makes them an attractive option for CO_2 capture. However, the enormous number of possible MOFs makes experimental synthesis and characterization impractical[14]. Therefore, using computers to model the interactions and adsorption between MOFs and the gasses is less resource demanding, allowing the screening of a large amount of MOFs in a relatively short amount of time.

In this regard, a wide range of modeling protocols can be applied, including density-functional theory (DFT) calculations[15], classical molecular dynamics (MD) simulations[16], and classical Grand-Canonical Monte-Carlo (GCMC) simulations[17]. DFT calculation is deemed to be the most accurate but comes with an extremely high computational burden. In a context of high throughput virtual screening (HTVS), classical simulations can be a particularly useful tool for obtaining the CO_2 capacity and selectivity of MOF banks containing thousands of MOFs[18]. In addition, with the development of large MOF databases (e.g., CoREMOF, hMOF), diverse machine learning (ML) techniques relying on crystal graph featurization of MOFs have been prevalent for data-driven materials design and larger materials space exploration. Successful examples include NU-800 for methane storage[19] and SBMOF-1[20] for Xe/Kr separation from HTVS pipelines.

Here, we develop an HTVS pipeline to screen MOFs for CO_2 separation from flue gas mixture by integrating GCMC simulations, surrogate ML model development, and user-defined MOF databank generation and screening. We first filter the CoREMOF database and acquire a subset of experimental MOFs for GCMC simulations to construct the labeled dataset constituting CO_2 adsorption capacity and selectivity. The labeled database is then utilized to finetune a pre-trained MOFTransformer model as the surrogate model for subsequent prediction tasks. Finally, to explore the hypothetical MOFs (i.e., generated *in silico* but not experimentally available), we generate a user-defined MOF bank and apply the surrogate model to preliminarily rank the putative MOF candidates. The selected ones are subjected to a final round of higher-fidelity GCMC simulations and proposed as the novel MOFs with superior performances. The computational workflow can be readily repurposed for different gas separation applications

using MOF materials and is anticipated to significantly facilitate the MOF materials discovery and design. From the computational workflow, we have successfully identified five MOFs that possess high adsorption capacity or selectivity from the filtered CoREMOF dataset. Moving forward towards a larger user-generated MOF database, one new MOF with higher performances has been identified, leading to a total of six potential MOFs for experimental synthesis and characterization for CO_2 capture, and hopefully making contributions to CO_2 reduction in real life.

2 Methods

2.1 Curation and Selection of CoREMOF Database

For the initial MOF database, we used the most recent CoREMOF 2019 database[21] taken from the MOFXdb website[58], which contains 12020 clean MOF structures. Before running GCMC simulations on the CoREMOF database, we need to filter out the MOFs based on their geometric properties and prior knowledge. In addition, to reduce the computational burden, we applied more stringent conditions. MOFs with a pore limiting diameter (PLD) smaller than 3.8 Å, a largest cavity diameter (LCD) larger than 10 Å, a unit cell with over 300 atoms, and a gravimetric surface area (GSA) less than 1000, are removed from subsequent calculations. The rationale is that the PLD should be large enough to accommodate both N_2 (D = 3.6 Å) and CO_2 (D = 3.3 Å) molecules for the calculation of selectivity, and that high LCD in MOFs loses high selectivity[22, 23]. Also, MOFs with high surface areas are expected to possess higher gas storage capacities. Finally, in total we narrowed the search space down to 1989 which is acceptable in the context of our computational resources. One thing to note is that, in the filtration process, the geometric properties were obtained directly from the MOFXdb database, and some of them are erroneous, and thus we used Zeo++[24], to recalculate the geometric properties using a N_2 probe size of 1.86 Å, in accordance with the diameter of N_2.

2.2 GCMC Simulations for CO_2 Adsorption Capacity and Selectivity

Grand Canonical Monte Carlo (GCMC) simulations can accurately capture the gas uptakes in porous materials. In our case, we implemented the GCMC simulations in RASPA2 simulation codes[25] to evaluate the CO_2 adsorption in a binary gas mixture of CO_2 and N_2 in a 0.15 to 0.85 ratio, mimicking the general flue gas composition from post-combustion of fossil fuels[26]. The

adsorption pressure and temperature were set to 1.0 bar and 298 K, respectively. The MOF frameworks were treated as rigid and the pair potentials were described by a Lennard-Jones (LJ) potential using the UFF force field[27] as it can be applied to a wide range of elements. The charges of MOFs were obtained using the Qeq method[28, 29] to describe the electrostatic interactions. The TraPPE force field[30, 31] was used for small molecules, which is suitable for studying phase equilibria. The cutoff distances for the interactions are 13 Å for both gas molecules and the MOF. To evaluate the performance of a MOF for CO_2 capture from flue gas, we used the adsorption capacity for CO_2 (N) and the selectivity for CO_2 over N_2 (S) as the metrics[32],

$$S = \frac{x_{CO_2}/x_{N_2}}{y_{CO_2}/y_{N_2}},$$

where x denotes the concentration of a gas in the MOF and y indicates the concentration of a gas in the environment. For a MOF to be optimal, one would expect high storage capacity and high selectivity.

2.3 Fine-tuning the MOFTransformer Model

Once we obtain the labeled data, we can develop a surrogate model for the prediction tasks of CO_2 adsorption capacity and selectivity of an unknown MOF. Preliminary results show that using geometric features failed to perform well due to the poor correlations. Thus, we employed more complicated featurization of MOFs and fine-tuned a pre-trained deep learning model to achieve this purpose. Crystal graphs have been employed for MOF embedding and have achieved good performances in predictions of diverse MOF properties, including adsorption, diffusion, outperforming other deep-learning models[33].

Among the deep learning models, the MOFTransformer emerges as the state-of-art model. It is a multi-model trained on 1 million hypothetical MOFs using an integrated atom-based graph and energy-grid embeddings[34]. Crystal graph data include nodes for MOF atoms and edges for connections between atoms, with the nodes and edges being characterized by vectors that correspond to the atoms and bonds in the crystal to capture the local chemical feature of MOFs[33]. Energy grid representation is very similar to a potential energy surface, which creates a 3D energy image to represent the global features of MOFs. Under the hypothesis that the MOFTranformer model has extracted a meaningful latent representation of the key structural

MOF properties, it can thus be finetuned for subsequent tasks. This transfer learning technique can be enormously beneficial when the labeled dataset is small[35].

We did the finetuning by converting all the MOFs into graph data and energy grid data, and the maximum super cell length was increased to 160 Å. The data was subdivided into 3 sets for a 0.8:0.1:0.1 train-validation-test split. The use of a validation partition was necessary to assist in monitoring the training process and avoiding overfitting due to our small data size. During training, the max epoch number was set to 50 and the batch size was 36. Two models were separately trained for CO_2 capacity and selectivity using the same data. The trained model was then used as a preliminary screening tool for a large MOF dataset.

2.4 MOF Database Generation and Optimization

A prerequisite for large-scale screening is the access to a large number of MOFs. Over the decades, several databases of MOFs have been developed for experimental MOFs (CoREMOF[21], CSD[36], etc.). Even with the large number of MOFs in these databases, especially CoREMOF, the MOFs exhibit poor topological diversity, and thus there is potentially a huge, unexplored MOF space that may open up a new avenue for MOF materials. In this context, computational screening offers a feasible way to efficiently explore MOFs in a short amount of time.

To build our own MOF dataset for CO_2 capture, the ToBaCCo codes[37-39] were used by combining different nodes and linkers to fit in a topology net to construct MOFs[40]. For the metal clusters, we decomposed the MOFs using MOFID[41] and found the three most abundant metals in the top-tier MOFs. The linkers were taken from the ToBaCCo codes excluding ones containing halogen, sulfur, and porphyrin moieties. In addition, we designed a set of NH_2 rich linkers based on the prior knowledge that NH_2 in MOFs may increase the CO_2 capture[42, 43]. The generated MOFs were pre-filtered so that the atoms in the unit cell do not exceed 400 atoms, based on the fact that large MOFs are not experimentally feasible to synthesize. In total, we obtained 5699 MOF structures, 3 times larger than the calculated dataset.

Since the ToBaCCo code generates MOFs purely relying on the geometry matching, the resulting structures may include severe atomic overlap or structure strains, which will lead to inaccuracy of the MOF geometric property calculations. Thus, we used molecular dynamics (MD) simulations to relax the unit cells of the generated MOFs. MD simulation analyzes physical

movements of atoms through simulating their interactions and dynamic evolution of the system. In our work, MD simulations were conducted using the Large-scale Atomic/Molecular Massively Parallel Simulator (LAMMPS)[44] developed by Sandia National Laboratory. The cif2lammps[45] tool was used to generate the input files for LAMMPS with implementation of the UFF4MOF[46] force field containing corrected parameters for copper-paddle wheel metal nodes to represent the interatomic interaction. For structure energy minimization, a multi-stage scheme was performed. First, the box size and atom coordinates were minimized simultaneously. Secondly, the atomic coordinates were relaxed by fixing the box. Then a short NPT equilibration was conducted followed by an additional round of energy minimization. The final structures were then converted back to CIF format to give the final hypothetical MOF dataset using the atomic simulation environment (ASE)[47].

3 Results

3.1 Identification of Top MOFs from CoREMOF Dataset

From the GCMC calculations on the selected CoREMOF dataset, the CO_2 adsorption capacity and selectivity adsorption in a simulated flue gas show a wide span among the 1411 MOF structures (excluding the pure inorganic solids). The adsorption capacity ranges from 0.0004 to 21.925 mol/kg and the selectivity ranges from 0.0013 to 707092.54 (**Figure 1a**). In addition, we observe that over 50% of the MOFs are located in the lower performance region, with selectivity lower than 100 and capacity lower than 5 kg/mol, indicating the challenges for finding the MOFs with exceptional performances in both capacity and selectivity. Even so, we are able to identify exceptional MOFs that are potential for experimental testing. To define the overall performance of MOFs, we took both the capacity and selectivity into consideration by using the Pareto front to indicate the best MOFs from the calculations (**Figure 1a**). The candidates possess either high selectivity, high capacity or balance between the two, and are shown in **Figure 1b**.

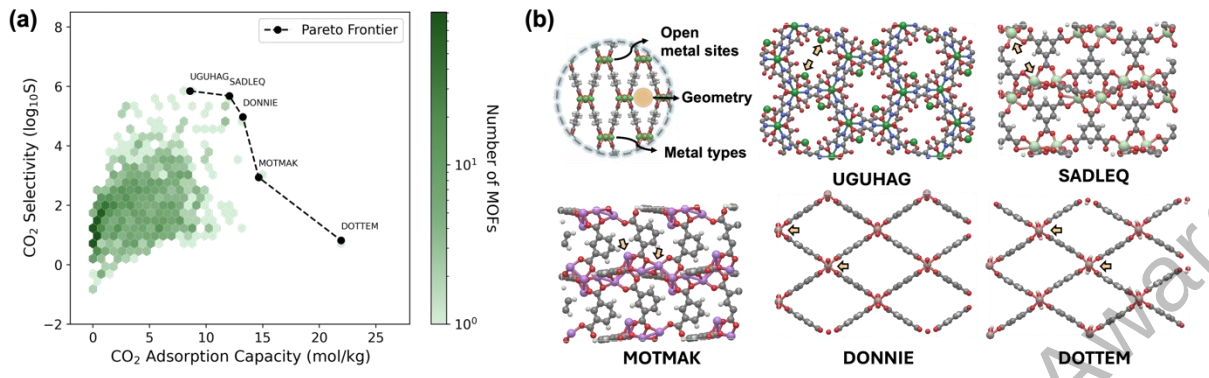

Figure 1. *Identification of the top performing MOFs for CO_2 separation. (a) Hexbin plot of the CO_2 adsorption capacity and the log10 selectivity of the MOFs in the CoREMOF dataset. (b) The visualization of the structures of top 5 MOFs on the pareto front. The structures are visualized by Avogadro software. The orange arrows indicate the open metal sites within the MOF structure.*

All five MOF candidates show characteristics of rich open metal sites (non-fully coordinated metal atoms) in the structures as indicated by the orange arrows, suggesting that creating open metal sites can potentially increase the selective CO_2 adsorption capacity[48]. This is in accordance with the previous reports that open metal sites can offer additional interaction sites for CO_2 to adsorb into the MOFs[49,50]. In the DONNIE and DOTTEM MOFs, it is found that the metals in these MOFs, Al and Ga, are shown to have strong interactions with CO_2[51]. In addition, all these MOFs represent small pores and channels that help accommodate CO_2 molecules. Overall, the results highlight the importance of metal types, geometric properties, and existence of open metal sites, in designing MOFs with high CO_2 capture performance in a simulated flue gas.

3.2 Can Simple Geometric Properties Indicate MOF Performance?

Following the creation of the Pareto front, we want to determine whether CO_2 adsorption capacity and selectivity can be inferred from simple geometric descriptors, such as surface area or pore diameters. If there is a strong correlation, then we can use these properties to predict the performance of a MOF without running the resource intensive simulations, which would allow us to conduct large throughput screenings. In total, we considered GSA, GAV, inverse density, LCD, PLD, and VF and computed their Pearson correlation coefficients (R) for each. However, we found a weak correlation between the CO_2 adsorption capacity and MOF geometry including GSA and GAV , while no correlation was found for selectivity with R values of around 0 for all (**Figure 2a**).

It is observed that MOFs with high GAV tend to have poor capacity and selectivity. The low correlation of the GAV with the adsorption capacity is counterintuitive, as one would expect that a large volume would allow for a higher adsorption capacity. This may happen because despite the large volume, the attraction is not strong enough to adsorb a sizable amount of CO_2 molecules. A similar explanation can also be applied to the GSA, which also has a low correlation with adsorption capacity. Combining the results with those from Section 3.1, we conclude that the CO_2 adsorption behavior within MOFs is a complex problem involving factors such as metal types, balance between geometric properties, and existence of open metal sites, and thus cannot be inferred by simple geometric properties.

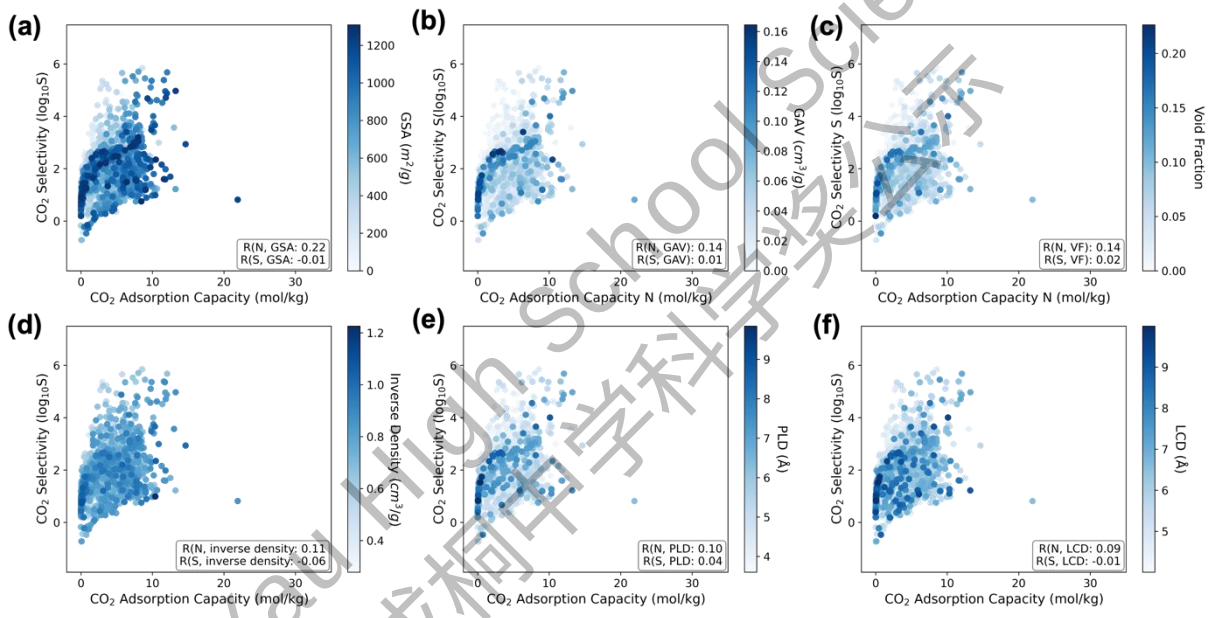

Figure 2. *Structure property analysis of the CO_2 capacity and selectivity. The color bars show the values of GSA (a), GAV (b), VF (c), inverse density (d), PLD (e), and LCD (e). The insets show the Pearson correlation coefficient R of each geometric property.*

3.3 Multi-Scale Screening of a Hypothetical MOF Dataset

Having successfully identified top MOFs from the selected CoREMOF database, we then set off to construct a user-defined hypothetical MOF dataset and to perform a multi-scale screening to find new high-performance MOFs. Within the multi-scale screening, we first develop a surrogate model built on MOFTransformer, then select a subset of MOFs with potential high performance based on the prediction, and finally propose new structures by an additional round of GCMC calculation.

3.3.1 Putative MOF Databank Generation and Optimization

To generate our own hypothetical MOF dataset, we employed the ToBaCCo codes, which is suitable for large-scale MOF generation. As indicated in Section 3.1, high-performance MOFs are usually rich in open metal sites and include specific metals. However, it is not feasible to create MOFs with all possible metal types when we generate the library, and thus we performed a metal abundance analysis in the good MOFs with capacity over 5 mol/kg and selectivity over 10. As shown in **Figure 3**, the most abundant metals are Mn, Cu, and Eu in these MOF, and thus all the metal clusters in ToBaCCo containing such metals are used for MOF generation. In addition, to increase the possibility of finding high-performance MOFs, we included a set of NH_2 rich linkers, which have been reported to increase CO_2 adsorption. All the metal nodes used in MOF generation and the general workflow are shown in **Figure 4a**. Among all the metal nodes, the 8c_Mn, 12c_Mn, 3c_Cu and 4c_Cu possess open metal sites, and may lead to superior MOFs for CO_2 capture from flue gas.

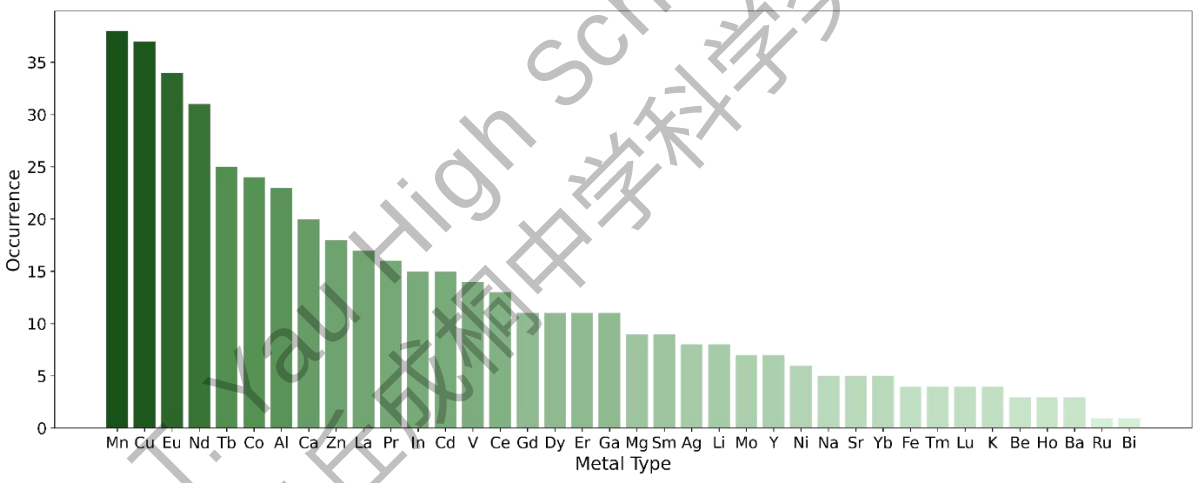

Figure 3. *Metal abundance analysis of the MOFs with high performance within the CoREMOF. The color scale indicates the occurrence of the metals in the subset.*

ToBaCCo purely relies on templates to assemble the MOFs, and thus includes unreasonable structures such as atomic overlaps and stretch bonds, which may negatively impact the embedding of MOFs in the following machine learning models. Thus, the structures need to be optimized using a short MD simulation in LAMMPS which would find the structure at its local minimum and equilibrium. For example, this simulation allows the disconnected and overlapping parts of the structure to develop into their natural state, allowing for a more precise

and accurate MOF featurization (**Figure 4b**). After structural optimization, we acquired 5395 MOFs in total for subsequent screening.

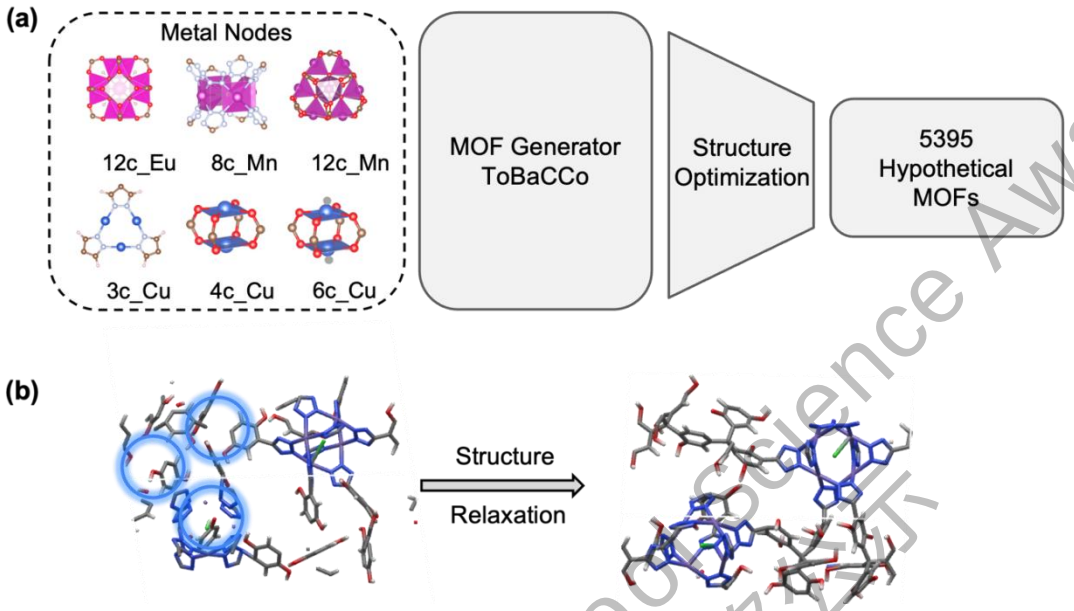

Figure 4. *Example showing the structure relaxation of a **ken** MOF. The as-generated structures have long atom distances (blue dashed circle), and a short MD simulation can eliminate the structural faults to generate an optimized structure for subsequent ML embedding.*

3.3.2 Development of a Surrogate Model for CO_2 Adsorption Capacity and Selectivity

We then aim to develop a machine learning surrogate model that can be used to predict the CO_2 capacity and selectivity of the user-defined MOF set. Based on the previous results, for a deep learning model to capture the performance of CO_2 adsorption of MOFs, we need to consider both the geometric and chemical properties of the MOFs. To this regard, MOFTransformer emerges as a useful tool, within which MOFs are featurized through atom-based graphs and energy grid representation, which capture the local and global features of the MOFs, respectively. It achieves state-of-art performances in a number of prediction tasks of MOFs, including gas adsorption, diffusion, and electronic properties. It is pretrained on 1 million hypothetical MOFs to capture the geometric and chemical properties of MOFs and can be fine-tuned for subsequent learning by a relatively small dataset.

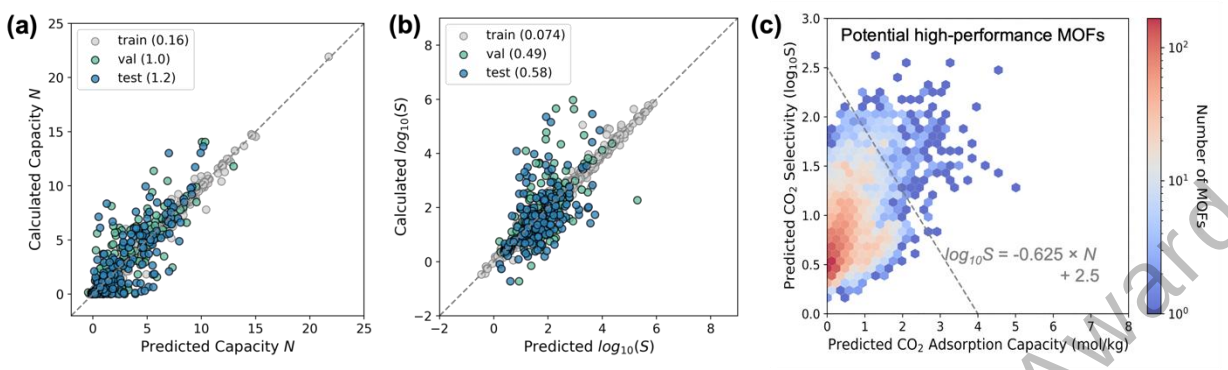

Figure 5. *Fine-tuning the MOFTransformer model for CO_2 capacity and selectivity prediction. Parity plots showing the predicted and calculated values of the training, validation and test sets for CO_2 capacity (a), selectivity trained on a log scale (b). The numbers in the legends are the mean absolute errors (MAE).*

The training dataset consists of all the calculated MOFs from the CoREMOF subset, and 303 randomly selected MOFs from the user-defined MOFs so that the model also learns the MOF features within our own dataset. In **Figure 5a-b**, we show the parity plot labeled data and prediction from the surrogate model for each subset of the data. The models show good prediction performances on the adsorption capacity and the log scale of selectivity, with MAE values being 1.2 and 0.58 for the test set. The results indicate that the MOFTransformer model can be feasibly fine-tuned to yield favorable performances for user-defined tasks, and that the surrogate models are ready for pre-screening of our own dataset.

3.3.3 Discovery of New MOFs For CO_2 Capture from Flue Gas

Using the fine-tuned MOFTransformer model, we performed a prediction on the MOF library we generated and visualized the results in **Figure 5c**. The majority of the predicted structures fall in the low performance region as we observed in the CoREMOF graph. Although we do not see predictions higher than the previously determined structures, the prediction only provides a general trend and some of the structures may still be promising to become good candidates. Therefore, we define the high-performance region to be the area above the arbitrary threshold, and the 273 MOFs located above the line are subjected to more accurate GCMC calculations (**Figure 5c**).

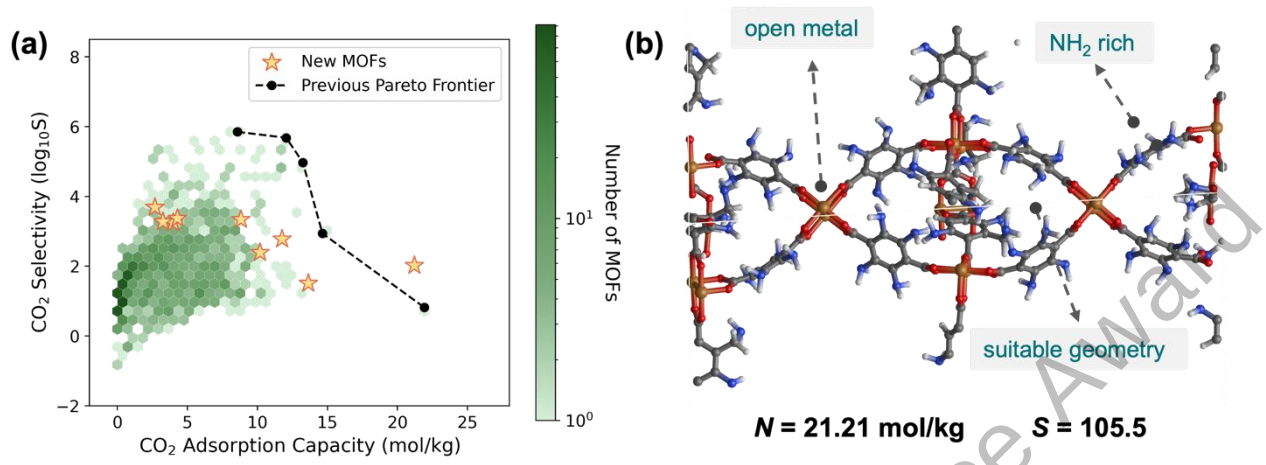

Figure 6. *Screening of the hypothetical MOF library. (a) The top-performing MOFs emerging from our hypothetical MOF library embedded in the previous CoREMOF dataset. (b) The illustration shows a new MOF that extends outside the previous pareto front possessing similar chemistry and geometry as the proposed ones from the CoREMOF dataset.*

Combining the previous calculated 303 structures, in total we obtained 576 calculated MOFs from our library. In **Figure 6a**, we plot the new MOFs that have an adsorption capacity N > 5 mol/kg and a log selectivity > 2 against the CoREMOF dataset as the potential MOFs. One candidate emerges beyond the Pareto front showing similar adsorption capacity as DOTTEM but with high selectivity (**Figure 6b**). From the MOF structure, we observe similar characteristics showing open metal sites and similar geometry as the DOTTEM in the CoREMOF library. In addition, this MOF also contains a number of NH_2 groups that may potentially aid the adsorption of CO_2, and narrow pore sizes may help increase the selectivity. These simple rules can be used as valuable references for experiments to design MOFs for CO_2 adsorption.

4 Conclusions and Discussions

In this study, we performed computational screening over a selected subset of the CoREMOF dataset and a user-defined hypothetical MOF subset for separating CO_2 from flue gas. GCMC simulations were performed under a CO_2/N_2 mixture to calculate the adsorption capacity and selectivity to infer superior MOFs. From the results, we identified 5 MOFs that possess high capacity or high selectivity, potentially for experimental testing. In addition, we found that the relationship between capacity or selectivity and MOF geometries is not monotonic, indicating

that other factors, such as metal types, open metal sites, pore geometries, complicates the design rules of MOFs.

As a step further to design unprecedented MOFs for CO_2 capture from flue gas, we generated a user-defined hypothetical MOF dataset based on the compositional analysis of the CoREMOF set, and then developed a surrogate model built upon MOFTransformer to perform the prediction tasks for new MOFs leveraging the calculated results for CoREMOF. Screening over the hypothetical database led to several new potential MOFs with high performances, with one emerging beyond the Pareto front with better performance than DOTTEM. The whole screening pipeline can be transferred to other properties of MOFs such as bandgaps, energy storage, etc. A simple structural analysis of the proposed existing MOFs and the newly defined MOF indicate that high-performance MOFs are usually characterized by rich open metal sites, suitable and narrow pores, and potential NH_2 groups in the framework. Such design rules can aid experimentalists in the endeavor of development of MOF materials for efficient CO_2 capture from flue gas.

In future, we plan to further enrich the compositional diversity of the hypothetical MOF dataset which can potentially lead to more potent MOF candidates[23, 39, 52, 53]. Also, simulating a flexible framework in a larger supercell would increase the simulation accuracy to provide higher quality data[54, 55]. We also note that the MOFTransformer model is data-hungry and requires sufficient data to provide good performances. Using an active learning pipeline[56] can potentially avoid the heavy computation in our current workflow. Last but not least, a structural analysis using unsupervised machine learning[57] to infer the geometric properties that constitute a good MOF would be significant for both experimentalists and computational chemists to design new MOFs.

References

[1] Solutions, C. f. C. A. E. Global Emissions. 2020.

[2] Papadis, E.; Tsatsaronis, G. Challenges in the decarbonization of the energy sector. *Energy 2020, 205*, 118025. DOI: https://doi.org/10.1016/j.energy.2020.118025.

[3] Pacala, S.; Socolow, R. Stabilization wedges: solving the climate problem for the next 50 years with current technologies. *Science 2004, 305* (5686), 968-972. DOI: 10.1126/science.1100103 From NLM PubMed-not-MEDLINE.

[4] Dziejarski, B.; Serafin, J.; Andersson, K.; Krzyżyńska, R. CO_2 capture materials: a review of current trends and future challenges. *Materials Today Sustainability 2023*, *24*, 100483. DOI: https://doi.org/10.1016/j.mtsust.2023.100483.

[5] Markewitz, P.; Kuckshinrichs, W.; Leitner, W.; Linssen, J.; Zapp, P.; Bongartz, R.; Schreiber, A.; Müller, T. E. Worldwide innovations in the development of carbon capture technologies and the utilization of CO_2. *Energy & Environmental Science 2012*, *5* (6), 7281-7305, 10.1039/C2EE03403D. DOI: 10.1039/C2EE03403D.

[6] Wang, X.; Song, C. Carbon Capture From Flue Gas and the Atmosphere: A Perspective. *Frontiers in Energy Research 2020, 8*, Review. DOI: 10.3389/fenrg.2020.560849.

[7] Rochelle, G. T. 3 - Conventional amine scrubbing for CO_2 capture. In *Absorption-Based Post-combustion Capture of Carbon Dioxide*, Feron, P. H. M. Ed.; Woodhead Publishing, 2016; pp 35-67.

[8] Raganati, F.; Miccio, F.; Ammendola, P. Adsorption of Carbon Dioxide for Post-combustion Capture: A Review. *Energy & Fuels 2021*, *35* (16), 12845-12868. DOI: 10.1021/acs.energyfuels.1c01618.

[9] Bui, M.; Adjiman, C. S.; Bardow, A.; Anthony, E. J.; Boston, A.; Brown, S.; Fennell, P. S.; Fuss, S.; Galindo, A.; Hackett, L. A.; et al. Carbon capture and storage (CCS): the way forward. *Energy & Environmental Science* 2018, *11* (5), 1062-1176, 10.1039/C7EE02342A. DOI: 10.1039/C7EE02342A.

[10] Introduction to Metal–Organic Frameworks. *Chemical Reviews* 2012, *112* (2), 673-674. DOI: 10.1021/cr300014x.

[11] Long, J. R.; Yaghi, O. M. The pervasive chemistry of metal–organic frameworks. *Chemical Society Reviews 2009, 38* (5), 1213-1214, 10.1039/B903811F. DOI: 10.1039/B903811F.

[12] Lu, W.; Wei, Z.; Gu, Z.-Y.; Liu, T.-F.; Park, J.; Park, J.; Tian, J.; Zhang, M.; Zhang, Q.; Gentle Iii, T.; et al. Tuning the structure and function of metal–organic frameworks via linker design. *Chemical Society Reviews 2014*, *43* (16), 5561-5593, 10.1039/C4CS00003J. DOI: 10.1039/C4CS00003J.

[13] Singh, G.; Lee, J.; Karakoti, A.; Bahadur, R.; Yi, J.; Zhao, D.; AlBahily, K.; Vinu, A. Emerging trends in porous materials for CO_2 capture and conversion. *Chemical Society Reviews 2020*, *49* (13), 4360-4404, 10.1039/D0CS00075B. DOI: 10.1039/D0CS00075B.

[14] Yin, X.; Gounaris, C. E. Computational discovery of Metal – Organic Frameworks for sustainable energy systems: Open challenges. *Computers & Chemical Engineering 2022*, *167*, 108022. DOI: https://doi.org/10.1016/j.compchemeng.2022.108022.

[15] Mancuso, J. L.; Mroz, A. M.; Le, K. N.; Hendon, C. H. Electronic Structure Modeling of Metal–Organic Frameworks. *Chemical Reviews 2020*, *120* (16), 8641-8715. DOI: 10.1021/acs.chemrev.0c00148.

[16] 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.; et al. The role of molecular modelling and simulation in the discovery and deployment of metal-organic frameworks for gas storage and separation*. *Molecular Simulation 2019, 45* (14-15), 1082-1121. DOI: 10.1080/08927022.2019.1648809.

[17] Ekberg, V.; Samways, M. L.; Misini Ignjatović, M.; Essex, J. W.; Ryde, U. Comparison of Grand Canonical and Conventional Molecular Dynamics Simulation Methods for Protein-Bound Water Networks. *ACS Physical Chemistry Au 2022, 2* (3), 247-259. DOI: 10.1021/acsphyschemau.1c00052.

[18] Gulbalkan, H. C.; Aksu, G. O.; Ercakir, G.; Keskin, S. Accelerated Discovery of Metal–Organic Frameworks for CO_2 Capture by Artificial Intelligence. *Industrial & Engineering Chemistry Research 2024, 63* (1), 37-48. DOI: 10.1021/acs.iecr.3c03817.

[19] Gomez-Gualdron, D. A.; Gutov, O. V.; Krungleviciute, V.; Borah, B.; Mondloch, J. E.; Hupp, J. T.; Yildirim, T.; Farha, O. K.; Snurr, R. Q. Computational Design of Metal-Organic Frameworks Based on Stable Zirconium Building Units for Storage and Delivery of Methane. *Chem Mater 2014, 26* (19), 5632-5639. DOI: 10.1021/cm502304e.

[20] Banerjee, D.; Simon, C. M.; Plonka, A. M.; Motkuri, R. K.; Liu, J.; Chen, X. Y.; Smit, B.; Parise, J. B.; Haranczyk, M.; Thallapally, P. K. Metal-organic framework with optimally selective xenon adsorption and separation. *Nature Communications 2016, 7*. DOI: 10.1038/ncomms11831.

[21] Chung, Y. G.; Haldoupis, E.; Bucior, B. J.; Haranczyk, M.; Lee, S.; Zhang, H.; Vogiatzis, K. D.; Milisavljevic, M.; Ling, S.; Camp, J. S.; et al. Advances, Updates, and Analytics for the Computation-Ready, Experimental Metal–Organic Framework Database: CoRE MOF 2019. *Journal of Chemical & Engineering Data 2019, 64* (12), 5985-5998. DOI: 10.1021/acs.jced.9b00835.

[22] Azar, A. N. V.; Velioglu, S.; Keskin, S. Large-Scale Computational Screening of Metal Organic Framework (MOF) Membranes and MOF-Based Polymer Membranes for $H2/N_2$ Separations. *ACS Sustainable Chemistry & Engineering 2019, 7* (10), 9525-9536. DOI: 10.1021/acssuschemeng.9b01020.

[23] Altintas, C.; Erucar, I.; Keskin, S. High-Throughput Computational Screening of the Metal Organic Framework Database for CH4/H2 Separations. *Acs Appl Mater Inter 2018, 10* (4), 3668-3679. DOI: 10.1021/acsami.7b18037.

[24] Willems, T. F.; Rycroft, C.; Kazi, M.; Meza, J. C.; Haranczyk, M. Algorithms and tools for high-throughput geometry-based analysis of crystalline porous materials. *Micropor Mesopor Mat 2012, 149* (1), 134-141. DOI: 10.1016/j.micromeso.2011.08.020.

[25] 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* (2), 81-101. DOI: 10.1080/08927022.2015.1010082.

[26] Cheng, Z.; Li, S.; Liu, Y.; Zhang, Y.; Ling, Z.; Yang, M.; Jiang, L.; Song, Y. Post-combustion CO_2 capture and separation in flue gas based on hydrate technology: A review. *Renewable and Sustainable Energy Reviews 2022, 154*, 111806. DOI: https://doi.org/10.1016/j.rser.2021.111806.

[27] Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *Journal of the American Chemical Society 1992, 114* (25), 10024-10035. DOI: 10.1021/ja00051a040.

[28] Ongari, D.; Boyd, P. G.; Kadioglu, O.; Mace, A. K.; Keskin, S.; Smit, B. Evaluating Charge Equilibration Methods To Generate Electrostatic Fields in Nanoporous Materials. *Journal of Chemical Theory and Computation 2019, 15* (1), 382-401. DOI: 10.1021/acs.jctc.8b00669.

[29] Rappe, A. K.; Goddard, W. A. Charge equilibration for molecular dynamics simulations. *The Journal of Physical Chemistry 1991, 95* (8), 3358-3363. DOI: 10.1021/j100161a070.

[30] Bai, P.; Tsapatsis, M.; Siepmann, J. I. TraPPE-zeo: Transferable Potentials for Phase Equilibria Force Field for All-Silica Zeolites. *J Phys Chem C 2013, 117* (46), 24375-24387. DOI: 10.1021/jp4074224.

[31] Maerzke, K. A.; Schultz, N. E.; Ross, R. B.; Siepmann, J. I. TraPPE-UA Force Field for Acrylates and Monte Carlo Simulations for Their Mixtures with Alkanes and Alcohols. *The Journal of Physical Chemistry B 2009, 113* (18), 6415-6425. DOI: 10.1021/jp810558v.

[32] Nath, K.; Wright, K. R.; Ahmed, A.; Siegel, D. J.; Matzger, A. J. Adsorption of Natural Gas in Metal–Organic Frameworks: Selectivity, Cyclability, and Comparison to Methane Adsorption. *Journal of the American Chemical Society 2024, 146* (15), 10517-10523. DOI: 10.1021/jacs.3c14535.

[33] Xie, T.; Grossman, J. C. Crystal Graph Convolutional Neural Networks for an Accurate and Interpretable Prediction of Material Properties. *Phys Rev Lett 2018, 120* (14). DOI: 10.1103/PhysRevLett.120.145301.

[34] Kang, Y. H.; Park, H.; Smit, B.; Kim, J. A multi-modal pre-training transformer for universal transfer learning in metal-organic frameworks. *Nature Machine Intelligence 2023, 5* (3), 309-318. DOI: 10.1038/s42256-023-00628-2.

[35] Ma, R.; Colón, Y. J.; Luo, T. Transfer Learning Study of Gas Adsorption in Metal−Organic Frameworks. *Acs Appl Mater Inter 2020, 12* (30), 34041-34048. DOI: 10.1021/acsami.0c06858.

[36] Li, A.; Perez, R. B.; Wiggin, S.; Ward, S. C.; Wood, P. A.; Fairen-Jimenez, D. The launch of a freely accessible MOF CIF collection from the CSD. *Matter 2021, 4* (4), 1105-1106. DOI: https://doi.org/10.1016/j.matt.2021.03.006.

[37] Colón, Y. J.; Gómez-Gualdrón, D. A.; Snurr, R. Q. Topologically Guided, Automated Construction of Metal-Organic Frameworks and Their Evaluation for Energy-Related Applications. *Cryst Growth Des 2017, 17* (11), 5801-5810. DOI: 10.1021/acs.cgd.7b00848.

[38] Anderson, R.; Gomez-Gualdron, D. A. Increasing topological diversity during computational "synthesis" of porous crystals: how and why. *Crystengcomm 2019, 21* (10), 1653-1665. DOI: 10.1039/c8ce01637b.

[39] Islamov, M.; Babaei, H.; Anderson, R.; Sezginel, K. B.; Long, J. R.; McGaughey, A. J. H.; Gomez-Gualdron, D. A.; Wilmer, C. E. High-throughput screening of hypothetical metal-

organic frameworks for thermal conductivity. *Npj Computational Materials 2023, 9* (1). DOI: 10.1038/s41524-022-00961-x.

[40] O'Keeffe, M.; Peskov, M. A.; Ramsden, S. J.; Yaghi, O. M. The Reticular Chemistry Structure Resource (RCSR) Database of, and Symbols for, Crystal Nets. *Accounts Chem Res 2008, 41* (12), 1782-1789. DOI: 10.1021/ar800124u.

[41] Bucior, B. J.; Rosen, A. S.; Haranczyk, M.; Yao, Z. P.; 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* (11), 6682-6697. DOI: 10.1021/acs.cgd.9b01050.

[42] Zhu, J.; Wu, L.; Bu, Z.; Jie, S.; Li, B.-G. Polyethyleneimine-Modified UiO-66-NH_2(Zr) Metal–Organic Frameworks: Preparation and Enhanced CO_2 Selective Adsorption. *ACS Omega 2019, 4* (2), 3188-3197. DOI: 10.1021/acsomega.8b02319.

[43] Justin, A.; Espín, J.; Kochetygov, I.; Asgari, M.; Trukhina, O.; Queen, W. L. A Two Step Postsynthetic Modification Strategy: Appending Short Chain Polyamines to Zn-NH_2-BDC MOF for Enhanced CO_2 Adsorption. *Inorganic Chemistry 2021, 60* (16), 11720-11729. DOI: 10.1021/acs.inorgchem.1c01216.

[44] Thompson, A. P.; Aktulga, H. M.; Berger, R.; Bolintineanu, D. S.; Brown, W. M.; Crozier, P. S.; Veld, P. J. I.; Kohlmeyer, A.; Moore, S. G.; Nguyen, T. D.; et al. LAMMPS-a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales. *Comput Phys Commun 2022, 271*. DOI: 10.1016/j.cpc.2021.108171.

[45] Anderson, R.; Gómez-Gualdrón, D. A. Large-Scale Free Energy Calculations on a Computational Metal – Organic Frameworks Database: Toward Synthetic Likelihood Predictions. *Chem Mater 2020, 32* (19), 8106-8119. DOI: 10.1021/acs.chemmater.0c00744.

[46] Coupry, D. E.; Addicoat, M. A.; Heine, T. Extension of the Universal Force Field for Metal-Organic Frameworks. *Journal of Chemical Theory and Computation 2016, 12* (10), 5215-5225. DOI: 10.1021/acs.jctc.6b00664.

[47] Hjorth Larsen, A.; Jørgen Mortensen, J.; Blomqvist, J.; Castelli, I. E.; Christensen, R.; Dułak, M.; Friis, J.; Groves, M. N.; Hammer, B.; Hargus, C.; et al. The atomic simulation environment—a Python library for working with atoms. *Journal of Physics: Condensed Matter 2017, 29* (27), 273002. DOI: 10.1088/1361-648X/aa680e.

[48] Marti, R. M.; Howe, J. D.; Morelock, C. R.; Conradi, M. S.; Walton, K. S.; Sholl, D. S.; Hayes, S. E. CO_2 Dynamics in Pure and Mixed-Metal MOFs with Open Metal Sites. *The Journal of Physical Chemistry C 2017, 121* (46), 25778-25787. DOI: 10.1021/acs.jpcc.7b07179.

[49] Britt, D.; Furukawa, H.; Wang, B.; Glover, T. G.; Yaghi, O. M. Highly efficient separation of carbon dioxide by a metal-organic framework replete with open metal sites. *Proc Natl Acad Sci U S A 2009, 106* (49), 20637-20640. DOI: 10.1073/pnas.0909718106 From NLM Medline.

[50] Kochetygov, I.; Bulut, S.; Asgari, M.; Queen, W. L. Selective CO_2 adsorption by a new metal–organic framework: synergy between open metal sites and a charged imidazolinium backbone. *Dalton Transactions 2018, 47* (31), 10527-10535, 10.1039/C8DT01247D. DOI: 10.1039/C8DT01247D.

[51] Hao, C.; Ren, H.; Zhu, H.; Chi, Y.; Zhao, W.; Liu, X.; Guo, W. CO_2-favored metal–organic frameworks SU-101(M) (M = Bi, In, Ga, and Al) with inverse and high selectivity of CO_2 from C2H2 and C2H4. *Separation and Purification Technology 2022, 290*, 120804. DOI: https://doi.org/10.1016/j.seppur.2022.120804.

[52] Ponraj, Y. K.; Borah, B. High-Throughput Computational Screening of Metal-Organic Frameworks for the Separation of Methane from Ethane and Propane. *J Phys Chem C 2021, 125* (3), 1839-1854. DOI: 10.1021/acs.jpcc.0c09117.

[53] Majumdar, S.; Moosavi, S. M.; Jablonka, K. M.; Ongari, D.; Smit, B. Diversifying Databases of Metal Organic Frameworks for High-Throughput Computational Screening. *Acs Appl Mater Inter 2021, 13* (51), 61004-61014. DOI: 10.1021/acsami.1c16220.

[54] Oktavian, R.; Goeminne, R.; Glasby, L. T.; Song, P.; Huynh, R.; Qazvini, O. T.; Ghaffari-Nik, O.; Masoumifard, N.; Cordiner, J. L.; Hovington, P.; et al. Gas adsorption and framework flexibility of CALF-20 explored via experiments and simulations. *Nature Communications 2024, 15* (1), 3898. DOI: 10.1038/s41467-024-48136-0.

[55] Zheng, B.; Gu, G. X.; Santos, C. d.; Neumann Barros Ferreira, R.; Steiner, M.; Luan, B. Simulating CO_2 diffusivity in rigid and flexible Mg-MOF-74 with machine-learning force fields. *APL Machine Learning 2024, 2* (2). DOI: 10.1063/5.0190372 (acccessed 8/20/2024).

[56] Osaro, E.; Mukherjee, K.; Colón, Y. J. Active Learning for Adsorption Simulations: Evaluation, Criteria Analysis, and Recommendations for Metal – Organic Frameworks. *Industrial &*

Engineering Chemistry Research 2023, 62 (33), 13009-13024. DOI: 10.1021/acs.iecr.3c01589.

[57] Moosavi, S. M.; Nandy, A.; Jablonka, K. M.; Ongari, D.; Janet, J. P.; Boyd, P. G.; Lee, Y.; Smit, B.; Kulik, H. J. Understanding the diversity of the metal-organic framework ecosystem. *Nature Communications 2020, 11* (1), 4068. DOI: 10.1038/s41467-020-17755-8.

[58] Bobbitt, N. S.; Shi, K.; Bucior, B. J.; Chen, H.; Tracy-Amoroso, N.; Li, Z.; Sun, Y.; Merlin, J. H.; Siepmann, J. I.; Siderius, D. W.; et al. MOFX-DB: An Online Database of Computational Adsorption Data for Nanoporous Materials. *Journal of Chemical & Engineering Data 2023, 68* (2), 483-498. DOI: 10.1021/acs.jced.2c00583.