Title
Computationally advancing the predictive power of the nanoporous materials genome

Author
Witman, Matthew

Publication Date
2019

Peer reviewed|Thesis/dissertation

Computationally advancing the predictive power of the nanoporous materials genome

by

Matthew D Witman

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Chemical Engineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Berend Smit, Chair
Professor Jeffrey Reimer
Professor Ali Mesbah
Professor Jeffrey Neaton

Spring 2019

Computationally advancing the predictive power of the nanoporous materials genome

Abstract

Computationally advancing the predictive power of the nanoporous materials genome

by

Matthew D Witman

Doctor of Philosophy in Chemical Engineering

University of California, Berkeley

Professor Berend Smit, Chair

Since tens of thousands of nanoporous materials have been synthesized and millions have been hypothesized, our ability to both predict novel materials and their performance for a given application in a computational setting is becoming increasingly important. Utilizing simulations that can screen for optimal material performance becomes invaluable since exhaustive experimental exploration of this vast chemical space can quickly become experimentally intractable from an expense and time efficiency standpoint. This thesis is therefore dedicated to advancing the computational tools and algorithms that can be used to both predict novel materials and evaluate their performance for clean energy technologies.

To the family, friends, mentors, colleagues, and advisers

whose guidance and support has made this endeavor possible.

Contents

List of Figures

List of Tables

Acknowledgments

Thank you to all of the collaborators who I've had the pleasure of working with throughout the course of this thesis, especially those who contributed directly to the work presented in this manuscipt: Berend Smit, Maciej Haranczyk, Sanliang Ling, Ben Slater, Peter Boyd, Sudi Jawahery, Senja Barthel, Nathan Mahynski, Kyriakos Stylianou, Samantha Anderson, Andrezj Gładysiak, and Lianheng Tong. It has been a pleasure collaborating with you and I am very thankful for your support.

Chapter 1

A brief introduction to the nanoporous materials genome

1.1 Nanoporous materials galore

The discovery of crystalline materials with microscopic pores (defined as 0.2-2 nm in size by the International Union of Pure and Applied Chemistry) has given rise to an entire field of scientific research. Some of the earliest investigations of zeolites revealed that the shape and size of these microscopic cavities could be used to catalyze reactions and separate chemical species, leading to their widespread adoption in the petrochemical industry.[1, 2] Zeolites have dominated (and still do) in mainstream industrial applications, but it was discovered starting in the 1990s that different chemistries and pore topologies could be accessed via different classes of porous materials such as metal-organic frameworks (MOFs),[3–6] zeolitic imidazolate frameworks (ZIFs),[7] covalent organic frameworks (COFs),[8] and porous polymer networks (PPNs).[9] This has led to an explosion of synthesizeable nanoporous materials over the past two decades, with thousands of experimentally validated crystal structures having been deposited in the Cambridge Crystallographic Database.[10] Despite the plethora of nanoporous material classes that exist, this thesis will focus on two specific types, MOFs and zeolites. This section briefly introduces these two classes of materials, and for more detailed background information the reader is referred to the review articles cited within.

Zeolites

As of the beginning of 2019, 234 zeolites have been reported by the International Zeolite Association, with a handful that exist as naturally occurring minerals. These microporous structures can typically be thought of as aluminosilicates (consisting of Si^{4+}, O^{2-}, and Al^{3+} ions) composed of tetrahedrally coordinated T-atoms. These tetrahedral units combine in a collection of secondary building units (SBUs) in various ways, yielding the wide range of pore sizes and topologies that are exhibited by the IZA database of zeolites.[11] An example of both a zeolite SBU and full framework is shown in Fig. 1.1.

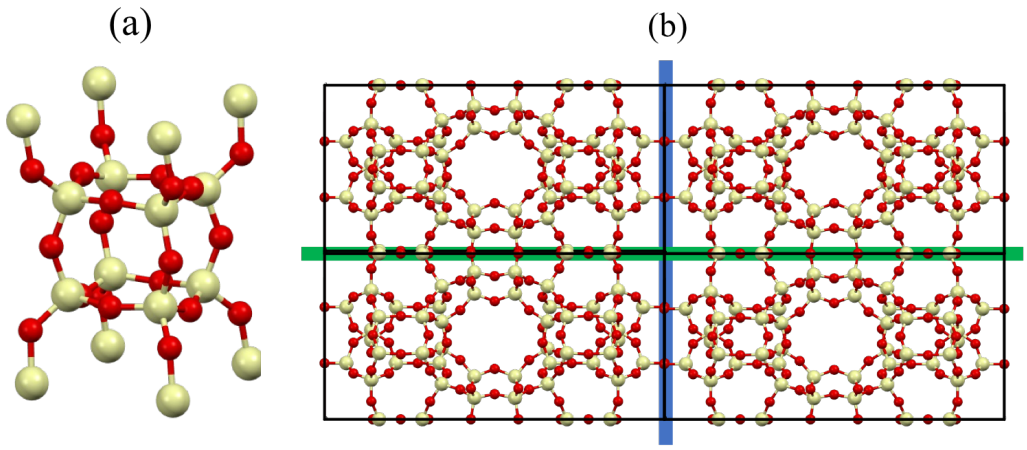


Figure 1.1: (a) A particular zeolite SBU known as a double 4 ring (D4R), which also demonstrates the fundamental zeolite unit: a Si atom with tetrahedral coordination to its neighboring O atoms. (b) Visualization of the zeolite MWW supercell.

Metal-organic frameworks

MOFs are a widely studied class of crystalline, nanoporous materials whose composition of metal or metal oxide nodes connected by organic linkers results in highly tunable structures with diverse chemical and geometric properties.[12] The discipline of reticular chemistry, or the rational design of MOFs from molecular building blocks in an underlying topology [13], has been utilized to design vastly diverse structures. Tens of thousands of MOF materials have been experimentally validated and their structural information submitted at the Cambridge Structural Database.[14] Hence reticular chemistry has been a useful tool, both experimentally and computationally, to tune these materials structural and chemical properties for a given application. Fig. 1.2 schematically demonstrates how different ligands and metal nodes can be combined to form materials with diverse structure and chemistry.

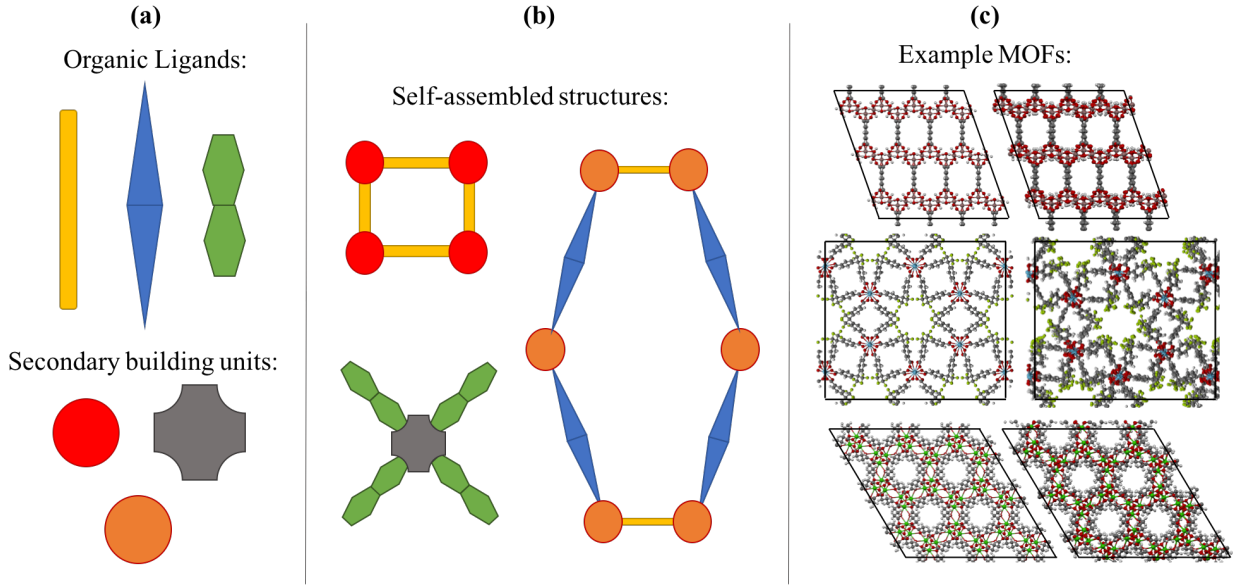


Figure 1.2: (a) Schematic visualization of the organic linkers and metal oxide nodes, or secondary building units (SBUs), from which MOFs are constructed and (b) self assemble into frameworks exhibiting diverse geometry and chemistry. (c) Three different synthesized MOFs are shown to demonstrate their structural diversity.

Hypothetical materials

In addition to those that have been synthesized in the laboratory, a plethora of structures have been hypothesized by following the molecular design rules associated with a particular class of materials. In one of the first examples of this large scale design of hypothetical nanoporous materials, Deem and coworkers used Monte Carlo simulations to identify low energy potential zeolite structures, resulting in a database of over 2,000,000 structures.[15, 16] The advent of MOFs gave rise to many more structure prediction algorithms,[17–22] the most advanced of which were topology based: SBU building blocks can be abstracted and overlaid on a template 3D net to construct the final hypothetical structure. Improvement to these structure building algorithms can still be achieved, since even the topology based approaches still do not have an automated way of dealing with particular families of MOFs (see Chapter 2 for a more detailed discussion). Using topological data analysis, the similarity in pore structures can be quantified amongst the materials in these combined databases and help shine light on the true pore space diversity that is spanned by these disparate structure building algorithms.[23] Ultimately, the presence of millions of hypothetical zeolites, MOFs, COFs, ZIFs, and PPNs necessitates an investigation of a large number of these structures understand the properties that yield optimal performance.

Proposed applications

The following is a non-exhaustive list of some applications for which nanoporous materials have been proposed. These range from the widely studied to the more recently proposed applications: methane storage,[24] hydrogen storage,[25] CO_2 removal from flue gas,[26] light hydrocarbon separations,[27] mixed-matrix membranes,[28] natural gas upgrading,[29] chemical sensing,[30] water harvesting,[31] water desalination,[32] heavy metal extraction for water purification,[33] gold extraction from wastewater,[34] quantum computing architectures,[35] scaffolds for nanoconfined systems,[36] single site catalysts,[37] drug delivery,[38] photocatalytic hydrogen production,[39] supercapacitors,[40] lithium ion battery anodes,[41] and the list goes on. From an applications stand point, this thesis will focus mainly on computational techniques that improve our ability to find and assess optimal materials in gas storage and separations.

1.2 A justification for computational materials design and simulation

The number of nanoporous materials, both synthesized and hypothesized, combined with the number of applications for which they are being studied, makes the identification of optimal materials an extremely daunting task. For an order of magnitude estimate, there are 10,000 synthesized nanoporous materials and 2,000,000 hypothesized structures contained in published databases (although the number of hypothetically enumerable structures is in fact much larger).[42] Simplistically assuming the only applications of interest are the 18 listed in the previous section (many of which in actuality contain further sub-applications that would warrant investigation), the upper bound on the number of experiments we conceivably would need to design understand each material and find the optimum for a given application would be on the order of 180,000. It takes weeks or months to create, characterize, and test a material in the laboratory. If we rely on 1,000 experimental research groups around the globe operating non-stop year round, taking 30 days to fully synthesize, characterize, and test a new material, we would have to wait 15 years to complete all these experiments. Matters are further complicated by the fact that the optimal material may lie within a hypothetical database and has not yet even been synthesized and reported. In such a case simulations are the only way to discover this materials utility. Clearly, if sufficiently accurate computational techniques exist to pre-screen, understand, and identify optimal hypothetical or existing candidates for a particular application, significant time and cost savings can be achieved in the materials design process.

1.3 Improving computational identification of optimal porous materials

This thesis is organized into the contributions we have made in the three major components of computational prediction of optimal nanoporous materials, the steps of which are summarized in Fig. 1.3. In the life cycle of optimal material identification, one must first generate novel hypothetical structures (Step 1 of Fig. 1.3) to study. Alternatively, rather than focusing on hypothetical materials, one can focus only on those that have already been synthesized.[43] However, by keeping synthesizability in mind when generating novel hypothetical structures, we can limit the structure prediction exercise to materials that are likely to be experimentally realized (Chapters 2, 3, and 4). Once we have a database of materials, we need to have efficient computational techniques that can efficiently predict a property of interest (Step 2 of Fig. 1.3). For example, we have extended a well-known enhanced sampling Monte Carlo (MC) technique that permits more efficient acquisition of thermodynamic adsorption data, thereby enabling a more holistic evaluation of adsorption performance from an industrial process perspective (Chapter 5). Armed with a database of materials and the necessary computational chemistry techniques to predict any properties of interest, we can finally evaluate their performance across the entire database and not only identify optimal materials for a given application but understand what properties enable this optimality (Step 3 of Fig. 1.3). In this work we demonstrate how we can advance such high-throughput screening studies for gas separations by simulating the frameworks as flexible (a more realistic approach than the rigid structure approximation), a topic discussed in Chapter 6. The final step (Step 4 of Fig. 1.3) of the computational materials design life cycle is to experimentally validate any *in silico* predictions that have been made. Alternatively, we can begin the entire process at Step 4, using previous experimental results to motivate the computational investigation of a particular family of materials.

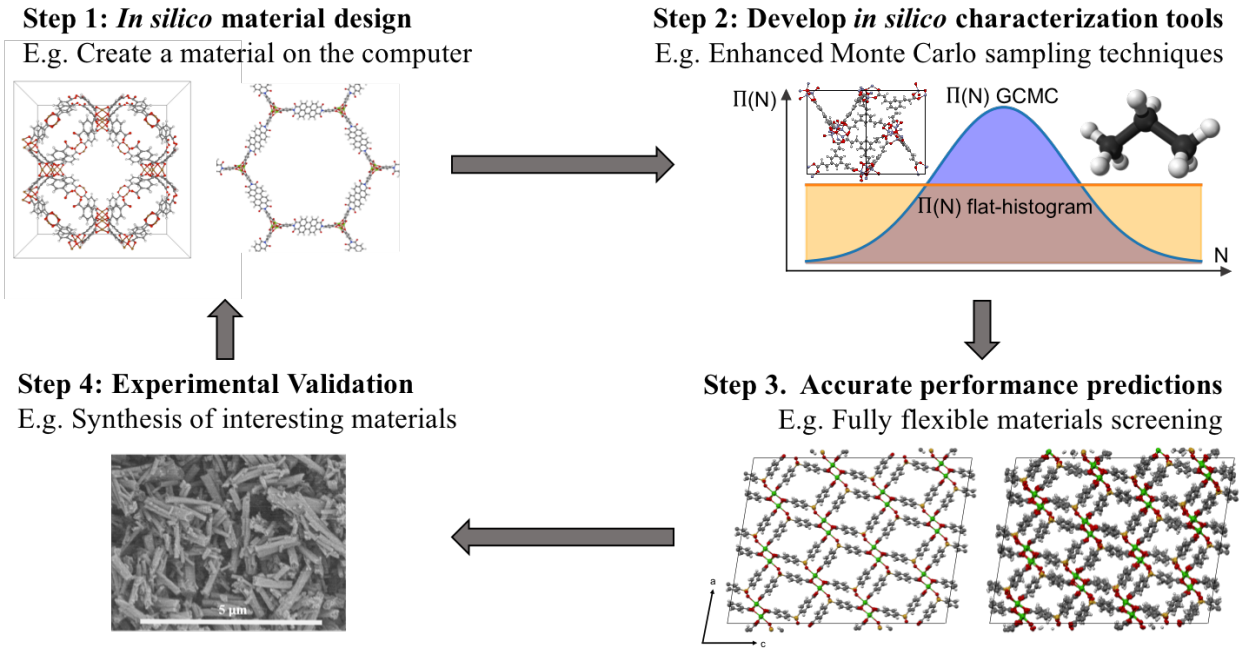


Figure 1.3: A typical process for the computational design and identification of optimal materials

Chapter 2

In silico materials design targeting experimentally realizable structures

This chapter focuses on the *in silico* design of MOFs exhibiting 1-dimensional rod topologies. We introduce an algorithm for construction of this family of MOF topologies, and illustrate its application for enumerating MOF-74-type analogs. Furthermore, we perform a broad search for new linkers that satisfy the topological requirements of MOF-74 and consider the largest database of known chemical space for organic compounds, the PubChem database. Our *in silico* crystal assembly, when combined with dispersion-corrected density functional theory (DFT) calculations, is demonstrated to generate a hypothetical library of open-metal site containing MOF-74 analogs in the 1-D rod topology from which we can simulate the adsorption behavior of CO_2. We finally conclude that these hypothetical structures have synthesizable potential through computational identification and experimental validation of a novel MOF-74 analog, Mg_2(olsalazine).

2.1 Reticular enumeration of MOFs: an extension to 1-D rod materials

The discipline of reticular chemistry, or the rational design of MOFs from molecular building blocks in an underlying topology [13], has been utilized to design vastly diverse structures and has been a useful tool, both experimentally and computationally, to tune these materials structural and chemical properties for a given application. Previous computational work has been directed towards generating large databases of hypothetical MOFs *in silico* through a variety of methods. MOFs can be abstracted by a special type of graph known as a periodic net that contains the information of their underlying topologies, and resources such as the RCSR [44] and EPINET [45] databases have been created to store this information. Martin and Haranczyk used the connectivity information of molecular building blocks and these catalogued nets to assemble hypothetical materials *in silico* for various classes of nanoporous materials [19] and a similar approach was later implemented by the Heine group.[46] The connection-based assembly strategy of Wilmer et al., which builds MOFs building block by building block, was shown to generate over 137,000 hypothetical MOFs with diverse geometric properties and linker functionalization.[18] Three hundred of these materials displayed a greater simulated methane storage capacity than any known material at the time. In addition to the assembly algorithm, a methodology for linker selection is also necessary for *in silico* design efforts. Previous methods to achieve this have relied on screening moderate-sized databases of commercially available molecules ($\sim$7,000,000),[47] using a genetic algorithm to investigate composition space of possible linkers,[48] or brute-force enumeration of linkers from fragments of known MOFs.[18] These studies involving enumerated hypothetical databases have proven useful for, among other things, elucidating structure-function relationships. Yet a noticeable limitation is that most, with a notable exception of the MIL-47 analogs of Wilmer, do not contain MOFs exhibiting a 1-D rod topology [49] in which the metal-oxide secondary building units (SBUs) extend as a rod infinitely in one direction. MOFs exhibiting this topological feature have demonstrated extremely interesting behavior including remarkable CO_2 capture in MOF-74 [50] and so called breathing frameworks [51, 52] which can yield stepped adsorption isotherms.

Yet thus far, reticular chemistry studies on 1-D rod MOFs have been almost exclusively limited to the experimental realm, with intense focus on exploring analogs of MOF-74 to increase its already remarkable potential for CO_2 capture. One of the original works in this area performed an isoreticular expansion of MOF-74 through exchange of the 2,5-dioxidobenzene-1,4-dicarboxylate (DOBDC) linker for ligands extended by an increasing number of benzene rings,[53] thereby increasing the channel volume of MOF-74, also known as Mg_2(DOBDC). Other efforts have aimed to tune or alter the open-metal site chemistry by utilizing different metal elements during synthesis [54, 55], i.e. M_2(DOBDC) [M = Mg, Fe, Co, Ni, Zn], or through functionalization of the open-metal site by appendage of amine based molecules.[56]

Our reticular chemistry study combines an application that has only been studied experimentally, the creation of MOF-74 analogs, with a computational method for the automated

generation of hypothetical analogs of MOFs exhibiting a 1-D rod topology. In place of brute-force enumeration of hundreds or thousands of structures across various topologies seen in previous *in silico* studies, we take a more directed approach to investigate the MOF-74 structure and vary the chemical composition of the ligand to alter its geometric and chemical properties. While this has been done previously for structures in the simpler MOF-5 topology,[47] MOF-74 requires a novel building algorithm since it exhibits 1-D SBU rods and complex connectivity between ligands and SBUs. Furthermore, in this study we explore the chemical space represented by the ∼60,000,000 distinct chemical species in the PubChem Compounds database [57] and exchange linkers found within this space with the DOBDC linker in the original structure. Only 61 molecules (0.0001% of this database) are identified as feasible MOF-74 analog linkers, which provides a concise starting point for attempts at experimental synthesis. One ligand in the set that was also commercially available, known by its pharmaceutical name olsalazine or 3,3'-azobis(6-hydroxybenzoate)salicylic acid, was used to successfully synthesize a MOF-74 analog. We significantly increase the impact of our *in silico* screening by demonstrating that novel, predicted structures are indeed synthesizable.

Hence we develop a hypothetical structure generation method using a geometric optimization routine which permits the rapid *in silico* assembly of 1-D rod MOF analogs (specifically MOF-74 structures in this study) while only using linkers guaranteed to represent an experimentally realizable molecule. The library of hypothetical MOF-74 analogs was structurally analyzed and screened for CO_2 capture potential by comparison to the performance of the original structure. Finally, the knowledge ascertained from this computational study was utilized to direct efforts to synthesize a particular MOF-74 analog from a commercially available linker, yielding an experimental structure in excellent agreement with the *in silico* generated and optimized structure. Thus we have introduced a set of experimentally realizable structures that can now be targeted in future synthetic efforts.

2.2 *In silico* generation of hypothetical MOF-74 analogs

The overall workflow for *in silico* generation, analysis, and experimental synthesis of hypothetical MOF-74 analogs consists of five steps and is summarized below:

1. **Identification of potential ligands**. Chemical substructure searching and ligand conformational analysis is performed to select potential substitutes for the MOF-74 DOBDC linker from the PubChem Compounds database.

2. **Crystal structure assembly**. Potential ligands are inserted into the MOF-74 topology through geometric optimization of the crystal structure.

3. **Dispersion corrected DFT**. Dispersion corrected DFT calculations are performed to optimize each assembled structure. REPEAT [58] derived partial atomic charges are obtained following the optimization, required for step 4.

4. **Grand Canonical Monte Carlo (GCMC) simulations**. CO_2 adsorption is simulated at flue gas conditions with GCMC for each structure in the library of MOF-74 analogs.

5. **Prioritization of synthetic targets**. Hypothetical crystal structures constructed from commercially available linkers are identified as synthetic priority targets. The hypothetical MOF-74 analog, hereon named Mg_2(olsalazine), was synthesized from the commercially available olsalazine linker to validate our approach of generating experimentally realizable crystal structures *in silico*.

The visualization of important structural features in MOF-74, shown in Fig. 2.1, is intended as a visual aid for the reader throughout this section. This crystal exhibits infinite 1-D hexagonal channels, where adjacent metal-oxide rods are bridged by a DOBDC linker. A DOBDC linker connects to each rod through its two carboxylic acid oxygen atoms and an adjacent phenolic oxygen.

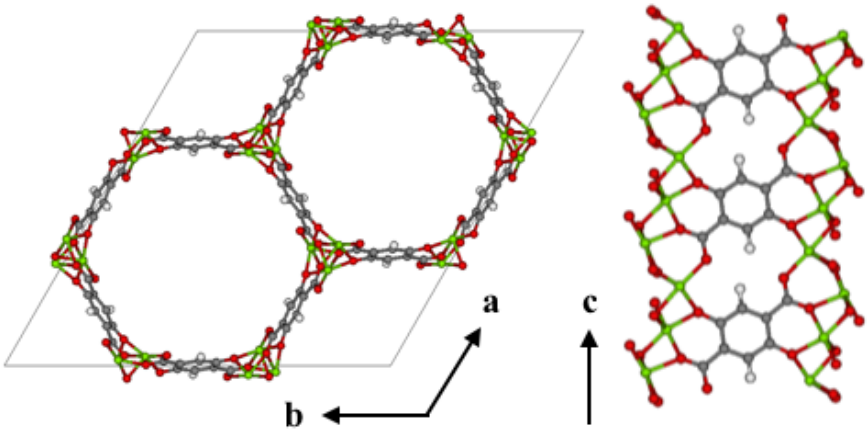


Figure 2.1: The *ab* face of Mg-MOF-74 illustrates the hexagonal channel shape. A slice of the channel wall demonstrates the extension of the 1-D metal oxide (Mg-O, Green-Red, respectively) rods in the c-direction and the connectivity of the DOBDC linker to two adjacent rods.

Identification of potential ligands

To create a library of hypothetical MOF-74 analogs requires identifying linkers which can feasibly be exchanged with the DOBDC linker in the original structure. The steps through which ligands were identified and chosen for MOF-74 enumeration is summarized in Fig. 2.2.

1. **Database Selection**. For this study we expand the scope of our exploration of chemical space from a commercial database such as eMolecules (www.emolecules.com),

sometimes referred to as the amazon.com for chemicals, to the larger PubChem Compounds database. This PubChem library of molecules consists of ∼160,000,000 user deposited structures, from which 60,000,000 are chemically validated and determined to be unique. This subset of unique molecules is known as the PubChem Compounds database and is approximately 9 times larger than the eMolecules database. While these PubChem compounds may not necessarily be commercially available, any ligand identified from this library should indeed represent a feasible, experimentally realizable chemical structure.

2. **Connectivity Filter**. The chemical connectivity of MOF-74 demonstrates greater complexity than the connectivity seen in the broad class of MOFs where the ligand attaches to the SBUs by two or more carboxylic acid groups. Here we require a phenol group directly adjacent to the carboxylic acid group in order to have the correct connectivity. We impose the constraint that a molecule must have exactly two of these identical substructures shown in Fig. 2.2 to satisfy the particular chemical connectivity of MOF-74. This substructure can be abstracted as a SMARTs string, or a textual representation of a substructure within a molecular graph, which can be readily identified by the open-source software package OpenBabel.[59] Thus each molecular entry in PubChem can be quickly analyzed and identified if it contains a particular SMART. After removing database entries that contained fragmented structures, salts, or net charged species and filtering for molecules that contained exactly two of the required SMARTs for MOF-74 connectivity, all but ∼400 molecules remained as potential ligands. Thus simply identifying molecules with MOF-74 connectivity indicates we can only investigate 0.0007% of the chemical space contained in the PubChem database.

3. **High Energy Conformer Filter**. Additional filtering was required to identify potential linkers that could adopt a conformation suitable for building a MOF-74 analog. The DOBDC linker is connected to a given metal rod SBU by three Mg-O bonds, one at each carboxylic acid oxygen and a third at the phenolic oxygen (see Fig. 2.1). This third connection point from the linker to the SBU adds the constraint that the terminal benzene ring which connects to the SBU is not free to rotate about the axis defined by the carboxylic acid bond vector. Since the original linker consists of just one benzene ring, any proposed linker from which we attempt to build a MOF-74 analog must have terminal benzene rings that lie in the same plane. An additional requirement is imposed in which the angle between the two vectors represented by the carboxylic acid carbon to benzene ring bond in both connection groups are required to have an angle of 180 degrees or 120 degrees. If the angle between the vectors is 180 degrees, then the assembled crystal structure will be a variant of the originally published MOF-74 structure.[60] If the angle between vectors is 120 degrees, then the assembled structure will be a variant of the m-MOF-74 structure.[54] These orientation constraints significantly affect the number of linkers that can be used to enumerate MOF-74 analogs. The Confab tool in OpenBabel, which systematically generates diverse, low-energy

conformers,[61] was used to generate a set of low energy conformers for each of the 400 remaining linkers. The criteria for a low energy conformer was set at 50 kJ/mol above the minimum energy conformer in the gas phase and was calculated via the Merck Molecular Force Field (MMFF94).[62] The molecules that passed the conformational test were those that contained at least one conformer in which the planes defined by the terminal benzene atoms were within +/- 10% of 180 degrees and carboxyl acid bond vectors that were within +/- 10% of 120 degrees or 180 degrees. Discarding molecules that could not realistically achieve the required conformational constraints of MOF-74 resulted in the removal of all but 61 structures from the set of potential ligands, yielding a collection of molecules representing 0.0001% of the PubChem database. While not a quantitative predictor of the synthesizability of a MOF-74 analog, the use of only low energy conformers suggests that the assembled MOF-74 analog may be experimentally realizable. Three of the 61 selected linkers in the PubChem database are shown in Fig. 2.3. Each ligand has access to a low energy conformer that exhibits both planarity between the terminal benzene rings and an angle between the two carboxylic acid C-C bonds of either 120 degrees or 180 degrees. All remaining linkers can be seen in the assembled MOF-74 analogs by visualizing the Crystallographic Information (CIF) files provided in Ref. [21].

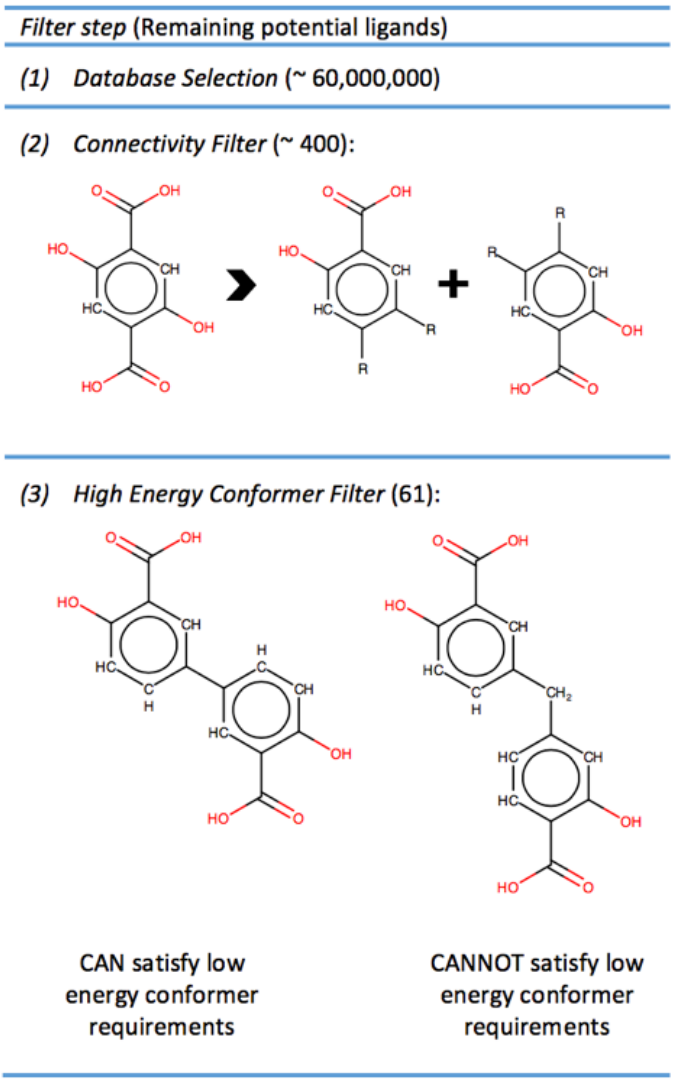


Figure 2.2: *Step 1*: PubChem Compounds database is chosen to screen ∼60,000,000 chemicals. *Step 2*: Filtering out molecules not appropriate for MOF ligands and those that do not satisfy MOF-74 chemical connectivity removes all but 0.0007% of compounds. *Step 3*: Molecules that cannot achieve a low energy conformer compatible with MOF-74 geometric requirements are filtered, thereby removing 84% of the remaining potential ligands.

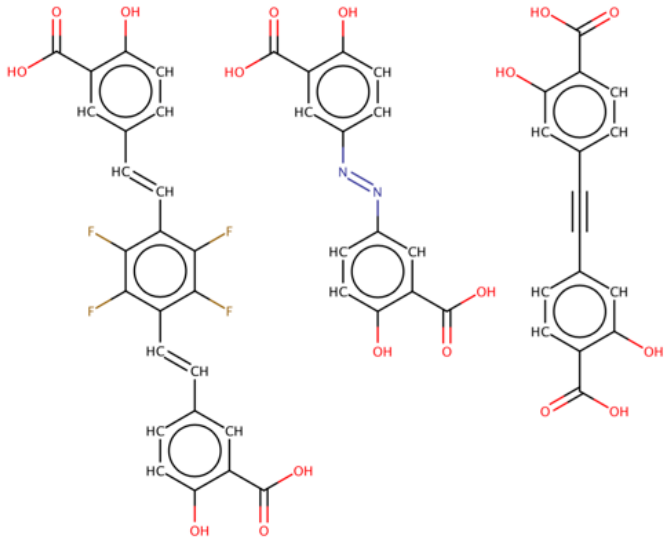


Figure 2.3: A sample of the 61 molecules found in the PubChem database that were identified as potential ligands for MOF-74 analogs. The middle linker, olsalazine, is utilized to synthesize a novel MOF-74 analog.

Crystal structure assembly

The **etb** net of MOF-74, visualized in Fig. 2.4, exhibits special complexities which preclude the use of previously developed methods for *in silico* crystal design.[18, 19] Three distinct problems arise. Firstly, not all edges in nets that abstract a 1-D rod MOF represent a linker or segment of a linker. As seen in Fig. 2.4, only some edges (highlighted in red) in the **etb** net abstract the ligand that connects one SBU rod to another, while others abstract edges that exist within the SBU rod itself (highlighted by thin black lines). Secondly, each node in **etb** does not abstract a single, discrete SBU in MOF-74 but rather they abstract only the metal atom contained within the SBU rod. For example, all nodes that can be visited without traversing a red edge represent the three metals contained by one rod within a single unit cell. Thirdly, attempting to define the c-component of the SBU rods central (a, b, c) coordinates becomes ambiguous since the rod transverses periodic boundary conditions in the c-direction. Thus the process of embedding the SBU rods in three dimensional space can significantly change depending on the selected linker. Upon exchange of the red edges in Fig. 2.4b with an alternative linker to DOBDC, the edge may in fact no longer be parallel to the ab face and all the SBU rods will shift relative to each other in the c-direction based on the new ligands geometry. These three difficulties necessitate a new methodology to quickly assemble crystal structures from the 61 selected ligands.

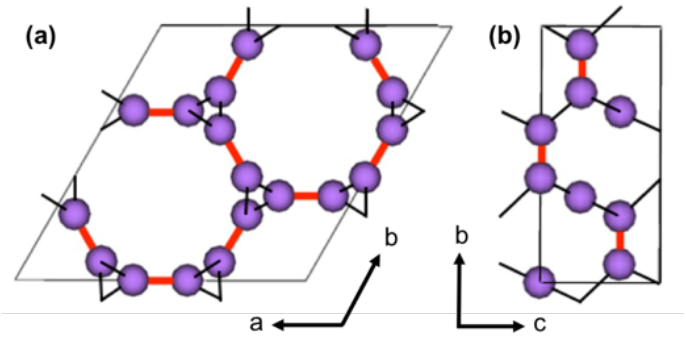


Figure 2.4: The **etb** net representation of MOF-74 where purple spheres represent nodes and lines represent edges in the graph. Nodes correspond to the metal atoms in each rod, red lines are edges that abstract a ligand in MOF-74, and thin black lines abstract connections within a single SBU rod. (a) View of the *ab* face of **etb** net. (b) View of the *bc* face of the **etb** net.

We overcome these challenges by simplifying our view of the MOF-74 net topology and by using a simple geometric optimization routine to assemble a new crystal structure for a given ligand. We identify three geometric constraints that must remain invariant after ligand exchange, schematically represented in Fig. 2.5. Firstly, the length of a ligand, L, is quantified as the distance between the two carboxylic acid carbons since they represent atoms that are immediately connected to the SBU rods. Thus they are termed the connection points of each rod. Upon enumerating a new crystal structure, we require that the magnitude of the distance between the connection points on two adjacent rods be equal to the length of the ligand. Secondly, the chemical environment within the SBU rod is invariant, so the relative coordinates of all Mg and O atoms in the rod are fixed. Thirdly, the a and b fractional coordinates of the center of all rods must remain constant to preserve the hexagonal packing of the crystal structure.

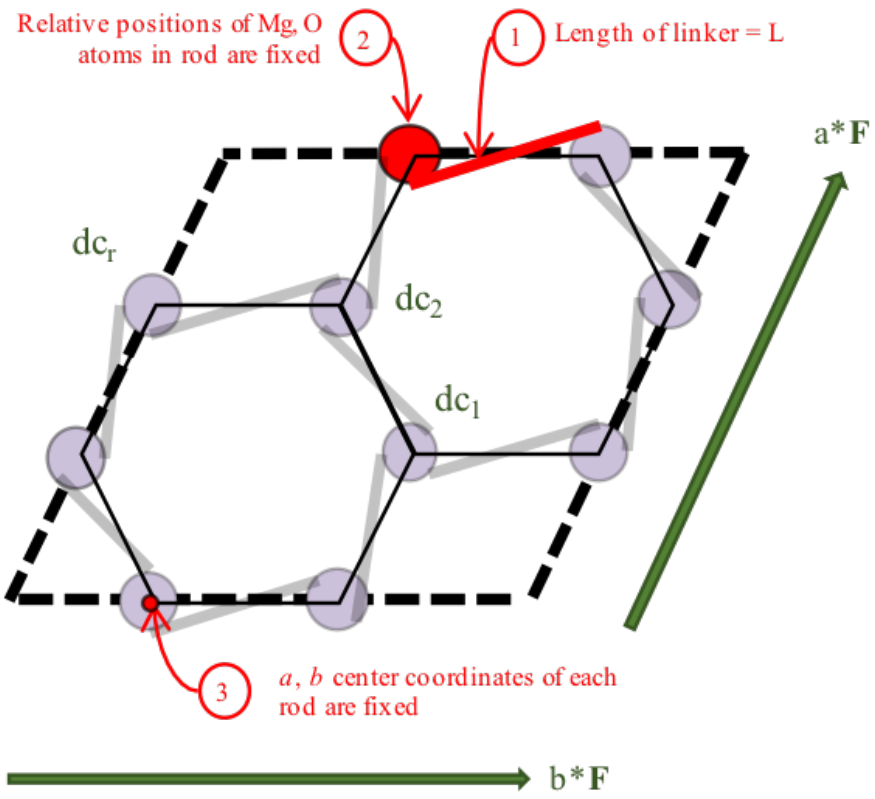


Figure 2.5: An abstracted view of the MOF-74 unit cell visually demonstrates the three constraints for building analog 1-D structures in red as well as the optimization variables in green $\{F, dc_1, dc_2, \ldots, dc_r\}$. The three constraints dictate how the SBU rods are able to position themselves relative to one another in the crystal framework.

By imposing these constraints, we can establish the framework of the unit cell by resolving the only two variables that can change for any given ligand. These optimization variables are visually distinguished in green in Fig. 2.5. Firstly, each of the r number of SBU rod may shift in the c-direction relative to its starting c-coordinate. Secondly, the unit cell will expand by some fractional value in both the a and b crystallographic directions. These variables we denote $\mathbf{dc} = \{dc_1, dc_2, \ldots, dc_r\}$ and F, respectively. The c lattice parameter cannot increase or decrease upon ligand exchange due to the connectivity of a 1-D rod MOF. Thus, given a linker of length L as an input, we can solve for the values of F and $\mathbf{dc}$ that satisfy the three constraints of Fig. 2.5. A continuum of (F, $\mathbf{dc}$) solutions exist; if F increases for example, $\mathbf{dc}$ must also change in order to maintain a fixed length between connection points on adjacent rods. The final task is to choose an (F, $\mathbf{dc}$) from the set of solutions that yields the most feasible crystal structure. This is accomplished by implementing a 3D-point cloud fitting of all carboxylic acid and phenolic oxygen atoms on the terminal benzene rings to their known connection coordinates on an SBU rod, as shown in Fig. 2.6. The (F, $\mathbf{dc}$) combination that

results in a point cloud fitting with the lowest root mean square displacement (RMSD) is selected as the optimal crystal framework. Thus the structural assembly method is fast and efficient, taking only a few seconds to generate all 61 hypothetical MOFs, three of which are visualized in Fig. 2.7.

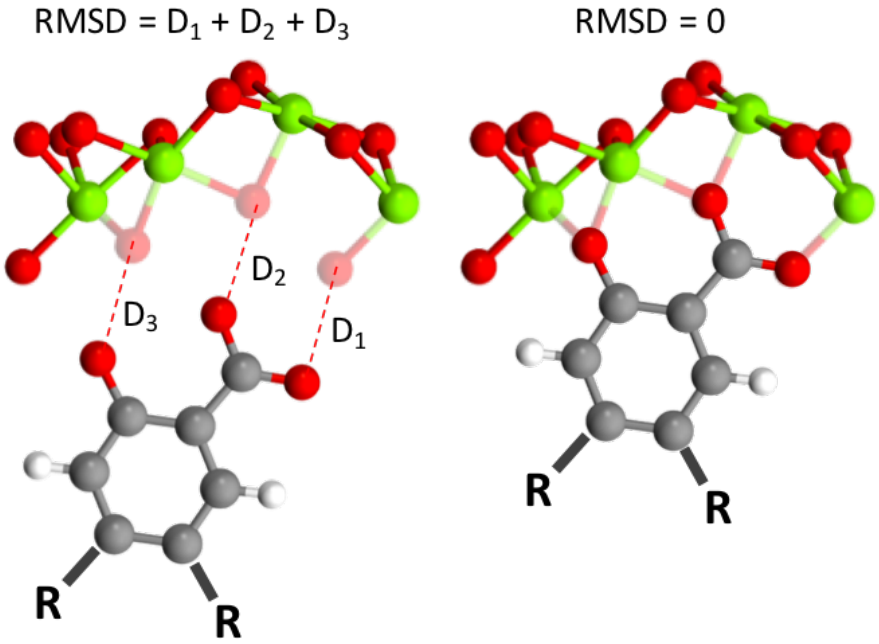


Figure 2.6: A schematic representation of the objective of the MOF-74 assembly algorithm. Using point cloud fitting to align the ligand as closely as possible to its connection points on the metal oxide rod, the optimal (F, **dc**) solution minimizes the RMSD between these connection points subject to the constraints shown in Fig. 2.5

.

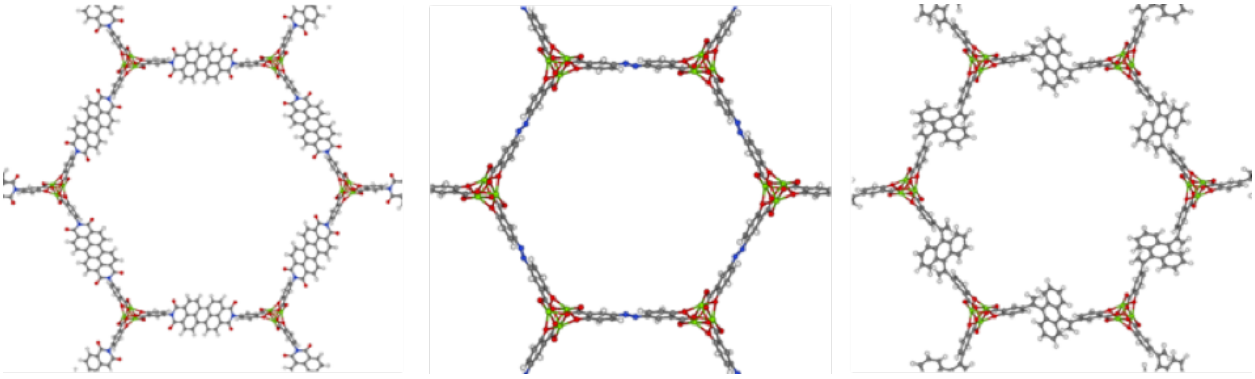


Figure 2.7: Visualization of three of the constructed hypothetical MOF-74 analogs.

Geometry relaxations and partial atomic charge analysis

The geometries, including both cell parameters and atomic positions, of the 61 hypothetical MOF-74 analogs generated from the aforementioned procedure are optimized at DFT level of theory, using the PBE functional,[63] together with Grimmes D3 dispersion correction [64] with the AxilrodTellerMuto three-body terms. The optimized lattice parameters of the original Mg-MOF-74, $a = 26.17$ Å and $c = 6.95$ Å , are in good agreement with experimental values ($a = 25.92$ and $c = 6.86$).[65] Similar settings were used in our previous studies on MIL-53 type MOFs [66, 67], in which good agreement with experimental structural and calorimetric data was obtained. All periodic DFT calculations were performed using the CP2K code which uses a mixed Gaussian/plane-wave basis set.[68, 69] We employed double-ζ polarization quality Gaussian basis sets and a 400 Ry plane-wave cutoff for the auxiliary grid, in conjunction with the Goedecker-Teter-Hutter pseudopotentials.[70, 71] A convergence threshold of 1.0×10^6 Hartree was used for all self-consistent field calculations. The structural optimizations were considered converged if the maximum force on all atoms falls below 0.534 kcal mol^{-1} $\mathring{A}^{-1}$ (4.5×10^{-4} Hartree Bohr^{-1}). All calculations were performed with the Γ-point approximation using a $1 \times 1 \times 2$ multiplication of the hexagonal primitive cell.

The partial atomic charge analysis was performed at the relaxed geometry, using the REPEAT method Campañá et al.,[58] which was recently implemented into the CP2K code based on a restrained electrostatic potential framework.[72] Our implementation of the REPEAT method in the CP2K code makes it possible to perform seamless geometry optimization and accurate partial atomic charge analysis of porous materials in the same software package. In addition, the parallel algorithms significantly accelerate the partial atomic charge analysis of large systems, for which the computational cost scales cubically with the cell length and density of grid points used for the electrostatic potential fitting. The REPEAT method calculates partial atomic charges from electrostatic potentials determined from DFT calculations, and only the grid points outside the van der Waals (vdW) radii of each atom were included in the fitting. In our calculations, we have used the vdW radii from the Universal Force Field (UFF),[73] which were also used by Campañá et al. in their original paper on the REPEAT method, with a scaling factor of 1.0. No other restraint was imposed during the REPEAT charge analysis. A detailed comparison of the derived charges from our implementation and from previous calculations are shown in Section 2.3.

Porosity characterizations and GCMC simulations

Geometric structure analysis of the MOF-74 analog set was completed using the Zeo++ software suite.[74] A probe radius of 1.65 Å which corresponds to the kinetic diameter of CO_2 was utilized to characterize the probe accessible surface area, the probe accessible volume, the largest included sphere, and the largest free sphere of all hypothetical structures. GCMC simulations at flue gas temperature of 313 K were utilized to measure the CO_2 adsorption capabilities of the MOF-74 analog library. Force field parameters were obtained

from previous studies [75] in which DFT single-point energy calculations were performed along a guest molecules approaching path of minimal repulsion to each unique atom type in MOF-74.

The energies were parameterized with the Buckingham potential to obtain highly accurate pairwise interaction parameters that are able to describe the repulsive interaction of CO_2 with the excess electron density at an open metal site, an essential requirement for capturing the adsorption behavior of CO_2 in MOF-74. These pairwise parameters are available for the atom types present in the original MOF-74 structure, however all MOF-74 analogs contain a unique atom type in the crystal that is not present in the original structure because all 61 ligands have more atom types than the original DOBDC linker. UFF was utilized to describe non-bonded, pairwise interactions between CO_2 and these atom types not found in the original MOF-74 structure.

Prioritization of synthetic targets

The olsalazine molecule was identified as one of the commercially available ligands in the set of hypothetical MOF-74 analogs and was therefore targeted for experimental synthesis.

2.3 High-throughput evaluation of hypothetical MOF-74 Analogs for CO_2 capture

Quality of structural assembly

Our geometric optimization routine was executed for each of the 61 potential linkers found within the PubChem Compounds database to generate a new hypothetical MOF-74 structure. The accuracy of this methodology in generating reasonable lattice geometries can be evaluated by observing the unit cell parameters both before and after DFT relaxation. Fig. 2.8 illustrates how DFT relaxation very marginally alters the lattice cell parameters. The accuracy of the generated crystal structures can be attributed to our pre-screening of linkers to find low energy conformers. Since the new a (and b) lattice cell parameters are dependent on the length of the proposed linker, using an energy minimized linker conformation results in a crystal structure that changes minimally following DFT relaxation. The error in the c-parameter tends to be slightly larger than the error in the a-parameter since it is constant in the crystal assembly step and cannot be pre-optimized by ligand conformational analysis.

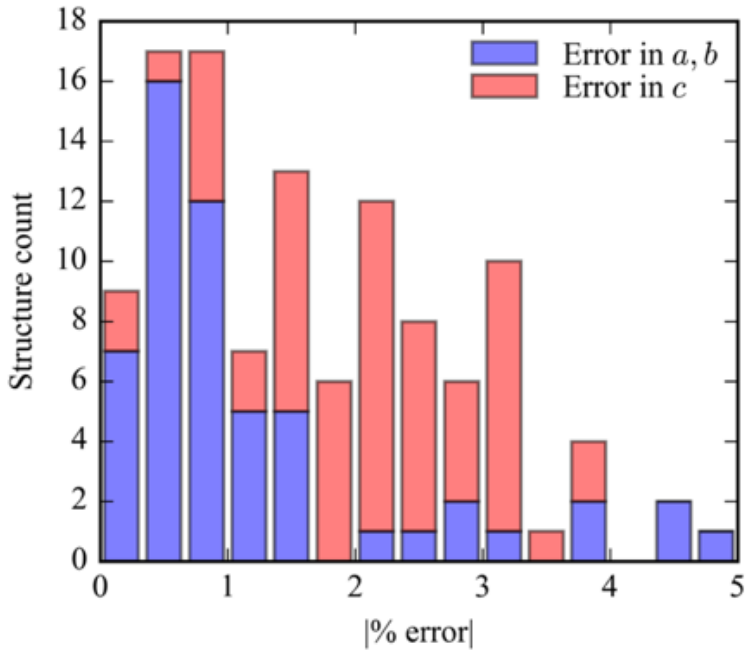


Figure 2.8: The absolute % error in the magnitudes of each lattice cell parameter from pre-DFT relaxed structures. DFT optimization results in very little change in the lattice cell parameters for all structures.

Geometric analysis of hypothetical structures

An interesting comparison can be drawn between the CO_2 accessible volume (AV) and the CO_2 accessible surface area (ASA), shown in Fig. 2.9a. After the volume of the unit cell reaches a critical volume of around 10,000 Å^3, linkers are able to produce variations in ASA because the length of the inserted linker permits variations in molecular structure, such as functional side groups or ring twisting in the interior of the linker, that still satisfy the constrained linker environment of the MOF-74 topology. For example, structures in the 10,000 Å^3 range have a 50% variation in their range of ASA, whereas at volumes below 10,000 Å^3 there is a much narrower, almost negligible distribution in ASA. This presents unique opportunities to make 1-D apertures that are accessible to extremely large guest molecules whose channel walls are much rougher than the original MOF-74 structure. These variations in channel wall shape, combined with the extremely large free spheres shown in Fig. 2.9b, suggest that the capture and separation of larger molecules could be a useful application for these analog structures. MOFs with such a large aperture have been previously synthesized,[53] and indeed one analog in our set, the novel Mg_2(olsalazine), was synthesized and characterized. The validation of this structure is described later in the manuscript. Due to the close packing of linkers in the crystallographic c-direction, all analogs strictly maintain channels with a dimensionality of one since the largest included sphere and largest free

spheres are almost exactly equivalent for all structures.

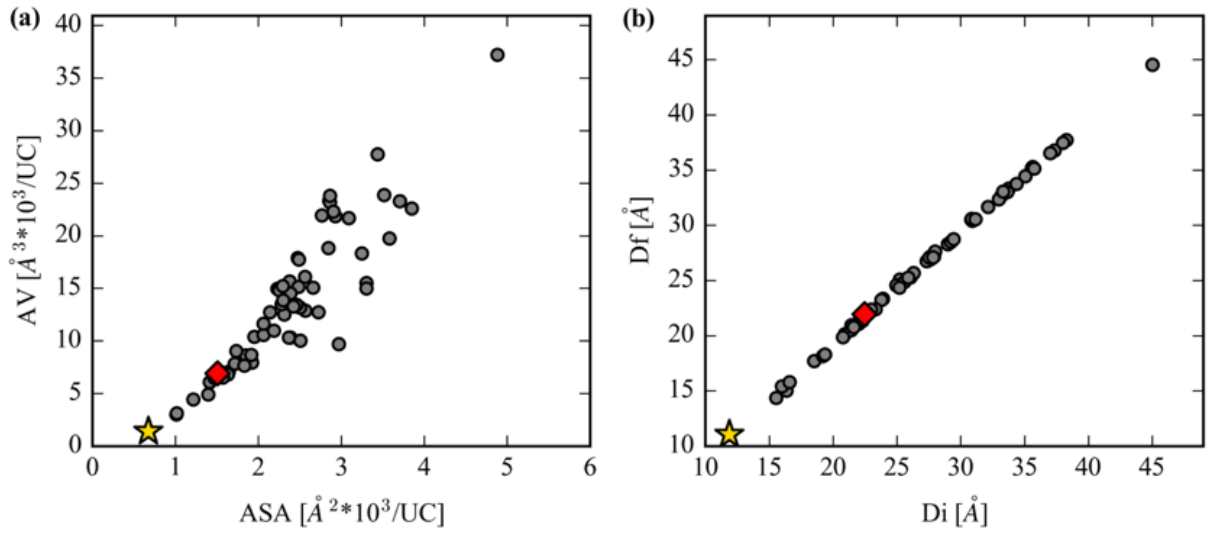


Figure 2.9: (a) The volume accessible to a 3.3 Å diameter probe versus the accessible surface area. (b) The free sphere, D_f, plotted against the largest included sphere, D_i, for the hypothetical MOF-74 library.

Benchmarking simulated CO_2 adsorption with previous results

Before using GCMC simulations to study CO_2 adsorption in the hypothetical set of analogs and gauge their potential to reduce the parasitic energy [76] of carbon capture, we need to benchmark our predicted adsorption in the original MOF-74 framework to previous results. A comparison of the partial charges between previously published work (VASP/REPEAT) [75] and this work (CP2K/REPEAT) are shown in Table 2.1 for the original MOF-74 structure.

The electrostatic potential generated by CP2K results in REPEAT derived charges that are 6% (Mg atom) and 8% (Oa atom) different than previous work. This difference only results in a small over-prediction of the CO_2 loading at low pressures on the CO_2 adsorption isotherm as seen in Fig. 2.10, thereby validating the approach we have taken to relax structures and obtain partial charges. We note small differences in the derived REPEAT charges exist. This is because the electrostatic potentials are sensitive to computational settings, and different codes employ different basis sets and pseudopotentials. In particular, for the implementation of the REPEAT method in CP2K, in order to use the same integration method and grid as those used for the calculation of the DFT electrostatic potential, Gaussian charge densities with a fixed width are used in place of point charges. Nevertheless, we do not expect these technical differences will result in qualitatively different pictures on the calculated adsorption isotherms.

Table 2.1: Summary of partial atomic charges derived from VASP (previous work) and CP2K (this work). The naming convention of atom types can be found in Ref. [75].

Atom type	VASP/REPEAT	CP2K/REPEAT
Mg	1.56	1.66
Oa	-0.9	-0.92
Ob	-0.75	-0.76
Oc	-0.9	-0.97
Ca	0.9	0.88
Cb	-0.31	-0.32
Cc	0.46	0.48
Cd	-0.23	-0.26
H	0.19	0.21

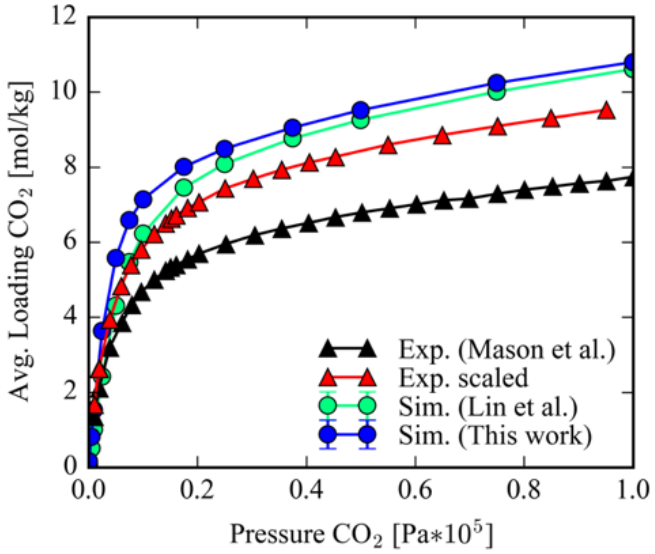


Figure 2.10: The CP2K relaxed structure and REPEAT derived atomic partial charges were validated by comparison of the CO_2 adsorption isotherm in the original MOF-74 structure with the previously published results of Lin et al. The experimental isotherm of Mason et al. is scaled due to 80% accessibility of the material.[50]

CO_2 capture in MOF-74 analogs

We can now proceed to screen the library of hypothetical analogs for CO_2 capture potential. The Henry coefficient versus AV plotted in Fig. 2.11a shows a declining trend but with significant scatter. Many of the structures achieve comparable Henry coefficients to the original structure even with a 6-fold increase in the AV. This is not surprising as the Henry coef-

ficient should not suffer significantly from simply increasing the pore size (see the following section for a toy model that predicts this trend), yet it confirms that many of the analogs can achieve the remarkable CO_2 uptake of the original MOF-74 in the limit of infinite dilution. At higher pore volumes the gravimetric uptake of CO_2 drops significantly, however. The gravimetric and volumetric dilution of the open-metal sites by inserting extended ligands into the framework is not compensated by the addition of any binding sites as strong as the under coordinated metals. Fig. 2.11b clearly demonstrates the decrease in performance in CO_2 loading after exchanging ligands to make a hypothetical MOF-74 structure. Furthermore, visualization of the CO_2 isotherms, color-coded by each structures accessible volume in Fig. 2.12b, reveals that most of the analog structures display condensation-like behavior rather than saturation-like behavior at high CO_2 pressures. Only the original MOF-74 structure and several of the smallest pore volume structures show any indication of approaching a saturation loading in the pressure range of 2.5 bar. Fig. 2.12a clearly demonstrates the superiority of the original MOF-74 structure for CO_2 capture applications.

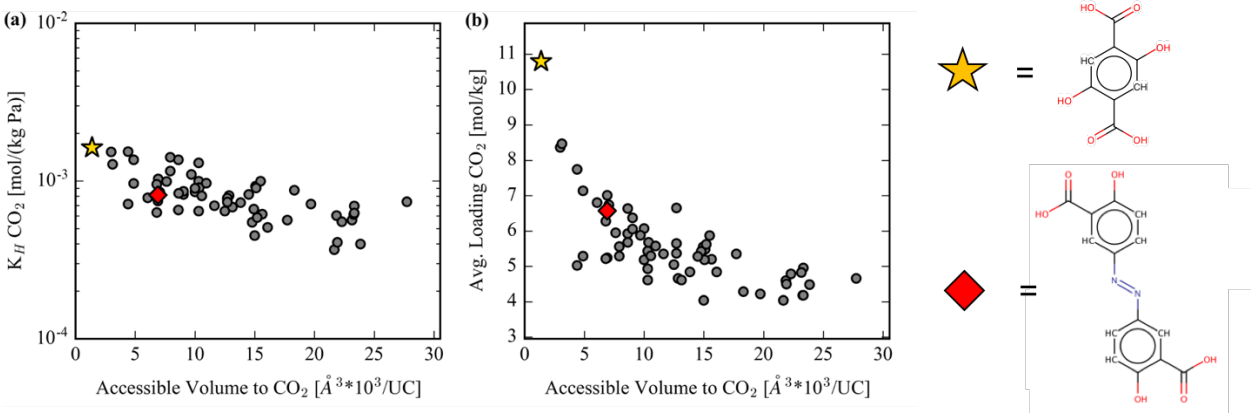


Figure 2.11: (a) CO_2 Henry coefficient in the hypothetical MOF-74 library. (b) Average CO_2 loading in MOF-74 analogs at 313 K and 1 bar. The gold star represents the original MOF-74 structure, and the red diamond represents the newly synthesized Mg_2(olsalazine).

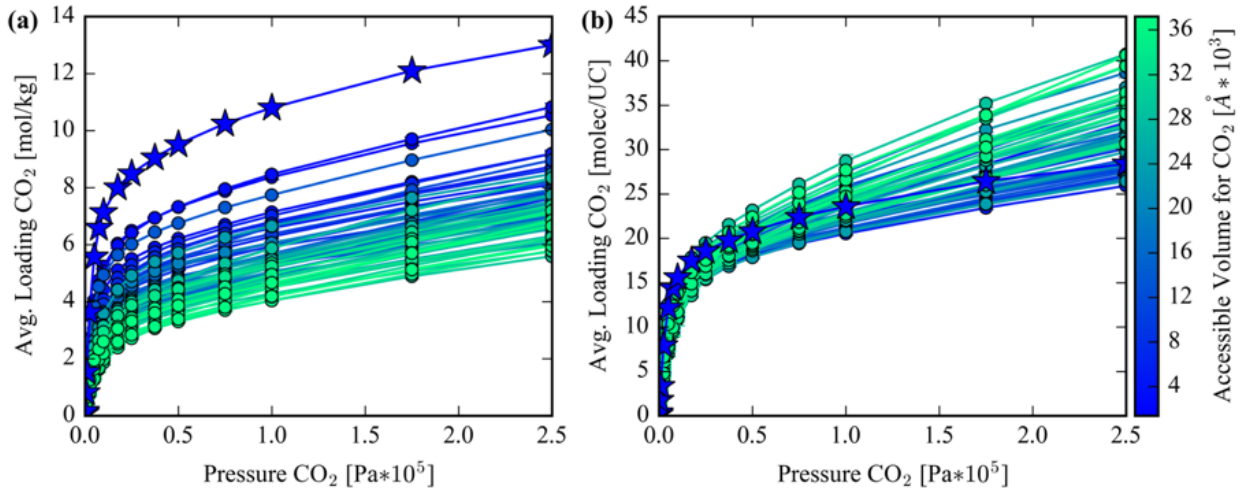


Figure 2.12: (a) CO_2 adsorption isotherms (in units of moles loaded per kilogram) in the hypothetical MOF-74 analog library color-coded by each structures CO_2 accessible volume. The isotherm with star markers represents the isotherm for the original MOF-74 analog. (b) CO_2 adsorption isotherms (in units of molecules loaded per unit cell) in the hypothetical MOF-74 analog library color-coded by each structures CO_2 accessible volume.

2.4 Modeling CO_2 Henry coefficient trends in the MOF-74 analogs

A toy model of CO_2 binding in MOF-74

The Henry coefficient, K_H, is computed via the Widom insertion technique, which is related to the ensemble averaged binding energy of a randomly inserted ghost particle inside the empty framework. This can be mathematically expressed via the following:

$$K_H = \beta \left\langle \exp\left(-\beta \Delta U^+\right) \right\rangle \tag{2.1}$$

where $\beta = (k_B T)^{-1}$ and ΔU^+ is the energy of the system after a random ghost particle insertion. This random ghost insertion of a CO_2 molecule is performed inside the MOF-74 framework, which can be approximated as occurring at either the very strongly binding open metal site (blue circles in Fig. 2.13a) or at the comparatively much weakly binding sites anywhere else inside the framework (empty space in Fig. 2.13a). Approximating the MOF-74 material as a simulation domain containing some volume fraction of strongly binding sites, f', and some volume fraction of weakly binding sites, f'', with binding energetics of U' and U'', respectively, the real MOF-74 material can be replaced by a very simple lattice model view of independent strong and weak binding sites shown in Fig. 2.13b. Based on

this approximate lattice model, the expectation value in Eqn. 2.1 reduces to the following analytical expression:

$$K_H = \beta \left[(1 - f'') \exp(-\beta U') + f'' \exp(-\beta U'') \right] \tag{2.2}$$

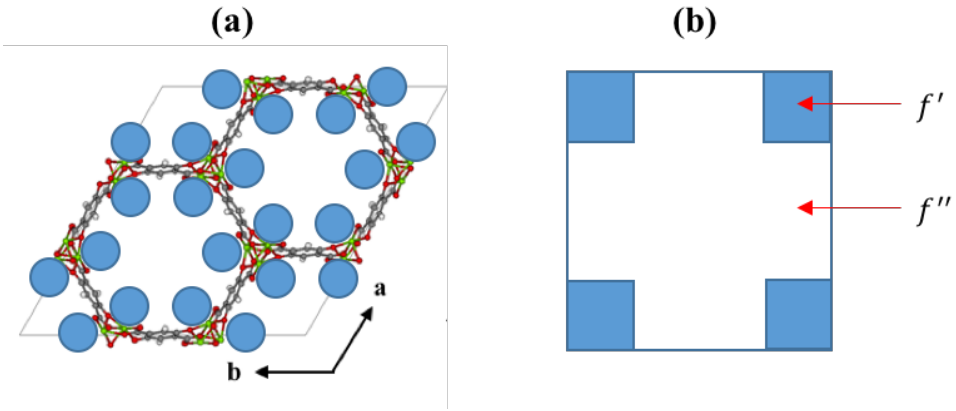


Figure 2.13: (a) A two binding site approximation of the MOF-74 material where the blue circles represent the volume corresponding to the strong binding sites and everywhere else is considered a weak binding site. (b) An equivalent representation of the approximated MOF-74 model containing some volume fraction of strongly binding sites, f', and weakly binding sites, f''.

Evaluating the model-based approximation of the Henry coefficient

We can plot the value of K_H as a function of f'' from Eqn. 2.2 to demonstrate the K_H trend shown in Fig. 2.11a. Since the volume occupied by strong binding sites remains constant in all MOF-74 analogs (the number of open metal sites per unit cell remains constant), the larger linkers that increase the AV correspond to an increase of f'' in our analytical model. As shown in Fig. 2.14, the trend in K_H dependence on f'' mimics the relatively weak dependence of the CO_2 Henry coefficient on the AV (Fig. 2.11a). It is not until extreme volumetric dilution of the strong binding sites that the Henry coefficient begins to drop precipitously, and such low volume fraction of strong binding sites were not generated in the library of hypothetical analogs.

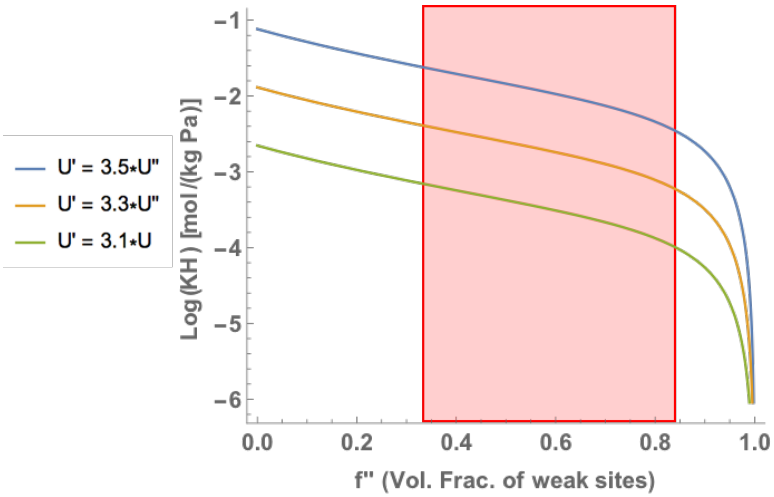


Figure 2.14: The solution to the analytic K_H expression in Eqn. 2.2 is shown as a function of the volume fraction of weak binding sites. Several solutions are shown for varying differences between the strong and weak binding site energies. The red region corresponds to a rough estimation of the values of f'' spanned by the library of MOF-74 analogs.

2.5 Synthesis and validation of a novel MOF-74 analog: Mg_2(olsalazine)

Synthesis procedure

The reaction of olsalazine sodium with $Mg(NO_3)_2{\cdot}6H_2O$ in a mixture of DMF:ethanol (ratio of 1:1) at 120 ^{o}C yields a homogeneous material based on yellow rod-type crystals as confirmed by scanning electron microscopy (SEM) images (Fig. 2.15a). The size of the isolated crystals was too small to permit structural resolution via single crystal X-ray diffraction. Despite this, PXRD confirms that the material is highly crystalline and the simulated pattern from the idealized structure of [Mg_2(olsalazine)] shows an excellent agreement with the experimental pattern of the as-made [Mg_2(olsalazine)$(DMF)_2$]·2DMF·$3H_2O$ (Fig. 2.15b). Additional characterization techniques (Ref. [21]) were also used to confirm the Mg_2(olsalazine) product. Mg_2(olsalazine) consists of hexagonal one-dimensional channels with dimensions of 23.3 Å (Mg-Mg distance, including van der Waals radii) which are larger than the channels resulted in Mg_2(DOPBDC) (18.4 Å) as shown in Fig. 2.15d. Similar to the original MOF-74 structure, concurrent work has interestingly shown that the olsalazine based MOF-74 structure can also be synthesized with different metals.[38]

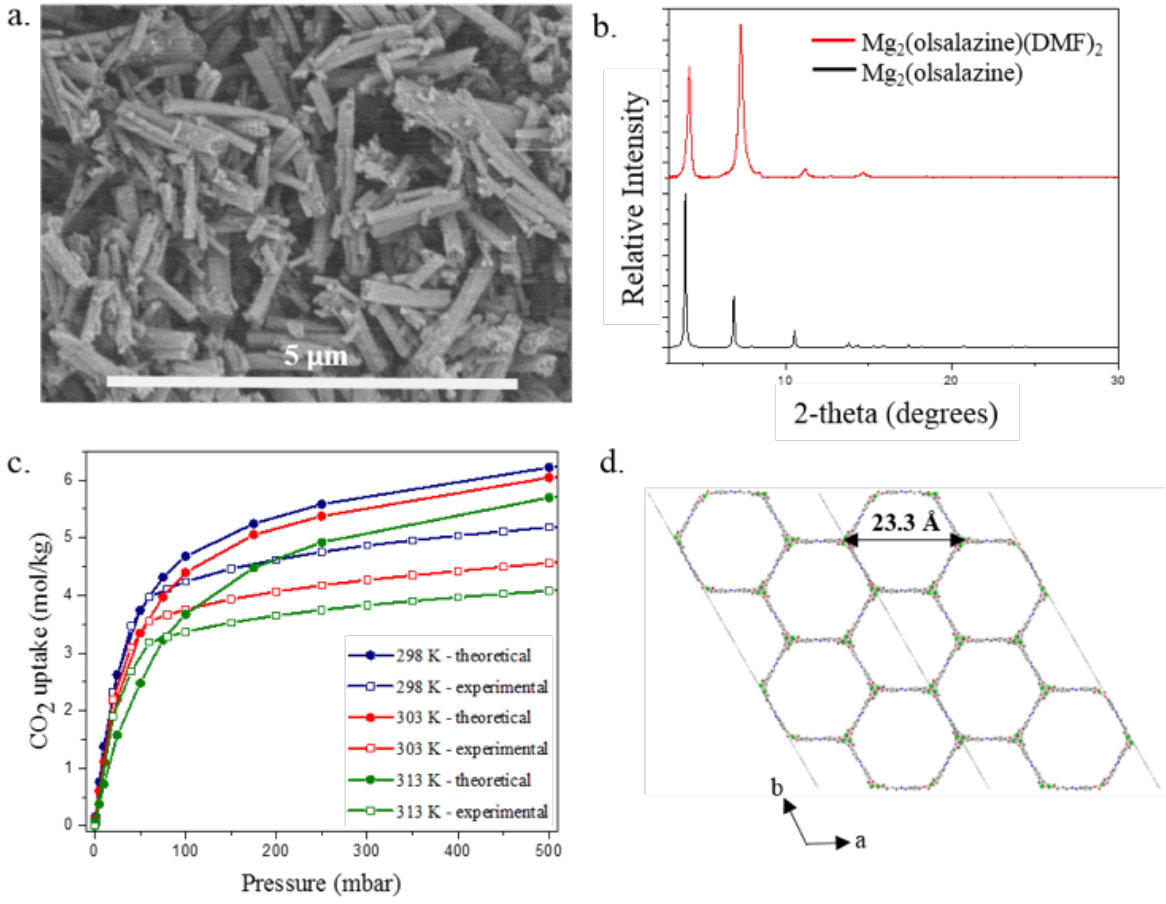


Figure 2.15: (a) SEM image of the as-made material reveals that the morphology of $[Mg_2(olsalazine)(DMF)_2]\cdot 2DMF\cdot 3H_2O$ crystals is based upon rods (scale bar: 5 μm); (b) comparison the simulated pattern from the idealized structure of $[Mg_2(olsalazine)]$ with the experimental pattern of the as made $[Mg_2(olsalazine)(DMF)_2]\cdot 2DMF\cdot 3H_2O$; (c) comparison of the theoretical and the experimental CO_2 adsorption isotherms collected at 298, 303 and 313 K and 1 bar; and (d) view of $Mg_2(olsalazine)$ along c-axis showing the formation of a 3-dimensional framework with hexagonal one-dimensional pores as it is observed in Mg-MOF-74.

Validation of CO_2 adsorption performance

CO_2 adsorption isotherms were experimentally measured to further characterize the material and to directly test our theoretical predictions of the CO_2 adsorption behavior. The theoretical and experimental isotherms are in good agreement, especially at low pressure, as shown in Fig. 2.15c. At higher pressures the experimental isotherms demonstrate that

the material has lower saturation loading than the theory predicts. The GCMC simulations are performed assuming a perfect, fully accessible crystal structure. From an experimental point of view, it is difficult to obtain a perfect crystal; defects that cause pockets of inaccessibility to CO_2 in the Mg_2(olsalazine) powder would give a lower saturation loading in the experimental material. In fact, the experimental BET surface area was calculated to be 2331 m^2/g, while the theoretical accessible surface area for N_2 was computed to be 2763 m^2/g with Zeo++, thereby suggesting that indeed we do not have 100% crystal accessibility in the Mg_2(olsalazine) powder. Ultimately, these experimental results demonstrate the theoretical methods used (*in silico* crystal design, DFT optimization, partial atomic charge analysis, force field parameterization, and GCMC simulations) not only correctly predict the crystal structure but also its CO_2 adsorption behavior.

2.6 Outlook: Discovering novel analogs of different classes of 1D rod MOFs

We have outlined a new, systematic method for enumerating MOFs exhibiting a 1-D rod topology. This method could be generally extended to systems other than the **etb** net and MOF-74, as we schematically visualize in Fig. 2.16. The geometric optimization approach permits the rapid and automated enumeration of analog structures, and our focus on searching for linkers in the PubChem database limited the structural assembly process to MOFs whose ligands are experimentally validated molecular species. The paucity of linkers found within PubChem (0.0001% of all molecules) that satisfy the strict geometric constraints of the MOF-74 topology provides a concise starting point for experimental efforts to synthesize analog structures. By providing an *in silico* methodology for generating 1-D rod MOFs as complex as MOF-74, we expand upon our ability to generate and predict properties of potentially interesting crystalline materials. Furthermore, we demonstrate the experimental validity of our *in silico* generated library by targeting the olsalazine linker for MOF synthesis. Mg2(olsalazine) was synthesized and its structural characterization presented excellent agreement with the *a priori* computationally predicted structure. Future work will focus on this targeted experimental synthesis of novel analogs of other 1D rod MOF structures.

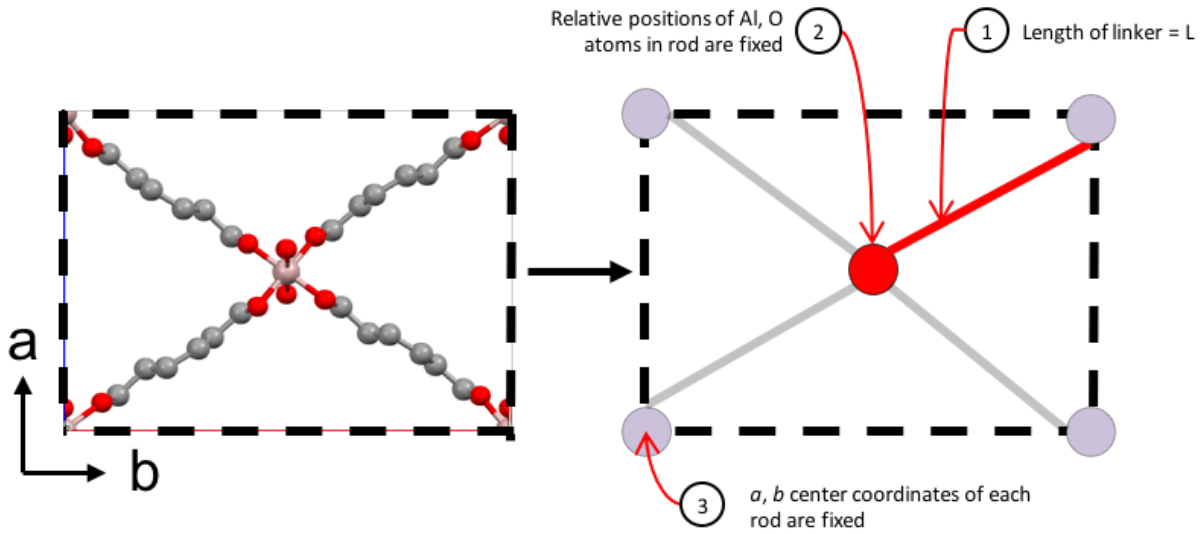


Figure 2.16: The simplification of MIL-53 to a set of geometric constraints and design variables shows how the building algorithm can be applied to MOFs other than MOF-74 in the etb net. In fact, this example demonstrates how no information about the underlying net of MIL-53 is necessary to set up a geometric optimization to enumerate analogs of the original structure.

Geometric analysis of the constructed hypothetical MOFs indicates that a wide range of surface areas are possible for a given accessible volume. This roughness and functionality of the channel walls, combined with the extremely large channel volumes, suggests that these MOFs could have greater potential for separations applications involving larger molecules than CO_2. Physical and chemical variations in the channel surface would provide regions of preferentially favorable interactions for adsorption of various large molecules, whereas for CO_2 capture, any new ligand in the analog library simply dilutes the open metal site density and does not provide any new, equally strong binding sites. Ultimately, we have introduced 61 hypothetical MOF-74 analogs and demonstrated their potential synthesizability. Future experimental and computational efforts are two-fold: synthesize additional hypothetical structures and investigate their potential for separation processes of large molecules.

Chapter 3

Rational materials design targeting high performance gas storage and separations

In the previous chapter, we presented the automated enumeration of a library of hypothetical MOF-74 analogs (also known as M_2(DOBDC) [M = Mg, Fe, Co, Ni, Zn]). In this chapter we focus on the *in silico* characterization and simulation of a rationally designed MOF-74 analog not represented in the aforementioned study. Hereon known as M_2(DHFUMA) [M = Mg, Fe, Co, Ni, Zn], this hypothesized material displays enhanced small molecule adsorption properties over the original M_2(DOBDC) series. Constructed from 2,3-dihydroxyfumarate (DHFUMA), an aliphatic ligand which is smaller than the aromatic 2,5-dioxidobenzene-1,4-dicarboxylate (DOBDC), the M_2(DHFUMA) framework has a reduced channel diameter, resulting in higher volumetric density of open metal sites and significantly improved volumetric hydrogen (H_2) storage potential. Furthermore, the reduced distance between two adjacent open metal sites in the pore channel leads to a CO_2 binding mode of one molecule per two adjacent metals with markedly stronger binding energetics. Through dispersion-corrected density functional theory (DFT) calculations of guest-framework interactions and classical simulation of the adsorption behavior of binary CO_2:H_2O mixtures, we theoretically predict the M_2(DHFUMA) series as an improved alternative for carbon capture over the M_2(DOBDC) series when adsorbing from wet flue gas streams. The improved CO_2 uptake and humidity tolerance in our simulations is tunable based upon metal selection and adsorption temperature which, combined with the significantly reduced ligand expense, elevates this material's potential for CO_2 capture and/or H_2 storage. The dynamical and elastic stability of Mg_2(DHFUMA) was verified by hybrid DFT calculations, demonstrating its significant potential for experimental synthesis. This chapter constitutes an unofficial adaptation of an article that appeared in an ACS publication [77]. ACS has not endorsed the content of this adaptation or the context of its use.

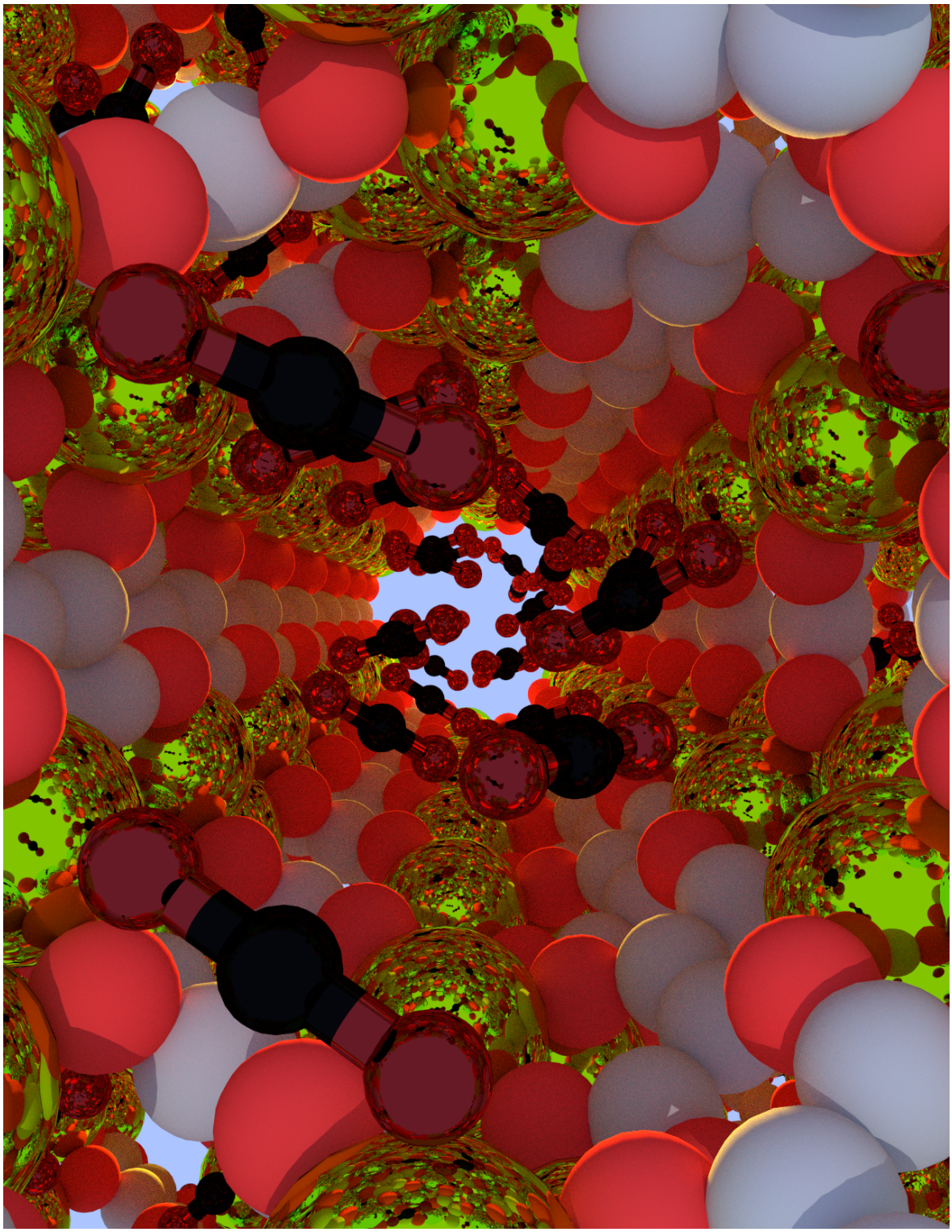


Figure 3.1: Artistic rendition of CO_2 adsorbing in Mg_2(DHFUMA)

3.1 Introducing a MOF with an ultra-high density of open metal sites

Despite the many advantages of MOFs that result from high tunability of chemistry and structure, the cost of MOF production is still a major factor that impedes their large-scale industrial applications. Apart from the capital investment in infrastructures, the cost of MOF production consists largely of raw materials (including metal salts and organic ligands) and processing, which includes but is not limited to non-reusable organic solvents and cost associated with activation. For MOF-74 with a molecular formula M_2(DOBDC) (M = Mg, Zn, Fe, etc and DOBDC = 2,5-dioxido-1,4-benzenedicarboxylate), the major cost of raw materials comes from organic ligand (i.e. DOBDC). Taking Mg-MOF-74 as an example, the cost of metal salts, usually $MgCl_2$, can almost be neglected, i.e. it accounts for only a small percentage of the expense of organic ligands (see Appendix A). Indeed, MOFs built from much cheaper raw organic ligands will need to be developed before they can be widely used in industry in large quantities. Generally speaking, larger and longer aromatic organic ligands are more expensive than smaller and shorter aliphatic ligands. However, the majority of the MOFs synthesized so far feature aromatic organic ligands, because the coordination-driven self-assembly of building blocks to produce porous crystalline MOFs requires the molecular precursor to be rigid and possess proper directionality.[78] Such properties are more likely to appear in conjugated organic ligands, e.g. DOBDC and BDC (BDC = 1,4-benzenedicarboxylate), both of which are frequently used in the synthesis of MOFs. On the other hand, most of the aliphatic ligands are flexible and do not have sites to form directional metal-ligand bonds, and they are less likely to form porous and crystalline solids with metal centres. Therefore, aliphatic ligands are rarely employed in MOF synthesis. Nevertheless, there are still several MOFs based on aliphatic ligands,[79–81] including the commercially available aluminium fumarate (Basolite A520).[82] However, there are no open metal sites in these MOFs, thereby limiting their CO_2 and H_2 storage potential at low and ambient pressures. It would be extremely useful to develop a MOF-74 analog featuring both open metal sites, which lead to enhanced adsorbate-adsorbent interactions and higher gas uptake at ambient pressure, and cheap aliphatic linkers, which lower the overall raw materials cost. To the best of our knowledge, all the MOF-74 analogs which have been experimentally synthesized so far were constructed from longer, aromatic organic linkers and are therefore likely to be more expensive with limited improvement on gas adsorption capacity in low to ambient pressure regimes.

Another popular approach to increase the gas adsorption capacity of MOFs is to synthesize MOFs with expanded pores and larger internal surface areas, e.g. by replacing the DOBDC linker in MOF-74 with longer linkers.[53] We investigated the effectiveness of pore expansion in MOF-74 analogs in one of our recent high-throughput screening studies,[21] whereby we developed a novel *in silico* crystal assembly algorithm that differed from previous approaches[18–20, 46, 47] to create a library of MOF-74 analogs which exhibit 1-D metal-oxide rod building units.[49] We found that the gravimetric uptake of CO_2 dropped

significantly in MOF-74 analogs with higher pore volumes due to the spatial and gravimetric dilution of the open metal sites which serve as the strong adsorption sites for CO_2 molecules. Thus the increased pore volume in these analogs (which results from construction with extended ligands) sacrifices gravimetric uptake and further complicates synthesis by introducing more complex organic molecule building units and by potentially reducing mechanical stability. Therefore larger pore sizes are not *always* desirable. While many efforts have been made to tune and improve upon the exceptional small molecule adsorption properties of the original MOF-74 framework,[54–56, 83, 84] we undertake a rational design approach to further improve the gas adsorption capabilities in MOF-74 analogs by increasing the density of open metal sites, e.g. by replacing the DOBDC linker with a smaller molecule. While DOBDC represents the smallest aromatic molecule that satisfies the topological requirement of MOF-74, an even smaller molecule can be identified from the thousands of aliphatic molecules which are smaller in size than DOBDC.

In this work, we rationally design *in silico* a MOF-74 analog based on a cheaper and commercially available aliphatic ligand, i.e. DHFUMA (DHFUMA = 2,3-dihydroxyfumarate), and simulate its H_2, CO_2 and H_2O adsorption properties, based on extensive previous work dedicated to describing the energetic interactions of small molecules in the MOF-74 framework.[75, 85–88] Namely, we predict significantly improved H_2 volumetric storage capacity, increased low pressure CO_2 adsorption, and higher CO_2:H_2O selectivity in the M_2(DHFUMA) series than the M_2(DOBDC) series. The cost (per mol) of DHFUMA is lower than that of DOBDC by more than 80% (see Appendix A), and the density of open metal sites in M_2(DHFUMA) is twice of that of M_2(DOBDC). Typical protocols used to synthesize M_2(DOBDC) have been tested and shown to result in a crystalline material that is not the desired M_2(DHFUMA) product (see Appendix A); however, calculation of the elastic constants and vibrational frequencies demonstrate the dynamical and mechanical stability of M_2(DHFUMA) and provide justification that the material can be synthesized. M_2(DHFUMA), if it can be synthesized in large quantities, has the potential to be a better candidate than M_2(DOBDC) for industrial applications including hydrogen storage and carbon capture.

3.2 Techniques for *in silico* evaluation of H_2 storage and CO_2 capture potential

In silico crystal design

Part of our recent work has focused on the *in silico* crystal design of 1-D rod MOFs.[21] The building blocks of these MOFs are embedded in three dimensional space by an optimization routine that is constrained by geometric rules that must hold for a 1-D rod MOF. Utilizing this method allows for facile substitution of DOBDC for DHFUMA into the MOF-74 framework and quickly creates an accurate starting crystal structure for DFT optimization. Fig. 3.2 demonstrates the analogous connectivity groups in DHFUMA and DOBDC. We

believe this to be the smallest possible ligand with which a MOF-74 analog can be constructed. Dispersion-corrected DFT optimization was performed to relax the M_2(DOBDC) and M_2(DHFUMA) frameworks and obtain partial atomic charges for each unique atom type in the framework.

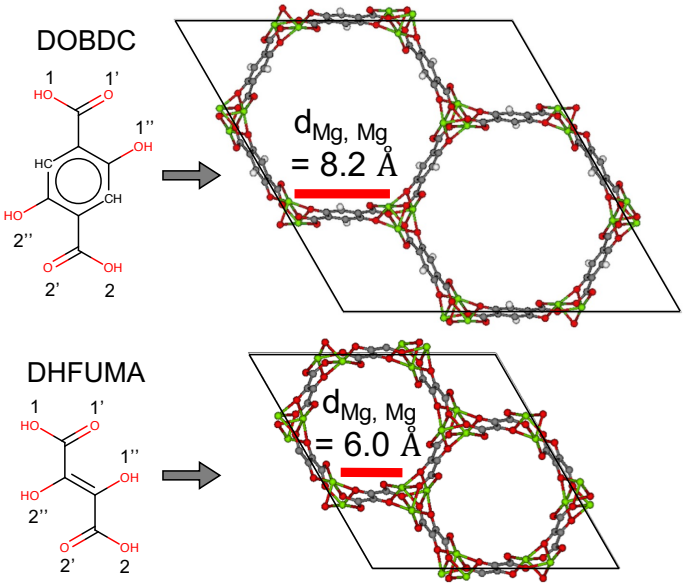


Figure 3.2: The DOBDC ligand and framework is visually compared to the DHFUMA ligand and framework. (1, 1', 1") oxygen atoms connect to one metal rod in the MOF-74 type framework and (2, 2', 2") connect to an adjacent metal rod. The distance between adjacent open metal sites is shown to be 2 Å smaller in the Mg_2(DHFUMA) framework.

DFT calculations

A majority of the periodic density functional theory calculations, including geometry and cell optimisations, have been performed using the CP2K code which uses a mixed Gaussian / plane-wave basis set.[68, 69] We have used both gradient corrected (i.e. PBE [63]) and hybrid density functional (i.e. PBE0 [89, 90] with 25% Hartree-Fock exchange) methods. It is known that a correct description of the dispersion interactions is important to predict the MOF structures and host-guest interactions in MOFs.[91, 92] In this work, we have used one of the most popular pairwise additive descriptions of the dispersion interactions as developed by Grimme et al., i.e. the D3 method [64] with the Axilrod-Teller-Muto three-body terms, in combination with the conventional PBE and PBE0 functionals. The same method was used in our previous work on MIL-53,[66, 67] UiO-66,[93] and MOF-74 [21] types of MOFs, and we achieved very good agreement between theory and experimental results on structures and calorimetric measurements. We note that a hybrid functional is necessary here to provide a correct description of the electronic structures and host-guest interactions of MOF-74

materials featuring M^{2+} cations with unpaired electrons, including Mn^{2+}, Fe^{2+}, Co^{2+}, Ni^{2+} and Cu^{2+}. The Hartree-Fock exchange calculations, which are part of the hybrid DFT functional PBE0, were performed and significantly accelerated using the auxiliary density matrix method (ADMM),[94] which enables us to consider relatively large systems (with the largest system containing 489 atoms) at hybrid DFT level. The partial atomic charge analysis was performed using the REPEAT method proposed by Campañá et al.,[58] which was recently implemented into the CP2K code based on a restrained electrostatic potential framework.[72] The REPEAT method calculates partial atomic charges from electrostatic potentials determined from DFT calculations, and only the grid points outside the van der Waals radii of each atom were included in the fitting. We have used partial atomic charges determined using the REPEAT scheme in our recent work on MOF-74, in which very good agreement was obtained between theory and experiment on the adsorption isotherms of CO_2 molecules.[21] The vibrational frequency and elastic constants calculations were performed using the CRYSTAL code[95, 96] with the B3LYP hybrid functional.[97] More details of the calculations are included in Appendix A.

Classical simulations and pore characterization

A critical component in the classical molecular simulation of nanoporous materials is the parameterization of classical potential energy functions (or force fields) that can accurately describe the energetics of guest-adsorbate systems. Many times off-the-shelf force fields such as UFF [73] or Dreiding [98] are used in lieu of a more accurate alternative, but this approach breaks down with MOFs that contain complex electronic structure features such as open metal sites.[75] Several different approaches have been used specifically to generate force fields which successfully describe gas interactions in MOF-74 type frameworks which contain these open metal sites.[75, 99, 100] Pham et al. used a many body polarization approach for the first time to classically capture the complex H_2 - open metal site potential energy interactions in the Mg_2(DOBDC) framework and later extended their force field development to the entire metal series.[88, 101] In this work, the model of Pham was used to model H_2 adsorption in Mg_2(DOBDC) and Mg_2(DHFUMA) and we refer the reader to these publications for further details. The Grand Canonical Monte Carlo (GCMC) simulations of H_2 with many body polarization were calculated with the RASPA2 software package.[102]

Mercado et al. used a recently developed approach to parameterize the potential energy surface for the isoreticular series M_2(DOBDC) [M = Mg, Mn, Fe, Co, Ni, Zn] such that classical molecular simulation of CO_2 and H_2O could be performed for the entire metal series for the first time.[87] The parameterization was performed by calculating DFT single point energies along the path of minimum repulsion between unique guest-host pairwise types to accurately capture the repulsive behavior between the guest and the excess electron density at the open metal sites. We adopt this force field parameterization for studying the adsorption properties of CO_2 and H_2O in DHFUMA and refer the reader to this publication for specific details and the parameters themselves. The unique types assigned to each atom in the DHFUMA crystal structure and their correspondence to the atom types of Mercado's

force field are shown in Appendix A, in addition to a justification for the transferability of the force field. Grand Canonical Monte Carlo (GCMC) simulations were executed to calculate adsorption isotherms and isosteric heats of adsorption of the frameworks under consideration. An annealing minimization scheme was used to determine the classical binding energy of adsorbates in all analogs. In this scheme, an NVT ensemble Monte Carlo simulation consisting of one adsorbate molecule is successively quenched from T = 298 K to T = 1 K. The potential energy of the final configuration in the T = 1 K simulation then corresponds to the classical binding energy. The porosity characterization of all frameworks was performed with the Zeo++ application using the high accuracy settings.[74, 103]

3.3 Physical characterization of the M_2(DHFUMA) series

Comparison of the M_2(DHFUMA) vs M_2(DOBDC) unit cells

We list our calculated lattice parameters of the all the MOFs considered in this work in Table 3.1, and we compare the data on M_2(DOBDC) with available experimental results (taken from Ref. [85]; see references therein) from which we find that the errors of our theoretically predicted lattice parameters of M_2(DOBDC) are within ∼1%. We also find that for the same metal, the a lattice parameter of M_2(DHFUMA) is proportionally smaller than that of M_2(DOBDC) by 27∼29%, and the c lattice parameter of M_2(DHFUMA) is almost the same as that of M_2(DOBDC), with the biggest difference to be only 0.16 Å (Cu and Zn). Indeed, the decreased unit cell volume of M_2(DHFUMA), i.e. by ∼50% in comparison with M_2(DOBDC), is mainly due to the shortening of the lattice parameter along the a and b axes, and a direct result of that is the doubling of the volumetric density of open metal sites. We will see that the shortening of the a lattice parameter and the doubling of the density of open metal sites in M_2(DHFUMA) have a significant effect on the optimal binding configuration of CO_2 in M_2(DHFUMA).

Porosity characterization

The channel geometry of the Mg analogs of the DHFUMA and DOBDC series were analyzed by Zeo++ to demonstrate the differences in porosity which are later shown to have a significant impact on the adsorption properties of the two frameworks. A probe radius of 1.65 Å was used which corresponds to the kinetic diameter of CO_2. Table 3.2 summarizes these important geometric quantities. We note that the distance between the centers of two metal rods opposite each other in a single hexagon of DHFUMA, i.e. the approximate diameter of a single channel, is equal to 12.6 Å. When accounting for the Van der Waals radii of the framework atoms in DHFUMA, the largest free and included spheres are close to half of this diameter as shown in Table 3.2. Interestingly, the typical diameter of single wall carbon

Table 3.1: Lattice parameters (in Å) of M_2(DOBDC) and M_2(DHFUMA) from theory and experiment.

Metal	DHFUMA (theory)		DOBDC (theory)		DOBDC (expt)		a/a'
	a	c	a'	c'	a'	c'	(theory)
Mg	18.86	6.88	26.17	6.95	25.89	6.87	72%
Mn	18.70	7.14	26.22	7.01	26.23	7.04	71%
Fe	18.93	6.75	26.11	6.85	26.10	6.85	73%
Co	18.67	6.77	25.91	6.82	25.89	6.81	72%
Ni	18.59	6.65	25.73	6.75	25.72	6.74	72%
Cu	18.85	6.13	25.84	6.29	26.00	6.26	73%
Zn	19.01	6.72	26.18	6.88	25.93	6.84	73%

nanotubes (SWNTs), depending on the chirality indices, can range from 6.2 (n+m = 8) to 12.2 (n+m = 18) for the lowest energy tube for each combination of chirality indices.[104]

Table 3.2: Accessible Surface Area (ASA), Accessible Volume (AV), Largest Included Sphere (D_I), Largest Free Sphere (D_F), open metal site volumetric density (ρ_{Mg}), and open metal site weight percent of two frameworks: Mg_2(DHFUMA) vs Mg_2(DOBDC)

Ligand	ASA [m^2/g]	AV [cm^3/g]	D_I [Å]	D_F [Å]	ρ_{Mg} [$Mg/Å^3$]	wt% Mg [%]
DOBDC	1782	0.350	11.8	11.1	0.0044	20.0
DHFUMA	1043	0.095	7.6	6.3	0.0084	25.2

We have therefore designed a MOF with a channel geometry that is essentially comparable to a SWNT but with a significantly higher degree of chemical diversity due to the presence of oxygen atoms and open metal sites decorating the inside of the channel (see Fig. A.1. The proximity and increased volumetric density of open metal sites in the DHFUMA structure compared to the DOBDC structure will later be shown to result in a new CO_2 binding configuration. The volumetric densities of open metal sites in Mg_2(DHFUMA) and Mg_2(DOBDC) are 0.0084 $Mg/Å^3$ and 0.0044 $Mg/Å^3$, respectively. One in every six atoms in DHFUMA is an open metal site, whereas one in every nine atoms in DOBDC is an open metal site. With nearly two times the volumetric metal site density and one and a half times the molar metal site density of DOBDC (in addition to the reduced inter-atomic distance between adjacent Mg atoms in each channel), DHFUMA contains a spatial configuration of open metal sites that is more favorable for hydrogen storage and CO_2 capture. Additional pertinent crystallographic data for Mg_2(DOBDC) and Mg_2(DHFUMA) is shown in Table A.4.

Predictions of material stability

To verify whether M_2(DHFUMA) is stable and therefore has the potential to be synthesized experimentally, we calculated the vibrational frequencies and elastic constants, taking Mg_2(DHFUMA) as an example. Our calculated vibrational frequencies and the full elastic matrix of Mg_2(DHFUMA) are shown in Appendix A. We find all the vibrational modes of Mg_2(DHFUMA) have positive frequencies, demonstrating its dynamical stability. We further verify the elastic stability of Mg_2(DHFUMA) against the Born stability criteria,[105] and we find the calculated elastic constants of Mg_2(DHFUMA) satisfy all the *necessary and sufficient* stability conditions (see Appendix A),[106] demonstrating Mg_2(DHFUMA) to be mechanically stable. We expect M_2(DHFUMA) based on other metals to have the same behavior and suggest these materials have the potential to be synthesized in future experiments. The results of the mechanical and dynamical stability calculations are not surprising, especially since the metal oxide rod M-O coordination environment is identical to M_2(DOBDC) and since DHFUMA is an experimentally validated ligand with a fully conjugated backbone exhibiting a planar geometry between the two connection groups (see Fig. 3.2). The synthetic difficulties arise in finding the necessary reaction conditions to yield the correct crystalline M_2(DHFUMA) product, the details of which are elaborated in Ref. [77].

3.4 Enhanced H_2 storage potential of Mg_2(DHFUMA)

The doubling of the volumetric density of open metal sites results in a factor of two increase in the simulated volumetric H_2 storage capacity of Mg_2(DHFUMA) over Mg_2(DOBDC) at cryogenic temperatures. Utilizing the many body polarization scheme implemented in the RASPA2 package and the polarizable model of Pham et al. [88] to compute H_2 potential energy interactions in Mg_2(DOBDC), we simulate the adsorption isotherms of both Mg_2(DOBDC) and Mg_2(DHFUMA) at 77 K. We assumed that the force field is transferable and adopt all model parameters of Pham with the exception of the frameworks' partial atomic charges for which we use the values derived from our REPEAT analysis which are summarized in Appendix A. Fig. 3.3(a) demonstrates a good agreement of our isotherm with the theoretical isotherm of Pham et al. and the experimental isotherm of Dietzel et al. (data extracted from Ref. [88] and Ref. [107], respectively) for Mg_2(DOBDC) at 77 K. Our simulated isotherm as generated by RASPA2 slightly over predicts the gravimetric uptake (by $\sim$20% at 1 bar) shown by the experimental results and simulated by Pham et al. which we further discuss in Appendix A.

Since the weight compositions of Mg in DHFUMA (25.2 wt %) and DOBDC (20.0 wt %) differ slightly, the amount of H_2 loaded per framework weight in DHFUMA is marginally better than DOBDC at low pressure but does not represent a remarkable improvement as shown in Fig. 3.3(a). The strong H_2 - open metal site interactions dominate the adsorption at low temperatures and pressures and weak H_2 - H_2 interactions are not sufficient to provide the strong cooperative binding effects observed with CO_2 which are later discussed in the

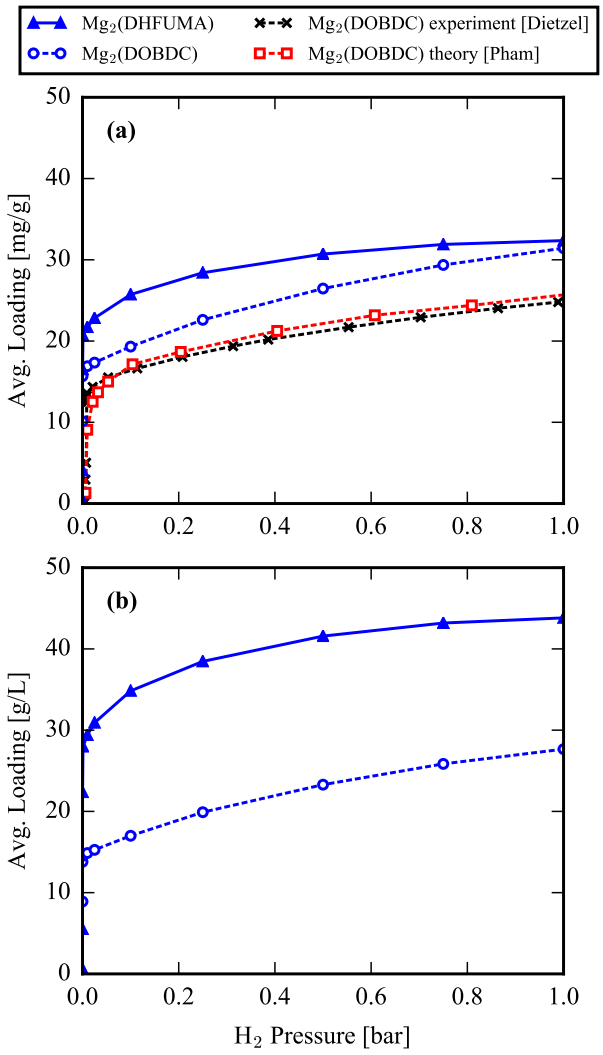


Figure 3.3: H_2 isotherms computed at T = 77 K. (a) Mg_2(DOBDC) isotherms computed in this work, by Pham et al. (extracted from Ref. [88]), and measured by Dietzel et al. (extracted from Ref. [107]) and the predicted Mg_2(DHFUMA) assuming a transferable force field. Isotherms are in units of amount adsorbed per framework mass. (b) Mg_2(DHFUMA) and Mg_2(DOBDC) simulated isotherms from this work in units of amount adsorbed per framework volume.

section on enhanced CO_2 heat of adsorption. In other words, H_2 gravimetric adsorption is not significantly improved in DHFUMA at low pressures, and the framework displays H_2 saturation behavior at significantly lower pressures than in DOBDC as one would expect from the reduced channel volume. Nevertheless, the advantage of Mg_2(DHFUMA) for H_2 storage lies exactly in this reduced channel volume and the doubling of volumetric open metal site density. As can be seen from Fig. 3.3(b), the H_2 storage capacity on a volumetric basis

(in which the amount loaded is expressed per total volume of adsorbent) is approximately twice that of Mg_2(DOBDC). Not only would a Mg_2(DHFUMA) based storage device require half the volume to achieve approximately the same H_2 storage by weight percent, the ligand is also drastically cheaper. At the cryogenic temperature of 77 K and extremley low pressure of 0.5 bar, Mg_2(DHFUMA) is predicted to achieve a volumetric uptake of 41.5 g H_2/L which is sufficient to surpass the DOE's 2020 H_2 volumetric storage target of 40 g H_2/L. We note that Mn-BTT (BTT^{3-} = 1,3,5-benzenetristetrazolate), one of the best performing MOFs for volumetric hydrogen uptake, has been reported to achieve uptake of 43 g H_2/L but at the higher pressure of 1.2 bar.[25]

3.5 Enhanced CO_2 affinity in the M_2(DHFUMA) series

DFT predicted binding geometries and energies

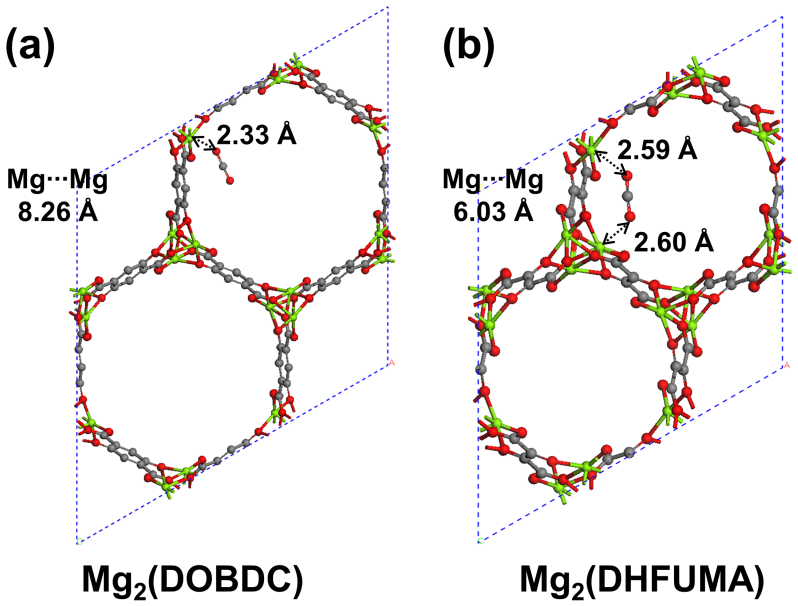


Figure 3.4: DFT optimised binding configurations of CO_2 in (a) Mg_2(DOBDC) and (b) Mg_2(DHFUMA).

Table 3.3: Binding energies (in kJ/mol) and relevant $O\cdots M$ binding distances (in Å) of CO_2 in M_2(DOBDC) and M_2(DHFUMA) from theory and experiment.

Metal	DHFUMA (theory)		DOBDC (theory)		DOBDC (expt)		E_b - E'_b
	E_b	$d_{O\cdots M}$	E'_b	$d'_{O\cdots M}$	E'_b	$d'_{O\cdots M}$	(theory)
Mg	50.1	2.59 / 2.60	41.3	2.33	43.5	2.27	8.8
Mn	40.5	2.64 / 2.75	29.2	2.57	31.7	2.51	11.3
Fe	40.9	2.56 / 2.76	30.0	2.43	33.2	2.29	10.9
Co	41.1	2.49 / 2.79	29.3	2.43	33.6	2.23	11.8
Ni	46.1	2.46 / 2.69	34.8	2.32	38.6	2.29	11.3
Cu	32.0	2.69 / 2.80	19.9	2.74	22.1	2.86	12.1
Zn	37.6	2.76 / 2.83	31.3	2.66	26.8	2.43	6.3

Table 3.4: Binding energies (in kJ/mol) and relevant $O\cdots M$ binding distances (in Å) of H_2O in M_2(DOBDC) and M_2(DHFUMA) from theory.

Metal	DHFUMA (theory)		DOBDC (theory)	
	E_b	$d_{O\cdots M}$	E'_b	$d'_{O\cdots M}$
Mg	87.4	2.19	88.6	2.16
Mn	73.5	2.26	73.3	2.29
Fe	77.5	2.20	77.1	2.21
Co	79.4	2.17	78.9	2.17
Ni	89.2	2.12	89.4	2.11
Cu	68.1	2.22	62.5	2.24
Zn	74.9	2.22	68.4	2.23

Taking Mg as an example, we show our theoretical optimised binding configuration of CO_2 in Mg_2(DOBDC) and Mg_2(DHFUMA) in Figs. 3.4a and 3.4b, respectively. From Fig. 3.4, we can find that a single CO_2 molecule has very different binding modes in Mg_2(DOBDC) and Mg_2(DHFUMA). In Mg_2(DOBDC), one terminal oxygen of CO_2 binds to Mg of Mg_2(DOBDC) with a short binding distance of 2.33 Å while the other terminal oxygen of CO_2 is aligned with the DOBDC linker and points towards the open pore space of Mg_2(DOBDC). However, in Mg_2(DHFUMA), because of the much shorter inter-chain $Mg\cdots Mg$ distance (i.e. 6.03 Å in comparison with 8.26 Å in Mg_2(DOBDC)), both terminal oxygen atoms of CO_2 are able to bind to two neighbouring Mg^{2+} cations simultaneously, with similar binding distances (i.e. 2.59∼2.60 Å). Such a unique binding mode results in a much enhanced binding energy of CO_2 in Mg_2(DHFUMA), i.e. 50.1 kJ/mol, which is 20% (8.8 kJ/mol) stronger than that in Mg_2(DOBDC). We show a detailed comparison of the binding energies and relevant $O\cdots M$ binding distances of CO_2 in M_2(DOBDC) and M_2(DHFUMA) with different metals in Table 3.3, in which we also list available experimental data on CO_2 adsorption in M_2(DOBDC).[108] We further looked at the adsorption of a

single H_2O molecule in both M_2(DOBDC) and M_2(DHFUMA) with different metals, and we show a detailed comparison of the binding energies and relevant O$\cdots$M binding distances in Table 3.4. Interestingly, the binding energies of a single H_2O molecule in M_2(DOBDC) and M_2(DHFUMA) with the same metal are almost identical. Taking Mg as an example, the binding energies of H_2O are 88.6 and 87.4 kJ/mol in Mg_2(DOBDC) and Mg_2(DHFUMA), respectively. This is because H_2O has only one central oxygen, and it interacts with both Mg_2(DOBDC) and Mg_2(DHFUMA) through a single-contact O$\cdots$Mg interaction. Comparing the whole series of CO_2 and H_2O adsorption in M_2(DOBDC) and M_2(DHFUMA) with different metals, we can find that the trend is the same, i.e. CO_2 tends to have much stronger binding in M_2(DHFUMA) than that in M_2(DOBDC) with the same metal, while H_2O tends to have almost the same binding strength in M_2(DHFUMA) and M_2(DOBDC) with the same metal. This would lead to improved selectivity of CO_2 in a CO_2:H_2O mixture in M_2(DHFUMA) than that in M_2(DOBDC).

Enhanced CO_2 adsorption

We demonstrate in Appendix A that the force field of Mercado can reproduce the *ab initio* potential energy landscape of CO_2 in our set of analogs as shown by the agreement of both binding energies and binding geometries. With confirmation of the force field's transferability, Henry coefficients of the M_2(DOBDC) and M_2(DHFUMA) structures were computed at temperatures of 313.0 K and 400.0 K, shown in Table 3.5.

Table 3.5: $K_H * 10^{-3}$ [mol/kg/Pa] of CO_2 in the M_2(DHFUMA) vs M_2(DOBDC) series at 313 K and 400 K

Metal	DHFUMA		DOBDC	
	313 K	400 K	313 K	400 K
Mg	10.7	0.22	1.56	0.064
Fe	1.8	0.07	0.20	0.017
Co	3.2	0.11	0.26	0.021
Ni	3.0	0.12	0.27	0.021
Zn	0.39	0.028	0.076	0.009

An order of magnitude increase is observed in DHFUMA structures over DOBDC structures for a given metal substitution. It is also worthwhile to note that, for a given metal substitution, M_2(DHFUMA) structure achieves the same order of magnitude (and only slightly lower) Henry coefficient at 400 K as its DOBDC counterpart at 313 K in all frameworks except for the Mg analogs. This large decrease in the free energy of a single adsorbed CO_2 molecule in DHFUMA is a direct result of the decreased potential energy of the one molecule per two open metal sites binding mode, which has been demonstrated in our DFT optimization and GCMC simulations. GCMC simulations were utilized to simulate the CO_2 uptake in M_2(DOBDC) and M_2(DHFUMA) structures and calculate isotherms for each material. The higher density of open metal sites and enhanced energy binding configuration results in larger

uptake at low pressures; however, the reduced channel volume results in quicker saturation of the DHFUMA adsorbent. The CO_2 isotherms in Mg_2(DOBDC) and Mg_2(DHFUMA) in Fig. 3.5 visualize this trend.

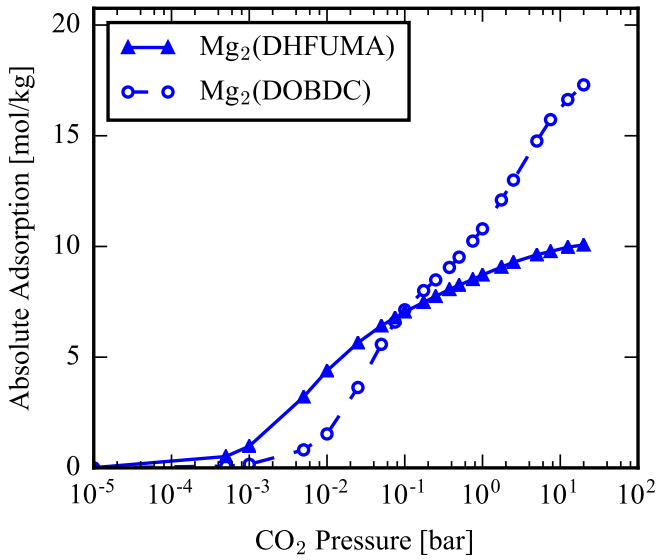


Figure 3.5: Absolute CO_2 adsorption in Mg_2(DHFUMA) vs the absolute CO_2 adsorption in Mg_2(DOBDC) at 313 K. The DHFUMA structure significantly outperforms the DOBDC structure in CO_2 uptake at low pressures but has far lower capacity of CO_2 in the limit of saturation.

Thus at low pressures DHFUMA performs significantly better in total CO_2 uptake but performs worse in total uptake at higher pressures. A detailed view of all CO_2 isotherms is provided in Appendix A. Due to the favorable enhancement of the binding energy, most DHFUMA analogs are able to capture significantly more CO_2 in any pressure range relevant to industrial CO_2 capture from flue gas where $P_{CO_2} = 0.15$ bar. Fig. 3.6 demonstrates the excess amount of CO_2 captured by each metal analog of DHFUMA in comparison to its DOBDC analogs across a pressure range applicable to flue gas conditions.

The excess value peaks at low pressures and then quickly drops to large negative values after the DHFUMA framework saturates with CO_2. It is also significant that DHFUMA analogs continue to load approximately 2 mol/kg more CO_2 at an elevated temperature of 400 K since high temperature adsorption can be used to mitigate competitive water adsorption, as will be seen in later discussion.

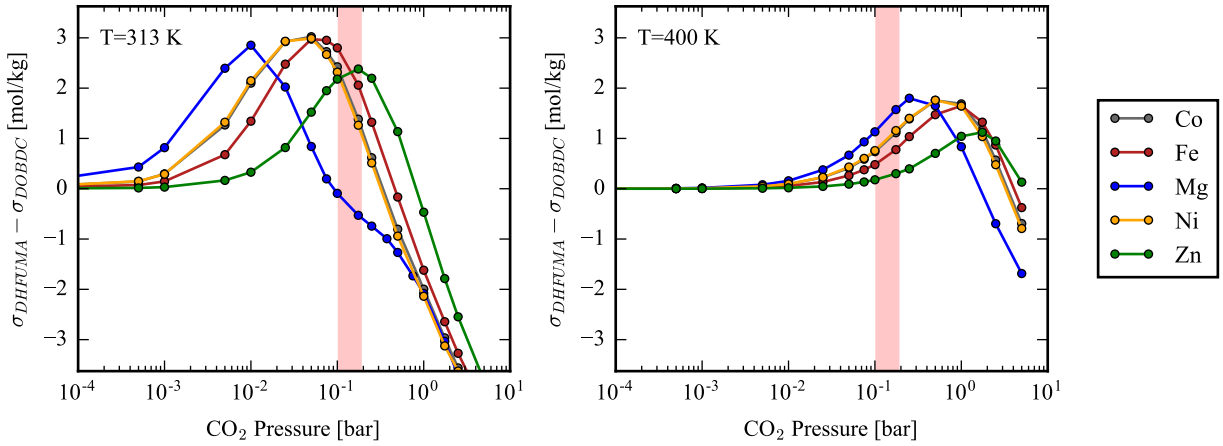


Figure 3.6: Absolute CO_2 adsorption in DHFUMA minus the absolute CO_2 adsorption in DOBDC (a) at 313 K and (b) at 400 K. The pressure region in pink corresponds to the typical partial pressure of CO_2 (P = 0.15 bar) in the exhaust from a coal fired power plant. For each metal a temperature exists between 313 K and 400 K which maximizes the excess CO_2 uptake in the DHFUMA structure.

Enhanced CO_2 heat of adsorption

The isosteric heat of adsorption, ΔH_{Ads}, as a function of loading is a measure of the enthalpy gained on average by adsorbing one additional molecule in the adsorbent system at a specified loading. Conversely, the isosteric heat of desorption, ΔH_{Des}, as a function of loading is the amount of enthalpy required to desorb one additional molecule at a specified loading. In Fig. 3.7 we observe an interesting feature of CO_2 adsorption in the DHFUMA structure that shows a monotonic increase in the heat of desorption as a function of loading from zero to saturation loading. A molecule that adsorbs when the framework is close to saturation (0.8∼0.9 molec./M^{2+}) releases nearly 10 kJ/mol more enthalpy than the first molecule to adsorb. In other words, the cooperative binding of CO_2 is very strong (due to the proximity of the primary binding sites) and increases in strength monotonically with loading. This leads to the perhaps non-intuitive property that the enthalpy penalty to desorb CO_2 *always decreases* as the loading decreases from saturation to empty framework. Notably, Mg_2(DHFUMA) does not exhibit this trend because the binding energy of one CO_2 molecule is so strong that cooperative adsorption is only favorable enough to maintain a constant ΔH_{Des} as a function of loading. The same phenomena does not exist with the DOBDC series. Only a 1∼2 kJ/mol increase in ΔH_{Des} is observed in all DOBDC analogs between the limit of 0 loading and the inflection point at one molecule per open metal site. Thus cooperative adsorption at loadings below 1 molecule per open metal site is negligible in the DOBDC frameworks when compared to the DHFUMA frameworks.

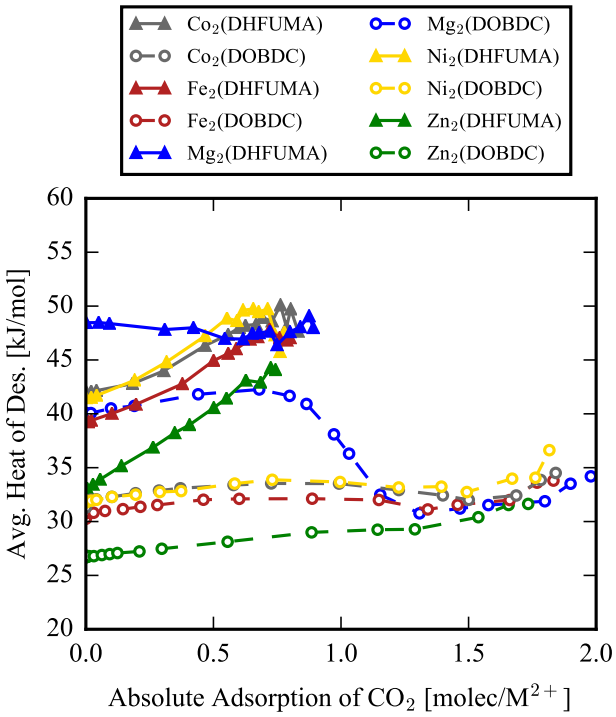


Figure 3.7: The heat of desorption as a function of loading for the M_2(DOBDC) and M_2(DHFUMA) analogs. The DHFUMA structures, with the exception of the Mg analog, exhibit a 7-10 kJ/mol increase in the ΔH_{Des} between the limit of zero loading and saturation.

Optimizing CO_2 capture in binary CO_2:H_2O mixtures

Enhanced CO_2 uptake in the range of partial pressures relevant to adsorption from a coal-fired flue stream does not necessarily indicate an improved potential for industrial scale CO_2 capture.[109] A multitude of other factors must be considered such as CO_2 selectivity, compression work of the CO_2 enriched waste stream, and the energy required for adsorbent regeneration, and these attributes can be quantified through a metric known as the parasitic energy.[76] However, the deleterious effects of water on CO_2 uptake is often overlooked when evaluating materials for CO_2 capture potential via the parasitic energy.[110] For this reason we investigated water adsorption in the DHFUMA and DOBDC series as well since creative strategies such as high temperature adsorption and low temperature desorption (HALD [111]) have been proposed to mitigate the parasitic energy when adsorbing CO_2 from humid flue gas streams. Pure component water isotherms and binding geometries are shown in Appendix A.

More relevant to carbon capture, we performed a CO_2:H_2O binary mixture analysis to investigate each structure's potential performance for carbon capture in the presence of water. GCMC simulations were performed at a fixed reservoir pressure of 0.15 bar, and

the molar composition of CO_2:H_2O of the reservoir was varied at different temperatures. Note that N_2 was not simulated in the mixture as its uptake in the MOF-74 framework series has been shown to be negligible in comparison to the uptake of CO_2 and H_2O at flue gas adsorption conditions.[86, 87]. The same stepped feature of water adsorption occurs in this binary analysis as in the pure component H_2O isotherms and, at a certain critical pressure, water condenses within the pore. In the case of this binary mixture analysis, the condensation of H_2O is sufficient to entirely remove any adsorbed CO_2 at equilibrium. Fig. 3.8 demonstrates the loss in CO_2 uptake capacity that occurs after the molar composition of water (at fixed total pressure) in the reservoir becomes too high.

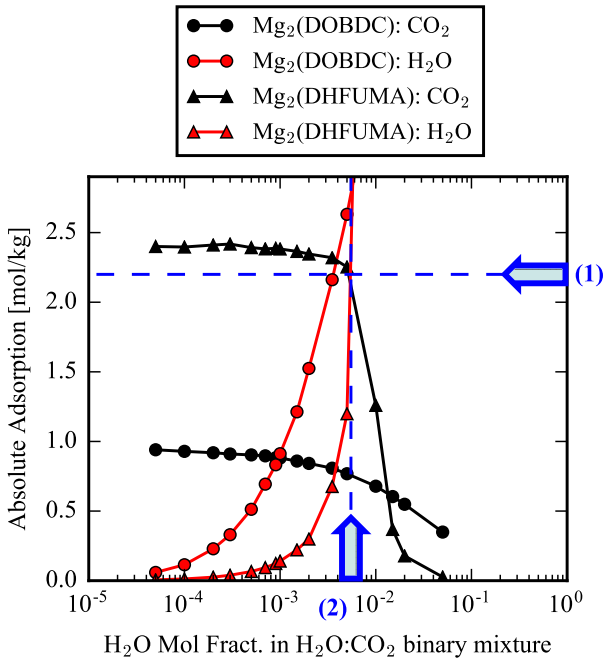


Figure 3.8: A mixture analysis of CO_2 and H_2O adsorption in the Mg analogs of DHFUMA and DOBDC. Each data point represents an equilibrated absolute adsorption loading from a GCMC simulation at 400 K and fixed total pressure of 0.15 bar while the molar composition of H_2O to CO_2 is varied between simulations. For each analog, two values are extracted, and (1) and (2) demonstrate the values extracted for Mg_2(DHFUMA). (1) Corresponds to the CO_2 uptake that is equal to 90% of the uptake in the limit of 0 mole fraction of H_2O. (2) Corresponds to the H_2O mole fraction at which the total CO_2 uptake has decreased by 10%.

Yet at higher temperature, the onset of the water step is shifted to a significantly higher mole fraction. A reduction in CO_2 uptake capacity follows from this temperature increase, yet this uptake loss is mitigated in the DHFUMA structure due to the enhanced CO_2 affinity. An entire summary of the mixture analysis for each structure at various temperatures is

presented in Appendix A. From one mixture analysis at a specified temperature, we can extract two values of importance which are visualized in Fig. 3.8: the water mole fraction just before water condensation occurs and the amount of CO_2 loaded at that specific water mole fraction. These represent competing process design variables. To increase the water mole fraction at which condensation occurs, we must raise the temperature, which consequently reduces the amount of CO_2 loaded. These two quantities are plotted for each analog structure across a range of adsorption temperatures (313K - 400K for M = [Co, Fe, Ni, Zn] and 400K - 473K for M = [Mg]) in Fig. 3.9.

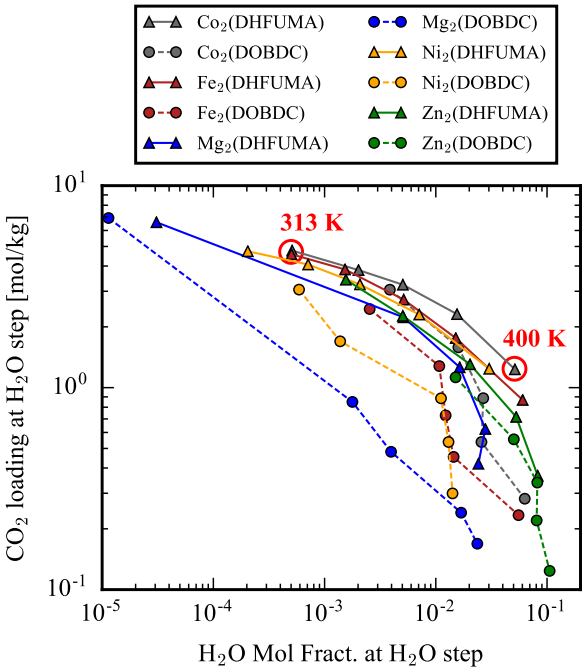


Figure 3.9: Mixture analysis of all analogs demonstrating the competing nature of CO_2 uptake and water tolerance. The y-axis corresponds to value (1) extracted from Fig. 3.8 and the x-axis corresponds to value (2). The ideal material for CO_2 capture would have a data point corresponding to 313 K at the top right corner of the plot.

From a carbon capture process design perspective, the ideal material would be located in the top right corner of Fig. 3.9 where the material loads large amounts of CO_2 in the presence of extremely high water mol fractions. Since the x and y quantities of Fig. 3.9 represent competing variables (but we desire to maximize both of them), we can interpret this summary as a problem of Pareto optimality and a Pareto frontier can be observed for Co_2(DHFUMA). This means that regardless of the adsorption temperature chosen for our capture process, there is no material that can simultaneously achieve a higher water tolerance and CO_2 uptake capacity at thermodynamic equilibrium than Co_2(DHFUMA). Therefore, regardless of the selected operating temperature, Co_2(DHFUMA) will have the

best uptake and water tolerance as is easily visualized in Fig. 3.9. In more physical terms, a Pareto optimal material in this context of CO_2 capture in the presence of humidity will be the material which delicately balances two factors. First, CO_2 uptake must remain the highest with increasing temperature, which fundamentally arises from the highest CO_2 Henry coefficient, which in turn arises from the strength of CO_2 interactions at the open metal site(s). Second, H_2O condensation must occur at the highest water mol fraction, which arises from a combination of the weakest possible H_2O interactions with the open metal site and the largest pore size. Hence we can see the competing nature of these two factors since the strength of CO_2 binding and H_2O binding at the open metal site are highly correlated, and the advantage of DHFUMA becomes immediately clear since we selectively strengthen the binding energetics of CO_2 across all metals due to the one molecule per two open metal site binding mode. Thus we also notably observe that each M_2(DHFUMA) represents a Pareto frontier over its DOBDC counterpart. For example, Ni_2(DOBDC) cannot maximize either water tolerance or CO_2 uptake above Ni_2(DHFUMA) regardless of our specification of the adsorption temperature.

3.6 Outlook: A challenging materials synthesis problem worth solving

We have demonstrated the *in silico* design of a new MOF-74 analog based on the aliphatic DHFUMA ligand. We predict exceptional small molecule adsorption properties via a combination of *ab initio* electronic structure calculations and classical molecular simulation techniques in conjunction with the extensive previous research efforts to create simulation methods that accurately predict guest molecule behavior in MOF-74 type frameworks. Namely we have predicted (1) a doubling of the volumetric storage capacity of H_2 in Mg_2(DHFUMA) over Mg_2(DOBDC) at the cryogenic temperature of 77 K and pressures below 1 bar (and can meet the 2020 DOE target of 40 g/L at $\sim$0.5 bar); (2) a marked enhancement of CO_2 uptake in low pressure regimes over the DOBDC analog series; (3) a selective increase of CO_2 binding energy (i.e. stronger CO_2 binding with no change in H_2O binding energy), the basis for which we propose an industrial scale CO_2 capture process inspired by the HALD scheme (see Appendix A). These enhancements are a direct result of the open metal site properties of the M_2(DHFUMA) frameworks.

This material contains double the volumetric density of open metal sites over DOBDC analogs, leading to a remarkable simulated volumetric H_2 storage capacity. Additionally, the distance between two adjacent open metal sites in each channel is reduced from 8.3 Å in DOBDC to 6.0 Å in DHFUMA. CO_2 binds to two open metal sites in the M_2(DHFUMA) framework, resulting in a significantly stronger binding energy than in M_2(DOBDC). The confined pore channel results in significant cooperative adsorption of CO_2, with an isosteric heat of adsorption that is $\sim$15 kJ/mol stronger at saturation than in the limit of zero loading. Furthermore, since H_2O still can only bind to one open metal site in M_2(DHFUMA), the *ab*

initio calculated quantity of $\Delta E_{bind,H_2O} - \Delta E_{bind,CO_2}$ in the DHFUMA series is typically ∼10 kJ/mol higher than in the DOBDC series, indicating that the DHFUMA series has more selective CO_2 binding energetics relative to H_2O. This enhanced selectivity for CO_2 is exploited in our classical GCMC simulations and adsorption at high temperatures is proposed, allowing for a theoretical process by which CO_2 can be captured in appreciable amounts (∼1-2 mol/kg) in the presence of non-trace amounts of water (∼0.1-1 mol %). The water tolerance and amount of CO_2 captured is dependent on metal choice and adsorption temperature. Finally, an 80% decrease in ligand expense (per mol) suggests that an M_2(DHFUMA) analog may in the future represent an economically improved path forward for large scale H_2 storage or CO_2 capture from flue gas.

The theoretical work in this paper should motivate efforts to experimentally synthesize M_2(DHFUMA) analogs and confirm our simulated volumetric H_2 storage capacity and adsorption behavior of CO_2:H_2O mixtures. We note the synthesis of M_2(DOBDC) is usually very challenging.[83] Thus far our efforts to synthesize M_2(DHFUMA) are unsuccessful. However, we do not see any obvious reason why M_2(DHFUMA) cannot be synthesized experimentally, considering the dynamical and mechanical stability of these materials as well as the availability and chemical stability of the metal and organic precursors. We also believe this framework could be useful for a variety of other separations or storage applications relevant to clean energy. In the future we plan to investigate a range of topics such as selective adsorption of components from light olefin mixtures which are small enough to fit into the DHFUMA channel network if the material can be synthesized.

Chapter 4

In silico materials prediction targeting 2D zeolites for more energy efficient separations

Having demonstrated the power of *in silico* techniques for predicting three dimensional porous crystals, a subsequent question arises. Which computational techniques will allow us to predict or gain insight into porous crystals that can be synthesised as two-dimensional (2D) frameworks? Scientific interest in 2D materials, ranging from graphene and other single layer materials to atomically thin crystals, is quickly increasing for a large variety of technological applications. While *in silico* design approaches have made a large impact in the study of 3D crystals, algorithms designed to discover atomically thin 2D materials from their parent 3D materials are by comparison more sparse. We hypothesize that determining how to cut a 3D material in half (i.e. which Miller surface is formed) by severing a minimal number of bonds or a minimal amount of total bond energy per unit area can yield insight into preferred crystal faces. We answer this question by implementing a graph theory technique to mathematically formalize the enumeration of minimum cut surfaces of crystals. While the algorithm is generally applicable to different classes of materials, we focus on zeolitic materials due to their diverse structural topology and because 2D zeolites have promising catalytic and separation performance compared to their 3D counterparts. We report here a simple descriptor based only on structural information that predicts whether a zeolite is likely to be synthesizable in the 2D form and correctly identifies the expressed surface in known layered 2D zeolites. The discovery of this descriptor allows us to highlight other zeolites that may also be synthesized in the 2D form that have *not* been experimentally realized yet. Finally, our method is general since the mathematical formalism can be applied to find the minimum cut surfaces of other crystallographic materials such as metal-organic frameworks, covalent-organic frameworks, zeolitic-imidazolate frameworks, metal oxides, etc. This chapter is an unofficial adaptation of an article that appeared in an ACS publication [112]. ACS has not endorsed the content of this adaptation or the context of its use.

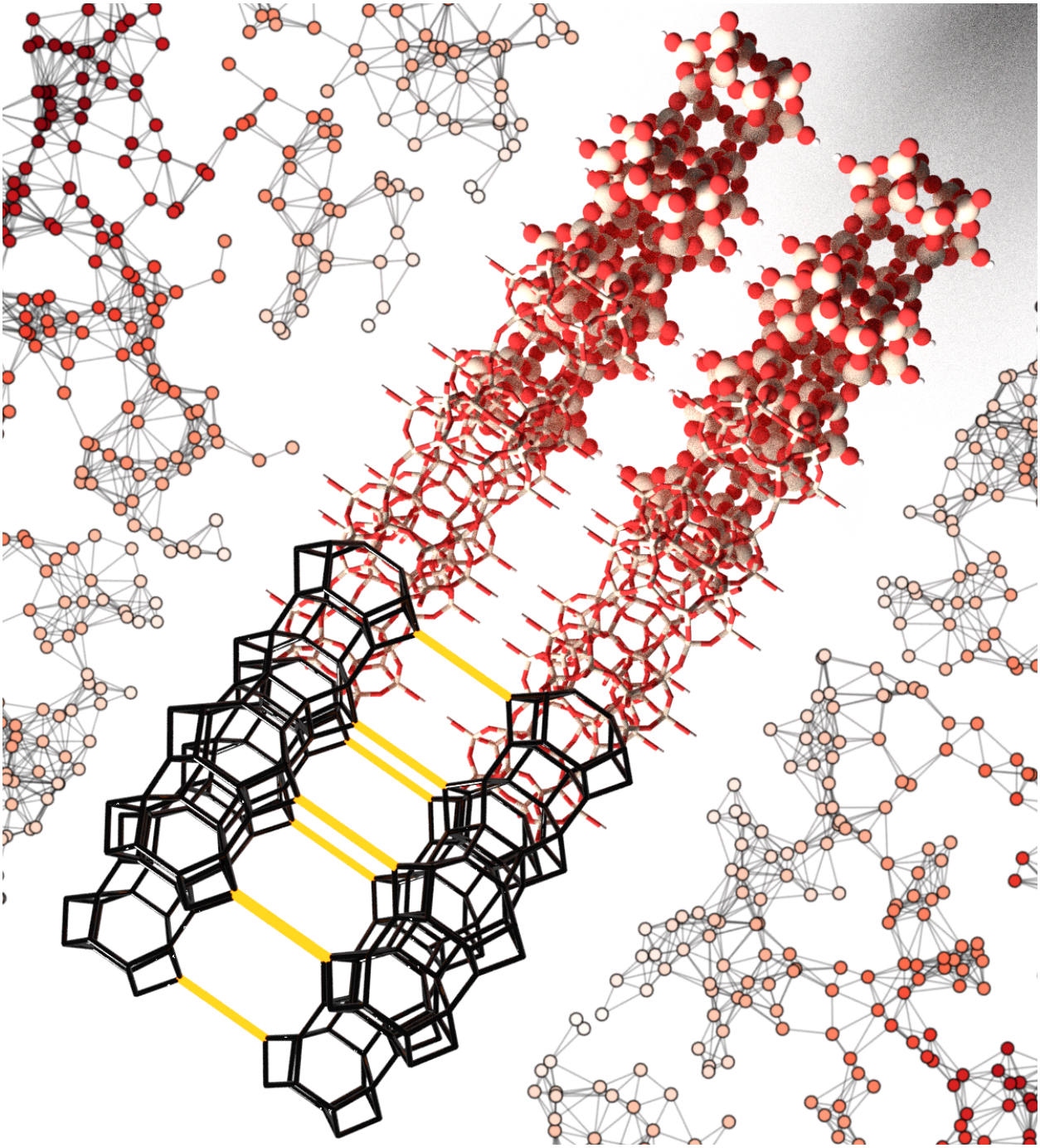


Figure 4.1: Artistic rendition of a graph theoretic interpretation of 2D zeolites

4.1 Investigating 2D zeolites for separations and catalytic applications

Two dimensional (2D) materials are quickly gaining attention as promising materials in a wide variety of applications.[113] While the pre-eminent 2D material (graphene) has a thickness of one atom,[114] more classes of 2D materials have been studied in recent times (i.e. van der Waals layered structures).[115] Atomically thin or layered crystals can also be considered 2D materials since they maintain long range connectivity in the direction of two unit cell vectors but not the third. We hypothesize that answering the following question can yield insights into preferred crystal faces and hence the propensity of crystals to form 2D-like structures: how can a 3D crystal be cut into two separate partitions by severing a minimum number of bonds or, given a pairwise potential, by severing the minimum total bond energy per unit area? Such concepts have already been used to rationalize the crystal dissolution of a zeolite by atomic force microscopy [116] and more recently it has been used in coarse grain crystal growth modeling.[117] Answering this question is, generally speaking, the analogous problem of minimal cuts in graph theory.[118] Given a connected graph, a minimum cut typically seeks to split the graph into two partitions by removing a minimum number of edges or, in the case of a weighted graph, removing edges whose total weight is minimal. Hence if a crystallographic material is interpreted as a graph with atoms for nodes and bonds for edges, we can use graph theory techniques to determine the surface termination for a given Miller plane that minimizes the number of cut bonds, i.e. the number of dangling bonds. As phrased here, this problem is not constrained to one particular class of materials. Initially however, zeolites present the perfect class of crystals to apply this minimum cut formalism since the Si-O bonds have an extremely narrow range of distances and hence each bond can be thought of as approximately equivalent in strength. For the remainder of this work we focus our discussion on zeolite materials. However, we stress that the formalism of this graph theory approach could be applied to many other crystalline materials, including but not limited to metal oxides, zeolitic-imidazolate frameworks (ZIFs), and metal-organic frameworks (MOFs), by introducing edge weights that address the different bond strengths in more complex materials.

Zeolites are crystalline solids whose microporous void structures have made them ubiquitous in a variety of commercial applications, most notably as catalysts in the petrochemical industry and as adsorbents in small molecule separation processes. It is estimated that the global zeolite molecular sieve market will surpass USD 35 billion by 2024.[119] Just as the majority of experimental zeolite research has focused on the performance of bulk 3D materials in catalysis, separations, and other applications, so too has the majority of computational research focused on 3D materials, from hypothetical structure generation [16, 20, 120] (including successful prediction of novel materials [21]) and synthetic descriptors [121–126] to performance prediction and high-throughput screening.[74, 127–129] However, the research on 2D zeolite materials has accelerated in recent times, especially over the past decade.[130–133] For example, 2D zeolites have demonstrated improved catalytic per-

formance [134–136] and shown potentially greater separations efficiency [32, 137–140] over their 3D counterparts. Both applications benefit from improved mass transfer in thin 2D zeolite materials. For catalysis, catalyst deactivation by coking, which is a major problem for 3D zeolite frameworks, can be significantly suppressed with 2D zeolite catalysts. For separations, the diffusion time for molecules to pass through the 2D zeolite molecular sieves can also be dramatically reduced. 2D zeolites have also been used as precursors to 3D zeolites for which no other synthetic procedure is known.[141] While the 2D form of a relatively small number of International Zeolite Association (IZA) [11] zeolites have been discovered, their significant potential in a variety of applications merits further investigations to uncover novel 2D zeolite structures. In this work we show that *in silico* design of 2D zeolites can help further this goal.

Computational investigations of 2D zeolite structures have also intensified recently, especially to understand the interactions between and the reassembly process of layered 2D precursors that are generated in the Assembly, Disassembly, Organization, Reassembly (ADOR) process.[142–145] We aim to build upon these computational studies to develop a high-throughput screening technique and a descriptor to understand whether an IZA zeolite is likely to demonstrate a stable 2D form. To do this we take advantage of some notable work on high-throughput surface characterization performed in the context of the Materials Project, for which the Pymatgen code has generated an efficient and user-friendly platform for generating surface slabs of inorganic crystalline materials.[146, 147] Integrating this platform with open source software for graph theory applications, we apply a technique known as the max-flow min-cut algorithm [118] coupled with an advanced recursive implementation to calculate the minimum cut for any particular surface.[148] Examining the statistics of this minimum cut across all IZA zeolites yields a simple structural descriptor that predicts whether the material is likely to be synthesized as a layered 2D zeolite. Furthermore, the formulation of this graph theory problem is flexible and can be used to mimic the chemistry of specific 2D zeolite synthesis methods such as the ADOR strategy [131] by re-weighting edges in a graph to model varying bond strengths. Thus we show that we can mathematically formalize the generation of zeolite surfaces using graph theory and use this information to make useful predictions regarding the synthesizability of 2D zeolites.

4.2 Minimal cut graph theory to mathematically analyze crystal surfaces

Motivation

We are interested in the fundamental question of how a 3D material can be cut into two separate partitions by removing a minimum number of bonds or, given pairwise potentials, the minimum total bond energy per unit area. Our assumption is that this minimum cut indicates that the surface formed is energetically preferred [149] or that the delamination of the 3D structures into 2D sheets may be facile in this direction. While the true surface energy

under synthesis conditions is extremely complex due to solvent, pH, structure directing agents (SDAs), etc., this minimum cut solution serves as a simple starting point from which an interesting structural descriptor will later be derived. However, we note that conceptual use of minimum cut surfaces has been able to predict and rationalize the crystal dissolution of a zeolite by atomic force microscopy [116] and more recently it has been used in coarse grain crystal growth modelling.[117] Before entering a detailed discussion of the methods, the solution to this minimum cut problem for zeolites is visualized in Fig. 4.2 for zeolite MWW where the minimum cut surface termination of the (001) and (100) surfaces are determined. For example, the (100) surface cannot be formed by removing less than two bonds per unit cell.

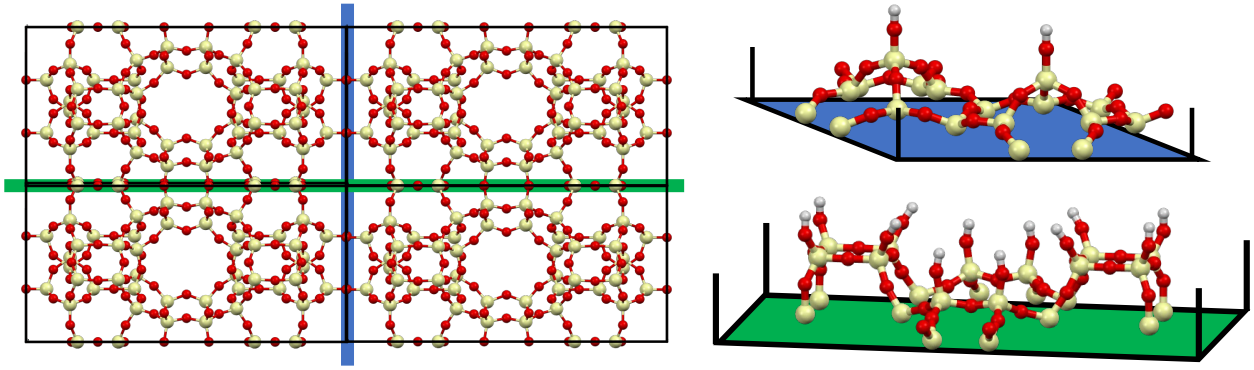


Figure 4.2: Zeolite MWW is shown along with the min cut surface terminations for both (001) = blue and (100) = green Miller planes. For either Miller plane, the surface termination shown represents the minimum number of bonds that can be cut while preserving the 2D periodicity corresponding to that particular Miller surface.

Within graph theory, the problem of determining the minimal cuts (see next section for formal definition) arises in many varieties and has been an active area of research for decades [150–156] with importance in a large number of applications.[157–161] We will combine previous graph theory work on minimal cuts [148] in a novel application: the automatic determination of minimum or near minimum cut surface terminations for any given Miller face of any given zeolite. Since a zeolite can be mathematically interpreted as a simple connected graph, we can use a minimum cut algorithm to solve the surface termination of a particular Miller face that cuts the minimal number of Si-O bonds. In this Methods section we start with mathematical notation and briefly outline the minimal amount of information necessary to understand a special type of graph cut, known as the minimum *s*-*t* edge cut, and provide references for additional details. After explaining how a "zeolite graph" can be created, we show how the minimum *s*-*t* cut problem can be applied to minimize the number of bond cuts necessary to terminate a particular Miller surface of a zeolite.

Preliminaries

The minimal amount of formalism is presented here in order to define the graph theory concepts utilized in this work, and the exact notation of Ref. [148] is used. We take $G = (V, E)$ to be an *edge-weighted, connected, directed graph* with a set of vertices (or nodes) V and a set of edges $E \subseteq V \times V$ where each pair of vertices is ordered, and $n = |V|$ is the order of the graph (number of nodes) and $m = |E|$ is the size of the graph (number of edges). An edge is denoted by its vertex pair $e = (u, v)$. The weight of an edge $e = (u, v)$, denoted w_e, is a numerical value associated with that edge, and the weights of all edges are given by the list $\mathbf{w} = (w_{e_1}, w_{e_2}, ..., w_{e_m})$. In this work the only possible edge weights will be $w_e = 0$, $w_e = 1$ or $w_e = \infty$, as explained later. Two different vertices in V may be specially distinguished as the source and target vertices, or s and t, respectively.

Now, a directed s-t path in G is any path which starts at s and ends at t, or more formally a sequence of vertices and edges of the form $s, (s, v_1), v_1, (v_1, v_2), v_2, ...v_{k-1}(v_{k-1}, t), t$. A generic s-t edge cut is a set of edges C belonging to E that, when removed from the graph, interrupts all paths from s to t. The value or weight of the cut, $w(C) = \sum_{e \in C} w_e$, is simply the sum over the weights of all edges in the cut. From here on we will use the definition of a *minimum cut* (or *min cut*), denoted as C_0, to describe an s-t edge cut whose weight $w(C_0)$ is a minimum among all possible s-t edge cuts. Henceforth we may drop the s-t for brevity since all graphs in this work have a source and target. Multiple minimum cuts can exist, in which case $\mathbf{C}_0$ is a set of min cuts. Finally, a *near-minimum cut* (or *near-min cut*) C_ϵ is an s-t edge cut whose weight is $w(C_\epsilon) \leq (1 + \epsilon)w(C_0)$, in which ϵ is a threshold to control how "near" a near-min cut must be to the min cut. Again, $\mathbf{C}_\epsilon$ is a set of near-min cuts if more than one exists. The following section gives visual examples of this formalism and briefly describes how the min cuts are calculated.

Determining the min and near-min cuts

The min cut of a directed, single-source, single-target graph is determined by computing the maximum flow of the graph.[162] Several algorithms can identify the max flow and in our work this problem is solved using the "shortest-augmenting path" as implemented in the python library NetworkX.[163] Once the max flow is known, the value of the min cut is known by the max-flow min-cut theorem, and a single C_0 solution can be easily determined.[118] Further details can be found in the extensive literature regarding the max-flow min-cut theorem, and we point the reader to Ref. [148] and references therein for more information. We show an edge-weighted, connected, directed graph with source s and target t in Fig. 4.3 to illustrate the min ($\mathbf{C}_0$) and near-min cuts ($\mathbf{C}_\epsilon$) that can be computed. Note that each black double-sided arrow represents two anti-parallel edges which effectively remove the directionality of all s-t paths in the graph. This is desired because materials' bonds have no directionality but the min cut algorithm in this work operates only on directed graphs. As a consequence the min cut value is modified to be $w(C) = \sum_{e \in C} w_e/2$ when anti-parallel edges exist to avoid double counting.

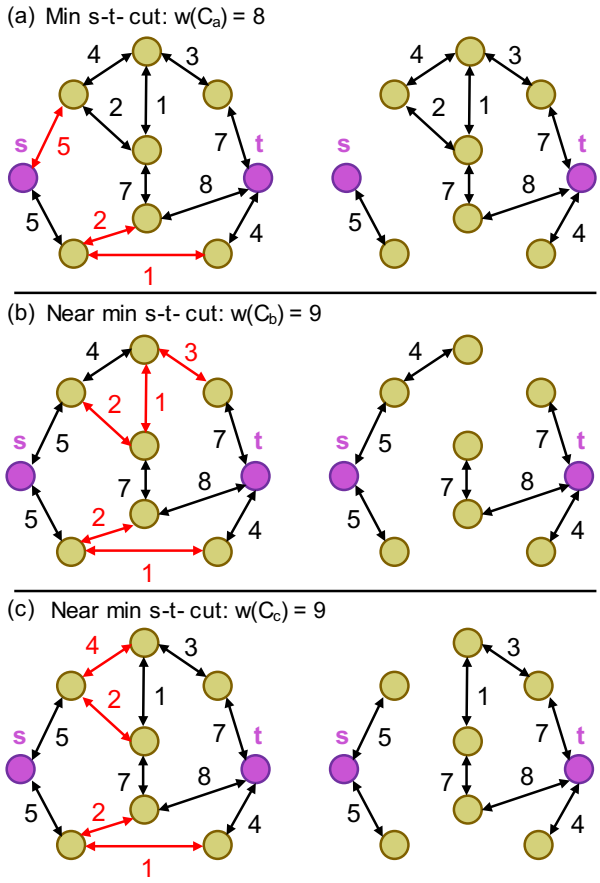


Figure 4.3: Sample graph with source and target nodes (purple) for which we seek to find all min and near-min cuts, where red edges represent those included in the cut. (a) Shows the min cut solution while (b-c) show near-min cut solutions. Hence $\mathbf{C}_0 = (C_a)$ and $\mathbf{C}_\epsilon = (C_a, C_b, C_c)$ when $\epsilon = 0.125$. More near-min cuts could be included in $\mathbf{C}_\epsilon$ with an increased value of ϵ.

Determining the first min cut (Fig. 4.3a) is relatively easy using the max-flow min-cut theorem. However, identifying all min and near min cuts (Fig. 4.3a, b and c) requires additional effort for which Balcioglu and Wood have proposed an elegant algorithm that can be easily implemented.[148] First, a single min cut C_0 is found by the max-flow min-cut theorem. Next, a recursive call to the max-flow min-cut algorithm is performed, but at each level of the recursion tree, the weight of a particular edge is modified to force its inclusion or exclusion from the previously identified min cut. We refer the reader to Section 2.2 and Fig. 2.2 of Ref. [148] for more specifics and to Appendix B for the Python implementation of this recursive function. This recursive search of possible cuts ultimately outputs all min cuts and any near min cuts based on the user specified value of ϵ. It should be noted that the computational feasibility of identifying $\mathbf{C}_\epsilon$ decreases for increasing ϵ since the number of

cuts may be exponential in the size of the graph; however, we are only interested in values of $\epsilon = 0$ in this work but still envision the use of non-zero ϵ in future work.

4.3 An algorithmic procedure to generate the minimal cut surfaces of zeolites

Creating an initial zeolite nanosheet

A "naive" zeolite surface slab (nanosheet) can easily be generated using open source materials science libraries, and for this work we utilize the generalized surface slab building feature of Pymatgen.[146, 147] The algorithm can build a surface slab of any Miller index from a bulk unit cell of any Bravais lattice, while advanced features can be used to build slabs where the two slab surfaces share an inversion point symmetry (when possible), to work with polar surfaces, and to determine symmetrically unique Miller faces. To generate our library of zeolite surfaces, we adopt the following procedure. We create a surface slab for a given Miller face for a given zeolite using Pymatgen and ensure that the slab contains Laue symmetry (when possible) so that both surfaces of the slab are symmetrically equivalent. A schematic representation is shown in Fig. 4.4 where, by Pymatgen convention, the slab's surface is parallel to the *ab* plane of the new unit cell and the *c*-direction corresponds to the vacuum.

We also require that the Pymatgen generated initial slab have N layers with translational symmetry where the thickness of the 2 ... N-2 layers must be greater than a certain cutoff distance. We chose a cutoff distance of 30 Å to minimize the self interactions of the two surfaces so that the slabs could be used for first principles calculation in future work. When enumerating all near min cuts, we will only finalize a cut C_0 or C_ϵ if at least one Si atom attached to any of the edges $e \in C_0, C_\epsilon$ is in Slab Bulk Layer 1, shown in Fig. 4.4. Imposing such a constraint eliminates the possibility of identifying equivalent cuts in the translationally symmetric layers, since a cut that only contained edges with Si atoms in Slab Bulk Layer 0 would be discarded due to the existence of an equivalent cut in Slab Bulk Layer 1. This also preserves the bulk character of Slab Bulk Layers 2 ... N-2 so that the surfaces remain sufficiently separated.

Application of min cuts to zeolites

A zeolite slab can be easily converted into a simple graph upon which the min cuts can be calculated. O atoms are ignored since all are two coordinated to exactly two different Si. Thus a zeolite can be interpreted as a simple connected graph where each node represents an Si atom with an edge between any two Si atoms that are connected by the same O. First, a zeolite slab is prepared and then interpreted as a simple graph which is schematically shown in Step 1 of Fig. 4.5. The real zeolite graph should have edges that represent bonds crossing periodic boundaries in two dimensions (as denoted by the dashed lines in this toy example)

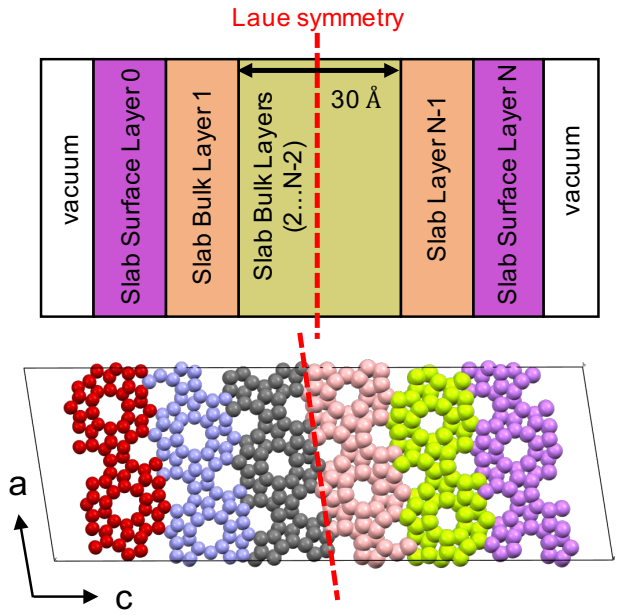


Figure 4.4: (Top) Each Miller surface slab contains N layers, where the slab thickness of the 2 ... N-2 bulk layers was chosen to be greater than 30 Å. (Bottom) The (103) Miller surface slab of MFI for N=6 is shown with oxygen atoms excluded where the color of each Si corresponds to its layer in the slab.

but none in the third dimension which corresponds to the direction perpendicular to the vacuum on either side of the slab. Any nodes in Step 1 that are less than 4-coordinated are under-coordinated Si atoms whose bonds have been removed in the initial slab generation. They are identified as the initial surface nodes in Step 2 and colored purple. On each side of the slab, these initial surface nodes are connected to a new single surface node, which is designated as the source or target node for a min s-t cut computation. The weights of these new edges must be set to infinity as shown in Step 3 to ensure that any C_0 and C_ϵ are independent of the initial surface given in Step 1. In Step 4, we show a solution to the min s-t cut that is identified by the algorithm, specifically the leftmost cut. There is no need to repeat the algorithm to identify the analogous cut on the right side of the slab since it is immediately determined by Laue symmetry or, in its absence, translational symmetry between the slab layers which was visualized in Fig. 4.4. The final step is to remove the partitions of the graph containing the source and target nodes, as well as passivate each dangling Si-O bond with an H atom. Hence we have identified a surface termination of the zeolite by disconnecting a minimum number of Si-O bonds since $w_e = 1$. It should be noted that the generated surface may express Si-OH, Si-$(OH)_2$, or Si-$(OH)_3$ groups to minimize the total number of removed bonds.

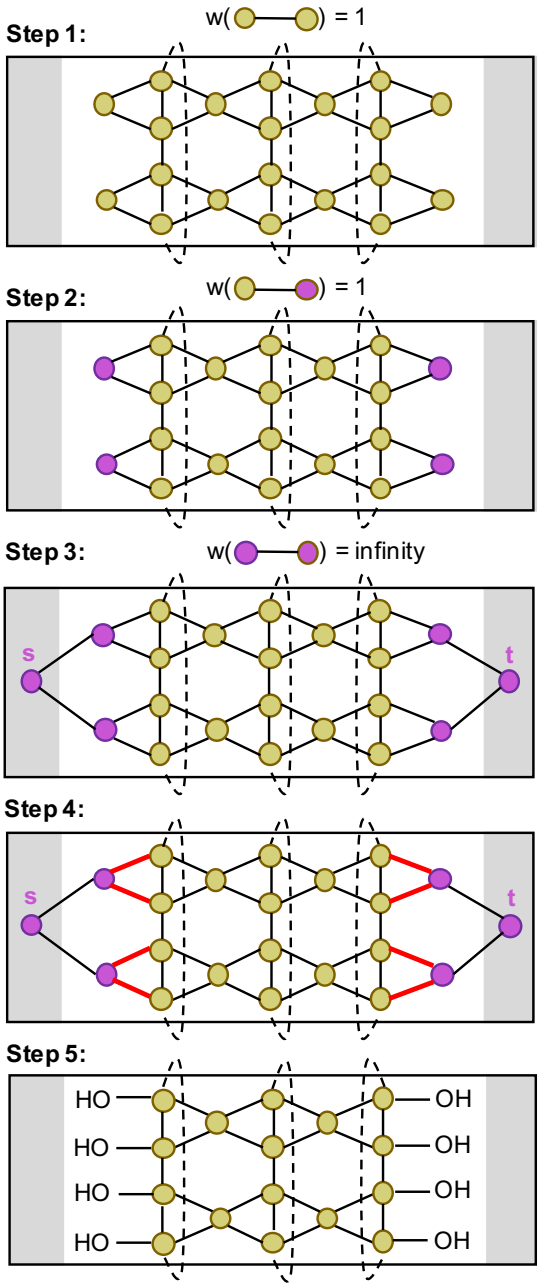


Figure 4.5: (Step 1) Schematic representation of a toy zeolite slab graph where nodes correspond to Si atoms and O atoms are replaced by a single edge. Note the dashed edges which are periodic when embedded in the 2D unit box. (Step 2) Under-coordinated surface nodes are identified. (Step 3) s and t nodes are connected to the surface nodes by edges with infinite weight. (Step 4) The min cut solution is found. (Step 5) The dangling bonds in the final structure are passivated with hydrogen.

We also note that the algorithm for finding min cuts using the max-flow min-cut theorem operates on *directed* graphs, but in this section the zeolite graph in Fig. 4.5 is presented as a simple *undirected* graph. However, finding a min cut in undirected graphs is straightforward provided the following standard transformation is performed.[156, 164] Every edge in the undirected graph can be replaced with two directed anti-parallel edges (Fig. 4.3), each with a weight equal to that of the original undirected edge. The previously discussed min cut algorithm can be executed on this analogous directed graph, and any min cut represents a cut of equal weight in the original undirected graph, provided the cut weight is modified to $w(C) = \sum_{e \in C} w_e/2$ to avoid double counting the additional anti-parallel edge.

The Lammps Interface code,[165] which interprets nanoporous materials as periodic graphs using Python's NetworkX package,[163] was used to convert zeolite structures into their analogous graphs, and Pymatgen was used to generate the initial surface slabs. We extended NetworkX's default max-flow min-cut computation with the previously described recursive algorithm of Ref. [148] to solve the min cut of the initial slab. Using GNU parallel [166] to run embarrassingly parallel jobs was sufficient to generate just one min cut slab for each symmetrically unique Miller surface up to maximum index of 2 of every IZA zeolite in just a few days on a 8 processor desktop. This yielded a total of ∼3700 slabs.

4.4 Predicting 2D zeolites from minimal cuts

Generation and characterization of IZA zeolite surfaces

The min cut for each Miller surface up to a maximum index 2 for each IZA (excluding any interrupted structures, denoted with a "-" by the IZA commission) was solved to generate a library of 2D nanosheets. Thus for each Miller face of each IZA we have calculated the $w(C_0)$ value, which is exactly equal to the number of edges that are cut to form the surface since $w_e = 1$ in this scheme. Since each slab is embedded in a new unit cell with the vacuum parallel to the c-direction as visualized in Fig. 4.4, $w(C_0)$ is converted to a cut density by division with the area of the face spanned by the ab unit cell vectors. Then each Miller surface is ranked by this surface density of cut edges, $\delta = w(C_0)/(|a \times b|)$ with units of Å^{-2}, as shown in Table 4.1 for the examples of EMT and MWW. From now on δ will be referred to as the min cut density, which can also be interpreted as a density of cut bonds when the weight of each edge in the graph is unity. Note that several Miller surfaces can have the same min cut density due to symmetry equivalence, and this was exploited to save computational time by only performing the min cut analysis on one of the symmetrically equivalent Miller planes as calculated by Pymatgen. Repeating the analysis in Table 4.1 would be extremely arduous if not impossible by manual/visual inspection of all IZA zeolites. Some additional analyses on the statistics of all IZA min cuts are shown in Appendix B to highlight the necessity of the automated and robust algorithmic approach provided in this work.

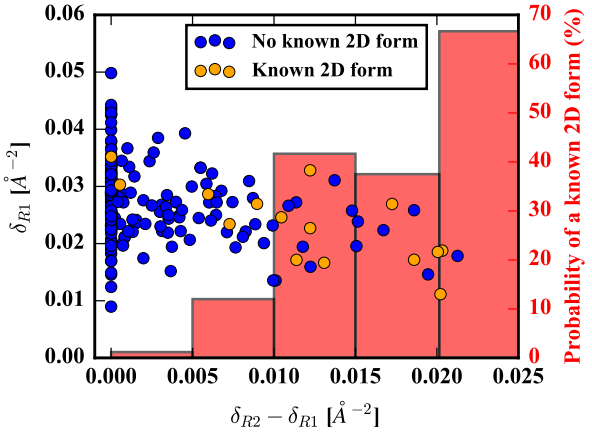
Table 4.1: Ranking of the EMT and MWW Miller surfaces based on their min cut density, δ [=] Å^{-2}. Surfaces ranked higher than 10 are omitted for clarity.

	EMT		MWW	
Rank	Face	δ	Face	δ
R1	(001)	0.0234	(001)	0.0112
R2	(100)	0.0248	(102)	0.0314
R3	(1$\bar{1}$0)	0.0248	(10$\bar{2}$)	0.0314
R4	(010)	0.0248	(1$\bar{1}$2)	0.0314
R5	(101)	0.0256	($\bar{1}$12)	0.0314
R6	(10$\bar{1}$)	0.0256	(012)	0.0314
R7	(1$\bar{1}$1)	0.0256	(01$\bar{2}$)	0.0314
R8	($\bar{1}$11)	0.0256	(100)	0.0331
R9	(011)	0.0256	(1$\bar{1}$0)	0.0331
R10	(01$\bar{1}$)	0.0256	(010)	0.0331
...				

Application: Predicting IZAs likely to grow in layered 2D form

We aim to have a predictor for zeolites which can grow in a stable, layered 2D form. A closer look at Table 4.1 reveals a major difference between the statistics of the min cut densities for MWW and EMT. For MWW, the difference in the min cut density between the R2 and R1 surface, $\delta_{R2} - \delta_{R1}$, is relatively large with a value of 0.02 whereas for EMT this quantity is practically zero. Now the question becomes whether such an outlying min cut density of the R1 surface indicates that it can be more easily isolated during crystal growth, leading to the formation of layered 2D zeolites. In other words, can one more readily find synthesis conditions/SDAs to achieve enhanced stability of the R1 surface relative to other surfaces or obstruct growth in the dimension orthogonal to R1? To investigate this question, we plot δ_{R1} vs $\delta_{R2} - \delta_{R1}$ for all IZA zeolites to elucidate an important trend shown in Fig. 4.6.

In this plot each data point corresponds to an IZA zeolite, where gold points have a known 2D form [130, 133, 167] and blue points have no known 2D forms. Picking a structure at random, one would have an approximate 7% chance of choosing an IZA zeolite that is known to exhibit a 2D form. However, when focusing on structures with the largest $\delta_{R2} - \delta_{R1}$ values, that probability increases significantly (up to $\approx 70\%$). In other words, $\delta_{R2} - \delta_{R1}$ is a *probabilistic descriptor* for identifying an IZA zeolite with a known 2D form. Physically, this simple structural descriptor says the following: it is more likely to find synthesis conditions or SDAs that block the growth of a zeolite in one crystallographic dimension when the face perpendicular to that dimension has a much lower minimum cut density than any other face. This leads to the formation of a 2D layered precursor and it is only the growth in the third crystallographic dimension that is blocked. To ensure that the descriptor actually predicts a 2D zeolite with the correct surface, we manually investigated all reports of the known 2D



Known 2D Form			No Known 2D Form		
IZA	Formula	δ_{R2} - δ_{R1}	IZA	Formula	δ_{R2} - δ_{R1}
CDO	$Si_{36}O_{72}$	0.020	CGF	$Co_4Ga_5P_9O_{36}$	0.021
MWW	$Al_{0.4}B_{5.1}Si_{66.5}O_{144}$	0.020	UOS	$Si_{55.5}Ge_{40.5}O_{192}$	0.019
FER	$Al_6Si_{30}O_{72}$	0.020	JSW	$Co_8Al_{16}P_{24}O_{96}$	0.019
HEU	$Al_8Si_{28}O_{72}$	0.018	SBN	$Ga_8Ge_{12}O_{40}$	0.017
CAS	$Al_4Si_{20}O_{48}$	0.017	GON	$Si_{32}O_{64}$	0.015
STI	$Al_{20}Si_{52}O_{144}$	0.013	BRE	$Al_4Si_{12}O_{32}$	0.015
PCR	$Si_{60}O_{120}$	0.012	EPI	$Al_6Si_{18}O_{48}$	0.015
RWR	$Si_{32}O_{64}$	0.012	AEN	$Al_{24}P_{24}O_{96}$	0.014
RRO	$Si_{18}O_{36}$	0.011	USI	$Co_4Ga_6P_{10}O_{40}$	0.012
AFO	$Al_{10}P_{10}O_{40}$	0.010	ZON	$Zn_8Al_{24}P_{32}O_{128}$	0.012
NSI	$Si_{24}O_{48}$	0.009	CFI	$Si_{32}O_{64}$	0.011
OKO	$Si_{68}O_{136}$	0.007	JRY	$Co_2Al_{10}P_{12}O_{48}$	0.011
MTF	$Si_{44}O_{88}$	0.006	AFS	$Mg_6Al_{22}P_{26}Si_2O_{112}$	0.010
MFI	$Al_nSi_{96-n}O_{192}$	0.0005	BPH	$Be_{14}P_{14}O_{56}$	0.010
SOD	$Al_6Si_6O_{24}$	0.0	STT	$Si_{64}O_{128}$	0.010

Figure 4.6: The plot of δ_{R1} vs $\delta_{R2} - \delta_{R1}$ where each data point corresponds to an IZA zeolite. Large values of $\delta_{R2} - \delta_{R1}$ lead to a much higher probability that an IZA zeolite has a known 2D form, indicating it is a probabilistic descriptor for the ability to form such structures. The 15 IZAs with known 2D form and the top 15 IZAs with no known 2D form are listed in order of decreasing descriptor value to help identify them in the figure. The chemical formula given by the IZA Commission [11] (omitting counter ions and SDAs) is also provided.

structures from the literature following the references in Ref. [130] and [133]. In all cases, the reported 2D structure [168–176] is formed such that the expressed surface corresponds to the same surface we identify as R1 in our high-throughput screening. To summarize, Fig. 4.6 clearly demonstrates that the R1 surface is much more likely to be expressed under given synthetic conditions than the R2 surface when $\delta_{R2} - \delta_{R1}$ is large, while the *absolute value* of δ_{R1} is less relevant for predicting layered 2D crystal growth.

We note real zeolite synthesis is a very complex process, which requires fine control of reaction conditions, including but not limited to the right reactants, a specific structure directing agent, controlled pH (of reaction medium), optimized reaction temperature and time, and the right mixing fraction and procedure. The experimental realization of our predicted 2D zeolite candidates also relies on the fact that the right reaction conditions need to be identified. This is beyond the scope of our current work. However, our predicted surface terminations of the Miller planes with the lowest cut densities may provide crucial insights on one of the most important reaction ingredients of zeolite synthesis, i.e. the structure directing agents. By computationally screening the binding strengths of conventional and unconventional SDA molecules with different Miller planes of a selected 2D zeolite candidate, one may identify a SDA molecule with much stronger interaction with the lowest cut density Miller plane than with other Miller planes, and therefore zeolite growth will be promoted along this unique direction and inhibited along other directions, which results in a 2D zeolite with a maximally expressed Miller surface. We also note that the life time of zeolite surfaces has been reported to inversely correlate with the density of surface dangling bonds (i.e.

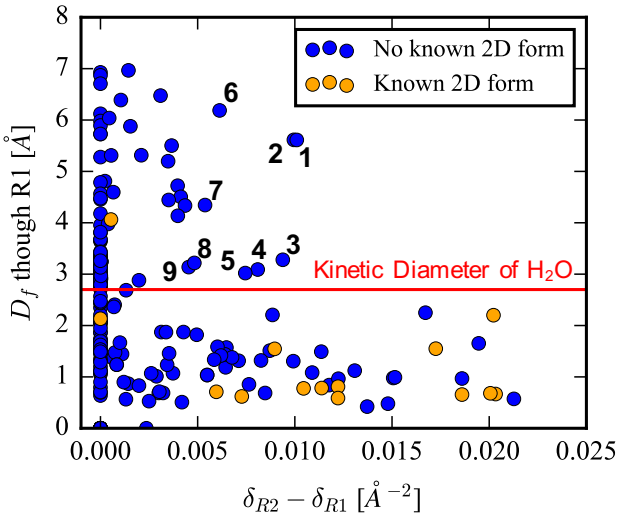
the cut density),[116] indicating that the lowest density cut Miller planes will have more time to interact with SDA molecules and have increased expression in the growth process, furthering the chance that a 2D instead of a 3D zeolite will be formed. It should be noted MFI represents an outlier since it is the only zeolite that can be synthesized in a stand-alone 2D form,[177] i.e. *not* as a layered 2D precursor, and also has $\delta_{R2} - \delta_{R1}$ approximately equal to zero. This 2D form is also achieved by synthesis methods unique to this structure, namely as a multilamellar precursor with surfactant.[130] Thus a large $\delta_{R2} - \delta_{R1}$ value more reflects the probability to find synthesis conditions to direct formation of 2D layered precursors but as expected does not represent the complex surfactant chemistry at the MFI surface that is utilized to direct its stand-alone 2D growth.

Finally, the structure corresponding to each data point is provided in Ref. [112]. These results provide a list of structures that can be immediately targeted experimentally because they are higher probability candidates to form novel 2D layered structures according to the statistics of currently known 2D layered zeolites. These large $\delta_{R2} - \delta_{R1}$ value structures serve as a starting point for future experimental and computational efforts to predict and investigate which synthetic conditions and SDAs may result in some 2D layered structure where the expressed surface is defined by our predicted R1 surface.

Application: Potential 2D zeolites for water desalination

The non-equilibrium molecular dynamics simulations of Jamali et al. demonstrated that using 2D zeolite nanosheets for water desalination could hypothetically provide large improvements in water permeation performance over current technologies.[32] We now look at all known zeolites with 2D form (as well as those in the "high probability zone" of Fig. 4.6) to determine which structures have potential for this application, i.e. are porous in the direction perpendicular to the R1 surface. Here we define a nanosheet as porous if the largest free sphere (or pore limiting diameter), D_f, in the crystallographic direction perpendicular to the R1 surface is larger than the kinetic diameter of water. D_f in the direction perpendicular to the R1 surface was calculated with Zeo++ [74] and is plotted vs. the descriptor in Fig. 4.7.

Only MFI satisfies this porosity criteria for all zeolites with a known 2D form. SOD and MWW have the next largest values for D_f but neither have pore limiting diameters large enough in the direction perpendicular to the R1 surface to allow water to pass from one side of the nanosheet to the other. However, several structures can be identified with relatively large $\delta_{R2} - \delta_{R1}$ value (indicating a higher probability of having a layered 2D form) that have a pore limiting diameter through the R1 surface larger than the kinetic diameter of water, some of which are listed in Fig. 4.7. Our analysis shows that these materials have the similar structural characteristic as other stable 2D layered zeolites and, given their potential for different types of separations, we hope that this work will stimulate a more targeted effort to synthesize them.



No Known 2D Form			
	IZA	Formula	δ_{R2}- δ_{R1}
1	AFS	$Mg_6Al_{22}P_{26}Si_2O_{112}$	0.010
2	BPH	$Be_{14}P_{14}O_{56}$	0.010
3	BOZ	$Be_{66.7}As_{25.3}O_{80.5}$	0.009
4	AFN	$Al_8P_8O_{32}$	0.008
5	ETL	$Si_{72}O_{144}$	0.007
6	MSE	$Al_{11.4}Si_{100.6}O_{224}$	0.006
7	SEW	$Si_{61.3}B_{4.7}O_{132}$	0.005
8	OBW	$Be_{22}Si_{54}O_{150}$	0.005
9	MON	$Al_{4.5}Si_{11.5}O_{32}$	0.005

Figure 4.7: The largest free sphere, D_f, in the direction perpendicular to the R1 surface plotted vs the descriptor $\delta_{R2} - \delta_{R1}$. For separations to occur in such 2D structures, this value must be larger than the kinetic diameter of the smallest species in the separation mixture. MFI is the only known 2D zeolite to achieve porosity through the nanosheet, but nine potential structures are highlighted that would also be porous and have not yet been discovered in a 2D form. The chemical formula given by the IZA Commission [11] (omitting counter ions and SDAs) is also provided.

Application: Predicting likely 2D zeolites from ADOR disassembly

There exist a variety of specific synthesis techniques for achieving a 2D form of an IZA zeolite.[130, 133] However, if one were to ask which materials may form layered 2D sheets for a specific synthesis method, it is possible in some situations to tailor the min cut algorithm to mimic the chemistry of a particular synthesis technique. Here we provide an example on the versatility of our algorithm to show how we can identify potential 2D zeolites formed

during the disassembly step of a special synthesis procedure, namely the ADOR strategy. In this technique, germanium preferentially occupies double 4 rings (D4R) sites which are selectively hydrolzyed upon acid treatment. With this knowledge, the weights of the edges in a zeolite graph should be set to zero if the edge contains one node in a D4R unit and one node outside the D4R unit. Removing the penalty to cut these edges mimics the selectivity of O-Ge bond hydrolyzation. This re-weighting of bonds attached to D4R units is shown schematically in Fig. 4.8a.

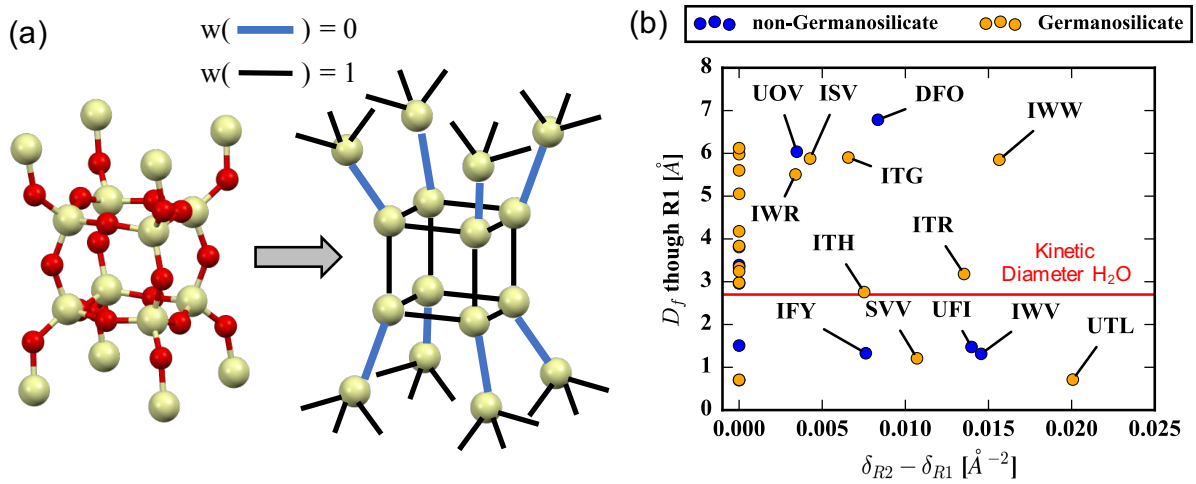


Figure 4.8: (a) A schematic showing the re-weighting of edges corresponding to bonds that are selectively hydrolyzed during ADOR disassembly. (b) The D_f vs. our descriptor, $\delta_{R2} - \delta_{R1}$, is plotted for IZA zeolites containing D4R motifs. Only structures where $\delta_{R1} = 0$ are considered, so a non-zero value of the descriptor indicates that *exactly one* Miller surface (the R1 surface) can be generated by *only* cleaving bonds connecting a D4R unit to the rest of the structure. The red line once again indicates the kinetic diameter of H_2O.

With this new weighting scheme, a min cut density of zero, or $w(C_0) = 0$, can occur for a particular Miller plane if all bonds in the min cut correspond to those that are selectively hydrolyzed during ADOR disassembly. Thus the first requirement for 2D sheet formation with this technique is that $\delta_{R1} = 0$. However, if there exist two min cut surfaces with $\delta = 0$ for two *different* Miller planes, it is evident that no 2D sheet could form as it would be hydrolyzed into discrete fragments that lack 2D periodicity. For 2D layers to be formed, it is sufficient to see that one and only one Miller surface has a min cut density of zero, or in other words, the second requirement is that $\delta_{R2} - \delta_{R1} \neq 0$. Fig. 4.8b shows the pore limiting diameter vs $\delta_{R2} - \delta_{R1}$ when the weighting modification of Fig. 4.8a has been applied to demonstrate promising materials for ADOR disassembly.

Here only IZA structures that have D4R units are shown and are color-coded by those that have been synthesized as germanosilicates. Ref. [131] determined, presumably by manual inspection, that "*the most suitable candidates for top-down synthesis of 2D zeolites [are]:*

ITG, ITH, ITR, IWR, IWW, SVV, UOS, and UTL ... [and] IWV", and our automated approach provides very similar prospective. Regarding non-germanosilicates, Ref. [131] proposes IWV as a potential material for ADOR disassembly if it can be synthesized as a germanosilicate, but excludes IFY, UFI, UOV, and ISV (which were likely excluded since they contain D3R or single 4 rings that if hydrolyzed, would destroy the 2 dimensionality). However, we additionally highlight as a target for the synthesis community that DFO can be included as a potential candidate if the germanosilicate version of the framework can be synthesized. UOS appears in Ref. [131] but not our list since our calculations revealed that *more than one* Miller surface can be formed by only hydrolyzing bonds connected to D4R units, or $\delta_{R2} = 0$. Clearly ITG, ITR, IWR, and IWW are all high potential candidate structures for separations applications since the 3D $\rightarrow$ 2D transformation via Ge-O hydrolysis results in an atomically thin sheet porous to water molecules. Finally, it should be highlighted that this re-weighting scheme can generally be applied to any structural motif in zeolites (e.g. double 6 rings, cages, etc.) to accommodate future disassembly techniques. Subsequent min cut calculations can then provide a formal way of determining whether a 2D sheet can be formed from a 3D structure by only breaking bonds connected to specific building blocks.

4.5 Outlook: Computationally directed search for 2D porous materials

We have applied a powerful graph theory technique to find and enumerate the minimum cut 2D surfaces of 3D crystallographic materials. The ability to calculate a Miller surface termination that minimizes the number of cut bonds (or the minimum total energy of cut bonds given a classical potential) provides a mathematical approach to identifying important surfaces, and we have specifically applied these ideas to zeolites in this work. For any given zeolite and Miller face, one can automatically and rigorously enumerate all of the minimal or near minimal cuts to create a library of 2D nanosheets. To our knowledge this is the first *in silico* design approach using graph theory to study crystal surfaces in such a high-throughput manner. While specifically applied to IZA zeolites in this project, this methodology has the potential to be applied to other crystallographic systems (i.e. ZIFs, metal oxides, MOFs, etc.) [178–181] to investigate various surface terminations in a formalized, high-throughput methodology. One could also, for example, perform an identical analysis for the database of aluminophosphates [182] and other zeotypes due to the generality of our graph theory approach.

Our algorithmic approach to zeolite surface generation not only yields a probabilistic descriptor for the likelihood that an IZA material has a known layered 2D form but also correctly identifies the expressed surface. From random selection one has $\sim 7\%$ chance of selecting a material with known 2D form. Using our descriptor, one can bias this selection probability to $\sim 70\%$. This indicates that materials with favorable descriptor values and *as of yet* unknown 2D form are the most likely to be discovered in layered 2D form upon

investigation of new synthesis conditions. We provide a list of structures which can guide experimental efforts for attempted synthesis of layered 2D zeolites with far higher probability than random search. We have furthermore demonstrated the versatility of the algorithm by predicting suitable candidates for a particular synthesis method, namely the 3D to 2D transformation during the disassembly step of ADOR. While ADOR is relatively new, Eliášová et al. commented that new methods to selectively design structural weakness at specific structural motifs (other than Ge-O in ADOR) and to selectively break such bonds will be critical to developing new 2D forms of known zeolites.[131] When such synthetic procedures are discovered and developed, our algorithmic approach will be invaluable to automatically identify likely materials for such 3D to 2D transformations from the large number of IZAs, non-IZAs, and hypothetical zeolites. Assuming sufficiently accurate classical potentials, a natural extension of this combined ADOR/min cut analysis would be to identify other inorganic crystals that are high probability candidates for exfoliation, or in other words exhibit a 3D to 2D disassembly with exactly one preferred Miller surface.[181]

Our formal approach to enumerating zeolite surfaces also opens a path for other computational studies that can be performed to better investigate and understand zeolite surfaces. For example, screening based approaches can be applied to identify SDAs or solvent conditions that energetically favor the min cut surface over higher density cut surfaces, leading to controlled structure growth. Informatics based identification of SDAs may be able to predict compounds that will lead to isolation of the 2D layered form of some of the high-potential candidates identified in this work. Ultimately, the theory applied here (and its continued development) will not only be important for the continued investigation of zeolite surfaces and identification of potential 2D zeolites, but will also provide a new methodology to examine surfaces of other classes of crystallographic materials.

Chapter 5

Advancing simulation techniques for evaluating adsorption thermodynamics

Monte Carlo simulations are the foundational technique for predicting thermodynamic properties of open systems where the process of interest involves the exchange of particles. Thus, they have been used extensively to computationally evaluate the adsorption properties of nanoporous materials and are critical for the *in silico* identification of promising materials for a variety of gas storage and chemical separation applications. In this work we demonstrate that a well-known biasing technique, known as "flat-histogram" sampling, can be combined with temperature extrapolation of the free energy landscape to efficiently provide significantly more useful thermodynamic information than standard open ensemble MC simulations. Namely, we can accurately compute the isosteric heat of adsorption and number of particles adsorbed for various adsorbates over an extremely wide range of temperatures and pressures from a set of simulations at just one temperature. We extend this derivation of the temperature extrapolation to adsorbates with intramolecular degrees of freedom when Rosenbluth sampling is employed. Consequently, the working capacity and isosteric heat can be computed for any given combined temperature/pressure swing adsorption process for a large range of operating conditions with both rigid and deformable adsorbates. Continuous thermodynamic properties can be computed with this technique at very moderate computational cost, thereby providing a strong case for its application to the *in silico* identification of promising nanoporous adsorbents. This chapter is adapted with permission from Witman, M.; Mahynski, N. A.; Smit, B. *J. Chem. Theory Comput.* **2018**, *14*, 61496158. Copyright 2018 American Chemical Society.

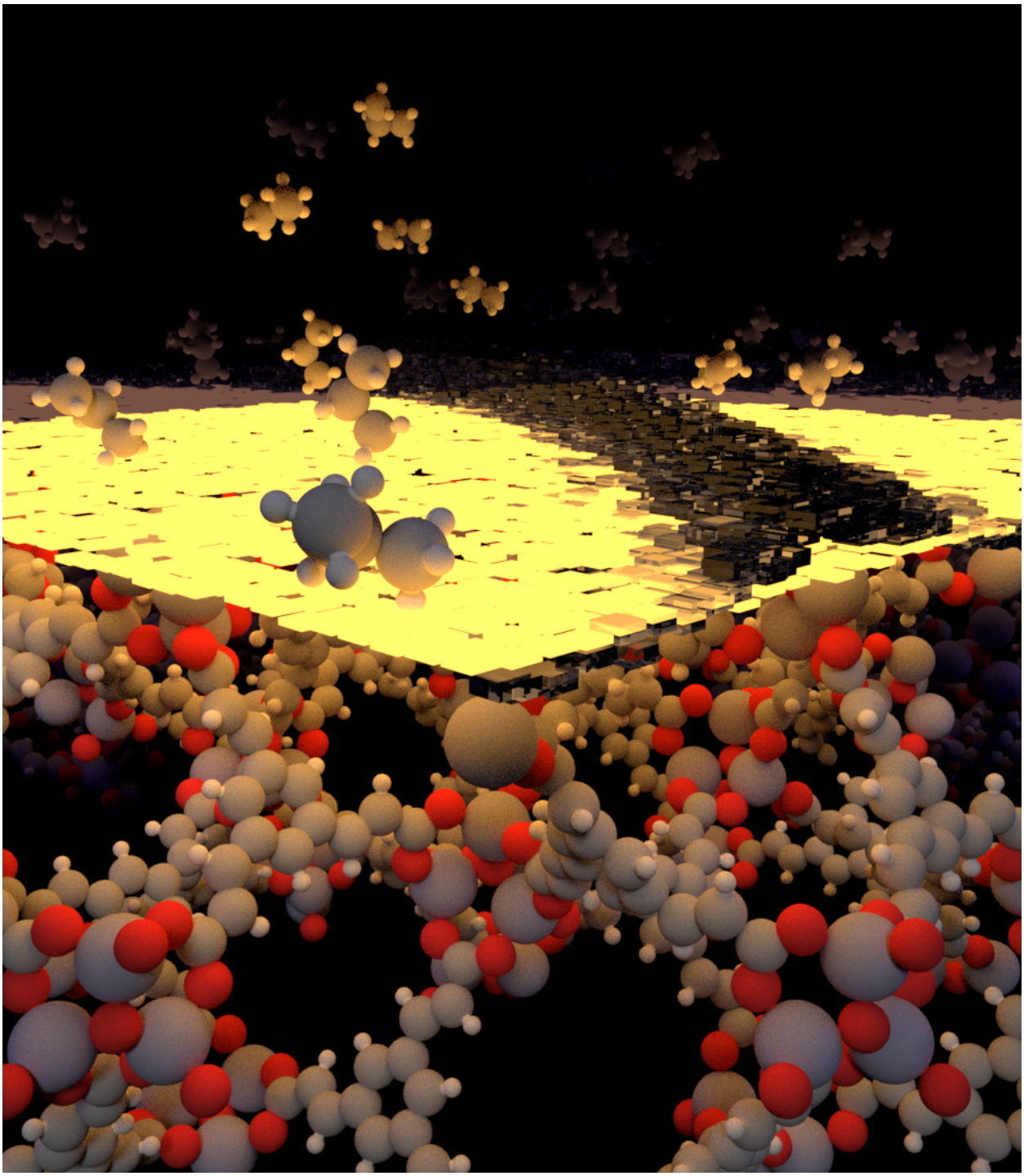


Figure 5.1: Artistic rendition of a flat-histogram illuminating the adsorption capabilities of porous materials

5.1 Beyond the grand canonical: flat-histogram Monte Carlo methods for simulating the thermodynamics of open systems

Open ensemble Monte Carlo (MC) simulations, in which particles can be added or removed from a simulation domain, are the technique of choice for studying many important thermodynamic phenomena such as phase coexistence and adsorption.[183] For example, Gibbs ensemble MC is particularly useful for predicting phase coexistence [184, 185] and grand canonical MC (GCMC) is often used to predict adsorption thermodynamics of fluids in porous media.[186, 187] Extensive work over the past several decades has focused on developing these methods, as well as biasing techniques to enhance sampling of systems where the standard acceptance probability of moves becomes prohibitively small. Some of the most prominent examples include Rosenbluth sampling,[188] configurational bias MC (CBMC),[189, 190] and continuous fractional MC [191] methods which allow effective sampling of deformable molecules and/or high density systems that otherwise cannot be simulated with standard GCMC particle insertion moves.

A different category of biasing techniques emerged when Wang and Landau (WL) developed a general technique to construct a bias on-the-fly enabling the sampling of low probability (high free energy) states that are otherwise not sampled during a normal MC simulation.[192] The goal of transition matrix Monte Carlo (TMMC) is closely related to that of the WL scheme, but there are distinct algorithmic differences between the two which are detailed extensively in the literature.[193] These are generally referred to as flat-histogram or density-of-states sampling techniques since the goal is to sample states characterized by a given collective variable with uniform probability.[194, 195] Often the two techniques are merged to exploit the advantages of each method while avoiding their individual disadvantages, and have been applied extensively to study phase coexistence and adsorption thermodynamics.[195–200] More complex applications of WL/TMMC have been developed recently, especially in the context of studying adsorption and phase coexistence in model systems that contain flexibility, multicomponent mixtures, and non-rigid molecules.[201, 202] Recently, an approach for extrapolating the free energy landscapes determined by WL/TMMC has presented an opportunity to predict continuous thermodynamic properties as a function of temperature by simply post-processing simulation data obtained at a single temperature.[203–205]

In this article, we build upon the aforementioned literature to demonstrate the effectiveness of a flat-histogram method for evaluating a porous material's performance in adsorption processes involving rigid and deformable adsorbates. By applying the recently developed temperature extrapolation procedure,[203] we can predict the thermodynamic adsorption properties of a porous material over a continuous range of temperature and pressure conditions from a simulation at just one temperature. For the first time, we extend this derivation of temperature extrapolation to simulations in which Rosenbluth sampling has been used to sample deformable molecules (e.g., a flexible chain with internal degrees of freedom). This

applicability to both rigid and flexible molecules, as well as the large temperature ranges over which thermodynamic properties can be extrapolated, demonstrates some significant advantages of this method compared to traditional GCMC for the *in silico* evaluation of porous materials. We additionally show how these techniques can be useful in practical applications. As an example, we consider the capture of gaseous species in nanoporous materials using combined temperature and pressure swing adsorption (T/PSA). If one would like to optimize the operating conditions for such an adsorption process, one needs to know the adsorption isotherms over a wide range of conditions. Using these flat-histogram and temperature extrapolation techniques, materials can be evaluated for T/PSA processes over a wide range of operating conditions (via determination of working capacities and isosteric heats) from simulation data obtained at just one temperature.

5.2 Flat-histogram Monte Carlo with temperature extrapolation for flexible adsorbates

A variety of studies have used flat-histogram approaches to determine the free energy landscape as a function of particle number in open ensembles.[196, 198, 206, 207] Due to the detailed discussion and derivation of these techniques elsewhere, we only briefly summarize the formalism to obtain the macrostate probabilities in the grand canonical ensemble. Then we show how analyzing the potential energy fluctuations at each macrostate allows determination of the free energy landscape at different temperatures by Taylor series expansion, and we subsequently extend this temperature extrapolation derivation to the case of deformable adsorbates. Finally we explain how to obtain the continuous thermodynamic properties of the adsorbate/adsorbent system that are important for predicting a material's utility as an adsorbent.

Thermodynamic background

We provide a brief review of the statistical mechanics equations that will be important for applying temperature extrapolation in the case when Rosenbluth sampling has been employed in conjunction with a TMMC biasing scheme. The total canonical partition function of an N particle system in a given volume, V, at thermodynamic temperature, $\beta = (k_B T)^{-1}$ (where k_B is Boltzmann's constant and T is the absolute temperature), is denoted $Q(N, V, \beta)$. It can be factored into an integration over kinetic, $Q_k(N, \beta)$, and configurational, $Q_c(N, V, \beta)$, degrees of freedom as shown in Eqn. 5.1.

$$Q(N, V, \beta) = Q_k(N, \beta) Q_c(N, V, \beta) = Q_k(N, \beta) \left(\frac{1}{N!} \int d\mathbf{r}^N \exp\left[-\beta E(\Gamma_{\mathbf{r}^N}) \right] \right) \tag{5.1}$$

Here E is the total potential energy and $\Gamma_{\mathbf{r}^N}$ is the configuration of the N particle system. More generally we can write $Q_k(N, \beta) = q(\beta)^N$, where $q(\beta)$ represents the integration over

the kinetic degrees of freedom of a single molecule. In the case of monatomic particles, $q(\beta) = \Lambda^{-3}$ where Λ is the thermal de Broglie wavelength; however, in what follows we keep the notation of $q(\beta)$ to more generally include polyatomic flexible molecules.

When this system can exchange particles with an infinite reservoir that imposes some constant chemical potential, μ, the system is described by the grand canonical ensemble where the grand canonical partition function is given by Eqn. 5.2.

$$\Xi(\mu, V, \beta) = \sum_{N=0}^{\infty} \exp(\beta\mu N)\, Q(N, V, \beta) \tag{5.2}$$

In this ensemble, the probability of observing N particles in the system is given by Eqn. 5.3:

$$\Pi(N; \mu, V, \beta) = \frac{\exp(\beta\mu N)\, Q(N, V, \beta)}{\Xi(\mu, V, \beta)} \tag{5.3}$$

By specifying μ, V, and T, one can use grand canonical MC simulations to sample the probability distribution in Eqn. 5.3 and compute, among other things, the expected number of particles in the system.

Grand canonical flat-histogram simulations

Flat-histogram techniques operate quite differently from the standard MC approach. They seek to bias the simulation so that a collective variable, or macrostate, is sampled uniformly, i.e., all macrostates are visited with equal probability and thus have a flat probability distribution. Therefore, macrostates that have a high free energy relative to other states become equally probable when the biasing function has been properly determined, and an efficient method to achieve this goal was pioneered by Wang and Landau.[192, 208, 209] When studying adsorption, a convenient macrostate variable is the number of particles in the system, N. Upon convergence of the WL algorithm, $\Pi(N; \mu, V, \beta)$ is known for each macrostate N, which can then be re-weighted to some determine $\Pi(N; \mu', V, \beta)$ at some new chemical potential μ' by simple post-processing.[202, 206] Alternatively, for single-component systems one can directly solve $Q(N, V, \beta)$ as the biasing function rather than $\Pi(N; \mu, V, \beta)$.[199]

Transition matrix Monte Carlo formulation

Typically the Wang-Landau approach builds up an initial estimate of $\Pi(N; \mu, V, \beta)$ quickly but converges rather slowly. Transition matrix Monte Carlo (TMMC) tends to do the opposite, so a simulations often start with a WL stage which is later switched to TMMC to accumulate the statistics that refine the exact values of the biasing function.[195] The starting point for TMMC is the statement of detailed balance for Monte Carlo particle exchange moves between macrostates:

$$\Pi(N; \mu, V, \beta) P(N \to N+1) = \Pi(N+1; \mu, V, \beta) P(N+1 \to N) \tag{5.4}$$

Here $P(N \rightarrow N+1)$ represents the probability that a proposed MC move takes the system from macrostate N to macrostate $N+1$, given that the system was already in macrostate N. The remaining task is to determine the values of $P(N)$. This is done by constructing the collection matrix (C-matrix). Each entry in the C-matrix is updated by evaluation of the *unbiased* acceptance rule, $\text{acc}(\Gamma_{\mathbf{r}^N} \rightarrow \Gamma_{\mathbf{r}^{N+1}})$ when a swap move is proposed. For example, every time a particle addition move is proposed, the N^{th} row and $(N+1)^{th}$ column is updated with the probability of accepting such a transition:

$$C(N, N+1) = C(N, N+1) + \text{acc}(\Gamma_{\mathbf{r}^N} \rightarrow \Gamma_{\mathbf{r}^{N+1}}) \tag{5.5}$$

Simultaneously, the probability that such a transition is rejected is also used to update the C-matrix.

$$C(N, N) = C(N, N) + [1 - \text{acc}(\Gamma_{\mathbf{r}^N} \rightarrow \Gamma_{\mathbf{r}^{N+1}})] \tag{5.6}$$

The same concept applies to updating the C-matrix for particle deletion moves, and since only single particle additions or deletions are proposed, the C-matrix is tridiagonal. The transition probability can be computed at the end of the simulation from the C-matrix via Eqn. 5.7.

$$P(N \rightarrow N+1) = \frac{C(N, N+1)}{\sum_{\Delta \in \{-1,0,1\}} C(N, N+\Delta)} \tag{5.7}$$

$P(N)$ may then be used to obtain $\Pi(N; \mu, V, \beta)$ via Eqn. 5.4. Regarding the implementation of this method, the simulation can be performed at any arbitrary value of μ. The macrostate probabilities can then be easily re-weighted to other values of the chemical potential to obtain $\Pi(N; \mu', V, \beta)$, which is extensively detailed elsewhere.[202, 206] It will be useful in subsequent sections to note that Eqn. 5.4 can be combined with Eqns. 5.1 and 5.3 such that the calculation of $\Pi(N; \mu, V, \beta)$ is re-expressed as a calculation of $\ln Q_c(N+1, V, \beta)$ in Eqn. 5.8.

$$\ln Q_c(N+1, V, \beta) = \ln Q_c(N, V, \beta) - \ln [q(\beta) \exp(\beta\mu)] + \ln \left[\frac{P(N \rightarrow N+1)}{P(N+1 \rightarrow N)}\right] \tag{5.8}$$

Temperature extrapolation of the free energy landscape

For TMMC, we expressed Eqn. 5.8 in terms of $\ln Q_c(N, V, \beta)$ to demonstrate how to most efficiently perform temperature extrapolation. The canonical partition function at some new temperature, β', is just the product of the kinetic and configurational partition functions at this new temperature, or $Q(N, V, \beta') = q(\beta')^N Q_c(N, V, \beta')$. The kinetic partition function is known exactly at this new temperature since $\ln q(\beta')^N = -3N \ln(\Lambda')$. The configurational partition function is not known analytically; however, it can be estimated at this new temperature by a Taylor series expansion truncated to some order m, as shown in Eqn. 5.9.

$$\ln Q_c(N,V,\beta') \approx \ln Q_c(N,V,\beta) + \sum_{n\geq 1}^{m} \frac{1}{n!} \frac{\partial^n \ln Q_c(N,V,\beta)}{\partial \beta^n} (\beta' - \beta)^n \tag{5.9}$$

Since $\ln Q_c(N,V,\beta)$ is a cumulant generating function and related to the fluctuations in the system's total potential energy, the derivative terms in the Taylor expansion can be evaluated by simply recording the potential energy from the simulation at the original temperature, β, and computing ensemble-averaged quantities, such as:

$$\begin{aligned}
\frac{\partial \ln Q_c(N,V,\beta)}{\partial \beta} &= -\langle E \rangle \\
\frac{\partial^2 \ln Q_c(N,V,\beta)}{\partial \beta^2} &= \left\langle (E - \langle E \rangle)^2 \right\rangle \\
\frac{\partial^3 \ln Q_c(N,V,\beta)}{\partial \beta^3} &= -\left\langle (E - \langle E \rangle)^3 \right\rangle \\
\frac{\partial^4 \ln Q_c(N,V,\beta)}{\partial \beta^4} &= \left\langle (E - \langle E \rangle)^4 \right\rangle - 3\left(\left\langle (E - \langle E \rangle)^2 \right\rangle\right)^2
\end{aligned} \tag{5.10}$$

Higher order terms may also be derived.[203] Using the Taylor series expansion to approximate $\ln q(\beta)$ at a different temperature would have been highly undesirable since the analytic solution is already known for any temperature.[204] This has been avoided by factoring the total partition function so that the Taylor approximation is only applied to $\ln Q_c(N,V,\beta)$. Finally, the extent to which $\ln Q_c(N,V,\beta)$ can be accurately extrapolated at each of the N macrostates depends on how well the moments of the potential energy distribution have converged throughout the simulation. The results section demonstrates that this extrapolation can be valid over a surprisingly large temperature range.

Extension to deformable adsorbates

The formalism presented thus far has implicitly assumed that the particles are rigid bodies with no internal degrees of freedom. Molecules with internal degrees of freedom often require additional sampling biases to explore configurational space effectively. One such approach we focus on is Rosenbluth sampling.[188, 189] Additional complexities must be accounted for in order to perform temperature extrapolation on TMMC simulations of deformable molecules using this sampling scheme, which we will now illustrate.

Rosenbluth sampling

The standard, unbiased acceptance rule for particle insertions in GCMC can become prohibitively small for the simulation of deformable adsorbates, since the vast majority of randomly generated configurations would yield an extremely high internal energy arising from intramolecular potentials (bonds, bends, torsions, etc.). Rosenbluth sampling presents an

efficient way to overcome this obstacle by biasing the growth of chain molecules during a MC move.[188, 190] In the case of flat-histogram MC, it is also necessary to take advantage of such a biasing scheme when accumulating statistics for the C-matrix. The acceptance rule for a MC particle insertion move using Rosenbluth sampling (in the notation of Ref. [183]) is given as follows:

$$\mathrm{acc}(\Gamma_{\mathbf{r}^N} \rightarrow \Gamma_{\mathbf{r}^{N+1}}) = \min\left[1, \frac{Vq(\beta)\exp(\beta\mu_B)}{(N+1)}\mathcal{W}^{ext}(\Gamma_{\mathbf{r}^{N+1}})\right] \tag{5.11}$$

where $\mathcal{W}^{ext}(\Gamma_{\mathbf{r}^{N+1}})$ is the Rosenbluth factor of the inserted molecule and μ_B is the chemical potential of a reservoir of "ideal chain" molecules (see the following section for details). In Rosenbluth sampling, only the chain's external (intermolecular and non-bonded intramolecular) interactions determine its acceptance probability, while reasonable trial configurations are generated using only the intramolecular bonded interactions. An in-depth discussion of this biasing technique and the algorithm for computing $\mathcal{W}^{ext}(\Gamma_{\mathbf{r}^{N+1}})$ is presented extensively in Ref. [183] and elsewhere.[210, 211]

Ideal chain partition function

In order to extrapolate the thermodynamic observables of a system of chain molecules simulated using Rosenbluth sampling,[183] the dependence of the ideal chain partition function on temperature must also be accounted for. An ideal chain is considered to have only intramolecular *bonded* interactions, and so its partition function represents the integration over each *bonded* degree of freedom in the molecule. We focus on propane (C_3H_8) for the remainder of this work which has two bonds and one angle potential such that the ideal chain partition function becomes:

$$Q_{IC,c}(\beta) = \int dr_1 \int dr_2 \int d\theta_1 \exp(-\beta(E_{bond}(r_1) + E_{bond}(r_2) + E_{angle}(\theta_1)). \tag{5.12}$$

$Q_{IC,c}(\beta)$ represents the configurational component of the partition function which has been separated from the integration of kinetic degrees of freedom, $q(\beta)$, such that the total ideal chain partition function becomes $Q_{IC}(\beta) = Q_{IC,c}(\beta)q(\beta)$. The default TraPPE description [212] of propane in the RASPA package [102] was used to evaluate the E_{bond} and E_{angle} terms. Such an integral can be easily handled directly, and the value of the ideal chain partition function can be found by numerical integration for any temperature. However, for larger adsorbates with many internal degrees of freedom, the ideal chain partition may be a much more difficult integral to evaluate. A simulation-based alternative to handle these cases is discussed in a subsequent section of this chapter.

Rosenbluth factor and the chemical potential

For the purposes of understanding temperature extrapolation of flat-histogram simulations with chain molecules, it is important to understand the relationship between the ensemble average of the Rosenbluth factor and the excess chemical potential [211]:

$$\left\langle \mathcal{W}^{ext}(\Gamma_{\mathbf{r}^{N+1}}) \right\rangle = \exp\left[-\beta(\mu - \mu_{IC})\right] \tag{5.13}$$

This equation states that the ensemble average of $\mathcal{W}^{ext}(\Gamma_{\mathbf{r}^{N+1}})$ samples the difference in chemical potential between the fully interacting chain and the ideal chain reference state, μ_{IC}. Consequently, the ensemble average of Rosenbluth trial insertions can be re-expressed as the ratio of the canonical partition functions normalized by the ideal chain partition function.

$$\left\langle \mathcal{W}^{ext}(\Gamma_{\mathbf{r}^{N+1}}) \right\rangle = \frac{Q(N+1,V,\beta)}{Q(N,V,\beta)Q_{IC}(\beta)} \tag{5.14}$$

This can be contrasted with the unbiased GCMC case where the ensemble average of trial insertions yields only the ratio of the canonical partition functions.

Chemical potential of deformable adsorbates

To perform temperature extrapolation of TMMC simulations with Rosenbluth sampling, it is critical to see how the chemical potential of a reservoir of ideal chain molecules depends on the ideal chain partition function. First, consider a reservoir of rigid ideal gas particles:

$$\beta\mu = \beta\mu^o + \ln(\rho) \tag{5.15}$$

Here ρ is the fluid density, which in the case of an ideal gas, is equal to $\beta\mathcal{P}$ where $\mathcal{P}$ is the pressure. For real fluids, we can account for the non-ideality of the reservoir by replacing $\mathcal{P}$ with the fugacity, $f = \phi\mathcal{P}$, after obtaining the fugacity coefficient, ϕ, from an equation of state. The reference state chemical potential, $\beta\mu^o = -\ln q(\beta)$, derives from the canonical partition function of a single ideal gas particle (which for a rigid particle is just the integration over its kinetic degrees of freedom).

If the reservoir consists of ideal chains, the chemical potential of such a fluid takes a similar form as Eqn. 5.15, but the reference state no longer accounts only for an integration over the kinetic degrees of freedom. Now the reference state must also account for the fact that the ideal chain has some potential energy interactions associated with its bonded internal degrees of freedom. Unlike the case of an ideal gas of rigid molecules, now there are interactions that cannot be neglected in the limit of zero density. Thus, the reference state chemical potential is shifted by the temperature dependent ideal chain configurational partition function to give the chemical potential of an ideal chain, μ_{IC}, in Eqn. 5.16.[183, 187]

$$\beta\mu_{IC} = -\ln Q_{IC}(\beta) = -\ln q(\beta) - \ln Q_{IC,c}(\beta) \tag{5.16}$$

The chemical potential of a reservoir of these ideal chain molecules, μ_B, is similar to Eqn. 5.15 where the reference state $\beta\mu^o$ has been replaced by $\beta\mu_{IC}$.

$$\beta\mu_B = [-\ln q(\beta) - \ln Q_{IC,c}(\beta) + \ln(\beta\phi\mathcal{P})] + C \tag{5.17}$$

Note we have introduced a temperature-dependent shift to the reservoir chemical potential in Eqn. 5.17, where $C = -\ln\langle\mathcal{W}_{IG}^{ext}(\beta)\rangle$. Known as the ideal gas Rosenbluth weight, $\mathcal{W}_{IG}^{ext}$ accounts for when non-bonded intramolecular interactions can contribute to an isolated chain's partition function due to the molecule's size (e.g., 1-5 pair interactions). It's inclusion in Eqn. 5.17 is derived in detail elsewhere.[187, 210, 211] By using propane as our flexible adsorbate we can presently ignore this possibility ($\langle\mathcal{W}_{IG}^{ext}(\beta)\rangle = 1$), which allows us to proceed with a slightly more concise derivation of the temperature extrapolation in subsequent sections. As in the case of rigid adsorbates, we can account for the non-ideality of the reservoir phase by replacing $\mathcal{P}$ with f and obtaining the fugacity coefficient from an equation of state. Alternatively, direct simulations of the bulk fluid can provide this information if a reliable equation of state is not readily available.[205]

Constructing and extrapolating the macrostate distribution

The TMMC formulation for chain molecules now becomes slightly more complex when using Rosenbluth sampling. This is because Rosenbluth sampling for particle insertions only yields the excess chemical potential relative to the ideal chain, a topic which we expand upon in the following section (this discussion closely relates to the formalism presented in Ref. [213] where the excess chemical potential of the ideal gas reference state of chain molecules was also computed in the context of a TMMC study). Thus, in order to properly compute the value of the total configurational partition function, one must add the contribution from the ideal chain when calculating $\ln Q_c(N, V, \beta)$ at each successive N macrostate, as shown in Eqn. 5.18.

$$\ln Q_c(N+1, V, \beta) = \ln Q_{IC,c}(\beta) + \ln Q_c(N, V, \beta) - \ln\left[q(\beta)\exp(\beta\mu_B)\right] + \ln\left[\frac{P(N \to N+1)}{P(N+1 \to N)}\right] \tag{5.18}$$

Thus, from a TMMC simulation at a chosen β, we may construct $\ln Q_c(N, V, \beta)$ using the transition probabilities, P, computed from the C-matrix; what remains is to then predict the macrostate distribution, $\ln\Pi(N, V, \beta')$, at arbitrary β' via extrapolation. Following Eqn. 5.3, assuming the fugacity coefficient is known from an equation of state, we obtain:

$$
\begin{aligned}
\ln \Pi(N;\mu,V,\beta') &\sim \beta'\mu N + \ln Q(N,V,\beta') \\
&\sim \left[\ln(\beta'\phi'\mathcal{P}) - \ln q(\beta') - \ln Q_{IC,c}(\beta')\right] N + \ln Q_c(N,V,\beta') + \ln Q_k(N,\beta') \\
&\sim \left[\ln(\beta'\phi'\mathcal{P}) - \ln Q_{IC,c}(\beta')\right] N + \ln Q_c(N,V,\beta') \\
&\sim \left[\ln(\beta'\phi'\mathcal{P}) - \ln Q_{IC,c}(\beta')\right] N + \ln Q_c(N,V,\beta) + \\
&\qquad \sum_{n\geq 1}^{m} \frac{1}{n!} \frac{\partial^n \ln Q_c(N,V,\beta)}{\partial \beta^n} (\beta' - \beta)^n \\
&\sim N \ln\left(\beta'\phi'\mathcal{P}\right) + \left[\ln Q_c(N,V,\beta) - N \ln Q_{IC,c}(\beta')\right] + \\
&\qquad \sum_{n\geq 1}^{m} \frac{1}{n!} \frac{\partial^n \ln Q_c(N,V,\beta)}{\partial \beta^n} (\beta' - \beta)^n
\end{aligned}
\tag{5.19}
$$

Following Eqn. 5.18, observe that $\ln Q_c(N,V,\beta)$ in the macrostate probabilities implicitly contains a contribution of $N \times \ln Q_{IC,c}(\beta)$ since it is constructed iteratively from $P(N)$ starting from the $N = 0$ state. Thus, the bracketed term in the final line explicitly shows that differences between the ideal chain partition functions at the simulation temperature and the temperature being extrapolated to contribute to the macrostate probabilities at a chosen $\mathcal{P}$ and ϕ. Note that in Eqn. 5.19, $Q_{IC,c}(\beta) = 1$ for rigid adsorbates. By summing the terms in Eqn. 5.19, we obtain an estimate of the grand canonical partition function (Eqn. 5.2), which serves to normalize the macrostate probabilities. Just as in the case of adsorbates with no internal degrees of freedom, one uses the moments of the total potential energy distribution (which now included all bonded and non-bonded energies) to evaluate terms in the Taylor series of $\ln Q_c(N,V,\beta)$.

Computing the ideal chain partition function's temperature dependence

Previously we described a situation in which we could numerically integrate the ideal chain partition function, $Q_{IC,c}(\beta)$, for propane since this requires an integration in only three dimensions. Now suppose the chain molecule of interest contains many internal degrees of freedom such as heptane, an alkane chain with seven monomer units. In this case, determining $Q_{IC,c}(\beta)$ would consist of an integration over 6 bonded interactions, 5 bend (angle) interactions, and 4 dihedral interactions based on the TraPPE force field implementation in RASPA. This high dimensional integral is too complex for numerical integration. Although we envision this quantity could be sampled by a Monte Carlo integration and the use of some modified Rosenbluth sampling scheme, a much easier shortcut is available to us. Since this quantity only serves as a reference state for the Rosenbluth sampling scheme, we do not strictly need to obtain the actual value of $Q_{IC,c}(\beta)$ to temperature extrapolate thermodynamic observables. Rather we only need to know its ratio between the simulation temperature, β, and the extrapolation temperature, β'.

To simulate an ideal chain, one can perform an NVT simulation of a single chain in an empty box. If the molecule has intramolecular non-bonded interactions, these should be turned off to satisfy the definition of an ideal chain. We can then extrapolate $\ln Q_{IC,c}(\beta)$ to some β' using a Taylor series approximation as shown in Eqn. 5.20.

$$\ln Q_{IC,c}(\beta') \approx \ln Q_{IC,c}(\beta) + \sum_{n\geq 1}^{m} \frac{1}{n!} \frac{\partial^n \ln Q_{IC,c}(\beta)}{\partial \beta^n} (\beta' - \beta)^n \tag{5.20}$$

Exploiting the cumulant generating properties of the canonical partition function, we can solve for the ratio of the ideal chain partition function between the simulation temperature and the extrapolation temperature by relating the partial derivatives in the summation to the total potential energy fluctuations, E_{tot}. Since only intramolecular bonded interactions exist in the simulation of the ideal chain, the total potential energy is just equal to the total intramolecular bonded interactions, or $E_{tot} = E_{int}$.

$$\begin{aligned}
\frac{\partial \ln Q_{IC,c}(\beta)}{\partial \beta} &= -\langle E_{int} \rangle \\
\frac{\partial^2 \ln Q_{IC,c}(\beta)}{\partial \beta^2} &= \left\langle (E_{int} - \langle E_{int} \rangle)^2 \right\rangle \\
\frac{\partial^3 \ln Q_{IC,c}(\beta)}{\partial \beta^3} &= -\left\langle (E_{int} - \langle E_{int} \rangle)^3 \right\rangle \\
\frac{\partial^4 \ln Q_{IC,c}(\beta)}{\partial \beta^4} &= \left\langle (E_{int} - \langle E_{int} \rangle)^4 \right\rangle - 3\left(\left\langle (E_{int} - \langle E_{int} \rangle)^2 \right\rangle\right)^2
\end{aligned} \tag{5.21}$$

These partial derivatives allow us to accurately evaluate the summation in Eqn. 5.20 assuming sufficient statistics in the energy fluctuations have been collected. For propane, we can compare the values obtained from a direct numerical integration of both $Q_{IC,c}(\beta)$ and $Q_{IC,c}(\beta')$ versus the method shown in Eqn. 5.20. Table 5.1 demonstrates that both methods yield excellent agreement. Just like the calculation of the ideal gas Rosenbluth weight for GCMC simulations of chain molecules, this simulation of the ideal chain must only be performed once for use in any subsequent flat-histogram temperature extrapolation analyses.

Table 5.1: The value of propane's ideal partition function as a function of some new temperature, T', is obtained from direct numerical integration (NI). The ratio of the partition functions obtained from NI is compared to the extrapolation method using potential energy fluctuations (Eqn. 5.20) measured at $T = 400$ K.

T' [K]	$Q_{IC,c}(\beta')$ (NI)	$\ln \frac{Q_{IC,c}(\beta)}{Q_{IC,c}(\beta')}$ (NI)	$\ln \frac{Q_{IC,c}(\beta)}{Q_{IC,c}(\beta')}$ (Eqn. 5.20)
350	0.00427	0.2003	0.2002
360	0.00446	0.1580	0.1580
370	0.00465	0.1169	0.1169
380	0.00484	0.0769	0.0769
390	0.00503	0.0380	0.0380
400	0.00522	0.0000	0.0000
410	0.00542	-0.0370	-0.0370
420	0.00562	-0.0732	-0.0731
430	0.00582	-0.1085	-0.1084
440	0.00603	-0.1430	-0.1428
450	0.00623	-0.1767	-0.1765

Extraction of thermodynamic properties

Temperature extrapolation of $\ln Q(N, V, \beta)$ yields continuous thermodynamic properties that cannot readily be obtained from standard GCMC simulations. For example, isotherms and isosteric heats of adsorption can be computed for any combination of reservoir (T,$\mathcal{P}$) state points via a computationally cheap post-processing of the $\ln Q(N, V, \beta)$ data. These are typically the first two important quantities considered when evaluating a material's potential as an adsorbent.

Isotherm prediction

The adsorption isotherm at a given temperature, β', and chemical potential, μ', can be resolved by computing the expectation number of particles in the system via Eqn. 5.22 for a range of pressures (or equivalently, chemical potentials).

$$\langle N \rangle_{\mu',V,\beta'} = \sum_{N} N \cdot \Pi(N; \mu', V, \beta') \tag{5.22}$$

The ability to calculate this quantity at any μ' (via re-weighting) and β' (subsequent extrapolation) from simulation data obtained at some different, single (μ,β) point results in significantly more predictive power than running simple GCMC simulations for individual (μ,β) state points.

Enthalpy of adsorption

Another important thermodynamic quantity for predicting an adsorbent material's utility is the isosteric heat (or enthalpy of adsorption). The isosteric heat of adsorption measures how much enthalpy is released when an additional molecule is adsorbed and plays an important role in calculating how much energy is required when cycling between adsorption and desorption conditions. The isosteric heat of adsorption can be determined from several statistical mechanics approaches. Since the standard energy-particle fluctuation method can suffer from poor statistical convergence, Vlugt et al. proposed determining the isosteric heat from a series of NVT simulations with Widom insertions (which just requires a simple post-processing step of our TMMC simulation data).[214] In this work, however, we chose to calculate the isosteric heat using the Clausius-Clapeyron equation. Isotherms can be obtained for arbitrarily small spacing between different T and P state points via the aforementioned post-processing and temperature extrapolation, therefore the derivative in the Clausius-Clapeyron equation (Eqn. 5.23) can be numerically evaluated.

$$q_{st}(N,T) = -RT^2 \left(\frac{\partial \ln \mathcal{P}}{\partial T} \right)_N \tag{5.23}$$

Thus, q_{st} as a function of N and T is known over the temperature range at which the temperature extrapolation is valid and the total enthalpy change between two states (N_1, T_1) and (N_2, T_2) is easily calculated by integrating Eqn. 5.23.[215] As noted in Ref. [215] one can replace $\mathcal{P}$ with the fugacity to account for the non-ideality of the vapor phase. While the isosteric heat often has only a weak dependence on temperature, it is interesting that this dependence can be explicitly computed via this method. In systems where the heat of adsorption has a strong dependence on temperature, the use of flat histogram simulations to compute isosteric heats via this method provides a highly attractive alternative to GCMC.

5.3 A simplified computational approach for generating flat-histogram probability distributions

Adsorbate/adsorbent system

Results in this work focus on the performance of the flat-histogram simulation of single-component adsorbate thermodynamic properties of CH_4, CO_2, and C_3H_8 in MOF-950. MOF-950, visualized in Fig. 5.2, was chosen as a model system because it was recently shown to have a high working capacity for methane.[216] CH_4 and CO_2 were modeled as rigid adsorbates with the united-atom TraPPE and EPM2 forcefields, respectively.[212, 217] C_3H_8 was modeled as a flexible adsorbate, also with the united-atom TraPPE forcefield. The molecule definition files for these adsorbates are provided in the standard RASPA installation. The framework was modeled as rigid, and the Universal Force Field parameters were adopted

for the MOF atoms.[73] Interaction parameters between the MOF and adsorbate were determined by Lorentz-Berthelot combining rules. The MOF-950 crystallographic information file is included in Ref. [112].

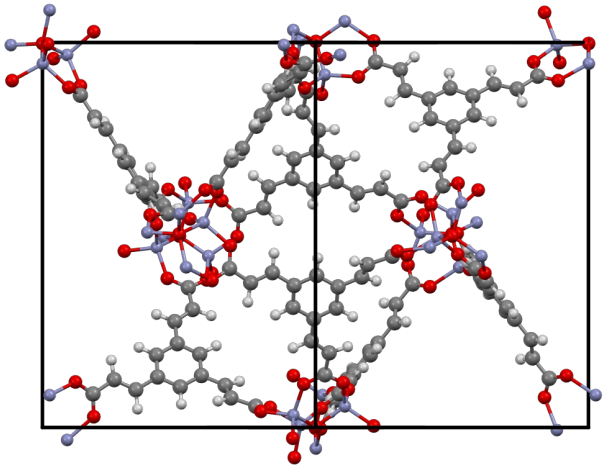


Figure 5.2: Visual representation of the MOF-950 framework with the viewing plane parallel to the 011 face.

Transition matrix MC implementation

Before using TMMC, one typically wants to obtain a *good* initial estimate of the macrostate probabilities by quickly exploring the entire macrostate space via the WL algorithm. While TMMC can technically operate with any initial biasing array, it will more efficiently explore macrostate space if the initial biasing array approximates the true macrostate probabilities which are roughly determined from a short WL run. In practice, macrostate space is also distributed in discrete chunks for parallel computation. The results of the individual simulations are then stitched back together to obtain the free energy across the entire macrostate space. Significant previous work has focused on determining the optimal way to distribute the macrostate blocks to each individual flat-histogram simulation.[218, 219] In this work we alternatively solve the macrostate probabilities, or $\ln Q(N, V, \beta)$, by running many simulations in the canonical ensemble for each value of N while performing Widom insertions.[183] By imposing constant N simulations, we can artificially make the "observation" of each macrostate equally likely by performing $N = 1 \ldots N_{max}$ different NVT simulations. Therefore, it is not necessary to ever construct or update a biasing function to make the sampling probability distribution flat as in WL/TMMC for effective macrostate exploration; however, we can still use the principles of TMMC to obtain the transition probabilities via Widom insertions and deletions. This is formally equivalent to taking a window size of 1, as has been already performed in Ref. [198]. In other words, performing a flat-histogram simulation with

a window size of 1 reduces the flat-histogram simulation to an N_{max} number of NVT simulations where the C-matrix statistics are consequently populated by Widom insertions and deletions. For simplicity, we henceforth refer to a TMMC simulation with window size of 1 as an NVT+W simulation. It should be noted that this ghost insertion approach in individual canonical ensemble simulations works well for supercritical fluids, which are the conditions of interest for high-throughput screening of materials' separation performance for light gases, but we have not explored its efficacy with temperature extrapolation in subcritical regimes where phase coexistence generates a bimodal macrostate distribution.

Simulation details

All simulations were performed with the RASPA code.[102] NVT+W simulations were performed to obtain the data necessary to generate temperature extrapolated isotherms and isosteric heats, and the accuracy of these temperature extrapolated thermodynamic properties are compared to standard GCMC simulations. For GCMC simulations, 5×10^4 cycles were utilized for both equilibration and production. Each NVT+W simulation was equilibrated for 5×10^3 cycles, and 4×10^3 cycles were used to accumulate the statistics necessary to construct the C-matrix, where translation/rotation/regrow/Widom moves were proposed in a ratio of 1/1/1/2. Cycles were defined such that this led to a total of at least $3.2 \times 10^4 \times N$ Widom insertions and deletions acquired for each NVT simulation. For each adsorbate, the Peng-Robinson equation of state was used to compute the fugacity coefficient of the reservoir fluid for a given temperature and pressure,[220] under the assumption that it faithfully captures the bulk behavior of the simulation model. The input files required to run the simulations are provided in Ref. [221].

5.4 Generating continuous thermodynamic observables to predict material performance

Isotherms of CH_4, CO_2, and C_3H_8 in MOF-950

Fig. 5.3 shows the isotherms computed from NVT+W and GCMC simulations for rigid CH_4 and CO_2 and deformable C_3H_8 in MOF-950. In this figure, open circles represent GCMC predicted loadings, while dashed lines represent the calculation of an isotherm at the color-coded temperature from an NVT+W simulation at that same temperature. Solid lines represent the temperature extrapolated isotherms from the NVT+W simulation data originally obtained at one single temperature, T_{sim}, using up to third order terms ($m = 3$) in Eqn. 5.9. The temperature extrapolated isotherms for CH_4 (Fig. 5.3a-b), CO_2 (Fig. 5.3c-d) and C_3H_8 (Fig. 5.3e-f) are generated from NVT+W simulations at $T_{sim} = 270$ K, $T_{sim} = 300$ K and $T_{sim} = 400$ K, respectively. In all cases the GCMC isotherms match the temperature extrapolated isotherms over an extremely large temperature extrapolation range, and the log-log representation of the data shows that the Henry coefficients match exactly. Interestingly,

the temperature extrapolation for CO_2 perfectly capture the non-Langmuirian shape (low pressure inflection) of the adsorption isotherms for all temperatures.

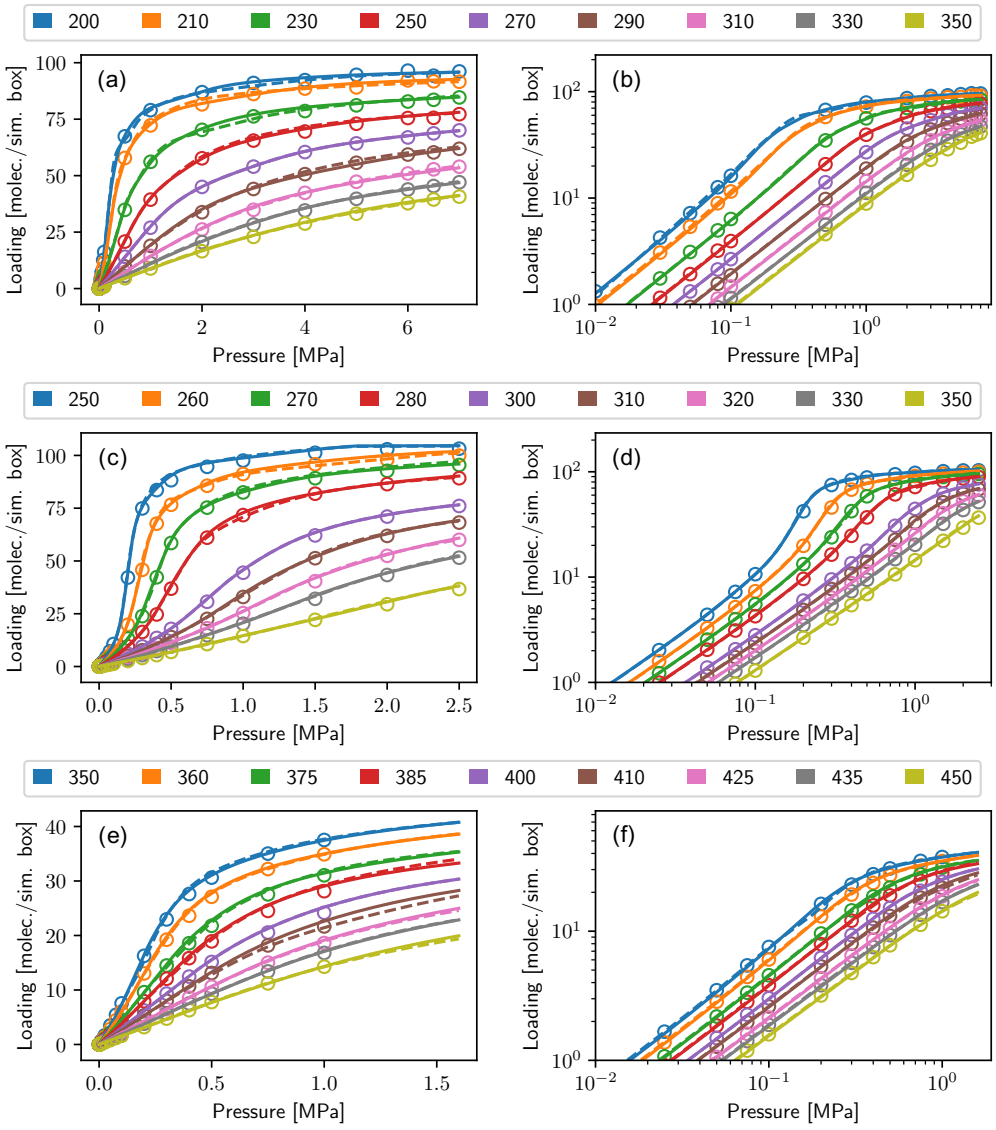


Figure 5.3: CH_4 (a-b), CO_2 (c-d), and C_3H_8 (e-f) isotherms are shown for MOF-950. (b), (d), and (f) are plotted on log-log axes to show the agreement between all calculations in the low-pressure, Henry's Law regime. Open circles represent GCMC simulations, dashed lines represent NVT+W calculated isotherms from a simulation at the specified temperature, and solid lines represent temperature extrapolated isotherms from an NVT+W simulation originally performed at a single temperature ($T_{sim} = 270$ K for CH_4, $T_{sim} = 300$ K for CO_2, and $T_{sim} = 400$ K for C_3H_8).

To evaluate a material's potential as an adsorbent, one is often interested in calculating working capacity, or the difference in uptake between the adsorption and desorption temperature and pressure. Since we can evaluate the number of particles adsorbed (Eqn. 5.22) for any given number of temperatures and pressures by post-processing of the $\ln Q_c(N, V, \beta)$ simulation data, the working capacity can be easily calculated via Eqn. 5.24.

$$n_{wc} = \langle N \rangle_{\mu_2 V \beta_2} - \langle N \rangle_{\mu_1 V \beta_1} \tag{5.24}$$

In a combined temperature and pressure swing adsorption (T/PSA) process, both temperature and pressure change between the adsorption and desorption conditions; however, an NVT+W simulation at only one temperature is needed to obtain any T/PSA working capacity, assuming that the Taylor series is sufficient to approximate the temperature dependence of $\ln Q_c(N, V, \beta)$. From an engineering perspective, this is a marked advantage of NVT+W simulations over unbiased GCMC since we can numerically compute nearly continuous working capacities as a function of $(\mu_2, \beta_2, \mu_1, \beta_1)$ state points and use this data as input into process simulation software.

Isosteric heats of CH_4, CO_2, and C_3H_8 in MOF-950

The isosteric heat of adsorption was computed for each adsorbate in MOF-950. Using the temperature extrapolated isotherms where $\langle N \rangle_{\mu VT}$ was generated on a grid of $\Delta T = 2$ K and $\Delta P = 250$ Pa intervals, the derivative in the Clausius-Clapeyron equation was numerically evaluated. Fig. 5.4 shows that the heat of adsorption predicted via this methodology produces the same results as GCMC simulations which use the particle-energy fluctuation method at a fixed temperature.[214]

However, an added benefit can be achieved by using the flat-histogram technique to solve the isosteric heats. Fig. 5.5 shows the isosteric heat of adsorption, $q_{st}(N, T)$, plotted as a continuous function of temperature and loading for CO_2 in MOF-950. This data is only accessible due to the evaluation of the Clausius-Clapeyron equation on the aforementioned grid of $\langle N \rangle_{\mu VT}$ values, which would otherwise take a large number of GCMC simulations at many different temperatures and pressures to obtain. As expected the temperature dependence of the isosteric heat is small, but interestingly the temperature extrapolated NVT+W data combined with the Clausius-Clapeyron analysis quantifies the subtle dependence of q_{st} on temperature. Another advantage of obtaining the data visualized in Fig. 5.5 is that the total change in enthalpy between any two states (termed the enthalpy of immersion in Ref. [215]) can be determined by numerical integration of the isosteric heat values along the path between the two states. Also note that, since we have determined $\ln Q(N, V, \beta)$, one can also compute the entropy of adsorption as a function of loading via the Helmholtz free energy, $F(N, V, \beta) = -k_B T \ln Q(N, V, \beta)$.

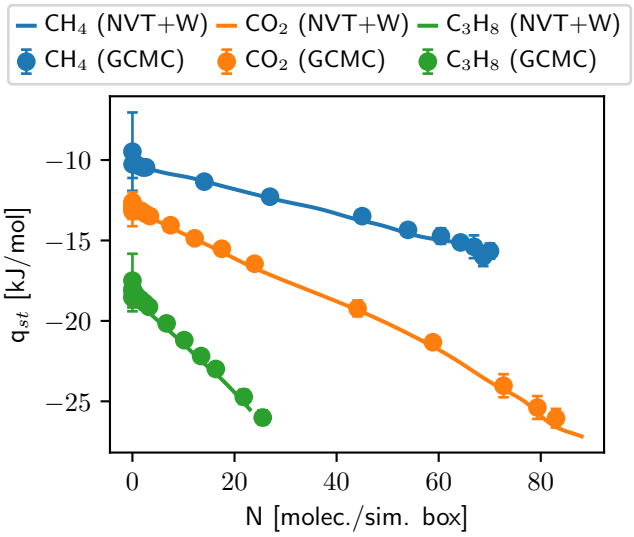


Figure 5.4: The isosteric heats of adsorption are plotted as a function of loading for CH_4 (blue), CO_2 (orange), and C_3H_8 (green) at temperatures of 270 K, 300 K, and 400 K, respectively.

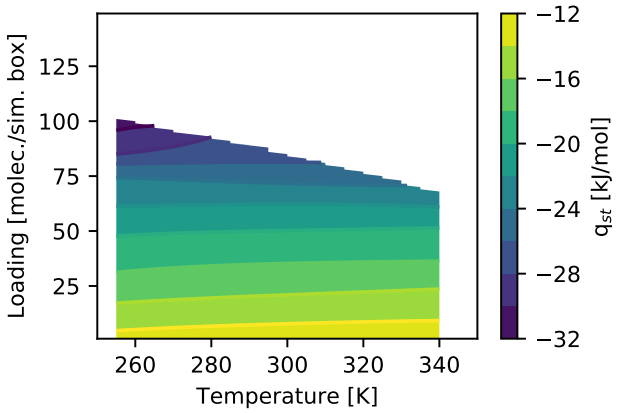


Figure 5.5: The full isosteric heat diagram, $q_{st}(N,T)$, is shown for CO_2 in MOF-950.

Statistical performance of temperature extrapolation

The temperature range over which the macrostate distribution can be extrapolated depends on the order at which the Taylor series approximation is truncated and how well the moments of the potential energy distribution have converged. To show the accuracy of the temperature extrapolation, we performed NVT+W simulations for a large number of β val-

ues to compute $\ln Q_{true}(N, V, \beta)$. Next we took the values from one particular temperature, β_{sim}, and extrapolated them to every other temperature, denoted by $\ln Q_{extrap,\beta_{sim}}(N, V, \beta)$. The relative error (RE) can be computed for each N macrostates and each extrapolation temperature β by Eqn. 5.25, and the results of this evaluation for all three adsorbates in MOF-950 are shown in Fig. 5.6. These are encouraging results for our ability to effectively screen continuous thermodynamic adsorption properties of light gases in the supercritical regime since the range over which the macrostates can be extrapolated is quite large. For the more difficult cases of deeply subcritical fluid regimes, future investigation on the efficacy of this temperature extrapolation is warranted.

$$RE(\beta, \beta_{sim}) = 100 \times \frac{\ln Q_{true}(N, V, \beta) - \ln Q_{extrap,\beta_{sim}}(N, V, \beta)}{\ln Q_{true}(N, V, \beta)} \tag{5.25}$$

Two characteristic decreases in performance can be observed from Fig. 5.6. First, the RE performance decreases for NVT states corresponding to very high particle densities, most notably with CO_2. However, the regions of highest percent error for CH_4 and CO_2 ($|RE| > 0.5\%$) correspond to states with density greater than the saturation loading as can be seen from the isotherm analysis in Fig. 5.3. It is important to choose N_{max} greater than the one's preliminary estimate of the saturation loading to avoid missing non-zero contributions at states that contribute significantly to the system's average properties, but the poor sampling well above the actual saturation loading has little impact on the results since they contribute negligibly (have sufficiently low value of $\Pi(N)$) for any relevant temperature and pressure values.

The more important characteristic performance decrease arises at the extremes of the temperature extrapolation range when a 1^{st} order Taylor approximation is used, i.e., $m = 1$ in Eqn. 5.9. This suggests using a 1^{st} order Taylor approximation is not sufficient to capture the curvature of $\ln Q(N, V, \beta)$ as a function of β as one might expect when extrapolating over a large range of $\Delta\beta$. Interestingly, the accuracy of the $m \geq 2$ extrapolations are essentially identical, indicating that the $m \geq 3$ order terms contribute negligibly to the expansion of $\ln Q(N, V, \beta)$ for these systems at these temperatures. For most of the extrapolation space, the error is within $\pm$ 0.5 % which leads to excellent prediction of thermodynamic properties across the entire temperature range.

Advantages of the flat-histogram approach

First, we highlight some of the significant advantages of these flat-histogram simulations over GCMC. Most importantly, flat-histogram simulations combined with temperature extrapolation yield the macrostate distribution as a function of adsorbed particle number over a large range of temperatures. Hence, thermodynamic observables can be predicted for a broad range of conditions from a computationally cheap post-processing of simulation data obtained at a single temperature for both rigid and, as we have demonstrated, deformable adsorbates. The temperature extrapolation was not only accurate in the tested range of $\Delta T = 100 - 150$ K, but could also be achieved with only low order moments ($m < 3$ in

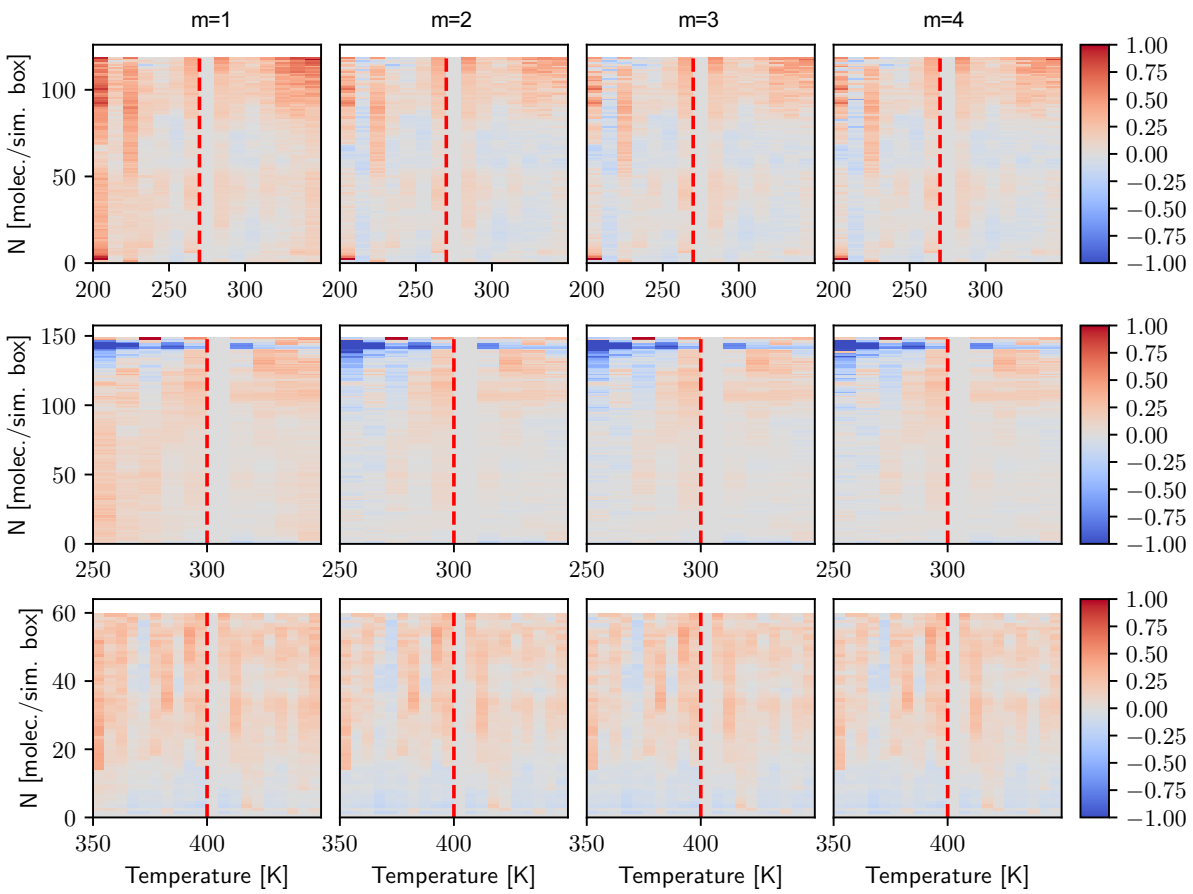


Figure 5.6: The first, second, and third rows of data correspond to the RE of temperature extrapolation for CH_4, CO_2, and C_3H_8, respectively. The data was extrapolated from the simulation temperatures denoted by red lines which are located at 270 K, 300 K, and 400 K for for CH_4, CO_2, and C_3H_8, respectively. Each column represents the Taylor approximation with largest order m (Eqn. 5.9). Finally, the color code is the RE of the extrapolation estimate, given by Eqn. 5.25.

Eqn. 5.9) of the potential energy distribution. This is highly encouraging as it indicates that applying this technique to study light gas adsorption under relevant supercritical (T, P) process conditions does not need computationally demanding convergence of higher order moments because the curvature of the partition function is sufficiently small and can be accurately described by a low order Taylor expansion.

Since this post-processing of simulation data can be performed for arbitrarily small differences in state points, this is a powerful technique if the thermodynamic properties of an adsorbent needs to be obtained for a continuous range of temperatures and pressures, e.g., as input for process optimization software. Consider in Fig. 5.3a that there are approximately

100 GCMC state points plotted, yet only approximately 100 total NVT+W simulations were required to generate significantly more useful data. An arbitrarily large number of GCMC simulations would be required to interpolate the expected number of particles as a function of temperature and pressure with the same accuracy as these flat-histogram methods. In this way, flat-histogram simulations can be vastly more efficient, depending on the goal of the simulation. Also, accumulating statistics at high density states becomes more efficient with the NVT+W approach since constant N simulations are imposed for high density macrostates and the thermodynamic properties of the system are extracted from Widom insertions. In contrast, equilibrating a GCMC simulation at an extremely high density can be slow to converge due to infrequent acceptance of the particle insertion moves. The major drawback of these flat-histogram methods compared to GCMC becomes evident if one only needs to compute $\langle N \rangle_{\mu,V,\beta}$ at a very limited number of state points.

Finally, we comment on the choice of using the NVT+W scheme, i.e., a TMMC scheme with a window size of 1, rather than a traditional binning scheme where each bin contains a range of macrostates. While optimal binning size is an important technical consideration, we chose to run TMMC in the NVT+W scheme to simplify several implementation details. The fundamental difference between the two implementations is that NVT+W simply uses Widom insertions to compute transition probabilities and uses many canonical ensemble simulations to quickly build up the initial estimate of the macrostate probabilities. WL/TMMC methods often require at least an initial pass with the WL algorithm to obtain a reasonably good starting estimate for the macrostate probabilities, enabling a simulation to visit high free energy states. Therefore no implementation of WL is necessary in a MC package when generating NVT+W data. Performing a traditional WL/TMMC simulation requires performing C-matrix updates on the fly, implementing biased acceptance rules, and updating biasing arrays, all of which has indeed been implemented before in freely available packages.[222] In the case of NVT+W, the MC code simply needs to write the unbiased acceptance criteria to disk, and all macrostate probability (TMMC) calculations can be done in post-processing. This may be useful to researchers working with multiple different MC codes for which there are no implemented WL/TMMC functionalities. A minor disadvantage of NVT+W is that additional simulation time is spent equilibrating each individual NVT simulation, an issue not experienced with WL/TMMC. Further investigation is also required before utilizing this scheme in deeply subcritical, phase coexistence regimes.

5.5 An outlook for flat-histogram simulations in high-throughput screening of adsorption thermodynamics

We have outlined the practical utility of flat-histogram Monte Carlo simulations that, when combined with temperature extrapolation, provide significantly more thermodynamic information regarding adsorption in nanoporous materials than traditional GCMC simulations.

The formalism for the temperature extrapolation of macrostate probabilities was extended for the first time to handle molecules with intramolecular degrees of freedom simulated using Rosenbluth sampling. This led to successful application of the method on both rigid and deformable adsorbates. Consequently, the working capacity and working enthalpy for a T/PSA process over an extremely wide range of conditions can also be ascertained from a simulation at only one temperature. Thus, the results presented here demonstrate the potential of these simulation techniques for *in silico* identification and study of high-performing adsorbents for realistic T/PSA processes. Mechanistically, the TMMC simulations were executed using a window size of 1 such that the problem reduces to a series of NVT simulations with Widom insertions used to compute the transition probabilities.

If the simulations community moves towards flat-histogram techniques, a material's performance can immediately be re-calculated when specified process conditions change. Rather than repeating a MC simulation at the new conditions, this can be achieved by simply reprocessing the free energy landscape from our original simulation. In other words we significantly increase the recyclability of our simulation results by performing flat-histogram MC simulations. Similar to the way that databases are currently being developed to store and benchmark results in computational material science,[223, 224] developing databases of free energy landscapes for different pure components or mixtures will provide more robust data as we continue to screen potential adsorbents for separations and storage applications. There also remain exciting opportunities to extend these methods to adsorption in nanoporous materials involving mixtures, for which a theoretical framework has already been developed.[198, 203] Upon further method development and application of more advanced sampling techniques to multi-component mixtures, these methods could lead to numerical, "model-free" optimization of the operating conditions that minimize the parasitic energy of separations in porous materials.[76, 110]

Chapter 6

Advancing the accuracy of high-throughput screening of adsorption thermodynamics

For applications of metal-organic frameworks (MOFs) in gas storage and separation, flexibility is often seen as a parameter that can tune material performance. In this work we aim to determine the optimal flexibility for the shape selective separation of similarly sized molecules (e.g. Xe/Kr mixtures). To obtain systematic insight into how the flexibility impacts this type of separation we develop a simple analytical model that predicts a material's Henry regime adsorption and selectivity as a function of flexibility. We elucidate the complex dependence of selectivity on a framework's intrinsic flexibility whereby performance is either improved or reduced with increasing flexibility, depending on the material's pore size characteristics. However, the selectivity of a material with the pore size and chemistry that already maximizes selectivity in the rigid approximation is continuously diminished with increasing flexibility, demonstrating that the globally optimal separation exists within an entirely rigid pore. Molecular simulations show that our simple model predicts performance trends that are observed when screening the adsorption behavior of flexible MOFs. These flexible simulations provide better agreement with experimental adsorption data in a high performance material that is not captured when modeling this framework as rigid, an approximation typically made in high-throughput screening studies. We conclude that, for shape selective adsorption applications, the *globally* optimal material will have the optimal pore size/chemistry *and* minimal intrinsic flexibility even though other non-optimal materials' selectivity can actually be improved by flexibility. Equally important, we find that flexible simulations can be critical for correctly modeling adsorption in these types of systems. This is an unofficial adaptation of an article that appeared in an ACS publication [225]. ACS has not endorsed the content of this adaptation or the context of its use.

6.1 High-throughput screening of adsorption capabilities

Metal-organic frameworks (MOFs) have garnered significant attention in the scientific community due to their potential applications ranging from gas storage and separations to catalysis and sensors.[25, 30, 226–229] The diversity of applications for which MOFs are currently being investigated arises primarily from their highly tunable chemical and geometric properties since the combinatorics of their nodal composition, consisting of secondary building units (SBUs) and ligands, leads to an enormous number of possible structures.[13] An interesting consequence of such structural diversity is that the flexibility of these materials can be exploited in various ways to yield improved performance in a variety of applications.[230] Focusing on gas storage and separations, flexibility has been experimentally and computationally examined in the contexts of breathing,[52, 231, 232] swelling,[17] rotatable and flexible linkers,[233, 234] shape memory,[235] complex lattice deformation,[236] subnetwork displacement,[237] negative gas adsorption,[238, 239] and its strong influence on diffusion,[240] all with the ultimate goal of exploiting these phenomena to obtain desirable adsorption properties. Such breadth presents interesting challenges and opportunities for experimental and computational characterization of flexibility in MOFs and for determination of how it can be utilized for the design of better gas adsorbents.

MOFs with dynamic constituents (i.e. rotatable and flexible ligands) [233] have pore spaces that can change in size and/or shape while the unit cell volume and shape remains constant. A natural question that arises is whether this phenomenon can be optimized for specific applications in gas separations, and our work highlights how controlling this intrinsic flexibility of MOFs can optimize these materials for shape selective adsorption. An example of such an application is the separation of Xe/Kr mixtures where the most selective materials have a pore size and shape commensurate with the adsorbates,[241] and thus we illustrate our findings in the context of the widely studied Xe/Kr separations.[242, 243]

To study the effect of intrinsic flexibility on the separation of tight fitting molecules we have developed a simple model that allows us to compute the selectivity in the Henry regime as a function of the flexibility of a material. Our model quantitatively and systematically demonstrates that intrinsic flexibility can either increase or decrease selectivity for a given material based on its pore size characteristics. However, an interesting consequence will follow: achieving the *globally* optimal separation performance necessitates not only having a material with the ideal pore size and chemistry but also finding materials that simultaneously minimize intrinsic flexibility. Next we compare these results to Xe/Kr adsorption data obtained from screening the CoRE MOF database [244] when each structure's flexibility is modeled via a classical force field (FF) recently used in the literature.[165] It is important to note that when high-throughput computational studies are employed to elucidate structure-property relationships,[18–21] they are almost exclusively performed using the rigid structure approximation (see details of Ref. [245] for the only exception to our knowledge). This approximation is considered a safe assumption when thermodynamic fluctuations are assumed

to average out the framework atoms' locations to their initial values, and so intrinsic flexibility is assumed to have no effect. However, by relaxing this approximation and using flexible simulations to screen a database of MOFs, we observe that flexibility has the same impact on adsorption behavior trends as shown by the analytical model. We then select several MOFs from the screening to study in further detail the effect of flexibility on each material's performance in Xe/Kr separations and finally show that flexibility is necessary to yield better agreement with experimental adsorption data of a high performance MOF system known as SBMOF-1.[243] Thus we not only quantitatively and systematically demonstrate the impacts of flexibility on shape selective adsorption through models and simulations but are able to answer the more fundamental question of whether controlling intrinsic flexibility can optimize material performance in this type of separation application.

6.2 Analytical model for adsorption selectivity as a function of flexibility

The effects of intrinsic flexibility in MOFs on Henry regime adsorption are elucidated by an analytical model of a flexible pore and direct simulations of flexible MOF materials. The details of the direct simulations are explained in the subsequent Methods section. For our analytical model, we construct a spherical pore of radius R_p consisting of a wall of carbon atoms. The visualization of this model is presented in Fig. 6.1. A continuum approximation [246] is invoked such that an adsorbate does not interact with discrete atomic centers but rather with a uniform surface density of atoms, $\eta = 1$ atom $/(\pi * 1.2\text{Å}^2)$. This surface density is chosen to be a slightly higher than that of graphene since adsorbates can interact with atoms beyond the pore wall in real MOF materials. This approximation will permit an integrable expression for the Henry coefficient, or the measure of a material's affinity for an adsorbate in the limit of infinite dilution.[183]
The adsorption energy of one adsorbate within the spherical pore is dependent only on the r coordinate due to spherical symmetry as shown in Fig. 6.1. Determination of this energy requires an integration of the host-adsorbate interaction across the pore surface, which in turn requires an expression for the distance between the adsorbate and any point on the surface, given by d_w in Eqn. 6.1.

$$d_w(r, \phi, R_p) = \sqrt{R_p^2 + r^2 - 2R_p r cos(\phi)} \tag{6.1}$$

When the interactions between the adsorbate and the pore wall are computed by a pairwise Lennard-Jones potential and smeared across the entire surface of the sphere rather than computed by discrete pairwise distances (as is the case in a direct simulation), the total adsorption energy takes the form of the integral in Eqn. 6.2,

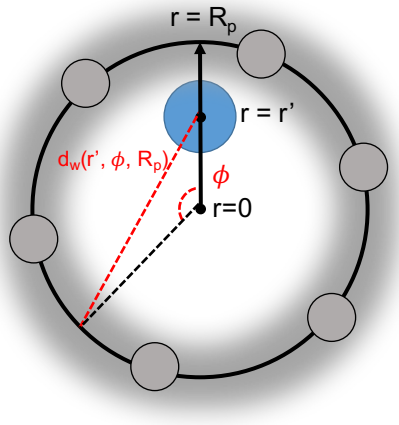


Figure 6.1: Visualization of a cross-section of the spherical pore model. The blue circle represents the adsorbate and the grey spheres represent the pore wall carbon atoms whose interaction energies are "smeared" across the surface of the sphere with uniform density η. Due to the spherical symmetry, the total adsorption energy of an adsorbate at a given location r' is simply an integral over the azimuth angle ϕ.

$$U_{tot}(r, R_p) = 2\pi\eta R_p^2 \int_0^\pi d\phi \ \sin(\phi) 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{d_w(r, \phi, R_p)} \right)^{12} - \left(\frac{\sigma_{ij}}{d_w(r, \phi, R_p)} \right)^{6} \right] \tag{6.2}$$

where ϵ_{ij} is the depth of the potential energy well and σ_{ij} is the distance between the adsorbate and the wall at which the potential energy is zero. Using this expression for the adsorbate interaction energy, we can now calculate the Henry coefficient, which is often computed in simulations via the expectation value in Eqn. 6.3,[183]

$$K_H = \beta \left\langle \exp(-\beta U^+) \right\rangle \tag{6.3}$$

where β is $(k_B T)^{-1}$ and U^+ is the total interaction energy of a randomly inserted ghost adsorbate. This expectation value can be rewritten as an integral over the pore volume by substitution of Eqn. 6.2 for U^+. The resulting Eqn. 6.4 yields an integrable formula for the rigid pore Henry coefficient, $K_{H,r}$, for a given radius R_p with spherical volume V_p.

$$K_{H,r}(R_p) = \frac{4\pi\beta}{V_p} \int_0^{R_p} dr \ r^2 \exp(-\beta U_{tot}(r, R_p)) \tag{6.4}$$

Next, flexibility is introduced into the model by allowing R_p to change according to a Gaussian distribution. Now the flexible Henry coefficient, $K_{H,f}$, is a function of the *mean* pore radius, $\langle R_p \rangle$, and the width of the Gaussian distribution of the radius, σ_p, as shown in Eqn. 6.5.

$$K_{H,f}(\langle R_p \rangle, \sigma_p) = \int_{R_{min}}^{R_{max}} dR \frac{1}{\sqrt{2\pi\sigma_p^2}} \exp\left(-\frac{(R - \langle R_p \rangle)^2}{2\sigma_p^2}\right) \cdot K_{H,r}(R) \tag{6.5}$$

The pore radius is bounded between R_{min} and R_{max}. Note that in the limiting case as σ_p approaches 0, the Gaussian distribution becomes a delta function, $\delta(\langle R_p \rangle)$, and we recover the rigid pore approximation such that $K_{H,f}(\langle R_p \rangle, 0) = K_{H,r}(\langle R_p \rangle)$. Hence the $K_{H,f}$ can be calculated over a range of average pore radii and distribution widths to demonstrate how the Henry coefficient depends on both the average size and strength of fluctuations inside a flexible pore. $K_{H,f}$ is calculated for both Xe and Kr ($K_{H,f}^{Xe}$ and $K_{H,f}^{Kr}$, respectively) and the infinite dilution selectivity of the flexible model, S_f follows as the ratio of these two quantities in Eqn. 6.6. The selectivity of the rigid model is just the ratio of the rigid Henry coefficients in Eqn. 6.7.

$$S_f(\langle R_p \rangle, \sigma_p) = \frac{K_{H,f}^{Xe}(\langle R_p \rangle, \sigma_p)}{K_{H,f}^{Kr}(\langle R_p \rangle, \sigma_p)} \tag{6.6}$$

$$S_r(R_p) = \frac{K_{H,r}^{Xe}(R_p)}{K_{H,r}^{Kr}(R_p)} \tag{6.7}$$

6.3 Simulation techniques to perform high-throughput screening of flexible structures

In addition to the analytical model, direct simulations are employed using various computational techniques to demonstrate the effects of intrinsic flexibility on the Henry regime adsorption properties in MOFs. Ideally one would like to exclusively use *ab initio* calculations to describe such flexibility, but these are prohibitively expensive for large systems and long simulation times that are required to obtain sufficiently accurate results for these materials. Therefore we rely on force field based molecular simulations and corroborate their performance with *ab initio* methods on a computationally feasible system. We elaborate the details of these simulation techniques starting with the description of the force fields used for classical molecular dynamics (MD) simulations of flexible materials, followed by a description of *ab initio* MD based on density functional theory (DFT). Finally we describe how the Henry coefficients are evaluated in these flexible materials, as well as the calculation of geometric properties which are necessary to evaluate how the pore sizes change in MOFs between the rigid approximation and flexible simulations.

Force field computed framework dynamics

There exists a large diversity of metal-ligand chemistry in the CoRE MOF database [244] (here we only study the materials for which density derived electrostatic and chemical, or DDEC, charges [247] have been obtained), the entirety of which cannot be described by

any single classical force field (Dreiding,[98] UFF,[73] BTW,[248] UFF4MOF,[249] etc.). In order to treat all materials on equal footing, we utilize the Universal Force Field (UFF) for all framework bonded potentials except for the coordination bonds between metals and organic species. For bond and angle potentials that are centered on the metal ions, the equilibrium harmonic bond length and equilibrium harmonic angle are modified to the crystallographic values read from the structure file.[165] This approximation is quite useful in that it allows us to simulate framework dynamics in the canonical ensemble, i.e. constant number of particles, volume, and temperature (NVT), for a vast majority of the CoRE MOF structures without significant aphysical distortions in the framework resulting from poor fits to UFF geometries. In a previous work we have demonstrated that this approximation is even sufficient to capture bulk moduli trends as calculated by DFT in various MOFs.[165] This force field is henceforth referred to as UFF-fix-metal (UFF-FM) and details on its implementation are presented in Appendix C.

To benchmark our classically generated dynamics using the UFF-FM approximation, we additionally performed an in-depth investigation of framework dynamics for one particular MOF, known as SBMOF-1 [250] with reference code KAXQIL in the Cambridge Structural Database (CSD) that displays excellent Xe selectivity.[243] We implement an additional classical force field which models metal-ligand coordination solely through Lennard-Jones interactions and electrostatics whereby dummy cation beads serve to delocalize the charge of the metal ion and preserve the octahedral geometry of the Ca^{2+} ions in the framework. While the model was originally developed to simulate solvation of cations,[251] recent work has illustrated this model's applicability in simulating MOF dynamics and deformation.[236] All potentials other than the metal-ligand interactions were modeled with standard UFF potentials, and the model is henceforth referred to as the UFF-cationic-dummy-model (UFF-CDM) with additional implementation details provided in Appendix C.

For generating an ensemble of structures from a molecular dynamics (MD) run over which the Henry coefficient can be computed, each CoRE MOF from Ref. [244] was simulated with UFF-FM in the NVT ensemble using the open source LAMMPS software package (`http://lammps.sandia.gov`).[252] The Nose-Hoover thermostat was used with a temperature of 298 K, and the structure was equilibrated for 30 ps, followed by a production run of 30 ps. During the production run, a framework configuration was saved every ps to give a total of 30 snapshots upon which the Henry coefficients could be calculated via the ensemble average in Eqn. 6.3. Justification for the number of snapshots necessary is elaborated in the *Porosity characterization* section. The same MD methods were used to simulate SBMOF-1 with UFF-CDM. Charges for UFF-FM simulations were taken from Ref. [244], while charges for UFF-CDM were derived from electronic structure calculations (see the subsequent section) and calculated according to the procedure discussed in Appendix C. Each MOF was simulated with periodic boundary conditions and a cutoff radius of 12.5 Å was imposed for non-bonded interactions. Supercells were generated such that the perpendicular components of the cell vectors were at least two times the cutoff radius.

Ab initio computed framework dynamics

All DFT calculations have been performed using the CP2K code, which uses a mixed Gaussian/plane-wave basis set.[68, 69] We employed double-ζ polarization quality Gaussian basis sets [253] and a 400 Ry plane-wave cutoff for the auxiliary grid, in conjunction with the Goedecker-Teter-Hutter pseudopotentials.[70, 71] All DFT calculations, including single point energies, geometry/cell optimizations and *ab initio* molecular dynamics simulations (AIMD), were performed using the PBE functional,[63] with Grimme's D3 van der Waals correction (PBE+D3).[64] This method was shown to give very good agreement with experimental structural data on several MOFs which we studied previously [67, 93] and on rare gas dimers and trimers.[64, 254] The counterpoise method [255] was used to correct for basis set superposition errors in all binding energy calculations. *Ab initio* molecular dynamics simulations within the Born-Oppenheimer approximation were performed in the canonical (NVT) ensemble at the PBE+D3 level of theory. A time step of 0.5 ps was used for the integration of the equation of motion. Different supercell sizes were considered for the AIMD simulations, and each AIMD simulation was performed for a duration of 10 ps (20,000 MD steps following 2,000 MD steps of equilibration run with a strong thermostat coupling) and at a temperature of 298 K, which was controlled by the canonical sampling through velocity rescaling thermostat [256] using a time constant of 50 fs. The initial structure was taken from the experimentally resolved crystal structure of SBMOF-1 and geometry optimized.[250]

For implementation of UFF-CDM, partial atomic charges are also needed (see Appendix C). The partial atomic charge analysis was performed using the REPEAT method proposed by Campañá et al.,[58] which was recently implemented into the CP2K code based on a restrained electrostatic potential framework.[72] We have used charges determined from this scheme recently in our grand canonical Monte Carlo simulations of CO_2 adsorption in MOF-74,[21, 77] and we obtained very good agreement with experiment on adsorption isotherms.

Flexible Henry coefficient calculation

The Henry coefficient, K_H, was previously defined in Eqn. 6.3 as the Boltzmann weighted average of the interaction energy of a randomly inserted ghost particle,[183] and it measures a material's affinity for an adsorbate in the limit of infinite dilution. This quantity can be simulated in porous materials by the Widom insertion method,[257] and in the rigid pore approximation the average in Eqn. 6.3 is calculated by attempting ghost particle insertions on the single framework configuration specified by experimental single crystal X-ray diffraction or by DFT optimization. For flexible materials, however, we compute this ensemble average over many different framework snapshots generated from a molecular dynamics simulation rather than just a single configuration. For both the rigid and the flexible framework simulations, the same force field is used to describe the non-bonded framework-adsorbate interactions. These non-bonded interactions are modeled with a pairwise Lennard-Jones potential where the framework atom ϵ and σ parameters are taken from the Universal Force

Field (UFF) [73] and the noble gas ϵ and σ parameters are taken from the force field of Boato.[258] Individual pairwise interaction parameters are obtained by Lorentz-Berthelot mixing rules. The Widom insertions for both the flexible and rigid Henry coefficients described above were performed in the RASPA software package [102] at a temperature of 298 K.

Porosity characterization

A geometric description of porous materials known as the pore size distribution (PSD) was calculated in the Zeo++ software package [74] using high accuracy settings for several selected flexible CoRE MOFs. At each snapshot from the NVT simulation, one PSD calculation was performed which produces a histogram of pore sizes. The overall PSD in a flexible material is the cumulative histogram of individual histograms from each NVT snapshot. A probe radius of 1.2 Å (smaller than the radius of either Xe or Kr) was used to ensure that both open pores and narrow constrictions are captured and can be visualized. In order to correctly perform the ensemble average in Eqn. 6.3, it is evident that the PSD must be converged, i.e. the cumulative, normalized distribution does not change with the addition of more NVT snapshots. With the SBMOF-1 system, performing a UFF-FM simulation of 10 ps with snapshots generated every 0.5 ps for a total of 20 snapshots produces a nearly identical PSD to the longer simulation procedure described previously, hence either is sufficient for performing the ensemble average.

The final PSD can then be mapped in a *semi-quantitative* way to the Gaussian variables (σ_p and $\langle R_p \rangle$) used in the analytical model described previously (Eqn. 6.5 and Eqn. 6.6). The difference is that atoms constituting the pore wall in a real MOF are not always carbon (as is imposed in the analytical model to allow it to be solvable) nor are pores in real systems perfectly spherical. Therefore some variation between the model and direct simulations is always expected when atoms constituting the pore walls in a given structure have different UFF parameters and do not form a perfectly spherical shell. Thus the limitation of the analytical model is that it cannot be used to directly map a computed PSD to exact values of the Henry coefficients and selectivity for a particular MOF. In other words, it does not replace the need to actually compute the Henry coefficient with Widom insertions on the accumulated snapshots from the NVT simulation. However, it does an excellent job of reproducing the trends of selectivity's dependence on flexibility as is seen in the following discussion sections.

6.4 Understanding the interplay between adsorbent flexibility and selectivity in chemical separations

Intrinsic flexibility effects on Henry regime adsorption behavior are discussed in five main points:

1. First, the analytical model is presented which quantitatively assesses the effects on Henry regime adsorption of tight-fitting molecules as a function of pore flexibility.

2. Next, the analytical model's results show that accounting for the systematic effects of intrinsic flexibility are critical for the design and/or identification of the best performing materials in this shape selective adsorption application.

3. Next, the CoRE MOF database is screened using flexible models to not only validate the conclusions drawn from the analytical model but also to demonstrate the discrepancies of the rigid pore approximation.

4. Next, four CoRE MOFs are selected to specifically detail the ways in which flexibility affects a material's potential for Xe/Kr separation.

5. Finally, the ideas developed thus far are applied to the SBMOF-1 system, and we demonstrate the necessity of using flexibility to obtain better agreement between experimental results and computational predictions of Xe/Kr adsorption properties.

Flexibility and the analytical model

We have developed both a rigid (Eqn. 6.4) and flexible (Eqn. 6.5) pore model to obtain insight into flexibility's effects on Xe/Kr Henry coefficients. The dependence of K_H on the pore radius and distribution width are shown in Fig. 6.2. The most interesting consequence of flexibility is that the pore sizes resulting in the largest $K_{H,r}$ are the most over-predictive of $K_{H,f}$ and the pore sizes with smaller $K_{H,r}$ are the most under-predictive of $K_{H,f}$. The size of these over-(under-)predictions is strongly dependent on the strength of pore size fluctuations (σ_p), which is shown in Fig. 6.2a (Xe) and Fig. 6.2b (Kr) when differing σ_p values are plotted. The over-predictions occur since there exists a finite probability that the pore deviates from the optimal size, even when the most probable configuration is the optimal size. The under-predictions occur since there exists a finite probability that the pore can adopt the optimal size, even though the most probable configuration is *not* the optimal size. Thus the main observation from this analysis is that, even if relatively small fluctuations ($\sigma_p < 0.4$ Å) in a flexible pore average out such that $\langle R_p \rangle$ is equivalent to R_p in the rigid approximation, the flexible pore Henry coefficient can differ significantly. We will see in subsequent discussion that, similar to the green curve in Fig. 6.2, $\sigma_p \approx 0.3$ Å for SBMOF-1.

More interesting than just K_H from an applications perspective is the infinite dilution selectivity of Xe/Kr, which was formulated for flexible and rigid pores in Eqn. 6.6 and Eqn. 6.7, respectively. The selectivity for the flexible pore model was calculated over the bounds 2.5 Å $< \langle R_p \rangle <$ 10.0 Å and 0.0 Å $< \sigma_p <$ 1.0 Å and mapped over this two dimensional space as shown in Fig. 6.3. It is immediately evident from Fig. 6.3a that, since the contours are not vertical lines, the rigid pore approximation will yield a different selectivity than a flexible pore (i.e. when $\sigma_p \neq 0$). Fig. 6.3b additionally reveals extremely sharp gradients in $S_f(\langle R_p \rangle, \sigma_p)$ when $\langle R_p \rangle$ is slightly smaller than the optimal size (~ 4.7 Å) for Xe selectivity.

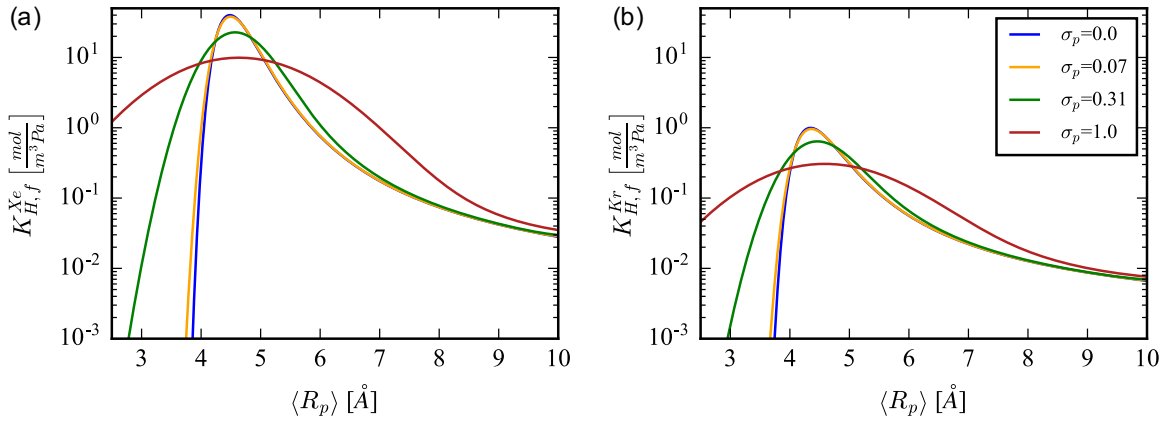


Figure 6.2: Broadening of the flexible Henry coefficient profile, $K_{H,f}(\langle R_p \rangle, \sigma_p)$, is calculated for increasing values of the pore size distribution width, $\sigma_p = \{0.0, 0.07, 0.31, 1.0\}$, for both (a) a Xe adsorbate and (b) a Kr adsorbate.

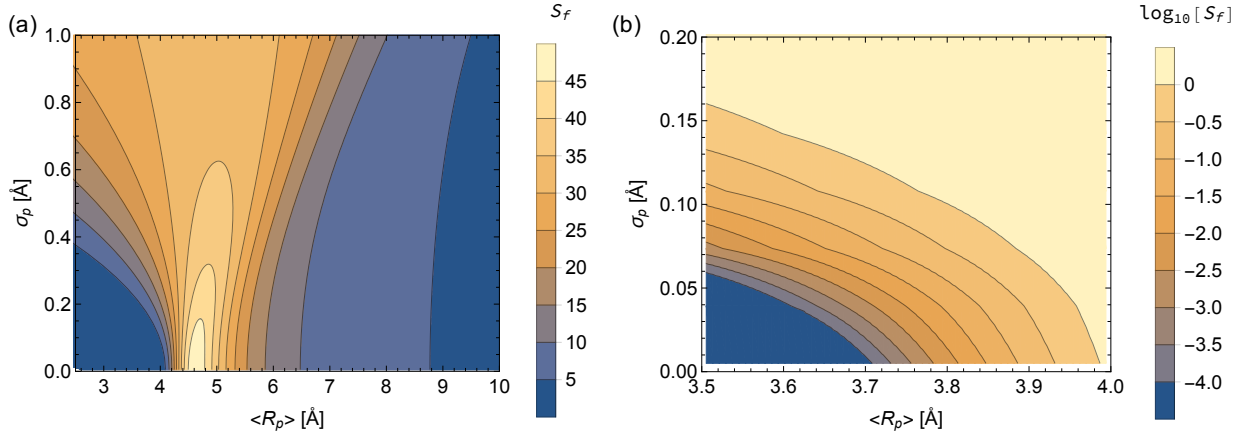


Figure 6.3: Selectivity of Xe to Kr mapped onto the average pore radius (x dimension) and the distribution width of the radius (y dimension) for the fluctuating pore model. The color scale in (a) is a linear scale of the flexible selectivity, S_f, whereas the color in (b) is the $\log_{10}$ scale of S_f to better emphasize the gradients that exist over small changes in $\langle R_p \rangle$ and σ_p.

There are three distinct situations where a flexible pore can result in significantly different predicted selectivity than a rigid pore. First, the optimal pore for Xe selectivity in this analytical model has $\langle R_p \rangle \approx 4.7$ Å and $\sigma_p = 0$ Å. Yet fluctuations on the order of $\sigma_p \approx 0.6$ Å can reduce the selectivity of such a pore by a factor of 2. Second, a rigid pore predicted to be Kr selective can very easily be non-selective or even Xe selective if the pore undergoes small fluctuations ($\sigma_p \approx 0.15$ Å). Third, predicted selectivity can be quite different when

flexible and rigid $\langle R_p \rangle$ are not equivalent. If thermal fluctuations in a flexible pore do not average out to the same value as the rigid pore, S_r can vastly over-predict S_f. For example, if a rigid pore has $R_p = 4.7$ Å and overestimates the average flexible pore $\langle R_p \rangle$ by 0.6 Å, the rigid selectivity could be reduced by a factor of 50. The sensitivity of adsorption properties to small changes in pore sizes has been shown in the literature when different DFT methods were used to optimize the crystal structure.[259–261] However, we are highlighting an entirely different situation where the average pore size changes due to flexibility, an effect which we will demonstrate for real MOF systems later on. Most interestingly, the global optimum for selectivity as a function of flexibility occurs for an entirely rigid pore. This is discussed further in the following section.

Flexibility and optimal separations

We have systematically demonstrated how we expect shape selective adsorption separations to have an intricate, non-trivial dependence on intrinsic flexibility, having chosen the specific application of Xe/Kr separations to illustrate our findings. Fig. 6.4 summarizes how this information can be exploited for the design and prediction of better separation materials. To construct this figure we have extracted S_f from Fig. 6.3 at constant σ_p and normalized the x-axis of $\langle R_p \rangle$ by $\langle R_p \rangle_{opt}$ such that the optimal selectivity for a given σ_p occurs at a value of 1. The global optimum of S_f in ($\langle R_p \rangle$,σ_p) space occurs when $\sigma_p = 0$.

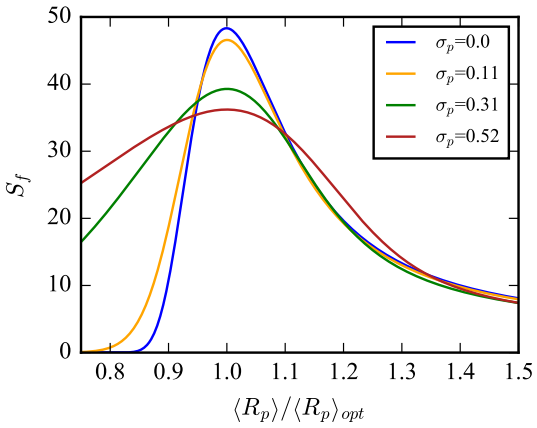


Figure 6.4: The flexible selectivity from the analytical model is plotted as a function of $\langle R_p \rangle$ divided by the optimal pore size, $\langle R_p \rangle_{opt}$, such that the point of optimal selectivity is normalized to 1 for all values of σ_p. This demonstrates that maximizing selectivity is achieved not simply by having the optimal pore size but also by simultaneously minimizing intrinsic flexibility. Although, interestingly, pore sizes that are smaller than the optimal size can have their selectivity significantly improved with increasing flexibility.

This clearly demonstrates that designing MOFs with the optimal combination of pore

size and chemistry is not sufficient to achieve a *global* optimum in performance; one must minimize intrinsic flexibility while simultaneously realizing the optimal average pore size to achieve the best separation. It also demonstrates that, while fluctuations reduce the selectivity of the optimally rigid pore, sufficiently large fluctuations expand the range of average pore sizes (or the number of MOFs) for which flexibility can improve selectivity. Nonetheless, the global optimum in $\langle R_p \rangle$ and σ_p space occurs for $\sigma_p = 0$. In other words, if one could design the perfectly optimal rigid pore with the ideal size and chemical composition, increasing flexibility would only lead to worse performance. These results provide a guiding principle in the rational design of nanoporous materials for shape selective separations and are validated when studying the effects of flexibility on Xe/Kr adsorption in real systems, i.e. the CoRE MOF database, which is the focus of subsequent discussion.

Flexibility in CoRE MOF screening

From the analytical model it is evident that flexibility is important when a material's pores approach the same size as the adsorbed molecules. For a correct screening it is therefore important to take framework flexibility into account, and to do this we calculate S_f with the reported CoRE MOF experimental structure as the starting framework configuration and input for UFF-FM. Three plots are instructive for comparison with the results from the analytical model.

Firstly, the Henry coefficient of Xe in the flexible material ($K_{H,f}^{Xe}$) is plotted versus the corresponding rigid Henry coefficient ($K_{H,r}^{Xe}$) in Fig. 6.5, and each material is color coded by their largest included sphere (D_i).[74] For materials with very large D_i the rigid and flexible calculations tend to give equitable results for the Xe Henry coefficient. However, for smaller pores and those that are the optimal size for Xe selectivity, orders of magnitude discrepancy can exist between the calculated Henry coefficients for the flexible structure and the rigid approximation. This discrepancy is attributed to pore size fluctuations, slight changes in $\langle R_p \rangle$, or a combination of both, all of which is clearly demonstrated by the results of the analytical model shown in Fig. 6.2 and Fig. 6.3.

Secondly, Fig. 6.6 plots the flexible selectivity versus the rigid selectivity, and data points are color coded by the material's $K_{H,f}^{Xe}$ such that the most interesting materials in terms of Xe affinity are colored yellow. Purple colored materials by comparison are expected to have very low infinite dilution uptake. The rigid approximation performs decently for some structures which remain on the $y = x$ line of parity. However, there is very noticeable scatter and considering flexibility can result in orders of magnitude deviation. Among top performing structures from the entire screening, S_r almost always overestimates S_f, supporting our conclusion that rigid pores are the globally optimal solution since, if a material already achieves the optimal pore size for a given chemistry, fluctuations will only reduce the selectivity. Medium performing structures tend to have both significant over-(under-)estimations, while the lowest performing rigid structures tend to exhibit a huge underestimation of the selectivity. This systematic dependency is visualized another way in Fig. 6.6b where the histogram shows that, among top performing structures, simulating flexibility reduces performance in

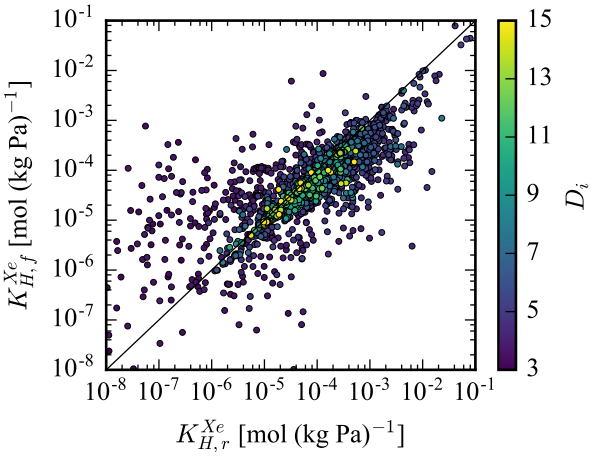


Figure 6.5: The flexible Xe Henry coefficient, $K^{Xe}_{H,f}$, is plotted versus the rigid Xe Henry coefficient, $K^{Xe}_{H,r}$. The color-coding of materials by their largest included sphere shows that the rigid approximation tends to perform worse as a material's pore size decreases.

∼95% of materials. By contrast, for the lowest performing MOFs, flexibility increases performance in ∼90% of the materials. The red highlighted regions in Fig. 6.6a show where the rigid and flexible simulations predict reverse selectivity. Thus a pore that was a good fit for Xe can become too small (lower right region), or one that was too small can become enlarged and favorable by intrinsic flexibility (upper left region) and result in a material with large selectivity for Xe. This systematic behavior has already been predicted by the solution of the analytical model shown in Fig. 6.3.

Finally, we plot the relative error of the rigid approximation (S_f/S_r) in $\log_{10}$ scale versus D_i in Fig. 6.7. We observe exactly what one might expect given the results of the analytical model. The divergence of S_f from S_r increases as the pore size decreases and is very significant in the range of D_i for optimal Xe selectivity. Additionally, the analytical model is mapped onto the screening results in Fig. 6.7 by converting $\langle R_p \rangle$ to D_i via Eqn. 6.8,

$$D_i = 2 \langle R_p \rangle - D_{carbon} \tag{6.8}$$

where $D_{carbon} = 3.4$ Å is the Van der Waals diameter of a carbon atom. The black, red, and purple lines represent S_f/S_r for σ_p values of 0.005, 0.04, and 0.35, respectively. The trend in the screening data is captured well by the analytical model despite its simplicity in assuming all pore wall atoms are carbon and are arranged in a perfectly spherical shell. Superimposing the analytical model in this way also does not account for when the flexible and rigid $\langle R_p \rangle$ are not identical for a particular material. This effect, which is demonstrated in subsequent discussion, would additionally change how the analytical model is mapped on Fig. 6.7. Most importantly, these analyses demonstrate that the screening results agree with the analytical model: the best rigid approximation materials (which ignore framework fluctuations) overestimate the selectivity when flexibility is considered.

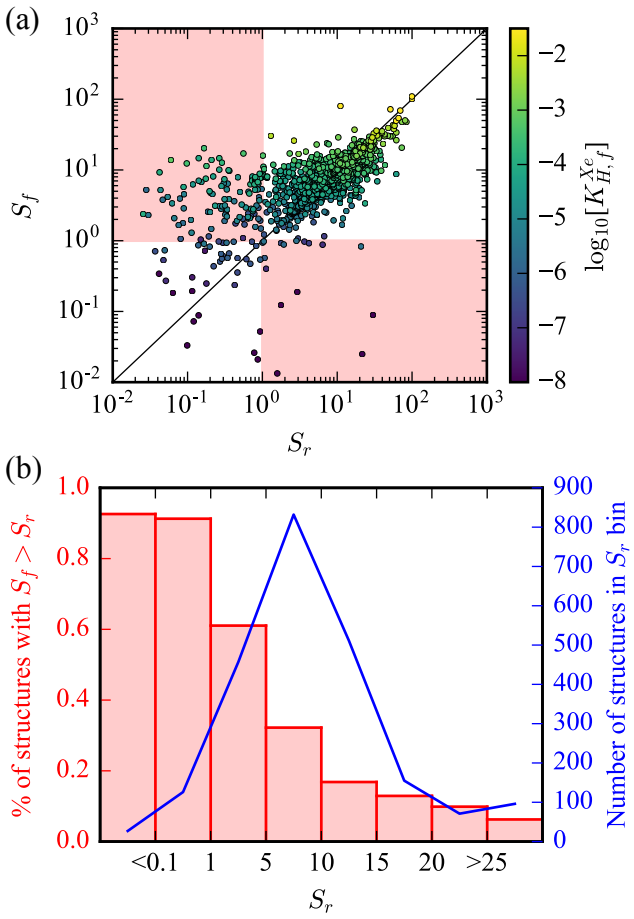


Figure 6.6: (a) Flexible selectivity is plotted versus rigid approximation selectivity for screened CoRE MOFs. Highlighted regions in red demonstrate areas where the rigid approximation and the flexible simulations predict reverse selectivity of one another. Materials are color coded by the $\log_{10}$ value of the flexible Xe Henry coefficient, $K^{Xe}_{H,f}$ [mol (kg Pa)$^{-1}$], to highlight the best flexible materials for Xe infinite dilution uptake. (b) A histogram view of the same data is shown. The red histogram shows that the larger the rigid selectivity becomes (i.e. the more optimal the pore size and chemistry), the more frequently that flexibility actually reduces the performance.

We caution that our screening calculations were *not* performed using a pocket-blocking algorithm since flexibility in points of constriction in the pore network makes the accessibility of pockets a function of time. Instead the screening results are only visualized for structures where $K^{Xe}_{H,r} >$ 1e-8 mol/(kg·Pa) as an approximate way to filter out non-porous structures. We closely investigate and explain the performance of selected structures, for which we manually ensure that pockets of inaccessibility do not exist, via comparison to the analytical model in the following section.

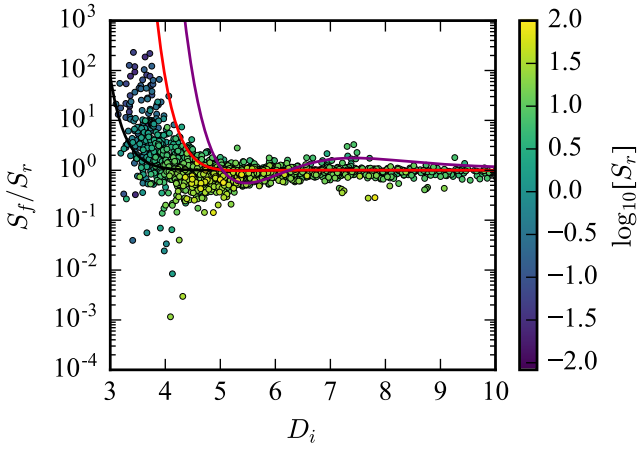


Figure 6.7: The relative error from the rigid structure approximation, S_f/S_r, is plotted in $\log_{10}$ scale against D_i. Each material is color-coded by $\log_{10}[S_r]$. The screening results are compared to the analytical model where the solid black, red, and purple lines represent $S_f(D_i, \sigma_p = 0.005)/S_r(D_i)$, $S_f(D_i, \sigma_p = 0.04)/S_r(D_i)$, and $S_f(D_i, \sigma_p = 0.35)/S_r(D_i)$, respectively.

Flexibility in selected CoRE MOFs

Several MOFs are selected from the screening study to illustrate in greater detail the various ways in which flexibility changes the selectivity by comparison to the rigid structure approximation. For these structures, NVT simulations show that the rigid pore approximation does not present a good statistical representation of the PSD when the host framework fluctuates at room temperature. Fig. 6.8 presents a visual analysis of flexibility in four selected CoRE MOFs and demonstrates how thermal fluctuations affect the PSD and thus the Xe/Kr selectivity. For each material we show the flexible/rigid PSD, as well as visualization of the rigid structure (left image) and a snapshot of the flexible structure (right image) from the UFF-FM MD run. The corresponding flexible and rigid Henry coefficients and selectivity are summarized in Table 6.1.

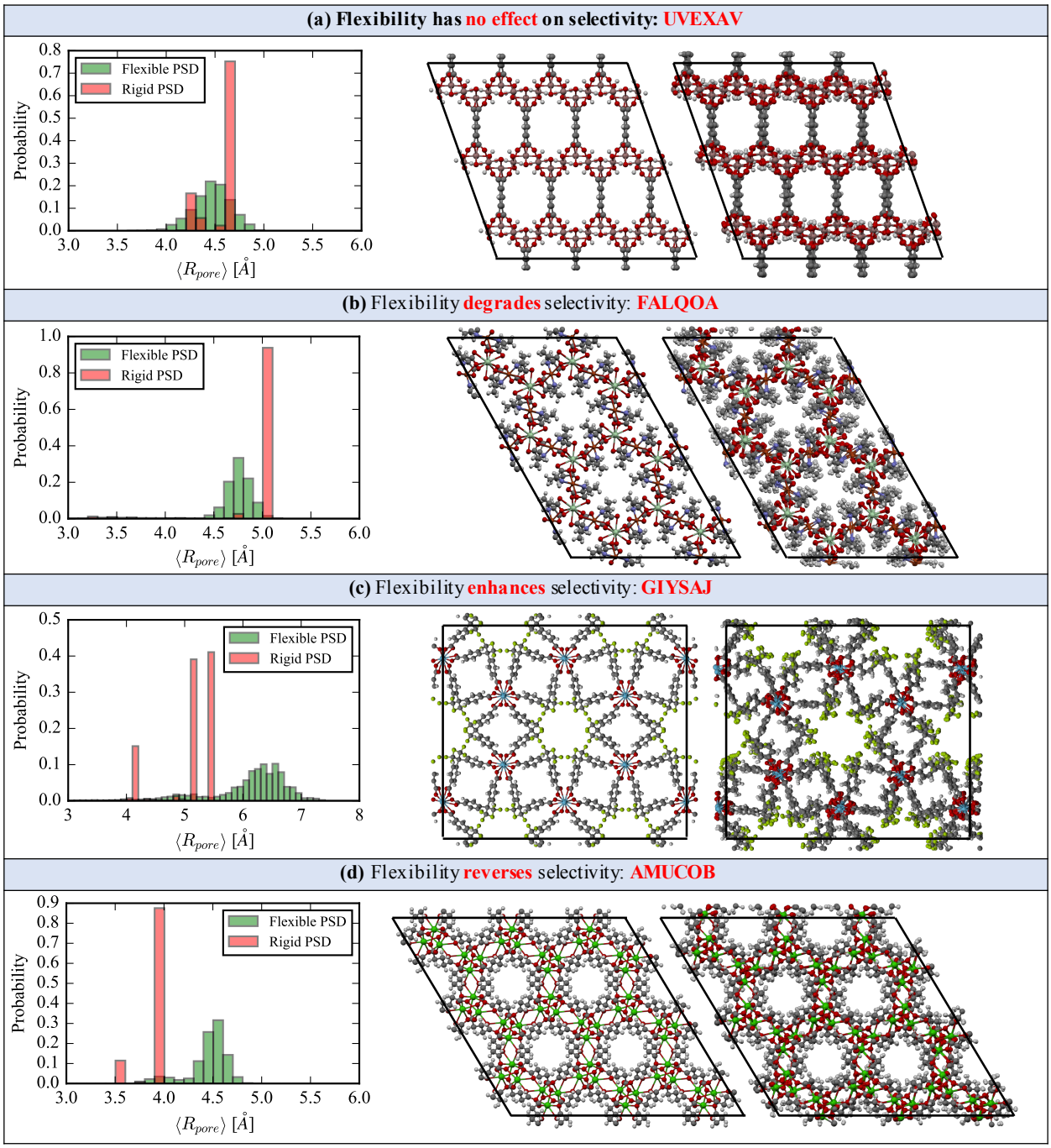


Figure 6.8: Evolution of the PSD after NVT dynamics is shown for four CoRE MOFs, with the experimental structure (left image) and a snapshot from the flexible simulation (right image). Sections (a-d) show specific cases for which flexibility has no effect, degrades, enhances, or reverses the Xe/Kr selectivity, respectively.

Table 6.1: Flexible and rigid Henry coefficients and selectivity for the four CoRE MOFs shown in Fig. 6.8.

CSD Code	$K_{H,f}$ [mol/(kg·Pa)]		$K_{H,r}$ [mol/(kg·Pa)]		S_f	S_r
	Xe	Kr	Xe	Kr		
UVEXAV	1.0e-2	3.5e-4	1.1e-2	3.8e-4	29	29
FALQOA	2.1e-3	4.5e-5	1.3e-2	1.6e-4	47	81
GIYSAJ	8.9e-4	3.0e-5	1.4e-3	5.3e-5	30	26
AMUCOB	4.3e-4	2.5e-5	2.9e-6	5.2e-6	17	0.56

In Fig. 6.8a, UVEXAV (MIL-120) [262] demonstrates a situation where the selectivity remains unaffected by flexibility. The rigid PSD reveals two pores that are very close in size, but notably the flexible PSD shows a continuous distribution with approximately the same mean, resulting in unchanged K_H and $S_f \approx S_r$. Fig. 6.8b highlights FALQOA,[263] a material for which flexibility reduces S_r. The rigid PSD demonstrates a pore that is of very favorable size for Xe adsorption which shrinks only slightly throughout the MD run. However, the sensitivity of the Henry coefficient to $\langle R_p \rangle$ is sufficient to lower $K_{H,f}^{Xe}$ by an order of magnitude and halve the selectivity when flexibility is considered. The most complex PSD in this set of examples is exhibited by GIYSAJ (RPF-4) [264] as shown in Fig. 6.8c for which flexibility serves to slightly enhance the selectivity. Flexible simulations show that the smallest pore can essentially be blocked due to ring rotation and the large pore shifts to a higher mean value. Both effects serve to decrease the Henry coefficients in the flexible structure but in such a way as to slightly raise the selectivity overall. AMUCOB (BioMIL-2) [265] in Fig. 6.8d presents arguably the most interesting behavior, as it is Kr selective under the rigid pore approximation but very Xe selective when flexibility is considered due to an overall increase in pore size. AMUCOB is just one of several CoRE MOFs to display this phenomena as previously discussed with Fig. 6.6. Additionally, we note that each structure's flexible selectivity is higher than the experimental selectivity of SBMOF-1, a material which is the focus of discussion in the following section.

Flexibility in SBMOF-1

The screening data and the analytical model demonstrate that when a given material already has the optimal pore size to maximize selectivity, increasing intrinsic pore flexibility only serves to reduce the selectivity performance. This effect can be shown in the context of a particular well known Xe/Kr separations system. Here our flexibility calculations are benchmarked against experimental results for the highly-performing Xe/Kr separation material, SBMOF-1, to once again show that pore flexibility reduces the performance in selectivity of optimal rigid materials. This system was chosen since it was identified as the best performer from a high-throughput screening study [242] but showed a large discrepancy between experimental and computational Henry coefficients when using the rigid pore approximation.[243]

Table 6.2: Henry coefficients and selectivity in SBMOF-1 is calculated for the rigid pore approximation and various framework dynamics methods. The "Description" column gives the simulation type (and number of unit cell replications used in the AIMD simulation), and the "Flexibility" column denotes a flexible simulation. Experimental data from Ref. [243] is included.

Description	Flexibility	K_H [mol/(kg·Pa)]		$S_{Xe/Kr}$
		Xe	Kr	
Experimental data	N/A	3.84e-4	2.37e-5	16
KAXQIL deposited structure	no	1.45e-2	2.70e-4	54
KAXQIL DFT optimized	no	1.03e-2	2.20e-4	47
AIMD (1x2x1)	yes	7.49e-3	1.85e-4	41
AIMD (2x2x1)	yes	6.80e-3	1.77e-4	38
AIMD (1x3x1)	yes	6.68e-3	1.72e-4	39
UFF-FM	yes	6.24e-3	1.67e-4	37
UFF-CDM	yes	3.18e-3	1.28e-4	25

Thus we focus on this structure to show how consideration of framework flexibility better aligns computations with experiments and reduces the optimal selectivity. DFT/PBE+D3, UFF-FM, and UFF-CDM were all used to model framework dynamics in SBMOF-1, and the resulting Henry coefficient and selectivity calculations are summarized in Table 6.2. The rigid structure approximation of SBMOF-1 (KAXQIL) has the optimal pore size for maximizing the Xe Henry coefficient and selectivity. However, considering flexibility reduces these quantities and leads to better agreement with experiments. UFF-CDM generated dynamics provide the best agreement with experiments and there exists a significant decrease in the Henry coefficients and selectivity when compared to the rigid pore approximation.

This decrease in Henry coefficients fundamentally arises from thermal disordering in SBMOF-1, visualized in Fig. 6.9. The impact of flexibility on the PSD for each dynamics method is quantitatively shown in Fig. 6.10. While the rigid pores are of optimal size for maximizing selectivity, the existence of flexibility in all cases leads to pore size fluctuations and a shrinking of average pore size away from its optimal value. As predicted by the analytical model, both effects can serve to reduce the Henry coefficients (of both Xe and Kr) and the overall selectivity of the material.

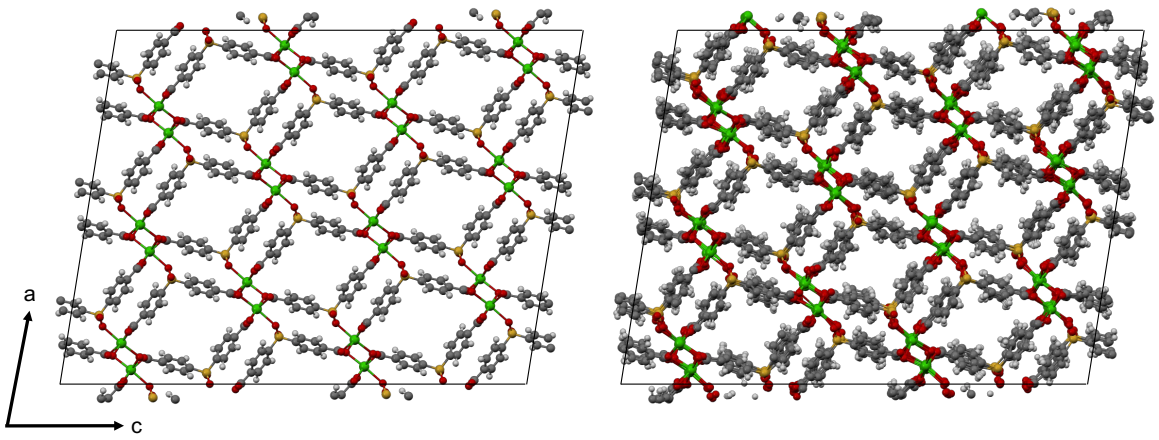


Figure 6.9: The DFT optimized cell (replicated 3x6x2) is depicted on the left while a snapshot from a MD trajectory generated with UFF-CDM is depicted on the right. Thermal disordering of the structure results in a change in the PSD accessible to Xe/Kr adsorbates.

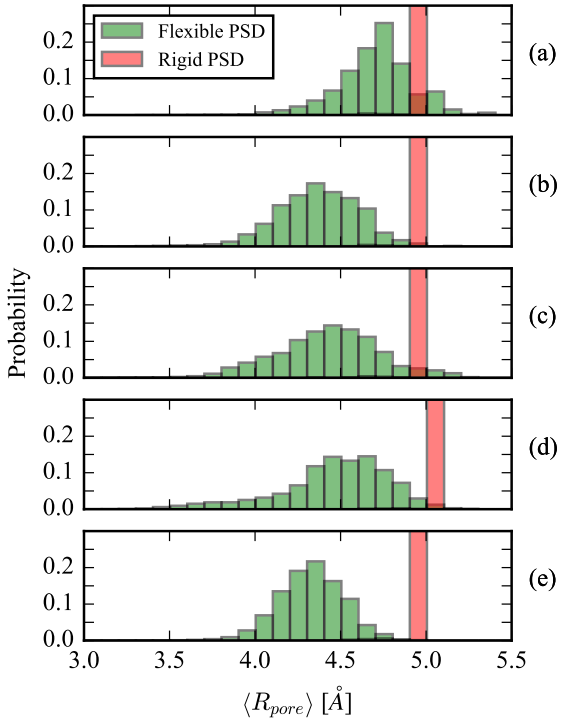


Figure 6.10: Evolution of the PSD is shown from the rigid structure (red histogram) to the flexible structure (green histogram). (a), (b), and (c) show the PSD for 1x2x1, 2x2x1, and 1x3x1 AIMD, respectively, where the rigid structure corresponds to the DFT minimized framework. (d) shows the PSD for UFF-FM with the rigid structure corresponding to the KAXQIL experimentally resolved framework, and (e) shows the PSD for UFF-CDM where the rigid structure is the DFT minimized framework.

For DFT/PBE+D3, only a small number of unit cell replications is computationally feasible. Interestingly, 1x2x1, 2x2x1, and 1x3x1 simulations all display different PSDs, suggesting that the flexibly determined PSD can be system size dependent since the system is simply too constrained when considering only one and two replications in certain crystallographic directions. Similar finite size effects have recently been reported where multiple unit cell replications were necessary to observe crystallographic deformations.[236] However, the key feature is that the mean pore size always decreases leading to reduced Henry coefficients of both adsorbates, which is accompanied by a reduction in the selectivity. The identical trend is observed in the larger classical simulations. Thus a major finding from this analysis is that, for all flexible simulation methods, the thermodynamic fluctuations *do not* average out to the rigid pore approximation. Different methods (classical vs. *ab initio*) for host framework dynamics yield similar evolution of the PSD's from the rigid to flexible simulations, and the computed $K_{H,f}$ is always in better agreement with experiments due to the decrease in $\langle R_p \rangle$.

It should be noted that the non-bonded parameters for guest-framework interactions

also have an effect on the predicted Henry coefficients and selectivity. We find that the Boato and UFF parameters for guest-framework interactions do not perfectly replicate the DFT/PBE+D3 potential energy surface which is shown in Appendix C. Regardless, the data presented here shows the large dependence of the adsorption properties on flexibly induced changes in the pore size.

6.5 Outlook: High-throughput screening studies of adsorption thermodynamics in flexible systems

The chemical tunability of MOFs is constantly exploited to create adsorption sites with a high selectivity. However, in the case of shape selective adsorption separations where the pore size and shape are commensurate with similarly sized adsorbates like Xe/Kr, the efficiency of these adsorption sites for separations relies on subtle geometric differences, which we have shown to be strongly influenced by flexibility. In this work we have systematically addressed the non-trivial dependence of selectivity on intrinsic flexibility, providing a comprehensive outlook for the design and/or identification of optimal shape selective separation materials in the context of Xe/Kr mixtures. Quantification of the effect of flexibility on the separation of Xe/Kr mixtures using an analytical model and molecular simulations shows that some materials' selectivity can be increased while others' decreased by flexibility. However, for those materials already displaying the optimal pore size for a given pore chemistry, increasing pore size fluctuations serve only to further decrease selectivity. Hence the design of a *globally* optimal separation must not only focus on optimizing pore size and chemistry but also on minimizing intrinsic flexibility. From an experimental design point of view, one must simultaneously aim for MOFs with ligands, coordination environments, and topologies that constitute the most rigid framework possible *and* the optimal pore size and chemistry to achieve maximum possible selectivity.

An equally important part of this work discusses the validity of the rigid framework approximation in shape selective adsorption applications. In most computational high-throughput screening studies, the MOF structure is assumed to be rigid for computational efficiency reasons and we investigated whether this assumption has biased the selection of optimal materials. Our results show that, among the top performing structures in the rigid framework approximation, simulations accounting for flexibility lead to a reduction of selectivity in ∼95% of these materials. Conversely, this screening also demonstrated that fluctuations alone can cause a material to reverse its selectivity when the average pore size is less than the optimal value for Xe adsorption. Among the lowest performing structures in the rigid framework approximation, ∼90% of these materials had their selectivity increased when accounting for framework flexibility. These results suggest that flexibility should be considered in shape selective screening studies for the highest degree of accuracy possible and to achieve the best ranking of high performance materials.

SBMOF-1, which was recently crowned as the material with the highest Xe/Kr selectivity

from a computational screening of rigid MOFs, was re-evaluated with flexibility using both *ab initio* and classical molecular dynamics calculations. This resulted in better agreement with experimental data. Naturally the question then arises if there are materials which can be experimentally demonstrated to have better Xe/Kr selectivity than SBMOF-1. We have presented four structures (CSD reference codes UVEXAV, FALQOA, GIYSAJ, and AMUCOB) identified in our screening study whose predicted S_f is higher than the experimental selectivity of SBMOF-1 and used each structure's PSD evolution in NVT simulations to observe how flexibility affects Xe and Kr Henry regime uptake. Through a variety of molecular dynamics methods, we demonstrated that flexibility always serves to reduce the Henry coefficients of Xe and Kr and selectivity in SBMOF-1, bringing better alignment between experimental results and computational predictions for this material's adsorption properties.

The concepts developed in this work can now be applied to other shape selective separations. Such an example would be CO_2/CH_4 separations where, like Xe/Kr mixtures, the kinetic diameters of both adsorbates are of similar size. Here one would expect intrinsic flexibility to affect the Henry regime selectivity as well. Similar to Xe/Kr separations, there would exist an optimal pore size and chemistry that maximizes the rigid selectivity for one of the adsorbates, but the more flexible this optimal pore is, the more the selectivity will be reduced. And while our analytical model only accounts for isotropic pore deformations, it would also be useful to develop another model in which an additional flexibility parameter characterizes the degree of anisotropy of the pore deformation since is CH_4 is an isotropic particle but CO_2 is not. This development of more intricate analytical models in conjunction with a similar screening of flexible CoRE MOFs could lead to new insights regarding the usefulness of MOFs for CO_2/CH_4 separations.

Finally, we re-emphasize that we have focused on predicting Henry regime selectivity. However, both the analytical model and the flexible screening methods could be extended to obtain selectivity at higher pressures with additional assumptions and more computational cost, respectively. In the cases of higher pressure, the adsorption behavior will then depend on the exact potential energy profile of the host as a function of the pore size, since less energetically favorable conformations of the framework can be stabilized with increasing chemical potential. Thus it is likely that new insights will be obtained for the selectivity dependence at higher pressures. Such future work will have two major benefits. First, we will understand if the trends observed in this work between flexibility and separation performance hold at high pressure (where many separations are performed). Second, we will be able to determine over what pressure ranges Henry's law is applicable in flexible materials and whether it differs from the Henry regime pressure range in the rigid pore approximation.

Appendix A

Additional Characterization of M_2(DHFUMA)

A.1 Additional computational details

For all periodic density functional theory (DFT) calculations using CP2K, we employed double-ζ polarization quality Gaussian basis sets [253] and a 600 Ry plane-wave cutoff for the auxiliary grid, in conjunction with the Goedecker-Teter-Hutter pseudopotentials.[70, 71] For hybrid DFT calculations, we have used cpFIT3 quality auxiliary basis sets for the Hartree-Fock exchange (HFX) calculations based on the auxiliary density matrix method (ADMM) approach,[94] and a cutoff radius of 4 Å for the HFX calculations using the truncated Coulomb potential.[266] Structural optimizations, including atomic coordinates and cell parameters, and total energy calculations of the periodic systems, were performed at density functional theory level, using both PBE[63] and PBE0[89, 90] functionals, together with Grimme's D3 van der Waals correction with the Axilrod-Teller-Muto three-body terms.[64] A convergence threshold of 5.0×10^{-6} Hartree was used for all self-consistent field (SCF) calculations. The structural optimizations were considered converged if the maximum force on all atoms falls below 0.534 kcal mol^{-1} Å^{-1} (4.5×10^{-4} Hartree Bohr^{-1}). Majority of the calculations were performed with the Γ-point approximation using a $1 \times 1 \times 2$ multiplication of the hexagonal primitive cell. Additional calculations were performed in a $1 \times 1 \times 3$ supercell to check the convergence of the binding geometries and energies with respect to the supercell sizes. The counterpoise method[255] was used to correct for basis set superposition errors (BSSEs) in all binding energy calculations.

Periodic density functional theory (DFT) calculations of the vibrational frequencies and elastic constants of Mg_2(DHFUMA) were performed using Crystal with the B3LYP functional, in a rhombohedral cell with a total of 6 symmetry-inequivalent atoms. Monkhorst-Pack k-point grids of ($8 \times 8 \times 8$) were used to sample the Brillouin zone. Gaussian-type basis sets of 8-511d1G, 8-411d11G and 6-311d11G qualities were used for Mg, O and C, respectively. The accuracy of the integral calculations (i.e. the TOLINTEG keyword) were

set to "7 7 7 7 16", and the "XLGRID" setting (i.e. extra large predefined grid) was used for the numerical integration of the exchange-correlation term. The vibrational frequencies calculations were performed at the Γ-point, and the default displacement of 0.001 Å on each atom was used to compute the numerical second derivatives of the total energy. For the calculation of the elastic constants, we have used the default settings on size of lattice deformation (0.01 Å) and number of points (i.e. 3, including the central point with zero displacement) for the numerical second derivatives calculations. We have used the default energy and force convergence criteria for vibrational frequency and elastic constants calculations as defined by Crystal. These convergence criteria are typically tighter than separate single point and geometry optimization calculations.

A.2 Calculated vibrational frequencies of Mg_2(DHFUMA)

Table A.1: Calculated vibrational frequencies (in cm^{-1}) of Mg_2(DHFUMA)

0.00	0.00	0.00	51.43	102.24	134.51
145.01	145.01	151.40	155.83	155.83	179.74
179.74	183.17	183.17	197.81	198.78	203.22
203.22	208.17	214.87	214.87	227.13	227.13
236.78	236.78	257.68	265.67	265.67	278.87
283.31	283.31	289.05	324.66	333.56	333.56
336.53	340.11	363.11	371.36	371.36	375.45
392.20	392.20	399.23	399.23	423.05	423.05
453.45	459.91	473.43	473.43	483.19	483.19
503.74	503.74	503.75	517.23	519.01	519.01
532.33	532.33	533.44	543.55	548.67	548.67
629.99	649.42	649.42	763.53	763.53	763.81
766.90	771.89	771.89	814.65	816.42	816.42
825.57	826.90	826.90	884.28	888.54	888.54
985.56	991.09	991.09	1186.45	1211.43	1211.43
1266.68	1279.78	1279.78	1358.82	1358.82	1417.73
1452.92	1454.46	1454.46	1550.64	1550.64	1553.73
1579.38	1579.38	1579.71	1608.71	1611.33	1611.33

A.3 Calculated elastic constants of Mg_2(DHFUMA)

Table A.2: Calculated elastic constants (in GPa) of Mg_2(DHFUMA)

59.614	47.303	26.739	-0.522	6.763	0
	59.614	26.739	0.522	-6.763	0
		61.869	0	0	0
			38.422	0	-6.763
				38.422	-0.522
					6.155

Only the upper triangular part of the symmetric elastic matrix is shown. Our calculated elastic constants satisfy all the *necessary and sufficient* Born stability conditions [105] as derived by Mouhat and Coudert [106] for a rhombohedral (II) system (Laue class $\bar{3}$):

$C_{11} > |C_{12}|$
$C_{44} > 0$
$C_{13}^2 < \frac{1}{2}C_{33}\,(C_{11} + C_{12})$
$C_{14}^2 + C_{15}^2 < \frac{1}{2}C_{44}\,(C_{11} - C_{12}) \equiv C_{44}C_{66}$

A.4 Structure files and charges

The DFT optimized structures for Mg_2(DHFUMA) and Mg_2(DOBDC) are included in the structures.zip file. The partial atomic charges derived from the REPEAT analysis in all M_2(DOBDC) and M_2(DHFUMA) frameworks is presented in Table A.3. Note that the labels of each atom type correspond to the atom types defined in the force field of Mercado et al. (see Fig. A.2).

Table A.3: REPEAT derived partial charges for A = M_2(DHFUMA) and B = M_2(DOBDC)

Atom	M = Mg		M = Fe		M = Co		M = Ni		M = Zn	
Type	A	B	A	B	A	B	A	B	A	B
M	1.619	1.66	1.372	1.343	1.306	1.306	1.386	1.384	1.224	1.228
Oa	-0.952	-0.921	-0.908	-0.824	-0.869	-0.797	-0.926	-0.857	-0.779	-0.72
Ob	-0.734	-0.763	-0.722	-0.725	-0.711	-0.714	-0.713	-0.734	-0.648	-0.659
Oc	-0.781	-0.969	-0.679	-0.814	-0.633	-0.793	-0.681	-0.844	-0.604	-0.75
Ca	0.838	0.881	1.002	0.946	0.979	0.925	0.997	0.98	0.829	0.808
Cb	0.01	-0.316	-0.065	-0.373	-0.072	-0.341	-0.063	-0.388	-0.022	-0.283
Cc	—	0.478	—	0.444	—	0.415	—	0.443	—	0.377
Cd	—	-0.257	—	-0.205	—	-0.205	—	-0.191	—	-0.197
H	—	0.207	—	0.208	—	0.204	—	0.207	—	0.196

A.5 Additional crystallographic data for Mg_2(DHFUMA) and Mg_2(DOBDC)

Additional crystallographic data for the DFT optimized Mg_2(DHFUMA) and Mg_2(DOBDC) structures are summarized in Table A.4 for the reader's convenience. All quantities were computed in Zeo++ with high accuracy settings.

Table A.4: Theoretical crystallographic data for Mg_2(DHFUMA) and Mg_2(DOBDC). Accessible quantities are computed for a probe radius of 1.65 Å

Property	Units	DHFUMA	DOBDC
Crystal density	[g/cm^3]	1.358	0.880
Cell volume	[Å^3/unit cell]	2120	4121
Accessible Surface Area	[m^2/g]	1043	1782
Accessible Surface Area	[m^2/cm^3]	1416	1568
Accessible Surface Area	[Å^2/unit cell]	272	646.3
Accessible Volume	[cm^3/g]	0.0945	0.350
Accessible Volume	[cm^3/cm^3]	0.130	0.308
Helium Void Fraction	[dimensionless]	0.484	0.716
Accessible Volume	[Å^3/unit cell]	272	1271
Largest Included Sphere	[Å]	7.6	11.8
Largest Free Sphere	[Å]	6.3	11.1

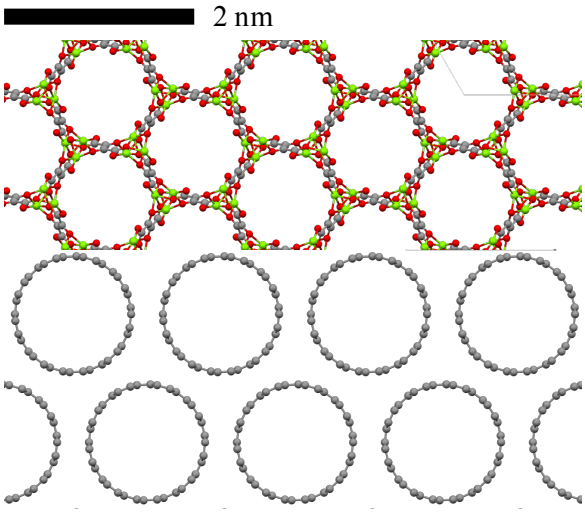


Figure A.1: An segment of the periodic supercell of Mg_2(DHFUMA) is drawn to scale next to bundle of (n=5,m=5) SWNTs

We note that the channel geometry of Mg_2(DHFUMA), as well as each metal analog of this framework, exhibits a striking similarity to the size and shape of carbon nanotubes.

Fig. A.1 illustrates a to-scale representation of a periodic superlattice of Mg_2(DHFUMA) adjacent to a bundle of (n=5,m=5) single wall carbon nanotubes (SWNTs) which were generated by TubeGen Online [*http://turin.nss.udel.edu/research/tubegenonline.html*]. Yet the similarities between the two end with the pore shape and channel dimensionality, since the MOF structure has diverse chemical functionality (i.e. open metal sites) along the channel walls and the pore network is perfectly aligned in one dimension due to the chemical bonding between adjacent channels.

A.6 Force field parameterization details

All force field parameters and the polarizable model of Pham et al. were used to model H_2 adsorption in Mg_2(DHFUMA) and Mg_2(DOBDC).[88] Several minor differences exist between our RASPA2 simulation and the implementation of Pham et al. that may describe the small discrepancy in the simulated isotherms for Mg_2(DOBDC). Firstly, the partial atomic charges of all framework atoms were taken from our REPEAT analysis (see Section A.4) which are not identical to those of Pham et al. Secondly, we accounted for quantum nuclear effects by a second order Feynmann-Hibbs correction term to the Lennard-Jones potential (equation A.1) as implemented in RASPA2, whereas Pham et al. used the fourth order correction term (equation A.2). Thirdly, Pham et al. utilize a modified dipole field tensor to damp dipole-dipole interactions that exhibit short range divergences. Finally, back polarization was not accounted for in our simulation. Our H_2 charges, H_2 Lennard-Jones parameters, framework Lennard-Jones parameters, and static point polarizabilities were identical to Pham et al.

$$U_{FH} = U + \frac{\beta\hbar^2}{24\mu}(U'' + \frac{2}{r}U') \tag{A.1}$$

$$U_{FH} = U + \frac{\beta\hbar^2}{24\mu}(U'' + \frac{2}{r}U') + \frac{\beta\hbar^4}{1152\mu^2}(U'''' + \frac{4}{r}U''' + \frac{15}{r^3}U') \tag{A.2}$$

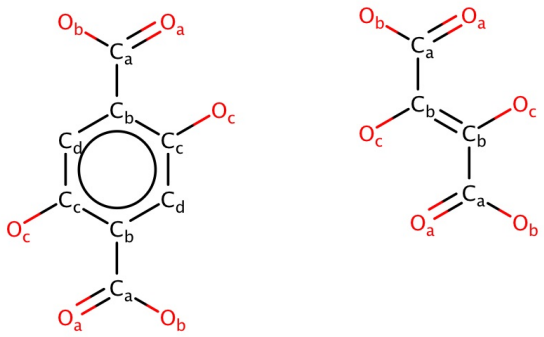


Figure A.2: Atom types assigned for DOBDC ligand as defined by Mercado's FF (left) and atom types assigned for DHFUMA ligand (right) in this work.

For modeling H_2O and CO_2 in the M_2(DHFUMA) and M_2(DOBDC) series, the Buckingham potential and force field parameters of Mercado et. al. were used.[87] We refer the reader to this publication for all force field parameters and details on its derivation and implementation. Each atom type in the DHFUMA ligand was assigned the analogous type to the DOBDC ligand, shown in Fig. A.2.

A.7 CO_2 force field transferability

Force field parameterizations of adsorbate interactions in nanoporous materials are often validated by their ability to reproduce *ab initio* or experimentally determined adsorbate binding energies and configurations. The agreement in the simulated binding energy between our *ab initio* calculations and the parameterized classical force field of Mercado et al. is shown in Table A.5.

Table A.5: *Ab initio* vs classical calculation of CO_2 binding energies [kJ/mol] in M_2(DOBDC) and M_2(DHFUMA) structures

Metal	DHFUMA			DOBDC		
	Ab initio	FF	% Diff.	*Ab initio*	FF	% Diff.
Mg	-50.1	-54.6	5	-41.3	-43.3	9
Fe	-40.9	-47.0	19	-30.0	-35.8	14
Co	-41.1	-49.8	16	-29.3	-34.1	21
Ni	-46.1	-49.1	5	-34.8	-36.5	6
Zn	-37.6	-39.1	1	-31.3	-31.7	4

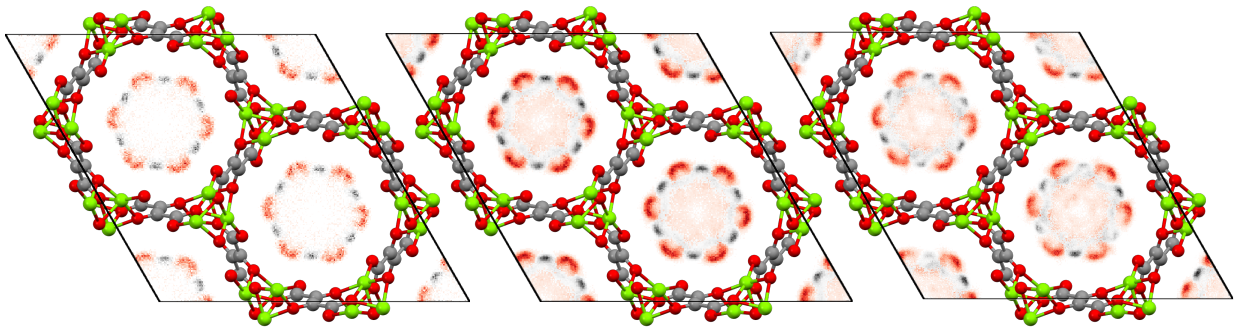


Figure A.3: Two dimensional elemental probability density plots of CO_2 taken from snapshots throughout a GCMC simulation at 313 K at approximate loadings of (a) 1 CO_2 molecule per unit cell; (b) 9 molecules per unit cell which corresponds to 0.5 molecules per open metal site; and (c) 15 molecules per unit cell which corresponds to saturation loading. Black represents Carbon and red represents Oxygen. The element with the higher normalized probability of occupation at that particular pixel determines the coloring.

This agreement is particularly important for accurate computation of gas adsorption behavior in the Henry and low-pressure regimes which are the most important regimes for CO_2 adsorption in the DHFUMA analog series. All DHFUMA analogs display significantly stronger binding than their DOBDC counterpart, a direct result of the optimized binding geometry afforded by the proximity of adjacent open metal sites. The same binding geometry is observed in the CO_2 probability density map generated by extracting various snapshots

of adsorbate configurations throughout the course of a GCMC in Mg_2(DHFUMA). These probability density maps are shown in Fig. A.3 and demonstrate that an Oxygen atom is most likely to be adjacent to an open metal site, whereas Carbon is most likely to be found directly between two oxygen atoms which are each bound to an open metal site. This agreement of CO_2 binding energies and binding geometries between classical GCMC simulations and *ab initio* simulations suggests that Mercado et al.'s classical force field parameterization is indeed transferable to this system and can reasonably reproduce the *ab initio* potential energy surface of CO_2 in the DHFUMA system.

A.8 CO_2 isotherms for M_2(DHFUMA) and M_2(DOBDC)

GCMC computed (T = 313 K and T = 400 K) CO_2 isotherms for all metal analogs of the DOBDC and DHFUMA variants of MOF-74 are presented, plotted in units of molecules CO_2 per unit cell and moles CO_2 per kg framework on normal-normal and log-log scales. The log-log plots particularly demonstrate the enhanced CO_2 uptake in the Henry and low-pressure regimes.

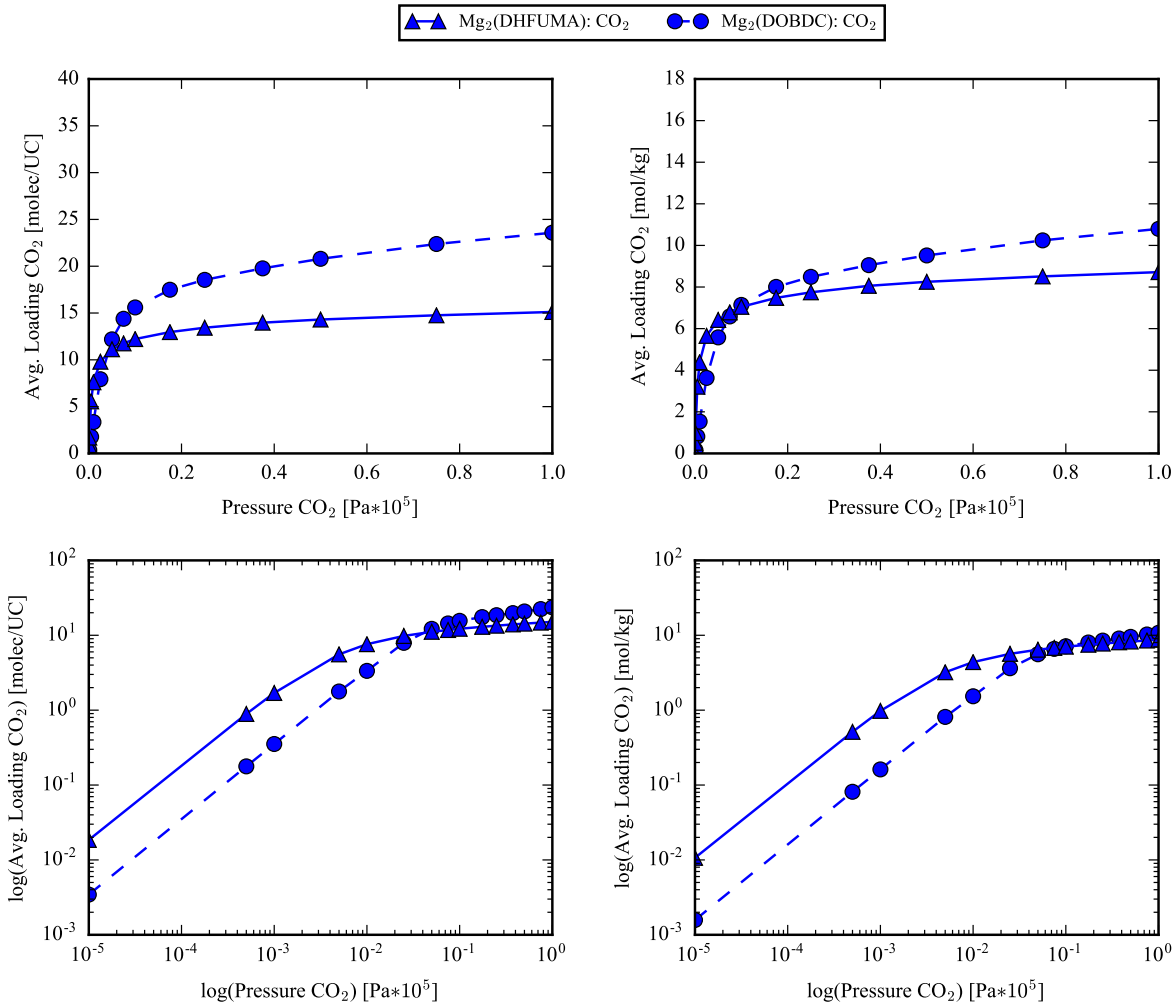


Figure A.4: CO_2 isotherms at T = 313 K for Mg_2(DOBDC) and Mg_2(DHFUMA)

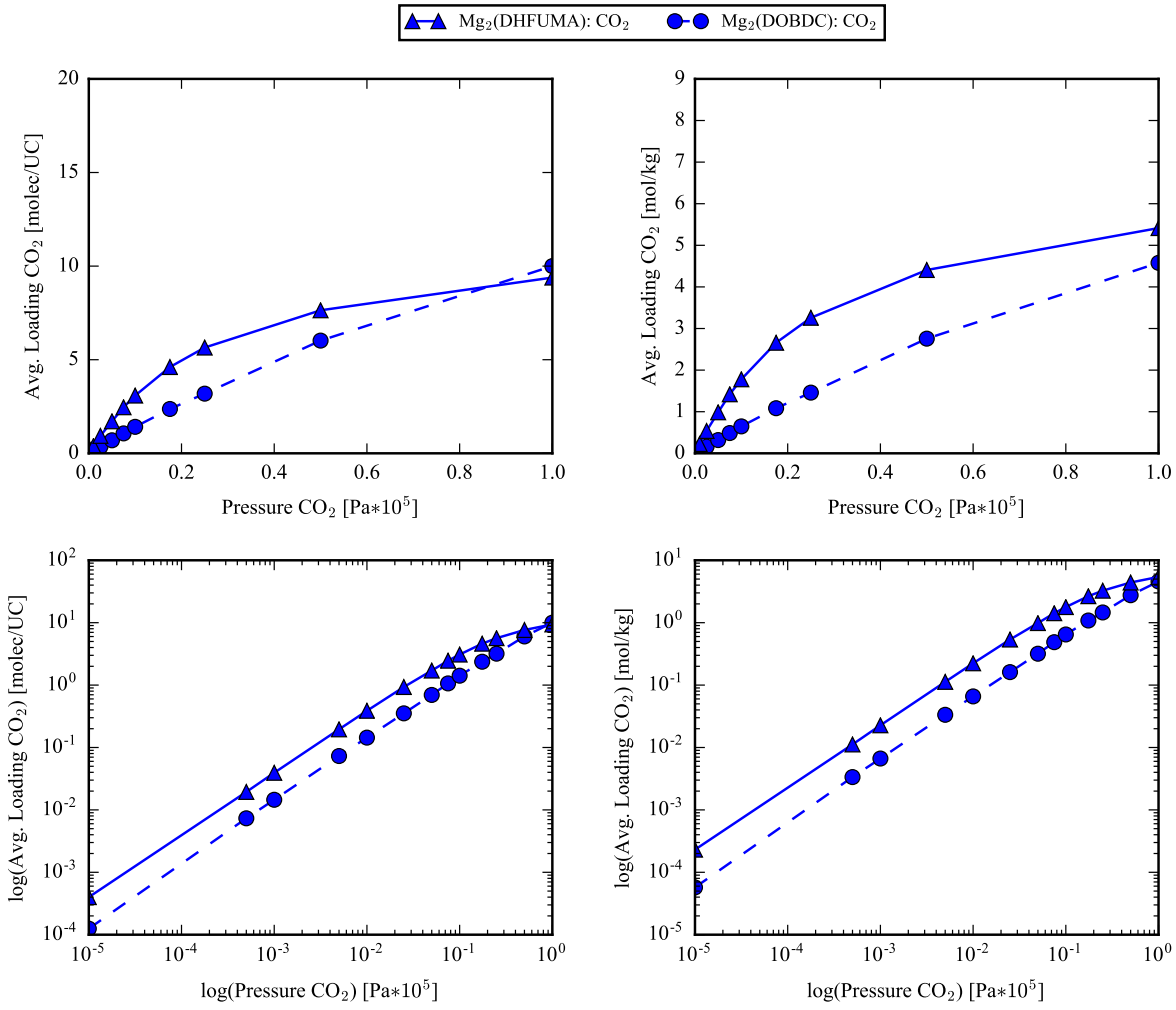


Figure A.5: CO_2 isotherms at T = 400 K for Mg_2(DOBDC) and Mg_2(DHFUMA)

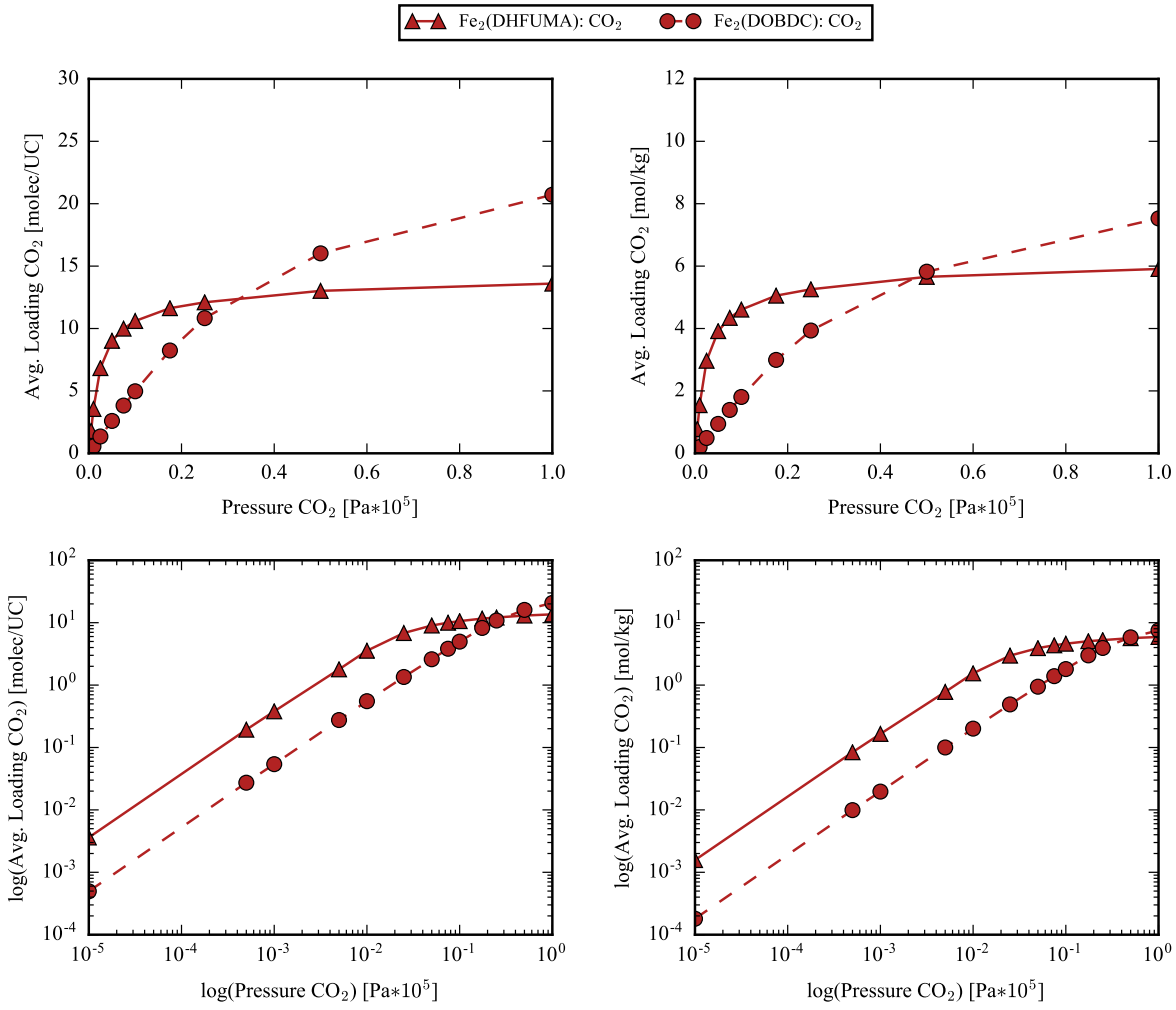


Figure A.6: CO_2 isotherms at T = 313 K for Fe_2(DOBDC) and Fe_2(DHFUMA)

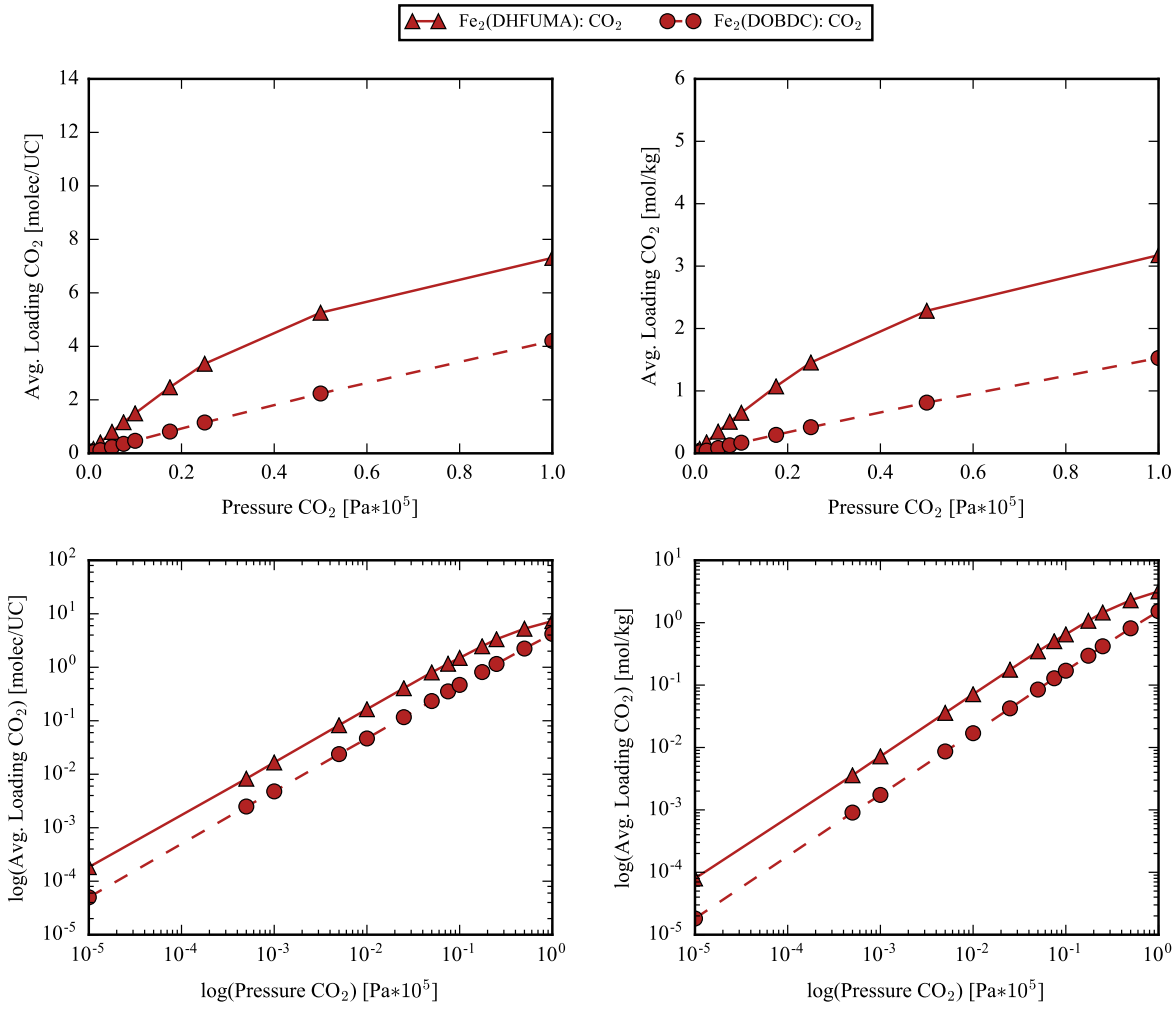


Figure A.7: CO_2 isotherms at T = 400 K for Fe_2(DOBDC) and Fe_2(DHFUMA)

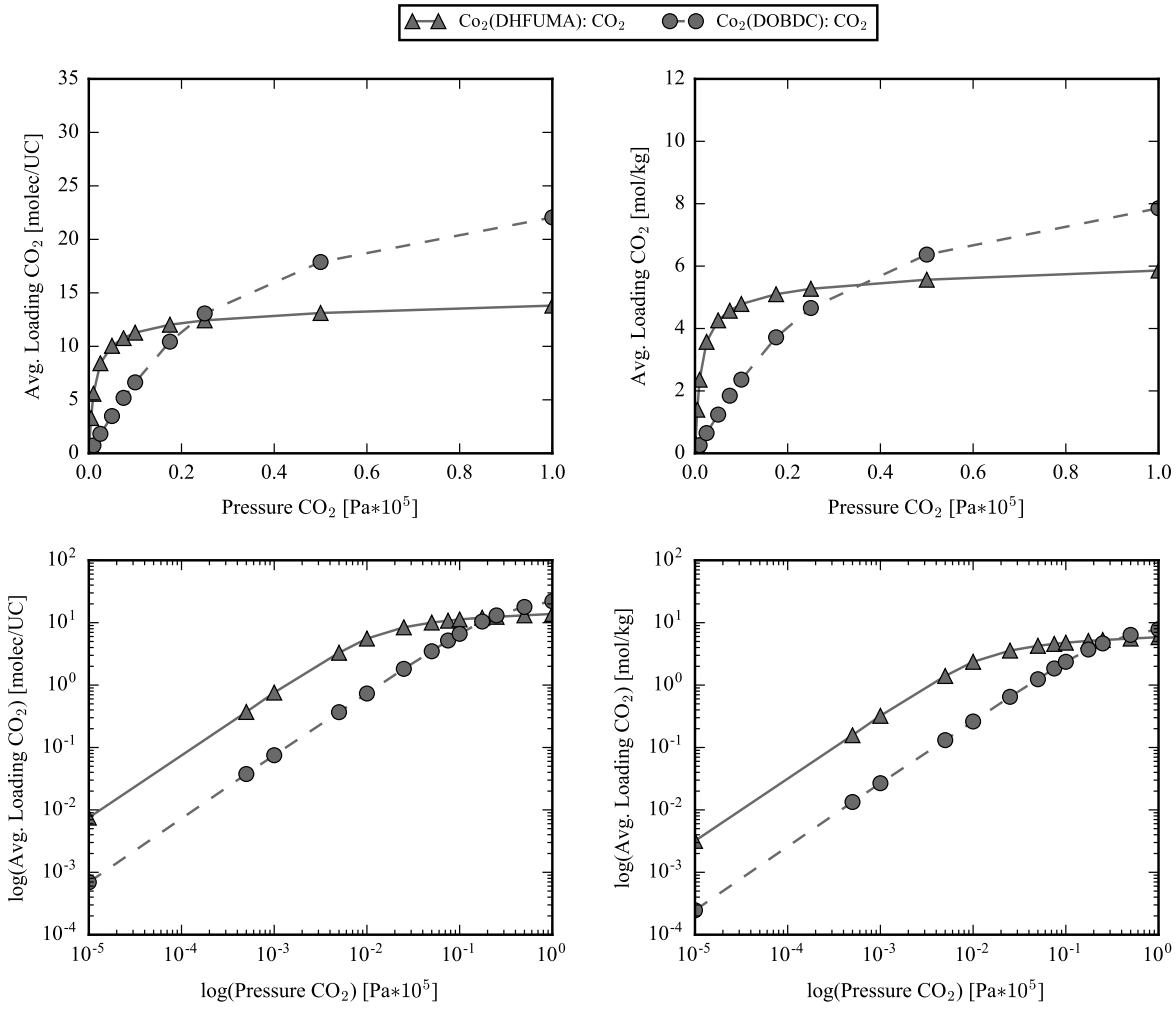


Figure A.8: CO_2 isotherms at T = 313 K for Co_2(DOBDC) and Co_2(DHFUMA)

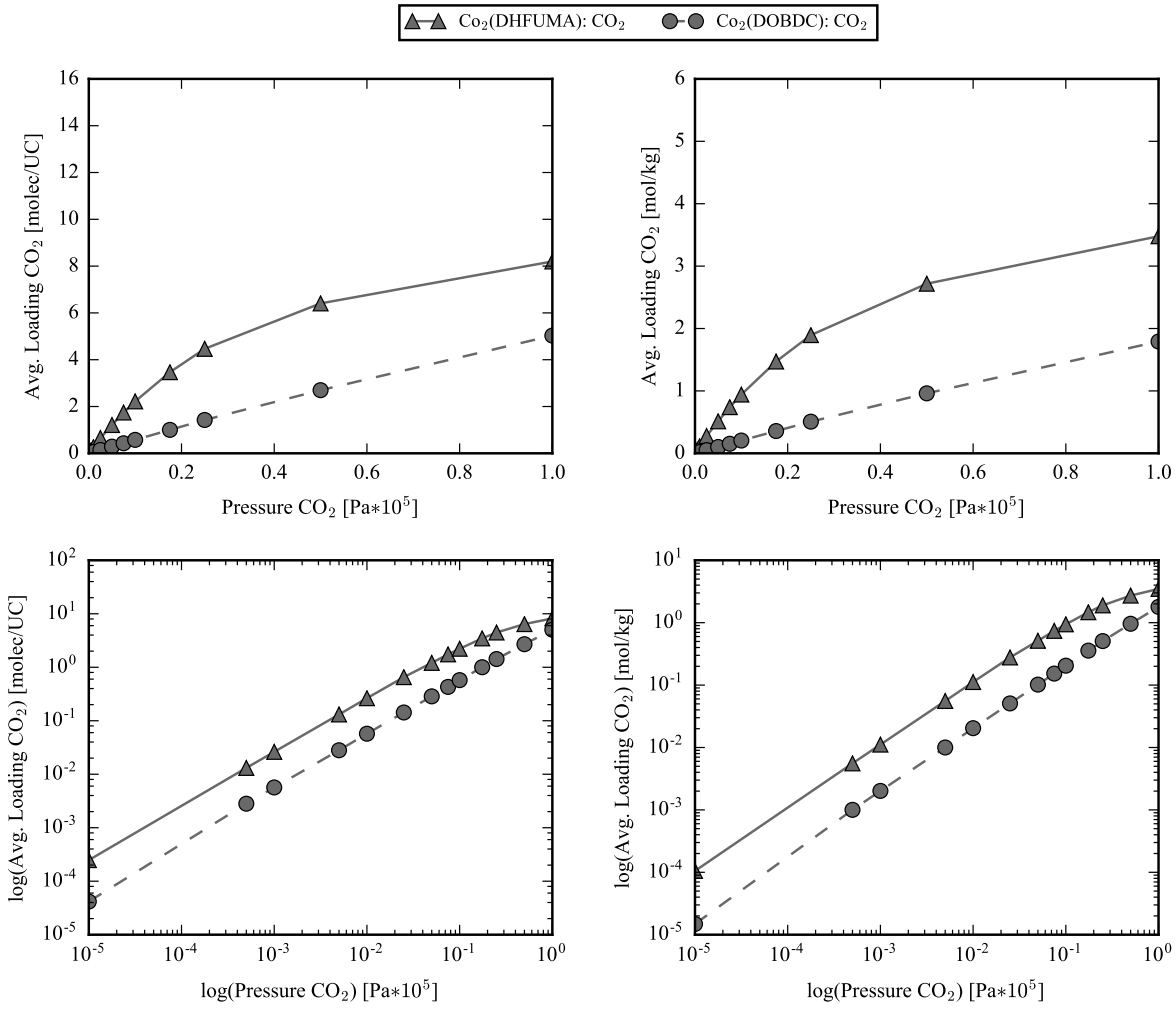


Figure A.9: CO_2 isotherms at T = 400 K for Co_2(DOBDC) and Co_2(DHFUMA)

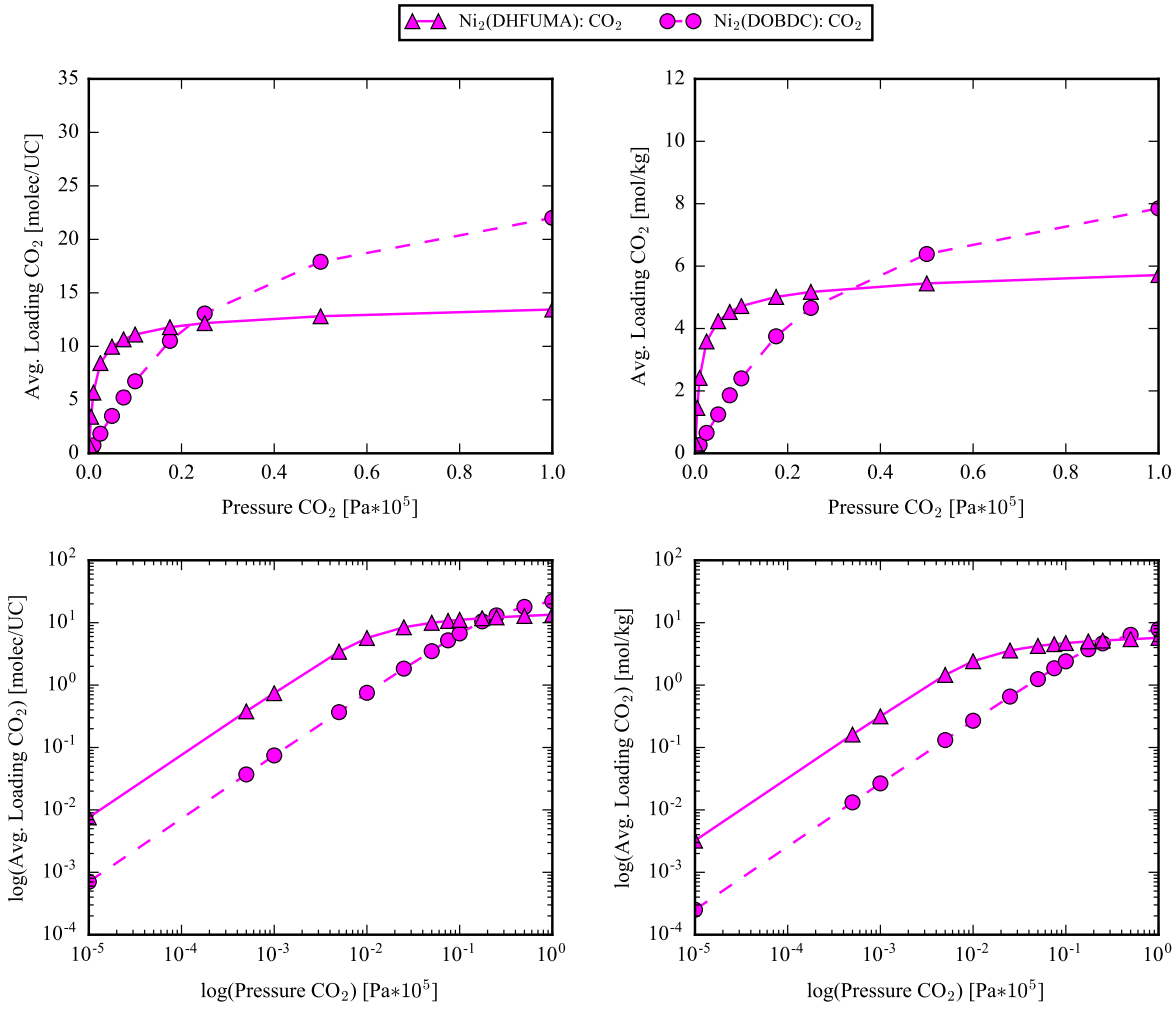


Figure A.10: CO_2 isotherms at T = 313 K for Ni_2(DOBDC) and Ni_2(DHFUMA)

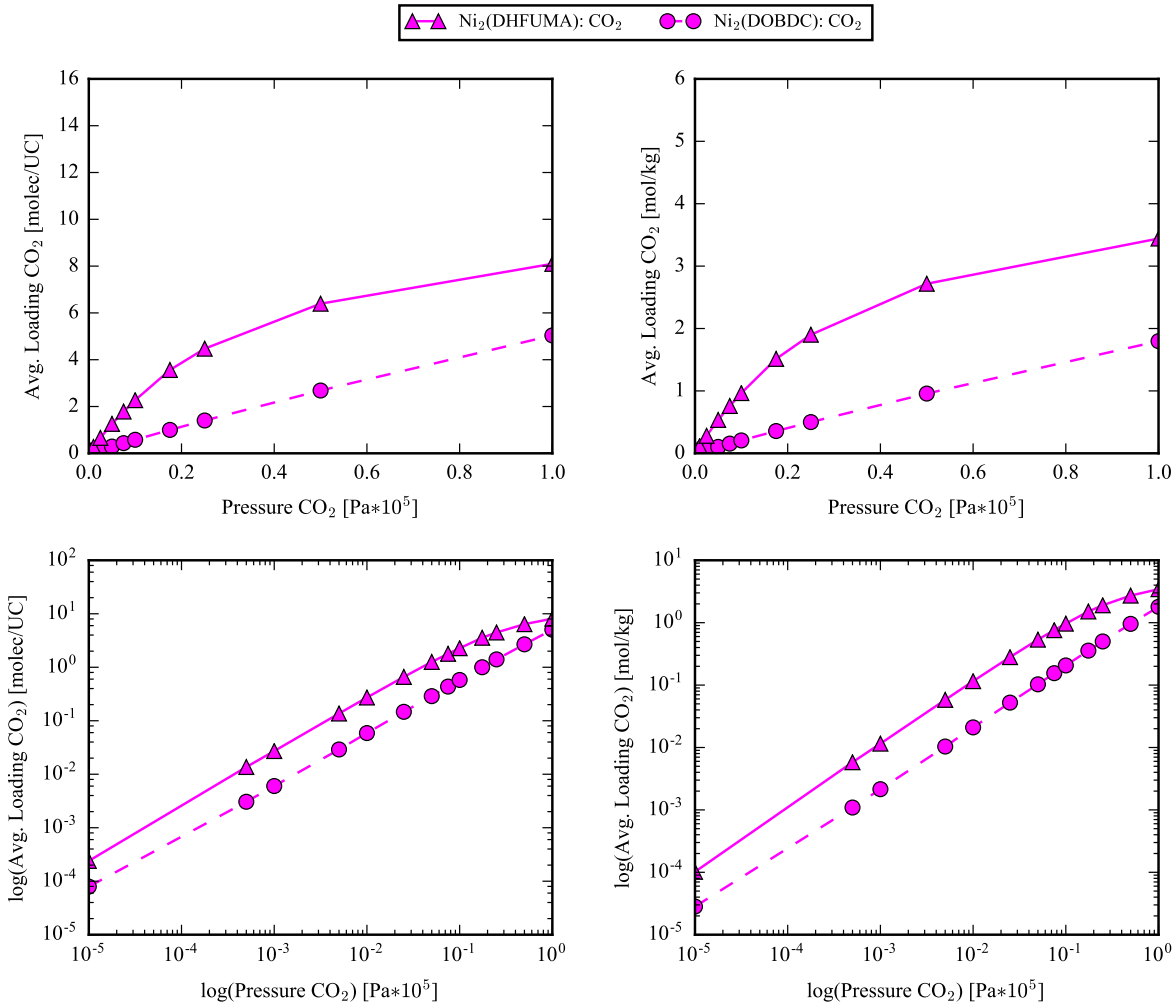


Figure A.11: CO_2 isotherms at T = 400 K for Ni_2(DOBDC) and Ni_2(DHFUMA)

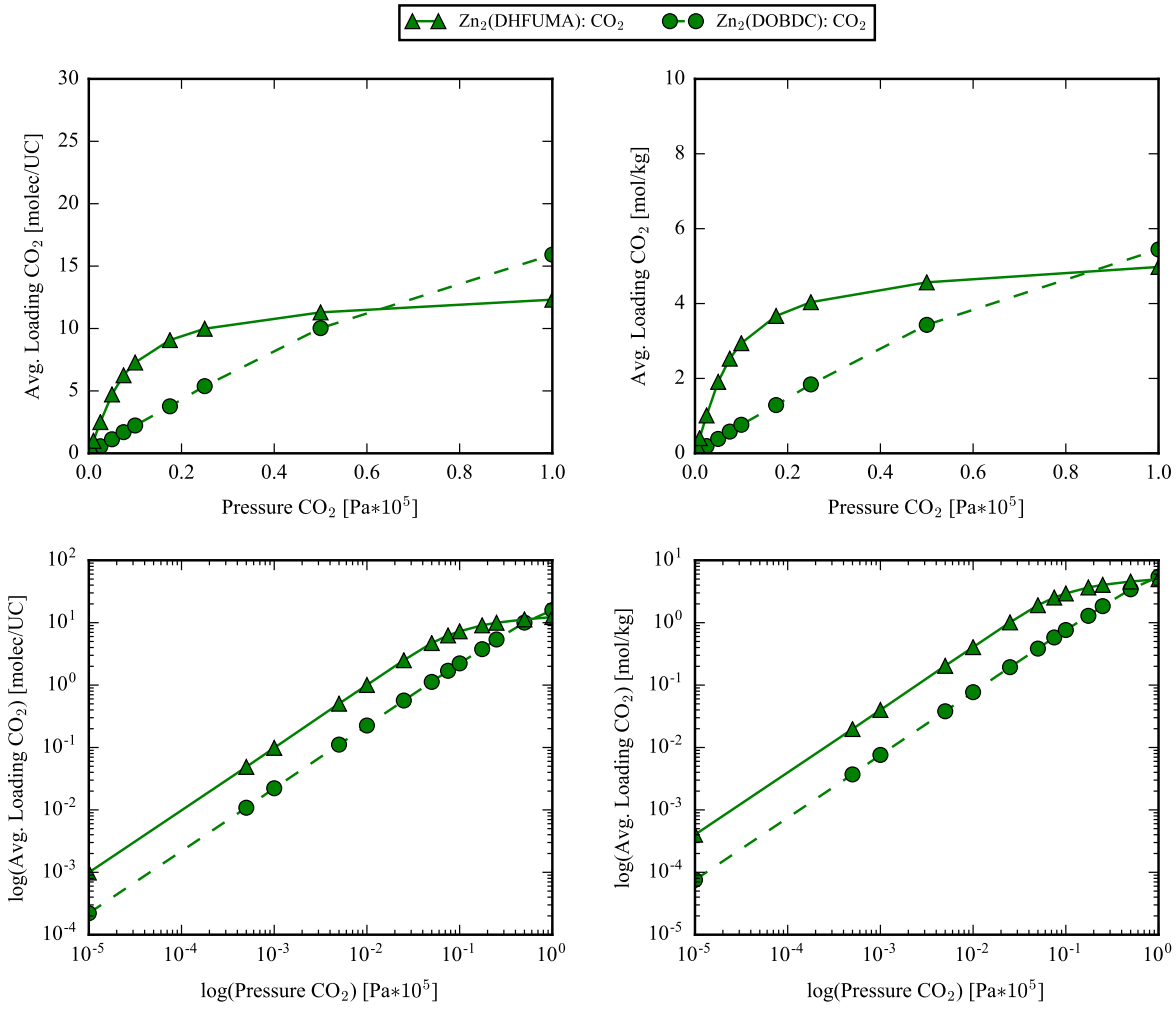


Figure A.12: CO_2 isotherms at T = 313 K for Zn_2(DOBDC) and Zn_2(DHFUMA)

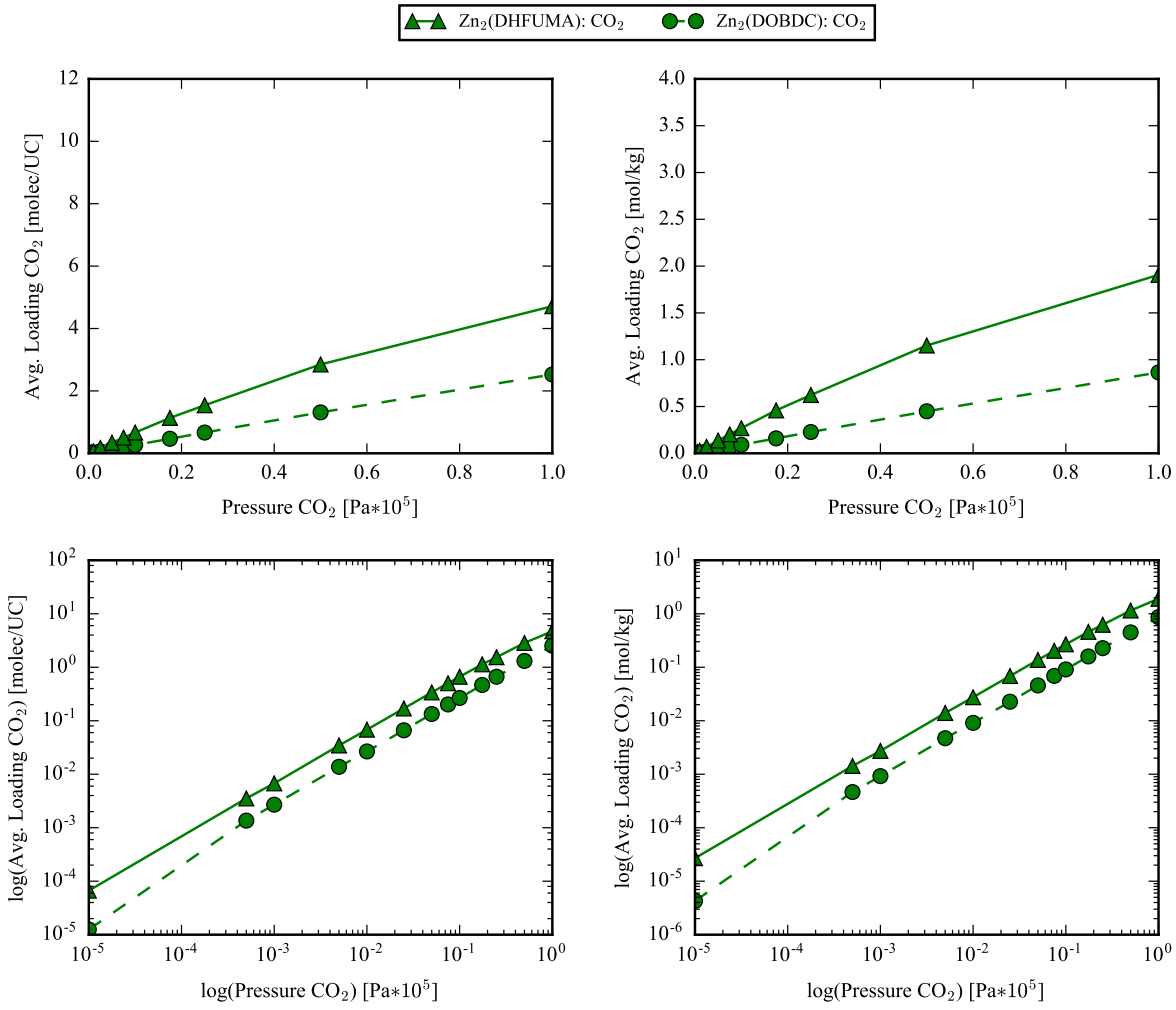


Figure A.13: CO_2 isotherms at T = 400 K for Zn_2(DOBDC) and Zn_2(DHFUMA)

A.9 H_2O isotherms for M_2(DHFUMA) and M_2(DOBDC)

GCMC computed (T = 313.0 K and T = 400.0 K) H_2O isotherms for all metal analogs of the DOBDC and DHFUMA variants of MOF-74 are presented. While DHFUMA condenses H_2O at a much lower external pressure than DOBDC due to the reduced pore size, the condensation point can be dramatically increased by raising the reservoir temperature. For example, the condensation pressure in Mg_2(DHFUMA) increases by almost three orders of magnitude when increasing the reservoir temperature from T = 313 K to T = 400 K. It is worth noting that at a temperature of T = 400 K, a majority of the DHFUMA analogs exhibit water condensation in the H_2O partial pressure range that could be reasonably achieved by drying a typical coal-fired flue gas (on the order of 1 mol% H_2O).

By utilizing the force field of Mercado et al., we are able to simulate the adsorption of H_2O in the M_2(DOBDC) and M_2(DHFUMA) frameworks to obtain a reasonable estimate of the pressure at which the step in this isotherm occurs. We assume the H_2O force field parameters are transferable as is the case with CO_2, yet we note that accurate molecular simulation of water is a challenging feat and an ongoing topic of research for decades. Simulation of water in confined spaces (such as the channels of DOBDC and DHFUMA) is also extremely challenging and extremely dependent of water model and force field parameters, as evidenced by a dearth of agreements between experimental and theoretical results for water adsorption in zeolites, MOFs, etc.[267] We expect three major trends to exist in the water adsorption isotherms for M_2(DOBDC) and M_2(DHFUMA). Firstly, we expect the condensation to occur in DHFUMA analogs at a lower pressure than their DOBDC counterpart since the pore channel is smaller and enhanced cooperative adsorption is stronger. Secondly, we expect the metal with strongest H_2O binding energy to exhibit a step at the lowest H_2O pressure. This trend matches the DFT binding energies of H_2O with the exception of the reversal of Mg_2(DHFUMA) and Ni_2(DHFUMA). Thirdly, the H_2O pressure at which the adsorption step occurs must increase with increasing temperature since higher temperatures shift the chemical potential of the external reservoir. Fig. A.14 demonstrates these three major trends that are at minimum qualitatively correct but in the best case would exactly reproduce the experimental adsorption trends across all metals in each structural series.

Despite the difficulties associated with molecular simulation of H_2O, Fig. A.14 demonstrates the the most important trend regarding our discussion of CO_2 adsorption in the presence of H_2O: high temperature adsorption increases the condensation pressure of H_2O in the M_2(DOBDC) and M_2(DHFUMA) frameworks.

The same binding configuration of H_2O is predicted by the Mercado force field and *ab initio* calculations. Fig. A.15 shows the elemental probability density map of H_2O in Mg_2(DHFUMA) generated from snapshots throughout a GCMC simulation (P = 0.5 Pa, T = 313 K), which corresponds to an average loading of about 30 H_2O molecules/UC (or $\sim$1.5 molecules/Mg^{2+}). The density map clearly demonstrates that the dominant binding mode is that of one Oxygen per open metal site.

This configuration agrees closely with the binding geometry of H_2O as calculated from our DFT optimization, shown in Fig. A.16.

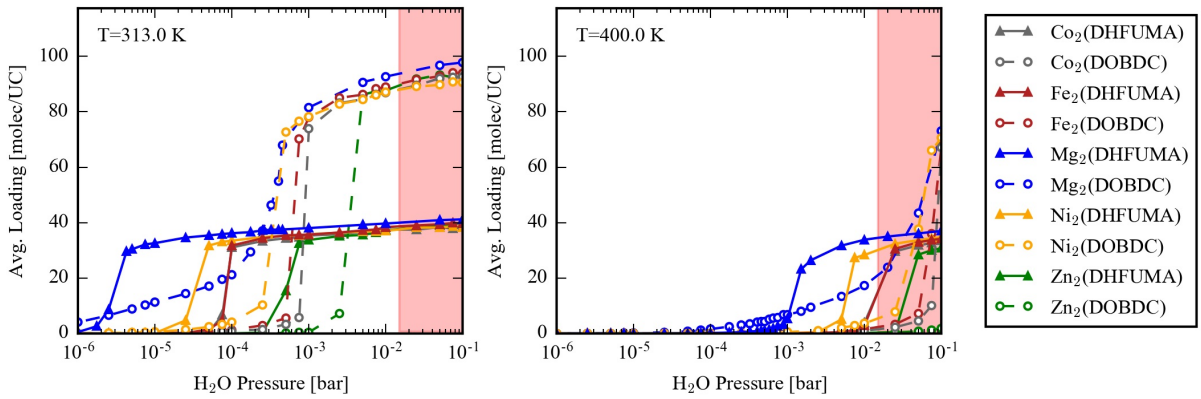


Figure A.14: Absolute H_2O adsorption in M_2(DOBDC) and M_2(DHFUMA) at (a) 313 K and at (b) 400 K. The pressure region in red corresponds a broad partial pressure range of H_2O (P = ∼0.01-0.1 bar) that one might expect in the scrubbed exhaust gas from a coal fired power plant.

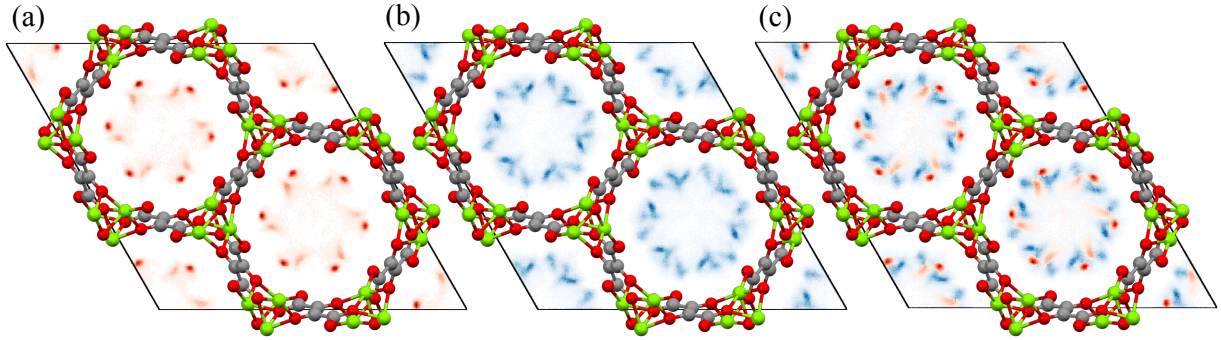


Figure A.15: Elemental probability density maps of H_2O taken from snapshots of a GCMC simulation at P = 0.5 Pa and T = 313 K, which corresponds to an approximate loading of 30 molecules/UC or ∼1.5 molecules/M^{2+}. Shown is (a) the elemental probability density of O alone; (b) the density of H alone; and (c) the density of both O and H overlaid where the coloring of each pixel is determined by the element with higher normalized probability of occupation at that grid point.

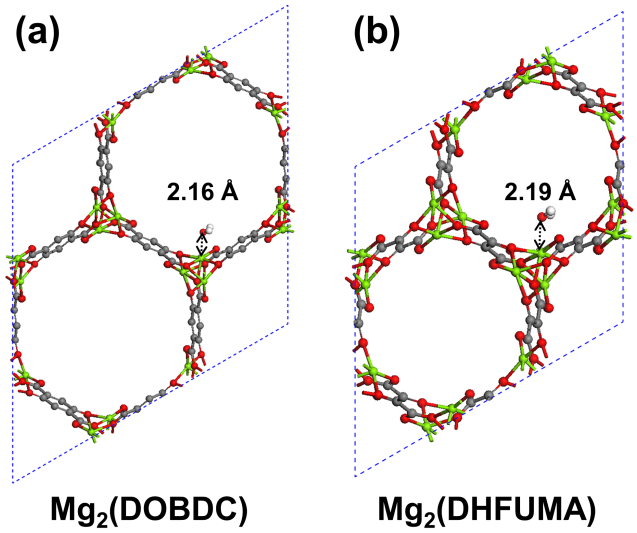


Figure A.16: DFT optimised binding configurations of H_2O in (a) Mg_2(DOBDC) and (b) Mg_2(DHFUMA).

A.10 H_2O:CO_2 mixture analysis for M_2(DHFUMA) and M_2(DOBDC)

GCMC simulations of a binary H_2O:CO_2 mixture at fixed total pressure of P = 0.15 bar for all analogs at varying temperatures are summarized below. Each data point corresponds to a single GCMC simulation with 2-components. The molar ratio of H_2O:CO_2 is varied at constant total pressure to generate the mixture analysis for each structure at each temperature. A stepped decrease in the CO_2 equilibrium uptake occurs at a critical mole fraction of water (i.e. the critical pressure at which water condenses in the channels of each structure).

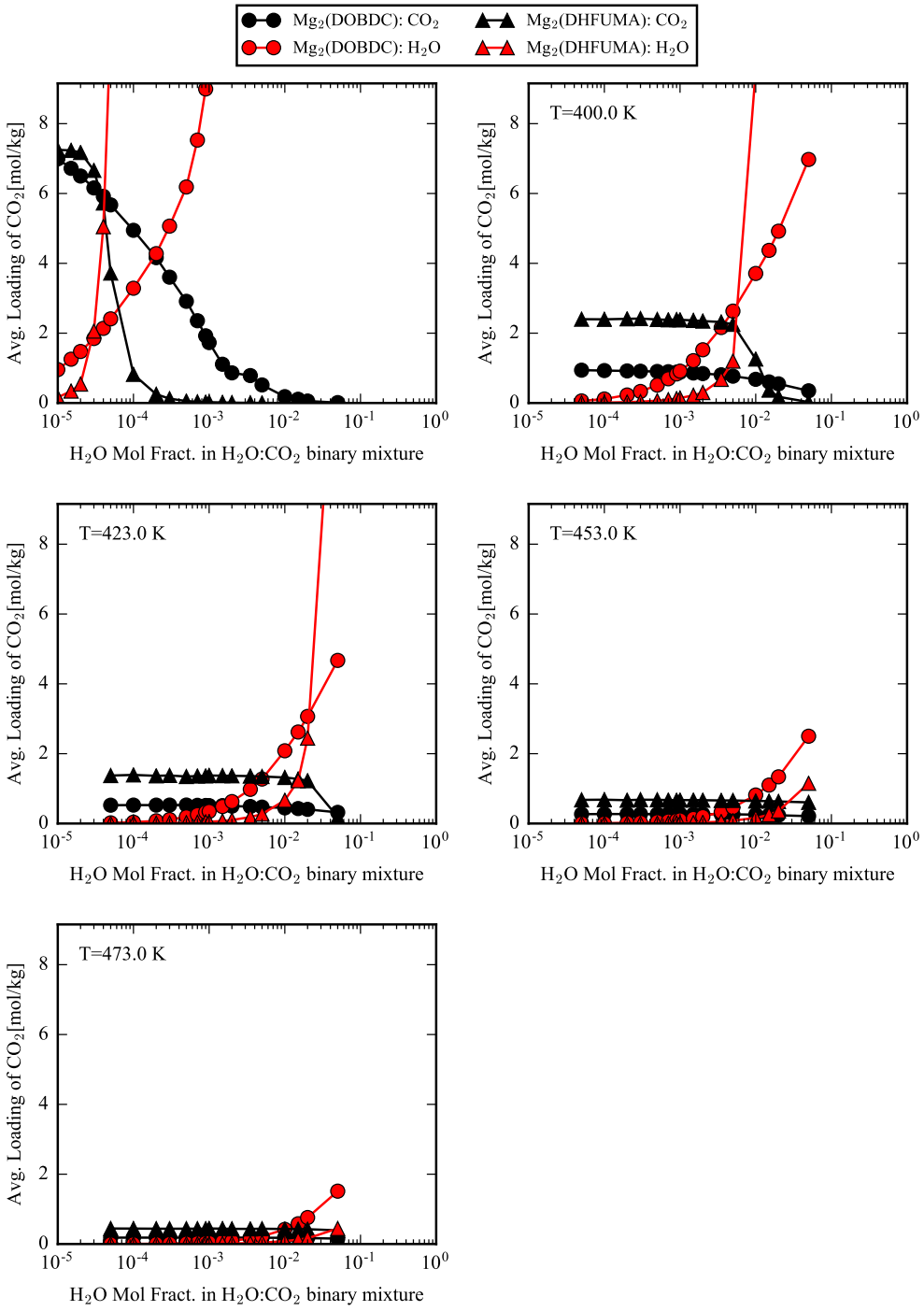


Figure A.17: H_2O:CO_2 mixture analysis for all Mg analogs. Each data point corresponds to one GCMC simulation at P_{total} = 0.15 bar and each subplot corresponds to a fixed temperature of T = 313, 400, 423, 453, or 473 K where the molar composition of the H_2O:CO_2 mixture is varied.

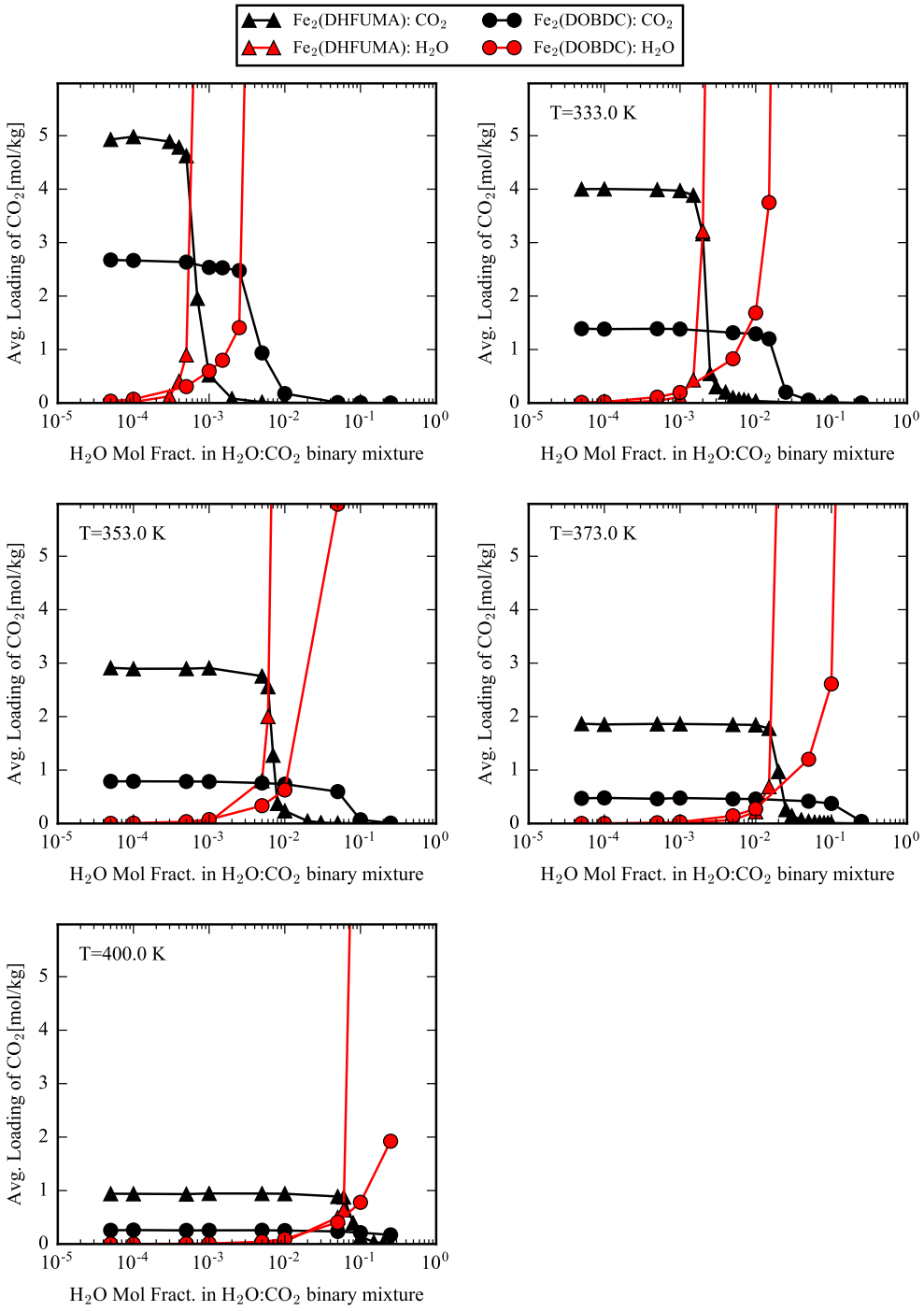


Figure A.18: H_2O:CO_2 mixture analysis for all Fe analogs. Each data point corresponds to one GCMC simulation at P_{total} = 0.15 bar and each subplot corresponds to a fixed temperature of T = 313, 333, 353, 373, or 400 K where the molar composition of the H_2O:CO_2 mixture is varied.

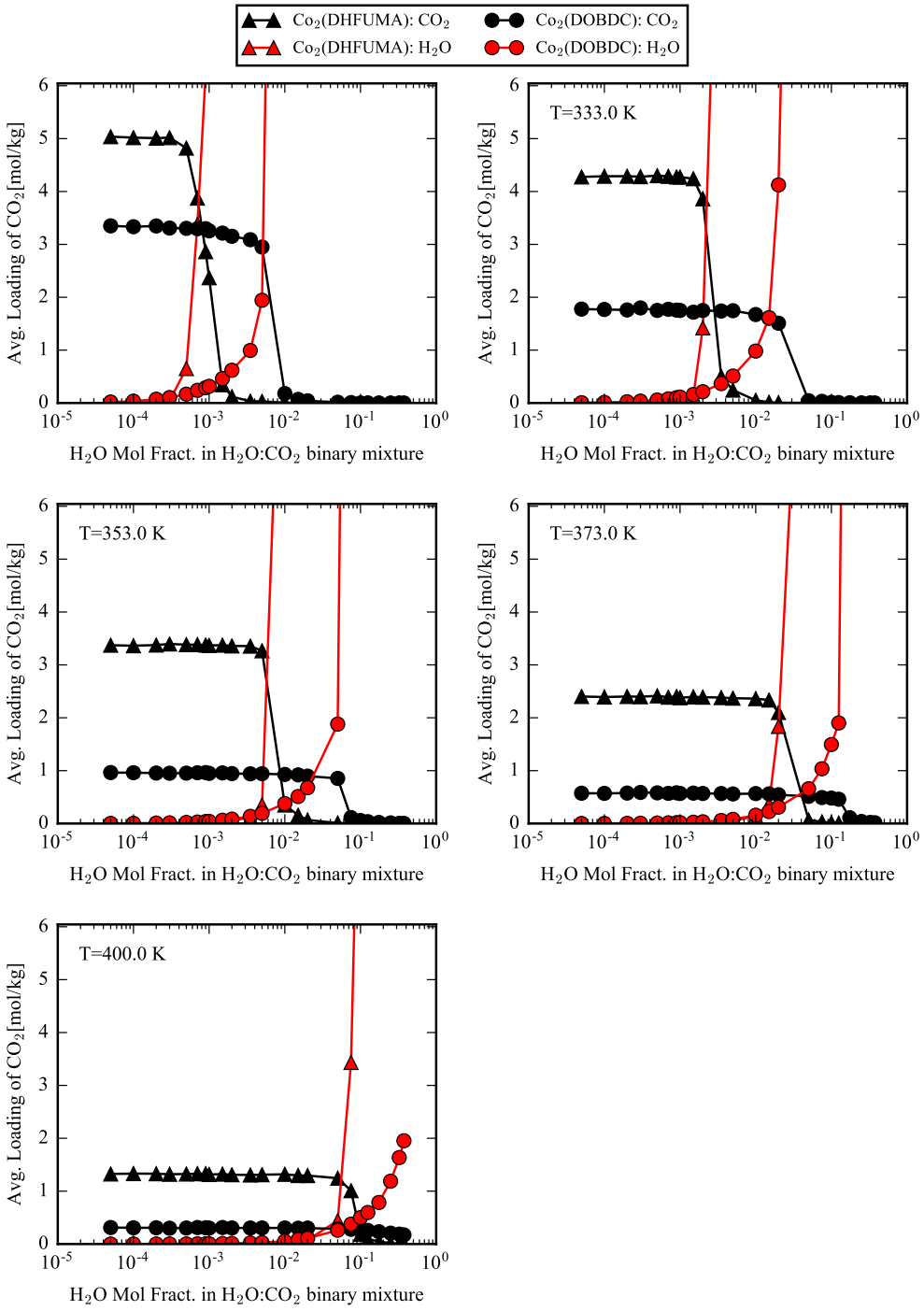


Figure A.19: H_2O:CO_2 mixture analysis for all Co analogs. Each data point corresponds to one GCMC simulation at P_{total} = 0.15 bar and each subplot corresponds to a fixed temperature of T = 313, 333, 353, 373, or 400 K where the molar composition of the H_2O:CO_2 mixture is varied.

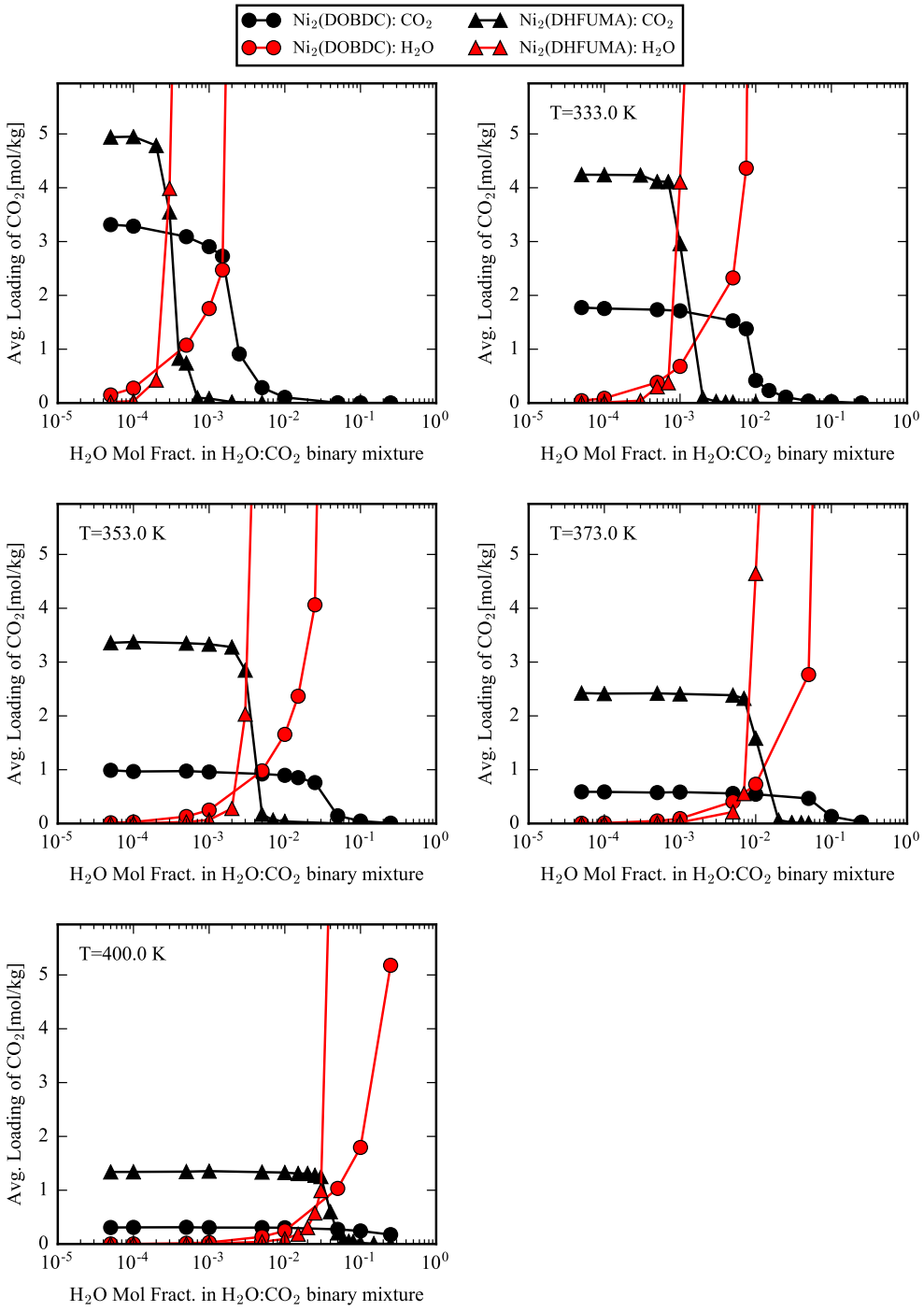


Figure A.20: H_2O:CO_2 mixture analysis for all Ni analogs. Each data point corresponds to one GCMC simulation at P_{total} = 0.15 bar and each subplot corresponds to a fixed temperature of T = 313, 333, 353, 373, or 400 K where the molar composition of the H_2O:CO_2 mixture is varied.

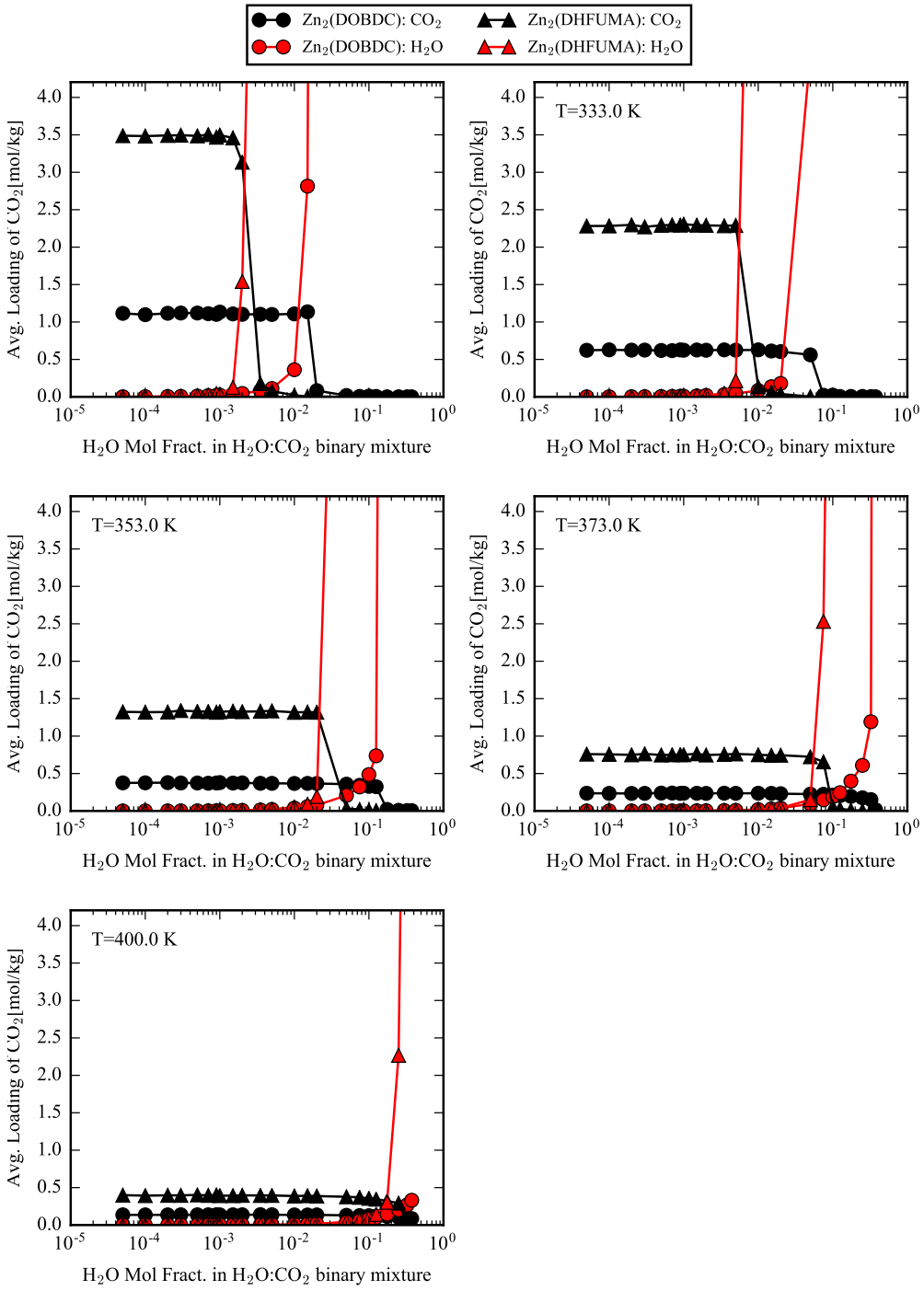


Figure A.21: H_2O:CO_2 mixture analysis for all Zn analogs. Each data point corresponds to one GCMC simulation at P_{total} = 0.15 bar and each subplot corresponds to a fixed temperature of T = 313, 333, 353, 373, or 400 K where the molar composition of the H_2O:CO_2 mixture is varied.

A.11 Improvements in raw material costs

A significant advantage of the M_2(DHFUMA) analog series is the drastically reduced raw materials cost (RMC) in comparison to the M_2(DOBDC) series. A non-bulk order price quote from Sigma Aldrich yields a cost of \$466.0/100g of ligand for DHFUMA which is significantly cheaper in comparison to the \$2412/100g of DOBDC [*http://www.sigmaaldrich.com*, accessed 8 August 2016]. This ligand RMC cost dominates the expense of the metal source, where prices range from \$26.35/100g for $MgCl_2$ to \$65.00/100g $NiCl_2$ (again taken from non-bulk order price quotes from Sigma Aldrich). As we have predicted the increased H_2 volumetric storage potential in Mg_2(DHFUMA) (see Fig. 3.3 in the main article) and the Pareto optimality of all M_2(DHFUMA) analogs over their DOBDC counterparts for CO_2 selectivity in the presence of H_2O (see Fig. 3.9) in the main article), we predict the DHFUMA analog series to be a cheaper, higher performing nanoporous material than M_2(DOBDC) for H_2 storage and CO_2 capture from wet flue gas if it can indeed be synthesized.

A.12 Proposed CO_2 capture process with M_2(DHFUMA)

We graphically illustrate a simple scheme for how one might design an industrial adsorption process if one of the top performing DHFUMA analogs were to be used for CO_2 capture from a mixture of N_2/CO_2 with non-trace amounts of H_2O. Fig. A.22 provides a general overview for how DHFUMA could be incorporated into an industrial scale CO_2 capture process from a coal-fired power plant flue stream. Fig. A.22(a) shows how a zeolite bed could be used to dry the flue gas to a permissible water content for DHFUMA while CO_2 adsorption occurs in the subsequent DHFUMA bed at high temperature. Desorption is illustrated in Fig. A.22(b) whereby the zeolite bed is heated to a high-temperature to desorb H_2O, and this desorbed water stream is cooled down just below the necessary temperature to condense in DHFUMA and thereby desorb CO_2. The ultimate result is a qualitative picture of how CO_2 capture might be performed at an industrial scale. The final amount of CO_2 captured and the amount of zeolite needed to dry the breakthrough of the stream to the minimum necessary water content would be optimized based upon experimental pilot-scale results that mimic the results of the section on "Optimizing CO_2 capture in binary CO_2:H_2O mixtures".

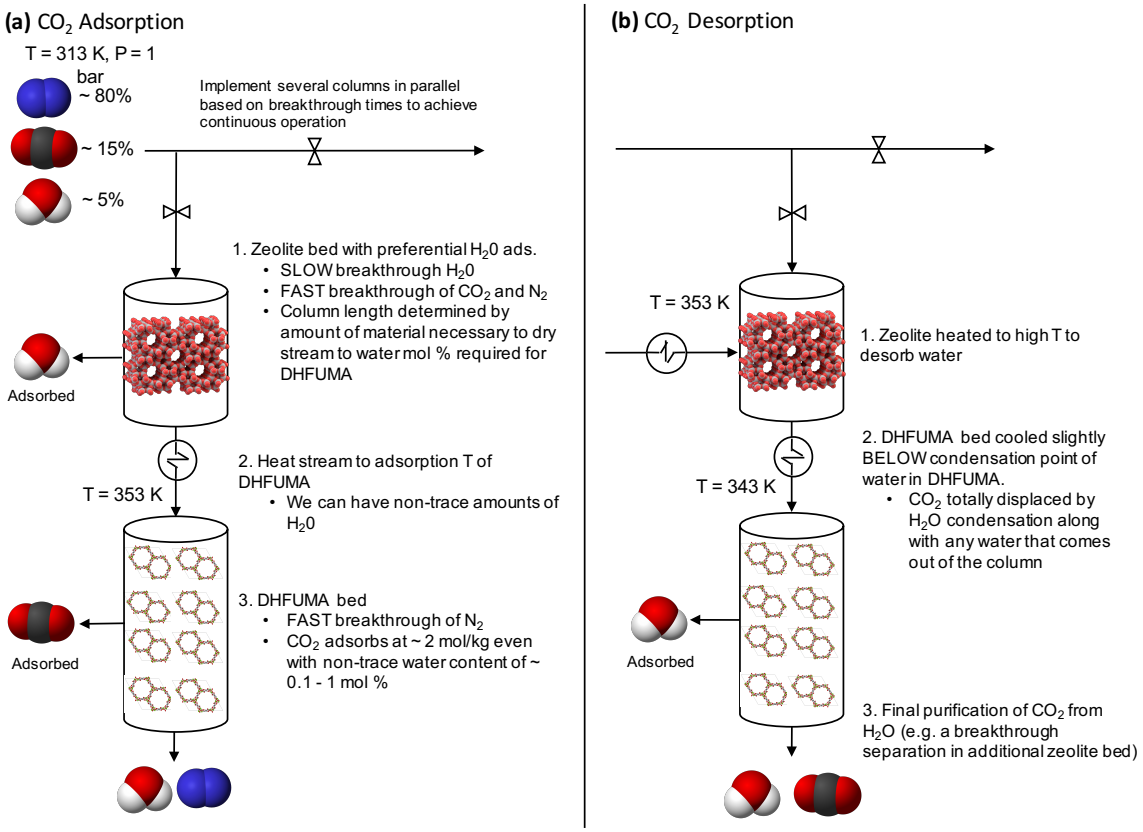


Figure A.22: (a) CO_2 adsorption. A cheap zeolite bed is utilized to reduce the water content to the maximum threshold for the DHFUMA bed, and high temperature adsorption of CO_2 occurs. (b) CO_2 desorption. The water that was originally adsorbed by the first zeolite bed is used to desorb the strongly bound CO_2 in the DHFUMA bed.

A.13 Attempted synthetic procedures

Our attempts to synthesize the family of M_2(DHFUMA) materials have for the moment been unsuccessful since the tested reaction conditions lead to crystalline products other than the desired structure. Details on the attempted synthesis procedures and the resulting products are discussed in the supporting information of Ref. [225].

Appendix B

Additional details on predicting 2D zeolites

B.1 Python implementation of the Balcioglu and Wood algorithm

A Python implementation of the Balcioglu and Wood algorithm[148] is provided in the "Nearmincut" github repository (`https://github.com/mwitman1/nearmincut.git`). This routine has been incorporated into the private version of the LammpsInterface[165] (`https://github.com/peteboyd/lammps_interface.git`) program which handles all other logistics of structure processing using both Pymatgen[146] and its own internal functionality. This new version may be publicly released in the near future, but for immediate inquiries regarding the use of this new functionality please contact the developers of LammpsInterface.

Formal definitions of min cuts are provided in the main manuscript, but a brief overview is provided here to understand the functionality of Nearmincut. Nearmincut accepts a Networkx graph and a source and target node to compute the min s-t edge cut. Denoted $w(C_0)$, this is the sum of the weights of all edges in the cut, where a cut is a set of edges that, if removed, interrupts all paths form s to t. If C_0 is the min cut, then $w(C_0) \leq w(C \in \mathbf{C})$, where $\mathbf{C}$ is all possible s-t cuts. Then the algorithm of Balcioglu and Wood recursively searches for all cuts in the set $\mathbf{C}_\epsilon$ whose weights have $w(C) \leq (1+\epsilon)w(C_0)$, where $\epsilon > 0$ is a user specified parameter that determines how "near" a cut is to the min cut. A trivial modification was introduced that also allows the user to output all cuts with weight $w(C) \leq k + w(C_0)$, where again $k > 0$ is a user specified parameter. A simple test script in the repository shows how to use the basic functionality of Nearmincut.

The recursive algorithm of Balcioglu and Wood is briefly summarized here and further details can be found in Ref. [148]. Consider the directed graph in Fig. B.1. Note that this graph is slightly different than the example presented in the manuscript since, instead of anti-parallel, directed edges between nodes, only a single directed edge exists. This example represents the identification of all cuts with $w(C) \leq k + w(C_0)$ where $k = 2$. Each graph

in Fig. B.2a represents the min cut found at each step of the recursive enumeration of the cutset, and Fig. B.2b visualizes the corresponding recursion tree. At each iteration edges have forced exclusion (by setting the weight to infinity) or forced inclusion (by adding an edge with infinite weight from both the source and target to both nodes in the edge). For example, after the first min cut is identified (Iter. 1) with $w(C_0) = 6$, e_{10} is excluded from the next possible solution by setting its weight to infinity. This produces a min cut with value of 8 in Iter. 2, which satisfies $w(C) \leq k + 6$. Hence we now go one level deeper into the recursion tree. Excluding e_4 does not produce a min cut with $w(C) \leq k + 6$ in Iter. 3. In Iter. 4 we must include e_4 and exclude e_3. In Iter. 4 one can see that the inclusion of e_4 is forced by setting $w(e_1) = w(e_8) = \text{inf}$. This again leads to a min cut with a value greater than 8, leading to termination at this node in the recursion tree. The examples are continued for Iter. 5 and Iter. 6. Upon completion of the recursion tree, one would find in addition that $w(e_{10}e_1) = 8$, yielding four possible solutions for all (near) min cuts.

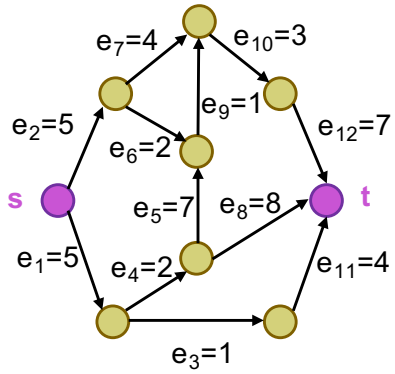


Figure B.1: A sample directed graph for which we seek to find all $C \in \mathbf{C}$ with $w(C) <= k + w(C_0)$ where $k = 2$.

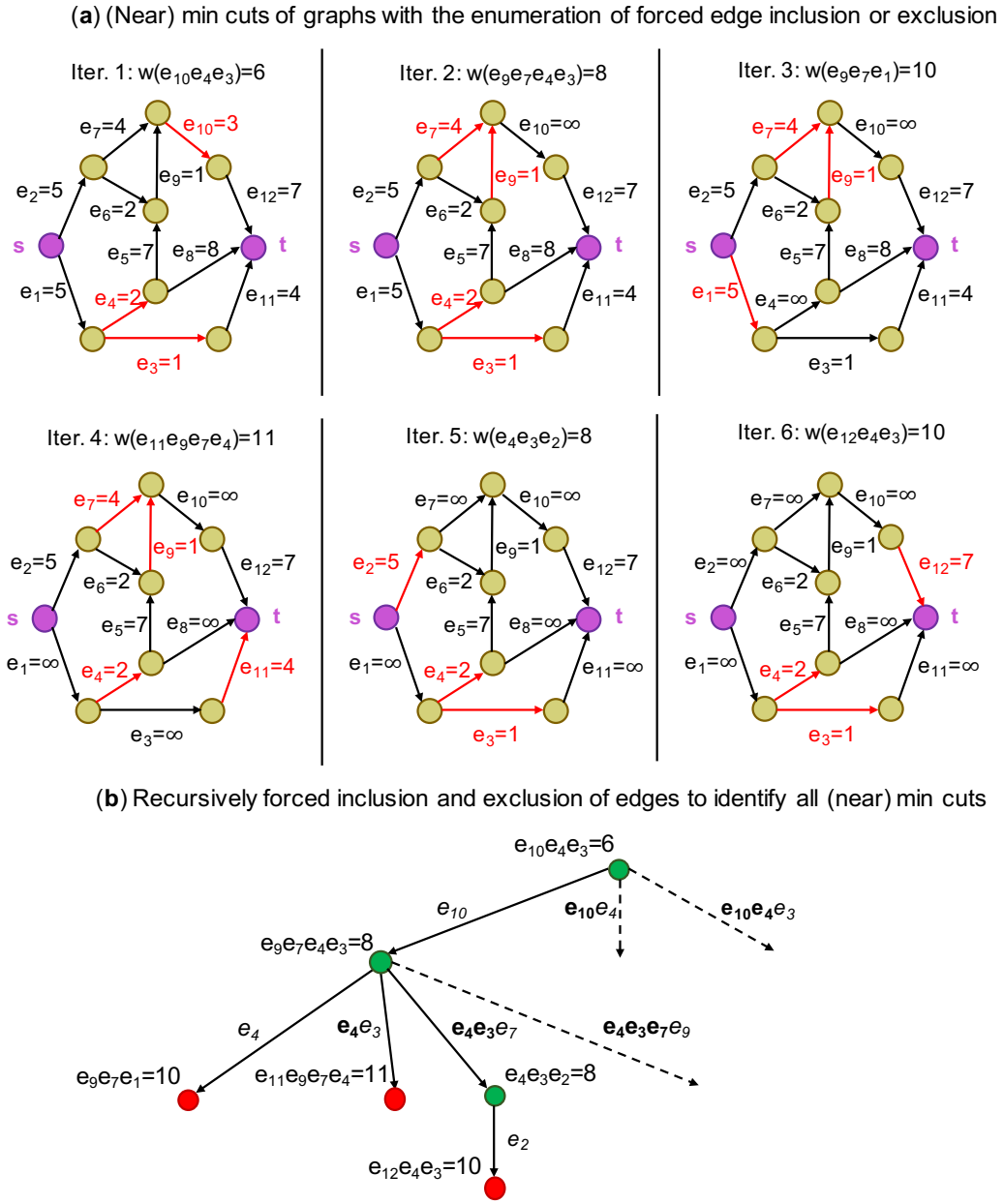


Figure B.2: (a) Visualization of the solution to the min cut problem at the first six iterations in the recursion tree. Red edges represent the min cut in each particular iteration. (b) The partially completed recursion tree where a node represents a cut and the connection to its child shows which edges have forced exclusion (italicized) and forced inclusion (bold) in the next iteration. The base case returns occurs when $w(C) > k + w(C_0)$ (red nodes) for $k = 2$. Dashed lines indicate further exploration of the recursion tree is needed since the parent node satisfies $w(C) \leq k + w(C_0)$

B.2 Statistics of min cuts for IZA zeolites

Fig. B.3 visualizes the ranking scheme demonstrated in Table 4.1 for all IZA zeolites in a single figure. To construct this figure, each IZA contributes one entry in each column of the histogram. For example, MWW contributes one purple unit in the (001) column since the (001) min cut density achieves the lowest cut density (rank R1) of all other Miller planes, which is also shown in Table 4.1. If two Miller surfaces are symmetrically equivalent, both are assigned the same rank for visualization purposes according to a dense ranking scheme (each surface's ranking number is 1 plus the number of items ranked above it that are distinct with respect to the ranking order). Anything above a rank of R16 is included in the R16 bin color scheme. The min cut density for each face of each zeolite, as well as the corresponding structure file, can be found in the supplementary data. According to Fig. B.3, the majority of structures have R1 ranks occurring for (001), (010), and (100) Miller faces, while a non-negligible number have R1 ranks occurring for (011), (101), and (110) faces. The general trend shows that, as might be expected, higher index Miller faces typically result in a higher min cut density. Yet clearly there are many outlying structures which have high index Miller faces with low ranks (R1, R2, or R3 for example). These subtleties highlight the necessity of an automated and robust approach for zeolite surface generation and screening: various IZA structures may require obtaining the min cut density of a relatively high index Miller plane to find the R1 surface. This task would be extremely arduous if not impossible by manual/visual inspection.

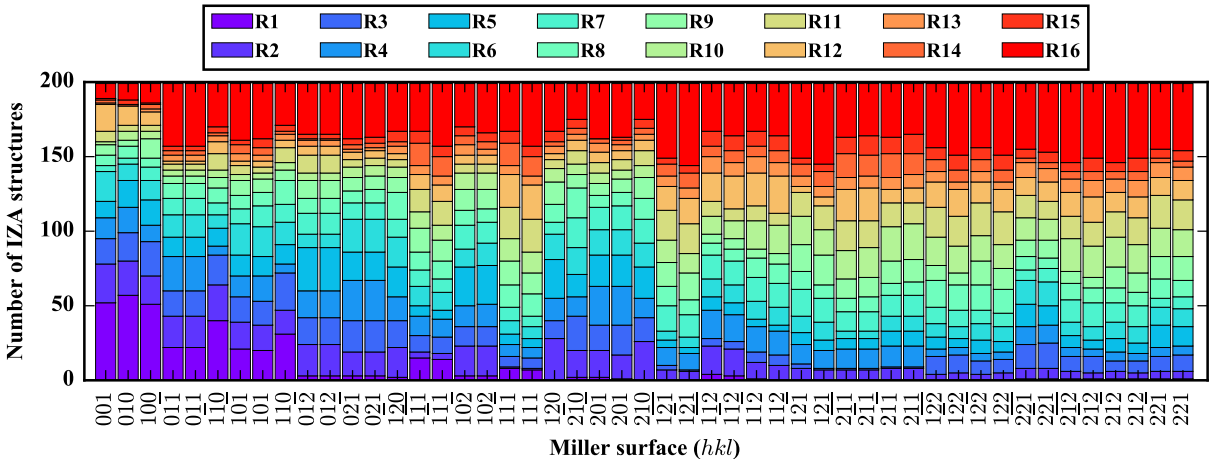


Figure B.3: The tabulated ranks of all 2-maximum index Miller surfaces of each IZA zeolite. Here the minimum cut density corresponds to a rank of R1, and the maximum density corresponds to R16 (all surfaces with a higher density than R16 are included in the R16 color-coding).

The power of the graph theory algorithm can also be appreciated by observing when it identifies min cuts that lead to faceted surfaces. When the min cut is formed by a highly

faceted surface, which is especially true for high index Miller planes in the case of all silica zeolites, it becomes visually clear that only an advanced graph theory based algorithm can identify the solution to the min cut. Take for example several min cuts for various Miller planes in MFI, shown in Fig. B.4. For high index Miller faces, the surface terminations become highly faceted to minimize the number of edges that are cut, but they are so complex that only an advanced algorithmic approach would be able to determine them.

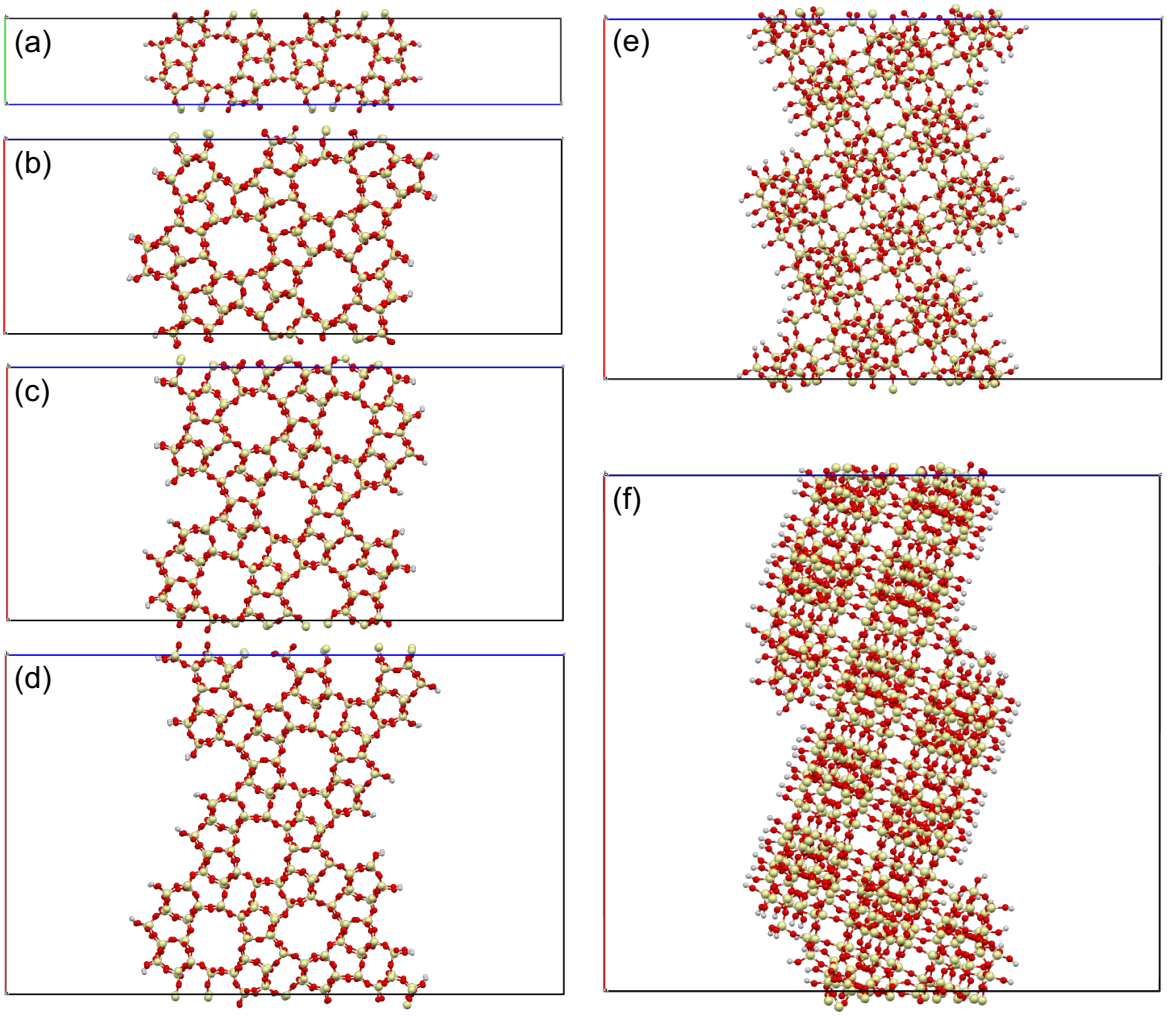


Figure B.4: (a-f) show the min cut surface termination for the (100), (101), (201), (301), (120), and (511) Miller planes of MFI, respectively.

B.3 Visualization of zeolites with high potential for a layered 2D form

The candidates with the largest $\delta_{R2} - \delta_{R1}$ values are visualized in Fig. B.5. The ranking of all IZA materials based on this descriptor is provided in supplementary data files as described in the following sections. Fig. B.5 displays the R1 plane horizontally and orthogonal to the viewing page. These R1 surfaces are the predicted surfaces that would be expressed if the zeolite can be synthesized in a layered 2D form.

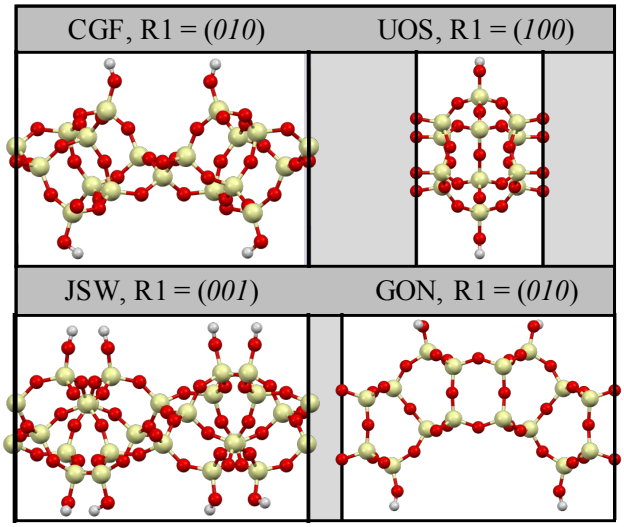


Figure B.5: Visualization of the proposed layered precursor for the four IZA structures with the largest $\delta_{R2} - \delta_{R1}$ values that have not yet shown experimental verification of a 2D form. The structures are shown with the R1 Miller surface horizontal and orthogonal with respect to the viewing.

B.4 Supporting data files

The data files associated with the min cut statistics of 2D zeolites can be found in Ref. [112]

Appendix C

Additional details for high-throughput screening of flexible materials

C.1 Generating simulation input files

To generate the necessary -data and -in required by LAMMPS to simulate the intrinsic flexibility of the CoRE MOFs with the UFF-FM approximation, first clone the git repository *https://github.com/peteboyd/lammps_interface*. From the command line one can generate the files for running NVT simulations (with the force field of UFF-FM as described below) at 298 K via the following:

python ⟨path/to/lammps_interface⟩/lammps_main.py --nvt --fix-metal --equilibration-steps 30000 --production-steps 30000 --dcd 1000 ⟨your_cif_file⟩.cif

The snapshots in the dcd trajectory file can then be easily converted to Crystallographic Information Files (CIFs) using the MDAnalysis Python package,[268] from which the Henry coefficient can be directly calculated in RASPA. Example files can be found in Ref. [225]

C.2 Additional simulation details

UFF and the 'fix-metal' approximation

The UFF fix-metal approximation (UFF-FM) is used to facilitate the screening of flexibility in the CoRE MOF materials by creating a force field applicable to any metal-ligand chemistry. To modify the UFF potential such that each metal ion is preserved in its crystal structure atomic arrangement, the harmonic bond equilibrium distance, R_0, for each bonded atom is modified to the reported crystallographic distance. Likewise, all angle potentials between atoms that share the metal ion as a neighbor were modified such that their equilibrium angle, θ_0, is equal to the angles found in the crystal structure, computed in the usual

way,

$$\theta_0 = cos^{-1}\left(\frac{\overrightarrow{v}_{JI} \cdot \overrightarrow{v}_{JK}}{|\overrightarrow{v}_{JI}| \cdot |\overrightarrow{v}_{JK}|}\right) \tag{C.1}$$

where $\overrightarrow{v}_{JI}$ and $\overrightarrow{v}_{JK}$ are the vectors pointing along the bonds from the metal ion (J) to atoms I and K respectively. θ_0 is then placed in the UFF equation for generalized non-linear angles

$$E_\theta = K_{IJK}\left[C_0 + C_1 cos\theta + C_2 cos2\theta\right] \tag{C.2}$$

where,

$$\begin{aligned} C_2 &= \frac{1}{4sin^2\theta_0} \\ C_1 &= -4C_2 cos\theta_0 \\ C_0 &= C_2\left(2cos^2\theta_0 + 1\right) \end{aligned} \tag{C.3}$$

The only caveat is if θ_0 computed in Equation C.1 is sufficiently close to 180°, in which case the linear function of the angle potential is used. The reader is referred to Ref. [73] for further details. The 'strength' of both the harmonic bonds and angles are kept the same as specified in the UFF recipe, and are not further modified by this approximation. All other bonded potentials (improper and dihedral torsions) in UFF have explicitly no parameters when metal ions are included as one of the central atoms in these potentials, thus these potentials are ignored in the 'fixed-metal' approximation.

UFF and the cationic dummy model

In the UFF-cationic-dummy-model (UFF-CDM) model of SBMOF-1, interactions between Ca^{2+} and linker atoms are purely non-bonded. Ca^{2+} atoms are modeled as a cationic dummy models parameterized by Duarte et al. which delocalizes charge in a way that imposes an octahedral coordination environment.[251] Lennard-Jones interaction parameters between Ca^{2+} atoms and linker atoms were computed using Lorenz-Berthelot mixing rules, with contributing interactions coming from UFF and Duarte et al. for linker and Ca^{2+} atoms, respectively. Because Ca^{2+} atoms in the SBMOF-1 structure do not carry the full 2+ charge seen in pure ionic form, charges on both central and dummy beads were scaled from what was presented by Duarte et al. in order to maintain the same overall charge on the entire cationic dummy complex as the Ca^{2+} charge as determined by REPEAT.[58] The UFF-CDM model of SBMOF-1 structure can capture Ca^{2+} ion coordination numbers of 6 or 7.

C.3 Scaling the potential energy landscape for better experimental agreement

In this section we discuss the potential energy surface (PES) of Xe/Kr adsorbates in SBMOF-1 as calculated by PBE+D3 and the Universal Force Field (UFF) to highlight the other

potential source of discrepancy between experimental results and computational predictions of adsorption properties in this system. In addition to changes in the PSD induced by flexibility, accurately parameterizing the strength of host-adsorbate interactions contributes to obtaining good agreement between computational and experimental Xe/Kr Henry coefficients. For the case of SBMOF-1 which is a very dense system, small errors in the potential well-depths can be compounded due to a large number of host-adsorbate interactions. For this reason we compared the Xe/Kr PES of SBMOF-1 when using UFF versus PBE+D3, and the comparison is captured in Figure C.1 where the UFF energies are plotted versus the DFT energies for 150 adsorbate locations in the void space. Interestingly, UFF over-predicts the strength of the minimum energy binding site of both Xe/Kr by the same factor. In order to better align the UFF PES with PBE+D3, we scale the UFF epsilon parameter of both Xe and Kr by 0.82 such that the minimum binding energies are equivalent between the two methods. After rescaling, the UFF binding energies are better aligned to reproduce the PBE+D3 energies.

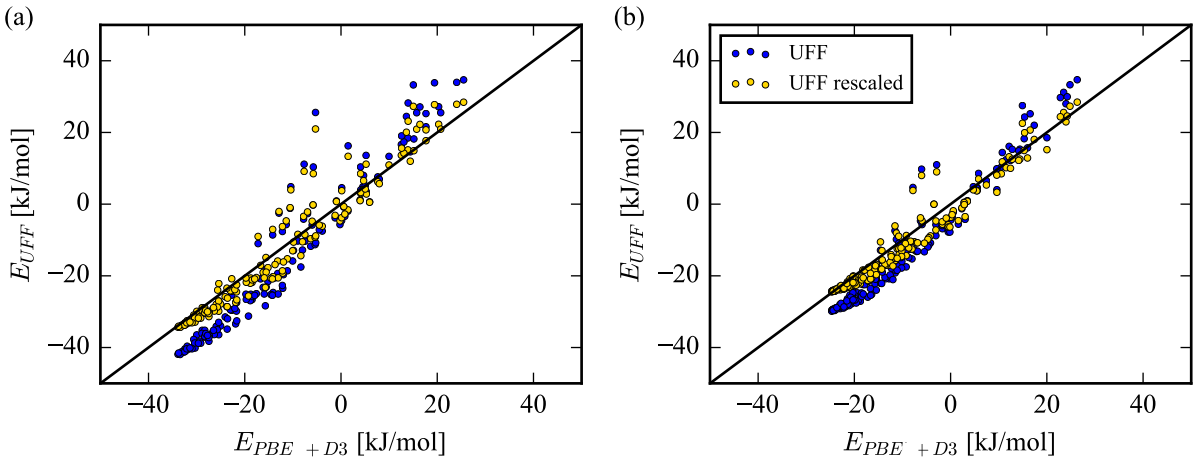


Figure C.1: The PES is shown by probing the void space of SBMOF-1 with 150 Xe adsorbates in (a) and 150 Kr adsorbates in (b). Pure UFF (blue circles) overestimates the binding energies of both adsorbates in comparison to PBE+D3. The UFF binding energy can be universally scaled by a constant factor of 0.82 (yellow circles) to align the minimum binding energies and provide better agreement of the PES between UFF and PBE+D3.

Using the scaled epsilon values of both Xe and Kr, the Henry coefficient calculations were repeated which results in closer agreement of computational and experimental Henry coefficients as would be expected since UFF overestimates the strength of Xe/Kr binding. As has been noted in the literature, no one particular DFT recipe appears to provide consistently better results in describing interactions between MOFs and noble gases.[269] However, this energy re-scaling serves to show that within the approximations of the PBE+D3 method we are able to combine both considerations of flexibility and adjustment of the UFF PES to obtain even better agreement with experiments as shown in Table C.1.

Table C.1: Comparison of Henry coefficients in SBMOF-1 for both AIMD and UFF-CDM generated dynamics with (**) and without (*) epsilon re-scaling of Xe/Kr by 0.82. The rigid structure for UFF-CDM is identical to the DFT minimized structure (and hence the rigid structure for the AIMD entries).

Simulation Type	$K_{H,f}$ [mol/(kg·Pa)]		$K_{H,r}$ [mol/(kg·Pa)]		S_f	S_r
	Xe	Kr	Xe	Kr		
2x2x1 AIMD (*)	6.80e-3	1.77e-4	1.03e-2	2.20e-4	38	47
2x2x1 AIMD (**)	1.52e-3	6.10e-5	2.33e-3	7.70e-5	25	30
3x6x2 UFF-DCM (*)	3.18e-3	1.28e-4	1.03e-2	2.20e-4	25	47
3x6x2 UFF-DCM (**)	7.25e-4	4.30e-5	2.33e-3	7.70e-5	17	30

Bibliography

(1) Flanigen, E. M.; Broach, R. W.; Wilson, S. T. In *Zeolites Ind. Sep. Catal.* Kulprathipanja, S., Ed.; Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, Germany, 2010, pp 1–26.

(2) Zimmermann, N. E. R.; Haranczyk, M. *Cryst. Growth Des.* **2016**, *16*, 3043–3048.

(3) Hoskins, B. F.; Robson, R. *J. Am. Chem. Soc.* **1990**, *112*, 1546–1554.

(4) Chui, S. S. *Science* **1999**, *283*, 1148–1150.

(5) Li, H.; Eddaoudi, M.; O'Keeffe, M.; Yaghi, O. M. *Nature* **1999**, *402*, 276.

(6) Yaghi, O. M.; Li, H. *J. Am. Chem. Soc.* **1995**, *117*, 10401–10402.

(7) Banerjee, R.; Phan, A.; Wang, B.; Knobler, C.; Furukawa, H.; O'Keeffe, M.; Yaghi, O. M. *Science* **2008**, *319*, 939–943.

(8) Cote, A. P.; Benin, A. I.; Ockwig, N.; O'Keeffe, M.; Matzger, A. J.; Yaghi, O. M. *Science* **2005**, *310*, 1166–1170.

(9) Lu, W.; Yuan, D.; Zhao, D.; Schilling, C. I.; Plietzsch, O.; Muller, T.; Bräse, S.; Guenther, J.; Blümel, J.; Krishna, R.; Li, Z.; Zhou, H.-C. *Chem. Mater.* **2010**, *22*, 5964–5972.

(10) Groom, C. R.; Bruno, I. J.; Lightfoot, M. P.; Ward, S. C. *Acta Crystallogr. Sect. B Struct. Sci. Cryst. Eng. Mater.* **2016**, *72*, 171–179.

(11) Baerlocher, C.; McCusker, L. Database of Zeolite Structures.

(12) Zhou, H.-C.; Long, J. R.; Yaghi, O. M. *Chem. Rev.* **2012**, *112*, 673–674.

(13) Yaghi, O. M.; O'Keeffe, M.; Ockwig, N. W.; Chae, H. K.; Eddaoudi, M.; Kim, J. *Nature* **2003**, *423*, 705–714.

(14) Moghadam, P. Z.; Li, A.; Wiggin, S. B.; Tao, A.; Maloney, A. G. P.; Wood, P. A.; Ward, S. C.; Fairen-Jimenez, D. *Chem. Mater.* **2017**, *29*, 2618–2625.

(15) Deem, M. W.; Pophale, R.; Cheeseman, P. A.; Earl, D. J. *J. Phys. Chem. C* **2009**, *113*, 21353–21360.

(16) Pophale, R.; Cheeseman, P. A.; Deem, M. W. *Phys. Chem. Chem. Phys.* **2011**, *13*, 12407.

(17) Mellot-Draznieks, C.; Serre, C.; Surblé, S.; Audebrand, N.; Férey, G. *J. Am. Chem. Soc.* **2005**, *127*, 16273–16278.

(18) Wilmer, C. E.; Leaf, M.; Lee, C. Y.; Farha, O. K.; Hauser, B. G.; Hupp, J. T.; Snurr, R. Q. *Nat. Chem.* **2011**, *4*, 83–89.

(19) Martin, R. L.; Haranczyk, M. *Cryst. Growth Des.* **2014**, *14*, 2431–2440.

(20) Boyd, P. G.; Woo, T. K. *CrystEngComm* **2016**, *18*, 3777–3792.

(21) Witman, M.; Ling, S.; Anderson, S.; Tong, L.; Stylianou, K. C.; Slater, B.; Smit, B.; Haranczyk, M. *Chem. Sci.* **2016**, *7*, 6263–6272.

(22) Mercado, R.; Fu, R.-S.; Yakutovich, A. V.; Talirz, L.; Haranczyk, M.; Smit, B. *Chem. Mater.* **2018**, *30*, 5069–5086.

(23) Lee, Y.; Barthel, S. D.; Dłotko, P.; Moosavi, S. M.; Hess, K.; Smit, B. *Nat. Commun.* **2017**, *8*, 15396.

(24) He, Y.; Zhou, W.; Qian, G.; Chen, B. *Chem. Soc. Rev.* **2014**, *43*, 5657–5678.

(25) Murray, L. J.; Dinca, M.; Long, J. R. *Chem. Soc. Rev.* **2009**, *38*, 1294.

(26) Sumida, K.; Rogow, D. L.; Mason, J. A.; McDonald, T. M.; Bloch, E. D.; Herm, Z. R.; Bae, T.-H.; Long, J. R. *Chem. Rev.* **2012**, *112*, 724–781.

(27) Bloch, E. D.; Queen, W. L.; Krishna, R.; Zadrozny, J. M.; Brown, C. M.; Long, J. R. *Science* **2012**, *335*, 1606–1610.

(28) Tanh Jeazet, H. B.; Staudt, C.; Janiak, C. *Dalt. Trans.* **2012**, *41*, 14003.

(29) Belmabkhout, Y. et al. *J. Mater. Chem. A* **2017**, *5*, 3293–3303.

(30) Kreno, L. E.; Leong, K.; Farha, O. K.; Allendorf, M.; Van Duyne, R. P.; Hupp, J. T. *Chem. Rev.* **2012**, *112*, 1105–1125.

(31) Kim, H.; Yang, S.; Rao, S. R.; Narayanan, S.; Kapustin, E. A.; Furukawa, H.; Umans, A. S.; Yaghi, O. M.; Wang, E. N. *Science* **2017**, *356*, 430–434.

(32) Jamali, S. H.; Vlugt, T. J. H.; Lin, L.-C. *J. Phys. Chem. C* **2017**, *121*, 11273–11280.

(33) Sun, D. T.; Peng, L.; Reeder, W. S.; Moosavi, S. M.; Tiana, D.; Britt, D. K.; Oveisi, E.; Queen, W. L. *ACS Cent. Sci.* **2018**, *4*, 349–356.

(34) Sun, D. T.; Gasilova, N.; Yang, S.; Oveisi, E.; Queen, W. L. *J. Am. Chem. Soc.* **2018**, *140*, 16697–16703.

(35) Graham, M. J.; Zadrozny, J. M.; Fataftah, M. S.; Freedman, D. E. *Chem. Mater.* **2017**, *29*, 1885–1897.

(36) Schneemann, A.; White, J. L.; Kang, S.; Jeong, S.; Wan, L. F.; Cho, E. S.; Heo, T. W.; Prendergast, D.; Urban, J. J.; Wood, B. C.; Allendorf, M. D.; Stavila, V. *Chem. Rev.* **2018**, *118*, 10775–10839.

(37) Rogge, S. M. J.; Bavykina, A.; Hajek, J.; Garcia, H.; Olivos-Suarez, A. I.; Sepúlveda-Escribano, A.; Vimont, A.; Clet, G.; Bazin, P.; Kapteijn, F.; Daturi, M.; Ramos-Fernandez, E. V.; Llabrés i Xamena, F. X.; Van Speybroeck, V.; Gascon, J. *Chem. Soc. Rev.* **2017**, *46*, 3134–3184.

(38) Levine, D. J.; Runčevski, T.; Kapelewski, M. T.; Keitz, B. K.; Oktawiec, J.; Reed, D. A.; Mason, J. A.; Jiang, H. Z. H.; Colwell, K. A.; Legendre, C. M.; FitzGerald, S. A.; Long, J. R. *J. Am. Chem. Soc.* **2016**, *138*, 10143–10150.

(39) Wang, J.-L.; Wang, C.; Lin, W. *ACS Catal.* **2012**, *2*, 2630–2640.

(40) Sheberla, D.; Bachman, J. C.; Elias, J. S.; Sun, C.-J.; Shao-Horn, Y.; Dinca, M. *Nat. Mater.* **2017**, *16*, 220–224.

(41) Dong, C.; Xu, L. *ACS Appl. Mater. Interfaces* **2017**, *9*, 7160–7168.

(42) Boyd, P. G.; Lee, Y.; Smit, B. *Nat. Rev. Mater.* **2017**, *2*, 17037.

(43) 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. *Chem. Mater.* **2014**, *26*, 6185–6192.

(44) O'Keeffe, M.; Peskov, M. A.; Ramsden, S. J.; Yaghi, O. M. *Acc. Chem. Res.* **2008**, *41*, 1782–1789.

(45) Hyde, S.; Delgado Friedrichs, O.; Ramsden, S.; Robins, V. *Solid State Sci.* **2006**, *8*, 740–752.

(46) Addicoat, M. A.; Coupry, D. E.; Heine, T. *J. Phys. Chem. A* **2014**, *118*, 9607–9614.

(47) Martin, R. L.; Lin, L. C.; Jariwala, K.; Smit, B.; Haranczyk, M. *J. Phys. Chem. C* **2013**, *117*, 12159–12167.

(48) Bao, Y.; Martin, R. L.; Simon, C. M.; Haranczyk, M.; Smit, B.; Deem, M. W. *J. Phys. Chem. C* **2015**, *119*, 186–195.

(49) Rosi, N. L.; Kim, J.; Eddaoudi, M.; Chen, B.; Keeffe, M. O.; Yaghi, O. M. *J. Am. Chem. Soc.* **2005**, *127*, 1504–1518.

(50) Mason, J. A.; Sumida, K.; Herm, Z. R.; Krishna, R.; Long, J. R. *Energy Environ. Sci.* **2011**, *4*, 3030–3040.

(51) Bourrelly, S.; Llewellyn, P. L.; Serre, C.; Millange, F.; Loiseau, T.; Férey, G. *J. Am. Chem. Soc.* **2005**, *127*, 13519–13521.

(52) Mason, J. A.; Oktawiec, J.; Taylor, M. K.; Hudson, M. R.; Rodriguez, J.; Bachman, J. E.; Gonzalez, M. I.; Cervellino, A.; Guagliardi, A.; Brown, C. M.; Llewellyn, P. L.; Masciocchi, N.; Long, J. R. *Nature* **2015**, *527*, 357–361.

(53) Deng, H.; Grunder, S.; Cordova, K. E.; Valente, C.; Furukawa, H.; Hmadeh, M.; Gandara, F.; Whalley, A. C.; Liu, Z.; Asahina, S.; Kazumori, H.; O'Keeffe, M.; Terasaki, O.; Stoddart, J. F.; Yaghi, O. M. *Science* **2012**, *336*, 1018–1023.

(54) Kapelewski, M. T.; Geier, S. J.; Hudson, M. R.; Stück, D.; Mason, J. A.; Nelson, J. N.; Xiao, D. J.; Hulvey, Z.; Gilmour, E.; FitzGerald, S. A.; Head-Gordon, M.; Brown, C. M.; Long, J. R. *J. Am. Chem. Soc.* **2014**, *2*, 12119–29.

(55) Wang, L. J.; Deng, H.; Furukawa, H.; Gándara, F.; Cordova, K. E.; Peri, D.; Yaghi, O. M. *Inorg. Chem.* **2014**, *53*, 5881–5883.

(56) McDonald, T. M. et al. *Nature* **2015**, *519*, 303–308.

(57) Kim, S.; Thiessen, P. A.; Bolton, E. E.; Chen, J.; Fu, G.; Gindulyte, A.; Han, L.; He, J.; He, S.; Shoemaker, B. A.; Wang, J.; Yu, B.; Zhang, J.; Bryant, S. H. *Nucleic Acids Res.* **2016**, *44*, D1202–D1213.

(58) Campañá, C.; Mussard, B.; Woo, T. K. *J. Chem. Theory Comput.* **2009**, *5*, 2866–2878.

(59) O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. *J. Cheminform.* **2011**, *3*, 1–14.

(60) Dietzel, P. D. C.; Blom, R.; Fjellvåg, H. *Eur. J. Inorg. Chem.* **2008**, 3624–3632.

(61) O'Boyle, N. M.; Vandermeersch, T.; Flynn, C. J.; Maguire, A. R.; Hutchison, G. R. *J. Cheminform.* **2011**, *3*, 8.

(62) Halgren, T. a. *J Comp. Chem.* **1999**, *20*, 720–729.

(63) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

(64) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. *J. Chem. Phys.* **2010**, *132*, 154104.

(65) Queen, W. L.; Brown, C. M.; Britt, D. K.; Zajdel, P.; Hudson, M. R.; Yaghi, O. M. *J. Phys. Chem. C* **2011**, *115*, 24915–24919.

(66) Ling, S.; Slater, B. *J. Phys. Chem. C* **2015**, *119*, 16667–16677.

(67) Nanthamathee, C.; Ling, S.; Slater, B.; Attfield, M. P. *Chem. Mater.* **2015**, *27*, 85–95.

(68) VandeVondele, J.; Krack, M.; Mohamed, F.; Parrinello, M.; Chassaing, T.; Hutter, J. *Comput. Phys. Commun.* **2005**, *167*, 103–128.

(69) Hutter, J.; Iannuzzi, M.; Schiffmann, F.; VandeVondele, J. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2014**, *4*, 15–25.

(70) Goedecker, S.; Teter, M.; Hutter, J. *Phys. Rev. B* **1996**, *54*, 1703–1710.

(71) Krack, M. *Theor. Chem. Acc.* **2005**, *114*, 145–152.

(72) Golze, D.; Hutter, J.; Iannuzzi, M. *Phys. Chem. Chem. Phys.* **2015**, *17*, 14307–14316.

(73) Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. *J. Am. Chem. Soc.* **1992**, *114*, 10024–10035.

(74) Willems, T. F.; Rycroft, C. H.; Kazi, M.; Meza, J. C.; Haranczyk, M. *Microporous Mesoporous Mater.* **2012**, *149*, 134–141.

(75) Lin, L. C.; Lee, K.; Gagliardi, L.; Neaton, J. B.; Smit, B. *J. Chem. Theory Comput.* **2014**, *10*, 1477–1488.

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

(77) Witman, M.; Ling, S.; Gladysiak, A.; Stylianou, K. C.; Smit, B.; Slater, B.; Haranczyk, M. *J. Phys. Chem. C* **2017**, *121*, 1171–1181.

(78) Cook, T. R.; Zheng, Y. R.; Stang, P. J. *Chem. Rev.* **2013**, *113*, 734–777.

(79) Xue, M.; Liu, Y.; Schaffino, R. M.; Xiang, S.; Zhao, X.; Zhu, G. S.; Qiu, S. L.; Chen, B. *Inorg. Chem.* **2009**, *48*, 4649–4651.

(80) Mueller, U.; Luinstra, G.; Yaghi, O. M. US Pat. 6 617 467., 2004.

(81) Rood, J. A.; Noll, B. C.; Henderson, K. W.; Dame, N. *Inorg. Chem.* **2006**, *45*, 5521–5528.

(82) Alvarez, E. et al. *Angew. Chem. Int. Ed.* **2015**, *54*, 3664–3668.

(83) Lee, K. et al. *J. Am. Chem. Soc.* **2014**, *136*, 698–704.

(84) Runčevski, T.; Kapelewski, M. T.; Torres-Gavosto, R. M.; Tarver, J. D.; Brown, C. M.; Long, J. R. *Chem. Commun.* **2016**, *2*, 8251–8254.

(85) Lee, K.; Howe, J. D.; Lin, L. C.; Smit, B.; Neaton, J. B. *Chem. Mater.* **2015**, *27*, 668–678.

(86) Dzubak, A. L.; Lin, L.-C.; Kim, J.; Swisher, J. A.; Poloni, R.; Maximoff, S. N.; Smit, B.; Gagliardi, L. *Nat. Chem.* **2012**, *4*, 810–816.

(87) Mercado, R.; Vlaisavljevich, B.; Lin, L.-C.; Lee, K.; Lee, Y.; Mason, J. A.; Xiao, D. J.; Gonzalez, M. I.; Kapelewski, M. T.; Neaton, J. B.; Smit, B. *J. Phys. Chem. C* **2016**, *120*, 12590–12604.

(88) Pham, T.; Forrest, K. A.; Mclaughlin, K.; Eckert, J.; Space, B. *J. Phys. Chem. C* **2014**, *118*, 22683–22690.

(89) Adamo, C.; Barone, V. *J. Chem. Phys.* **1999**, *110*, 6158.

(90) Ernzerhof, M.; Scuseria, G. E. *J. Chem. Phys.* **1999**, *110*, 5029–5036.

(91) Poloni, R.; Smit, B.; Neaton, J. B. *J. Phys. Chem. A* **2012**, *116*, 4957–64.

(92) Nazarian, D.; Ganesh, P.; Sholl, D. S. *J. Mater. Chem. A* **2015**, *3*, 22432–22440.

(93) Ling, S.; Slater, B. *Chem. Sci.* **2016**, *7*, 4706–4712.

(94) Guidon, M.; Hutter, J.; Vandevondele, J. *J. Chem. Theory Comput.* **2010**, *6*, 2348–2364.

(95) Dovesi, R.; Orlando, R.; Erba, A.; Zicovich-Wilson, C. M.; Civalleri, B.; Casassa, S.; Maschio, L.; Ferrabone, M.; De La Pierre, M.; D'Arco, P.; Noël, Y.; Causà, M.; Rérat, M.; Kirtman, B. *Int. J. Quantum Chem.* **2014**, *114*, 1287–1317.

(96) Dovesi, R.; Saunders, V. R.; Roetti, C.; Orlando, R.; Zicovich-Wilson, C. M.; Pascale, F.; Civalleri, B.; Doll, K.; Harrison, N. M.; Bush, I. J.; D'Arco, P.; Llunell, M.; Causà, M.; Noël, Y. CRYSTAL14 User's Manual., 2014.

(97) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623–11627.

(98) Mayo, S. L.; Olafson, B. D.; Goddard, W. A. *J. Phys. Chem.* **1990**, *94*, 8897–8909.

(99) Chen, L.; Morrison, C. A.; Düren, T. *J. Phys. Chem. C* **2012**, *116*, 18899–18909.

(100) Wahiduzzaman, M.; Walther, C. F. J.; Heine, T. *J. Chem. Phys.* **2014**, *141*, doi: 10.1063/1.4892670.

(101) Pham, T.; Forrest, K. A.; Banerjee, R.; Orcajo, G.; Eckert, J.; Space, B. *J. Phys. Chem. C* **2015**, *119*, 1078–1090.

(102) Dubbeldam, D.; Calero, S.; Ellis, D. E.; Snurr, R. Q. *Mol. Simul.* **2016**, *42*, 81–101.

(103) Pinheiro, M.; Martin, R. L.; Rycroft, C. H.; Haranczyk, M. *CrystEngComm* **2013**, *15*, 7531–7538.

(104) Hedman, D.; Reza Barzegar, H.; Rosén, A.; Wågberg, T.; Andreas Larsson, J. *Sci. Rep.* **2015**, *5*, 16850.

(105) Born, M. *Math. Proc. Cambridge Philos. Soc.* **1940**, *36*, 160.

(106) Mouhat, F.; Coudert, F.-X. *Phys. Rev. B* **2014**, *90*, 224104.

(107) Dietzel, P. D. C.; Georgiev, P. A.; Eckert, J.; Blom, R.; Strässle, T.; Unruh, T. *Chem. Commun.* **2010**, *46*, 4962–4964.

(108) Queen, W. L. et al. *Chem. Sci.* **2014**, *5*, 4569–4581.

(109) Smit, B.; Reimer, J. R.; Oldenburg, C. M.; Bourg, I. C., *Introduction to Carbon Capture and Sequestration*; Imperial College Press: 2014.

(110) Huck, J. M.; Lin, L.-C.; Berger, A. H.; Shahrak, M. N.; Martin, R. L.; Bhown, A. S.; Haranczyk, M.; Reuter, K.; Smit, B. *Energy Environ. Sci.* **2014**, *7*, 4132–4146.

(111) Joos, L.; Lejaeghere, K.; Huck, J. M.; Van Speybroeck, V.; Smit, B. *Energy Environ. Sci.* **2015**, *8*, 2480–2491.

(112) Witman, M.; Ling, S.; Boyd, P.; Barthel, S.; Haranczyk, M.; Slater, B.; Smit, B. *ACS Cent. Sci.* **2018**, *4*, 235–245.

(113) Bhimanapati, G. R. et al. *ACS Nano* **2015**, *9*, 11509–11539.

(114) Novoselov, K. S. *Science* **2004**, *306*, 666–669.

(115) Novoselov, K. S.; Mishchenko, A.; Carvalho, A.; Castro Neto, A. H. *Science* **2016**, *353*, aac9439.

(116) Brent, R.; Cubillas, P.; Stevens, S. M.; Jelfs, K. E.; Umemura, A.; Gebbie, J. T.; Slater, B.; Terasaki, O.; Holden, M. A.; Anderson, M. W. *J. Am. Chem. Soc.* **2010**, *132*, 13858–13868.

(117) Anderson, M. W.; Gebbie-Rayet, J. T.; Hill, A. R.; Farida, N.; Attfield, M. P.; Cubillas, P.; Blatov, V. A.; Proserpio, D. M.; Akporiaye, D.; Arstad, B.; Gale, J. D. *Nature* **2017**, *544*, 456–459.

(118) Ford Jr, L. R.; Fulkerson, D. R., *Flows in networks*; Princeton university press: 2015.

(119) Hexaresearch Zeolite Molecular Sieve Market Size To Grow Beyond $35 Billion By 2024., 2016.

(120) Earl, D. J.; Deem, M. W. *Ind. Eng. Chem. Res.* **2006**, *45*, 5449–5454.

(121) Lewis, D. W.; Willock, D. J.; Catlow, C. R. A.; Thomas, J. M.; Hutchings, G. J. *Nature* **1996**, *382*, 604–606.

(122) Sastre, G.; Cantin, A.; Diaz-Cabañas, M. J.; Corma, A. *Chem. Mater.* **2005**, *17*, 545–552.

(123) Sartbaeva, A.; Wells, S. a.; Treacy, M. M. J.; Thorpe, M. F. *Nat. Mater.* **2006**, *5*, 962–965.

(124) Li, Y.; Yu, J.; Xu, R. *Angew. Chem. Int. Ed.* **2013**, *52*, 1673–1677.

(125) Schmidt, J. E.; Deem, M. W.; Davis, M. E. *Angew. Chem. Int. Ed.* **2014**, *53*, 8372–8374.

(126) Brand, S. K.; Schmidt, J. E.; Deem, M. W.; Daeyaert, F.; Ma, Y.; Terasaki, O.; Orazov, M.; Davis, M. E. *Proc. Natl. Acad. Sci.* **2017**, *114*, 5101–5106.

(127) Kim, J.; Martin, R. L.; Rübel, O.; Haranczyk, M.; Smit, B. *J. Chem. Theory Comput.* **2012**, *8*, 1684–1693.

(128) Thornton, A. W.; Winkler, D. A.; Liu, M. S.; Haranczyk, M.; Kennedy, D. F. *RSC Adv.* **2015**, *5*, 44361–44370.

(129) Evans, J. D.; Coudert, F. X. *Chem. Mater.* **2017**, *29*, 7833–7839.

(130) Roth, W. J.; Nachtigall, P.; Morris, R. E.; Čejka, J. *Chem. Rev.* **2014**, *114*, 4807–4837.

(131) Eliášová, P.; Opanasenko, M.; Wheatley, P. S.; Shamzhy, M.; Mazur, M.; Nachtigall, P.; Roth, W. J.; Morris, R. E.; Čejka, J. *Chem. Soc. Rev.* **2015**, *44*, 7177–7206.

(132) Xu, L.; Wu, P. *New J. Chem.* **2016**, *40*, 3968–3981.

(133) Roth, W. J.; Gil, B.; Makowski, W.; Marszalek, B.; Eliášová, P. *Chem. Soc. Rev.* **2016**, *45*, 3400–3438.

(134) He, Y.; Nivarthy, G.; Eder, F.; Seshan, K.; Lercher, J. *Microporous Mesoporous Mater.* **1998**, *25*, 207–224.

(135) Choi, M.; Na, K.; Kim, J.; Sakamoto, Y.; Terasaki, O.; Ryoo, R. *Nature* **2009**, *461*, 828–828.

(136) Koekkoek, A. J. J.; Kim, W.; Degirmenci, V.; Xin, H.; Ryoo, R.; Hensen, E. J. M. *J. Catal.* **2013**, *299*, 81–89.

(137) Varoon, K.; Zhang, X.; Elyassi, B.; Brewer, D. D.; Gettel, M.; Kumar, S.; Lee, J. A.; Maheshwari, S.; Mittal, A.; Sung, C.-Y.; Cococcioni, M.; Francis, L. F.; McCormick, A. V.; Mkhoyan, K. A.; Tsapatsis, M. *Science* **2011**, *334*, 72–75.

(138) Schnell, S. K.; Wu, L.; Koekkoek, A. J. J.; Kjelstrup, S.; Hensen, E. J. M.; Vlugt, T. J. H. *J. Phys. Chem. C* **2013**, *117*, 24503–24510.

(139) Zhong, J.-Q.; Wang, M.; Akter, N.; Kestell, J. D.; Boscoboinik, A. M.; Kim, T.; Stacchiola, D. J.; Lu, D.; Boscoboinik, J. A. *Nat. Commun.* **2017**, *8*, 16118.

(140) Jeon, M. Y. et al. *Nature* **2017**, *543*, 690–694.

(141) Mazur, M.; Wheatley, P. S.; Navarro, M.; Roth, W. J.; Položij, M.; Mayoral, A.; Eliášová, P.; Nachtigall, P.; Čejka, J.; Morris, R. E. *Nat. Chem.* **2015**, *8*, 58–62.

(142) Palin, L.; Croce, G.; Viterbo, D.; Milanesio, M. *Chem. Mater.* **2011**, *23*, 4900–4909.

(143) Zhao, Z.; Zhang, W.; Ren, P.; Han, X.; Müller, U.; Yilmaz, B.; Feyen, M.; Gies, H.; Xiao, F.-S.; De Vos, D.; Tatsumi, T.; Bao, X. *Chem. Mater.* **2013**, *25*, 840–847.

(144) Trachta, M.; Bludský, O.; Čejka, J.; Morris, R. E.; Nachtigall, P. *ChemPhysChem* **2014**, *15*, 2972–2976.

(145) Trachta, M.; Nachtigall, P.; Bludský, O. *Catal. Today* **2015**, *243*, 32–38.

(146) Sun, W.; Ceder, G. *Surf. Sci.* **2013**, *617*, 53–59.

(147) Tran, R.; Xu, Z.; Radhakrishnan, B.; Winston, D.; Sun, W.; Persson, K. A.; Ong, S. P. *Sci. Data* **2016**, *3*, 160080.

(148) Balcioglu, A.; Kevin Wood, R. In *Netw. Interdiction Stoch. Integer Program.* Kluwer Academic Publishers: Boston, 2003, pp 21–49.

(149) Crabtree, J. C. Computer Simulation of Carbon Dioxide Adsorption and Transport in Zeolites., Ph.D. Thesis, University of Bath, 2014.

(150) Edmonds, J.; Karp, R. M. *J. ACM* **1972**, *19*, 248–264.

(151) Picard, J.-C.; Queyranne, M. In *Comb. Optim. II*, 1980, pp 8–16.

(152) Stoer, M.; Wagner, F. *J. ACM* **1997**, *44*, 585–591.

(153) Provan, J. S.; Shier, D. R. *Algorithmica* **1996**, *15*, 351–372.

(154) Karger, D. R. *J. ACM* **2000**, *47*, 46–76.

(155) Bruglieri, M.; Maffioli, F.; Ehrgott, M. *Discret. Appl. Math.* **2004**, *137*, 311–341.

(156) Yeh, L.-P.; Wang, B.-F.; Su, H.-H. *Algorithmica* **2010**, *56*, 297–312.

(157) Colbourn, C. J. *Ann. Oper. Res.* **1991**, *33*, 1–15.

(158) Wood, R. *Math. Comput. Model.* **1993**, *17*, 1–18.

(159) D'Auriac, J. C. A.; Preissmann, M.; Leibniz-Imag, A. S. *Math. Comput. Model.* **1997**, *26*, 1–11.

(160) Leighton, T.; Rao, S. *J. ACM* **1999**, *46*, 787–832.

(161) Boykov, Y.; Kolmogorov, V. *IEEE Trans. Pattern Anal. Mach. Intell.* **2004**, *26*, 1124–1137.

(162) Ford, L. R.; Fulkerson, D. R. *Can. J. Math.* **1956**, *8*, 399–404.

(163) Hagberg, A. A.; Schult, D. A.; Swart, P. J. In *Proc. 7th Python Sci. Conf.* Pasadena, CA USA, 2008, pp 11–15.

(164) Vazirani, V. V.; Yannakakis, M. In *Automata*; Chapter 30, 1992; Vol. 623, pp 366–377.

(165) Boyd, P. G.; Moosavi, S. M.; Witman, M.; Smit, B. *J. Phys. Chem. Lett.* **2017**, *8*, 357–363.

(166) Tange, O. *USENIX Mag.* **2011**, *36*, 42–47.

(167) Chen, Y.; Huang, S.; Wang, X.; Zhang, L.; Wu, N.; Liao, F.; Wang, Y. *Cryst. Growth Des.* **2017**, *17*, 5465–5473.

(168) Ikeda, T.; Akiyama, Y.; Oumi, Y.; Kawai, A.; Mizukami, F. *Angew. Chem. Int. Ed.* **2004**, *43*, 4892–4896.

(169) Marler, B.; Camblor, M.; Gies, H. *Microporous Mesoporous Mater.* **2006**, *90*, 87–101.

(170) Wheatley, P. S.; Morris, R. E. *J. Mater. Chem.* **2006**, *16*, 1035.

(171) Roth, W. J.; Kresge, C. T. *Microporous Mesoporous Mater.* **2011**, *144*, 158–161.

(172) Marler, B.; Ströter, N.; Gies, H. *Microporous Mesoporous Mater.* **2005**, *83*, 201–211.

(173) Wang, Y. X.; Gies, H.; Lin, J. H. *Chem. Mater.* **2007**, *19*, 4181–4188.

(174) Moteki, T.; Chaikittisilp, W.; Shimojima, A.; Okubo, T. *J. Am. Chem. Soc.* **2008**, *130*, 15780–15781.

(175) Luo, H. Y.; Michaelis, V. K.; Hodges, S.; Griffin, R. G.; Román-Leshkov, Y. *Chem. Sci.* **2015**, *6*, 6320–6324.

(176) Schreyeck, L.; Caullet, P.; Mougenel, J.; Guth, J.; Marler, B. *Microporous Mater.* **1996**, *6*, 259–271.

(177) Zhang, X.; Liu, D.; Xu, D.; Asahina, S.; Cychosz, K. A.; Agrawal, K. V.; Al Wahedi, Y.; Bhan, A.; Al Hashimi, S.; Terasaki, O.; Thommes, M.; Tsapatsis, M. *Science* **2012**, *336*, 1684–1687.

(178) Slater, B.; Ling, S. *Nat. Mater.* **2017**, *16*, 501–502.

(179) Cliffe, M. J.; Castillo-Martínez, E.; Wu, Y.; Lee, J.; Forse, A. C.; Firth, F. C.; Moghadam, P. Z.; Fairen-Jimenez, D.; Gaultois, M. W.; Hill, J. A.; Magdysyuk, O. V.; Slater, B.; Goodwin, A. L.; Grey, C. P. *J. Am. Chem. Soc.* **2017**, *139*, 5397–5404.

(180) Zhu, Y.; Ciston, J.; Zheng, B.; Miao, X.; Czarnik, C.; Pan, Y.; Sougrat, R.; Lai, Z.; Hsiung, C.-E.; Yao, K.; Pinnau, I.; Pan, M.; Han, Y. *Nat. Mater.* **2017**, *16*, 532–536.

(181) Mounet, N.; Gibertini, M.; Schwaller, P.; Merkys, A.; Castelli, I. E.; Cepellotti, A.; Pizzi, G.; Marzari, N. *Nat. Nanotechnol.* **2018**, *13*, 246–252.

(182) Yu, J.; Xu, R. *Chem. Soc. Rev.* **2006**, *35*, 593.

(183) Frenkel, D.; Smit, B., *Understanding Molecular Simulations*; Elsevier: San Diego, 2002.

(184) Panagiotopoulos, A. Z. *Mol. Phys.* **1987**, *61*, 813–826.

(185) Siepmann, J. I.; Karaborni, S.; Smit, B. *Nature* **1993**, *365*, 330–332.

(186) Snurr, R. Q.; June, R. L.; Bell, A. T.; Theodorou, D. N. *Mol. Simul.* **1991**, *8*, 73–92.

(187) Smit, B. *Mol. Phys.* **1995**, *85*, 153–172.

(188) Rosenbluth, M. N.; Rosenbluth, A. W. *J. Chem. Phys.* **1955**, *23*, 356–359.

(189) Siepmann, J. I.; Frenkel, D. *Mol. Phys.* **1992**, *75*, 59–70.

(190) Frenkel, D.; Mooji, G.; Smit, B. *J. Phys. Condens. Matter* **1992**, *4*, 3053–3076.

(191) Shi, W.; Maginn, E. *J. Chem. Theory Comput.* **2007**, *3*, 1451–1463.

(192) Wang, F.; Landau, D. P. *Phys. Rev. E* **2001**, *64*, 056101.

(193) Wang, J.-s.; Swendsen, R. H. *J. Stat. Phys.* **2002**, *106*, 245–285.

(194) Yan, Q.; Faller, R.; de Pablo, J. J. *J. Chem. Phys.* **2002**, *116*, 8745–8749.

(195) Shell, M. S.; Debenedetti, P. G.; Panagiotopoulos, A. Z. *J. Chem. Phys.* **2003**, *119*, 9406–9411.

(196) Errington, J. R. *J. Chem. Phys.* **2003**, *118*, 9915–9925.

(197) Errington, J. R. *Phys. Rev. E* **2003**, *67*, 012102.

(198) Shen, V. K.; Errington, J. R. *J. Chem. Phys.* **2005**, *122*, DOI: 10.1063/1.1844372.

(199) Ganzenmüller, G.; Camp, P. J. *J. Chem. Phys.* **2007**, *127*, 154504.

(200) Desgranges, C.; Delhommelle, J. *J. Chem. Phys.* **2012**, *136*, 184108.

(201) Maerzke, K. A.; Gai, L.; Cummings, P. T.; McCabe, C. *J. Phys. Conf. Ser.* **2014**, *487*, 012002.

(202) Mahynski, N. A.; Shen, V. K. *J. Chem. Phys.* **2016**, *145*, 174709.

(203) Mahynski, N. A.; Blanco, M. A.; Errington, J. R.; Shen, V. K. *J. Chem. Phys.* **2017**, *146*, 074101.

(204) Mahynski, N. A.; Errington, J. R.; Shen, V. K. *J. Chem. Phys.* **2017**, *147*, 054105.

(205) Mahynski, N. A.; Errington, J. R.; Shen, V. K. *J. Chem. Phys.* **2017**, *147*, 234111.

(206) Errington, J. R.; Shen, V. K. *J. Chem. Phys.* **2005**, *123*, 164103.

(207) Rane, K. S.; Murali, S.; Errington, J. R. *J. Chem. Theory Comput.* **2013**, *9*, 2552–2566.

(208) Wang, F.; Landau, D. P. *Phys. Rev. Lett.* **2001**, *86*, 2050–2053.

(209) Landau, D. P.; Tsai, S.-H.; Exler, M. *Am. J. Phys.* **2004**, *72*, 1294–1302.

(210) Smit, B.; Siepmann, J. I. *J. Phys. Chem.* **1994**, *98*, 8442–8452.

(211) Vlugt, T. J. H.; Krishna, R.; Smit, B. *J. Phys. Chem. B* **1999**, *103*, 1102–1118.

(212) Martin, M. G.; Siepmann, J. I. *J. Phys. Chem. B* **1999**, *103*, 4508–4517.

(213) Chopra, R.; Truskett, T. M.; Errington, J. R. *J. Phys. Chem. B* **2010**, *114*, 16487–16493.

(214) Vlugt, T. J. H.; García-Pérez, E.; Dubbeldam, D.; Ban, S.; Calero, S. *J. Chem. Theory Comput.* **2008**, *4*, 1107–1118.

(215) Myers, A. L. In *Chem. Thermodyn. Ind.* 2004, pp 243–252.

(216) Jiang, J.; Furukawa, H.; Zhang, Y.-B.; Yaghi, O. M. *J. Am. Chem. Soc.* **2016**, *138*, 10244–10251.

(217) Harris, J. G.; Yung, K. H. *J. Phys. Chem.* **1995**, *99*, 12021–12024.

(218) Shell, M. S.; Debenedetti, P. G.; Panagiotopoulos, A. Z. *J. Phys. Chem. B* **2004**, *108*, 19748–19755.

(219) Escobedo, F. A.; Martinez-Veracoechea, F. J. *J. Chem. Phys.* **2008**, *129*, DOI: 10.1063/1.2994717.

(220) Peng, D.-Y.; Robinson, D. B. *Ind. Eng. Chem. Fundam.* **1976**, *15*, 59–64.

(221) Witman, M.; Mahynski, N. A.; Smit, B. *J. Chem. Theory Comput.* **2018**, *14*, 6149–6158.

(222) Paluch, A. S.; Shen, V. K.; Errington, J. R. *Ind. Eng. Chem. Res.* **2008**, *47*, 4533–4541.

(223) Shen, V.; Siderius, D.; Krekelberg, W.; Hatch, H. NIST Standard Reference Simulation Website.

(224) Pizzi, G.; Cepellotti, A.; Sabatini, R.; Marzari, N.; Kozinsky, B. *Comput. Mater. Sci.* **2016**, *111*, 218–230.

(225) Witman, M.; Ling, S.; Jawahery, S.; Boyd, P. G.; Haranczyk, M.; Slater, B.; Smit, B. *J. Am. Chem. Soc.* **2017**, *139*, 5547–5557.

(226) Liu, J.; Thallapally, P. K.; McGrail, B. P.; Brown, D. R.; Liu, J. *Chem. Soc. Rev.* **2012**, *41*, 2308–2322.

(227) Mason, J. A.; Veenstra, M.; Long, J. R. *Chem. Sci.* **2014**, *5*, 32–51.

(228) Ranocchiari, M.; van Bokhoven, J. A. *Phys. Chem. Chem. Phys.* **2011**, *13*, 6388.

(229) Li, J.-R.; Sculley, J.; Zhou, H.-C. *Chem. Rev.* **2012**, *112*, 869–932.

(230) Schneemann, A.; Bon, V.; Schwedler, I.; Senkovska, I.; Kaskel, S.; Fischer, R. A. *Chem. Soc. Rev.* **2014**, *43*, 6062–6096.

(231) Férey, G.; Serre, C. *Chem. Soc. Rev.* **2009**, *38*, 1380.

(232) Bousquet, D.; Coudert, F.-X.; Boutin, A. *J. Chem. Phys.* **2012**, *137*, 044118.

(233) Lin, Z.-J.; Lü, J.; Hong, M.; Cao, R. *Chem. Soc. Rev.* **2014**, *43*, 5867–5895.

(234) Simon, C. M.; Braun, E.; Carraro, C.; Smit, B. *Proc. Natl. Acad. Sci.* **2017**, *114*, E287–E296.

(235) Sakata, Y.; Furukawa, S.; Kondo, M.; Hirai, K.; Horike, N.; Takashima, Y.; Uehara, H.; Louvain, N.; Meilikhov, M.; Tsuruoka, T.; Isoda, S.; Kosaka, W.; Sakata, O.; Kitagawa, S. *Science* **2013**, *339*, 193–196.

(236) Jawahery, S.; Simon, C. M.; Braun, E.; Witman, M.; Tiana, D.; Vlaisavljevich, B.; Smit, B. *Nat. Commun.* **2017**, *8*, 13945.

(237) Maji, T. K.; Matsuda, R.; Kitagawa, S. *Nat. Mater.* **2007**, *6*, 142–148.

(238) Krause, S.; Bon, V.; Senkovska, I.; Stoeck, U.; Wallacher, D.; Többens, D. M.; Zander, S.; Pillai, R. S.; Maurin, G.; Coudert, F.-X.; Kaskel, S. *Nature* **2016**, *532*, 348–352.

(239) Evans, J. D.; Bocquet, L.; Coudert, F.-X. *Chem* **2016**, *1*, 873–886.

(240) Haldoupis, E.; Watanabe, T.; Nair, S.; Sholl, D. S. *ChemPhysChem* **2012**, *13*, 3449–3452.

(241) Chen, L. et al. *Nat. Mater.* **2014**, *13*, 954–960.

(242) Simon, C. M.; Mercado, R.; Schnell, S. K.; Smit, B.; Haranczyk, M. *Chem. Mater.* **2015**, *27*, 4459–4475.

(243) Banerjee, D.; Simon, C. M.; Plonka, A. M.; Motkuri, R. K.; Liu, J.; Chen, X.; Smit, B.; Parise, J. B.; Haranczyk, M.; Thallapally, P. K. *Nat. Commun.* **2016**, *7*, ncomms11831.

(244) Nazarian, D.; Camp, J. S.; Sholl, D. S. *Chem. Mater.* **2016**, *28*, 785–793.

(245) Sarkisov, L.; Martin, R. L.; Haranczyk, M.; Smit, B. *J. Am. Chem. Soc.* **2014**, *136*, 2228–2231.

(246) Cox, B. J.; Thamwattana, N.; Hill, J. M. *Proc. R. Soc. A Math. Phys. Eng. Sci.* **2007**, *463*, 477–494.

(247) Manz, T. A.; Sholl, D. S. *J. Chem. Theory Comput.* **2010**, *6*, 2455–2468.

(248) Bristow, J. K.; Tiana, D.; Walsh, A. *J. Chem. Theory Comput.* **2014**, *10*, 4644–4652.

(249) Addicoat, M. A.; Vankova, N.; Akter, I. F.; Heine, T. *J. Chem. Theory Comput.* **2014**, *10*, 880–891.

(250) Banerjee, D.; Zhang, Z.; Plonka, A. M.; Li, J.; Parise, J. B. *Cryst. Growth Des.* **2012**, *12*, 2162–2165.

(251) Duarte, F.; Bauer, P.; Barrozo, A.; Amrein, B. A.; Purg, M.; Åqvist, J.; Kamerlin, S. C. L. *J. Phys. Chem. B* **2014**, *118*, 4351–4362.

(252) Plimpton, S. *J. Comput. Phys.* **1995**, *117*, 1–19.

(253) VandeVondele, J.; Hutter, J. *J. Chem. Phys.* **2007**, *127*, 114105.

(254) Tran, F.; Hutter, J. *J. Chem. Phys.* **2013**, *138*, 204103.

(255) Boys, S.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553–566.

(256) Bussi, G.; Donadio, D.; Parrinello, M. *J. Chem. Phys.* **2007**, *126*, 014101.

(257) Widom, B. *J. Chem. Phys.* **1963**, *39*, 2808–2812.

(258) Boato, G.; Casanova, G. *Physica* **1961**, *27*, 571–589.

(259) Watanabe, T.; Sholl, D. S. *Langmuir* **2012**, *28*, 14114–14128.

(260) Lawler, K. V.; Hulvey, Z.; Forster, P. M. *Phys. Chem. Chem. Phys.* **2015**, *17*, 18904–18907.

(261) Nazarian, D.; Camp, J. S.; Chung, Y. G.; Snurr, R. Q.; Sholl, D. S. *Chem. Mater.* **2016**, acs.chemmater.6b04226.

(262) Hajjar, R.; Volkringer, C.; Loiseau, T.; Guillou, N.; Marrot, J.; Férey, G.; Margiolaki, I.; Fink, G.; Morais, C.; Taulelle, F. *Chem. Mater.* **2011**, *23*, 39–47.

(263) Zhuang, G.-l.; Chen, W.-x.; Zeng, G.-n.; Wang, J.-g.; Chen, W.-l. *CrystEngComm* **2012**, *14*, 679–683.

(264) Gándara, F.; Andrés, A. D.; Gómez-Lor, B.; Gutiérrez-Puebla, E.; Iglesias, M.; Monge, M. A.; Proserpio, D. M.; Snejko, N. *Cryst. Growth Des.* **2008**, *8*, 378–380.

(265) Miller, S. R.; Horcajada, P.; Serre, C. *CrystEngComm* **2011**, *13*, 1894–1898.

(266) Spencer, J.; Alavi, A. *Phys. Rev. B* **2008**, *77*, 193110.

(267) Peng, X.; Lin, L.-C.; Sun, W.; Smit, B. *AIChE J.* **2015**, *61*, 677–687.

(268) Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O. *J. Comput. Chem.* **2011**, *32*, 2319–2327.

(269) Demir, H.; Greathouse, J. A.; Staiger, C. L.; Perry IV, J. J.; Allendorf, M. D.; Sholl, D. S. *J. Mater. Chem. A* **2015**, *3*, 23539–23548.