Computational Simulations to Aid in the Experimental Discovery of Ice Recrystallization Inhibitors and Ultra-Microporous Metal Organic Frameworks

Phil De Luna

Thesis submitted to the
Faculty of Graduate and Postdoctoral Studies
In partial fulfillment of the requirements for the
Masters of Science degree in Chemistry

Department of Chemistry and Biomolecular Sciences
Faculty of Science
University of Ottawa

Declaration of Co-Authorship

I, Phil De Luna, hereby declare that this thesis incorporates material that is the product of cooperative research as follows,

Chapter 3 is in collaboration with Dr. Michael Fernandez, Jennie Briard, Nick Trefiak, Dr. Robert Ben, and Dr. Tom Woo. Michael was responsible for the development of the QSAR model as well as provided me with assistance and mentorship. Jennie Briard was my main liaison with the Ben group and provided much of the experimental work as well as gathering and cataloguing previous experimental data. Nick Trefiak wrote the descriptor generation code which I used for this project. Dr. Ben and Dr. Woo were the supervisors of the project and provided guidance.

Chapter 4 is in collaboration with Shyamapada Nandi, Dr. Ramanathan Vaidhyanathan, Dr. Thomas Daff, and Dr. Tom Woo. Shyamapada was responsible for all experimental synthesis and characterization of the MOFs, Dr. Vaidhyanathan provided guidance and supervision, Dr. Thomas Daff was my direct mentor and wrote the program used for adsorption analysis, and Dr. Tom Woo provided guidance and supervision.

I am aware of the University of Ottawa's Faculty of Graduate and Postdoctoral Studies policy on authorship and I certify that I have properly acknowledged the contribution of other researchers to my thesis. I certify that, with the above qualification, this thesis, and the research to which it referees, is the product of my own work.

Acknowledgements

This section of the thesis is typically where the writer gets to bare their soul and show some earnest emotion, which is a stark contrast to the exact preciseness of a scientific document such as a thesis. I have spent the last two years of my life pursuing a Masters degree and this experience has truly been a transformative one for me. Many of the lessons I have learned were beyond that of an academic nature.

The first of these lessons was, *do not let your expectations define your experiences.* I began with eager naiveté about what graduate school would be like. These expectations were quickly shattered with the reality of working on projects and problems with no discernable solution. There were many projects and ideas which I had spent much effort and time on that ended up unceremoniously drifting in the wind. It took a long time for me to equate these "failures" into "lessons". This leads me to the second lesson, *it is not that you failed; it is what you have learn from your failure.* The work you are about to read is merely a fraction of what I have done over the past two years. Much of the simulations provided more questions than they did answers, some outright failed in their own right, but every single one provided insight into the problem. Learning how to find that insight is a skill that I will never let go. My last lesson is, *everyone has different strengths and weaknesses and success cannot be had in solitary.* I have had the immense fortune to have met many intelligent and creative people in the Woo Lab. All of them have unique skills and experiences which greatly compliment the other. I have learned a level of teamwork and collaboration that I have not seen before. For these lessons I am forever thankful.

There are a countless number of people who I wish to acknowledge for supporting me throughout the past two years. A proper acknowledgement would span several volumes and condensing this was a huge challenge but I would like to acknowledge three sets of people, my family, my friends, and my research group.

First, I would like to thank my parents Joy and Art. The amount of sacrifice they endured to provide a better life for me and my sister is something that I constantly remind myself. Moving across the world with no money and no friends for just the *chance* at a better life for your kids is one of the most admirable and selfless acts I know. Next, I would like to acknowledge my little sister, Liana. Being, 11 years apart in age is a difficult thing when trying to relate. I only wish that I was around more for her formative years, every time I visit home I see a new person, a better person, a smart young girl with her own interests, passions, and opinions. I hope that I can provide an example for her, but really I know that she has the potential to surpass me in every facet of life and honestly I am excited to see her do it.

Secondly, I would like to thank my friends and squad of the Osgoode Barn. To Schrambo, Patrice, Swayze, Zona, Homie Tom, Wafleek, G-Eazy, D-Money, and Yvonne, thank you for being there for me and always supporting the YOLO. I will never forget the concerts we put on, the charity gala we organized, and endless nights of porch beers. To my girlfriend Danielle, thank you for keeping me sane and reminding me that there is at least some thing redeemable in me to warrant having you by my side.

Thirdly, I must acknowledge my research group for all their support. To Bianca, Sarah, Moshtagh, Daff, Michael, Carlos, Pete, Mo, Sean, Jason, Burns, Hana, and Chris. I am so grateful to have had the opportunity to work with you every day. I consider each of you not only a

colleague but also a friend. You have all taught me so much and I will fondly remember the laughter, conversation, debates, and memories which we shared within Marion 017. Lastly, I would like to acknowledge my supervisor Dr. Tom Woo. I have nothing but respect and admiration for him as a scientist. Being able to switch fields and topic of research mid-career is a difficult thing, to do it and produce the calibre of research that Tom has is astounding. But perhaps most importantly I would like to thank Tom for his (sometimes brutal) honesty with me. He once told me that he did not believe me fit to continue in this field. To be honest that comment shook me very deeply. I had spent a lot of time questioning my role in academia, my abilities in myself, and my worth as a scientist. Then one day I realized that I absolutely love research. I love what I do and I will be spending the rest of my life pursuing my passion. So I would like to thank Tom for his words, they have only renewed my motivation and determination in this field.

Abstract

In this thesis computational chemistry has been used to accelerate experimental discovery in the fields of ice recrystallization inhibitors for cryopreservation and ultra-microporous MOFs for carbon dioxide capture and storage.

Ice recrystallization is one of the leading contributors to cell damage and death during the freezing process. This occurs when larger ice crystal grains grow at the expense of smaller ones. Naturally occurring biological antifreeze molecules have been discovered but only operate up to -4°C and actually exasperate the problem at temperatures lower than this. Recently, the group of Dr. Robert Ben have been successful in synthesizing small organic molecules which are capable of inhibiting the growth of ice crystals during the freezing process. They have built a library of diverse compounds with varying functionalities and activity. Chemical intuition has been unsuccessful in finding a discernable trend with which to predict activity. Herein we present work where we have utilized a quantitative structure activity relationship (QSAR) model to predict whether a molecule is active or inactive. This was built from a database of 124 structures and was found to have a positive find rate of 82%. Proposed molecules that had yet to be synthesized were predicted to active or inactive using our method and 9/11 structures were indeed active which is strikingly consistent to the 82% find rate. Our efforts to aid in the discovery of these novel molecules will be described here.

Metal organic frameworks (MOFs) are a relatively new class of porous materials which have taken the academic community by storm. These three-dimensional crystalline materials are built from a metal node and an organic linker. Depending on the metals and organic linkers

used, MOFs can possess a vast range of topologies and properties that can be exploited for specific applications. Ultra-microporous MOFs possess relatively small pores in the range of 3.5 Å to 6 Å in diameter. These MOFs have some structural advantages compared to larger pored MOFs such as molecular sieving, smaller pores which promote strong framework-gas interactions and cooperative effects between guests, and longer shelf-life due to small void volumes and rigid frameworks. Here we present newly synthesized ultra-microporous MOFs based on isonicotnic acid as the organic linker with Ni and Mg as the metal centre. Despite having such small pores, Ni-4PyC exhibits exceptionally high CO_2 uptake at high pressures. Furthermore, Mg-4PyC exhibits novel pressure dependent gate-opening behaviour. Computational simulations were employed to investigate the origin of high CO_2 uptake, predict high pressure (>10bar) isotherms, quantify CO_2 binding site positions and energies, and study uptake-dependent linker dynamics. This work hopes to provide experimentalists with some explanation to these interesting unexplained phenomena and also predict optimal properties for new applications.

Table of Contents

List of Figures

Chapter 3: 3D-QSAR Accelerated Discovery of Ice Recrystallization Inhibitors

Chapter 4: Simulation of Novel Ultra-Microporous MOFs

Chapter 5: Conclusions

List of Tables

List of Abbreviations

- **AAAS** American Association for the Advancement of Science
- **ABSL** Automated Binding Site Locator
- **AFP** Antifreeze Proteins
- **AM1** Austin Model 1
- **AUC** Area Under Curve
- **BA** Biological Antifreeze
- **BDC** benzene-dicarboxylic acid
- **BET** Brunauer-Emmet-Teller
- **CDCC** Cambridge Crystallographic Data Centre
- **CIF** Crystallographic Information File
- **CoMFA** Comparative Molecular Field Analysis
- **CSS** Carbon Sequestration and Storage
- **DFT** Density Functional Theory
- **DHBC** Dihydroxybenzoic acid
- **DIPY** Dipyridine
- **DMA** Dimethylacetamide
- **DMF** Dimethylformamide

- **DOBPDC** 2,5-dioxidobenzene-1,4,dicarboxylate
- **ESP** Electrostatic Potential
- **FP** False Positives
- **GA** Genetic Algorithm
- **GCMC** Grand Canonical Monte Carlo
- **GFA** Genetic Function Approximation
- **GFP** Green Fluorescent Protein
- **GRIND** Grid-Independent Descriptors
- **HECA** Hydrogen Energy California Project
- **HF** Hartree-Fock
- **HOMO** Highest Occupied Molecular Orbital
- **IGCC** Integrated Gasification Combined Cycle
- **IISER** Indian Institute of Science, Education, and Research
- **IRI** Ice Recrystallization Inhibitor
- **L-J** Lennard-Jones
- **LOO** Leave-one-out
- **LUMO** Lowest Unoccupied Molecular Orbital
- **MACC** Maximum Auto- Cross- Correlation
- **MC** Monte Carlo
- **MD** Molecular Dynamics
- **MGS** Mean Grain Size
- **MLR** Multilinear Regression
- **MM** Molecular Mechanics
- **MMEN** N,N'-dimethylethylenediamine
- **MMFF** Merck Molecular ForceField
- **MNDO** Modified Neglect of Diatomic Overlap
- **MOF** Metal Organic Framework
- **MSD** Mean-squared Displacement
- **NMP** N-Methylpyrrolidone
- **NMR** Nuclear Magnetic Resonance
- **PAW** Plane Augmented Wave
- **PBE** Perdew-Burke-Ernzerhof exchange-correlation functional
- **PBS** Phosphate Buffer Solution
- **PES** Potential Energy Surface
- **PLSR** Partial-Least Squares Regression
- **PM3** Parametric Method 3
- **PSA** Pressure Swing Adsorption
- **QLL** Quasi-Liquid Layer

- **QSAR** Quantitative Structure Activity Relationship
- **QSPR** Quantitative Structure Property Relationship
- **RBCs** Red Blood Cells
- **REPEAT** Repeating Electrostatic Potential Extracted Atomic
- **SBU** Secondary Building Unit
- **SIESTA** Spanish Initiative for Electronic Simulations with Thousands of Atoms
- **TCEP** Texas Clean Energy Project
- **TGA** Thermogravimetric Analysis
- **TN** True Negatives
- **TP** True Positives
- **TSA** Temperature Swing Adsorption
- **UFF** Universal Force Field
- **VASP** Vienna Ab Initio Software Package
- **vdW** van der Waals
- **VSA** Vacuum Swing Adsorption
- **ZIFs** Zeolitic Imidazolate Frameworks

1. Introduction and Thesis Goals

Computational chemistry is a sub discipline of chemistry which uses theory and computer simulation to solve chemical problems. While this field has made dramatic leaps and bounds since its inception midway through the last century, it is still far from explaining most chemical phenomena independent of experimental evidence. However, it has also lead to discoveries and explanations which would not have been possible with experiment alone. Indeed, computational simulation and experiment currently exist in balanced equilibria, each providing a role in the pursuit of scientific discovery. My thesis consists of two distinct parts, each with projects concerning different chemical topics, but the theme of computational simulation aiding experiment remains consistent in both. The first part is the use of simulation to aid in the screening and discovery of small organic ice recrystallization inhibitor (IRI) molecules. The second part involves probing unique features of newly synthesized ultra-microporous metal-organic frameworks (MOFs) with respect to carbon dioxide adsorption. My thesis hopes to display the utility of computational chemistry in application to two important fields, medicinal cryogenics and materials for carbon capture and sequestration.

1.1 Ice Recrystallization Inhibitors

The preservation of biological materials such as blood and organ tissues is an important issue in cryogenics. Better preservation of these materials would result in longer shelf-life for blood, greater storage and transport capacities, and would help alleviate the shortage of blood necessary for transfusions. The largest contributor to cell death during the freezing process is

ice recrystallization. This section will discuss this phenomenon, how nature has evolved natural antifreeze proteins, and how organic chemists have designed small molecules which exhibit ice recrystallization inhibition activity.

1.1.1 Ice Recrystallization

Ice recrystallization is a phenomenon during the freezing process where larger ice crystal grains grow at the expense of smaller ice crystals grains.[1] Depending on the temperature and pressure, water molecules can arrange in different ways giving rise to different ice polymorphs. The hydrogen bonding of water determines the properties and phase of ice as well as the axis of growth. The most common naturally found form of ice below 0° C and at atmospheric pressure is a hexagonal lattice crystal form called ice I_h. (Figure 1.1) At atmospheric pressure and 0°C ice grows fastest laterally along the hexagonal plane giving rise to sheets of ice crystals.[2,3]

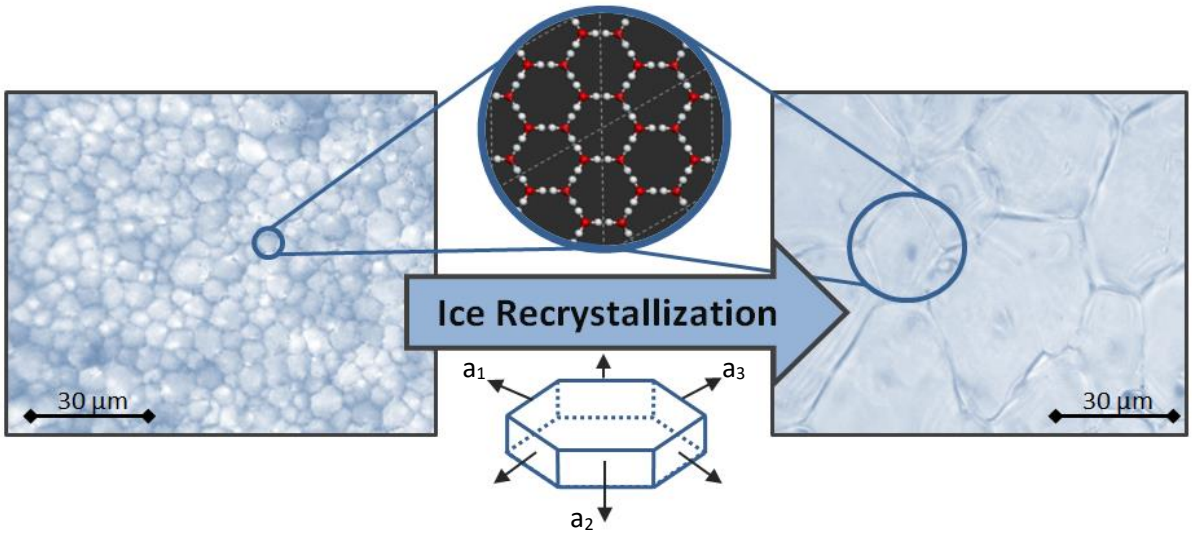


Figure 1.1: Schematic representation of ice recrystallization. Hexagonally shaped smaller ice crystals (left) in comparison to larger ice crystals (right) at the same scale. Shown in the center is the hexagonal crystal structure of ice I_h. An illustration of the lattice unit and the direction of ice growth along the a-axis are also shown.

There are two explanations for ice recrystallization, boundary grain migration which involves the movement of water molecules from one crystal to another, and Ostwald ripening which is a thermodynamic process at the interface of crystalline ice and bulk liquid water.

In crystalline ice, a grain boundary can be considered the point at which differently oriented ice crystal grains meet.[4] The boundaries of ice grains tend to be curved where smaller ice grains are convex with higher surface energy and larger ice grains are concave with lower surface energy. Grain boundary migration is the movement of individual water molecules from the higher energy ice grains (smaller, convex boundaries) to lower energy ice grains (larger, concave boundaries). This growth of larger ice crystals at the expense of smaller ones gives an overall reduction in energy of the system.[4,5] This process concerns itself only with the solid phase which contains the distinct ice grains, but does not account for the aqueous environment around ice during freezing.

Ostwald ripening is a thermodynamically driven process where larger ice crystals grow in favour of smaller ice crystals due to interaction energies of the ice crystal/bulk-water interface. During the formation of ice in aqueous solution a semi-ordered layer exists between the highly ordered ice lattice and disordered bulk water called the quasi-liquid layer (QLL).[6,7] This layer has been well studied experimentally using a wide variety of methods where the thickness has been found to be temperature dependent.[8–12] In addition to experimental work, computational insights on the structure and dynamics of the QLL was provided by Karim at al. via molecular dynamics simulations.[13] Water molecules at the surface of the ice at the QLL will be energetically less stable compared to the ordered water molecules within the crystal. Thus, small ice crystals with a smaller surface area are thermodynamically less favoured, due to the

greater presence of less stable water molecules per volume, than larger ice crystals. During the Ostwald ripening process individual water molecules move from the surface of smaller ice crystals, through the QLL into the bulk water, and then onto the surface of larger ice crystals which results in a reduction in the free energy of the system.[5,14,15] In contrast to grain boundary migration, Ostwald ripening accounts for the different phases of water but evidence points to both processes being present during ice recrystallization.

1.1.2 Consequences of Ice Recrystallization

Ice recrystallization has the greatest impact in the fields of frozen foods and cryopreservation of biological material. In terms of the food industry, much research has been done to examine the effects of freezing on food in hopes of retaining freshness and providing longer shelf life.[16] One such well-studied example is that of ice recrystallization in ice cream which include mechanistic studies[17], experimental characterization[18], and even computer simulations.[19] The size of the ice crystal has a significant control over the texture and taste of ice cream where larger ice crystal grains lead to have a more coarse texture and overall reduction in quality.[20] Despite an entire field of research dedicated to ice cream research, ice recrystallization has yet to be observed in situ and a thorough molecular level understanding is still far from complete.

In terms of the medical field, ice recrystallization is a leading contributor to cellular damage and death during cryostorage.[21] Cryopreservation is a process where cells, tissues, or other biological materials are cooled (~ -190^{o}C) in order to halt any chemical or enzymatic activity which might lead to degradation or damage. During the freezing process there are

usually three mechanisms of cell death, cell rupture due to external cell membrane damage, necrosis, and cold induced apoptosis. Some contributing stress factors include ionic imbalances around the cell membrane, membrane phase transitions and alterations, free radical production and accumulation, water solidification, cell volume excursions, hyperosmolarity, and protein denaturation.[22] Dehydrating the cell during the freezing process will reduce the presence of lethal intracellular ice. However, extreme dehydration also shrinks the cell and increases the concentration of electrolytes within the cell which is also lethal. A key factor in cell death are the cooling and warming rates.[23] It has been shown that too slow of a cooling rate leads to imbalances between extracellular and intracellular solutions and too fast of a cooling rate results in the rapid formation of intracellular ice and thus ice recrystallization upon warming.[23] Addition of cryoprotectants such as dimethyl sulfoxide and glycerol decrease the concentration of electrolytes and preserve cell volumes to avoid shrinking.[24] These cryoprotectants penetrate through the cell membrane and displace intracellular water but are toxic due to disruption of intracellular signaling[25a] and need to be removed from the sample when it is thawed.[25b] This is a time consuming and costly process especially in the context of blood cell storage where quick thawing is essential in an emergency setting. Thus, drastic efforts have been made to develop non-toxic small molecule ice recrystallization inhibitors to reduce and replace toxic cryoprotectants.

1.1.3 Biological Antifreeze Proteins and Thermal Hysteresis

Nature has evolved its own set of biological antifreeze (BA) proteins found in a variety of fish, insects, plants, and bacteria which have been able to survive at sub-zero temperatures.[26] These BAs are a complicated class of structurally diverse compounds which make native ice

recrystallization inhibition difficult to understand and mimic. There have been two proposed antifreeze activities, thermal hysteresis and ice recrystallization inhibition. In terms of BAs, the more researched of these activities is thermal hysteresis.

Thermal hysteresis is a phenomenon where the freezing point of a solution is selectively depressed compared to the melting point. That is, the solution will experience an inhibition of ice growth despite cooling to temperatures below the bulk melting point.[27] This is accomplished by BAs irreversibly binding to the surface of the ice crystal. The binding of BA's is favoured on the sides of the ice crystal rather than on the top or bottom which inhibits lateral ice sheet growth along the *a-axis*. (Figure 1.2)

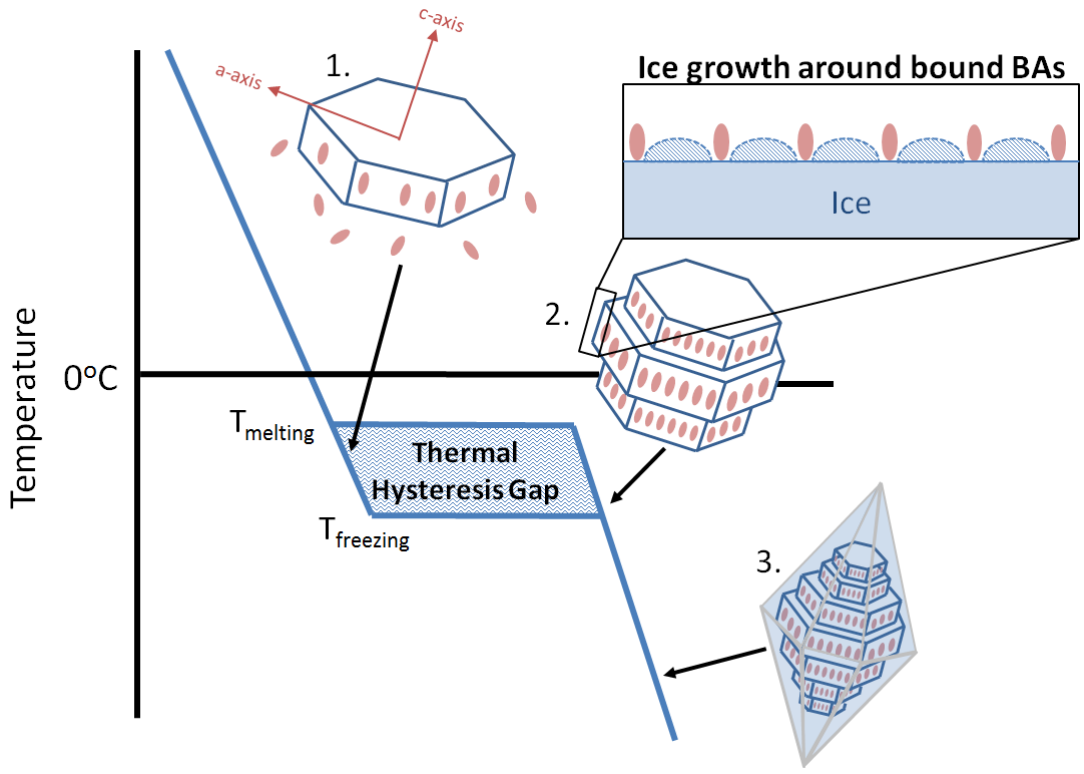


Figure 1.2: Graphical representation of thermal hysteresis and its effect on crystal growth.(1) Biological antifreezes (represented as red oblong shapes) begin to preferentially bind to the side of the crystal (shown as blue hexagonal shapes). (2) As the BAs bind, crystal growth along the a-axis is halted and thus crystal growth begins on the c-axis above and below the crystal. The ice grows in between adjacent BAs on the surface resulting in a curved crystal surface on the side. (3) Far below the thermal hysteresis gap the ice crystals form sharp spicules.

As the temperature decreases bellow the hysteresis freezing point, ice begins to grow at the top and bottom of the ice crystal creating sharp spicule shapes.[28,29] Thus, below the thermal hysteresis gap, the characteristic shape of the ice crystal is deformed causing more elongated and sharper needle-like crystals further exacerbating damage to cells.[30,31] Since cryogenic applications typically operate at temperatures far below the thermal hysteresis gap (-80 ^{o}C to -190 ^{o}C) BAs with high thermal hysteresis activity are unsuitable for use in medical applications. Thus, while BA are non-toxic and can inhibit the growth of ice crystals, synthetic molecules which can inhibit ice recrystallization without irreversibly binding to the ice surface are imperative.

1.1.4 Ice Recrystallization Inhibition

Ice recrystallization inhibition is the activity of BAs to inhibit ice growth without binding directly to the ice crystal surface. While there have been proposed mechanisms for how this occurs (which will be presented in the Section 1.1.5) it is still unclear as to how some molecules exhibit IRI activity while others do not.

The most common method for experimentally assessing IRI activity is called the splat-cooling assay.[32] The basis of this technique examines the differences in size of individual ice grains. First, the sample solution is dropped onto a -80 ^{o}C precooled aluminum block at a height of 2 meters such that it freezes as a thin circular wafer.[32] Alternatively, the solution can be sandwiched between two coverslips and then frozen,[33] but the former is more common. Next, the samples are annealed at a temperature below 0 ^{o}C and the ice crystal size distribution is measured using Domain Recognition Software.[34] The ice crystal size is quantified as the mean

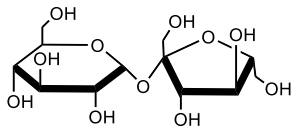
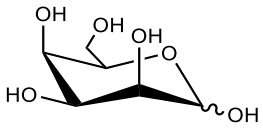
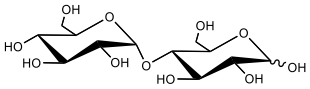
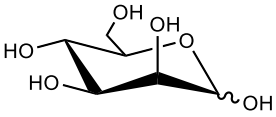
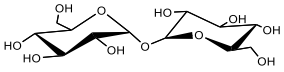
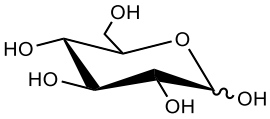
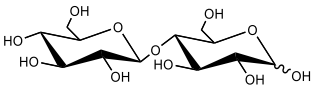
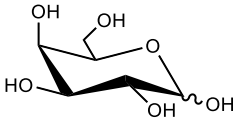
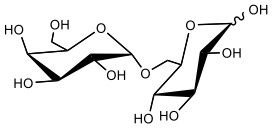
largest ice grain dimension along any axis[35] or by measuring the mean ice grain area.[15,34] This is a common procedure which allows for quantitative comparisons between samples of various structure.

There has been a wealth of research in developing protein analogues[36], peptide[37], glycopeptide[38], and synthetic polymers[39] as IRI active molecules. For further information, a great review on this subject has been published by the Ben Lab.[40] However, with regards to the scope of this thesis only small molecule ice recrystallization inhibitors will be discussed.

1.1.5 Small Molecule IRI Activity

The use of small molecules for ice recrystallization which are not peptide or polymer based has only recently been discovered by the pioneering work of the Ben Lab. The first reported study of IRI active small molecules was the testing of carbohydrates for IRI activity by Tam *et al.*[41] The use of carbohydrates for IRI activity was inspired by previous work where a more hydrated carbohydrate moiety in of a synthetic BA analogue was found to play a significant role in increased IRI activity.[42] Here a set of four monosaccharides and five disaccharides of varying hydration number were tested for their IRI activity. Their hydration numbers were calculated from density and ultrasound experiments.[43] (Table 1.1) This study found a strong correlation between hydration number, which is the number of tightly bound water molecules within the first hydration shell, and IRI activity. The optimal concentration of IRI molecules was found to be 22mM which was low enough that any negative effects due to high viscosity was not an issue, but high enough to retain sufficient IRI activity.

Table 1.1: Hydration numbers and percent mean ice grain size relative to a PBS standard for various monosaccharides and disaccharides. Notice that as hydration number increases, IRI activity improves (lower %MGS).

Disaccharides	Hydration Number	% MGS	Monosaccharides	Hydration Number	% MGS
 Sucrose	13.9 ± 0.3	89	 D-Talose	7.7 ± 0.2	88
 Maltose	14.5 ± 0.3	85	 D-Mannose	8.1 ± 0.2	85
 Trehalose	15.3 ± 0.3	80	 D-Glucose	8.4 ± 0.2	75
 Lactose	15.3 ± 0.3	72	 D-Galactose	8.7 ± 0.2	64
 Melibiose	15.5 ± 0.3	64			

In both carbohydrate classes of monosaccharide and disaccharide, the molecules with a greater number of tightly bound water molecules showed moderate IRI activity. Tam *et al*. have proposed that carbohydrates with high hydration numbers, and thus greater interaction with water molecules, disrupt the transfer of bulk water to the QLL resulting in a disruption of ice recrystallization.[41] It was postulated that the carbohydrate molecules aggregated at the interface of the QLL and bulk water. Molecules with a higher hydration number will tightly bind to more water molecules which disrupts the hydrogen-bonding framework of bulk water. This disorder in the bulk water results in an energetic increase associated with the transfer of water

from the bulk to the QLL. Therefore, it was proposed molecules with greater hydration numbers exhibited ice recrystallization inhibition by disrupting the transfer of water from the bulk to the QLL limiting the growth of the ice crystal.

To further test the hypothesis of hydration shell, the Ben lab continued by developing carbohydrate derivatives of D-galactose which was the native monosaccharide from the work of Tam *et al.* that presented the best IRI activity.[41,44] These derivatives had differing stereochemistry and hydroxyl group substitutions about the sugar ring. However, it was found that all derivatives had poor to weak IRI activity and none were as active as native D-galactose.

In 2012 another more successful study of small IRI active molecules by the Ben Lab was published on the use of carbohydrate-based non-ionic surfactants and hydrogelators.[45] (Figure 1.3) Surfactants are amphiphilic molecules which contain hydrophilic and hydrophobic portions simultaneously and hydrogelators are molecules which aggregate in aqueous solution sequestering bulk-water to form hydrogels.

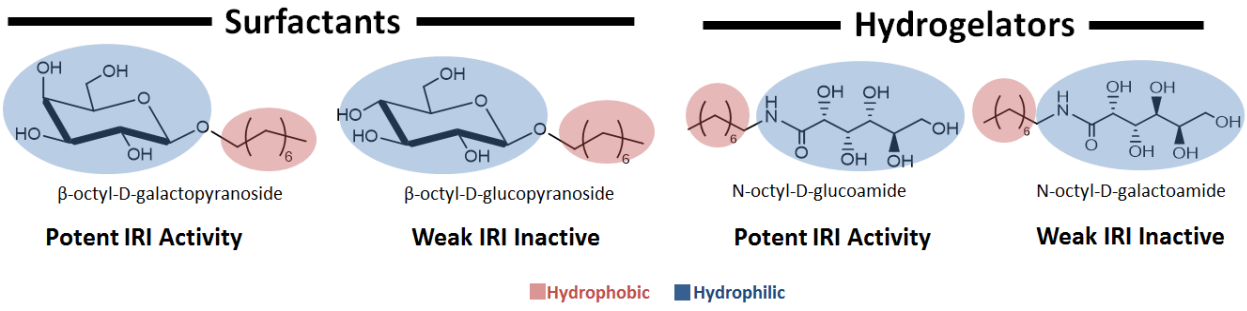


Figure 1.3: Chemical structures of carbohydrate-based small IRI molecules. (left) Closed ring β-octyl-D-galactopyranoside and β-octyl-D-glucopyranoside surfactants and (right) open chain N-octyl-D-glucoamide and N-octyl-D-galactoamide hydrogelators are shown with their hydrophobic aryl chain portions highlighted in red and the hydrophilic carbohydrate portions highlighted in blue.

The surfactants were based off of native carbohydrate molecules that have been functionalized with long alkyl chains. The carbohydrate portion is quite hydrophilic with many

hydroxyl groups while the substituted long alkyl chain is hydrophobic. It was found that IRI activity was very sensitive where just a change in the equatorial or axial position of a hydroxyl group affected IRI activity immensely. For example, it was found that β-octyl-D-galactopyranoside was highly IRI active even at concentrations as low as 11 mM. But β-octyl-D-glucopyranoside was only weakly IRI active at high concentrations of 44 mM. The only difference between these two molecules is the equatorial vs. axial position of the hydroxyl group on C5. As for the hydrogelators, it was found that *N*-octyl-D-glucoamide was a potent IRI active molecule even at 0.5 mM which was the lowest active concentration of these sets of carbohydrates. Conversely, it was found that *N*-octyl-D-galactoamide had weak IRI activity. This was an interesting case where in the closed ring form of carbohydrates the galactose derivative was a potent IRI active molecule but in the open chain form the glucose derivative was the IRI potent molecule.[45] In all of the molecules studied it was found that there was no thermal hysteresis activity which indicated that ice binding was not necessary to halt ice growth. This was validated by solid-state NMR measurements as well as thermal hysteresis experiments.[45] IRI active small molecules have since been a large topic of research with development in lysine-based surfactants and hydrogelators[46] and truncated *C*-linked glycopeptides.[47]

1.1.6 Small Molecule IRIs in Red Blood Cells

Very recently, the Ben Lab has been able to test and confirm potent IRI activity of phenolic-glucosides on human red blood cells (RBCs) for the first time.[48] As briefly mentioned in Section 1.1.2, high concentrations of the cryoprotectant glycerol (~40%) are typically added to RBCs as the standard practice in clinical cryopreservation. Furthermore a slow freezing rate of 1 ^{o}C/min to -80 ^{o}C is required to ensure the survival of RBCs post-thaw. [49] Before a transfusion

can occur the glycerol must be removed from the plasma to levels where intracellular glycerol concentrations are less than 1% or intravascular hemolysis can occur.[50] This time-consuming and costly process is a huge hindrance in providing emergency care. The phenolic-glucosides presented by Capicciotti *et al.* showed the utility of small molecule IRIs in freezing of RBCs with drastically decreased concentration of glycerol. Structures and IRI activities of the para-substituted phenyl-glucosides are shown in Figure 1.4.

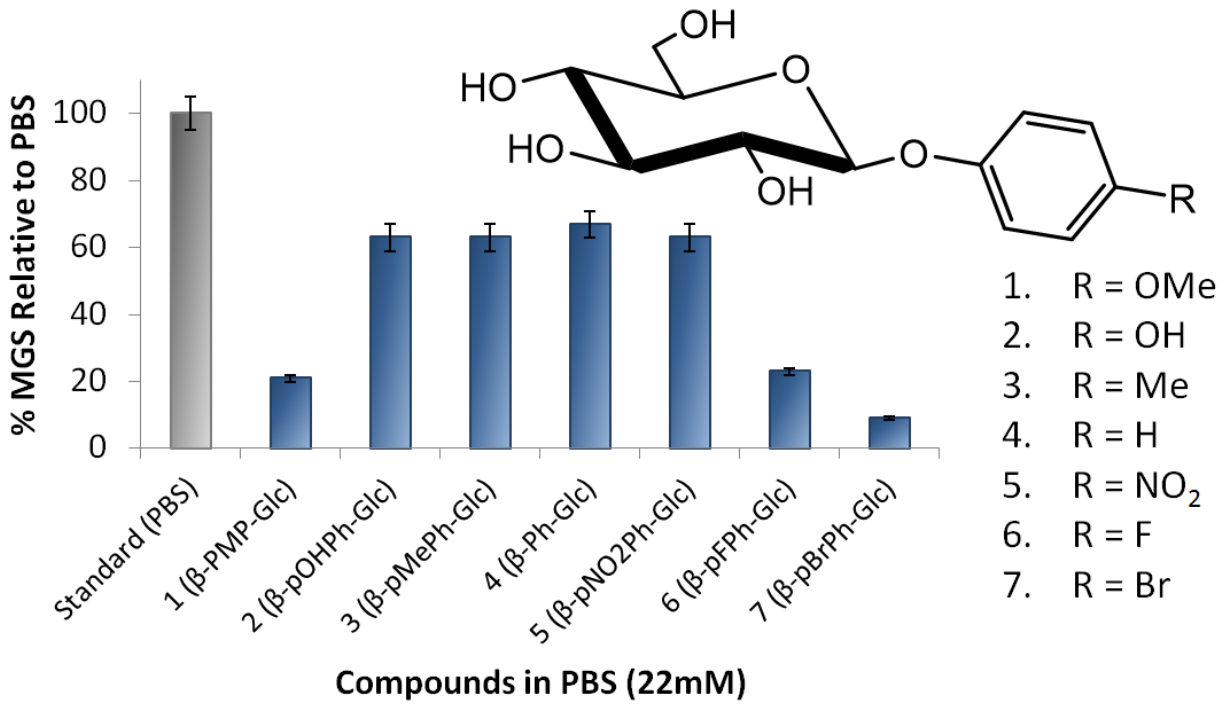


Figure 1.4: Graph of the IRI activity as expressed by the % mean grain size of a set of phenyl-glucosides which have been substituted at the para-position of the phenyl group. (Right) The structure of the phenyl-glucoside derivatives are shown along with their substitution groups.

The best performance came from β-pBrPh-Glc, where only addition of 30 mM to the sample allowed for the reduced use of only 15% glycerol and a recovery rate of 70-80% of post-thaw RBC.[48] Interestingly, β-pBrPh-Glc was a part of a set of other substituted phenolic-glucosides that all had drastically different IRI activities. By simply substituting the bromine at

the para position to a methyl, hydroxyl, hydrogen, or nitro group the percent mean grain size rose from 10% to over 60%.[48]

1.1.7 Structure Activity Relationships in IRI Molecules

The relationship between molecular structure and biological activity is something that has been long established and utilized in the field of drug discovery.[51] The basic assumption is that molecules with similar structure will also possess similar activities. Of course, an activity can be complicated with many competing factors, so any structure-activity relationship is more concerned about discovering *trends* which can point in the direction of improved activity. In order to compare the structure and function one needs a set of structures to correlate how changing the structure affects the activity. In terms of BAs such as polypeptides there have been structure-function relationships established by mutating portions of the protein and then testing for activity. Through the mutation of a native antifreeze protein Laursen *et al*. looked at the effect of neutral vs. polar amino acids,[52] charged amino acids,[53] and methyl and hydroxyl groups[54] in antifreeze activity. In these cases of structure-activity relationships were *qualitative* in nature, meaning they relied on the *qualitative* observations to determine a correlation or trend. However, there exist also *quantitative* structure-activity relationships which rely on quantitative data. For example, work described in Section 1.1.5 on hydration number vs. IRI activity of small molecules constitutes a quantitative structure-activity relationship. Quantitative-structure activity relationship (QSAR) models are computational methods which relate structure and activity using machine learning and regression techniques. The advantage here is the ability to distinguish trends from structures which may not be clear via chemical intuition. Furthermore, vastly larger databases of data can be computed and analyzed for high-

throughput screening in a fraction of the time. QSAR techniques have seen much success in the fields of drug discovery, toxicology, and protein activity studies.[55] The variety in structures and activities of small molecules can be increased via functionalization and this large search space demands a more efficient method of discovery. To the best of our knowledge, there have been no QSAR models constructed for IRI molecules to date. A more detailed description QSAR modelling will be discussed in Section 2.3.2**.**

1.1.8 Project Motivations and Goals

The field of using small molecules as IRIs is in its infancy with great potential for discovery. Currently, the mechanisms by which ice recrystallization inhibition works is uncertain with evidence pointing in multiple directions such as ice binding, QLL disruption, and other processes. Complicating the issue further is the fact that simple functionalizations and stereochemical substitutions can have a drastic effect on IRI activity. Finally, the search space for functionalization is potentially quite large as more complicated functional groups and compounds are tested for IRI activity. It would be quite expensive in terms of time and materials to synthesize, characterize, and test every possible compound. Thankfully computational chemistry offers an attractive alternative to real life screening, virtual high-throughput prediction.

The main goal of this project will be to develop a predictive model in order to help accelerate the experimental discovery of small IRI molecules. The Ben Lab has synthesized and tested a large catalogue of small IRI molecules, many of which have yet to published. This experimental data will be essential to creating a predictive quantitative structure-activity

relationship (QSAR) model. Essentially the goal is to determine a correlation between the structures and IRI activities of a database of diverse compounds in order to build a model which can classify a proposed compound as either IRI active or IRI inactive.

1.2 Metal Organic Frameworks

Metal organic frameworks (MOFs) are an exciting class of porous crystalline material which has invigorated the scientific community in the past decade. Since the emergence of this material in the late 90's, MOFs have endured unparalleled research interest in terms of the number of publications and the breadth of topics covered. These molecular self-assembled materials have extremely high surface areas and enormous porosity with up to 90% free volume.[56] MOFs lie at the intersection of inorganic and organic chemistry as the components of a MOF are the inorganic metal secondary building unit (SBU) and the organic linker. These structures have an extraordinary amount of variability in terms of the metal SBU and the organic linker used to construct the MOF. This allows for precise control in tuning desired characteristics for certain applications.

The term metal organic framework was first popularized by the work of Yaghi *et al.*[57,58] but the idea of a porous coordination three dimensional polymer was developed by the early simultaneous and independent work of Robson,[59] Moore,[60] and Zaworotko.[61] The design of metal organic frameworks first came about by trying to emulate the structures of microporous aluminosilicate materials called Zeolites.[58] Zeolites are naturally occurring inorganic materials made up of aluminum, oxygen, and silicon which have been known for centuries.[62] The porosity of zeolite materials has been utilized by the petroleum and chemical industry for separations,

catalysis, and waste management.[63–65] However, due to their limited chemical make-up of aluminum, oxygen, and silicon these materials have a limited amount of variability in terms of their connectivity, topology, and geometry. To compare with MOFs, the Cambridge Crystallographic Data Centre (CDCC) is composed of almost 14% MOF structures totalling 105,862 structures as of February 2015[66] whereas there are only about 200 zeolites from the Database of Zeolite Structures.[67] Additionally, while zeolites have been known for decades before MOFs, the number of publications per year for MOFs have risen at an exponential rate surpassing zeolites for the in 2014 and is on pace to surpass it again this year.[68] (Figure 1.5)

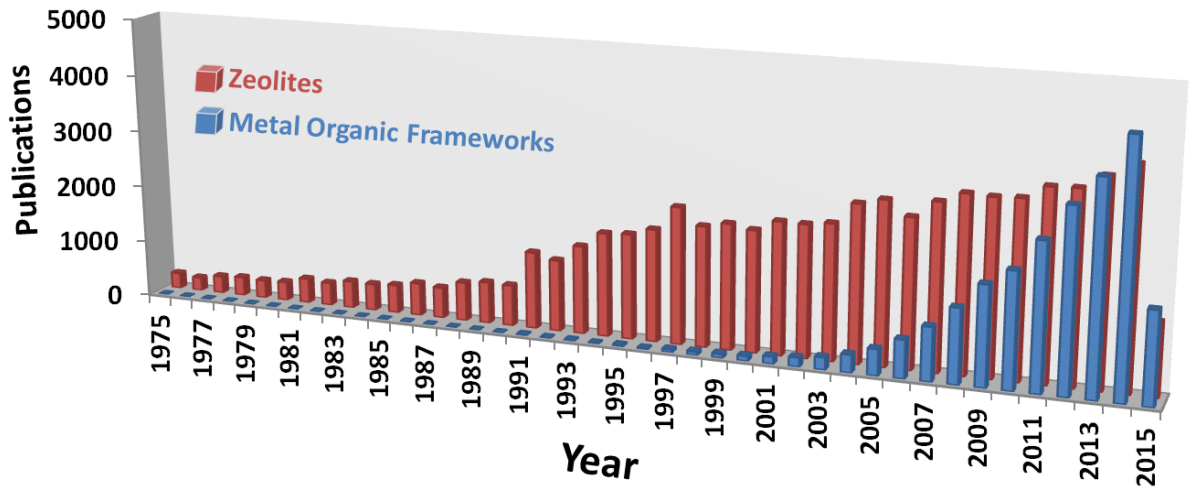


Figure 1.5: Graph of the number of zeolite vs. metal organic framework publications by year. Data taken from Web of Science academic search engine by Thomson Reuters.[68]

The structural design of MOFs can be rationalized by the term reticular synthesis.[69] The basis behind reticular synthesis is the self-assembly of rigid molecular building blocks (in MOFs the organic linker and metal SBU) into predetermined ordered structured networks (nets) via strong bonds. Perhaps the quintessential example of this is the simple cubic network MOF-5.[57] This MOF was prepared from Zn(II) metal and 1,4-benzene-dicarboxylic acid (BDC) to yield an octahedral SBU linked by organic benzene linkers. (Figure 1.6)

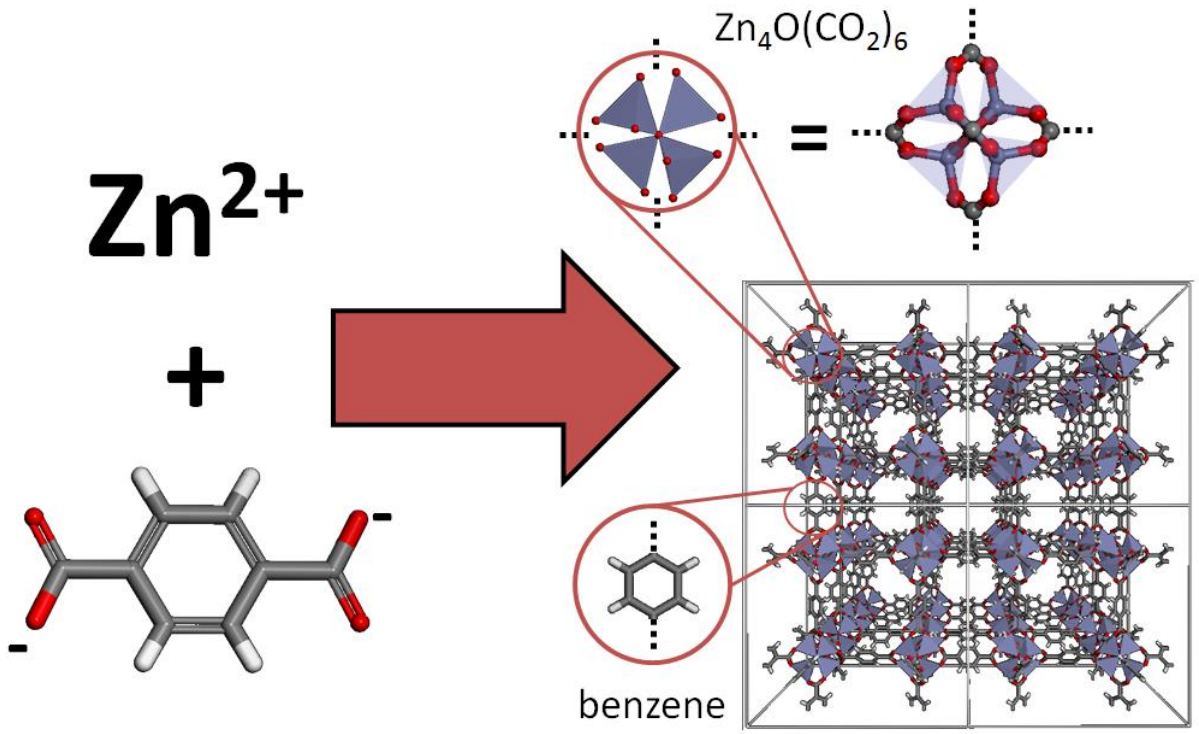


Figure 1.6: Schematic of the Zn(II) metal and 1,4-benzene-carboxylate (BDC) combining to form MOF-5 (right) with Zn in purple, carbon in grey, oxygen in red, and hydrogen in white. Clearly depicted are the metal SBU ($Zn_4O(CO_2)_6$ cluster) and organic linker (benzene). The dotted lines represent the connection vectors between the metal SBU and organic benzene linker. MOF-5 forms a simple cubic network where the metal SBU are the corners of the cube and the organic linker is the side.

The stability of MOF-5 can be rationalized through studying its simple primitive *c*ubic unit cell, aptly termed the *PCU* topology. Due to the Zn cluster and benzene being fairly large and rigid units with stable shapes (a truncated tetrahedral in the metal SBU and a planar slat in the organic linker) MOF-5 was one of the first examples of a rigid, stable and extremely porous material.

The variability of MOF design hinges on the fact that the only necessities are a metal ion with vacant sites and a bridging organic ligand. The more common metal ions are Cu^{+}, Cu^{2+}, Ag^{+}, Zn^{2+}, Co^{2+}, Al^{2+}, Ni^{2+}, and Mg^{2+} and other first row transition metals[56] while the organic linkers can encompass cyanide[70], triazole[71], oxalate[72], imidazole[73], and 1,2,4,5-tetracarboxylates molecules[74]. Additionally, the organic linker on a MOF can be functionalized post-synthetically

to provide even more unique structural possibilities.[75] Indeed, there has even been development of hypothetical frameworks in hopes of searching the combinatorial space of MOFs.[76] In fact, this is an active and important project within the Woo Lab with millions of hypothetical MOFs developed to date. The almost limitless combination of metal ions and organic linkers available means extraordinary control on the structural and chemical features of the MOF.

1.2.1 Ultra-Microporous MOFs

One of the greatest advantages of MOFs over other porous materials is the ability to control pore size based on the size of the organic linkers. This control has been shown through the work of Yaghi *et al.* through the IRMOF-74 series of MOFs (I to IX) whereby the pore size was systematically increased from 14 Å to 98 Å simply by increasing the size of the organic linker.[77] (Figure 1.7) This was done in an isoreticular fashion where the topology and connectivity all remain the same between the series of MOFs. Some MOFs in this series have pore sizes so large that even biological molecules such as vitamin B_{12}, myoglobin, and green fluorescent protein (GFP) can pass through. Currently, IRMOF-74-XI is the largest MOF that has been published with a pore size of 98 Å.

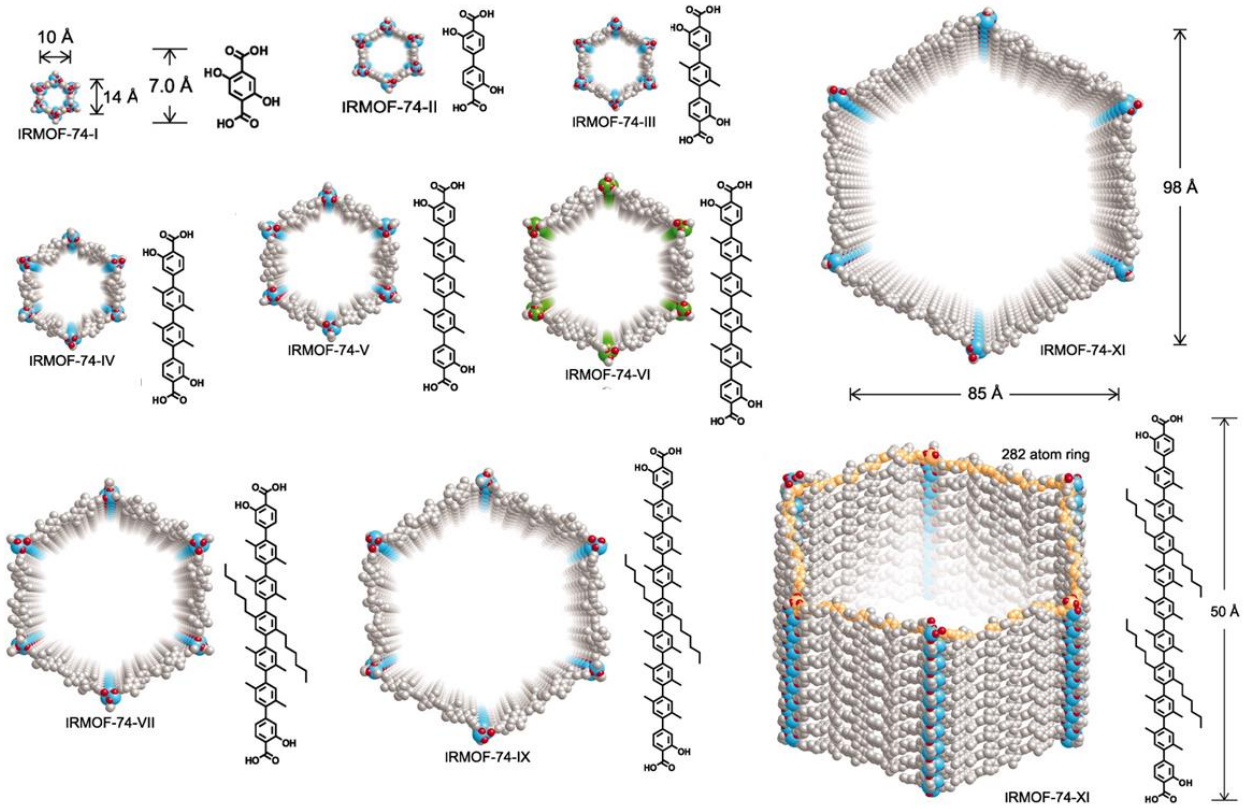


Figure 1.7: Representations of the pores of IRMOF-74-I, II, III, IV, V, VI, VII, IX, XI from the work of Yaghi et al.[78] These series of MOFs is an example of reticular synthesis whereby the connectivity and topology of all MOFs in the series are the same. Image reproduced from reference 78.

This paper showcases a series of MOFs that encapsulate the mesoporous (20 – 500 Å pores) and microporous (6 – 20 Å pores) regimes that the majority of MOFs belong to. However, there exists another class of MOFs with smaller pore sizes called ultra-microporous MOFs (2 – 6 Å pores).

The term ultra-microporous metal-organic framework first appeared in literature in 2007 with the work of Zhou *et al*.[79] The particular MOF synthesized was PCN-13 which was composed of a Zn_4O metal cluster coordinated to three water molecules and six organic 9,10-anthracenedicarboxylic acid to give the $Zn_4O(H_2O)_3(COO)_6$ metal SBU and an anthracene organic linker. This MOF exhibited very small pore sizes of 4.97 Å. Additionally there was a significant amount of hydrogen (46 cm^3/g) and oxygen (67 cm^3/g) adsorption but very little nitrogen or

carbon dioxide adsorption.[79] This selectivity of hydrogen and oxygen over nitrogen and carbon dioxide is important in selective gas adsorption and gas separations[80] which is an important application which will be presented in the next chapter. While at first glance it might seem that large pores would be more conducive to high uptake due to more space for adsorbents to fit, small pored MOFs actually have characteristics which make them excellent gas sorbents.[81] For example, the Woo Lab in collaboration with the Shimizu Group was able to demonstrate an ultra-microporous MOF, $Zn_2(Atz)_2(ox)$ exhibited strong framework-gas interactions and enhanced cooperative effects between adsorbed gas molecules that was facilitated by the small pores.[82] This study utilized computational simulation to show that the ultra-micropores of 4 Å promoted preferential carbon dioxide binding at low pressures due to strong van der Waals dispersion interactions, high density of binding sites, and cooperative guest orientations. Furthermore, ultra-micropores are small enough to provide molecular sieving abilities which allows for the selective diffusion of guest molecules through the pores based on size.[83] That is, a pore that is sufficiently small will allow for a larger gas molecule to fit tightly while a smaller gas molecule can pass through would exhibit molecular sieving. Finally, ultra-microporous MOFs can be made up of small rigid organic ligands which lead to structurally robust MOFs with longer shelf-life and easy synthesis due to the use of a single organic ligands.[84–87] However while ultra-microporous MOFs have shown excellent gas adsorption capabilities at low pressure, they often have relatively low saturation limits at high pressure due to the small pores and limited pore volume.

1.2.2 Flexible and Gate-Opening MOFs

The ability for a MOF to dynamically transform its structure in response to external stimuli such as guest adsorption,[88] light irradiation,[89] temperature,[90] and mechanical force[91] is a phenomena which is unique and uncommon for other crystalline solid-state materials. These flexible MOFs are also termed soft porous crystals due to the soft interaction between the metal and the organic linker which allows for flexibility in structure.[92] The possibility of these types of materials were first envisioned in 1998 with the introduction of three "generations" of porous materials.[93] The first generation are materials which collapsed upon removal of the guest molecule and thus did not maintain permanent porosity. The second generation are materials that maintained permanent porosity upon guest removal and can be analogous to other rigid nanoporous materials such as zeolites. The third generation are materials that are capable of reversible structural transformations while also maintaining a highly ordered network and permanent porosity.[93] These reversible structural transformations in MOFs are typically a crystal-to-crystal phase transition. An example of a third generation porous material are "breathing" MOFs which show a change in pore volume upon the adsorption and desorption of guest molecules.[94] While these MOFs present an interesting case, as of 2014 a little less than 100 compounds in the Cambridge Structural Database display phase transitions which can be considered flexible.[95]

The types of structural transformations in flexible MOFs can be distilled into four modes, breathing, swelling, linker rotation, and subnetwork displacement.[95] Breathing can be considered the reversible transition of MOFs whereby the framework atoms are substantially displaced leading to a change in unit cell volume. In this case the space groups and lattice

parameters of the open and closed phases may be different but the same. The archetypal MOF of this type is the MIL-53(M) family of MOFs $[M(bdc)(OH)]_n$ which are made up of different metal centres (Al,[96] Fe,[97] Cr,[98] Sc,[99] Ga,[100]) and 1,4-benzenedicarboxylate as the organic linker. In this specific case, MIL-53 reversibly adsorbs water molecules at room temperature which leads to a shrinking of the pores.

Swelling can be considered an enlargement or shrinking of the MOF unit cell volume without a change unit cell shape so the cell lengths might change by the unit cell angles would be unchanged. A prototypical swelling MOF would be MIL-88A which is a based off a trimeric $M_3O(H_2O)_2OH^{6+}$ (M = Fe^{3+}, Cr^{3+}) SBU connected with fumaric acid linkers.[101] In this case the structural changes are guest dependent with increases in unit cell volume as a function of different alcohol adsorptions. Here, the activated empty structure of MIL-88A had a unit cell volume of 1135 $Å^3$ but then increased to 1840 $Å^3$, 1970 $Å^3$, 2090 $Å^3$, and 2110 $Å^3$ after loading of *n*-butnaol, ethanol, methanol and water respectively.

Linker rotation is the continuous rotation of an organic linker around a specific axis such that a pore volume may change but the unit cell size, topology, and lattice parameters remain the same. One recent example of guest-dependent linker rotation of phenyl ring linkers is that by Jenkins *et al*.[102] Here a series of MOFs were synthesized with Cu(I) and 4,4'-(1,4-(xylene)-diyl)bis(1,2,4-triazole) as the ligand. Depending on the guest solvent molecules of dimethylformamide (DMF), dimethylacetamide (DMA), N-Methylpyrrolidone (NMP), and water the MOF would crystallize to in different conformations where the phenyl rings are rotated either parallel or perpendicular to the pore channel axis. Molecular dynamics simulations were also used to analyze the energetics and dynamic rotation of the phenyl ring.

It was found that the framework-framework energy cost of rotating the linker rings are offset by the stronger framework-solvent interaction energy.[102]

Subnetwork displacement occurs when different subnetworks within a MOF which are held together by weak forces rather than strong chemical bonds are shifted with respect to one other. The three main cases of subnetwork displacement occur in interpenetrated three dimensional frameworks, interdigitated layered frameworks, or stacked two dimensional frameworks.[103–105] Subnetwork displacement can be pressure dependent in terms of the guest providing a gate opening/closing effect. For example, the 2D MOF $[Cu(dhbc)_2(dipy)]$ is a two dimensional MOF made up of a copper metal center, 2,5-dihydroxybenzoic acid (dhbc) and 4,4'-dipyridine (dipy) organic linkers which forms stacked sheets of 2D framework. It was found that the MOF showed a gate-opening after a certain pressure threshold was reached for various gases such as nitrogen, oxygen, and methane.[106]

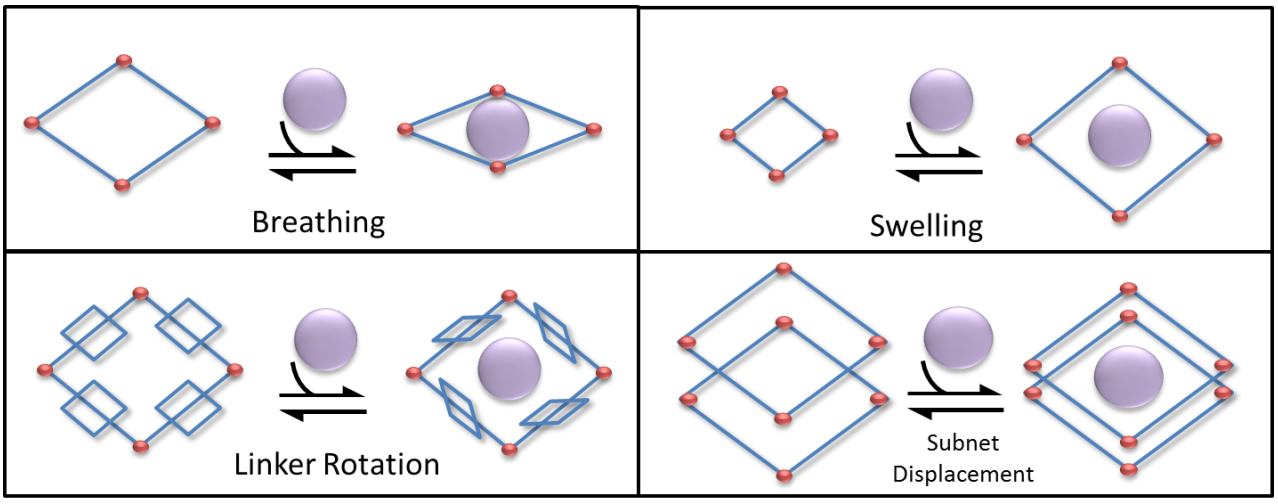


Figure 1.8: Representation of the four different flexible framework modes. Breathing, swelling, and subnet displacement all feature a change in unit cell volume whereas linker rotation may not.

1.2.3 Applications

Due to the explosion of research interest in the past few years, an immense number of possible applications for MOFs are being discovered. Indeed, there have been great strides in the application of MOFs for methane storage,[107] carbon dioxide capture,[108] hydrogen storage,[109] hydrocarbon and alcohol separation,[110] thin film devices,[111] optics,[112] chemical sensors,[113] luminescence,[114] heterogeneous catalysis,[115] and biomedicine.[116] MOFs represent an interesting venue for truly functional design and with the stability and chemistry of MOFs not completely understood, the field is at a "*try it, see if it works, what is it good for?*" stage. For the sake of conciseness only the application of MOFs for carbon capture and sequestration will be discussed further.

1.2.4 Carbon Capture and Storage

According to a World Energy Council Report, population growth and rising standards of living across the world will at least double global energy demand by 2050.[117] The burning of fossil fuels remains the most inexpensive form of energy but is also the leading cause of anthropologic CO_2 emissions. Humanity recently reached a milestone of 400 ppm CO_2 emitted globally per month in March, 2015.[118] (Figure 1.9) To make matters worse, recent studies have shown that the climate change effects due to CO_2 emissions are irreversible for 1000 years even after CO_2 emissions halt.[119] Some dire effects include rainfall reductions and droughts akin to the "dust bowl" era which lead to the Great Depression in North America, thermal warming of the ocean leading to global average sea level rise of 0.4 – 1.0m over the coming century, and

mass extinction of aquatic life.[119] Thus, materials for carbon capture and storage in order to combat climate age are one of the defining problems of this scientific generation.

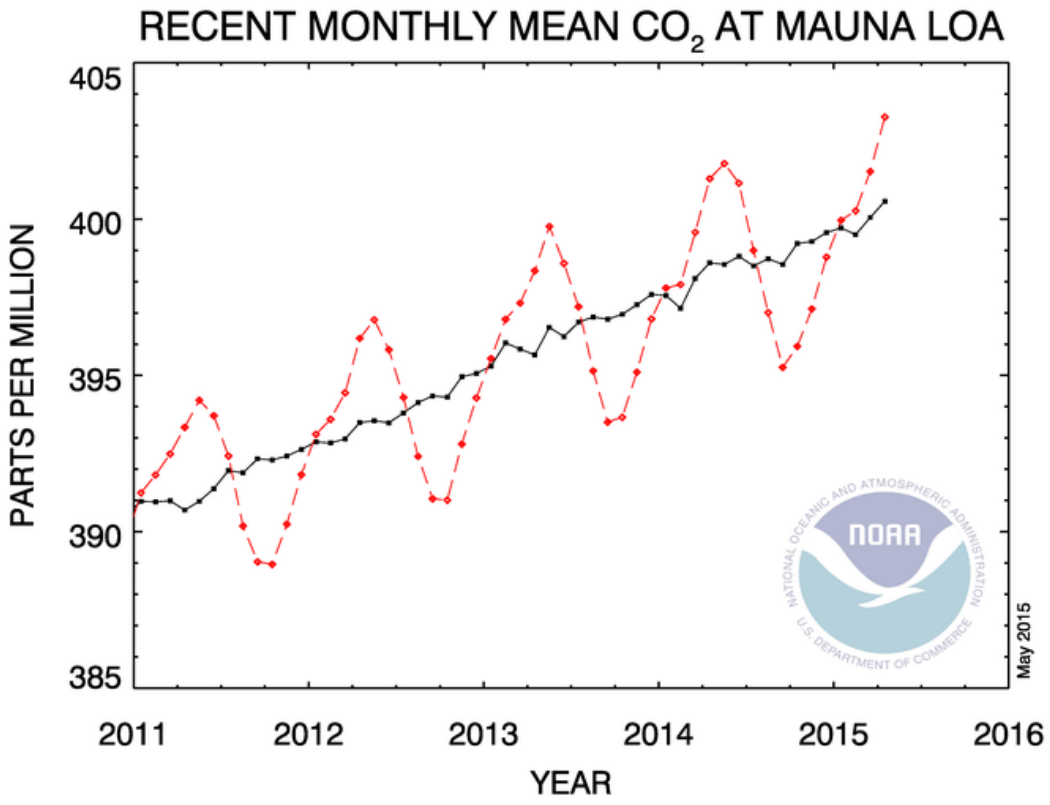


Figure 1.9: Increasing CO_2 levels in parts per million (ppm) from the last four complete years taken from Mauna Loa Observatory in Hawaii. The red line represents the monthly mean values centered on the middle of each month. The black line is a correction for the average seasonal cycle.[118]

One of the largest and detrimental sources of CO_2 emissions are that of energy production plants which burn fossil fuels such as coal and natural gas. In Canada alone from March 2014 to February 2015 the production of energy from the burning of fossil fuels was equivalent to 112,796 gigawatt hours.[120] When adjusted to the current electricity regulation of 420 tonnes of carbon dioxide per gigawatt hour[121] for CO_2 emissions from coal power plants, this means Canada emitted 47.37 megatonnes of CO_2 in the last year from generating electricity alone.

While there has been great progress in moving to renewable energy sources in Canada, cheap energy production via fossil fuels will remain the global standard among developing nations.

One promising strategy for mitigating the environmental damage and limiting the emissions of CO_2 is called carbon sequestration and storage (CSS). In CSS, carbon dioxide is captured from the combustion exhaust (flue) streams of power plants on site, shipped to off-site reservoirs, and then permanently stored underground. The most costly and energy intensive part of the CSS process (~70%) the capture of CO_2.[122] Amine scrubbing is a well-established technology first patented in 1930[123] which currently stands as the industry standard for large scale CO_2 capture. In this process, the flue gas is bubbled through an aqueous amine solution near ambient temperature. The CO_2 chemically reacts with the amines to form carbamates. The regeneration of the amine and removal of CO_2 is accomplished by heating the amine solution to $100^o - 120^oC$ where the products are pure CO_2 which is then compressed for sequestration and water which is condensed and reused. The large capital costs with respect to implementing this process, utility costs in terms of heating the solution, and added containment risks with a liquid sorbent have kept this technology from being widely implemented in industry.[124] Solid sorbent materials present a more attractive alternative to the current use of amine solutions. The biggest advantage of solid sorbents is that there is no need to heat a solution since water has a high heat capacity and most solids utilize physisorption rather than chemisorption.[125] Physisorption of CO_2 in solid state adsorbent materials means that the CO_2 adsorbs via weak dispersive force such as van der Waals rather than chemically reacting and forming strong bonds in chemisorption. Essentially, the energy needed to regenerate CO_2 is lowered since there is no need to break a strong chemical bond. Another

primary advantage is that solid sorbents eliminate the need to heat water. For example when comparing the common MOF HKUST-1 (heat of adsorption of CO_2 = 35 kcal/mol)[126] to an amine solution (heat of adsorption = 90 kcal/mol)[124] it is clear that less energy is required to regenerate CO_2. Furthermore, the specific heat capacity of most MOFs (e.g HKUST-1, 0.8 $J(g\cdot^oC)^{-1}$)[127] are much lower than that of liquid amines (3 – 4 $J(g\cdot^oC)^{-1}$)[128] which means the energy required to heat the material for CO_2 regeneration is significantly lower for MOFs.

Throughout literature there are predominately three different methods for carbon capture: pre-combustion CO_2 capture, post-combustion CO_2 capture, and oxyfuel combustion.[108] In pre-combustion CO_2 capture the fuel is cleaned of any carbon dioxide before combustion such that the *fuel is pure H_2* and thus this process results in zero CO_2 emission. This process begins with the gasification of coal whereby coal is oxidized by oxygen in the presence of water to produce hydrogen gas and carbon monoxide. (**Eq. 1**)

$$3C\ (coal) + O_2 + H_2O \rightarrow H_2 + 3CO \qquad (\textbf{Eq. 1})$$

The carbon monoxide is then reacted with water vapor in a water gas shift reaction where the main products are CO_2 and H_2 which is commonly referred to as shifted syn gas. (**Eq. 2**)

$$CO + H_2O \rightarrow CO_2 + H_2 \qquad (\textbf{Eq. 2})$$

This gasification process is typically done at higher pressures of 5 – 40 bar and elevated temperatures of 40^{o}C and yields a fuel gas stream composed of 20 – 40% CO_2 and 60 – 80% H_2. [129] At this point CO_2 is separated from H_2 and then is consequently used as the fuel to be combusted in electricity generation with the only combustion product being water. Pre-combustion is perhaps the most developed method for industrial application with many

integrated gasification combined cycle (IGCC) plants near completion or already running in the United States such as Mississippi Power's 524 MW Kemper Project,[130] California's 300 MW Hydrogen Energy California Project (HECA),[131] and Summit's 400 MW Texas Clean Energy Project (TCEP).[132] Unfortunately all these power plants currently use solvent based carbon capture techniques but research on solid state pre-combustion sorbent materials are ongoing.[133] Furthermore, pre-combustion CSS must be integrated into the power plant system before the combustion cycle and thus only new power plants can use this process which increases the initial capital costs immensely. This also means that this process cannot be retrofitted to existing power plants. Unfortunately, replacing existing power plants with new ones is prohibitively expensive since existing plants were financed with an expected lifetime of, on average, 40 years.

Post-combustion capture, as the name implies, is the capture of CO_2 from the flue gas after the combustion of fuel at pressures of approximately 1 bar and temperatures of 40 – 60^oC.[134] This post-combustion flue gas is composed of ~ 15% CO_2, ~75% N_2, ~6% H_2O, and ~4% O_2[135] and thus selective capture of CO_2 over N_2 is essential. The main advantage of this method is that it can retrofitted to existing power plants and has the greatest potential for global impact in areas where the capital cost of CSS is too high. The greatest challenge in the implementation of this technology is that the CO_2 capture technology is not energy efficient enough to make it feasible. Currently there is intense interest in developing new solid sorbents, such as MOFs, to make post-combustion CO2 capture feasible. The ideal solid sorbent is one that is economical, highly selective for CO_2, with high CO_2 adsorption capacities, energetically

favourable regeneration, possesses long-term stability to heat and water, and allows for rapid diffusion of gas through the solid.[136]

Oxyfuel combustion is the newest method for CSS and involves igniting coal in a pure oxygen environment which results in two primary products, water (which is removed) and almost pure CO_2 gas which is captured. . First, oxygen is separated from air which is primarily an O_2/N_2 separation. Next, this pure oxygen is diluted with CO_2 from the combustion flue gas to a partial pressure of 0.21 bar in order to keep the temperature of fuel combustion manageable.[137] Finally combustion occurs with CO_2 (65 wt%) and water vapour (35 wt%) as the main exhaust gases. CO_2 is subsequently captured within a sorbent. Oxyfuel combustion boasts the highest CO_2 capture rates of all three methods with higher than 95% captured.[138] Additionally, in oxyfuel combustion oxygen is about 20% of the gas stream whereas in post-combustion CO_2 capture CO_2 is only 10-15%.[138] Similar to pre-combustion CSS, oxyfuel combustion cannot be retrofit to current power plants and require large capital costs. A further challenge exists in the separation of O_2 from the air which is currently done cryogenically and is an energetically expensive process.[139]

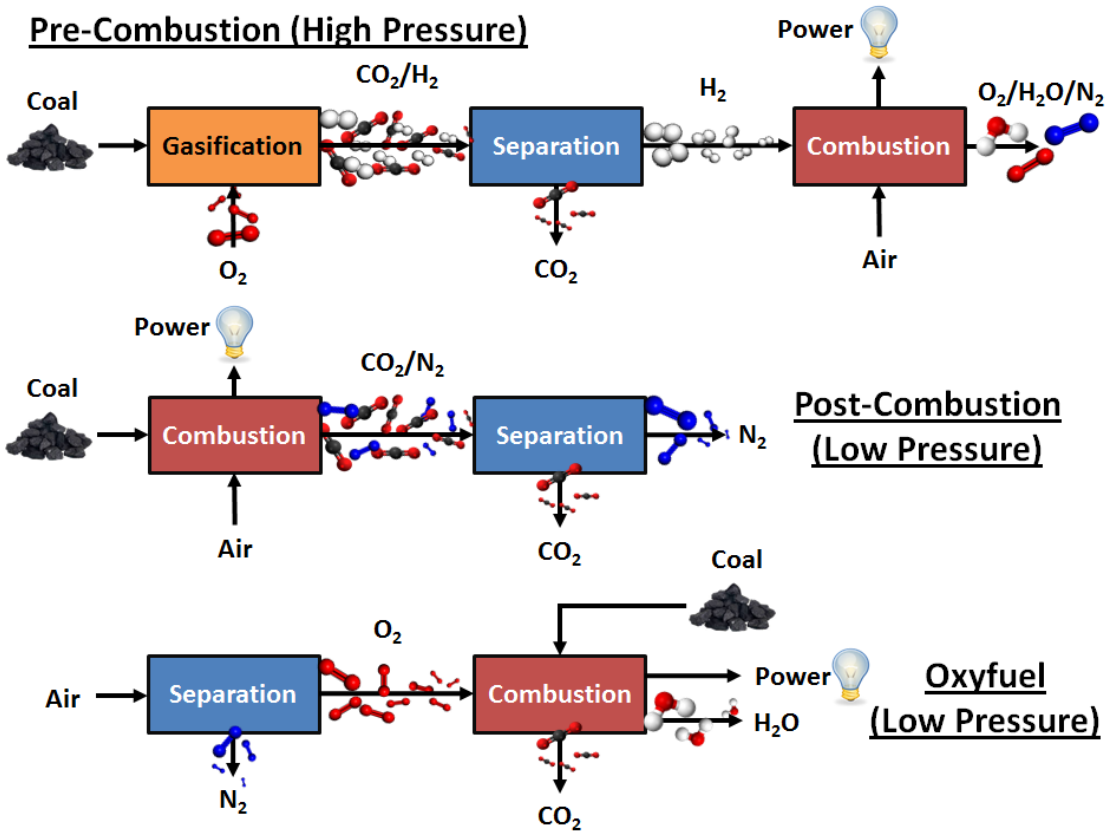


Figure 1.10: Schematic flowchart representation of the different methods of carbon capture and sequestration. First is pre-combustion carbon capture where the main separation is CO_2/H_2 at high pressure. Second is post-combustion carbon capture where the main separation is CO_2/N_2 at low pressure. Third is oxyfuel combustion where the main separation is O_2/N_2 at low pressure.

Once CO_2 is captured by the aforementioned methods, the material must be regenerated – that is, the CO_2 must be removed from the adsorbent for transport and storage. In liquid amine solutions, this is done by heating. With regards to solid sorbents like MOFs, the regeneration is done by altering temperature and/or pressure via temperature swing adsorption (TSA), pressure swing adsorption (PSA), or vacuum swing adsorption (VSA).[140] These processes hinge on the fact that a material uptakes different amounts of CO_2 at different temperatures and pressures. Each MOF has different uptake characteristics which allows for the desorption conditions to be optimized to fit each unique MOF. In TSA, the temperature is increased from ambient temperature to some optimal temperature in order to desorb the CO_2

molecules. The desorption process works two fold, first the heat allows for the CO_2 to break the physical adsorption interaction, second it allows for the pressure inside the adsorbent to increase which in turn pushes the CO_2 out of the MOF. A purge is done after equilibrium to try and remove any trace amounts of CO_2 that remain and then the MOF is cooled for subsequent adsorption. In PSA, the intake gas is pressurized to pressures above atmospheric pressure before it runs through the adsorbent material. This is especially useful in pre-combustion capture where the gasification of coal results in an initially pressurized gas. At high pressure the MOF will adsorb faster and usually reaches full saturation. Next, the inlet value is closed and the outlet is opened causing a pressure gradient from high pressure down to atmospheric pressure. The concept for VSA is very similar to PSA however instead of an initial pressure above atmospheric pressure, the initial pressure is atmospheric. In order to desorb the gas a vacuum is applied to bring the pressure to sub atmospheric levels. The largest energetic cost in all these processes would be the heating, gas compression, or vacuum. Fortunately, the adsorption and the desorption conditions can be optimized for the specific solid sorbent used in order to reduce this cost.[141]

Clearly, optimizing a CSS system is complex and requires the influence of many different factors to be considered. To that effect, there have been many different metrics proposed by various groups to evaluate the suitability for a solid adsorbent material in CSS. Three widely used metrics include the selectivity of gases[142], the working capacity of the MOF[143], and the parasitic energy.[144] The selectivity of a gas can either be kinetic, based on the size of guests, or thermodynamic, based on the physical properties of the guests.[145] Kinetic separation typically relies on separating molecules based on size. Ultra-microporous MOFs have pore sizes in the

range of some gas molecules which allow them to exhibit molecular sieving capabilities. However, these MOFs typically lack high saturation capacity due to low surface areas and low diffusion rates due to small pore openings. In thermodynamic separation the difference in adsorption enthalpies of certain gases are utilized such that some gases are preferred over others. For example CO_2 has a polarizability of 29.1 x 10^{-25} cm^{-3} compared to N_2 with a polarizability of 17.4 x 10^{-25} cm^{-3} and a MOF with exposed cation sites or polar organic linkers may selectively bind CO_2 over the more inert N_2.[146] The selectivity is typically calculated as a ratio of adsorbed gas over partial pressure for the two gases, (**Eq. 3**)

$$S = \left(\frac{q_1}{p_1}\right) \Big/ \left(\frac{q_2}{p_2}\right) \qquad \textbf{(Eq. 3)}$$

where q_i is the amount of adsorbed gas *i* and p_i represents the partial pressure of gas *i.*

The working capacity is defined as the amount of CO_2 which can be captured given the specific temperature and pressures of an adsorption cycle.[147] This is the difference in quantity adsorbed at the adsorption pressure/temperature and the quantity adsorbed at the desorption pressure/temperature. Ideally, one would want the largest working capacity over a small pressure interval. One way this can be accomplished is via a flexible MOF where the *closed* conformation has very low adsorption capacity and the *open* conformation has a very high adsorption capacity. (Figure 1.11) Further exploration of this will be present in the coming chapters regarding Mg-4PyC.

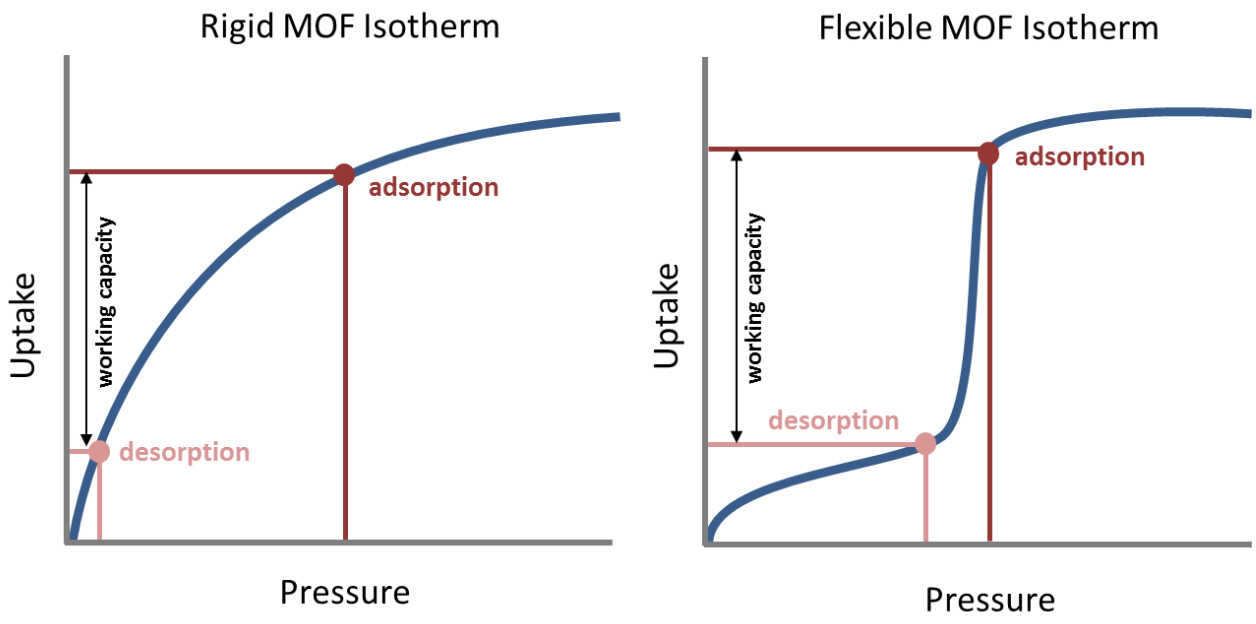


Figure 1.11: Isotherms of CO_2 in an idealized rigid MOF (left) and flexible MOF (right). An isotherm is a graph of uptake with respect to pressure at a fixed temperature. Isotherms are particularly useful in conveying adsorption for PSA applications. The working capacity is the difference of uptake between adsorption and desorption pressures. Notice that due to the stepped isotherm for a flexible MOF the pressure difference is much smaller to gain the same working capacity as that of a rigid MOF.

Another metric that was recently proposed was to quantify the energy needed for a CSS process in terms of the burden this process would have on a power plant. This metric has been coined the parasitic energy.[144] Here the parasitic energy takes into account the energetic cost of two processes, the amount of heat energy needed to recovery CO_2 from the MOF and the energy needed to compress the CO_2 for transport to 150 bar. The amount of heat needed is termed the thermal energy requirement and is defined as the sum of energy needed to heat the MOF to desorption temperature and the energy required to supply the heat of adsorption, (**Eq. 4**)

$$Q = \frac{C_p m_{sorbent}(T_{final} - T_{flue}) + (\Delta q_{CO_2} \Delta h_{CO_2} + \Delta q_{N_2} \Delta h_{N_2})}{m_{CO_2}} \quad \textbf{(Eq. 4)}$$

where c_p is the specific heat capacity of the MOF, $m_{sorbent}$ is the mass of the MOF, T_{final} is the temperature of desorption, T_{flue} is the temperature of adsorption, $\Delta q_{(i)}$ is the working capacity of the gas, $\Delta h_{(i)}$ is the heat of adsorption of the gas, and m_{CO2} is the mass of CO_2. The parasitic energy is thus defined as, (**5**)

$$E_{eq} = 0.75\,\eta_{T_{final}} Q + W_{comp} \qquad \textbf{(Eq. 5)}$$

where $0.75\eta_{Tfinal}$ is the excess thermal energy of the power plant and the factor 0.75 is the typical efficiency of a turbine, η is the Carnot efficiency, and W_{comp} is the work of staged compression to 150 bar. This work by Smit *et al.* was used to screen 330,000 porous zeolites in order to reduce the parasitic energy of current industry standards. Reduced parasitic energy means lower energy cost, higher power plant efficiency, and ultimately decreased monetary costs.

1.2.4.1 Metal Organic Frameworks for CSS

Currently, no MOF or any solid sorbent has been implemented in a large scale CSS plant as an ideal material has yet to be discovered to make it economically feasible. However, there has been fervent research dedicated to this topic with hundreds of MOFs being explored for this application.[148] Due to the ability for MOFs to be fine-tuned for certain applications the design principles which have evolved for the different types of capture methods are quite different. For pre-combustion CO_2 capture, a screening done by Long *et al.* showed that MOFs with a higher amount of exposed metal cation sites (also known as open-metal site MOFs) showed the best promise for PSA applications with working capacities as high as 8.6 mol/kg and CO_2/H_2 selectivities as high as 860.[129] They found MOFs to be even better adsorbents than

Zeolite 13X and activated carbon JX101 which are currently used in some industrial applications for H_2 purification. Unfortunately, many MOFs which exhibit high H_2 uptake are also ones with open metal sites which are prone to degradation via water.[149] These are metal centres with free coordination sites which typically have a higher affinity to water than other guest gases which results in decreased uptake.[150,151] With regards to post-combustion CO_2 there have been many recent efforts in incorporating amines into the organic linker of a MOF in order to chemically bind CO_2. A recent study published last year by Yaghi *et al.* showcased IRMOF-74-III that was functionalized with a primary amine which was capable of selectively capturing CO_2 while retaining stability in the presence of water (65% relative humidity).[152] This was an important discovery because one of the largest challenges with MOFs for CSS is their susceptibility to water degradation in humid environments. Another study published recently in Nature elucidated a chemisorption mechanism which induced a phase change in the MOF mmen-M_2(dobpdc) (M = Mg, Mn, Fe, Co, Zn) at small pressure swings.[153] This MOF is made up of a metal centre, 2,5-dioxidobenzene-1,4,dicarboxylate (dobpdc) linker, and functionalized with N,N'-dimethylethylenediamine (mmen) at an open metal site. Here a unique adsorption mechanism was characterized for the first time by spectroscopic, diffraction, and computational studies which showed that upon a certain CO_2 pressure the CO_2 molecules insert into metal-amine bonds causing the amines to reorient themselves into chains of ammonium carbonate. These studies are a few chosen examples of MOF research which shed light on how broad and diverse this field of research really is.

1.2.5 Probing MOFs with Molecular Simulation

Molecular simulation is an invaluable tool in probing the characteristics of MOFs, providing explanations for phenomena that are difficult to probe experimentally, or even high-throughput screening of MOFs. The field of molecular simulation for MOFs has grown exponentially within the past few years and is the main research focus of the Woo Lab. With regards to molecular simulation there are two main levels of theory, quantum mechanical and molecular mechanical. With large periodic systems such as MOFs, quantum mechanical analysis is typically too costly for use of any dynamic studies. However, *ab initio* calculations are routinely utilized for the determination of charges,[154] electronic structure analysis,[155] open-metal site studies,[156] and more. With respect to gas adsorption, molecular mechanical methods such as molecular dynamics and grand canonical Monte Carlo (GCMC) simulations are routinely used.

The backbone behind any classical calculation is the forcefield – the set of interatomic potentials which describe the bonded and non-bonded energies of a molecular system. This is based off of molecular mechanics where the molecule is classically treated as a system of "balls" and "springs". The Universal Force Field (UFF) is one of the most used force fields for molecular simulation.[157] Metal organic frameworks have been studied so intensively by the molecular simulation community that last year an extension of the UFF made specifically for MOFs called UFF4MOF was released.[158] Forcefield parameterization and development for MOFs is currently a large field of research with forcefields developed for MOFs such as Cu-BTC,[159] zeolitic imidazolate frameworks (ZIFs),[160] and flexible MIL-53(Al)[161] to name a few.

Once an appropriate forcefield has been chosen for the system, Grand Canonical Monte-Carlo (GCMC) simulations are typically used to simulate the adsorption isotherms for various gases. An adsorption isotherm is a plot of the equilibrium uptake vs. pressure at a specific temperature. The adsorption isotherm is typically the standard by which many materials are compared in terms of adsorption and guest uptake. This well-established method has been able to quantitatively match experimental data and reliably predict isotherms at pressures, temperatures, and mixtures not easily accessible with experiment.[162] GCMC simulations for MOFs will also be the main topic of a further chapter so the discussion will be limited here but further reading is available through the many published reviews of recent years.[163,164] In addition to GCMC, molecular dynamics (MD) have been used to simulate the gas diffusion of MOFs,[165] dynamic effects of guests with open metal sites,[166] and study of flexible MOFs.[167] Further thorough details on these two methods will be covered in a subsequent chapter.

The number of possible MOFs is nearly infinite due to the combination of metals, organic linkers, and also potential functionalization. To this extent there has been much interest in developing databases of hypothetical MOFs and then screening them in a high-throughput fashion. The first development of building hypothetical structures appeared in 2000 with the work of Mellot-Draznieks *et al*. – the automated assembly of secondary building units or AASBU.[168] This energy minimization approach randomly places building units within a unit cell at specific interactions sites which then are annealed with Monte Carlo simulations. Conversely, geometric based approaches to structure building have also been explored. For example, 137,953 hypothetical MOFs were constructed based on structures of known MOFs via a bottom approach by Wilmer *et al.*[76] The construction of hypothetical MOFs is a large focus in the Woo

Lab with over 1.3 million MOFs created to date with topology based MOF generation and genetic algorithm functionalization. With such a massive set of structures, efforts have been made to more efficiently screen these MOFs. In this sense there is currently an emerging field of research in developing quantitative structure-property relationship (QSPR) models in order to screen and predict MOF properties. This is another topic of focus in the Woo Lab with QSPR models being utilized to screen for high-performing MOFs with respect to CO_2 capture,[169] methane storage,[170] and development of QSPR descriptors for gas adsorption.[171]

1.2.6 Project Motivations and Goals

There will be two distinct projects which will be presented in this thesis under the topic of MOFs. These projects are much more intimate studies of specific newly synthesized MOFs rather than the high-throughput screening studies, development of screening methods or forcefields, and MOF generation projects of the Woo Lab. The two projects are a direct result of collaboration with experimentalist Dr. Vaidhyanathan Ramanathan from the Indian Institute of Science, Education and Research (IISER), Pune. Dr. Ramanathan approached the Woo Lab with two interesting MOFs based on isonicotonic acid as the organic linker. The first MOF, Ni-4PyC was an ultra-microporous MOF which showed exceptionally high CO_2 uptake and selectivities at high pressure relevant to pre-combustion CO_2 capture. This work has been tentatively accepted, following requested revisions, to the journal *Science Advances* published by the American Association for the Advancement of Science (AAAS) publishing group. The second MOF, Mg-PyC is an ultra-microporous MOF which exhibits a stepped isotherm and gate-opening characteristics due to the rotation of the organic ligand. This work is currently on-going. The main goals for these projects are to use computational simulation to probe

experimental phenomena and provide explanations to the unique characteristics of these two novel MOFs. These projects hope to display the synergy between experiment and simulation whereby both approaches to the problem were necessary in order to develop a thorough and concise scientific story.

1.3 Summary of Chapters

A brief summary of chapters will be provided here. Chapter two, *Theoretical Methods*, will discuss the relevant theoretical methods for the computational simulations used. Topics discussed in further detail include conformational searching, QSAR models, periodic DFT, GCMC simulations, molecular dynamics, etc. The computational methods discussed will be separated based on the topic of either ice recrystallization inhibition or metal organic frameworks.

Chapter three, *3D-QSAR Prediction of Small Molecule Ice Recrystallization Inhibitors*, will encompass the work done with respect to QSAR prediction of small molecule ice recrystallization inhibitors. The structures and activities of the initial set of small molecules provided by the Ben Lab will be presented. Generation of three dimensional structures from two dimensional ChemDraw format and subsequent conformational searching will be also be discussed. Next, three dimensional grid independent descriptor generation from quantum mechanical data will be explained which was accomplished with help from previous work done by Nick Trefiak. A brief discussion on the development of the QSAR model will be presented which was done by former post-doc Michael Fernandez. Finally, proposed structures, QSAR results, and experimental validation done by the work of Jennie Briard from the Ben Lab will be discussed.

Chapter four, *Ni-PyC: An Ultra-microporous MOF for Pre-combustion CO_2 Capture and Hydrogen Purification*, will discuss the work collaborative done with the MOF Ni-PyC. First geometry optimizations of Ni-PyC were done. Then GCMC simulations were performed and the binding sites of CO_2 molecules were determined. Next, high pressure isotherms were simulated and then subsequently validated experimentally. Further discussions will be done on the calculation of CO_2/H_2 selectivity, CO_2 heat of adsorptions, diffusion coefficients via molecular dynamics, structural characterization of pore sizes, and MD simulations of pore accessibility.

Chapter five, *Mg-PyC: Pressure Dependent Porosity via Organic Linker Rotation*, will present the efforts to probe the dynamic gate-opening effect of Mg-PyC. An exhaustive conformational search was performed with respect to linker orientation at the DFT level. Next efforts were made to modify the UFF forcefield in order to fit the potential to reproduce DFT level energies. This was done in order to try and capture the linker rotation as accurately as possible. Several tests done to validate this modified potential will be discussed also. Finally, molecular dynamics simulations to probe linker rotation movement with respect to CO_2 loading were done.

Chapter six, *Conclusions, Outlook, and Future Work,* will summarize the content of the thesis as concisely as possible. Reflections on the impact of the work done as it relates to the cryogenics and MOF communities will be discussed. Finally, future work will be proposed with some promising initial results of on-going work.

1.4 References

(1) Capicciotti, C.; Doshi, M.; Ben, R. *Ice Recrystallization Inhibitors: From Biological Antifreezes to Small Molecules*; Wilson, P., Ed.; InTech: New York, NY, 2013.

(2) Fletcher, N. In *The Chemical Physics of Ice (1st. ed.)*; Cambridge University Press: London, 1970; pp 23–48.

(3) Devries, A. L.; Lin, Y. *Biochim. Biophys. Acta - Protein Struct.* **1977**, *495* (2), 388.

(4) Alley, R. B.; Perepezko, J. H.; Bentley, C. R. *J. Glaciol.* **1986**, *32* (112), 415.

(5) Pedersen, C.; Mihranyan, A.; Strømme, M. *PLoS One* **2011**, *6* (9).

(6) Fletcher, N. *Philos. Mag.* **1962**, *7* (74), 255.

(7) Fletcher, N. *Philos. Mag.* **1968**, *18* (156), 1287.

(8) Dosch, H.; Lied, A.; Bilgram, J. H. *Surf. Sci.* **1996**, *366* (1), 43.

(9) Döppenschmidt, A.; Butt, H. J. *Langmuir* **2000**, *16* (16), 6709.

(10) Sadtchenkoa, V.; Ewing, G. E. *J. Chem. Phys.* **2002**, *116* (11), 4686.

(11) Kahan, T. F.; Reid, J. P.; Donaldson, D. J. *J. Phys. Chem. A* **2007**, *111* (43), 11006.

(12) Goertz, M. P.; Zhu, X. Y.; Houston, J. E. *Langmuir* **2009**, *25* (12), 6905.

(13) Karim, O. A.; Haymet, A. D. J. *J. Chem. Phys.* **1988**, *89* (11), 6889.

(14) Pronk, P.; Infante Ferreira, C. A.; Witkamp, G. J. In *Journal of Crystal Growth*; 2005; Vol. 275.

(15) Budke, C.; Heggemann, C.; Koch, M.; Sewald, N.; Koop, T. *J. Phys. Chem. B* **2009**, *113* (9), 2865.

(16) Archer, D. L. *Int. J. Food Microbio.*, **2004**, *90*, 127–138.

(17) Adapa, S.; Schmidt, K. A.; Jeon, I. J.; Herald, T. J.; Flores, R. A. *Food. Rev. Int.* **2000**, *16* (3), 259.

(18) Regand, A.; Goff, H. D. *Food Hydrocoll.* **2003**, *17* (1), 95.

(19) Ben-Yoseph, E.; Hartel, R. W. *J. of Food Eng.*, **1998**, *38*, 309–329.

(20) Cook, K. L. K.; Hartel, R. W. *Compr. Rev. Food Sci. Food Saf.* **2010**, *9* (2), 213.

(21) Baust, J. M.; Vogel, M. J.; Van Buskirk, R.; Baust, J. G. *Cell Transplant.* **2001**, *10* (7), 561.

(22) Bissoyi, a; Nayak, B.; Pramanik, K.; Sarangi, S. K. *Biopreserv. Biobank.* **2014**, *12* (1), 23.

(23) Mazur, P.; Leibo, S. P.; Chu, E.H.Y, *Exp. Cell Research*, **1972**, 71(2), 345-335.

(24) Fuller, B. J. *Cryo-Letters*, **2004**, *25*, 375–388.

(25) a) Song, Y. C.; Khirabadi, B. S.; Lightfoot, F.; Brockbank, K. G.; Taylor, M. J. *Nat. Biotechnol.* **2000**, *18* (3), 296.

b) Pegg, D. E. *Methods Mol. Biol.* **2007**, *368*, 39.

(26) Davies, P. L.; Baardsnes, J.; Kuiper, M. J.; Walker, V. K. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **2002**, *357* (1423), 927.

(27) Barrett, J. *Int. J. of Biochem. and Cell Bio*., **2001**, *33*, 105–117.

(28) Kristiansen, E.; Zachariassen, K. E. *Cryobiology*, **2005**, *51*, 262–280.

(29) Raymond, J. A.; Wilson, P.; DeVries, A. L. *Proc. Natl. Acad. Sci. U. S. A.* **1989**, *86* (3), 881.

(30) Carpenter, J.; Hansen, T. *Proc. Natl. Acad. Sci.* **1992**, *89* (19), 8953.

(31) Chao, H.; Davies, P. L.; Carpenter, J. F. *J. Exp. Biol.* **1996**, *199* (Pt 9), 2071.

(32) Knight, C. A.; Hallett, J.; DeVries, A. L. *Cryobiology* **1988**, *25* (1), 55.

(33) Smallwood, M.; Worrall, D.; Byass, L.; Elias, L.; Ashford, D.; Doucet, C. J.; Holt, C.; Telford, J.; Lillford, P.; Bowles, D. J. *Biochem. J.* **1999**, *340*, 385.

(34) Jackman, J.; Noestheden, M.; Moffat, D.; Pezacki, J. P.; Findlay, S.; Ben, R. N. *Biochem. Biophys. Res. Commun.* **2007**, *354* (2), 340.

(35) Gibson, M. I.; Barker, C. A.; Spain, S. G.; Albertin, L.; Cameron, N. R. *Biomacromolecules* **2009**, *10* (2), 328.

(36) Tam, R. Y.; Rowley, C. N.; Petrov, I.; Zhang, T.; Afagh, N. A.; Woo, T. K.; Ben, R. N. *J. Am. Chem. Soc.* **2009**, *131* (43), 15745.

(37) Corcilius, L.; Santhakumar, G.; Stone, R. S.; Capicciotti, C. J.; Joseph, S.; Matthews, J. M.; Ben, R. N.; Payne, R. J. *Bioorg. Med. Chem.* **2013**, *21* (12), 3569.

(38) Van der Wal, S.; Capicciotti, C. J.; Rontogianni, S.; Ben, R. N.; Liskamp, R. M. J. *Medchemcomm* **2014**, *5* (8), 1159.

(39) Budke, C.; Koop, T. *ChemPhysChem* **2006**, *7* (12), 2601.

(40) Capicciotti, C.; Doshi, M.; Ben, R. *Ice Recrystallization Inhibitors: From Biological Antifreezes to Small Molecules*; **2013**.

(41) Tam, R. Y.; Ferreira, S. S.; Czechura, P.; Ben, R. N.; Chaytor, J. L. *J. Am. Chem. Soc.* **2008**, *130* (51), 17494.

(42) Czechura, P.; Tam, R. Y.; Dimitrijevic, E.; Murphy, A. V; Ben, R. N. *J. Am. Chem. Soc.* **2008**, *130* (10), 2928.

(43) Galema, S. A.; Hoiland, H. *J. Phys. Chem.* **1991**, *95* (13), 5321.

(44) Balcerzak, A. K.; Ferreira, S. S.; Trant, J. F.; Ben, R. N. *Bioorganic Med. Chem. Lett.* **2012**, *22* (4), 1719.

(45) Capicciotti, C.; Leclère, M.; Perras, F. *Chem. Sci.* **2012**, *3* (5), 1408.

(46) Balcerzak, A. K.; Febbraro, M.; Ben, R. N. *RSC Adv.* **2013**, *3* (10), 3232.

(47) Trant, J. F.; Biggs, R. A.; Capicciotti, C. J.; Ben, R. N. *RSC Adv.* **2013**, *3* (48), 26005.

(48) Capicciotti, C. J.; Kurach, J. D. R.; Turner, T. R.; Mancini, R. S.; Acker, J. P.; Ben, R. N. *Sci. Rep.* **2015**, *5*, 9692.

(49) Scott, K. L.; Lecak, J.; Acker, J. P. *Transfusion Medicine Reviews*, **2005**, *19*, 127–142.

(50) Hess, J. R.; Greenwalt, T. G. *Transfusion Medicine Reviews*, **2002**, *16*, 283–295.

(51) Hansch, C.; Leo, A.; Livingstone, D. J. *Pestic. Biochem. Physiol.* **1996**, *56* (1), 78.

(52) Wen, D.; Laursen, R. A. *J. Biol. Chem.* **1992**, *267* (20), 14102.

(53) Wen, D.; Laursen, R. A. *J. Biol. Chem.* **1993**, *268* (22), 16396.

(54) Zhang, W.; Laursen, R. A. *J. Biol. Chem.* **1998**, *273* (52), 34806.

(55) Perkins, R.; Fang, H.; Tong, W.; Welsh, W. J. *Environ. Toxicol. Chem.* **2003**, *22* (8), 1666.

(56) James, S. L. *Chem. Soc. Rev.* **2003**, *32* (5), 276.

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

(58) Yaghi, O. M.; Li, G.; Li, H. *Nature*, **1995**, *378*, 703–706.

(59) Abrahams, B. F.; Hoskins, B. F.; Michail, D. M.; Robson, R. *Nature*, **1994**, *369*, 727–729.

(60) Gardner, G. B.; Venkataraman, D.; Moore, J. S.; Lee, S. *Nature*, **1995**, *374*, 792–795.

(61) Subramanian, S.; Zaworotko, M. J. *Angew. Chemie, Int. Ed.* **1995**, *34* (19), 2127.

(62) Rees, L. V. C. Introduction to Zeolite Science and Practice. *Zeolites*, 1992, *12*, 767.

(63) Kulprathipanja, S. *Zeolites in Industrial Separation and Catalysis*; 2010.

(64) Vermeiren, W.; Gilson, J. P. *Top. Catal.* **2009**, *52* (9), 1131.

(65) Wang, S.; Peng, Y. *Chem. Eng. Journal*, **2010**, *156*, 11–24.

(66) Cambridge Crystalogrphic Database Centre. How many MOFs are there in the CSD? http://www.ccdc.cam.ac.uk/SupportandResources/Support/pages/SupportSolution.aspx?supportsolutionid=275 (accessed May 8, 2015).

(67) International Zeolite Association. Database of Zeolite Structures http://www.iza-structure.org/databases/ (accessed May 8, 2015).

(68) Thomson Reuters. Web of Knowledge Search http://apps.webofknowledge.com (accessed May 8, 2015).

(69) Yaghi, O. M.; O'Keeffe, M.; Ockwig, N. W.; Chae, H. K.; Eddaoudi, M.; Kim, J. *Nature* **2003**, *423* (6941), 705.

(70) Dunbar, K. R.; Heintz, R. A. In *Progress in Inorganic Chemistry*; 1997; pp 283–391.

(71) Ling, Y.; Chen, Z.-X.; Zhai, F.-P.; Zhou, Y.-M.; Weng, L.-H.; Zhao, D.-Y. *Chem. Commun. (Camb).* **2011**, *47* (25), 7197.

(72) Zhong, R. Q.; Zou, R. Q.; Pandey, D. S.; Kiyobayashi, T.; Xu, Q. *Inorg. Chem. Commun.* **2008**, *11* (9), 951.

(73) Ortiz, A. U.; Boutin, A.; Fuchs, A. H.; Coudert, F.-X. *J. Phys. Chem. Lett.* **2013**, 1861.

(74) Fang, Q. R.; Zhu, G. S.; Xue, M.; Zhang, Q. L.; Sun, J. Y.; Guo, X. D.; Qiu, S. L.; Xu, S. T.; Wang, P.; Wang, D. J.; Wei, Y. *Chem. - A Eur. J.* **2006**, *12* (14), 3754.

(75) Song, X.; Kim, T. K.; Kim, H.; Kim, D.; Jeong, S.; Moon, H. R.; Lah, M. S. *Chem. Mater.* **2012**, *24* (15), 3065.

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

(77) Deng, H.; Grunder, S.; Cordova, K. E.; Valente, C.; Furukawa, H.; Hmadeh, M.; Gándara, F.; Whalley, A. C.; Liu, Z.; Asahina, S.; Kazumori, H.; O'Keeffe, M.; Terasaki, O.; Stoddart, J. F.; Yaghi, O. M. *Science* **2012**, *336* (6084), 1018.

(78) 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.

(79) Ma, S.; Wang, X.-S.; Collier, C. D.; Manis, E. S.; Zhou, H.-C. *Inorg. Chem.* **2007**, *46* (21), 8499.

(80) Li, J.-R.; Sculley, J.; Zhou, H.-C. *Chem. Rev.* **2012**, *112* (2), 869.

(81) Düren, T.; Sarkisov, L.; Yaghi, O. M.; Snurr, R. Q. *Langmuir* **2004**, *20* (7), 2683.

(82) Vaidhyanathan, R.; Iremonger, S. S.; Shimizu, G. K. H.; Boyd, P. G.; Alavi, S.; Woo, T. K. *Science* **2010**, *330* (6004), 650.

(83) Ma, S.; Sun, D.; Wang, X.-S.; Zhou, H.-C. *Angew. Chemie Int. Ed.* **2007**, *46* (14), 2458.

(84) Dinca, M.; Yu, A. F.; Long, J. R. *J. Am. Chem. Soc.* **2006**, *128* (27), 8904.

(85) Forrest, K. A.; Pham, T.; Hogan, A.; McLaughlin, K.; Tudor, B.; Nugent, P.; Burd, S. D.; Mullen, A.; Cioce, C. R.; Wojtas, L.; Zaworotko, M. J.; Space, B. *J. Phys. Chem. C* **2013**, *117* (34), 17687.

(86) McCormick, L. J.; Duyker, S. G.; Thornton, A. W.; Hawes, C. S.; Hill, M. R.; Peterson, V. K.; Batten, S. R.; Turner, D. R. *Chem. Mater.* **2014**, *26* (15), 4640.

(87) Wu, X.; Yuan, B.; Bao, Z.; Deng, S. *J. Colloid Interface Sci.* **2014**, *430*, 78.

(88) Kim, T. K.; Suh, M. P. *Chem. Commun. (Camb).* **2011**, *47* (14), 4258.

(89) Modrow, A.; Zargarani, D.; Herges, R.; Stock, N. *Dalton Trans.* **2011**, *40* (16), 4217.

(90) Liu, Y.; Her, J. H.; Dailly, A.; Ramirez-Cuesta, A. J.; Neumann, D. A.; Brown, C. M. *J. Am. Chem. Soc.* **2008**, *130* (35), 11813.

(91) Ito, H.; Muromoto, M.; Kurenuma, S.; Ishizaka, S.; Kitamura, N.; Sato, H.; Seki, T. *Nat. Commun.* **2013**, *4* (May), 2009.

(92) Horike, S.; Shimomura, S.; Kitagawa, S. *Nat. Chem.* **2009**, *1* (9), 695.

(93) Kitagawa, S.; Kondo, M. *Bull. Chem. Soc. Jpn.* **1998**, *71* (8), 1739.

(94) Férey, G.; Serre, C. *Chem. Soc. Rev.* **2009**, *38* (5), 1380.

(95) Schneemann, a; Bon, V.; Schwedler, I.; Senkovska, I.; Kaskel, S.; Fischer, R. a. *Chem. Soc. Rev.* **2014**, 6062.

(96) Boutin, A.; Coudert, F.-X.; Springuel-Huet, M.-A.; Neimark, A. V; Férey, G.; Fuchs, A. H. *J. Phys. Chem. C* **2010**, *114* (50), 22237.

(97) Jia, J.; Xu, F. J.; Long, Z.; Hou, X. D.; Sepaniak, M. J. *Chem. Commun.* **2013**, *49* (41), 4670.

(98) Mulder, F. M.; Assfour, B.; Huot, J.; Dingemans, T. J.; Wagemaker, M.; Ramirez-Cuesta, A. J. *J. Phys. Chem. C* **2010**, *114* (23), 10648.

(99) Chen, L.; Mowat, J. P. S.; Fairen-Jimenez, D.; Morrison, C. A.; Thompson, S. P.; Wright, P. A.; Düren, T. *J. Am. Chem. Soc.* **2013**, *135* (42), 15763.

(100) Volkringer, C.; Loiseau, T.; Guillou, N.; Férey, G.; Elkaïm, E.; Vimont, A. *Dalton Trans.* **2009**, No. 12, 2241.

(101) Mellot-Draznieks, C.; Serre, C.; Surblé, S.; Audebrand, N.; Férey, G. *J. Am. Chem. Soc.* **2005**, *127* (46), 16273.

(102) Murdock, C. R.; McNutt, N. W.; Keffer, D. J.; Jenkins, D. M. *J. Am. Chem. Soc.* **2014**, *136* (2), 671.

(103) Bureekaew, S.; Sato, H.; Matsuda, R.; Kubota, Y.; Hirose, R.; Kim, J.; Kato, K.; Takata, M.; Kitagawa, S. *Angew. Chem. Int. Ed. Engl.* **2010**, *49* (42), 7660.

(104) Maji, T. K.; Matsuda, R.; Kitagawa, S. *Nat. Mater.* **2007**, *6* (2), 142.

(105) Kanoo, P.; Matsuda, R.; Higuchi, M.; Kitagawa, S.; Maji, T. K. *Chem. Mater.* **2009**, *21* (24), 5860.

(106) Kitaura, R.; Seki, K.; Akiyama, G.; Kitagawa, S. *Angew. Chemie - Int. Ed.* **2003**, *42* (4), 428.

(107) Getman, R. B.; Bae, Y.-S.; Wilmer, C. E.; Snurr, R. Q. *Chem. Rev.* **2012**, *112* (2), 703.

(108) 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* (2), 724.

(109) Suh, M. P.; Park, H. J.; Prasad, T. K.; Lim, D.-W. *Chem. Rev.* **2012**, *112* (2), 782.

(110) Wu, H.; Gong, Q.; Olson, D. H.; Li, J. *Chem. Rev.* **2012**, *112* (2), 836.

(111) Bétard, A.; Fischer, R. A. *Chem. Rev.* **2012**, *112* (2), 1055.

(112) Wang, C.; Zhang, T.; Lin, W. *Chem. Rev.* **2012**, *112* (2), 1084.

(113) Kreno, L. E.; Leong, K.; Farha, O. K.; Allendorf, M.; Van Duyne, R. P.; Hupp, J. T. *Chem. Rev.* **2012**, *112* (2), 1105.

(114) Cui, Y.; Yue, Y.; Qian, G.; Chen, B. *Chem. Rev.* **2012**, *112* (2), 1126.

(115) Yoon, M.; Srirambalaji, R.; Kim, K. *Chem. Rev.* **2012**, *112* (2), 1196.

(116) Horcajada, P.; Gref, R.; Baati, T.; Allan, P. K.; Maurin, G.; Couvreur, P.; Férey, G.; Morris, R. E.; Serre, C. *Chem. Rev.* **2012**, *112* (2), 1232.

(117) World Energy Council. *2015 World Energy Issues Monitor*; London, UK, 2015.

(118) US Department of Commerce, NOAA, E. S. R. L. ESRL Global Monitoring Division - Global Greenhouse Gas Reference Network http://www.esrl.noaa.gov/gmd/ccgg/trends/global.html (accessed May 11, 2015).

(119) Solomon, S.; Plattner, G.-K.; Knutti, R.; Friedlingstein, P. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (6), 1704.

(120) Statistics Canada. CANSIM - 127-0002 - Electric power generation, by class of electricity producer, monthly (megawatt hour) http://www5.statcan.gc.ca/cansim/a47 (accessed May 11, 2015).

(121) CEPA Environmental Registry. Reduction of Carbon Dioxide Emissions from Coal-Fired Generation of Electricity Regulations https://www.ec.gc.ca/lcpe-cepa/eng/regulations/detailreg.cfm?intReg=209 (accessed May 11, 2015).

(122) IPCC. *IPCC special report on carbon dioxide capture and storage*; 2005; Vol. 2.

(123) Rochelle, G. T. *Science* **2009**, *325* (5948), 1652.

(124) Romeo, L. M.; Bolea, I.; Escosa, J. M. *Appl. Therm. Eng.* **2008**, *28* (8-9), 1039.

(125) Yeh, J. T.; Resnik, K. P.; Rygle, K.; Pennline, H. W. In *Fuel Processing Technology*; **2005**; Vol. 86, pp 1533–1546.

(126) Wang, Q. M.; Shen, D.; Bülow, M.; Lau, M. L.; Deng, S. *Microporous Mesoporous Mater.* **2002**, *55*, 217.

(127) Mu, B.; Walton, K. S. *J. Phys. Chem. C* **2011**, *115* (46), 111027103506007.

(128) Weiland, R. H.; Dingman, J. C.; Cronin, D. B. *J. Chem. Eng. Data* **1997**, *42* (5), 1004.

(129) Herm, Z. R.; Swisher, J. A.; Smit, B.; Krishna, R.; Long, J. R. *J. Am. Chem. Soc.* **2011**, *133* (15), 5664.

(130) Mississippi Power. Kemper Project http://www.mississippipower.com/about-energy/plants/kemper-county-energy-facility/home.cshtml (accessed May 11, 2015).

(131) Southern California Carbon Sequestration Research Consortium. Hydrogen Energy California Project (HECA) http://www.socalcarb.org/heca.html (accessed May 11, 2015).

(132) Summit Power. Texas Clean Energy Project http://www.texascleanenergyproject.com/ (accessed May 11, 2015).

(133) Plasynski, S. I.; Litynski, J. T.; McIlvried, H. G.; Srivastava, R. D. Progress and New Developments in Carbon Capture and Storage. *Critical Reviews in Plant Sciences*, 2009, *28*, 123–138.

(134) Jassim, M. S.; Rochelle, G. T. In *Industrial and Engineering Chemistry Research*; 2006; Vol. 45, pp 2465–2472.

(135) Granite, E. J.; Pennline, H. W.; Hargis, R. A. *Ind. Eng. Chem. Res.* **2000**, *39* (4), 1020.

(136) Drage, T. C.; Snape, C. E.; Stevens, L. A.; Wood, J.; Wang, J.; Cooper, A. I.; Dawson, R.; Guo, X.; Satterley, C.; Irons, R. *Journal of Materials Chemistry*, **2012**, *22*, 2815.

(137) Kather, A.; Scheffknecht, G. *Naturwissenschaften*, **2009**, *96*, 993–1010.

(138) Rezvani, S.; Huang, Y.; McIlveen-Wright, D.; Hewitt, N.; Wang, Y. *Fuel* **2007**, *86* (14 SPEC. ISS.), 2134.

(139) Cornelissen, R. L.; Hirs, G. G. *Energy Conversion and Management*, **1998**, *39*, 1821–1826.

(140) Tlili, N.; Grévillot, G.; Vallières, C. *Int. J. Greenh. Gas Control* **2009**, *3* (5), 519.

(141) Mulgundmath, V.; Tezel, F. H. In *Adsorption*; **2010**; Vol. 16, pp 587–598.

(142) Li, T.; Chen, D.-L.; Sullivan, J. E.; Kozlowski, M. T.; Johnson, J. K.; Rosi, N. L. *Chem. Sci.* **2013**, *4* (4), 1746.

(143) Li, B.; Wen, H. M.; Wang, H.; Wu, H.; Tyagi, M.; Yildirim, T.; Zhou, W.; Chen, B. *J. Am. Chem. Soc.* **2014**, *136* (17), 6207.

(144) 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* (7), 633.

(145) Li, J.-R.; Sculley, J.; Zhou, H.-C. *Chem. Rev.* **2012**, *112* (2), 869.

(146) BAE, Y.; LEE, C. *Carbon N. Y.* **2005**, *43* (1), 95.

(147) Zhang, W.; Huang, H.; Zhong, C.; Liu, D. *Phys. Chem. Chem. Phys.* **2012**, *14* (7), 2317.

(148) 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* (2), 724.

(149) Schoenecker, P. M.; Carson, C. G.; Jasuja, H.; Flemming, C. J. J.; Walton, K. S. *Ind. Eng. Chem. Res.* **2012**, *51* (18), 6513.

(150) Dinča, M.; Dailly, A.; Liu, Y.; Brown, C. M.; Neumann, D. A.; Long, J. R. *J. Am. Chem. Soc.* **2006**, *128* (51), 16876.

(151) Zhou, W.; Wu, H.; Yildirim, T. *J. Am. Chem. Soc.* **2008**, *130* (46), 15268.

(152) Fracaroli, A. M.; Furukawa, H.; Suzuki, M.; Dodd, M.; Okajima, S.; Gándara, F.; Reimer, J. A.; Yaghi, O. M. *J. Am. Chem. Soc.* **2014**, *136* (25), 8863.

(153) McDonald, T. M.; Mason, J. A.; Kong, X.; Bloch, E. D.; Gygi, D.; Dani, A.; Crocellà, V.; Giordanino, F.; Odoh, S. O.; Drisdell, W. S.; Vlaisavljevich, B.; Dzubak, A. L.; Poloni, R.; Schnell, S. K.; Planas, N.; Lee, K.; Pascal, T.; Wan, L. F.; Prendergast, D.; Neaton, J. B.; Smit, B.; Kortright, J. B.; Gagliardi, L.; Bordiga, S.; Reimer, J. A.; Long, J. R. *Nature* **2015**, *519* (7543), 303.

(154) Campañá, C.; Mussard, B.; Woo, T. K. *J. Chem. Theory Comput.* **2009**, *5* (10), 2866.

(155) Kim, D.; Lee, T. B.; Choi, S. B.; Yoon, J. H.; Kim, J.; Choi, S. H. *Chem. Phys. Lett.* **2006**, *420* (1-3), 256.

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

(157) Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. *J. Am. Chem. Soc.* **1992**, *114* (25), 10024.

(158) Addicoat, M. A.; Vankova, N.; Akter, I. F.; Heine, T. *J. Chem. Theory Comput.* **2014**, *10* (2), 880.

(159) Zhao, L.; Yang, Q.; Ma, Q.; Zhong, C.; Mi, J.; Liu, D. *J. Mol. Model.* **2011**, *17* (2), 227.

(160) Han, S. S.; Kim, D.; Jung, D. H.; Cho, S.; Choi, S. H.; Jung, Y. *J. Phys. Chem. C* **2012**, *116* (38), 20254.

(161) Vanduyfhuys, L.; Verstraelen, T.; Vandichel, M.; Waroquier, M.; Van Speybroeck, V. *J. Chem. Theory Comput.* **2012**, *8* (9), 3217.

(162) Yang, Q.; Zhong, C. *J. Phys. Chem. B* **2006**, *110* (36), 17776.

(163) Düren, T.; Bae, Y.-S.; Snurr, R. Q. *Chem. Soc. Rev.* **2009**, *38* (5), 1237.

(164) Yang, Q.; Liu, D.; Zhong, C.; Li, J.-R. *Chem. Rev.* **2013**, *113* (10), 8261.

(165) Skoulidas, A. I. *J. Am. Chem. Soc.* **2004**, *126* (5), 1356.

(166) Lin, L. C.; Kim, J.; Kong, X.; Scott, E.; McDonald, T. M.; Long, J. R.; Reimer, J. A.; Smit, B. *Angew. Chemie - Int. Ed.* **2013**, *52* (16), 4410.

(167) Parkes, M. V.; Demir, H.; Teich-Mcgoldrick, S. L.; Sholl, D. S.; Greathouse, J. A.; Allendorf, M. D. *Microporous Mesoporous Mater.* **2014**, *194*, 190.

(168) Mellot Draznieks C; Newsam, J.; Gorman, A.; Freeman, C.; Férey, G. *Angew. Chem. Int. Ed. Engl.* **2000**, *39* (13), 2270.

(169) Fernandez, M.; Boyd, P. G.; Daff, T. D.; Aghaji, M. Z.; Woo, T. K. *J. Phys. Chem. Lett.* **2014**, *5* (17), 3056.

(170) Fernandez, M.; Woo, T. K.; Wilmer, C. E.; Snurr, R. Q. *J. Phys. Chem. C* **2013**, *117* (15), 7681.

(171) Fernandez, M.; Trefiak, N. R.; Woo, T. K. *J. Phys. Chem. C* **2013**, *117* (27), 14095.

2. Theoretical Methods

Often times when it comes to computational chemistry, results seem almost like magic to those who are not intimately involved with the intricacies of each grueling calculation. There are usually many hours of mental acrobatics that go into the determination of a single number. This chapter hopes to shed some theoretical background coupled with some historical context in order to give the reader an idea of why each method was appropriate for each application. The organization of this chapter will be split into two parts, first theoretical methods necessary for the ice recrystallization inhibitor project will be discussed, and secondly theoretical methods utilized for the MOF based projects will be displayed.

2.1. Ice Recrystallization Inhibition Theoretical Methods

2.1.1. Quantum Mechanics

The physical laws of motion, as proposed by Isaac Newton, are the foundation for classical mechanics and allow one to solve for the *motion* of a system. Unfortunately, such laws fail when applied to describing the properties and behavior of microscopic particles (e.g. atoms and molecules). The theories of quantum mechanics are the physical laws of motion that govern microscopic particles. One of the postulates of quantum mechanics states that the state of a quantum mechanical system is completely defined by the state wave function $\Psi(r,t)$ which depends on the particle's position and time.[1] The wave function contains all the information about the system and by solving for it one can attain the values for any observable. The time-

independent Schrödinger equation (**Eq. 2.1**) relates the wave function to the energy of the system and can be written as,

$$\hat{\mathrm{H}}\psi(r) = E\psi(r) \qquad \textbf{(Eq. 2.1)}$$

where ψ is the time independent wave function, E is the energy, and the Hamiltonian ($\hat{\mathrm{H}}$) is also now independent of time. The Hamiltonian is the operator corresponding to the total energy of the system. While there have been many approximations to the Schrödinger equation it can only be solved exactly for one electron systems such as hydrogen.

2.1.1.1. Semi-Empirical Method

The Hartree-Fock method (HF) is the basis for many computational chemistry methods.[2] Importantly, in order to solve the Schrödinger equation the method treats the interactions between electrons in an average way. Such an approximation allows the replacement of the electron-electron repulsion term within the electron Hamiltonian with a one-electron potential energy term. In particular, each electron moves within an averaged charge density field from the other electrons. Thus, it is not surprising that the main drawback is that HF does not properly account for the instantaneous interactions between electrons. This instantaneous interaction is called electron correlation and is very important in bond breaking or formation processes and in weakly interacting systems. Another large drawback is the computational cost due to manipulating and calculating complicated integrals.

The Semi-Empirical Method is a way of approximating or neglecting some of the complicated integrals by parameterizing them to experimental data. In semi-empirical methods, only the valence electrons of the system are considered since these are the electrons most

important to chemical reactivity. The core electrons are to be included into the nuclear core and this reduces the integrals which correspond to core electrons. Next, semi-empirical parameterizes some integrals based on experimental data of specific atoms. Different semi-empirical methods parameterize to different properties, for example the modified neglect of diatomic overlap (MNDO) method is based on spectroscopic data of H, C, N, and O,[3] the Austin Model 1 (AM1) emphasizes dipole moments, ionization potentials, and geometries of molecules,[4] and the Parametric Method 3 (PM3) was parameterized to a large number of chemical properties such as heat of formations.[5] The greatest advantage of semi-empirical methods is their speed of calculation which results in even large systems being calculated. However, the accuracy of semi-empirical methods can be questionable with respect to exotic systems which differ greatly from the set of structures used in parameterization.

In the context of screening for the optimal geometry of small ice recrystallization inhibitors, one requires a method which has a good balance between accuracy and computational cost. Unsurprisingly, semi-empirical methods where the first quantum mechanical methods to be used in the study of antifreeze proteins (AFPs). In 2002 Cheng *et al.* studied the binding of a type II antifreeze protein from sea raven to ice and calculated their interaction energies using the AM6 and PM3 methods.[6] They also found computational results revealed a weakening of the ice lattice upon AFP binding. Since then, semi-empirical methods have been used to study the AFP-ice-solvent system,[7] the role of nonpolar amino acid functional groups in natural and synthetic AFPs,[8] and size dependence of AFPs with respect to interactions with ice.[9]

2.1.1.2. Density Functional Theory

The fundamental concept of Density Functional Theory (DFT) is that the ground-state energy of the system can be calculated from the total electron density. In order to calculate the energy from the density the Kohn-Sham formalism was developed.[10] A key step was to assume that the total electronic energy can be represented as a summation of energy terms each of which were dependent on the electron density. Thus the exact ground-state energy of a molecule based off the electron density could be written as: **(Eq. 2.2)**

$$E_{exact} = E^T + E^V + E^J + E^{XC} \quad \textbf{(Eq. 2.2)}$$

where E^T is the kinetic energy due to the movement of electrons, E^V is the nuclear-electron attraction potential energy term, E^J is the electron-electron repulsion term, and E^{XC} is called the exchange-correlation. Notably, the E^{XC} is the only unknown in the above expression and must be approximated. Additionally, E^{XC} can be expressed as the summation of two terms, the exchange term E^X and the correlation term E^C. As well as devising a way to break up the exact energy, the Kohn-Sham theorem also states that the exact ground-state density could be represented as a linear combination of one-electron densities called the Kohn-Sham orbitals. A basis set is a set of mathematical functions used to form linear combinations to describe the Kohn-Sham orbitals. With respect to molecular systems, atom localized Gaussian-type functions is used. To evaluate the energy from the electron density various functionals of the electron density have been developed.

Perhaps the most widely used functional is the exchange-correlation functional B3LYP which is known as a hybrid functional because it contains elements of Hartree-Fock as well as

Becke's gradient-corrected exchange terms.[11] The B3LYP method is ubiquitous in the study of biological systems[12] and is typically employed in calculating the reaction pathway for various enzymes.[13] In context to antifreeze molecules, B3LYP has actually been used to optimize geometries of various polyols in a structure-activity relationship (QSAR) study of carbohydrates and antifreeze activity.[14] Indeed, B3LYP methods have been extensively used to optimize geometries of compounds in the use of QSAR to model trends in toxicity of nitroaromatics,[15] antimalarial activity,[16] and antioxidant activity[17] to just name a few. While B3LYP shows good chemical accuracy, the evaluation of the Hartree-Fock exchange term is prohibitively expensive for periodic calculations on solid state systems. Thus the functional of choice for solid state physics is the Perdew-Burke-Ernzerhof (PBE) exchange correlation functional[18] and its other variants which have seen widespread use in materials research.

2.1.2. Molecular Mechanics

In Molecular Mechanics (MM) each molecule is modeled as an assortment of balls (atoms) attached together by springs (bonds). A chemical bond, much like a spring, has a regular length and resistance to being stretched, bent, or twisted. The main issue with MM is that it does not consider electrons but rather treats an atom as a ball of charge. Thus it is not capable of looking at electronic properties like bond breaking or formation. On the other hand, MM methods are very fast in terms of calculation time and can handle larger systems in comparison to higher level *ab-initio* calculations such as the semi-empirical method.

2.1.2.1. Force Fields

A forcefield is a collection of mathematical terms and parameters used to describe the potential energy of a molecule. These functions and parameters can be derived from experiment (semi-empirical forcefields) or high level quantum mechanical calculations (ab initio forcefields). The functional form of a forcefield is a summation of the energy of bonded and non-bonded atoms. The bonded, or intramolecular terms are typically that of energy contributions due to bonds stretching, angle bending, and torsional twisting. The non-bonded, or intermolecular terms include weak forces or van der Waals interactions, and electrostatic contributions due to charges. More sophisticated terms exist which include added parameters for specific systems.

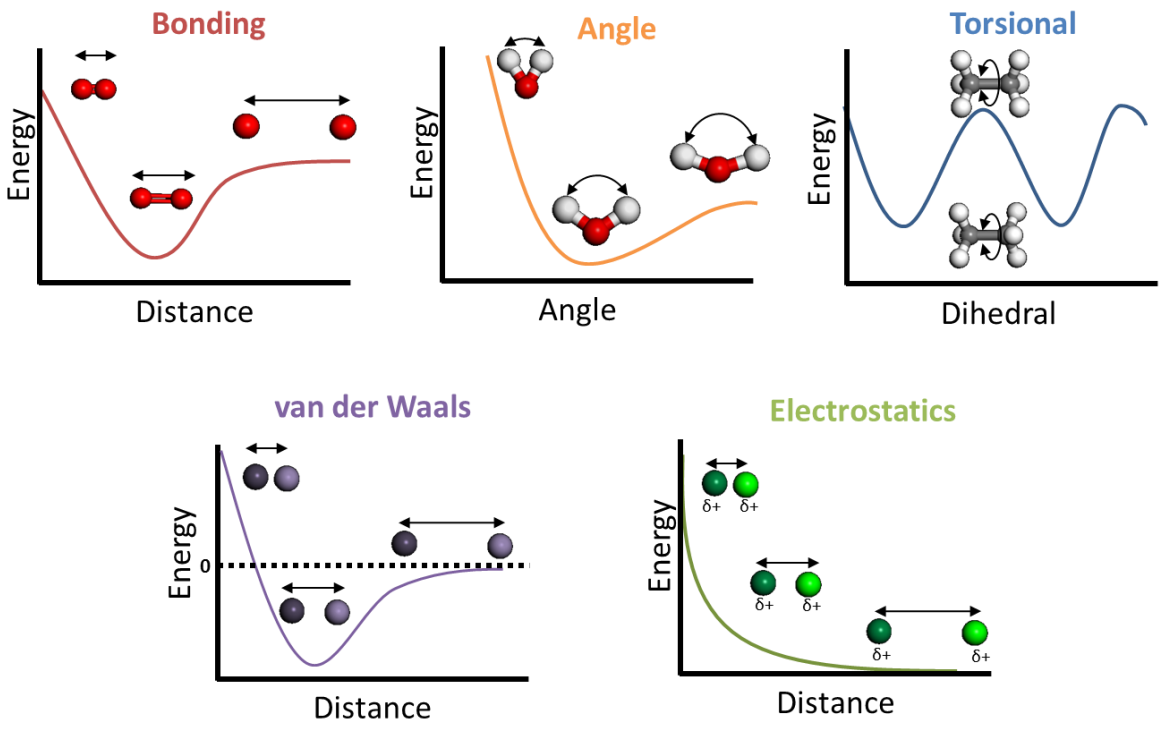


Figure 2.1: The intramolecular (bonding, angle, and torsional) potential energy functions and the intermolecular (van der Waals and electrostatic) potential energy functions.

A common functional form for the potential energy *V(r)* that serves as the basis of many forcefields used is, (**Eq. 2.3**)

$$V(r^{(i)}) = \sum_{bonds} k_b(l - l_0)^2 + \sum_{angles} k_a(\theta - \theta_0)^2 + \sum_{torsions} \frac{1}{2} V_n[1 + \cos(n\omega - \gamma)] + \sum_{j=1}^{N-1} \sum_{i=j+1}^{N} \left\{ 4\varepsilon_{i,j} \left[\left(\frac{\varepsilon_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \right\} \quad \textbf{(Eq. 2.3)}$$

- First term (Bonding Energy): This harmonic potential accounts for the energy between covalently bonded atoms. K_b is the stretching constant, *l* is the bond length, and l_0 is the equilibrium bond length.
- Second term (Angle Energy): This harmonic potential accounts for the energy of angle distortion. K_a is the angle force constant, ϑ is the bond angle, and ϑ_0 is the equilibrium bond angle.
- Third term (Torsional Energy): This cosine potential accounts for the energy of torsional rotation. V_n is often referred to as the barrier height of the potential, or the maximum energy value of the cosine function. *n* is the multiplicity which determines the number of minimum points in the function within a 360° rotation. ω is the torsional angle and γ is the phase factor which determines the angle at which the minimum energy value is.
- Fourth Term (van der Waals Energy): This Lennard-Jones 12-6 function accounts for the energy due to weak dispersive forces as well as short range repulsive forces due to steric interactions. This potential is the most popular form due to only two parameters, σ (separation where the energy is zero) and ε (the well depth). r is the separation between the two atoms. There are other potential forms such as the Buckingham potential which is more accurate but also has more parameters.

- Fourth Term (Electrostatic Energy): This Coulomb law function accounts for the electrostatic energy due to the point charges which rise from unequal electron distribution. ε_0 is the electric constant, q is the partial atomic charge on each atom, and r is the distance between atoms.

2.1.3. Conformational Searching

The conformation of a molecule – the arrangements of atoms in space purely by torsional rotation about single bonds, is crucial to its physical, chemical, and biological properties. Conformational searching is a way to find a molecule's preferred orientation which usually means the most stable and lowest energy conformation. The importance of molecular conformation can be traced back to the foundations of organic stereochemistry. In 1950, D.H.R Barton first published work which showed the reactivity of substituted cyclohexanes was highly dependent on the equatorial or axial position of the substituents.[18] This was one of the first cases which confirmed that the conformation of a molecule affected its chemical reactivity. It has since been well established that the stereochemistry is extremely important to a molecule's reactivity – an example is a flexible drug molecule which can alter its conformation through bond rotation to fit into a protein's active site.

The basis of a conformational search is finding energy minimum points on the potential energy surface (PES) of a molecule. Take for example the prototypical case of butane which which can exist in the *fully eclipsed-, gauche-, eclipsed-,* and *anti-* conformations based on the rotation of the dihedral bond. In this simple case calculating the energy as function of dihedral angle would be trivial and it would be found that the lowest energy conformation is the *anti-*

conformer where the methyl groups are staggered as far away from each other as possible. However, for larger molecules the presence of many dihedral bonds exponentially increase the search space. It would be ideal to determine all energy minimum points on the PES in a conformational search but some molecules are of such complexity that this is not possible. Furthermore, finding the global minimum, or the lowest overall energy point on the PES, is not necessarily the best answer. While this conformation might be the lowest in energy, perhaps it is not the "active" conformation. For example, a flexible drug molecule may exist in one global mininmum conformation in aqueous solution but when bound to an enzyme active site it may alter its orientation to a different local energy minimum which better binds to the enzyme. Thus, it is very important to sample and keep multiple conformations for analysis during the search.

Conformational searching can be traditionally divided into a few categories: systematic search algorithms, model building models, distance geometry, and random stochastic approaches. A brief description of the methods will be provided with a more in-depth discussion on the Monte Carlo random approach which was utilized for the IRI project. There have been many more conformational searching techniques developed which utilize genetic algorithms,[19] molecular dynamics,[20] and potential energy surface contour tracing[21] to name a few. For the sake of brevity they will not be discussed here.

Systematic: A systematic search makes sequential changes to the conformation by rotating torsional bonds by some fixed increment with all bond lengths and angles fixed. In this algorithm after every change in the torsional angle the geometry is energy minimized to obtain the associated minimum energy conformation. The search is terminated when all possible

dihedral angle combinations have been generated and minimized. An example would be a grid search of an alanine dipeptide, a common amino acid. This molecule only has two torsional angles denoted ϕ and ψ which can vary. A systematic grid search yields a Ramachandran map which is a common way to visualize the flexibility of proteins.[22] A figure of this Ramachandran map is shown in Figure 2.2. Systematic searches are typically reserved for only very small molecules as the cost of calculation experiences a factorial increase with the size of molecule.

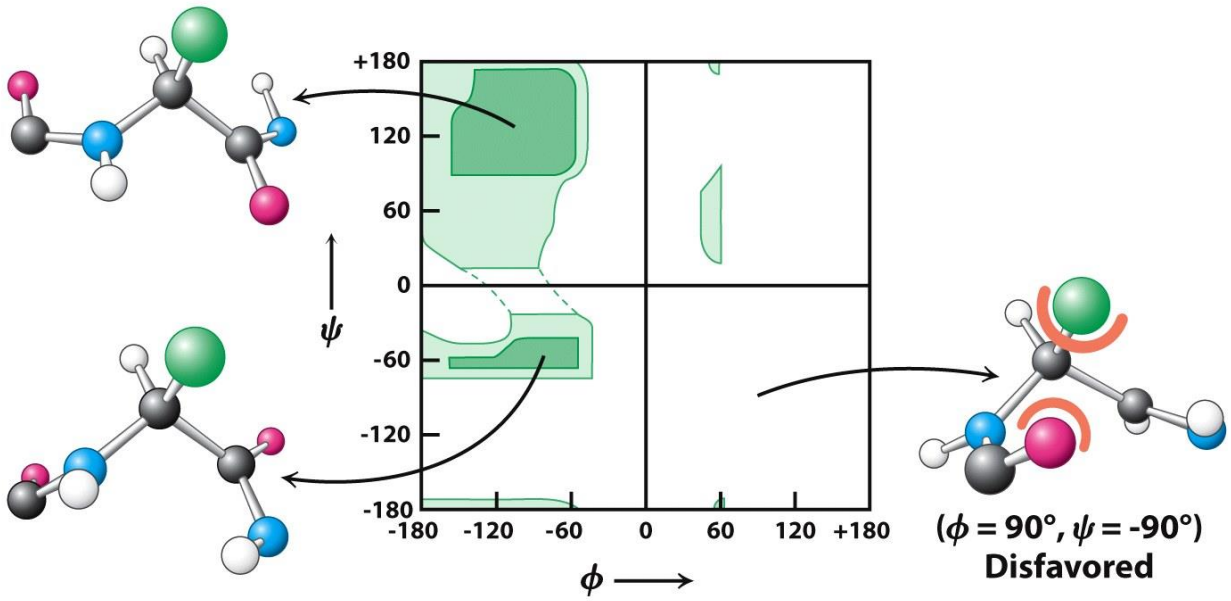


Figure 2.2: Ramachadran plot of alanine dipeptide which shows the energy contour plot based on dihedral rotation. Dark green are the lowest energy configurations, light green are 2.0 kcal/mol higher in energy and white are unflavoured regions where optimized conformations were not found. A systematic search would be used in this case to test every single dihedral angle conformation. Figure taken from the undergraduate textbook *Biochemistry, 7th Ed.*[23]

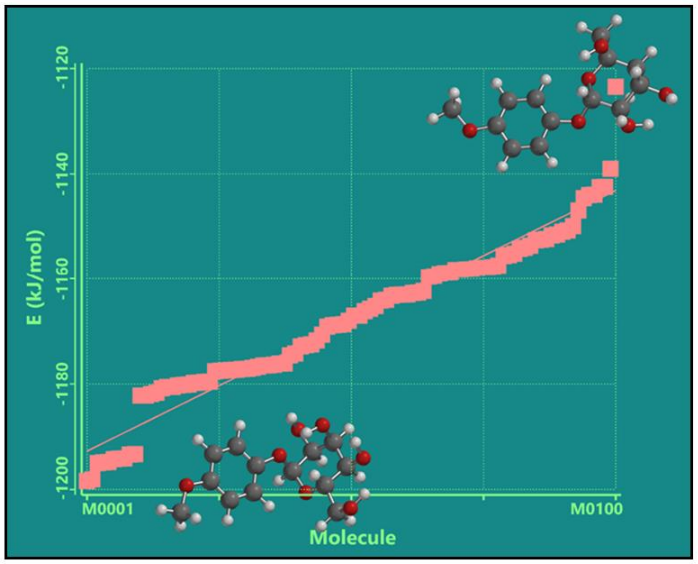
Model Building: In a model building approach, the conformational search is done through a bottom up approach by building a molecule from molecular fragments.[24] Fragments are joined at dihedral bonds which allow for the partitioning of the molecule for energy minimizations. Here the molecular fragments are optimized before being connected. They are then connected at various dihedral angles and a systematic search is performed on that specific

torsion. This approach limits the number of dihedral angles needed to be tested and is particularly useful for cyclic fragments which have a more limited number of conformations. Automated model building approaches have been widely used into prediction of large protein structures such as Lysozyme and Cyanase.[25]

Distance Geometry: Distance geometry, first introduced by Crippen and Havel in 1983, uses the distance between all pairs of atoms to describe the structure of the molecule.[26] There are $N(N-1)/2$ interatomic distances in a molecule which can be represented using an $N \times N$ matrix. Distance geometry randomly generates various distance matrices which are converted into a conformation. Of course, many randomly generated combinations of distances are geometrically invalid and a special mathematical algorithm must be employed. First a matrix of upper and lower bounds of the intermolecular distance is calculated. Next a procedure called *triangle smoothing* refines this matrix. Finally a set of matrix operations called *embedding* is performed to turn this matrix into Cartesian coordinate form.

Random Search: One random search algorithm utilizes a Monte Carlo scheme in order to search the conformational space. First we start with an initial structure with randomly generated dihedral angles. Then an energy minimization is done and the energy of this conformation is calculated. Next the dihedral angels are randomly perturbed and then the resulting conformation is energy minimized. If the energy of the new conformation is lower, it is automatically accepted, if the energy is higher, then this structure is either accepted or rejected based on whether the acceptance criterion (which is the Boltzmann factor, $e^{-\Delta E/kT}$, of the energy difference between the initial and generated structure) is higher than a random number between 0 and 1. This is done until no new structures can be obtained or if the difference of

energy between the new and old conformation is below a certain cut-off. Monte Carlo conformational searches are now established methods with widespread use in protein folding[27,28] and other biological systems. Monte Carlo sampling is only one type of Random Search and there exist others which are better for finding global minima. In the context of the IRI work, the conformational search was implemented on a small flexible IRI molecule as shown below in Figure 2.3. Included is a read out of the generated energy data from the conformational search. Notice that the energy starts very low (-1194.13 kJ/mol) then the second conformer is generated which is higher in energy and is thus rejected. Conformers 4 and 5 are deemed as the same minima and are also higher than the starting structure and thus are rejected. Conformer 12 has a lower energy than the starting structure and is accepted as the low energy conformers. This process continues until a user specified number of final conformers is obtained, in this case 100. Included in Figure 2.3 is a graph of the conformers with respect to their energy. The highest and lowest energy conformations are shown.



Using Monte-Carlo algorithm.

Using rotatable bonds from rule normal set.

Conf (100)	Energy kJ/mol	Remark
1	-1194.130	
2	-1176.352	
3	-1174.576	
4	-1194.145	Same Min.
5	-1194.145	Same Min.
6	-1164.517	
7	-1194.125	
8	-1163.348	
9	-1163.354	Same Min.
10	-1163.345	Same Min.
11	-1163.050	
12	-1194.731	!New Best!
13	-1164.477	
14	-1179.533	
15	-1179.617	
16	-1178.609	
17	-1172.500	
18	Rejected	No Minima
19	-1179.578	Duplicate
20	-1168.693	
21	-1172.499	Prev. Min.
22	-1179.137	
23	-1152.591	
24	-1179.753	
25	-1178.405	Duplicate
26	-1179.627	Prev. Min.
27	-1179.622	Same Min.
28	-1179.623	Same Min.
29	-1179.622	Same Min.
30	-1178.537	Prev. Min.

Figure 2.3: (Left) A graph of the 100 accepted conformations with respect to energy found from a conformational search. (Right) Raw data output of the simulation which displays the change in energy with respect to the conformation. Remarks are included to indicate whether a conformation was accepted or rejected. Created from the Spartan program.[29]

2.1.4. Quantitative Structure Activity Relationships

Quantitative Structure-Activity Relationship (QSAR) models are a way of relating chemical structure with a specific chemical activity. It is a regression model which is built by relating a set of descriptor variables with a response variable which could be the activity of a molecule. QSAR is first used to create a model that relates these variables, and then is used to predict the activities of new molecules. The relationship between structure and property can be expressed numerically in an equation of the general form, (**Eq. 2.4**)

$$a = f(p) \qquad \textbf{(Eq. 2.4)}$$

where *a* is the activity of the molecule, *p* are the structure-derived properties, and *f* is some numerical function. The seminal work of Hansch in 1969 was the first to introduce the concept of quantitative structure-activity relationships in which the hydrophobicity (shown as the octanol-water partition coefficient, log *P*) was found useful in the prediction of various biological activities. Since then QSAR has been extensively in drug development and areas such as toxicity, enzyme inhibition, ligand-receptor binding, mutagenesis, and more.[30]

QSAR modelling can be distilled into three steps, (1) collect a training set of chemical compounds with experimentally measured activity, (2) choose or design a descriptor that can properly relate chemical structure to activity, and (3) apply statistical methods that correlate changes in structure with changes in activity. Caution must be taken when collecting data for a training set as this experimental data is the foundation upon which the QSAR model is built.[31] First, one must be careful to make sure that the experimental activity is reproducible, consistent, and accurate. Second, the range of structures should be diverse enough to span the

range of structural and chemical space relevant to the activity of study. Third, the activities should be distributed fairly evenly throughout the data – that is, both active and inactive molecule should be included, not just purely active molecules. Fourth, the number of structures should be large enough to draw statistical stability and relevant correlations (i.e a QSAR model cannot be made with only two entries for the training set.) Once a reliable set of data has been acquired the challenge now lies in finding an appropriate descriptor which can best relate the molecular structure to the activity in question.

2.1.4.1. Descriptors

A descriptor can be any variable which can relate the structure to the activity in question. This encompasses a very wide variety of properties which can be classified as constitutional, topological, geometric, electrostatic, quantum-chemical, and thermodynamic. A constitutional descriptor is the simplest type which does not account for electronic structure or spatial geometry. Some examples include atom counts, molecular weight, and number of rotatable bonds.[32] Topological descriptors describe the bonding information and connectivity between atoms in a molecule, many connective indices have been proposed as topological descriptors.[33] Geometric or spatial descriptors take the entire three dimensional geometric structure into account and include things such as the molecular surface area, molecular density, and volume.[34] Electrostatic descriptors involve describing the molecule with respect to its electronic structure and include thing such as partial atomic charges and the polarizability. [35] Quantum chemical descriptors are calculated using *ab initio* methods and include things such as the highest occupied molecular orbital (HOMO) or lowest occupied molecular orbital (LUMO) energies.[36] Thermodynamic descriptors are perhaps the oldest descriptors and encompass

empirical data such as heats of formation, hydrophobicity, and solvation free energies.[32] Typically descriptors based on higher levels of theory typically result in up to thousands of specific descriptors to describe the molecular structure. The number of descriptors needs to be high enough to provide distinct information between structures but also small enough that over-fitting does not occur. To this end there have been many optimization algorithms such as genetic algorithms in which to minimize the number of descriptors for subsequent use in generating a QSAR model.[37]

2.1.4.2. Alignment Independent Descriptors

The current state of QSAR development is highly focused on three-dimensional QSAR (3D-QSAR) due to the great importance of stereochemistry and spatial effects being a large determinant in activity. This form of QSAR deals with the three dimensional arrangement of atoms rather than a one dimensional or two dimensional representation. The most common way of describing the three dimensional surface is via a molecular field lattice. Comparative Molecular Field Analysis (CoMFA) is perhaps the most widely researched of these lattice-based methods.[38] In a CoMFA model alignment between similar portions of the molecules is extremely important. This usually means this is limited to a training set of molecules with the same base structure and is useful in studies of varied functionalization. The first step to constructing a CoMFA model is to align the training set molecules to a common reference point and placed within a 3D grid. Next the steric and electrostatic fields are calculated for the molecule by taking the curvature of the vdW surface and the electrostatic potential respectively at every grid point surrounding the molecules. The fluctuations in these steric and electrostatic fields serve as the descriptors to be correlated with the activity. Unfortunately, the greatest

drawback to the CoMFA approach is that the alignment step may not be appropriate for diverse classes of molecules with no obvious alignment hypothesis and can be highly subjective to user input. There have been many efforts to develop 3D grid independent descriptors but none have been more successful as the work of Pastor *et al.* called the GRid-Independent Descriptors (GRIND).[39] The GRIND approach calculates a maximum auto- and cross-correlation (MACC) which is essentially the product of interaction energies between two nodes in the 3D grid. It then discretizes this product according to distance and then only keeps the maximum values. Calculating the interaction between points of the molecular field eliminates the alignment dependence. The products of the nodes are the descriptors which can now be correlated to the activity. This method has been highly popular in drug discovery,[40] catalysis,[41] and even material discovery.[42]

2.1.4.3. Statistical Correlations Methods

Once a set of descriptors have been calculated, statistical methods must be used in order to build a correlation between these descriptors and the activity. Multilinear Regression (MLR) is a one of the approaches used in QSAR studies. Multilinear regression makes the assumption that the activity being modelled is a linear function of the structural descriptors. This function takes the general form as shown, (**Eq 2.5**)

$$y_i = w_0 + w_1 x_{i1} + w_2 x_{i2} + w_3 x_{i3} + \cdots \qquad \textbf{(Eq 2.5)}$$

where y_i is the activity being modeled, x_i are the structural descriptors, and w are the fitted coefficients which are calculated using various statistical regression techniques. Often the number of descriptors can be larger than the training set and MLR utilizes a stepwise regression

algorithm in order to optimize the number of descriptor variables. This method either adds, deletes, or both add and deletes descriptor variables until the greatest correlation is obtained. At each step the model is tested by the squared correlation coefficients, R^2 and the variables are either added or deleted to maximize the R^2. MLR has a few disadvantages which include inability to handle intercorrelated data, the assumption that the data has no noise, and requiring more activity observations than descriptor variables. Partial-least squares regression (PLSR) is method which overcomes many of the shortcomings of MLR and works by maximizing the covariance between the dependent activity variable and the original independent structural descriptor variables. This statistically condenses the vastness of the data while also maintaining the variation of the data set.[43] In cases where the structure-property relationship is not linear or where the property being modeled is a category or class than other regression techniques such as kernel, neural network, or support vector machine methods can be used.[44]

QSAR offers an attractive way to use computation and statistics to find correlations between structure and physicochemical activity which human chemical intuition may not be able to do. Furthermore, as computational power only increases, these "machine-learning" algorithms will become more capable of extracting novel correlations from complicated structures and activity data sets. Indeed, the original GRIND 3D-QSAR paper published in 200 has been cited 408 times as of June 2015 as found on Google Scholar. There have been 3D-QSAR models built to study substrate-protein recognition,[45] histone deacetylase inhibitors,[46] and hydrophobic interactions in potassium channel blockers.[47] To date there have been no IRI studies utilizing any QSAR models. However, the examples above show that QSAR can be used

to study substrate binding, inhibitors, and hydrophobic interactions all of which are aspects to ice recrystallization inhibition.

2.2. Metal Organic Framework Theoretical Methods

2.2.1. Periodic Density Functional Theory

DFT methods for molecular systems have already been described in Section 2.1.1.2. However, there is a distinct difference between DFT used to model discrete molecular systems and periodic solids. Most commonly, the types of basis set used to describe the orbitals are different. In the periodic solid case, plane wave basis sets are more convenient to use although there are periodic methods which use Gaussian basis sets. Plane wave basis sets utilize the Fourier theorem which states that a periodic function may be expressed as a sum of sine and cosine functions. (Figure 2.4) Therefore rather than being composed of Gaussian type functions, the basis set is a combination of plane wave sine or cosine functions such that the interference allows for greater maxima centered around the nucleus. With plane wave basis sets periodicity is intrinsic to the function, is more suited for solid-state calculations. Furthermore, since the core electrons are not involved in binding, excitations, or conductivity they are no longer explicitly treated. Instead the core electrons are bundled into the nucleic potential as an approximation. This combination of nucleic potential and implicit core electron potential is referred to as an *atomic pseudopotential*. Another added benefit of plane wave basis sets is that rather than expanding the basis set in Gaussian type functions, the only thing needed to increase quality is increasing the frequency cut-off of the plane waves.

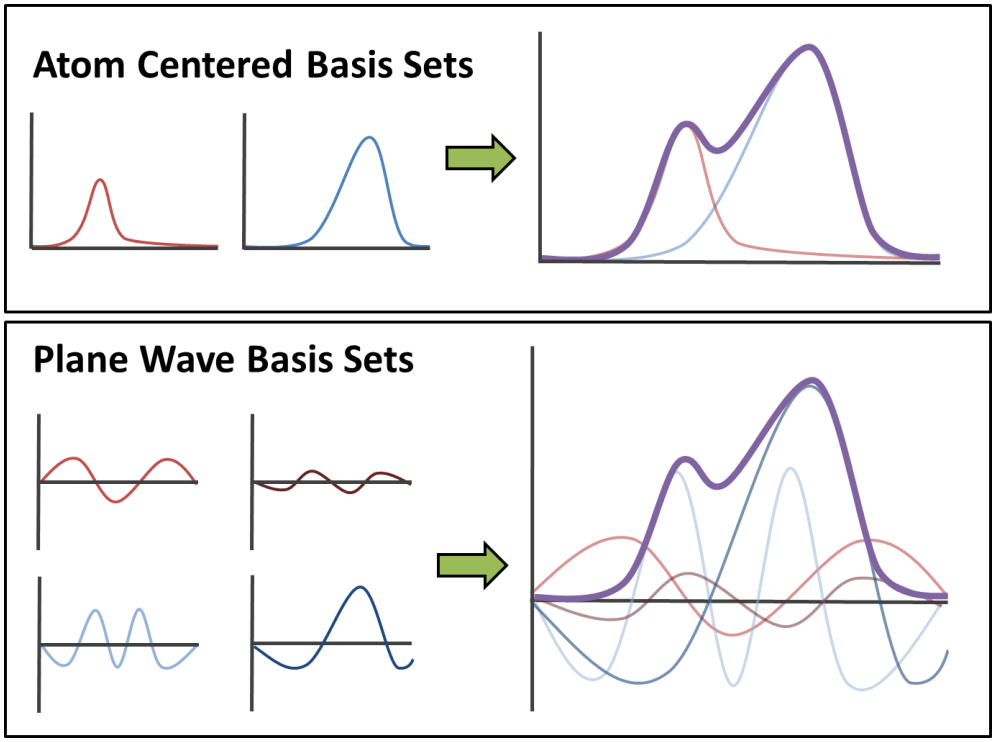


Figure 2.4: A representation of the combination of atom centered basis sets (Gaussian-type functions) and plane wave basis sets (sine/cosine type functions) to form the orbital wavefunction.

The primary uses for periodic DFT for these projects are geometry optimization and calculation of the electrostatic potential (ESP). As previously mentioned, there is often disorder from an experimental crystallographic information file. (CIF) X-ray diffraction is an experimental technique which measures the electron density whereby larger atoms with higher electron density results in greater perturbation within the diffraction pattern. Hydrogen is the lightest element with only one electron which is localized away from the nucleus which leads hydrogen to be essentially non-observable in X-ray diffraction experiments. Therefore, the hydrogen positions from an experimental CIF file must be calculated based on the atoms with which hydrogen is attached to and their known bond lengths and angles. Hydrogen atoms are then placed within these positions and then optimized using DFT in order to get the as accurate lowest energy geometry as possible. Furthermore, while something may crystallize in a certain

orientation with solvent, this may not be the "true" conformation of an activated MOF. Many flexible MOFs exist in different conformations with different cell vectors as described in section 1.2.2. Therefore, it is often wise to optimize all atom positions as well as the cell vectors in order to explore as much of the potential energy surface as possible. Once the hydrogen positions (and other atom positions if necessary) are relaxed, the resulting optimized geometry is used for subsequent GCMC simulations and assumed to be rigid. For the majority of MOFs this is a reasonable approximation since most MOFs are rigid and do not exhibit any flexible behaviour. But this assumption fails for flexible or dynamic MOFs which will be displayed in a further chapter.

Currently the group utilizes two different DFT programs which are both widely used in materials simulation, the Vienna Ab Initio Software Package (VASP)[48] and the Spanish Initiative for Electronic Simulations with Thousands of Atoms (SIESTA).[49] The main difference between VASP and SIESTA is the utilization of plane-wave basis sets (VASP) vs. atom-centered basis sets (SIESTA) for calculation in the solid state. The work presented in this thesis regarding DFT has been done with VASP which possess accurate pseudopotentials for most elements. For all calculations the Perdew-Burke-Ernzerhof (PBE)[69] exchange-correlation functional was used with an energy cutoff of 400 eV.

2.2.2. REPEAT Periodic Charge Calculations

The electrostatic interaction between atoms can be described as the static electric force felt between the two atoms. This is the net combination between an atom's positive nuclear charge and the negative charge arising from its local electron density. When considering

molecular mechanics simulations, the electron density is not explicitly treated but rather the electrostatic interactions are approximated by assigning a partial atomic point charge to the atom. The electrostatic potential (ESP) is the potential energy that arises from charge distribution. The ESP is a continuous property through three dimensional space and that for a molecule is determined from the electron density and nuclear charges. From the DFT wavefunction the ESP is calculated and a fitting procedure is done to produce ESP values at grid points surrounding the surface of the molecule typically at the van der Waals radii. In order to capture the electrostatics of the ESP in point charge form, partial atomic charges must be derived in order to recreate the QM calculated ESP. There have been many methods to derive partial atomic charges from the ESP of molecular systems but it was a significant challenge for periodic systems due to the ambiguity of a reference state in an infinite system. The Repeating Electrostatic Potential Extracted ATomic (REPEAT)[50] charge method was developed in the Woo Lab by Dr. Carlos Campana to address this very issue. Since the introduction of the REPEAT method in 2009, many studies have utilized this approach for calculations in porous crystalline materials. While there have been other methods used to calculate partial atomic charges for MOF framework atoms, REPEAT has been tested against them and provide the best match to the quantum mechanically calculated ESP.[51] The details and implementation of the REPEAT method is beyond the scope of this thesis but it is important to mention that before this method partial atomic charges were not derivable from the electrostatic potential of periodic systems.

2.2.3. Grand Canonical Monte-Carlo Simulations

Statistical mechanics attempts to relate the microscopic world to the macroscopic world through probability and statistical averages. Grand Canonical Monte Carlo (GCMC) simulations are a powerful statistical technique commonly used to study interfacial phenomena such as guest molecules adsorbing into a porous material. "Grand canonical" references the ensemble being be sampled and "Monte Carlo" references the random sampling of this ensemble. Consider a MOF system with *N* number of guest molecules confined in a unit cell of fixed volume *(V)* with a constant temperature (*T)*. A collection of replicas of this system which are all in contact with each other is called an *ensemble*. Within the grand canonical ensemble the temperature (*T)*, the volume (*V*), and the chemical potential (μ) remain constant within each replica while the number of guest molecules (*N)* change. The chemical potential (μ*)* is the measure of how much the free energy of a system changes if you add or remove a set number of particles of one species while keeping the number of another species as well as the temperature and pressure constant. For an ideal gas this can be expressed as the following equation, (**Eq. 2.6**)

$$\mu_{ideal\ gas}(p,T) = \mu^{0}_{ideal\ gas} + RT \cdot ln\frac{p}{p^{0}} \qquad \textbf{(Eq. 2.6)}$$

where P_o is the total pressure, $\mu^{0}_{ideal\ gas}$ is the standard chemical potential of the gas at total pressure, *R* is the gas constant, *T* is the temperature, and *P* is the partial pressure. With respect to a mixture of species, such as a MOF and its guests, the chemical potential is the *slope* of the potential energy vs. the number of guests in the MOF and is dependent on the quantity of guest molecules. The grand canonical ensemble allows the MOF to be "open" and allows for

guests to enter and exit the system with the chemical potential remaining the same regardless the number of guests. Systems tend to move from high chemical potential to lower chemical potential. For example, consider an empty MOF surrounded by guest molecules with high concentrations in the gas phase. In this case the guest molecules will move from areas of high concentration (the surface) to low concentration (within the bulk MOF) thus minimizing the chemical potential. One of the ways to sample this ensemble system is the Monte Carlo method.

Monte Carlo (MC) is a random sampling method which takes its name from the Monte Carlo Casino in Monaco where random chance wins the day. Its utility has been recognized in many fields such as economics, computer science, physics, biology and chemistry. The goal of MC is to generate various configurations or states of a system by making random perturbations to the positions of the guest molecule and then sample properties of interest such as energy. If one is exploring a potential energy surface, it is the energy minima which represent the most likely configurations. In chemistry, it is important to bias sampling towards these more probable configurations which is called *importance sampling*. This is done via the Metropolis Monte Carlo algorithm, an example for sampling the Canonical Ensemble is outlined below:

1. A guest molecule is randomly chosen.
2. The molecule's position will be randomly perturbed via a trial move. Some trial moves include translation, rotation, exchanges and even jumps.
3. The potential energy of the new configuration is calculated typically via a molecular mechanics forcefield.

4. If the energy of the new configuration is lower than the energy of the old configuration then this configuration is automatically accepted and this energy is added to the ensemble average (average energy of all accepted configurations).
5. If the energy of the new configuration is higher than the energy of the old configuration it is accepted based on a Boltzmann probability. The Boltzmann probability is equal to $e^{-\Delta U/kT}$ where ΔU is the change in energy between old and new configurations, *k* is the Boltzmann constant, and *T* is the temperature. This factor can take a value between 0 and 1 where the larger the change in energy the smaller the acceptance probability. A random number is generated between 0 and 1 and if this number is smaller than the acceptance probability, the configuration is accepted and added to the ensemble average.
6. This process is repeated for a predetermined number of steps. These GCMC steps must be high enough that energy of the system and number of guest molecules converge to equilibrium.

Grand Canonical Monte Carlo (GCMC) differs from MC in that GCMC includes the trial moves of guest molecule insertion and deletion and the acceptance probabilities are slightly different. This allows the quantity of guest molecules to fluctuate and therefore samples the grand canonical ensemble. Figure 2.5 showcases some of the trial moves possible.

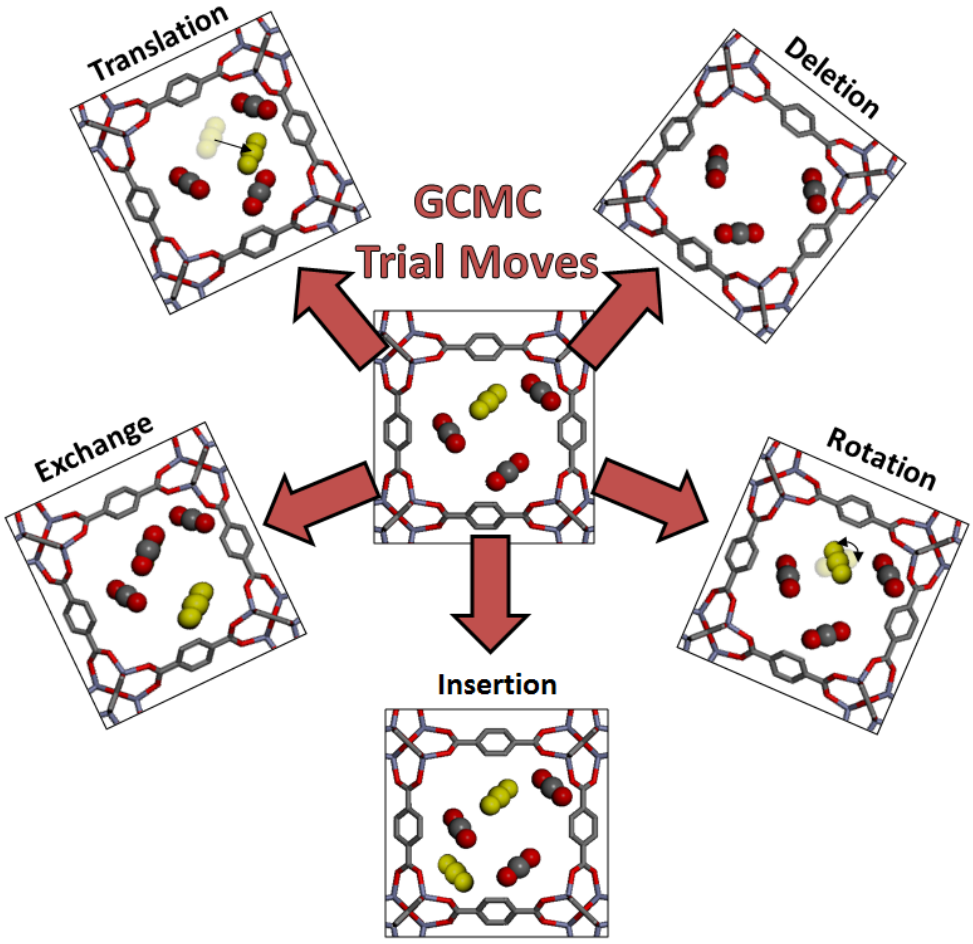


Figure 2.5: Graphical representation of the different types of GCMC moves of a guest CO_2 molecule within a rigid framework MOF. These moves include translation, deletion, rotation, insertion, and exchange. At each configuration the potential energy is calculated and either accepted or rejected based on the Boltzmann probability.

The simulation starts with the MOF assumed to be fixed and rigid with no guest gas molecules and a predetermined pressure of the guest which determines the chemical potential in the gas phase (μ_{gas}). As the simulation progresses, it is more favorable for the gas molecules to be inserted into the MOF since this results in lower chemical potential by molecules moving from high concentration (gas phase) until the chemical potential within the MOF reaches that of the lower concentration (adsorbed phase). As more steps occur and the number of guests increases such that the gas reaches the specified temperature and pressure conditions of the simulation. At this point equilibrium has been reached and the number of guest gas molecules

at equilibrium is now known at a specific pressure and temperature and thus the amount of adsorbed guest can be calculated. An appropriately high number of simulation steps (~10^7) must be used in order to achieve this equilibrium. Furthermore, once equilibrium is reached the simulation must be sufficiently long enough to statistically average in a meaningful way. For example, if it takes 100,000 steps to reach equilibrium and the simulation only has 150,000 steps total, then only the 50,000 steps after equilibrium has been reached (called the production steps) are counted for the thermodynamic average. The beauty of this technique is that in spite of being a purely statistical method, the results can be highly accurate compared to experimental isotherms. Therefore GCMC simulations are commonly used to modelling gas adsorption isotherms for porous materials such as MOFs.[52] The Woo Lab uses its own GCMC code called FastMC which was developed by Ph.D student Peter Boyd. This code utilizes input files of the same format as those used in the molecular dynamics program DL_POLY.[73]

2.2.4. Automatic Binding Site Locator (ABSL)

The binding sites of a guest are defined as most probable locations to find a guest within a host. In the context of CO_2 capture in MOFs, determining the binding sites is important in identifying the structural and chemical features which promote optimal adsorption characteristics. Often times it is challenging to locate CO_2 binding sites with experimental techniques such as x-ray diffraction. This is because the physisorbed CO_2 molecules are weakly bound and move around between binding sites within the MOF pores which leads to weak diffraction. This disorder is a problem when determining a crystal structure of a MOF which includes the physisorbed CO2 binding sites. The Woo Lab in collaboration with the Shimizu group were the first to experimentally observe CO_2 binding sites physisorbed within a MOF and

then validate those binding sites using computational simulations.[53] Since then the Woo Lab has developed a program to automatically determine the location of CO_2 binding sites within MOFs.

The Automatic Binding Site Locator (ABSL) is a program which identifies the CO_2 binding sites from the maxima in the guest probability distribution from GCMC simulations. The GCMC simulations provide a spatial probability distribution of the guest molecules in a MOF under specified temperatures and partial pressure of the guest molecule. Thus, a spatial probability distribution constructed from the GCMC simulation gives the probability per unit volume that a CO_2 will occupy that region of space during the simulation. However, due to the randomness of the GCMC simulation, these probabilities are usually quite fluxional which leads to noisy data. One solution to this problem would be to increase the number of GCMC runs where in theory a higher degree of sampling would result in smoother data. However, this approach is impractical and inefficient when it comes to high-throughput studies. Another challenge lies in overlapping binding sites, where the probability distribution shows two binding sites which are both probable and valid but spatially overlap. To this effect the Woo Lab has developed the ABSL which takes the noisy probability data from GCMC simulations and uses a Gaussian filter and an equitable binning algorithm to smooth the data in order to condense the local maxima. (Figure 2.6)

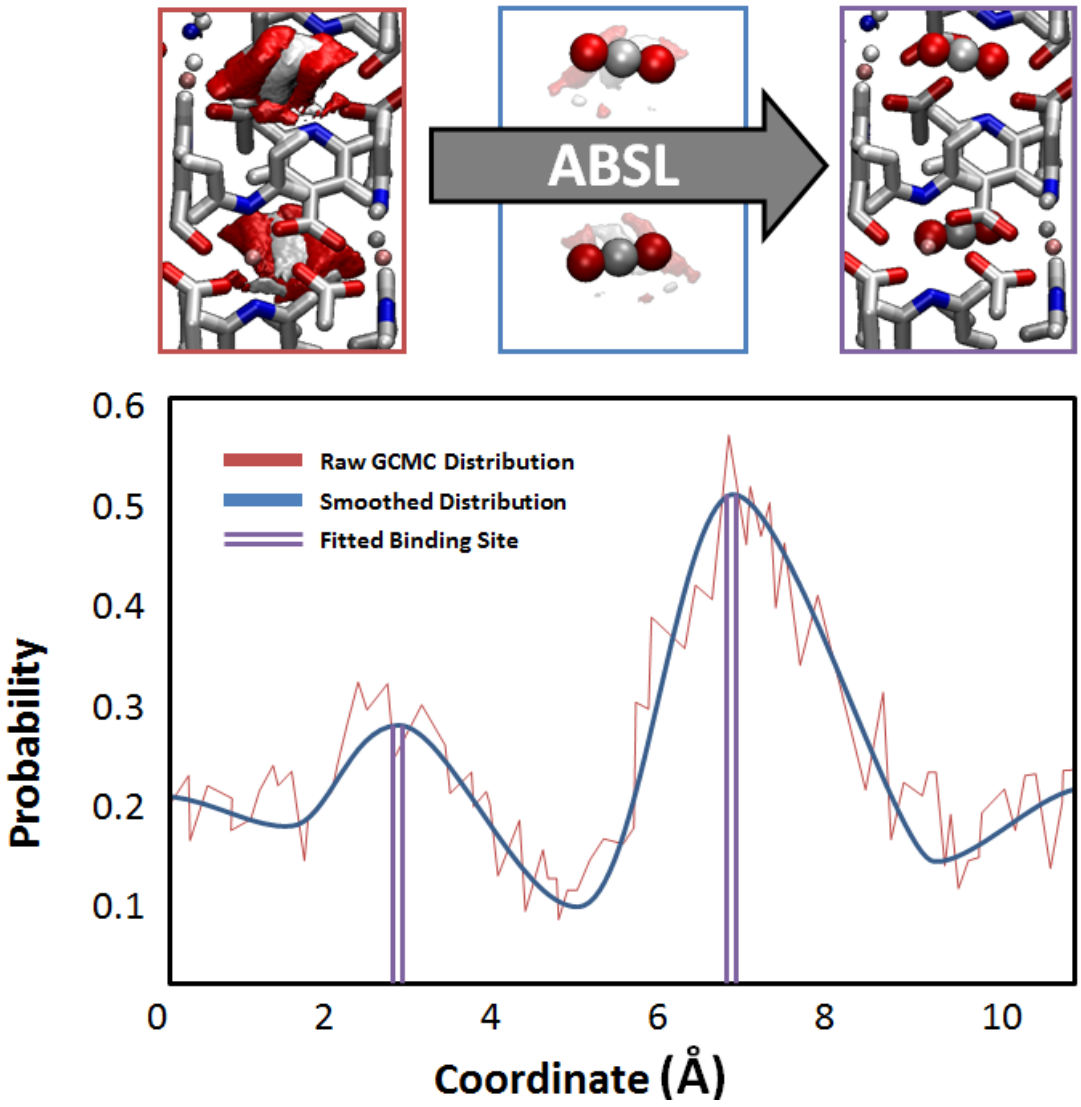


Figure 2.6: Top: The probability distribution of CO_2 displayed as an isosurface (left) and the fitting of CO_2 molecules (middle) to give the binding sites within the MOF (right). Bottom: A plot of CO_2 probability with respect to position. A Gaussian filter is applied to smooth the raw probability distribution (red). The resulting smoothed distribution (blue) can be used to fit the binding site locations (purple).

Next it fits a CO_2 molecule to the maxima of the smoothed probability distribution in such a way that accounts for alignment. For linear molecules such as CO_2, the centre atom (in this case carbon) or the center of mass for diatomic molecules is fitted to a maxima. Then a bonded atom (oxygen) is aligned such that it is in the direction of the next nearest maxima in the probability distribution. Once the CO_2 is placed in the binding site region, a short optimization

using molecular mechanics methods is done to allow the CO_2 to relax to its local energy minimum. Next, the occupancy of the binding site as well as its binding energy is calculated in terms of its electrostatic and van der Waals contributions. The binding energy per CO_2 molecule is defined as, (**Eq. 2.7**)

$$E_{binding} = E(\mathrm{MOF} + n\mathrm{CO_2}) - E(\mathrm{MOF}) - nE(\mathrm{CO_2}) \qquad \textbf{(Eq. 2.7)}$$

where the configurational energies for the MOF with the guest CO_2included in the simulations are used for $E(\mathrm{MOF} + n\mathrm{CO_2})$. $E(\mathrm{MOF})$is the configurational energy of the MOF with no guests. $nE(\mathrm{CO_2})$ is the configurational energy of one CO_2 molecule times *n* number of CO_2 molecules. Thus, ABSL provides data on the location of the binding sites, how favourable one binding site is over another, and also energetic data which can be used to determine whether the CO_2 is interacting via electrostatic forces or steric forces.

2.2.5. Structural Property Determination

A necessity in the study of nanoporous materials is determining structural features such as the accessible surface area and accessible volume. The accessible surface area was first defined by Lee and Richard[54] as the surface traced by the centre of a spherical probe as it is rolled along the atomic van der Waals surface. This describes the surface area which a guest molecule is able to freely access. (Figure 2.7)

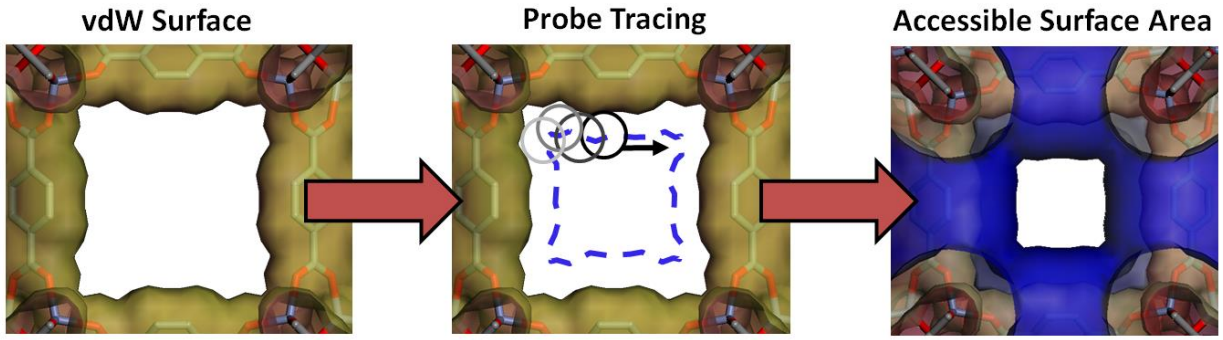


Figure 2.7: A representation of how the accessible surface area is calculated. First the vdW surface of the framework is calculated. Next a probe of a specific size (in this case 1.72 Å for CO_2) traces along the vdW surface area. The accessible surface area is calculated as the surface generated from the center of a tracing probe.

By extension, the accessible volume is the volume accessible by the centre of this probe. The probe radius can be adjusted to fit the size of guest. Some typical values for probe radii are 1.42 Å for H_2O, 1.72 Å for CO_2, and 1.82 Å for N_2. The most common accessible surface area algorithms are based off a Monte Carlo integration approach. Here probes are randomly placed within the volume and are only accepted if the distance to the closest framework atom is greater than or equal to the distance of the van der Waals radii plus the radius of the probe. The accessible surface area has been shown by the work of Frost *et al.* to have a strong correlation to amount of hydrogen adsorbed at moderate pressures.[55] Thus, fast calculation of the surface area may be used as a descriptor for high-throughput screening of MOFs for various adsorption applications. FAP^{3}S utilizes the software ZEO++ for all calculations of accessible surface are and pore size distributions.[56]

2.2.6. Fully Automated Adsorption Analysis for Porous Solids (FA^{3}PS)

FA^{3}PS is a software program which incorporates many different computational methods including those previously mentioned under one umbrella in order to provide adsorption

analysis data. FA^{3}PS reduces user error while speeding up calculation time by efficiently combing different programs into one simple streamlined system. This program was written and implemented within the Woo Lab by former post-doc Dr. Tom Daff. Before the conception of FA^{3}PS the generation of an isotherm for a single metal-organic framework was a long process which involved user intervention between each calculation step. However, it is now all fully automated in a modular way so that many different adsorption features can be calculated and analyzed in different ways.

The entire process starts with one file, the Crystallographic Information File (CIF), which contains the structure of the crystal with atomic positions and periodic cell vectors. Many CIF structures include the solvent within the crystal structure of MOF where the solvent is inside the pore. This is called the inactive MOF because the pores have not been emptied for maximum adsorption capabilities. Often times there is rotational disorder within the CIF file due to the resolution of x-ray diffraction. Some examples include rotations of methyl groups, rotations of benzene or pyridine linkers, and other functional groups. Often times there must be a manual "cleaning" of the CIF file as taken from experiment. This includes deleting solvent and removing rotational disorder while trying to maintain as high level of symmetry as possible. This is done in a materials visualization program such as Materials Studio[57] which is a commercial software which the Woo Lab uses.

A typical FA^{3}PS calculation has the following steps, (1) A periodic DFT calculation on the MOF for geometry optimization and subsequent calculation of the electrostatic potential (ESP), (2) partial atomic charges calculated from the ESP via the REPEAT method, (3) Grand Canonical Monte Carlo (GCMC) calculations to simulate guest adsorption, (4) the Automated Binding Site

Locator (ABSL) which fits guest binding sites to GCMC maxima, and (5) physical structure property determination such as surface areas and pore sizes via the ZEO++ code. Once the GCMC simulation is done FA^{3}PS utilizes the data to calculate the uptake, heat of adsorption, and excess uptake. A graphical representation of the FA^{3}PS program is shown in Figure 2.8.

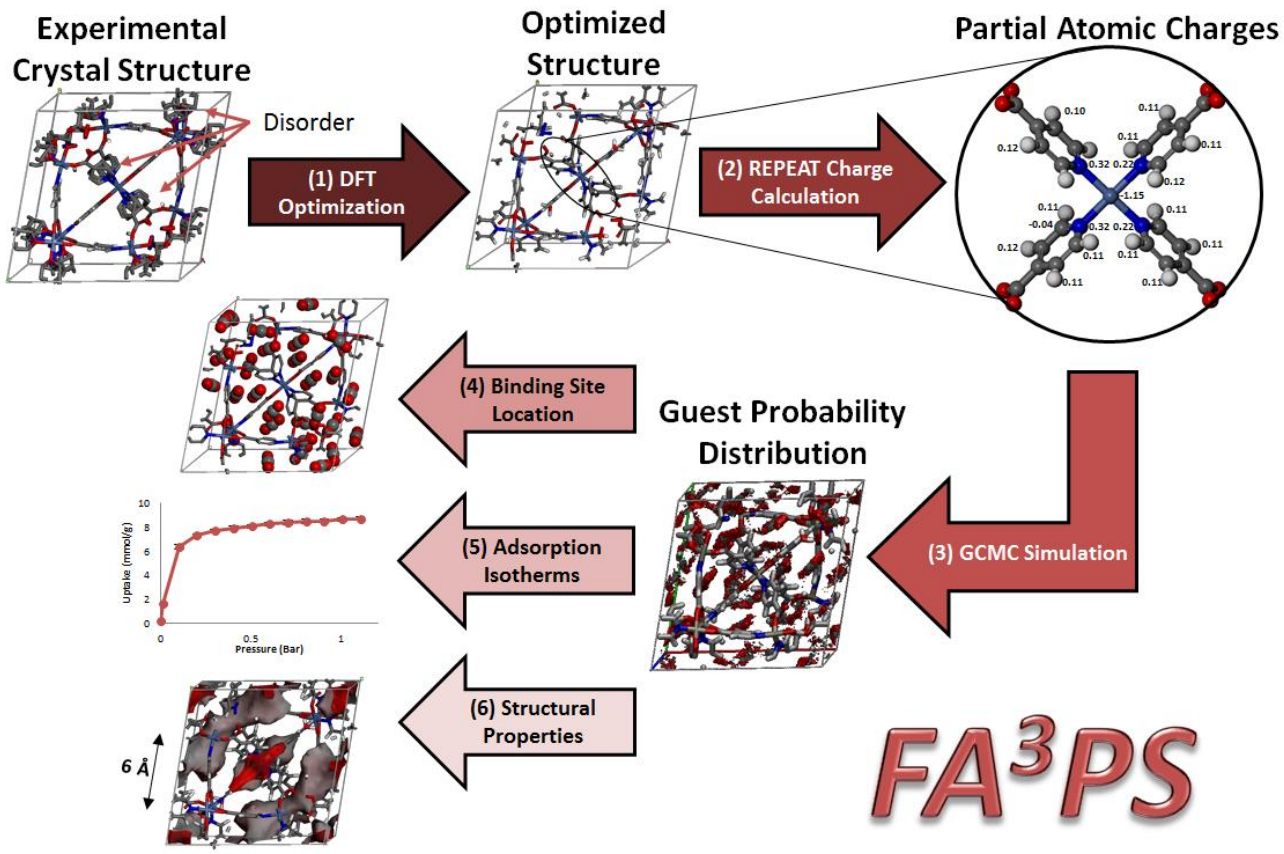


Figure 2.8: Graphic representation of the workflow of the FA^{3}PS program. (1) A DFT optimization is done to relax hydrogen positions and find an optimized structure with no disorder. (2) The REPEAT method is used to calculate partial atomic charges on the framework atom. (3) GCMC simulations are performed with a specific guest. (4) The binding sites are calculated from maxima in the guest probability distribution. (5) Thermodynamic properties and the adsorption are calculated from the average number of guests from the GCMC simulation. (6) ZEO++ is used to calculate structural properties such as accessible surface area, pore size, etc.

2.2.7. Molecular Dynamics Simulations

All of the computational methods so far described have been time independent. Conversely, molecular dynamics is the simulation of a molecular system propagating with respect to time. In GCMC simulations configurations are generated randomly, in DFT configurations are generated as a result of an optimization process, but in molecular dynamics atomic positions are determined through integrating Newton's laws of motion. This allows for a trajectory of the system which is comprised of the position, velocity, and forces acting on all atoms at any point in time. The timescale of molecular dynamics is typically on the ps-ns scale, but even at these time scales a wealth of information is available.

When molecular dynamics (MD) is referenced in chemistry it is usually discussed in terms of classical molecular dynamics where the atoms follow Newtonian or classical laws of motion and the potential energy surface is determined by a forcefield. This potential energy surface can be calculated using molecular mechanics methods, quantum mechanical methods, or both combined.[58] There exists *ab-initio* molecular dynamics[59] methods uses quantum mechanical methods such as DFT to calculate the potential energy surface during a molecular dynamics simulation. However, these are expensive computationally and are usually limited to smaller systems. On the other hand, classical MD is capable of modelling very large systems such as proteins in explicit aqueous environments.

With classical MD, the atomic nuclei are moved according to Newton's second law where the force on a particle is equal to its acceleration multiplied by its mass. In order to model motion the initial positions and forces on the atoms must be known. The initial positions are

typically gained from crystal structure coordinates and the forces are determined from the gradient of the potential energy surface as calculated via quantum mechanics or the molecular mechanics forcefield. The force is the first derivative of the potential energy function and the acceleration of an atom can be found from the forces. From here the position at a later time ($t + \Delta t$) can be determined from the initial position, the initial velocity, the change in time (time step) and the acceleration. The time step used is very important, two small of a time step and the simulation will take too long to compute. Too large a time step and the simulation experiences instabilities and will not accurately determine the motion of the atoms. A standard time step employed in MD simulations is typically 1 femtosecond. Once the particle moves to its new position the energy of this configuration is calculated and the process begins anew. This process is repeated until the time of simulation has been satisfied, typically when properties of interest have converged.

A typical ensemble used is the NVT ensemble where the number of particles remains constant, the volume remains constant, and the temperature remains constant. This is a general scheme for an MD simulation but there have been many algorithms utilized to speed up this process such as the Verlet algorithm and the leap-frog algorithm. For further general reading on general molecular dynamics simulations please read the following references, Chapter 6 in Molecular Modelling: Principles and Applications by Andrew R. Leach[60] and Introduction to Molecular Dynamics Simulation.[61]

2.2.7.1. Diffusion Coefficient

The diffusion coefficient is a measure of how easily a gas such as CO_2 can diffuse through a media such as a porous MOF where the higher the diffusion coefficient the faster a CO_2 can travel with respect to time. Diffusivity is a measurement of length squared over time and an example is carbon dioxide has a diffusion coefficient of 0.0016 mm^2/s in water.[62] This is an important property with respect to MOFs for CO_2 capture because an efficient adsorbent should be able to absorb and desorb the gas *quickly*. The diffusion coefficient is one of the properties which can be calculated from the position data from a molecular dynamics simulation. This is done by relating the displacement of molecules and time. The mean-squared displacement (MSD) is the average squared displacement for all *N* molecules during the simulation. This average property is calculated as, (**Eq 2.8**)

$$MSD(t) = \frac{1}{N}\langle\sum_{i=1}^{N}|\boldsymbol{r_i}(t) - \boldsymbol{r_i}(0)|^2\rangle = \langle|\boldsymbol{r}(t) - \boldsymbol{r}(0)|^2\rangle \qquad \textbf{(Eq. 2.8)}$$

where $r_i(0)$ corresponds to a starting position at time t = 0 and $r_i(t)$ corresponds to a position at time t. The diffusion coefficient, *D* is defined as the slope of MSD with respect to time, (**Eq. 2.9**)

$$\lim_{t\to\infty}\frac{d\langle|\boldsymbol{r}(t)-\boldsymbol{r}(0)|^2\rangle}{dt} = 6D \qquad \textbf{(Eq. 2.9)}$$

where 6 is a numerical constant dependent on the dimension of the system. (2, 4, or 6 for 1, 2, or 3 dimensional diffusion). An important consideration is that the MD simulation must be sufficiently long (~ 1 nanosecond) for the slope of MSD vs. time to be linear.

With respect to gas diffusion within MOFs and similar porous materials, molecular dynamics have been a long established simulation technique to calculate the diffusion

coefficients with many examples in literature. In 2005, Skoulidas and Sholl were the first to publish any diffusion coefficients of light gases in MOFs either experimentally or computationally.[63] They studied Ar, CH_4, CO_2, N_2 and H_2 diffusion in MOF-5 using MD simulations. Experimental validation of this technique arrived when Rosenbach *et al*. calculated the diffusion coefficient of methane in MOFs MIL-47(V) and MIL-53(Cr) and found excellent agreement with Quasi-Elastic Neutron Scattering (QENS) experiments.[64] In a comparative study of silicate, C168 scharzite, and IRMOF-1 Barbarao and Jiang used MD to calculate the diffusion coefficients of CO_2 and CH_4 with good agreement to experiment and found that IRMOF-1 had the best diffusion of CH_4 but silicate was better for CO_2.[65] Thus, MD calculated diffusion coefficients have proven to be in good agreement with experiment in terms of small gases in MOFs.

2.3. Software Details

Ice Recrystallization Inhibition: All three-dimensional structures were constructed using the Spartan Program.[29] An initial geometry optimization was performed using the Merck Molecular Force Field (MMFF),[66] the conformational search was done using a Monte Carlo algorithm within Spartan at the PM6[67] semi-empirical level of theory. The quantum mechanical electron density and electrostatic potential was calculated at the B3LYP/6-311g(d,p) level of theory using the Gaussian 09 software.[68] The maximum auto and cross-correlation grid independent descriptors were calculated with an in house code first developed by former Ph.D student Nick Trefiak and modified by me for this project. The QSAR model regression techniques were done by an in house code developed by former post-doc Dr. Michael Fernandez.

Metal Organic Framework Studies: All MOFs and CIF files were visualized and manipulated with Materials Studio.[57] The adsorption characteristics were done with FA^{3}PS which is an in-house wrapper program used for the calculation and analysis of adsorption properties. All periodic DFT calculations were performed on the Vienna Ab-Initio Simulation Program (VASP)[48] using the PBE exchange-correlation functional.[69] Plane Augmented Wave (PAW) pseudopotentials[70,71] were used in a plane wave basis set with empirical dispersion corrections of Grimme[72] and spin polarization calculations. The partial atomic charges were calculated using the REPEAT method.[50] GCMC simulations were performed with an in-house code developed by Ph.D student Peter Boyd, FastMC. The binding sites were calculated using the Automatic Binding Site Locator (ABSL) as discussed in section 2.4.4. Binding site relaxation and diffusion coefficient calculations were done using the molecular dynamics program DL_POLY.[73] Molecular dynamics simulations of flexible MOFs were done with the GROMACS program.[74]

Further technical details will be elucidated within the coming chapters.

2.4. References

(1) McQuarrie, D. A. *Quantum Chemistry*; University Science Books, 2008.

(2) Hartree, D. R. **1957**.

(3) Dewar, M. J. S.; Thiel, W. *J. Am. Chem. Soc.* **1977**, *99* (15), 4899.

(4) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107* (13), 3902.

(5) Stewart, J. J. P. *J. Comput. Chem.* **1989**, *10* (2), 209.

(6) Cheng, Y.; Yang, Z.; Tan, H.; Liu, R.; Chen, G.; Jia, Z. *Biophys. J.* **2002**, *83* (4), 2202.

(7) Liu, K.; Wang, Y.; Tan, H.; Chen, G.; Tong, Z. *Front. Biol. China* **2007**, *2* (2), 180.

(8) Wierzbicki, A.; Knight, C. A.; Salter, E. A.; Henderson, C. N.; Madura, J. D. *Cryst. Growth Des.* **2008**, *8* (9), 3420.

(9) Liu, K.; Jia, Z.; Chen, G.; Tung, C.; Liu, R. *Biophys. J.* **2005**, *88* (2), 953.

(10) Kohn, W.; Sham, L. J. *Phys. Rev.* **1965**, *140* (4A).

(11) Becke, A. D. *Phys. Rev. A* **1988**, *38* (6), 3098.

(12) Zhao, Y.; Truhlar, D. G. *Phys. Chem. Chem. Phys.* **2005**, *7* (14), 2701.

(13) Siegbahn, P. E. M. The performance of hybrid DFT for mechanisms involving transition metal complexes in enzymes. *Journal of Biological Inorganic Chemistry*, 2006, *11*, 695–701.

(14) Baruch, E.; Belostotskii, A. M.; Mastai, Y. *J. Mol. Struct.* **2008**, *874* (1-3), 170.

(15) Yan, X. F.; Xiao, H. M.; Gong, X. D.; Ju, X. H. *J. Mol. Struct. THEOCHEM* **2006**, *764* (1-3), 141.

(16) Saghaie, L.; Sakhi, H.; Sabzyan, H.; Shahlaei, M.; Shamshirian, D. *Med. Chem. Res.* **2012**, *22* (4), 1679.

(17) Rasulev, B. F.; Abdullaev, N. D.; Syrov, V. N.; Leszczynski, J. *QSAR Comb. Sci.* **2005**, *24* (9), 1056.

(18) Barton, D. H. R. *Experientia* **1950**, *6* (8), 316.

(19) Judson, R. S.; Jaeger, E. P.; Treasurywala, A. M.; Peterson, M. L. *J. Comput. Chem.* **1993**, *14* (11), 1407.

(20) Schlitter, J.; Engels, M.; Kruger, P. In *Journal of Molecular Graphics*; 1994; Vol. 12, pp 84–89.

(21) Van Schaik, R. C.; van Gunsteren, W. F.; Berendsen, H. J. C. *J. Comput. Aided. Mol. Des.* **1992**, *6* (2), 97.

(22) RAMACHANDRAN, G. N.; RAMAKRISHNAN, C.; SASISEKHARAN, V. *J. Mol. Biol.* **1963**, *7*, 95.

(23) Berg, J. M.; Tymoczko, J. L.; Stryer, L. *Biochemistry*, 7th ed.; W.H. Freeman & Company, 2010.

(24) Gibson, K. D.; Scheraga, H. A. *J. Comput. Chem.* **1987**, *8* (6), 826.

(25) Perrakis, A.; Morris, R.; Lamzin, V. S. *Nat. Struct. Biol.* **1999**, *6* (5), 458.

(26) Havel, T. F.; Kuntz, I. D.; Crippen, G. M. *J Theor Biol* **1983**, *104*, 359.

(27) Li, Z.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U. S. A.* **1987**, *84* (19), 6611.

(28) Hansmann, U. H.; Okamoto, Y. *Curr. Opin. Struct. Biol.* **1999**, *9* (2), 177.

(29) Wavefuntion Inc. Spartan '14, 2015.

(30) Hansch, C.; Leo, A.; Livingstone, D. J. *Pestic. Biochem. Physiol.* **1996**, *56* (1), 78.

(31) Perkins, R.; Fang, H. *Environ. …* **2003**.

(32) Todeschini, R.; Consonni, V. *Handbook of Molecular Descriptors*; 2000; Vol. 11.

(33) Lipkus, A. H. *J. Chem. Inf. Comput. Sci.* **2001**, *41* (2), 430.

(34) Baumann, K. *J. Chem. Inf. Comput. Sci.* **2002**, *42* (1), 26.

(35) Stanton, D.; Jurs, P. *Anal. Chem.* **1990**.

(36) Karelson, M.; Lobanov, V. S.; Katritzky, A. R. *Chem. Rev.* **1996**, *96* (3), 1027.

(37) Gao, H.; Lajiness, M. S.; Drie, J. Van. In *Journal of Molecular Graphics and Modelling*; 2002; Vol. 20, pp 259–268.

(38) Cramer, R. D.; Patterson, D. E.; Bunce, J. D. *J. Am. Chem. Soc.* **1988**, *110* (18), 5959.

(39) Pastor, M.; Cruciani, G.; McLay, I.; Pickett, S.; Clementi, S. *J. Med. Chem.* **2000**, *43* (17), 3233.

(40) Artese, A.; Cross, S.; Costa, G.; Distinto, S.; Parrotta, L.; Alcaro, S.; Ortuso, F.; Cruciani, G. Molecular interaction fields in drug discovery: Recent advances and future perspectives. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 2013, *3*, 594–613.

(41) Aguado-Ullate, S.; Guasch, L.; Urbano-Cuadrado, M.; Bo, C.; Carbó, J. J. *Catal. Sci. Technol.* **2012**, *2* (8), 1694.

(42) Rofouei, M. K.; Salahinejad, M.; Ghasemi, J. B. *Fullerenes, Nanotub. Carbon Nanostructures* **2014**, *22* (7), 605.

(43) Katritzky, A.; Maran, U.; Lobanov, V.; Karelson, M. *J. Chem. Inf. Comput. Sci.* **2000**, *40* (1), 1.

(44) Katritzky, A. R.; Kuanar, M.; Slavov, S.; Hall, C. D.; Karelson, M.; Kahn, I.; Dobchev, D. A. *Chem. Rev.* **2010**, *110* (10), 5714.

(45) Cianchetta, G.; Singleton, R. W.; Zhang, M.; Wildgoose, M.; Giesing, D.; Fravolini, A.; Cruciani, G.; Vaz, R. J. *J. Med. Chem.* **2005**, *48* (8), 2927.

(46) Ragno, R.; Simeoni, S.; Rotili, D.; Caroli, A.; Botta, G.; Brosch, G.; Massa, S.; Mai, A. *Eur. J. Med. Chem.* **2008**, *43* (3), 621.

(47) Ermondi, G.; Visentin, S.; Caron, G. *Eur. J. Med. Chem.* **2009**, *44* (5), 1926.

(48) Hafner, J. Ab-initio simulations of materials using VASP: Density-functional theory and beyond. *Journal of Computational Chemistry*, 2008, *29*, 2044–2078.

(49) Soler, M.; Artacho, E.; Gale, J. D.; Garc, A.; Junquera, J.; Ordej, P.; Daniel, S. *J. Phys. Condens. Matter* **2002**, *14* (11), 2745.

(50) Campañá, C.; Mussard, B.; Woo, T. K. *J. Chem. Theory Comput.* **2009**, *5* (10), 2866.

(51) Watanabe, T.; Manz, T. A.; Sholl, D. S. *J. Phys. Chem. C* **2011**, *115* (11), 4824.

(52) Yang, Q.; Liu, D.; Zhong, C.; Li, J.-R. *Chem. Rev.* **2013**, *113* (10), 8261.

(53) Vaidhyanathan, R.; Iremonger, S. S.; Shimizu, G. K. H.; Boyd, P. G.; Alavi, S.; Woo, T. K. *Science* **2010**, *330* (6004), 650.

(54) Lee, B.; Richards, F. M. *J. Mol. Biol.* **1971**, *55* (3), 379.

(55) Frost, H.; Duren, T.; Snurr, R. Q. *J. Phys. Chem. B* **2006**, *110* (19), 9565.

(56) Willems, T. F.; Rycroft, C. H.; Kazi, M.; Meza, J. C.; Haranczyk, M. *Microporous Mesoporous Mater.* **2012**, *149* (1), 134.

(57) Segall, M.; Linda, P.; Probert, M.; Pickard, C. *Accelrys San Diego, CA* **2002**.

(58) Field, M. J.; Bash, P. a.; Karplus, M. *J. Comput. Chem.* **1990**, *11* (6), 700.

(59) Marx, D.; Hutter, J. *Mod. methods algorithms quantum …* **2000**, *1*, 301.

(60) Leach, A. R. *Molecular Modelling: Principles and Applications*; 2001.

(61) Allen, M. *Comput. Soft Matter From Synth. Polym. to …* **2004**, *23* (2), 1.

(62) Cussler, E. L. *Diffusion: Mass Transfer in Fluid Systems*; 1997; Vol. Second.

(63) Skoulidas, A. I.; Sholl, D. S. *J. Phys. Chem. B* **2005**, *109* (33), 15760.

(64) Rosenbach Jr., N.; Jobic, H.; Ghoufi, A.; Salles, F.; Maurin, G.; Bourrelly, S.; Llewellyn, P. L.; Devic, T.; Serre, C.; Ferey, G. *Angew. Chemie, Int. Ed.* **2008**, *47* (35), 6611.

(65) Babarao, R.; Hu, Z.; Jiang, J.; Chempath, S.; Sandler, S. I. *Langmuir* **2007**, *23* (2), 659.

(66) Halgren, T. A. *J. Comput. Chem.* **1996**, *17* (5-6), 490.

(67) Stewart, J. J. P. *J. Mol. Model.* **2007**, *13* (12), 1173.

(68) Frisch, M. J. et al. Gaussian 09, Revision A.02. *Gaussian 09, Revision A.02*, 2009.

(69) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77* (18), 3865.

(70) Blöchl, P. E. *Phys. Rev. B* **1994**, *50* (24), 17953.

(71) Kresse, G. *Phys. Rev. B* **1999**, *59* (3), 1758.

(72) Grimme, S. *J. Comput. Chem.* **2006**, *27* (15), 1787.

(73) Forrester, T. R.; Smith, W. DL_POLY ver. 2.18 (C.L.R.C. Daresbury Laboratory, Daresbury, UK, 1995).

(74) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, flexible, and free. *Journal of Computational Chemistry*, 2005, *26*, 1701–1718.

3. 3D-QSAR Accelerated Discovery of Ice Recrystallization Inhibitors

This project originated as a collaborative effort between the Ben Lab and Woo Lab to accelerate the discovery of ice recrystallization inhibitors. The largest cause of cell death during the freezing process is due to ice recrystallization. Developing molecules which can inhibit ice recrystallization is a prominent challenge in current cryogenics. The Ben Lab has been able to develop small molecules with varying structures and IRI activities. Unfortunately, chemical intuition has failed to find a discernable trend in structure and activity. Thus, we have used molecular simulation to develop a three dimensional quantitative structure-activity relationship (3D-QSAR) model to predict whether a molecule is IRI active or inactive. The following chapter will detail the results of our efforts. I was specifically responsible for generating three dimensional structures, performing conformational searches, calculating the electron density and electrostatic potential via DFT, and developing the 3D descriptors. Former post-doc Michael Fernandez was responsible for building the QSAR model, validation, and testing.

As previously mentioned in Section 1.1, in recent years there has been significant progress in the development of IRI molecules from protein anti-freeze analogues to small molecules. Much of this pioneering work has been done by the Ben Lab.[1–5] There have been several experimental structure-function studies of anti-freeze glycoproteins which have been able to display ice recrystallization activity without the detrimental thermal hysteresis activity.[6–10] However, the large size and considerable challenges with the synthesis and preparation of antifreeze proteins have limited their ability to be screened in a high-throughput fashion.

Recently, the Ben Lab was able to demonstrate that small-molecule carbohydrate based surfactants and hydrogelators possessed IRI activity.[11] Since this discovery it was found that amphiphilic nature is important to IRI activity where a balance between hydrophobic and hydrophilic functionality was essential.[4] In that work it was found that the presence of long hydrophobic alkyl chains increased IRI activity. Since then, the Ben Lab has been busy synthesizing and testing many small molecule ice recrystallization inhibitors, many of which have been unpublished. While there have been preliminary non quantitative structure-function studies to reveal the importance of the amphiphilic nature, the diversity of the molecular set demands a more rigorous quantitative structure-activity relationship (QSAR) analysis to be done.

3.1. Experimental Design and Synthesis

When initial experimental design of IRI molecules began it became apparent that hydrophobic and hydrophilic portions were essential. The small molecule library in this thesis consists of the hydrophilic portion based on open and closed chain carbohydrates. Initially, long alkyl chains were used as the hydrophobic portion on these small molecules, but in order to increase hydrophobicity a new class of small IRI molecules were envisioned whereby the alkyl chains were replaced with an aryl ring. (Figure 3.1)

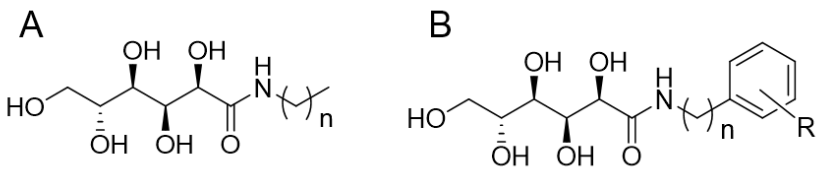
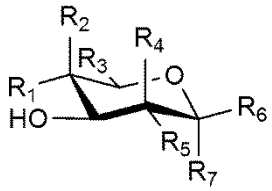


Figure 3.1: The general structure of small molecule carbohydrate based amphiphilic hydrogelators and surfactants. The hydrophobic moiety was first based on long alkyl chains (A) and then functionalized aryl rings (B).

In addition to providing increased hydrophobicity, the aryl ring provides a wealth of functionality potential simply by altering the substituent on the aryl ring. This provided a route to study the effects of functional groups on the hydrophobic moiety as well. The experimental synthesis, characterization, and IRI testing was done by various members of the Ben Lab throughout the years. These experimental synthetic details of each molecule are beyond the scope of this thesis. Compounds were characterized using NMR and consistently tested using the standardized splat cooling assay method as described in section 1.1.4. The IRI activity was measured as the ratio between the ice grain size of a 22mM solution of the tested compound and a 22mM phosphate buffer solution (PBS) standard. For example, if the grain size of ice in a 22 mM solution of PBS was found to be 30 µm, and the grain size of ice in a 22 mM solution of an IRI molecule was 15 µm, then that IRI molecule would have an IRI activity of 50% mean grain size. To develop the 3D-QSAR model we used a total of 124 unique and structurally diverse IRI molecules that were experimentally synthesized and tested. A full list of all structures and IRI activities expressed at percent mean grain size is presented below. Cells highlighted in blue and green are compounds used in QSAR training and test sets respectively.

Table 3.1: Structures and IRI activities for the 124 small IRI molecules as synthesized, characterised, and tested by the Ben Lab. IRI activity is measured as percent mean grain size to a PBS standard. All concentrations used were 22 mM.

Base Structure



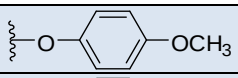
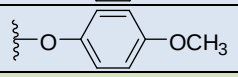
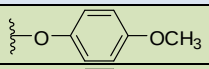
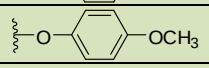
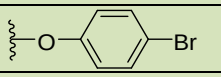
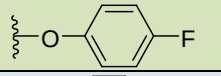
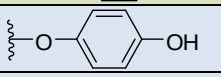
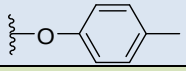
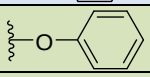
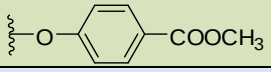
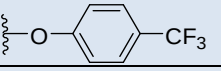
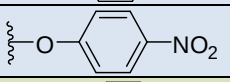
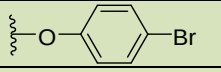
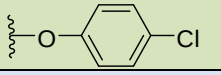
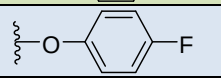
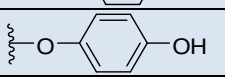
Entry	R_1	R_2	R_3	R_4	R_5	R_6	R_7	IRI Activity % MGS
1	H	OH	CH_2OH	H	OH		H	78.303
2	OH	H	CH_2OH	H	OH		H	23.302
3	H	OH	CH_2OH	H	OH	H		76.44
4	OH	H	CH_2OH	H	OH	H		63.209
5	OH	H	CH_2OH	H	OH		H	76.261
6	OH	H	CH_2OH	H	OH		H	10.27
7	OH	H	CH_2OH	H	OH		H	14.426
8	OH	H	CH_2OH	H	OH		H	61.798
9	OH	H	CH_2OH	H	OH		H	62.260
10	OH	H	CH_2OH	H	OH		H	64.436
11	OH	H	CH_2OH	H	OH		H	49.132
12	OH	H	CH_2OH	H	OH		H	33.221
13	OH	H	CH_2OH	H	OH		H	60.892
14	H	OH	CH_2OH	H	OH		H	6.169
15	H	OH	CH_2OH	H	OH		H	21.09
16	H	OH	CH_2OH	H	OH		H	79.658
17	H	OH	CH_2OH	H	OH		H	81.060

Table 3.1 Continued:

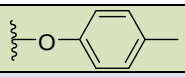
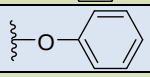
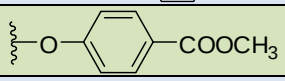
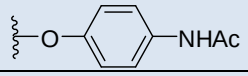
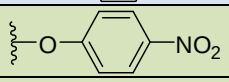
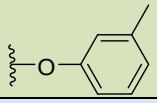
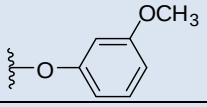
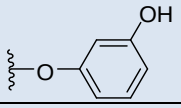
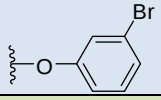
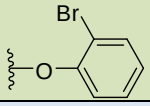
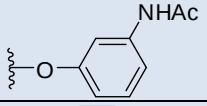
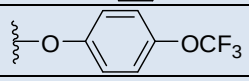
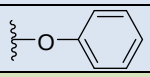
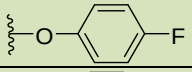
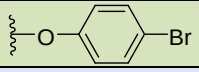
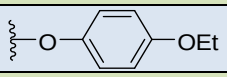
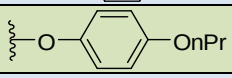
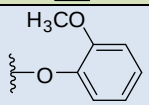
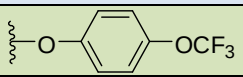
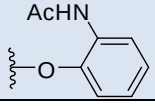
18	H	OH	CH_2OH	H	OH		H	88.314
19	H	OH	CH_2OH	H	OH		H	79.061
20	H	OH	CH_2OH	H	OH		H	29.891
21	H	OH	CH_2OH	H	OH		H	81.855
22	H	OH	CH_2OH	H	OH		H	57.878
23	H	OH	CH_2OH	H	OH		H	45.615
24	OH	H	CH_2OH	H	OH		H	79.061
25	OH	H	CH_2OH	H	OH		H	80.745
26	OH	H	CH_2OH	H	OH		H	90.061
27	H	OH	CH_2OH	H	OH		H	35.944
28	H	OH	CH_2OH	H	OH		H	71.169
29	H	OH	CH_2OH	H	OH		H	72.106
30	OH	H	CH_2OH	H	OH		H	57.368
31	OH	H	CH_2OH	H	OH	H		69.723
32	OH	H	CH_2OH	H	OH	H		80.695
33	OH	H	CH_2OH	H	OH	H		75.401
34	OH	H	CH_2OH	H	OH		H	31.806
35	OH	H	CH_2OH	H	OH		H	39.655
36	OH	H	CH_2OH	H	OH		H	54.439
37	H	OH	CH_2OH	H	OH		H	72.310

Table 3.1 Continued:

38	H	OH	CH_2OH	H	OH		H	27.052

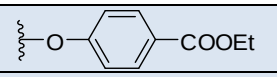
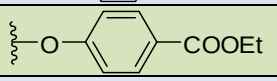
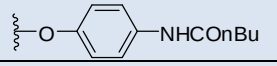
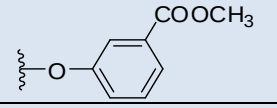
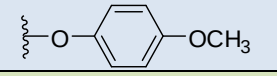
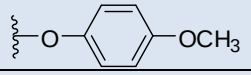
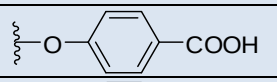
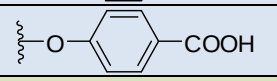
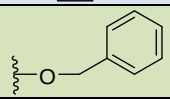
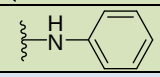
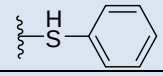
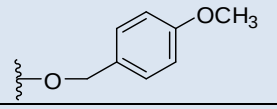
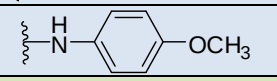
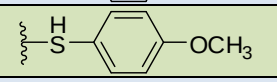
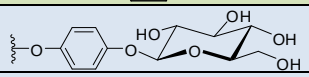
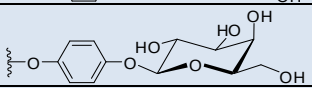
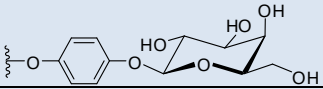
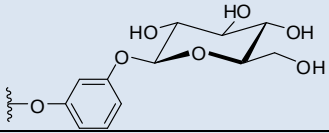
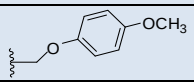
39	H	OH	CH_2OH	H	OH		H	76.330
40	H	OH	CH_2OH	H	OH		H	35.140
41	OH	H	CH_2OH	H	OH		H	5.844
42	OH	H	CH_2OH	H	OH		H	17.770
43	H	OH	CH_2OH	H	OH		H	56.891
44	OH	H	H	H	OH		H	47.109
45	OH	H	CH_2OH	OH	H		H	77.702
46	OH	H	CH_2OH	OH	H	H		73.754
47	OH	H	CH_2OH	H	OH		H	73.311
48	H	OH	CH_2OH	H	OH		H	82.457
49	OH	H	CH_2OH	H	OH		H	71.858
50	OH	H	CH_2OH	H	OH		H	55.328
51	OH	H	CH_2OH	H	OH		H	55.328
52	OH	H	CH_2OH	H	OH		H	69.141
53	OH	H	CH_2OH	H	OH		H	62.115
54	OH	H	CH_2OH	H	OH		H	79.970
55	OH	H	CH_2OH	H	OH		H	34.100
56	H	OH	CH_2OH	H	OH		H	79.228
57	OH	H	CH_2OH	H	OH		H	58.008
58	OH	H	CH_2OH	H	OH		H	60.967
59	OH	H		H	OH	OH (α/β mixture)		82.510

Table 3.1 Continued:

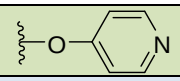
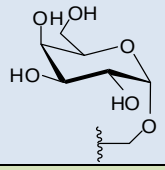
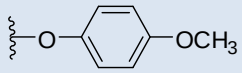
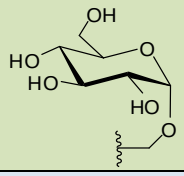
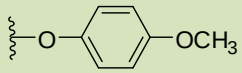
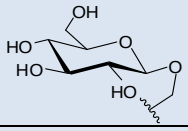
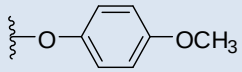
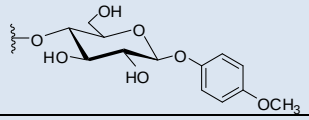
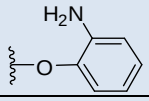
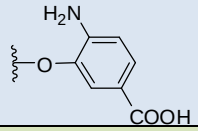
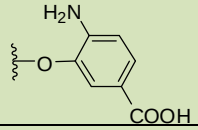
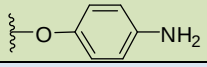
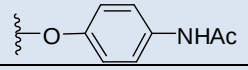
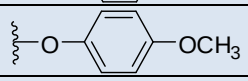
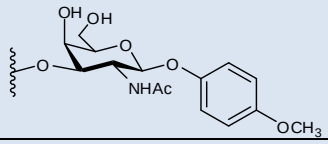
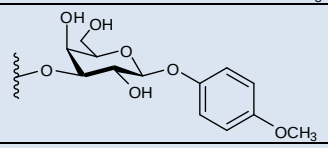
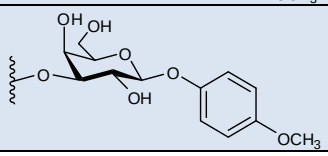
60	OH	H	CH_2OH	H	OH		H	51.586
61	OH	H		H	OH		H	66.096
62	OH	H		H	OH		H	66.766
63	OH	H		H	OH		H	61.557
64	H	OH	CH_2OH	H	OH		H	44.525
65	H	OH	CH_2OH	H	OH		H	83.909
66	H	OH	CH_2OH	H	OH		H	40.541
67	OH	H	CH_2OH	H	OH		H	88.603
68	OH	H	CH_2OH	H	OH		H	88.813
69	OH	H	CH_2OH	H	OH		H	70.163
118	H	OH	H	H	OH		H	47.109
121	H	OH	CH_2OH	H	OH		H	50.624
122	OH	H	CH_2OH	H	OH		H	77.798
123	H	OH	CH_2OH	H	OH		H	64.710

Table 3.1 Continued:

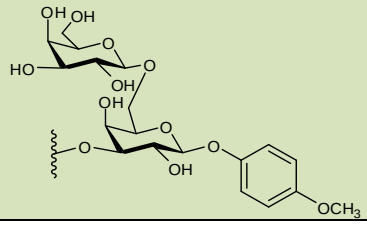
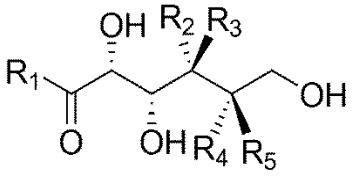
124	H	OH	CH_2OH	H	OH		H	59.067

Table 3.1 Continued:

Base Structure



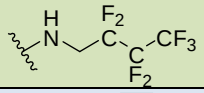
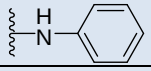
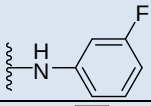
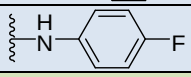
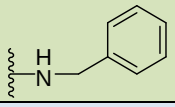
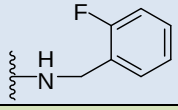
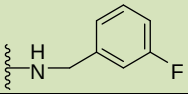
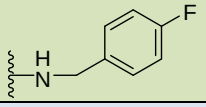
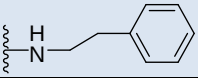
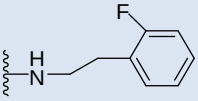
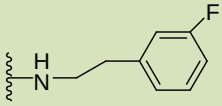
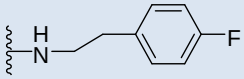
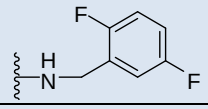
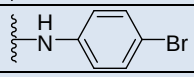
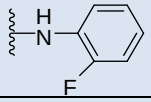
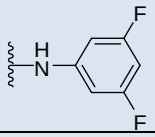
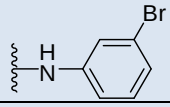
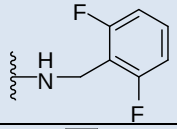
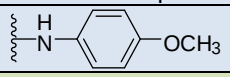
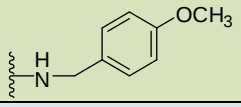
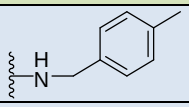
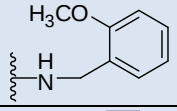
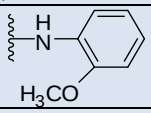
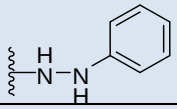
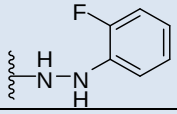
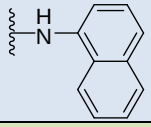
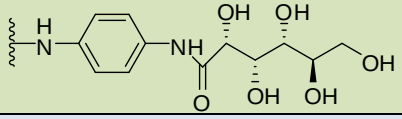
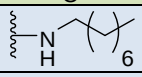
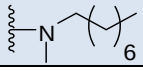
Entry	R_1	R_2	R_3	R_4	R_5	IRI Activity % MGS
70		OH	H	H	OH	32.8
71		OH	H	H	OH	65.7
72		OH	H	H	OH	9.41
73		OH	H	H	OH	90.3
74		OH	H	H	OH	38.1
75		OH	H	H	OH	35.4
76		OH	H	H	OH	30.7
77		OH	H	H	OH	19.7
78		OH	H	H	OH	29.7
79		OH	H	H	OH	20.6
80		OH	H	H	OH	30.2
81		OH	H	H	OH	69.0

Table 3.1 Continued:

82		OH	H	H	OH	9.7
83		OH	H	H	OH	75.0
84		OH	H	H	OH	3.0
85		OH	H	H	OH	4.4
86		OH	H	H	OH	40.0
87		OH	H	H	OH	12.6
88		OH	H	H	OH	3.9
89		OH	H	H	OH	6.5
90		OH	H	H	OH	5.045
91		OH	H	H	OH	72.808
92		OH	H	H	OH	100
93		OH	H	H	OH	81.1
94		OH	H	H	OH	68.7
95		OH	H	H	OH	103.8
97		OH	H	H	OH	66.74
98		H	OH	H	OH	72.17
100		OH	H	H	OH	71.87

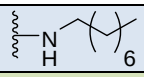
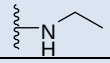
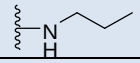
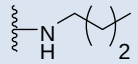
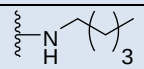
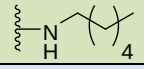
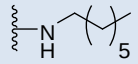
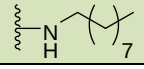
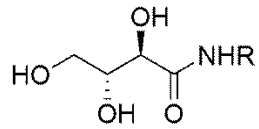
108		OH	H	H	OH	12.11
109		OH	H	H	OH	52.5
110		OH	H	H	OH	102.5
111		OH	H	H	OH	104.3
112		OH	H	H	OH	87.3
113		OH	H	H	OH	22.7
114		OH	H	H	OH	12.4
115		OH	H	H	OH	5.9
116		OH	H	H	OH	18.5
117		OH	H	H	OH	8.6

Table 3.1 Continued:



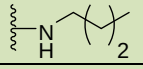
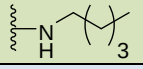
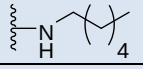
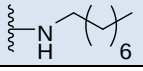
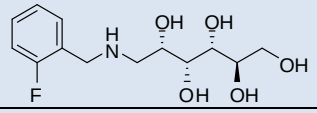
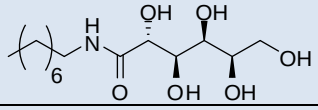
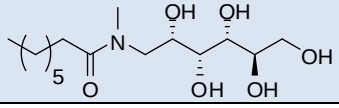
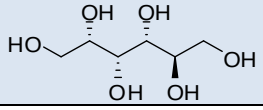
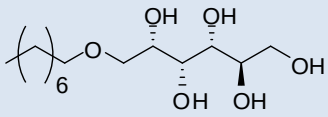
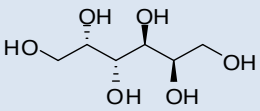
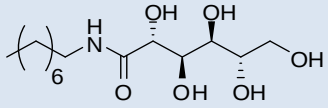
Entry	R	IRI Activity % MGS
104		71.222
105		73.597
106		68.666
107		56.342

Table 3.1 Continued:

Entry	Structure	IRI Activity % MGS	Entry	Structure	IRI Activity % MGS
96		30.5	**103**		72.37
99		66.26	**119**		82.31

101		59.84	**120**		87.92
102		21.99			

While there is a wealth of data available within this library of experimental molecules, there are no straightforward structural features that clearly correlate to the IRI activity. In terms of stereochemistry of axial vs. equatorial positions there seem to be no discernable trend. For example, by simply altering the position of the C4 hydroxyl group from axial (**1**) to equatorial (**2**) for a pair of identical molecules there is a drastic improvement of IRI activity from 78.3% to 23.3% MGS. However, if looking at another pair identical molecules, altering the C4 hydroxyl group from axial (**14**) to equatorial (**5**) results in a decrease in IRI activity from 6.2% to 76.3% MGS. There also seems to be no correlation to electronegativity of the substituent group. Take for example a set of molecules where the only variance is the substituent on the para position of the aryl group. There seems to be no trend between increasing electronegativity of the functional groups OH (**8**), Br (**5**), Cl (**6**), F (**7**) and IRI activity of 61.8%, 76.3%, 10.3%, and 14.4%, respectively. Even the length of the alkyl chain is also not well correlated. By increasing the size of the substituent in terms of butyl (**113**), pentyl (**114**), hexyl (**115**), heptyl (**116**), and nonyl (**117**) resulted in IRI activities of 22.7%, 12.4%, 5.9%, 18.5% and 8.6%. Clearly, human chemical intuition is not sufficient to guide the design of IRI molecules. Further exacerbating the problem is that the details and mechanisms by which these molecules function is still unknown, and likely to be very difficult to ascertain. Thus, other methods such

as quantitative structure activity relationships offer a way to determine correlations that chemical intuition cannot.

The main goal for this project is to help accelerate the experimental discovery of IRI molecules. It is important to provide a preliminary screening of proposed molecules so that only those most likely to be active are synthesized and tested. This would enable experimentalists to more rapidly find molecules of interest while also saving time and resources.

3.2. Computational QSAR Model Implementation

To construct a QSAR model, the molecular structure and experimental activity are required. The data set of 124 structures provided by the Ben lab possessed both favourable and less than ideal characteristics in terms of building a QSAR model. First, the IRI activity of the molecules ranged from a MGS of 3% corresponding to exceptional IRI activity range to a MGS of 104%, which essentially showed activity worse than the pure buffered solution. Secondly, the library of structures is sufficiently diverse with many different functional groups, stereochemical configurations, and molecular weights. On the other hand, the data set was heavily skews towards poorly active molecules. Of the 124 molecules, only 27 are considered highly active (%MGS < 30%), 13 are considered moderately active (%MGS < 70%), and 65 are considered poorly or not active (% MGS > 70%). (Figure 3.2)

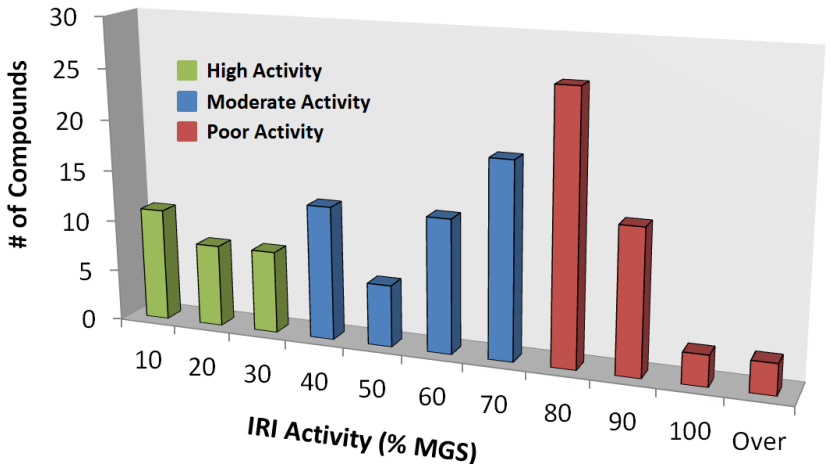


Figure 3.2: A histogram displaying the distribution of IRI activity within the initial set of 124 molecules tested for activity.

The goal of a QSAR study is to correlate the structural feature or fingerprints to the IRI activity, in this case the MGS relative to the PBS buffered solution. Perhaps the greatest importance to a successful QSAR model that can accurately predict activity is the quality of the descriptor. However, what features of the molecular structure, if any and what descriptors that correlate these features to the activity are not known *a priori*. Thus, in building a QSAR model, one may use many descriptors such as the constitutional, topological, geometric and others as described in section 2.1.1.1. Three dimensional fingerprints offer the highest level of structural information and model interpretation capability as shown by its extensive use in form of the comparative molecular field analysis (CoMFA) method in drug design.[12,13] CoMFA works by representing a molecule by their steric and electrostatic fields. These parameters are taken from calculating the energies of steric (van der Waals potential) and electrostatic (Coulombic potential) interaction between the molecule of interest and a "probe atom" placed at nodes of a consistently spaced three-dimensional lattice. These steric and electrostatic fields are calculated for all the molecules in the data-set and used as a descriptor for further analysis. The

greatest drawback to this method is that all molecules within the set are highly alignment dependent. That is, comparison of the molecular field is only valid if the fields are calculated from the same reference orientation. Thus, if the molecules are so diverse that they do not have the same "parent" structure then the comparisons between molecules will be challenging. Unfortunately, in the case of the IRI molecule library, the molecules are sufficiently diverse that consistent alignment could not be attained throughout the set. Therefore there was a need to turn to alignment independent descriptors for this application.

Grid-Independent Descriptors (GRIND) calculate an alignment independent descriptor by calculating the maximum product between any two nodes on the surface and then plotting them to specific distance bins to create unique fingerprints.[14] By only looking at the distances between nodes and not their orientation within space, GRIND eliminates the need to have all molecules oriented the same way before processing. In this specific work the maximum pair products of properties (surface curvature and electrostatic potential) at nodes in the van der Waals surface were calculated to generate a unique fingerprint. With this distance-between-features scheme the fingerprint corresponds to the interaction between two nodes of a certain distance on the molecular surface. More details on this will be described in Section 3.2.3.

3.2.1. Three-Dimensional Structure Generation

The initial set of 124 structures with experimental data was given in the format of 2D ChemDraw structures. Thus, 3D structures had to be generated and the correct conformation had to be determined. The models were generated using the Spartan program[15] and optimized using the Merck Molecular Force Field (MMFF)[16] before a conformational search was done. The

conformational search was done using a Monte-Carlo algorithm with the moves being rotations of bonds. The energy was evaluated at the PM6 level of theory.[17] This procedure was carried out until 100 conformers were generated. In order to maintain similarity within the structures the conformational search was done for a base molecule and then this structure was functionalized rather than a conformational search done on all functionalized molecules separately. Finally, single point DFT calculations at the B3LYP/6-311G(d,p) level of theory were performed using the Gaussian 09 software[18] to calculate the total electronic density distribution as well as the electrostatic potential.

3.2.2. Molecular Surface Recognition

The concept of using the curvature at the van der Waals surface as a descriptor of molecular shape was first proposed by Fontaine *et al.*[14] The curvature gives a metric to describe how convex or concave a certain node is on the molecular surface. The curvature calculation is done in a few steps which will be outlined below and a graphical representation is provided in Figure 3.3.

1. **A molecular surface is calculated**: This is done via isosurface values calculated form a charge density distribution from DFT methods.[19] An isosurface is a surface generated from points of a constant value. It is calculated by first identifying the grid points with an electron density value which corresponds to the vdW surface. These nodes are connected to define a molecular surface.
2. **The nearest-neighbours of a node are selected:** The curvature of a node directly relies on its orientation with respect to its nearest neighbours. Therefore the

nearest neighbours have to be identified at a certain cut-off. Typically an empirically determined Euclidean distance of 6 Å is used as a sufficient cut-off limit for small molecules.[14]

3. **Calculation of curvature coefficient*:*** The curvature coefficient is calculated for each nearest neighbour N_i from the following equation, (**Eq. 3.1**)

$$C_f = \cos(\alpha) = \frac{xx' + yy' + zz'}{\sqrt{x^2+y^2+z^2}\cdot\sqrt{x'^2+y'^2+z'^2}} \qquad \textbf{(Eq. 3.1)}$$

where x, y, and z are the components of the vector perpendicular ($\vec{v}$) to the surface node *R*. x', y', and z' are the components of the vector between the surface node *R* and its nearest neighbour N_i and α is the angle between those two vectors. If these two vectors are perpendicular the surface between the node *R* and its nearest neighbour is planar and the curvature will be zero. If α is above or below π/2 then the surface is concave ($C_f > 0$) or convex ($C_f < 0$) respectively depending on how the vector v is determined. The total curvature is defined as the median of the partial curvature coefficients for all the nearest neighbours which allows for less disturbance by extreme values.

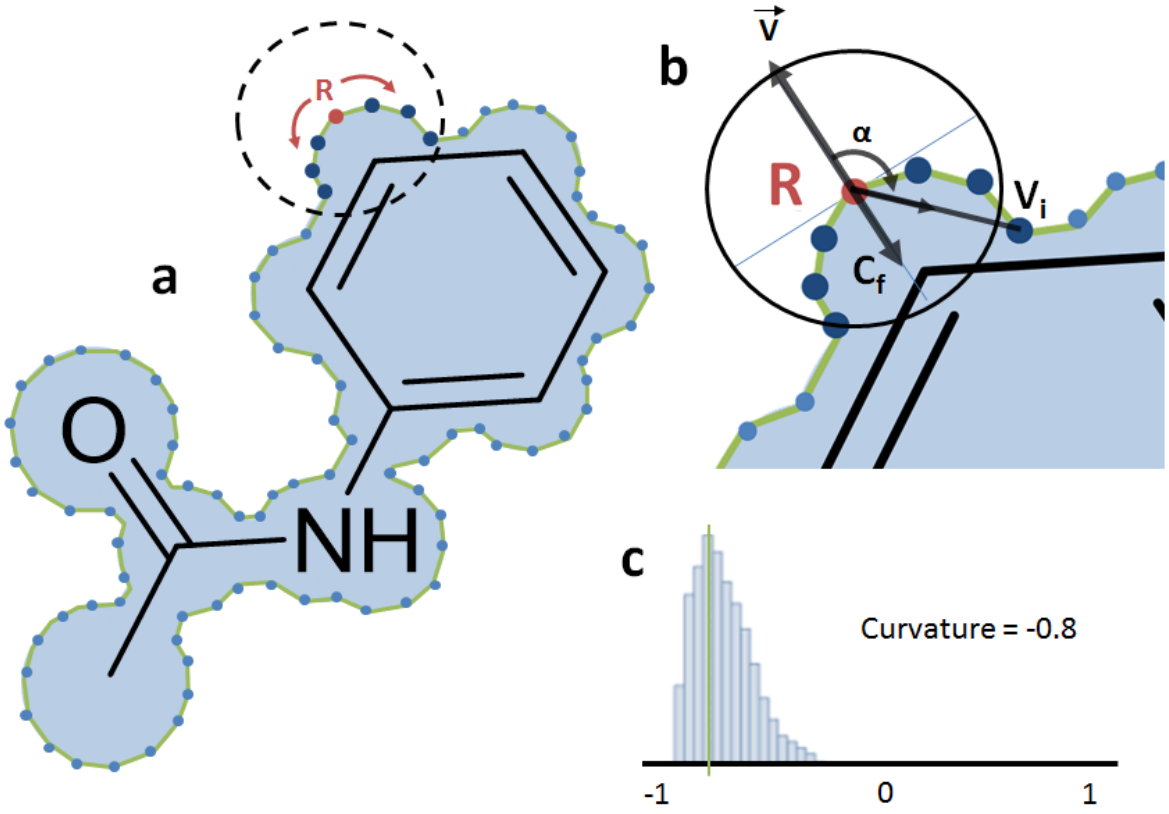


Figure 3.3: Schematic representation of a curvature calculation. (**a**) First the nearest-neighbours for each surface node R are found within a cut-off of 6 Å. (**b**) The partial curvature coefficient (Cf) is calculated for each nearest neighbour (N_i). (**c**) The total curvature is defined as the median of the Cf distribution for all nearest neighbours partial curvatures.

3.2.3. Fingerprint Generation

The molecular surface was generated from the DFT-derived electronic density distribution. The molecular surface was defined by the total electron density isosurface value of 0.0135 electrons/Bohr^3, which was empirically shown to correspond to the vdW surface.[19,20] The values were generated at nodes separated 0.176 Å apart on a rectangular grid. This was done for both the total electronic density where the values at the nodes are the curvature and for the electrostatic potential where the values of the nodes are the electrostatic potential at that node. This allows for the steric and electrostatic features to be represented on the

molecular surface. This was calculated using an in-house code that was first written by former Ph.D student Nick Trefiak to study catalysts and then modified to allow for large molecules for this project to be run. The maximum product of the values at the surface nodes are now used as maximum auto- and cross correlation (MACC) scores to develop alignment-independent fingerprints. These maximum products are assigned to a distance bin according to the node separation whereby the number of bins was 200. Thus, at each specific distance bin is the maximum curvature-curvature, esp-esp, or curvature-esp product between nodes of the molecular surface. This procedure was derived from the GRIND method as described by Pastor et al.[14] and the MACC equations are listed below.

Maximum auto-correlations (*MAC*):

$$MAC_{r_o}{}^{Curvature} = max\left[\left\{\left|\delta_{ij} \times \rho_i^{curv} \times \rho_j^{curv}\right|\right\}_{i=1\ldots(N-1);j=(i+1)\ldots N}\right] \quad \textbf{(Eq. 3.2)}$$

$$MAC_{r_o}{}^{ESP} = max\left[\left\{\left|\delta_{ij} \times \rho_i^{ESP} \times \rho_j^{ESP}\right|\right\}_{i=1\ldots(N-1);j=(i+1)\ldots N}\right] \quad \textbf{(Eq. 3.3)}$$

Maximum cross-correlations (MCC):

$$MCC_{r_o}{}^{Curv-ESP} = max\left[\left\{\left|\delta_{ij} \times \rho_i^{Curv} \times \rho_j^{ESP}\right|\right\}_{i=1\ldots N;j=1\ldots N}\right] \quad \textbf{(Eq. 3.4)}$$

where δ_{ij} is the distance between the nodes *i* and *j*, ρ_i^{curv}, ρ_j^{curv}, ρ_i^{ESP}, ρ_j^{ESP}, are the values of the curvature of the vdW surface and electrostatic potential (ESP) at the nodes *i* and *j* on the molecular surface with a total number of node *N* with r_o as the distance bins calculated from the minimum to maximum distance between nodes. This quantum-mechanical derived descriptor allows for the independence of alignment between molecules because the distance

between features is not dependent on orientation or directionality. Thus, each maximum product represents the interaction between nodes at the molecular surface and the combined maximum products at each distance bin form what is known as a correlogram. The correlogram formed from the maximum auto- and cross- correlation data form a unique chemical fingerprint representing the molecular structure. (Figure 3.4)

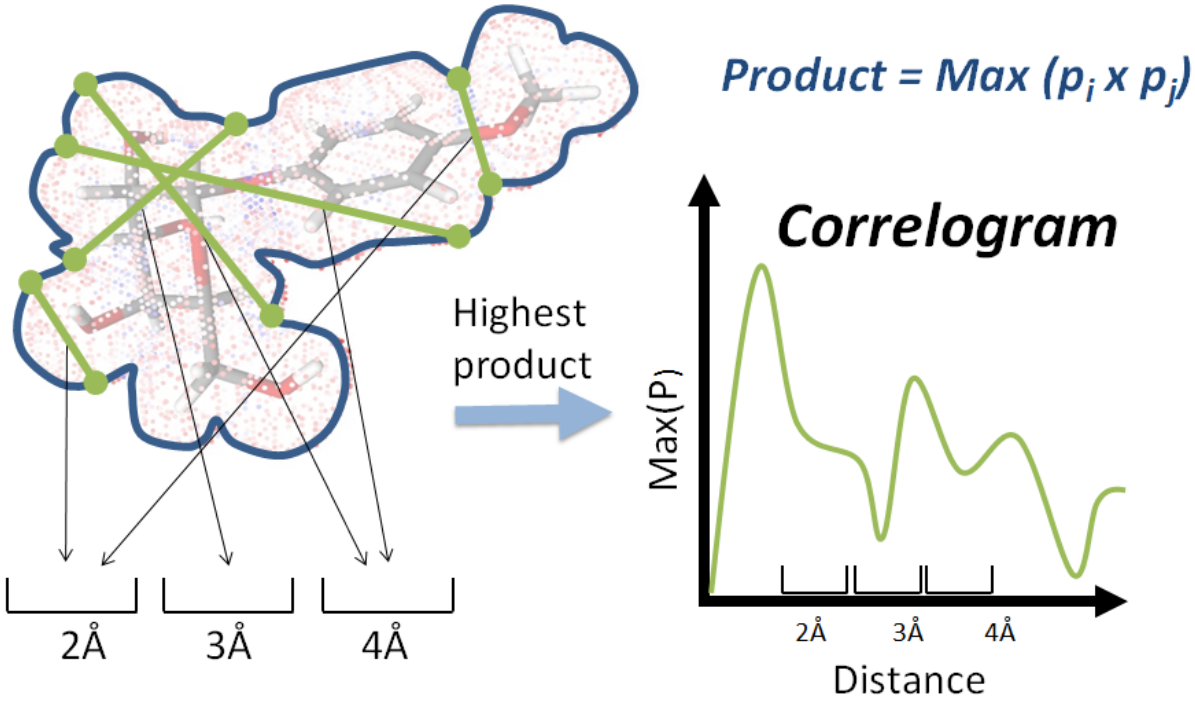


Figure 3.4: Schematic representation of the calculation of the molecular fingerprint via the GRIND method. (Left) The interaction is calculated as the product of the curvature of the electron density or electrostatic potential. The MACC transform gives the correlogram (Right) which is a histogram of activity with respect to distance of maximum interaction.

Once descriptors for all 124 molecules in the set were prepared they had to be partitioned into a training set to build the QSAR model and a test set to test its predictive potential. The training set is comprised of 84 compounds (68% of the total set) and the test set is comprised of 40 compounds (32% of the total set). The training and test sets were chosen at random but were ensured to contain the same ratio of active to inactive molecules. Details of these structures are found in Table 3.1 where those highlighted in blue are the training set and

those highlighted in green are the test set. Partial least square regression (PLSR) models[21] were calibrated using the GRIND descriptors of the training set. The number of descriptors from this method was quite high with ~150 descriptors for each molecule. A descriptor is one of the maximum auto- or cross- correlation values. For example, the maximum product of the electronic density curvature and electrostatic potential at a value of 8Å is considered a descriptor. The number of features (variables in the linear equation) needs to be small enough such that over-fitting does not occur. Thus, PLSR alone can produce poor results and an additional feature selection step must be implemented. To this effect a genetic algorithm (GA)[22] was utilized to optimize the descriptors to reduce the number to only 23 descriptors with which to build the QSAR model.

3.2.4. Genetic Algorithm Feature Selection

The genetic algorithm (GA) approach to descriptor selection has been widely used and proved to be quite effective.[23,24] This method is based on the principle of Darwinian evolution where by optimal features are carried down through generations. First, a random set of descriptors is selected from the original pool, in this specific case these are the maximum cross and auto correlations for each IRI molecule. Next, a multivariate regression technique is used to develop a QSAR equations for the set. This is repeated for another random set of descriptors. These sets of descriptors form what is called an “individual”. A first generation population of 100 individual sets of descriptors is created along with the QSAR equations for each individual. Then each individual equation is ranked using a lack-of-fitness function.[23,24] The fitness function

used for the GA in this project was an F-score (a measure of accuracy), defined as the harmonic mean of the precision and sensitivity, (**Eq. 3.5**)

$$F = 2 \cdot \frac{precision \cdot sensitivity}{precision + sensitivity} \quad \textbf{(Eq 3.5)}$$

where precision is equal to the ratio of true positives over the sum of true positives and false positives (TP/(TP + FP)) and sensitivity is the ratio of true positives over the sum of true positives plus false negatives (TP/(TP + FN). A true positive is a molecule that was predicted to be active and is actually active, a false positive is a molecule that was predicted to be active and is actually not active, and a false negative is a molecule that was predicted to be inactive but is actually active. Two QSAR "parent" equations can be mated to produce "offspring" QSAR equations by randomly taking descriptors from both parents. These offspring will take the place of the worst ranked QSAR equations in the pool and this process is repeated until an optimal set of descriptors are found. In this specific work each individual set of descriptors was represented as a binary string with a size equal to the number of descriptors. Initially a population of 100 binary individual models was randomly selected for the first generation. Each bit in the binary string had a probability of 0.1 to have the value 1 which corresponds to the descriptor being in the model. Mating yielded a 90% offspring success during evolution and mutation had a 10% offspring success meaning that mating resulted in 90% of the resulting models performing well with the fitness function while only 10% of mutated models did well. A total of 100 independent GA runs were performed with a cutoff of 1000 generation or when 90% of the generations reached the same fitness score. This method was written in python by

Dr. Michael Fernandez using a generic GA implementation with the NIPALS algorithm for PLSR analysis.[26]

The resulting optimal QSAR model was found to contain 26 features. The linear equation is of the form $y = m_1x_2 + m_2x_2 + m_3x_3 \ldots + m_nx_n + c$ where m is the optimized coefficient for feature *i,* x is the feature which takes the form of the maximum steric-steric, electrostatic-electrostatic, or steric-electrostatic interaction at a certain distance. The optimal QSAR formula is shown below in Figure 3.5.

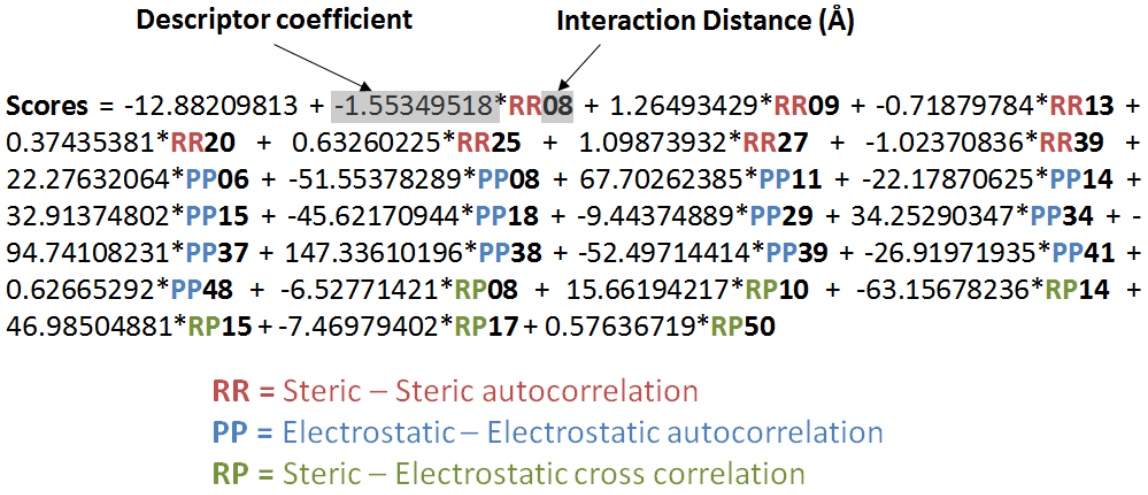


Figure 3.5: Optimal QSAR linear equation as found after 100 GA runs. RRX, PPX, and RPX are the maximum steric-steric autocorrelation, electrostatic-electrostatic autocorrelation, and steric-electrostatic cross correlation at distance X in angstroms respectively.

Thus, to calculate a score for a molecule all one needs are the unique structural fingerprint as calculated at the specified interatomic distances. Once again a binary classifier was built so that the score will determine whether the molecule is active or inactive. The model was *not* built such that the score determines a specific IRI activity. The resulting score was built to lie between 0 and 1 where 0 is inactive and 1 is active and the cutoff for classification is 0.5

where any molecule below 0.5 is deemed to be IRI inactive and any molecule above 0.5 is deemed IRI active.

3.2.5. Cross-Validation

The QSAR model was made to be a classifier whereby the IRI activities were classified as either active (% MGS < 70) and inactive (% MGS > 70). It was found that this model successfully identified 80% of the IRI active compounds in the training set (the compounds used to make the model) with precision of 0.8. Next, the same procedure was performed on the remaining 40 compounds (test set) and consistent performance was shown with 83% of active molecules were successfully identified with a similar precision of 0.8. The quality of fit of the training set of the model is measured by its R^2, (**Eq. 3.6**)

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(x_i - y_i)^2}{\sum_{i=1}^{N}(x_i - \bar{y})^2} \qquad \textbf{(Eq 3.6)}$$

where N is the total number of compounds, x_i is the predicted IRI activity, y_i is the experimental IRI activity of compound i, and $\bar{y}$ is the average experimental IRI activity. When R^2 is computed on the training set it provides a measure of "goodness of the fit" for the model. This can expected to be high since it is calculated on the set of molecules that the model was built on. A better indication of the model is the R^2 computed for the IRI activities of the test set.

In order to further test the model, a cross validation was performed. In cross validation, the total set of 124 molecules is partitioned into small subsets where the whole QSAR analysis is performed. This is done multiple times to reduce variability and gain confidence in the model.

The subsets are randomly chosen and include molecules from both training and tests sets. For this specific project, an exhaustive leave-one-out (LOO) cross-validation was performed. Here a single molecule is left out from the set, then the model is refitted, and finally the predictive activity for the molecule that was left out is compared to its actual activity. This is repeated until each molecule has been omitted once. The sensitivity, (the ratio of true positives over the sum of true positives and false negatives), specificity, (the ratio of true negatives over the sum of false positives and true negatives) and precision (the ratio of true positives over the sum of true positives and false positives) were calculated for the training, cross-validation, and test sets. The sensitivity was above 0.8 for training, cross-validation, and test sets. For specificity, only the test set (specificity = 0.72) had a value below 0.8. Finally, precision yielded good results for training (0.89), satisfactory results for cross-validation (0.78), and poor results for the test set (0.57).

Another statistical analysis measure used in QSAR studies is the area under the curve AUC) plot. This plot gives the probability that a classifier will rank a randomly chosen active molecule higher than a randomly chosen inactive molecule. The AUC is calculated from a plot of sensitivity, and specificity, which is the ratio of true negatives over the sum of false positives and true negatives (TN/FP +TN). This looks at the relationship between the true positive rate and true negative rate where the optimal curve would have a sensitivity of 1 (true positive rate of 100%) at any specificity value (true negative rate). AUC plots for the training set and test sets are shown below in Figure 3.6. It was found that the AUC for training and test sets were 0.834 and 0.830 respectively.

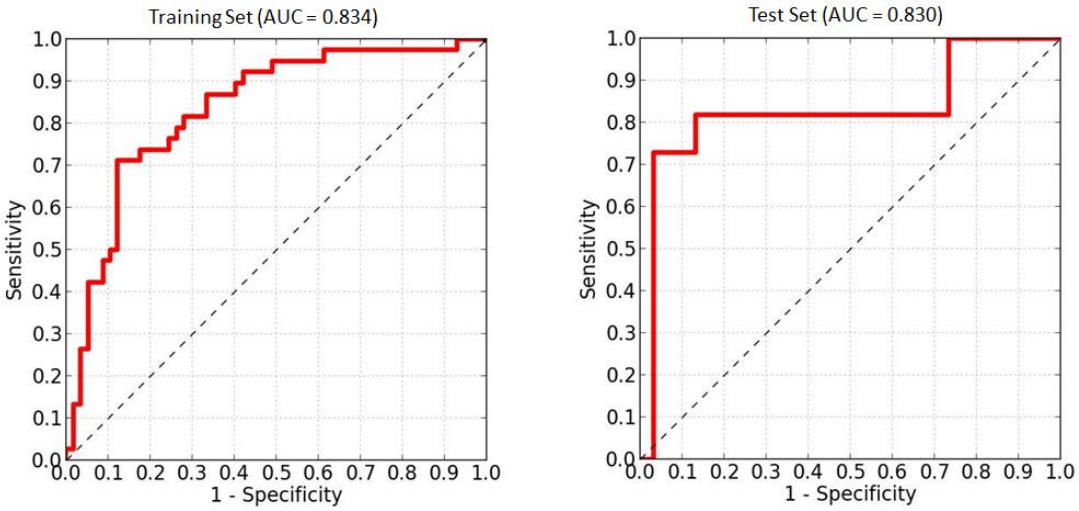


Figure 3.6: Area under the curve (AUC) plots for the training set and test sets. This is a graph of sensitivity vs. specificity and provides the probability that the model will properly classify molecules based on their activity.

3.2.6. Interpretation of 3D-QSAR Model

With the GRIND approach it is possible for the most relevant interactions (or features) to be traced back to the chemical structure. The optimum PLS model reveals that the features of greatest importance corresponds to the ESP-ESP interactions at 6.0 and 10.0 Å; and curvature-ESP interactions at 3.0 Å. (Figure 3.7) Interestingly, it was found that both active and inactive compounds exhibit prominent ESP-ESP interactions between the substituent in the aryl ring and the hydroxyl group of the carbohydrate ring. Meanwhile, the IRI active compound showed prominent ESP-ESP and curvature-ESP interactions between the aryl group and the substituent of the aryl group, while the inactive compound showed the same type of but between the hydroxyl groups of the carbohydrate group.

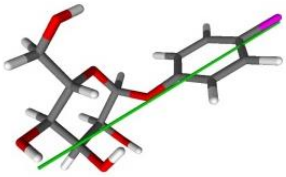
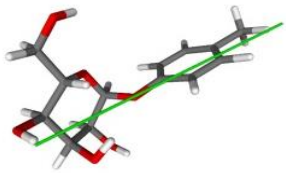
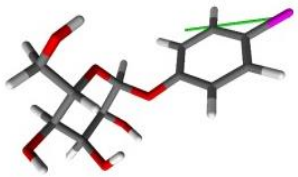
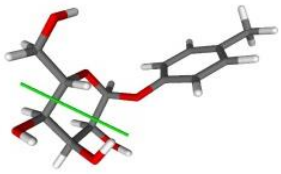
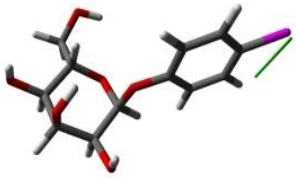
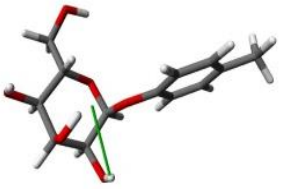
Strong IRI Activity (10.27% MGS)	**Weak IRI Activity (88.31% MGS)**
	
ESP-ESP ~10 Å	
	
ESP-ESP ~6 Å	
	
Curvature-ESP ~3 Å	

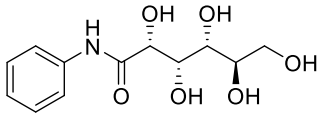
Figure 3.7: Most relevant features and their specific correlations for two molecules from the test set (left: **6**, right: **18**) that were predicted, and experimentally validated, to have good IRI activity and poor IRI activity.

3.3. Experimental Application of QSAR Predictions

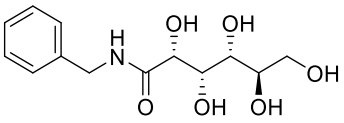
Once an optimal QSAR model was obtained and tested using the experimental set of 124 molecules, it was time to determine whether this model could classify activities of molecules which had yet to be made. To this effect, a library of 24 new molecules was proposed by the Ben Lab (Table 3.2). These molecules were specifically chosen for their ease of synthesis as many of the substituted aryl groups, such as 2-fluoroaniline (**Cpd. A1**), are commercially available. The same procedure was used to determine their predicted activity; a conformational search was performed, DFT level calculations were performed to calculate the electron density and electrostatic potential, GRIND descriptors were calculated with MACC transforms, and PLSR

were performed to fit the QSAR model with optimized coefficients. A list of the proposed molecules and their scores is shown below.

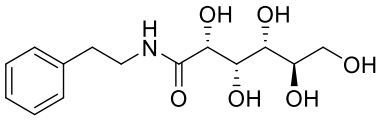
Table 3.2: Structures and computed QSAR scores of the proposed phenyl-alditol structures as provided by the Ben Lab. Those highlighted in green were predicted to be active, those in blue were predicted to be inactive.



Cpd.	Aryl Substituent(s)	QSAR Score
A1	2-F	-39.68
A2	4-F	-3.04
A3	4-CF_3	1.12
A4	3-CF_3	0.79
A5	2-OH	1.05
A6	3-OH	0.44
A7	3-OCF_3	0.76
A8	2-OCF_3	0.27
A9	2-Me	-0.07
A10	4-Me	0.93



Cpd.	Aryl Substituent(s)	QSAR Score
A11	2-Me	1.46
A12	2-CF_3	0.69
A13	3-CF_3	0.88
A14	4-CF_3	0.84
A15	4-OH	0.21
A16	2-Cl	-66.18
A17	4-Cl	-17.27
A18	2-OCF_3	1.16
A19	4-OCF_3	0.57
A20	2,4-F	-66.79



Cpd.	Aryl Substituent(s)	QSAR Score
A21	None	0.49
A22	2-F	0.01
A23	3-F	-0.57
A24	4-F	-0.18

The model was built such that the QSAR score would lie between 1 and 0 where the classifier score cut-off is 0.5 – 1 for IRI active and 0 – 0.5 for IRI inactive. Of the 24 molecules, almost half yielded scores outside the specified range of 0 – 1. However, it should be noted that only 4 molecules had scores drastically far outside from the range and all were inactive molecules. These numbers suggest that many of the compounds were not captured in our QSAR model, but as a classifier the absolute score is not as important as a compound's relative position to the cut-off. Thus, while some of these compounds were outside of the range of the model, they were still able to be classified as active or inactive. The reason for these outlier

values may be over-fitting with the descriptors, the binary nature of the model, insufficient size of training set, or diversity of proposed molecules compared to those in the training set.

These structures are based on three main base structures whereby the only difference is an increase in carbon linkages between the phenyl and secondary amine. The differences in the structures are functionalization of the benzene ring. The functional groups of CF_3, OCF_3, F, Cl, Me, and OH were chosen due to the fact that some of the best IRI activity in the training set had these functional groups (**84** – **90** in Table 3.1). Furthermore, these molecules were relatively simple to synthesize and thus could be made and tested within a reasonable timeframe.

The 24 structures were classified as either IRI active or IRI inactive where 13 of the proposed set were predicted to be active. In terms of the classification any molecule with a predicted percent mean grain size of less than 70% compared to PBS standard was considered active and those greater than 70% were considered inactive. 11 molecules that were predicted to be active were made and synthesized as well as 2 molecules which were predicted to be inactive. A graph of the experimentally measured % MGS for each of the molecules synthesized is shown below. (Figure 3.8)

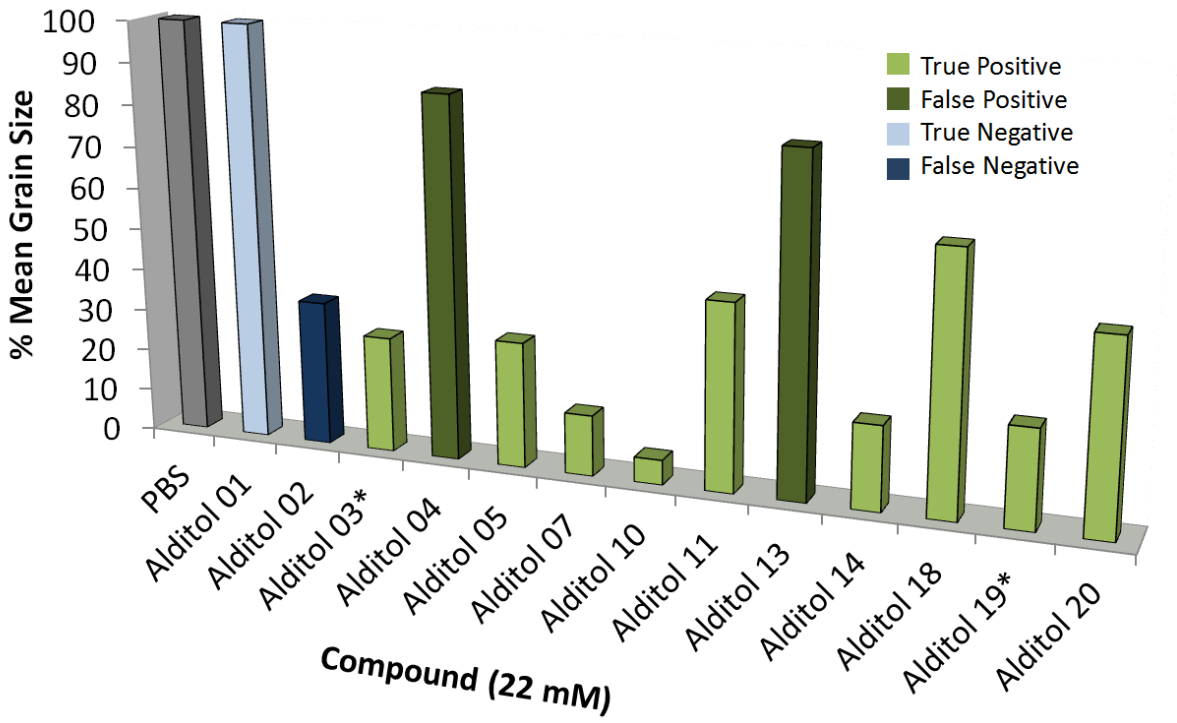


Figure 3.8: Graph of experimental percent mean grain size (MGS) of an ice crystal compared to the standard PBS solution. Highlighted in light green are the true positive compounds predicted to be IRI active and were experimentally found to be active, dark green are the false positive compounds which were predicted to be IRI active but had % MGS above 70%. Light blue are the true negative compounds predicted to be IRI inactive and gave experimental % MGS above 100% and dark blue are the false negative compounds predicted to be IRI inactive but gave % MGS below 70%. Compounds marked with an asterisk where tested at 11mM concentration.

It was found that nine of the compounds (82%) predicted as IRI active actually exhibited experimental IRI activity <70%. This precision rate is remarkably consistent with the training and test set precisions of 83%, which utilized experimental data from the set of 124 previously synthetized molecules. The molecule displaying the best activity was Compound **A10** (Table 3.2) – a 4-methylaniline-alditol. In fact, the two compounds which were deemed as inactive gave a mean grain size of 86.6% (**A4**) and 80.2 % (**A13**) which are still smaller than the PBS standard. Furthermore, compounds **A1** and **A2** which were predicted to be inactive were also experimentally synthesized and tested. It was found that **A1** indeed showed an inactive molecule with a % MGS greater than 100%. Surprisingly, compound **A2** was found to be active

with a MGS of 34.6%. While this may seem like a cause of concern, the goal is to identify true positives not false negatives. To this extent the model has been proven to predict activity with a success rate of greater than 80%. The greatest accomplishment of this model is the ability to forgo time consuming synthesize on compounds which would end up being inactive. The initial test and training set had only about 20% of the molecules as highly IRI active, which means the majority of the compounds synthesized used valuable time and resources which could have been better spent. By using this model as a pre-experimental screen, an experimentalist can focus solely on synthesizing molecules which are more likely to be active.

3.4. Conclusions

In conclusion, 3D-QSAR modeling provides an indispensable method to analyze the ice recrystallization activity of small molecules and provides a route to efficiently accelerate the discovery of new bioactive compound libraries. Given a diverse data set of 124 molecules, the built 3D-QSAR model showed accuracies >80% and revealed interesting intramolecular interaction patterns associated to the IRI activity. These patterns are characterized by prominent ESP-ESP and curvature-ESP interactions between the aryl group and the substituent of the aryl group of the active compounds, whilst the inactive compounds showed similar interactions but among the hydroxyl groups of the carbohydrate group. The 3D-QSAR model was then used to rationally design a small library of novel IRI active compounds with promising results, which exhibits similar precision rates to that of calibration and validation, albeit missing a few compounds.

3.5. References

(1) Capicciotti, C. J.; Kurach, J. D. R.; Turner, T. R.; Mancini, R. S.; Acker, J. P.; Ben, R. N. *Sci. Rep.* **2015**, *5*, 9692.

(2) Capicciotti, C.; Leclère, M.; Perras, F. *Chem. Sci.* **2012**, *3* (5), 1408.

(3) Capicciotti, C.; Doshi, M.; Ben, R. *Ice Recrystallization Inhibitors: From Biological Antifreezes to Small Molecules*; Wilson, P., Ed.; InTech: New York, NY, 2013.

(4) Balcerzak, A. K.; Febbraro, M.; Ben, R. N. *RSC Adv.* **2013**, *3* (10), 3232.

(5) Tam, R. Y.; Ferreira, S. S.; Czechura, P.; Ben, R. N.; Chaytor, J. L. *J. Am. Chem. Soc.* **2008**, *130* (51), 17494.

(6) Chao, H.; DeLuca, C.; Davies, P. *Protein Sci.* **1994**, *3* (10), 1760.

(7) Chakrabartty, A. *J. Biol. Chem.* **1989**, *264* (19), 11307.

(8) Wen, D.; Laursen, R. *J. Biol. Chem.* **1993**, *268* (22), 16396.

(9) Wen, D.; Laursen, R. *J. Biol. Chem.* **1992**, *267* (20), 14102.

(10) Ewart, K. V; Lin, Q.; Hew, C. L. *Cell. Mol. Life Sci.* **1999**, *55* (2), 271.

(11) Capicciotti, C. J.; Leclère, M.; Perras, F. A.; Bryce, D. L.; Paulin, H.; Harden, J.; Liu, Y.; Ben, R. N. Potent inhibition of ice recrystallization by low molecular weight carbohydrate-based surfactants and hydrogelators. *Chemical Science*, 2012, *3*, 1408.

(12) Cramer, R. D.; Patterson, D. E.; Bunce, J. D. *J. Am. Chem. Soc.* **1988**, *110* (18), 5959.

(13) Kim, K. H.; Greco, G.; Novellino, E.; Federico, N. *Perspect. drug Discov. Des.* **1998**, *12-14*, 257.

(14) Fontaine, F.; Pastor, M.; Sanz, F. *J. Med. Chem.* **2004**, *47* (11), 2805.

(15) Wavefuntion Inc. Spartan ’14, 2015.

(16) Halgren, T. A. *J. Comput. Chem.* **1996**, *17* (5-6), 490.

(17) Stewart, J. J. P. *J. Mol. Model.* **2007**, *13* (12), 1173.

(18) Frisch, M. J. et al. Gaussian 09, Revision A.02. *Gaussian 09, Revision A.02*, 2009.

(19) Walker, P. D.; Arteca, G. A.; Mezey, P. G. *J. Comput. Chem.* **1991**, *12* (2), 220.

(20) Bondi, A. *J. Phys. Chem.* **1964**, *68* (3), 441.

(21) Geladi, P.; Kowalski, B. R. *Anal. Chim. Acta* **1986**, *185*, 1.

(22) Holland, J. H. *Adaptation in Natural and Artificial Systems*; The University of Michigan Press, 1975.

(23) Forrest, S. *Science (80-.).* **1993**, *261* (5123), 872.

(24) Clark, D. E.; Westhead, D. R. *J. Comput. Aided. Mol. Des.* **1996**, *10* (4), 337.

(25) Rogers, D.; Hopfinger, a J. *J. Chem. Inf. Model.* **1994**, *34* (4), 854.

(26) Lorber, A.; Wangent, L. E.; Kowalski, B. R. *J. Chemom.* **1987**, *1*, 19.

4. Simulation of Novel Ultra-Microporous MOFs

This chapter outlines the work done in collaboration with the experimentalist group of Dr. Ramanathan Vaidhyanathan at the Indian Institute of Science Education and Research in Pune, India. Two MOFs will be discussed here, Ni-4PyC and Mg-4PyC – both are ultra-microporous MOFs with unique properties and different topologies. Ni-4PyC exhibits optimal features for pre-combustion carbon capture and hydrogen purification and will be presented in section 4.1. This work was recently submitted to *Science Advances* and is tentatively accepted pending revisions. Mg-4PyC is an interesting case of dynamic linker rotation facilitating a gate-opening conformational change with respect to CO_2 loading and will be presented in section 4.2. This work is currently in preparation for submission. For both projects I was responsible for all computational studies which include adsorption analysis, molecular dynamics simulations, and DFT optimizations. Most of these calculations were performed using FA^{3}PS, the adsorption analysis program developed by former post-doc Dr. Tom Daff. All experimental work such as synthesis, characterization, and stability studies were performed by Shyamapada Nandi of the Vaidhyanathan Lab.

4.1. A Single Ligand Ultra-Microporous MOF for Pre-Combustion CO_2 Capture and Hydrogen Purification

As previously mentioned in Section 1.2.4, carbon capture and sequestration technologies are currently limited by the adsorbent materials used during the capture process. There are different ways in which CO_2 can be captured in the context of energy generation. Metal-organic

frameworks represent an attractive material which can be tailor made for specific adsorption and desorption conditions. Ultra-microporous MOFs have been demonstrated to provide excellent gas separation characteristics but poor saturation limits due to their small pores as explained in Section 1.2.1. Thus, the design and synthesis of an ultra-microporous MOF which also exhibits high uptake capacities at high pressures poses an enigmatic challenge.

Ni-4PyC is an ultra-microporous MOF with uncommonly high CO_2 uptake capacity (8.2 mmol/g at 10 bar and 313K) despite having small pores and extraordinary CO_2/H_2 selectivity at high pressure. This material was found to be stable to humidity, exhibit favorable CO_2 diffusion, and optimal heats of adsorption for pressure swing adsorption conditions. Additionally, this is made with only a single small readily available ligand, 4-pyridylcarboxylate, and is easily synthesizable in a one-pot fashion. Simulation was used to probe the unusually high adsorption of CO_2 through binding site determination.

4.1.1. Experimental Synthesis, Characterization, and Adsorption Properties

Ni-4PyC was prepared at both a milligram scale and 10 gram scale by mixing nickel carbonate and pyridine-4-carboxylic acid as the precursors in a solution of tetrahydrofuran, methanol, and water and heated to 150^{o}C for 72hrs. The resulting crystal structure was determined from a single crystal using x-ray crystallography. The resulting structure is a three dimensional MOF built from corner-sharing nickel dimers and an octahedral nickel center. The nickel dimers (referenced as Ni(1) and Ni(2)) and octahedral metal center (referenced as Ni(3)) are coordinated by pyridyl carboxylate (PyC) linkers as well as terminal and bridging water

groups. The topology can be likened to a six-connected cubic network with two types of channels as well as a cage system. The structure of Ni-4PyC is shown below in Figure 4.1.

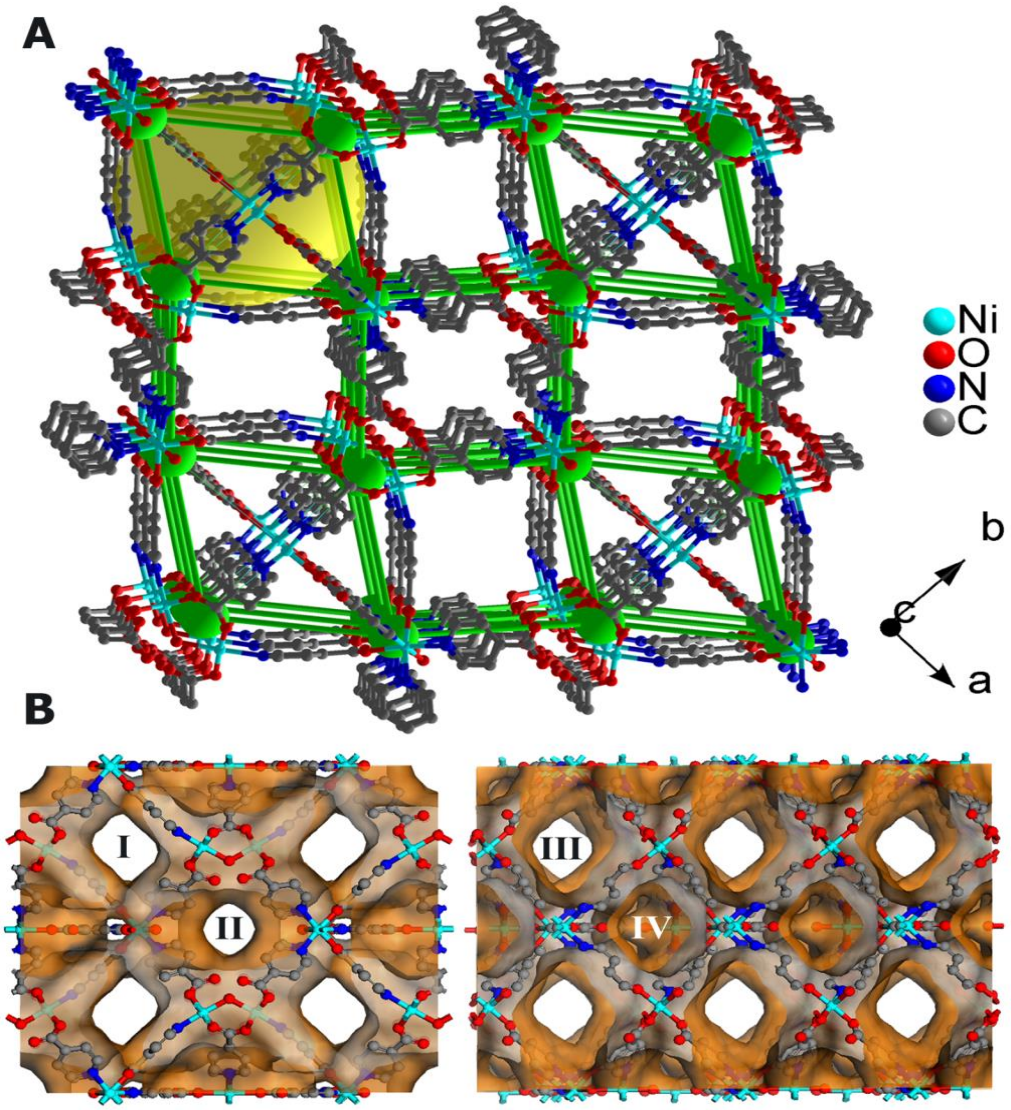


Figure 4.1: (A) Three dimensional structure of Ni-4PyC as visualized by the crystallography software OLEX.[1] The distorted cubic framework is shown in green whereby the Ni dimers are reduced to a spherical node and are connected by the PyC linkers shown as green rods. The yellow ball traces the cages of the structure which are lined by the octahedral Ni located in the middle of the pores. (B) The Connolly surface area representations of Ni-4PyC using a robe radius of 1.4 Å which corresponds to the radius of water molecule and is the default for most crystallographic software. The channels labelled I and III are interconnected and run along the a- and c-axis respectively, II is a one-dimensional channel along the c-axis, and IV represents the cage and is lined by terminal water molecules coordinated to the nickel centre.

The framework shows an interesting system of interconnected channels and cages. Two dimensional channels I and III (Figure 4.1) are interconnected along the a- and c-axis and

possess dimensions of 7.8 x 7.8Å and 7.5 x 7.5Å from atomic centre of mass. Channel II (Figure 4.1) is one dimensional running along the c-axis with dimensions of 6.7 x 6.7Å. Cage IV (Figure 4.1) are made up of nickel dimers arranged into a square formation capped by nickel octahedral on either side of the near spherical cage with dimensions of 12 x 12 x 12 Å.

In order to experimentally determine the surface area adsorption of N_2 at 77K was done and calculated with the Brunauer-Emmet-Teller (BET) theory. BET determined surface areas are the standard for experimentally measuring surface areas for MOFs.[2] It was found that Ni-4PyC had a modest experimental BET surface area of 945 m^2/g. A positron annihilation lifetime spectroscopy (PALS)[3] experiment as well as adsorption based NLDFT[4] calculations were performed to determine the pore sizes. It was found that a Ni-4PyC possess a biomodal pore distribution with two pores of 3.5 – 3.9Å and 4.4 – 4.8 Å (considering vdW volume) which is consistent with the values from crystal structure. As a comparison to larger pored MOFs, NU-100[5] which has a staggering BET surface area of 6,143 m^2/g and pore sizes of 13.4 – 48Å.

For a MOF with such low surface area and small pore sizes, one would expect the uptake to be quite low. However, for an ultra-microporous material, Ni-4PyC exhibits an extraordinarily high CO_2 uptake capacity of 10.8 mmol/g at 195 K, 5.5 mmol/g at 273K, and 3.6 mmol/g at 303K and 1 bar. (Figure 4.2B) Perhaps more importantly, this MOF exhibits virtually no H_2 or N_2 adsorption at 273K or 303K (Figure 4.2A) which shows that this MOF is capable of selectively capturing CO_2 over H_2 or N_2. The heats of adsorption (HOA) were determined experimentally from adsorption isotherms at -10^oC, 0^oC, +10^oC and +30^oC via virial fits and a DFT model to be 35 kJ/mol at zero-loading to 28 kJ/mol at higher loadings. (Figure 4.2C) This shows Ni-4PyC is in the optimal HOA range of 25-30kJ/mol[6] for pressure swing adsorption which enables easy

removable of CO_2. Furthermore there is almost perfect agreement with simulated and experimental uptake at high pressures at 298K. (Figure 4.2 D) The details of how the simulated uptake was calculated will be discussed in the next section.

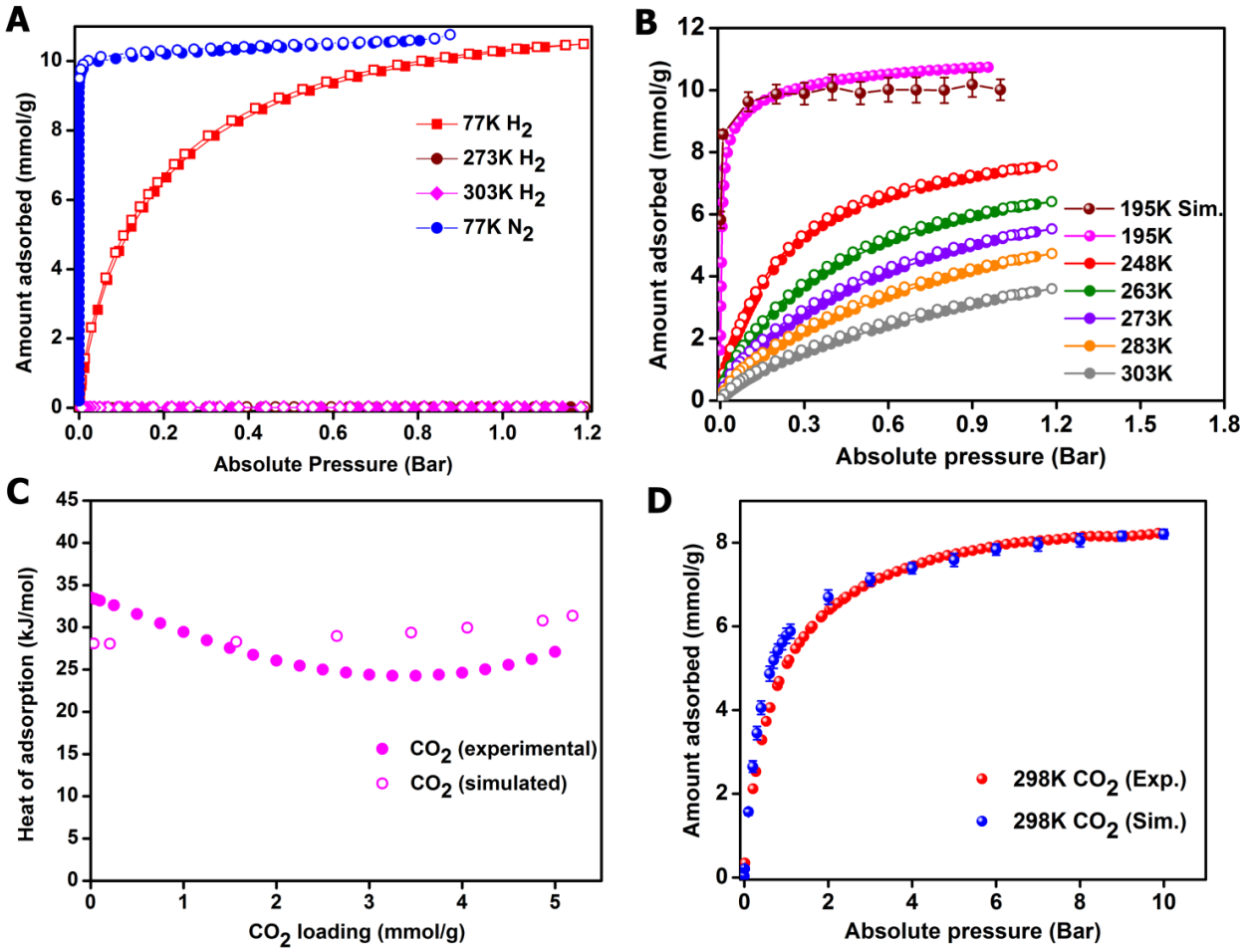


Figure 4.2: (A) Experimental H_2 and N_2 adsorption isotherms for Ni-4PyC at 77K, 273K, and 303K. The N_2 adsorption isotherm at 77K displays the saturation of Ni-4PyC at ~11 mmol/g. Filled circles represent adsorption, open circles represent desorption. (B) The CO_2 adsorption isotherms at various temperatures. Note the simulated uptake for 195K is shown with relatively good agreement. (C) The HOA plot as a function of CO_2 loading determined experimentally from -25°C to 30°C (filled circles) and from GCMC simulations at 25°C (open circles). (D) Experimental and simulated high pressure CO_2 adsorption isotherms for CO_2 at 298K. Initially, the simulated uptake preceded the experiment and was found to be a perfect match.

Beyond excellent adsorption abilities, an ideal MOF for CO_2 capture must be highly stable and exhibit fast diffusion of CO_2 through its pores. To this end, stability and recyclability experiments were performed. (Figure 4.3A) A sample of Ni-4PyC was exposed to steam for 160

hours and PXRD patterns remained essentially unchanged. Furthermore Ni-4PyC was also exposed to a stream of humidified CO_2 at 30% relative humidity for over 48 hours but exhibited no loss in adsorption capabilities as the isotherm was the same prior and after to treatment. (Figure 4.3B) In terms of stability with respect to pressure, it was found that the original porous structure was retained after being subjected to 70 bar pressure for 24 hours. Finally, the shelf-life was tested and Ni-4PyC was found to retain >90% of its porosity even after exposure to ambient air for over 6 months. In terms of recycling, Ni-4PyC exhibited smooth adsorption-desorption characteristics as observed from thermogravimetric analysis (TGA) cycling experiments. Essentially CO_2 is consistently adsorbed and desorbed using the same sample up to ~6.5% weight at a constant temperature of 35°C. (Figure 4.3C)

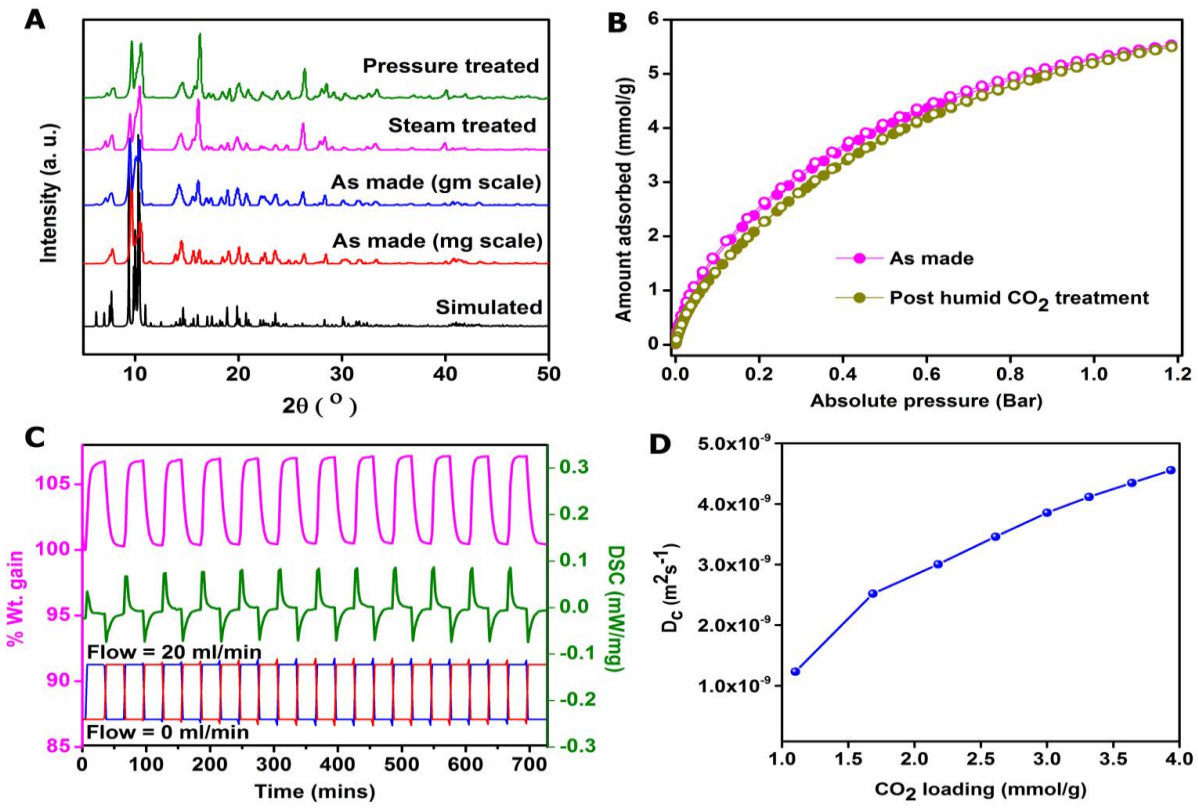


Figure 4.3: (A) Powdered x-ray diffraction plots showing the stability of Ni-4PyC to pressure, steam, and the homogeneity between the mg and gm scales. (B) CO_2 adsorption isotherms of Ni-4PyC at 273K displaying no loss of adsorption characteristics

even after 48hours exposure to 30% relative humidity CO_2. Filled circles are adsorption and open circles are desorption. (C) TGA cycling data at 308K. (D) Diffusion coefficient as a function of CO_2 loading taken from adsorption isotherms at 273K.

Diffusion is an important property for PSA conditions where smaller pores of an ultra-microporous MOF might pose a challenge by severely restricting the speed of adsorption and desorption. The experimental kinetics and diffusion of CO_2 within Ni-4PyC was determined via a high resolution rate of adsorption experiment. This was carried out at 273K in the pressure range of 0-1bar of pure component CO_2 and the average diffusion coefficient was found to be $3.03 \times 10^{-9} m^2/s$. (Figure 4.3D)

At this point experimentalists have shown a MOF with seemingly excellent characteristics for CO_2/H_2 separations which would be important in hydrogen purification and pre-combustion carbon dioxide capture. However, there remained many questions which experiment could not answer alone. Why does this MOF have such high uptake despite such small pores? Where are the CO_2 molecules binding and what are the interactions associated here? Are all the pores fully accessible? In order to answer these questions computational simulations were performed.

4.1.2. Computational Simulations

4.1.2.1. Geometry Optimization

Before any simulations could be performed it was imperative that a proper optimized structure of Ni-4PyC was used. The initial experimental crystal structure of Ni-4PyC had disorder with respect to the orientation and direction of the pyridine carboxylate SBU. This disorder was a result of discrepancies of the SBU binding to the metal center – on one hand the metal centre could bind via the nitrogen of the pyridine ring and on the other hand the metal could bind to

the oxygen of the carboxylic acid group. Thus, various combinations of the organic linker orientation were made in Materials Studio[7] and then optimized using periodic density functional theory (DFT).[8] Most of the possible conformations had steric overlap between the pyridine hydrogens which would be unphysical. (Figure 4.4A) Only two conformations did not have this steric overlap. (Figure 4.4B,C) The CO_2 uptake was calculated for these structures from GCMC simulations performed using FA^{3}PS and it was found that the greatest difference in uptake in the isotherm from 0-1 bar was only 0.1 mmol/g. Therefore between these two structures the one with least symmetry was used. Geometry optimizations were performed starting from high quality experimental X-ray structures with all atoms and unit cell parameters optimized. Periodic DFT calculations were performed with the VASP code[9–11] using the PBE exchange-correlation functional[12]. PAW pseudopotentials[13,14] were used in a plane wave basis set with a kinetic energy cut-off of 520 eV. All calculations were spin polarized and only the Γ-point was sampled. Empirical dispersion corrections of Grimme[15] were included in both energy and force calculations with the default scaling factor of 0.75, as parameterized by Grimme, for the PBE functional. The structures of the different configurations are shown below in Figure 4.4.

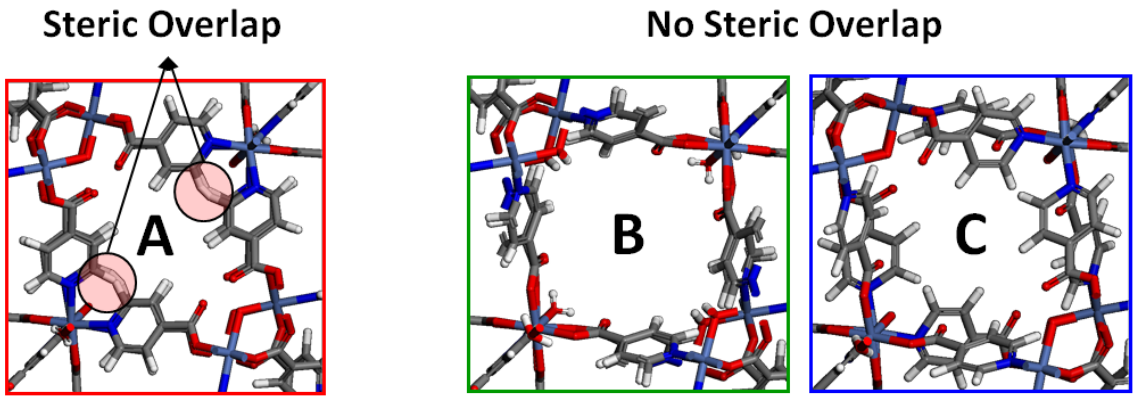


Figure 4.4: (A) Linker orientation with steric overlap of hydrogens in the pyridine ring. (B) Linker orientation with no steric overlap but higher symmetry. (C) Linker orientation with lower symmetry, this structure was used for subsequent simulations.

4.1.2.2. Simulated Adsorption Isotherms

Once a reasonable structure was found, GCMC calculations were performed in order to simulate the isotherms. Two approximations used were that the framework was held fixed while the gas guest molecules were assumed to be rigid. The electrostatic energetic contributions were determined by partial atomic charges assigned to each atom calculated with the REPEAT method[16] using the DFT derived electrostatic potential. Dispersive and steric repulsive interactions were included by a 12-6 Lennard-Jones (L-J) potential for each atom. The ε and σ parameters for the framework were taken from the Universal Force Field (UFF).[17] H_2 Lennard-Jones parameters, the H-H bond length (0.742 Å), and point charges for the five-site H_2 model were taken from work by Belof *et. al*[18] which is a hydrogen potential for condensed phase simulation. These parameters have been used in the simulation of H_2 in to study the adsorption of H_2 in ultra-microporous MOFs[19] and are shown below in Table 4.1.

Table 4.1. Forcefield parameters for the 5-site H_2 model taken from by Belof *et. al*[18]. H2E corresponds to the true atomic positions of hydrogen atoms, H2G coincides with the center-of-mass site, and H2N contains the additional Lennard-Jones sites.

Atom	R / Å	Q / e	ε / kcal mol^{-1}	σ / Å
H2E	-0.371	0.3732	0.0000000	0.0000
H2N	-0.329	0.0000	0.0080798	2.3406
H2G	0.000	-0.7464	0.0175899	3.2293
H2N	0.329	0.0000	0.0080798	2.3406
H2E	0.371	0.3732	0.0000000	0.0000

The ε and σ parameters of CO_2 were taken from García-Sánchez *et al.*[20] which were developed to fit experimental adsorption isotherm data in zeolite frameworks. The C-O bond length (1.149 Å) and partial charges on CO_2 atoms (C = +0.6512*e*, O = -0.3256*e*) were taken

from the potential by Harris and Yung[21]. Lennard-Jones parameters of all atom types are given in Table 4.1.

Table 4.2. Lennard-Jones parameters for framework atoms from the UFF forcefield, CO_2 guest molecules.

Forcefield	Atom	ε / kcal mol^{-1}	σ / Å
UFF	C	0.1050	3.4309
UFF	O	0.0600	3.1181
UFF	N	0.0690	3.2607
UFF	Ni	0.0150	2.5248
García-Sánchez *et al.*	O (CO_2)	0.1702	3.0170
García-Sánchez *et al.*	C (CO_2)	0.0595	2.7450

GCMC simulations were performed with an in-house code, FastMC via the FA^{3}PS program. The number of production steps used was 10^7 after an initial equilibration stage of 10^6 steps for each gas pressure point on the isotherm. The Monte Carlo algorithm utilized equal probabilities for the moves of guest displacement, insertion, and deletion. A cut-off of 12.5 Å was used for long range interactions which were calculated using a Ewald summation. For pressures less than 1 bar, the ideal gas pressure was used in the Monte Carlo guest insertion and deletion criteria. Conversely, pressures greater than 1 bar was corrected for fugacity by evaluating the uptake based on pressures fitted to the Peng-Robinson Equation of State.[22] A 2x2x3 super-cell was used for the GCMC simulations. A similar approach has been used by the Woo Lab successfully to analyze CO_2 binding within an amine-functionalized MOF.[23]

The adsorption isotherms were calculated for Ni-4PyC and matched experiment quite well as shown in Figure 4.2B,D. However, the initial set of adsorption data for CO_2 was actually far below that of simulation. (Figure 4.5) Simulation gave an uptake of 6.50 mmol/g at 0.9 bar

and 273K and the experimental data was 3.14 mmol/g at the same pressure and temperature. In order to solve for this discrepancy many tests were performed, different forcefields for CO_2 as well as the framework atoms of the MOF were tried, charges were set to zero, and vdW interactions were also set to zero, but despite this the simulated uptake was still more than double that of experiment. Of course, in order to have confidence in any simulations that are done they must match experimental data as much as possible. It was at this point that the experimentalists were informed of our simulations and asked if it was at all possible to better activate the MOF. Before a MOF can act as an adsorbent it must be activated, this means that the solvent within the pores of the MOF have to be removed. This is typically done with vacuum and high temperature. Sometimes if not enough time is given to the activation process some residual solvent may remain in the pores which blocks the channels and limits the adsorption capabilities of the MOF. At this point adsorption experiments were redone after a more thorough activation of Ni-4PyC. It was found that the uptake increased from 3.14 mmol/g (0.9 bar, 273 K) to 5.6 mmol/g (0.9 bar, 273 K) which is much more in line with simulation.

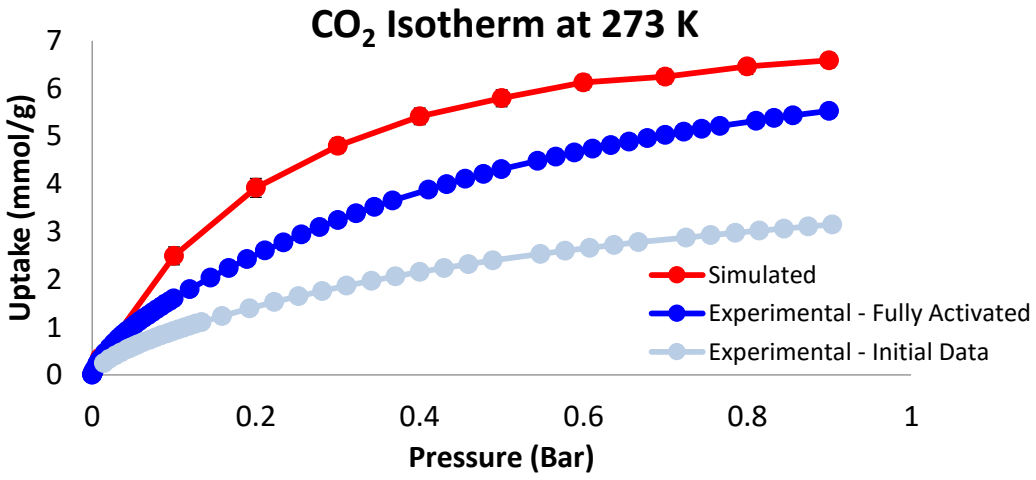


Figure 4.5: Adsorption isotherm of Ni-4PyC. (Red) simulated isotherm from GCMC simulations. (Light blue) Initial experimental data which showed far lower adsorption capacity than simulation. (Blue) Experimental adsorption isotherm after sample was fully activated and shows better agreement with the simulated isotherm.

One can expect simulation to be higher especially at the lower pressure regime of (0.1 – 1 bar) because there are always experimental impurities, crystal defects, and trace amounts of solvent which will cause a decrease in uptake. However, simulation assumes an ideal case with a perfect crystal and solvent-free pores. This represented a case where simulation demanded the review of experimental data and in turn led to an enhancement of adsorption features. Another way to validate simulation is to compare the isoteric heat of adsorption. This can be calculated from the output of a GCMC simulation via the Green-Kubo fluctuation theory expression, (**Eq**. **4.1**)

$$q_{isosteric} = \frac{\langle UN\rangle - \langle U\rangle\langle N\rangle}{\langle N^2\rangle - \langle N\rangle\langle N\rangle} + RT \qquad \textbf{4.1}$$

where *N* is the number of CO_2 gas molecules in the GCMC simulation, *U* is the total configuration energy for the CO_2 molecules, *R* is the gas constant, and *T* is the temperature .[24] The HOA is shown to match well with experiment as shown in Figure 4.2C.

Now that our simulation methods were established showing good match in terms of adsorption isotherms and heat of adsorptions, it was proposed that the high pressure regime be modeled. Initially, Ni-4PyC showed exceptionally high CO_2 saturation capacity of 10.8 mmol/g at low temperatures of 195 K. Our simulations showed an isotherm which was in excellent agreement for this temperature and pressure range of 0-1 bar. (Figure 4.2B) This high saturation capacity suggested that Ni-4PyC may also have high CO_2 capacity at higher pressures and higher temperatures which are conditions relevant for pre-combustion CO_2 capture. Thus, GCMC simulations were first performed to calculate the adsorption isotherm at high pressure regime from 0 – 10 bar and 298K. It was found that a predicted Ni-4PyC uptake capacity of 8.2

mmol/g at 10 bar which is quite high. This finding resulted in the need to confirm with experimental results. As an additional test the probability distributions of CO_2 molecules were compared at low temperature and low pressure (195 K and 1 bar) and standard temperature and high pressure (298K and 40 bar). It was found that there were no significant differences in the GCMC probability distributions between the two conditions. (Figure 4.6)

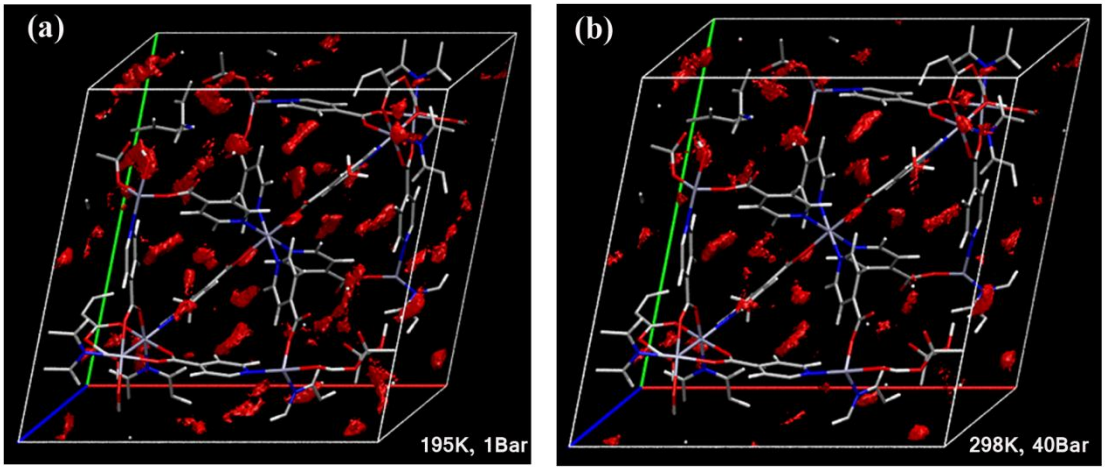


Figure 4.6. Probability densities of CO_2 center of mass as determined via GCMC calculations at a) 195 K and 1 bar and b) 298 K and 40 bar. The isosurface representation has an isovalue of 0.04 molecules/Å^3. The densities are very similar between low temperature, low pressure, and high temperature, high pressure.

Typically high pressure isotherms are not performed when first characterizing a MOF due to cost. However, our simulations suggested that this was necessary. Indeed, the experimental adsorption isotherm was found to match almost exactly to Ni-4PyC. (Figure 4.2D) Once again, simulation preceded experiment to aid in the discovery of suitable applications for Ni-4PyC.

4.1.2.3. Binding Site Location

The high-uptake of Ni-4PyC was so unique that it warranted further molecular level investigation utilizing molecular simulation for the determination of the adsorption sites of CO_2

in order to understand how Ni-4PyC can accommodate so many CO_2 molecules. A similar analysis of binding sites of ZnAtzOx MOF has been done before by the Woo Lab which showed excellent agreement between the simulated CO_2 positions and those experimentally determined via x-ray crystallography.[25] The binding sites were determined using ABSL as part of the FA^{3}PS program as outlined in Section 2.4.1.4. The binding site energies were calculated with equation 2.7. The ranking of binding sites were based on interaction energy and the occupancy of the binding sites with respect to the probability distributions. For determination of the binding energies, single point calculations were performed with interaction energies subdivided into dispersion and electrostatic contributions.

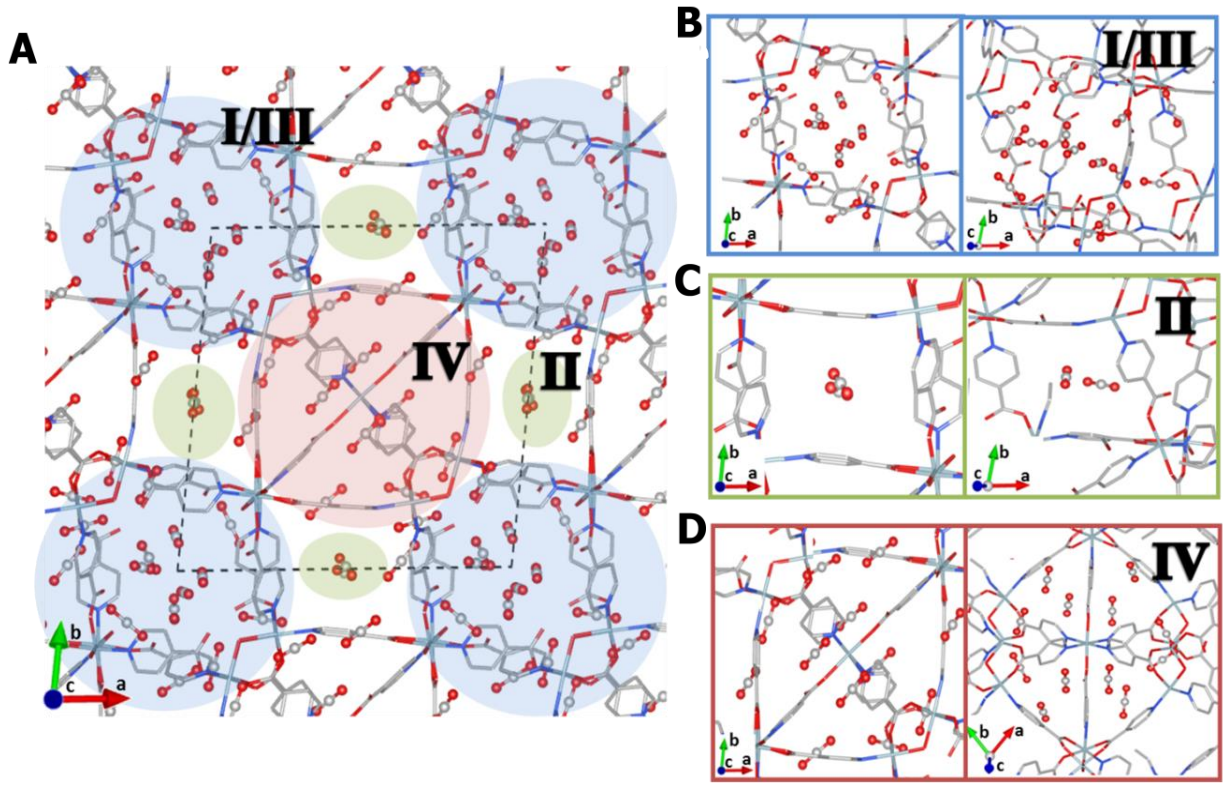


Figure 4.7: CO_2 binding sites of Ni-4PyC as represented via the VESTA[26] visualization software. (A) A view looking down the c-axis of Ni-4PyC showing the top 30 distinct CO_2 binding sites as calculated using ABSL at 195 K and 1 bar. There were found to be three unique binding region noted in blue (I/III), green (II), and red (II). Blue corresponds to the 2D channels, green to the 1D channels, and red corresponds to the spherical cage. Zoomed in images of each binding region are shown on the right (B),(C), (D).

The experimental low temperature (195 K) saturation limit of 10.8 mmol/g corresponds to approximately 28 CO_2 molecules per unit cell. The strongest 30 binding sites were found using ABSL and all had binding energies ranging from -24.0 to -32.8 kJ/mol. (Figure 4.7) It was found that three main binding site regions existed which encompassed the three main pore systems. The first region labelled **I/III** in Figure 4.7 is the intersection of the 2D channels as shown in Figure 4.1. The second region labelled **II** is the 1D channel, and the third region labelled **IV** is the spherical cage. The strongest binding sites were located in the spherical cage of binding region **IV.** Within this region the CO_2 molecules are wedged between the aromatic planes of two PyC ligands, with the greatest binding energies ranging from -30.0 to 32.8 kJ/mol. Conversely, the weakest binding sites were found within the 1D channel of binding region **II**, which have binding energies of -24.0 to -27.2 kJ/mol. Interestingly, the binding in this region was found to have virtually no electrostatic component – in other words the binding was due almost entirely to dispersion interactions. This contrasts the binding sites in the other two regions (**I/III** and **IV**) whose binding energies were composed of up to 22% electrostatic interactions.

Such a high density of binding sites signals the possibility of cooperative binding effects where the bound guests molecules interact favorably with other bound guests. Thus, the cooperative binding energy was calculated from single point energy calculations. The cooperative energy was calculated as, (**Eq. 4.2)**

$$E_{cooperative} = E(\mathrm{MOF} + n\mathrm{CO_2}) - E(\mathrm{MOF} + (n-1)\mathrm{CO_2}) \qquad \textbf{4.2}$$

where the configurational energy of the MOF with *n* number of CO_2 molecules in the unit cell is $E(\mathrm{MOF} + n\mathrm{CO_2})$. The term $E(\mathrm{MOF} + (n-1)\mathrm{CO_2})$ is the summation of the binding energies of *n-1* number of CO_2 molecules in the unit cell as such, (**Eq. 4.3**)

$$\sum_{i=1}^{n-1} E(\mathrm{MOF} + n\mathrm{CO_2}) - E(\mathrm{MOF}) - nE(\mathrm{CO_2}) \qquad \textbf{4.3}$$

By occupying the 28 most stable binding sites with CO_2 which corresponds to the experimental saturation uptake, we find there is a significant cooperative binding energy of 5.2 kJ/mol per guest molecule. In other words, with the 28 binding sites occupied there is a net stabilization of 146 kJ/mol due to favorable CO_2-CO_2 interactions. Interestingly, after 29 guest molecules (just beyond the low temperature saturation limit) the cooperative binding energy begins to diminish as additional CO_2 molecules interact unfavorably with existing guest molecules.[27] These results suggest that cooperative binding plays a significant role in the high CO_2 uptake capacities observed in Ni-4PyC. A summary of the cooperative binding energies of the 9 lowest ranked binding site CO_2 molecules are shown below in table 4.3.

Table 4.3. Cooperative CO_2-CO_2 energies with respect to the number of molecules loaded.

n CO_2 per unit cell	Total Cooperative Energy (kJ/mol)	Cooperative Energy Per CO_2(kJ/mol)
22	-108.7	-4.9
23	-118.5	-5.1
24	-120.0	-5.0
25	-121.2	-4.8
26	-123.8	-4.8
27	-136.3	-5.0
28	-146.6	-5.2
29	-152.9	-5.3
30	-145.3	-4.8

4.1.2.4. CO_2/H_2 Separation

When assessing a material for separations of gas mixtures, it is often important that the adsorption of a mixture of gas be explored and not just their pure components. However, experimentally determining the adsorption of a mixture of gas is non-trivial and is currently very difficult to do. The problem lies in the fact that once a mixture is adsorbed, the ratios of the gases within the mixture that has been adsorbed is difficult to experimentally determine. Thus, computational simulation is the best alternative to analyzing a binary mixture of gases being adsorbed by the MOF. The working capacity of a PSA swing from 10 to 1 bar (as described in section 1.2.4) and selectivity (**Eq. 1.3**) of CO_2 over H_2 was calculated from a binary GCMC simulation where both guest molecules were present within the GCMC simulation at the same time. This was done by specifying the partial pressures of each gas molecule with a ratio of 40:60 and 20:80 (CO_2: H_2) which is an industrially relevant mixture comparable to that found in pre-combustion intake gas[28] and then evaluating the uptake with this mixture using the GCMC methods. The simulated PSA working capacities of Ni-4PyC (using a desorption pressure of 1 bar) was compared to the working capacities of the recently reported industrial benchmarks zeolite 13X and activated carbon JX101, and two of the top performing MOFs identified for this application, MgMOF-74 and CuBTTri.[29] (Figure 4.8A,B). Similarly, a comparison of the CO_2/H_2 selectivities to these materials were also calculated at the two H_2/CO_2 ratios. (Figure 4.8C,B)

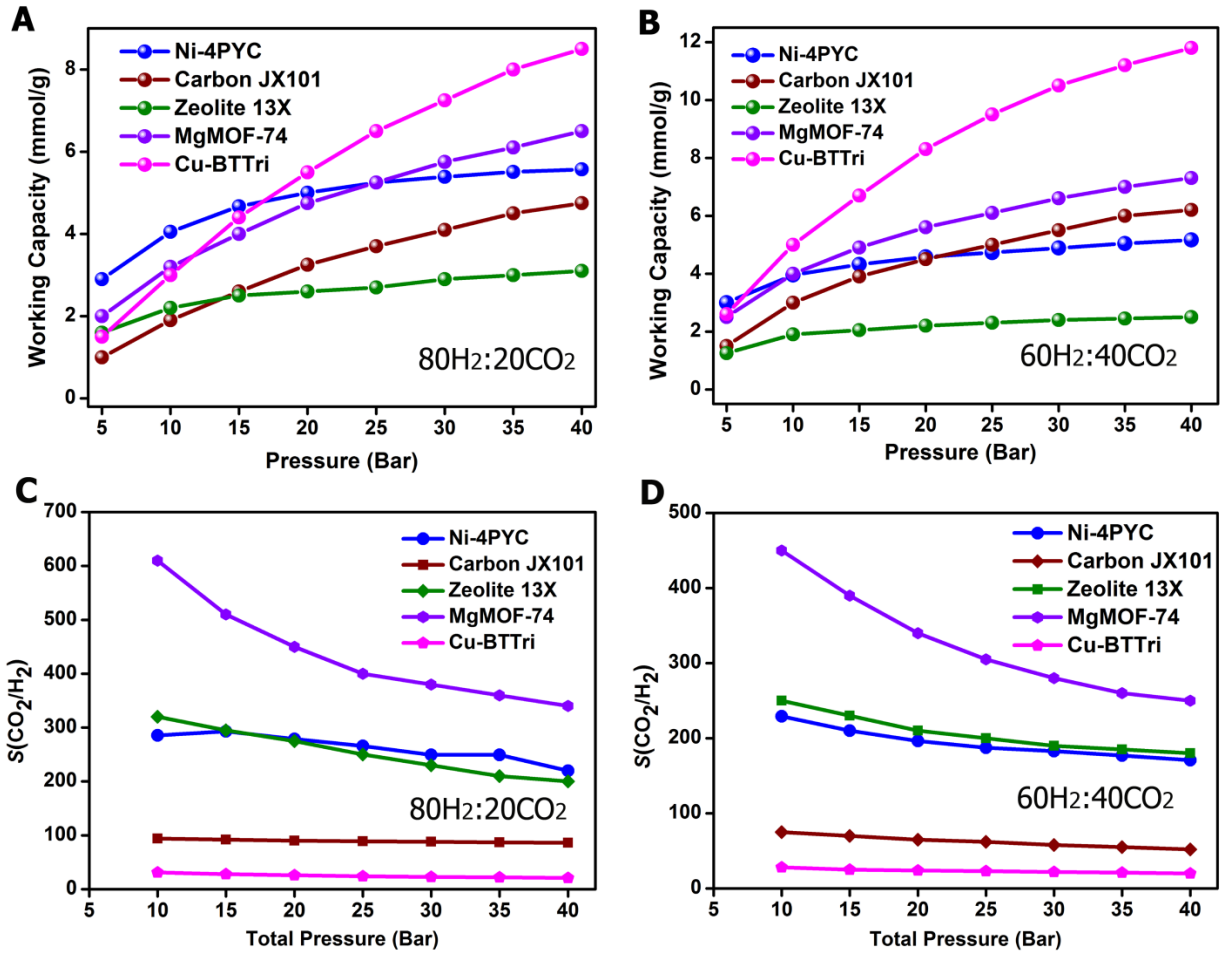


Figure 4.8: Working capacities and selectivities of Ni-4PyC. The simulated working capacities of Ni-4PyC compared to that of several other materials used for this application and high performing MOFs at $80H_2$:$20CO_2$ (A) and $60H_2$:$40CO_2$ (B) gas mixtures at 313 K and desorption pressure of 1 bar. The selectivity of Ni-4PyC vs. the same materials for $80H_2$:$20CO_2$ (A) and $60H_2$:$40CO_2$ (B) gas mixtures are also shown. Data for activated carbon JX101, zeolite 13X, Mg-MOF-74 and Cu-BTTri are taken from reference 29.

At low CO_2 concentrations (80% H_2, 20% CO_2), Ni-4PyC has the largest working capacity up to an adsorption pressure of 15 bar, but remains amongst the top performers in this respect throughout the pressure range. Only the MOF CuBTTri has a significantly higher working capacity at pressures greater than 25 bar. However, CuBTTri has a very poor H_2/CO_2 selectivity, the lowest of all the materials compared, making it unsuitable for practical use.[29] It was found that at higher CO_2 concentrations (60% H_2, 40% CO_2), the working capacity of Ni-4PyC is less competitive. Nonetheless, when compared to zeolite 13X, which is used industrially for PSA

based CO_2 scrubbing of natural gas, Ni-4PyC has an almost identical selectivity but roughly double the working capacity throughout the pressure range. Compared to the high performance activated carbon JX101, Ni-4PyC has a higher working capacity for the 80:20 gas mixture, and a comparable working capacity for the 60:40 gas mixture throughout the whole pressure range. However, Ni-4PyC has a CO_2/H_2 selectivity that is at least 2.5 times better than JX101 for both gas compositions. The only material which performed better was MgMOF-74 which has one of the highest working capacities at all pressures and both gas compositions. Moreover, in all cases it also has the highest CO_2/H_2 selectivity, outperforming Ni-4PYC by at least 50% in this respect. Despite the favorable adsorption properties, MgMOF-74 is not hydrolytically stable due to the presence of open metal sites, which limits its practical use.

4.1.2.5. Accessible Surface Area

The accessible surface area is an important feature that warrants study via simulation. The void volume (0.0434 cm^3/g) and surface area (1193.16 m^2/g) were calculated using the Zeo++ code[30,31] with a probe radii (1.72 Å) corresponding to CO_2 gas molecules. This simulated surface area is in good agreement with the experimentally determined BET surface area of 945m^2/g. The accessible volume as measured by a solvent probe radius of 1.72 Å shows the presence the 1D channels which running through the unit cell. The center pore was found to be accessible via the channels at 1.00 Å which was also found to contain the strongest binding sites as mentioned in the main text. Depending on the orientation of the CO_2 molecule, it should be able to access this cage via the channels. The accessibility of this spherical cage will be discussed in the next section. Interestingly, the accessible surface area drops from 1194 m^2/g in

the empty MOF, to <1 m^2/g in the MOF with the 28 lowest energy CO_2 binding sites occupied demonstrating almost full CO_2 saturation within Ni-4PyC.

4.1.2.6. Molecular Dynamics Simulations

So far most of the calculations involved either geometry optimizations using periodic DFT or GCMC simulations. However, molecular dynamics played a key role in answering a few questions. First MD simulations were utilized to study the kinetics of Ni-4PyC, more specifically calculating its diffusion coefficient. Secondly, experimentalists had initially believed that the spherical cage pore was not accessible. However, MD simulations were used to prove that the spherical cage pore in Ni-4PyC was indeed accessible by CO_2 molecules.

Molecular dynamics simulations were performed with DL_POLY[32] in order to calculate the diffusion coefficients as described in Section 2.4.2.1. The MD simulation was done at 298.15 K and 1 bar with 0.2 ns of equilibration, 1 ns for the production run, and a time step of 0.001 ps with an NVT ensemble. This was done at flue gas conditions with a binary mixture of $15CO_2$:$85N_2$ in a 2x2x3 supercell with 69 CO_2 molecules which corresponds to a 2.10 mmol/g loading. Simulation gave a diffusion coefficient of $3.73x10^{-9}m^2/s$ at 298K under the flue gas compositions. This is excellent agreement with the experimentally determined average CO_2 diffusion coefficient of $3.03x10^{-9}m^2/s$. This diffusivity is comparable to those observed in some of the microporous MOFs, ZIF-8: $8x10^{-10}$; MIL-53(Cr): ~$5x10^{-8}$; MOF-5: 1.17×10^{-9} and MOF-177: $2.3\times10^{-9}m^2s^{-1}$ at 298K.[33–37] The diffusion coefficient of Ni-4PyC is also two orders of magnitude higher than that of zeolite-13X which is currently used in PSA scrubbers for natural gas purification.[37]

In order to validate whether the CO_2 molecules are accessible to the middle cage, molecular dynamics simulations were performed with DL_POLY.[32] Two configurations were tested, one where the channels were saturated and the cage was empty and vice versa. This was done at 313 K and 10 bar to model high pressure adsorption with 0.2 ns of equilibration, 1 ns for the production run, and a time step of 0.001 ps with an NVT ensemble. In both simulations the CO_2 molecules diffuse into and out of the cages. It was found that cage to channel diffusion occurred throughout the simulation time length while channel to cage diffusion occurred almost instantaneously. Snapshots of the simulation are showed below in Figure 4.9.

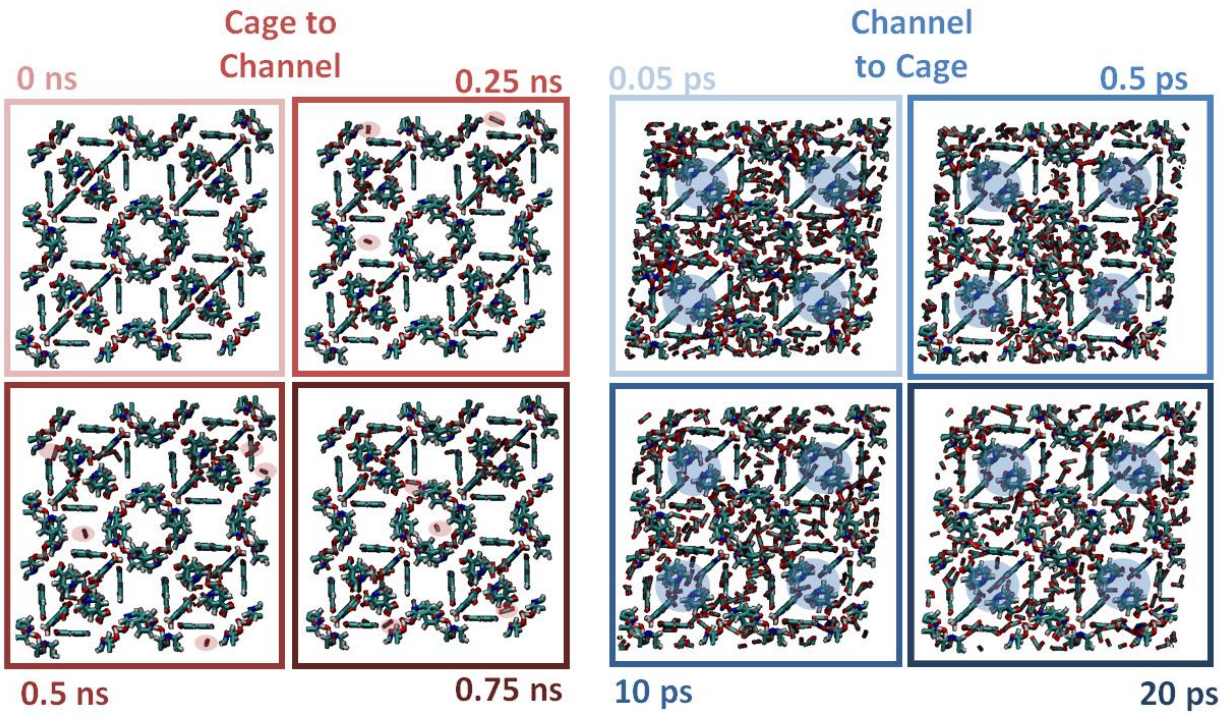


Figure 4.9. Snapshots from the MD simulation of CO_2 diffusing from the cage to the channel and vice versa. Highlighted in red are the CO_2 molecules which have diffused from the cage to the channel and highlighted in blue are the cages with CO_2 molecules inside. This was visualized using VMD.[38]

4.1.3. Conclusions

This section presented Ni-4PyC which is an exceptional pre-combustion CO_2 capture material. This ultra-microporous MOF was built from a small and readily available ligand, can have highly favorable adsorption/desorption characteristics for gas separation processes, despite having pores <6Å in size and a modest surface area (945m2/g). Additionally it has optimal working capacities and CO_2/H_2 selectivities for PSA based pre-combustion CO2 capture that are competitive with the best known MOFs for that application. Simulations were integral in the study of this MOF by ushering increased activation and investigation of high-pressure adsorption. Simulated binding sites suggest that strong cooperative guest-guest interactions, in part, allow for the exceptional 8.2mmol/g CO2 uptake capacity of NIi-4PyC at 10bar, 298K. In addition to possessing favorable gas adsorption properties, Ni-4PyC also exhibits excellent stability and recyclability – properties that are critical for practical operation in gas separation processes. Following 160 hours of steam treatment and 24 hours of exposure to 70 bar pressure, Ni-4PyC structure remains unchanged. Moreover, Ni-4PyC retains its CO_2 adsorption properties following exposure to water. The simple, single ligand synthesis and isolation to the gram scale suggests that potential industrial-level scale ups should also be straight forward. With all these features and considering Ni-4PyC is built from inexpensive and readily available components, it is an attractive candidate for a variety of hydrogen purification applications. Such comprehensive performance with clear synthetic advantages should prompt revisiting ultra-microporous MOFs using small ligands as a design target for solid sorbents for gas separation applications.

4.2. Gas Specific Porosity in a Dynamic Gate-Opening MOF via Coordination Flexibility

The ability to tune a material's porosity based on the amount of guest adsorbed has been a long sought after feature especially in PSA applications. As previously described in section 1.2.4, the desired feature is a large change in uptake adsorbed over a very small pressure range which is usually only possible through a flexible framework MOF. In guest dependent structural flexibility, the interaction energies of the MOF (~20 – 50kJ/mol) are sufficient enough to cause a conformational change.[39] This phenomena has been presented at both low pressure and high pressure regimes, each with their advantages. At low pressure regimes flexible MOF could provide a "switch" used to detect specific gases above or below a certain partial pressure.[40] At high pressure regimes these soft porous materials can be used for selective gas separation membranes.[41] So far, gate opening gas separation MOFs have been built from relatively long linkers with micropores in the range of 8 - 15Å. As shown in the last section, ultra-microporous MOFs built from a single ligand exhibit optimal CO_2 capture characteristics compared to larger MOFs. Unfortunately, ultra-microporous MOFs made from small single ligands are usually quite rigid which inherently results in no flexibility. Thus, being able to incorporate gate-opening behaviour into an ultra-microporous MOF which exhibits excellent CO_2 adsorption capabilities would be paramount.

4.2.1. Experimental Synthesis, Characterization, and Adsorption Properties

One avenue for incorporating flexibility into a MOF would be to alter the coordination strength of the linker with the metal. The strength of coordination between the metal and

organic linker mediates how rigid a structure is. In this work the concept of hard-soft acid-base theorem was used to design a MOF whereby the metal was chosen to be a hard lewis acid and the organic ligand was chosen to be a soft borderline base in order to finely tune the flexibility of a MOF. This has been done at both the mg scale and also up to 10 grams.

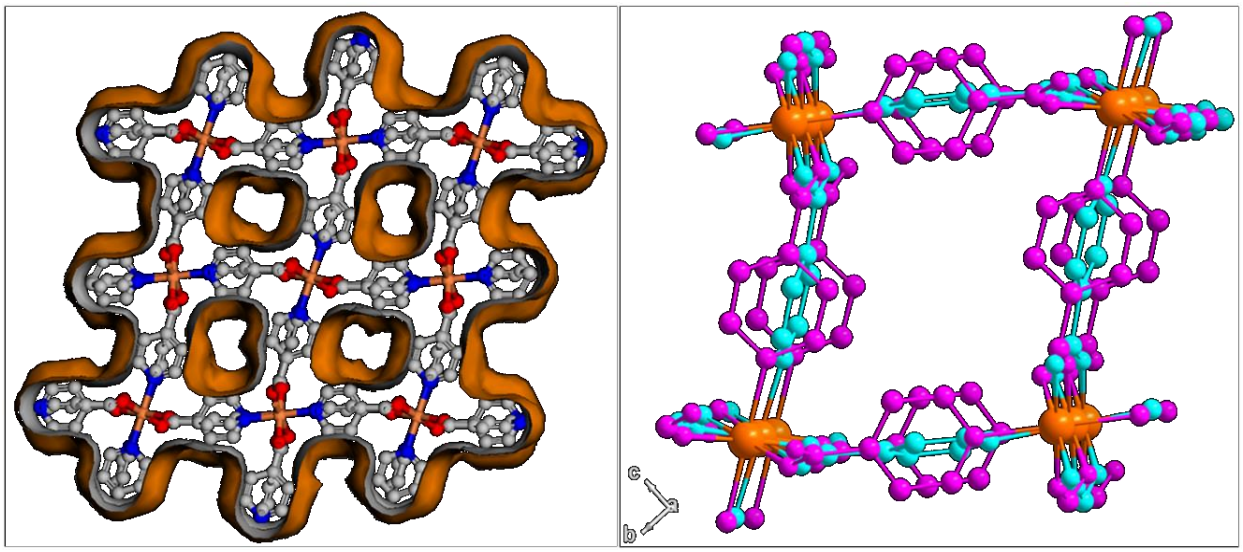


Figure 4.10: (Left) The experimentally determined single crystal structure of Mg-4PyC from x-ray diffraction down the a-axis. The Connolly surface is represented in brown with a probe radius of 1.4Å. The Mg centers are the corners linked together by the PyC ligands into a cubic three-dimensional lattice. Adjacent PyC units are oriented perpendicular with respect to each other. Mg is in orange, carbon in grey, nitrogen in blue, and oxygen in red. (Right) A single channel of Mg-4PyC has been shown where the PyC linkers have been colour coded to indicate the different orientations of linker as well as the different Mg-N distances within the lattice. The purple molecules have Mg(1) atoms (orange) bonding to PyC units such that the Mg(1)-N bond distance is 2.19Å and the cyan molecules have Mg(2) bonding to PyC units such that the Mg(2)-N bond distance is 2.22Å.

The structure of Mg-4PyC is that of a cubic three-dimensional lattice built from linking the Mg centers (which serve as the corners) to 4-PyC (which serves as the sides). (Figure 4.10) There are two crystallographically unique Mg centres connected by bridging carboxylate which form adjacent chains at the corners of the square lattice to form a cubic topology. Alternate PyC rings are rotated by ~90^{o} with respect to each other in an alternating fashion which results in an ultra-microporous MOF with a one-dimensional channel along the a-axis with dimensions 5.5 x 6.5Å. Upon crystallization the pores are filled with DMF guest molecules which help

template the porosity of the MOF. The carboxylate groups form shorter Mg-O bonds with a length of 2.05Å whereas the pyridyl nitrogens form longer Mg-N bond length of 2.19Å and 2.22Å for each of the crystallographically unique Mg centers.

One might expect an MOF with only an ultra-microporous one-dimensional channel to have negligible to no porosity. As expected, the 77K and 298K N_2 adsorption isotherms showed absolutely no uptake. Surprisingly, the 195K and 298K CO_2 isotherm showed adsorption saturation uptakes of 4.7 mmol/g and 2.4 mmol/g respectively. Despite the ultra-micropores made from a single rigid ligand, Mg-4PyC was capable of displaying dynamic gate-opening properties which are dependent on CO_2 pressures. The dynamic gate opening behaviour occurs within the 0-1 bar regime at two different pressure steps and is presently solely with CO_2. At 273K the first step in uptake occurs at ~0.1 bar from 0.3 mmol/g to 1.7 mmol/g indicating the first conformational change. As the CO_2 pressure is further increased to ~0.3 bar the uptake jumps from 1.8 mmol/g to 2.6 mmol/g signalling another conformational change. Finally, as the pressure is further increased there is a gradual increase in uptake to a saturation uptake of 3.2 mmol/g at 1.2 bar. This gate opening behaviour is consistent across different temperature ranges of 298, 283, 273, 263, and 248K. However, a trend occurs whereby as the temperature decreases, the gate opening occurs at lower and lower pressures to the point where absolutely no gate opening behaviour is shown at 195K. CO_2 isotherms at various temperatures are shown below in Figure 4.11.

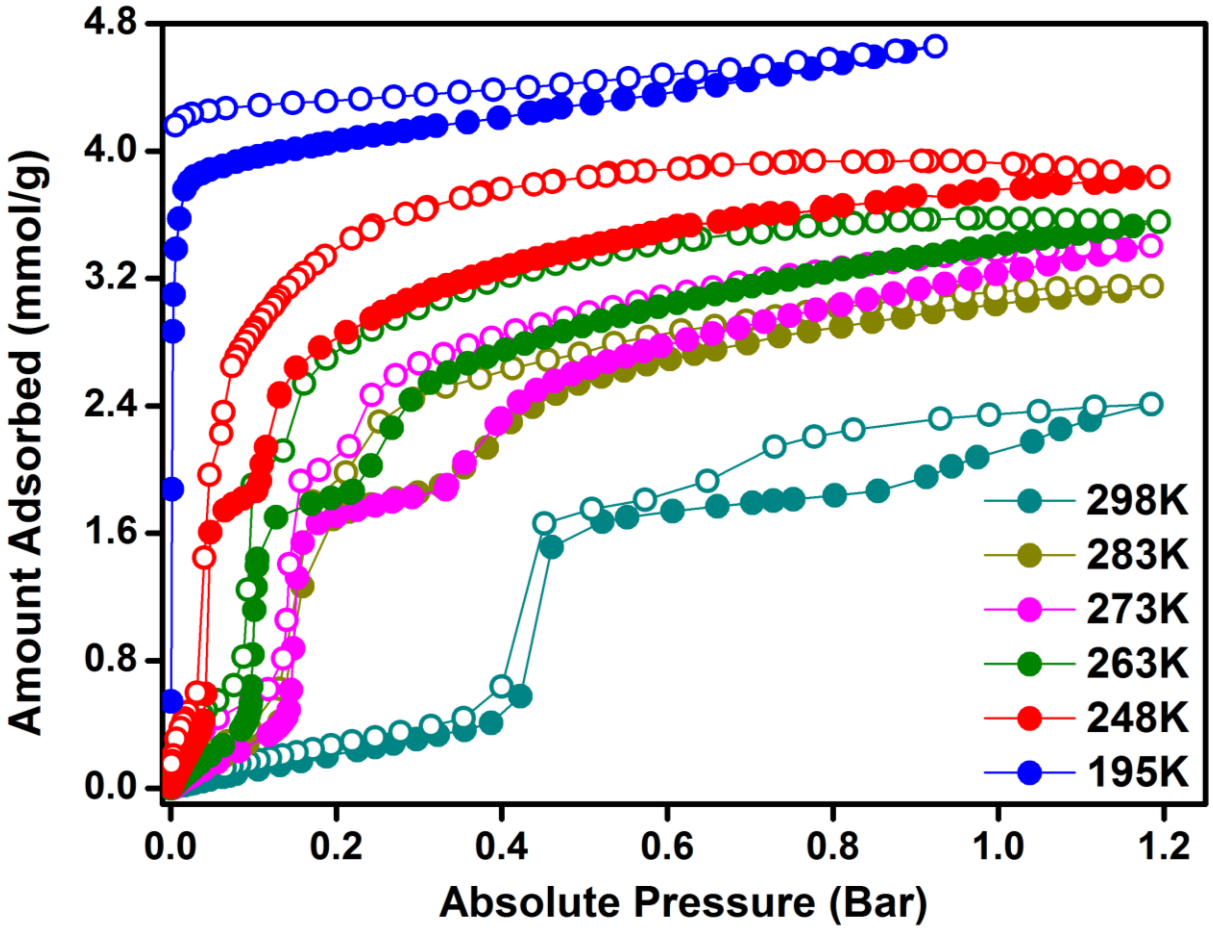


Figure 4.11: Adsorption-desorption isotherms for CO_2 adsorption within Mg-4PyC at various temperatures. Note the gate opening at 0.1 bar and 0.3 bar pressures for the 273K isotherm (pink). The pressure difference between the gate opening steps shrinks as temperature increases to the point where 195K exhibits not gate opening at all.

Due to the variable gate opening with respect to temperature, experiments were carried out to determine whether the dynamic conformational change was due to temperature effects. Single crystal and powder x-ray diffraction studies were carried out at various temperatures and analyzed to see whether there was a large change to the unit cell due to temperature. A single crystal x-ray diffractometer was used to determine a plot of the cell parameters vs. temperature shows that the a- and c-axis cell parameters showed subtle change while the b-axis showed none at all. Although there was a small change in the monoclinic beta angle from 101° to 98°, this minor change can be attributed to structural change due to solvent

loss. Futhermore, PXRD measurements were done in from room temperature to 548K under vacuum. The plot show an absence of any major structural changes. There are however very slight peak shifting (11.0^{o} to 11.35^{o}) due to structural contraction due to loss of DMF solvent and lattice expansion at high temperatures. Surprisingly, Mg-4PyC remains crystalline and completely intact up to temperatures as high as 550^{o}C which is very unusual for a Mg-pyridyl based MOF. Further stability experiments included a TGA cycling experiment. This showed that Mg-4PyC had an initial two-step solvent loss of ~15% from RT to 300^{o}C and exceptional stability up to 480^{o}C. Furthermore TGA was performed for a sample that was freshly prepared and one that had been activated and rehumidified which is a 24hr soak in methanol followed by 180^{o}C heating for 12 hours, then exposure to ambient air for 7 days. No loss in structural stability, even in the presence of water, was observed. Lastly, Mg-4PyC lies on the border of being non-porous or porous due to the small size of its ultra-micropores. With such a small pore any conformational change can cause increased or decreased accessibility of guests. This behaviour was monitored by the changes in the CO_2 self-diffusion coefficients as a function of CO_2 loading. A rate of adsorption experiment was performed on the ASAP2020HD instrument at 273K in the pressure range of 0-1bar and 8 different pressure points were used to determine the diffusion coefficients. As expect from the isotherm, the self-diffusion coefficient showed a significant jump ($9.9x10^{-9}$ to $6.5x10^{-8} m^2s^{-1}$) at the low pressure gate opening point (0.1 bar) and a relatively lower jump ($7.5x10^{-8}$ to $9.1\ x10^{-8} m^2s^{-1}$) at the higher pressure gate opening (0.3bar). Above this pressure, the diffusion steadily increases to a value of $1.2\ x10^{-7} m^2s^{-1}$ giving almost two orders of magnitude increase in diffusion as we go from lowest partial pressure to 1 bar. Experimental plots of the PXRD, single crystal X-ray diffraction, and diffusion results are shown in Figure 4.12.

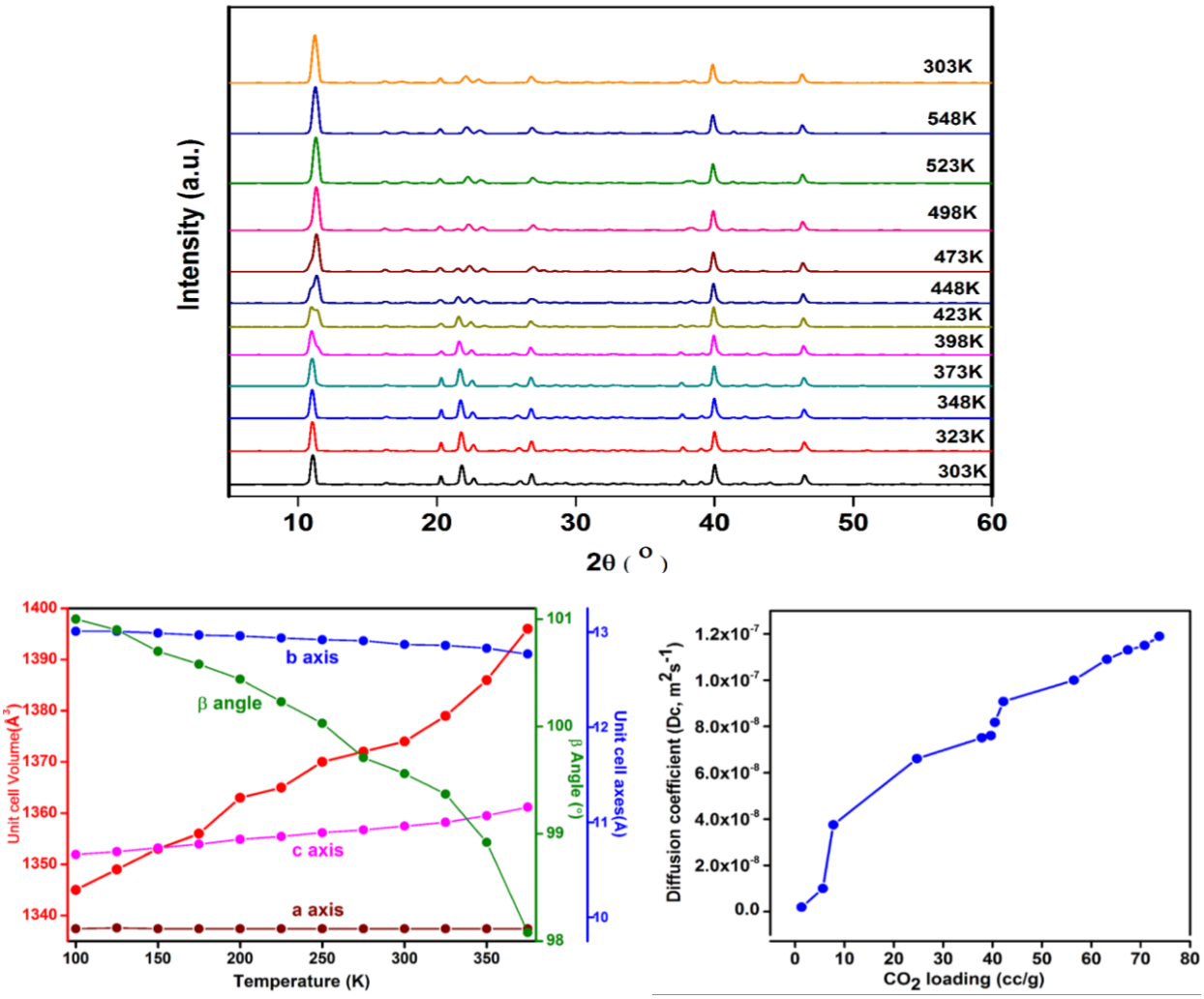


Figure 4.12: (Top) Powder x-ray diffraction (PXRD) plots of Mg-4PyC at temperatures from 303K to 548K showing no major structural change. (Bottom Left) A plot of the change in single crystal unit cell parameters of Mg-4PyC as a function of temperature obtained from single crystal x-ray diffraction. (Bottom right) Plot of the diffusion coefficient of CO_2 as a function of CO_2 loading. Notice the jumps in diffusion at 10 to 30 cc/g and 35 to 42 cc/g loading which correspond to 0.1 and 0.3 bar respectively.

These interesting experimental results demand a more in-depth molecular investigation on the structural dynamics of this gate opening phenomenon. The ability of Mg-4PyC to exhibit this behaviour is a result of the components used to make this MOF. This is facilitated by the coordination flexibility about the Mg-N bonds. Mg^{2+} is a hard lewis acid and pyridyl is a borderline soft base. Hard-soft acid-base theory states that hard acids will bind stronger to hard bases and vice versa. The mismatch between hardness of the metal centre and organic linker of

Mg-4PyC allows coordination flexibility which results in structural changes via rotation of the pyridyl ring via the Mg-N bond. In order to properly analyze this MOFs behaviour, computational simulation was necessary to elucidate details on the dynamic gate-opening via the rotation of the organic linker.

4.2.2. Computational Simulations

4.2.2.1. Searching Conformational Space

Crystal structure shows organic PyC linker existing in two rotational states, one where the plane of the aromatic ring is parallel to the channel along the a-axis and one where the aromatic ring exists perpendicular to it. When the aromatic ring is parallel to the channel it is in an "open" configuration where the pore size is larger and conversely when the aromatic ring is perpendicular to the channel it is in a "closed" configuration where the pore size is smaller. In order to test this hypothesis, a fixed rotational scan of one linker in the unit cell from the crystal structure to the "open" and "closed" configurations was performed and the simulated uptake was calculated via GCMC simulations. Reported uptakes were taken at 1 bar and 298K. As shown in Figure 4.13, the simulated CO_2 uptake increases as the perpendicular linker from crystal structure is rotated 90^o to an orientation parallel to the a-axis and similarly the uptake decreases going to the closed perpendicular orientation.

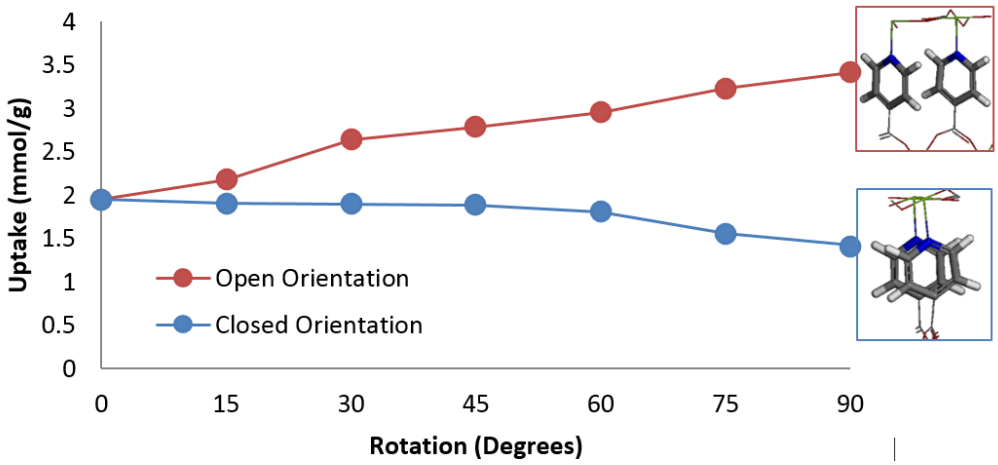


Figure 4.13: Simulated uptake as a function of linker rotation from crystal structure to linkers parallel to the a-axis channel (open orientation) and crystal structure to linkers perpendicular to the a-axis channel (closed orientation). Uptake was evaluated at 1 bar and 298K.

These simulations show that there is a relation between CO_2 uptake and linker rotation which supports the hypothesis that the stepped isotherm is due to conformational changes due to linker rotation. A simulated isotherm was calculated for both "extreme" orientations where the linkers parallel to the a-axis is the open pore structure and the linkers perpendicular to the a-axis is the closed pore structure. It was found that the uptake was quite high for the open pore structure with an uptake of about 5 mmol/g at 1 bar and 273 K and the closed pore structure had absolutely no uptake at all. This open pore structure was optimized with DFT using VASP to reduce the steric clash between hydrogens of the pyridine within the linkers. The optimized structure (Figure 4.14) gave an uptake of 3.5 mmol/g at 1 bar and 273 K which is very close to experiment at similar pressure and temperature conditions. The optimized structure shows the linkers slightly rotated in a way that optimizes pore size but minimizes steric interaction between linkers as shown in Figure 4.14.

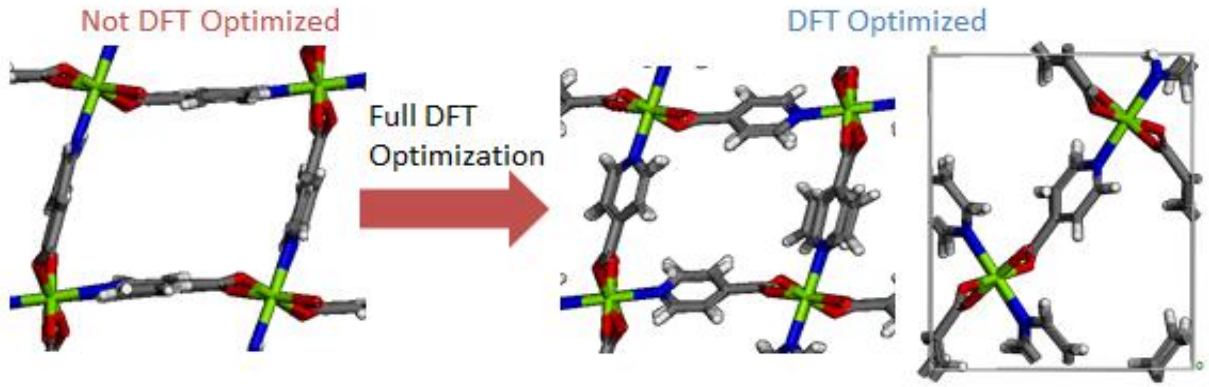


Figure 4.14: The pre-optimized open pore structure (left) where the linker rings are fully parallel to the a-axis and the corresponding DFT optimized structure (right) where the linker rings are slightly rotated in order to reduce steric interactions between hydrogens of the linkers.

Next, in order to search the conformational space of linker rotations an exhaustive set of conformers with different combinations of linker rotations was devised. This set of conformers was optimized using DFT in order to discover the low energy conformations that can exist with respect to linker rotation. The conformations were made from fully closed to fully open conformations of Mg-4PyC. The set of pre-optimized conformations are shown in Figure 4.15.

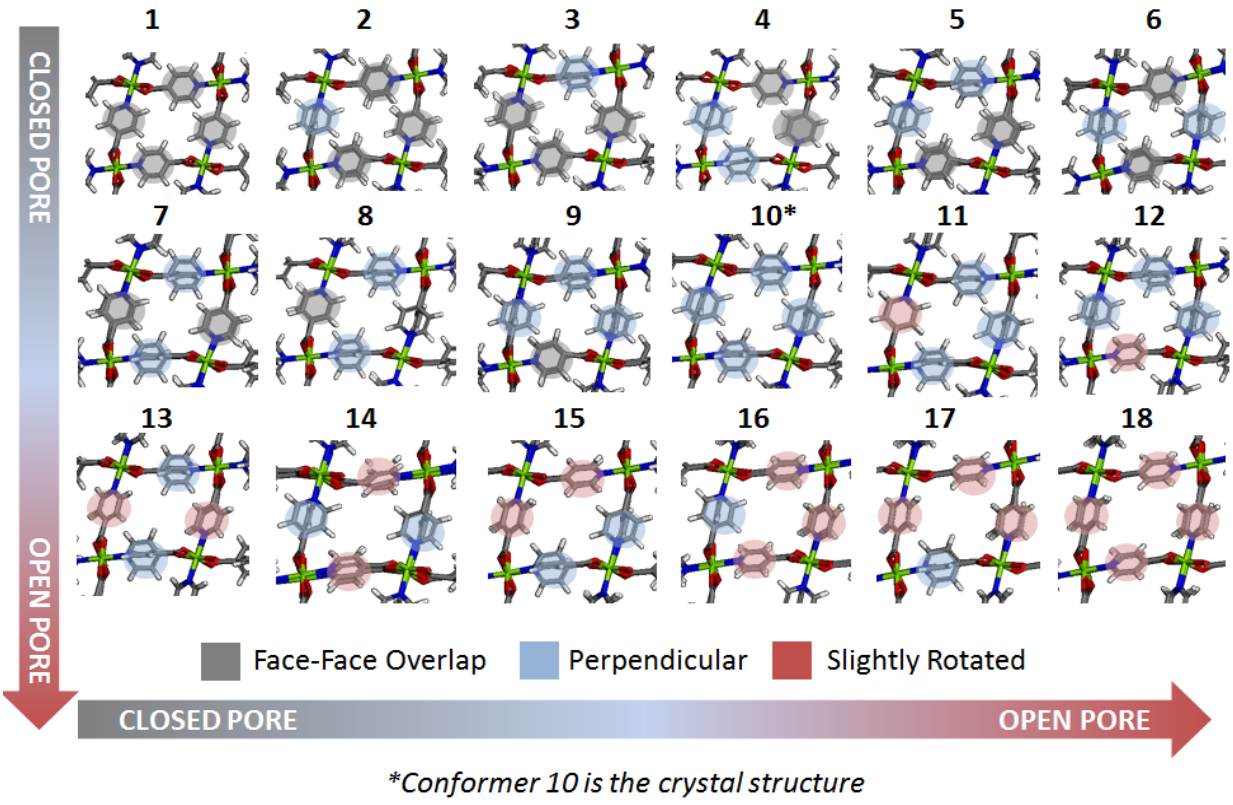


Figure 4.15: The 18 various trial conformers before DFT optimization. The structures are colour coded to show the orientation of the linkers whereby grey are linkers perpendicular to the a-axis channel as shown in the closed pore structure, blue is mixed linkers as shown in the crystal structure, and red is slightly rotated orientation where the linkers are parallel to the a-axis channel as shown in the open pore structure.

It was found that regardless of starting orientation all conformers optimized to five main structures. Two of these conformers contained linkers perpendicular to the a-axis in the "closed" conformation and showed no CO_2 uptake. The relative energies for all of these conformers were calculated and it was found that the crystal structure was lowest in energy (Table 4.4). This may be due to the fact that crystallization with DMF in the pores allows for the Mg-4PyC to relax to a lower energy conformation via a templating effect. A summary of the optimized conformers is shown in Figure 4.16.

Table 4.4. Relative energies of the optimized conformers as found from the conformational search. All conformations optimized to five lowest energy structures.

Pre-Optimized Conformer	Relative Energy (kcal/mol)	Optimized Conformer
1	20.24	1 - Closed Pore
2	10.11	4 - 2 Gates Closed
3	10.11	4 - 2 Gates Closed
4	9.98	4 - 2 Gates Closed
5	-0.27	10 - Crystal Structure
6	2.69	7 - 1 Gate Opened
7	2.84	7 - 1 Gate Opened
8	2.91	7 - 1 Gate Opened
9	2.80	7 - 1 Gate Opened
10	0.00	10 - Crystal Structure
11	2.49	7 - 1 Gate Opened
12	-0.40	10 - Crystal Structure
13	-0.52	10 - Crystal Structure
14	-0.35	10 - Crystal Structure
15	-0.11	10 - Crystal Structure
16	0.09	10 - Crystal Structure
17	2.28	7 - 1 Gate Opened
18	11.50	18 - Open Pore

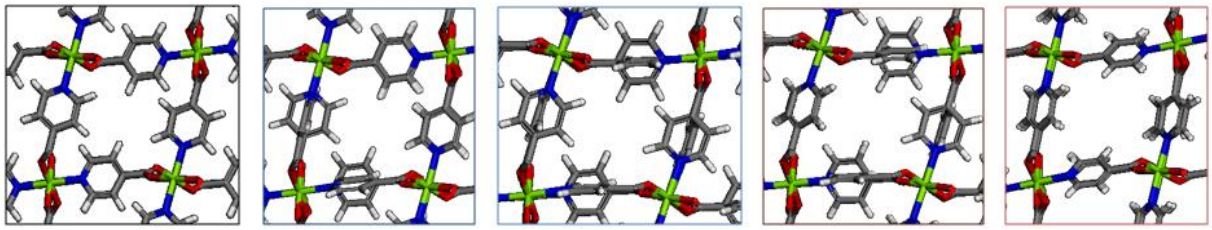


Figure 4.16. DFT optimized lowest energy conformations of Mg-4PyC from closed (left) to open (right).

CO_2 isotherms were calculated at 273 K of all lowest energy conformations and compared to the experimental isotherm. It was found that the crystal structure conformation yielded an isotherm which was in good agreement with the first step of the experimental isotherm from 0.18 bar to 0.3 bar. Furthermore, the optimized open pore structure had a saturation uptake at 1 bar (3.5 mmol/g) similar to that of experiment at the same pressure (3.35 mmol/g). To further investigate the open pore structure, the isotherm of the open pore

structure was calculated at 195 K and compared to the experimental isotherm. At low temperature the experimental type 1 isotherm exhibits no steps and thus should be one conformation. Due to the high experimental uptake of 4.6 mmol/g at 0.92 bar (195K) it stands to reason that this conformation is the open pore structure. The open pore structure gave a simulated isotherm which is in great agreement with experiment as shown in Figure 4.17.

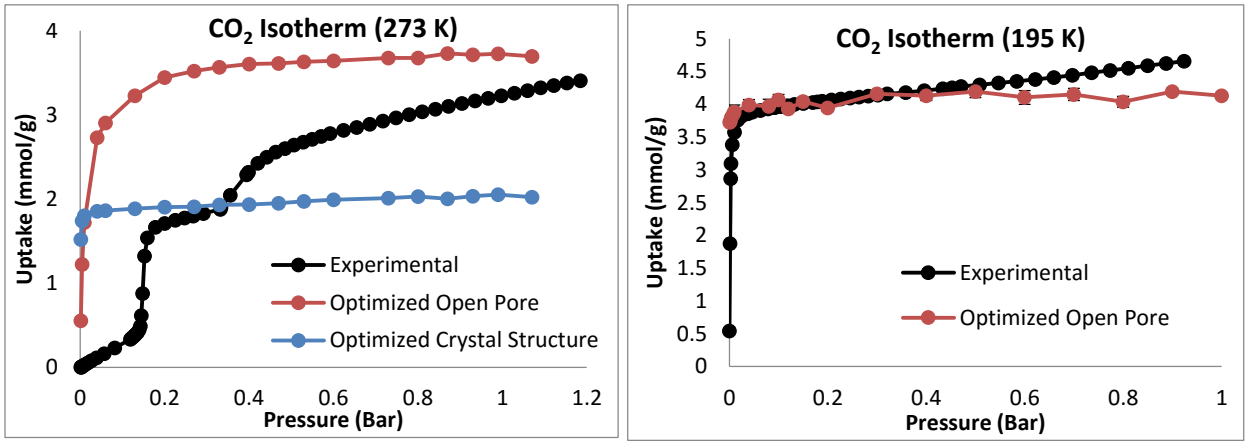


Figure 4.17: CO_2 isotherms at 273 K (left) and 195 K (right) comparing experimental isotherms to computationally optimized structures of the crystal structure and open pore conformers.

The strong match of computational to experimental isotherms shows confidence in finding structures which correspond to the first step in the isotherm (0.2 bar to 0.35 bar) as well as the final structure (1.2 bar). However, many questions remain about this interesting MOF's dynamics which cannot be explained by static structural representations. First, is the conformational change truly dependent on CO_2 loading? Secondly, are the conformational changes gradual or more abrupt discrete shifts? Third, what are the conformational orientations of the linkers during the shift in uptake? Lastly, what is happening at low pressure in the experimental isotherm (0 – 1.8 bar)? In order to answer these questions molecular dynamics simulations were utilized in order to gain information from a dynamic representation of the system rather than static one.

4.2.2.2. Developing an Accurate Potential for Linker Rotation

All MD simulations were performed using the GROMACS[42] program which allows for custom parameters and the use of a more accurate Buckingham potential.[43] The Universal Force Field (UFF)[44] is a flexible interaction potential that is commonly employed for the study of MOFs. UFF includes bond stretching, bond bending, and bond dihedral torsion for intramolecular interactions and a Lennard-Jones potential for non-bonded interactions. This forcefield served as the base with which to modify in order to develop a more accurate potential for linker rotation. The ε and σ parameters of CO_2 were taken from García-Sánchez et al.[20] which were developed to fit experimental adsorption isotherm data in zeolite frameworks. The C-O bond length (1.149 Å) and partial charges on CO_2 atoms (C = +0.6512e, O = -0.3256e) were taken from the potential by Harris and Yung.[21] A GROMACS utility script was used to convert Lennard-Jones ε and σ parameters into Buckingham a, b, c potential parameters for the framework and any guests. All intermolecular interaction potential parameters are listed below in Table 4.5.

Table 4.5: Intermolecular potentials of the framework and guests used in MD simulation.

	Lennard-Jones		Buckingham		
Atom Type	**sig (kcal/mol)**	**eps (nm)**	**a**	**b**	**c**
Framework					
C_2	0.3430851	0.439614	71549.3	31.1607	0.00287
H	0.2571134	0.184219	29982.5	41.5801	0.00021
C_R	0.3430851	0.439614	71549.3	31.1607	0.00287
Mg	0.2691405	0.464735	75637.9	39.7219	0.00071
N_R	0.3260689	0.288889	47018.2	32.7869	0.00139
O_2	0.3118146	0.251208	40885.3	34.2857	0.00092
CO_2					
Cx	0.05948	0.2745	44676.2	179.737	0.00000
Ox	0.17023	0.3107	50567.9	62.802	0.00003
DMF					
O_3	15.9994	-0.68	142899	36.1175	0.00236
N_2	14.0067	0.04	115719	32.8947	0.00335
C_3	12.0107	-0.11	44920.3	32.8947	0.00130
H_1	1.0079	0.06	20507.1	42.7631	0.00012
C_4	12.0107	0.5	71449.3	28.5088	0.00488
H_2	1.0079	0	20507.1	42.7631	0.00012

Before any MD simulations could be performed the UFF potential was fitted to DFT. First was obtaining a standard at the DFT level with which to compare to. A torsional scan was performed and the single point energies were calculated using DFT in order to determine the potential energy surface (PES) of the linker rotation. A toy system (Figure 4.17.) was used where all the pyridine rings were deleted and the dangling bonds were capped with hydrogen except one. This was done in order to isolate the energetic contributions to be purely from the rotation of one ring and in effort to eliminate any steric contributions from neighbouring rings.

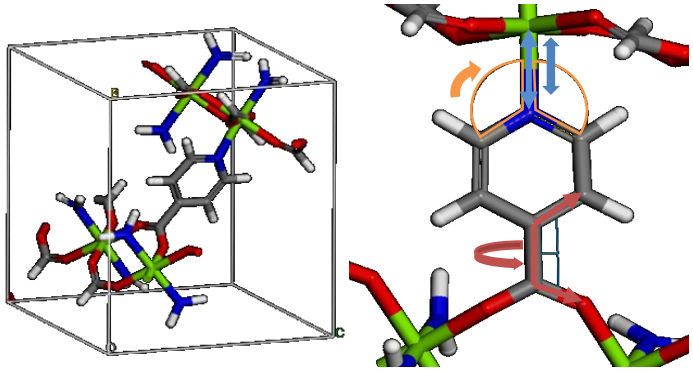


Figure 4.18: (Left) A toy system of Mg-4PyC used to calculate the potential energy surface of linker rotation. (Right) The intramolecular modes of importance with respect to linker rotation. The Mg-N-C angle is shown in orange, the Mg-N bond is shown in blue and the O-C-C-C dihedral angle is shown in red.

A target potential was calculated by subtracting the potential as calculated by DFT and the potential as calculated by UFF with all parameters contributing to the rotation turned off. The most important potential contributions are that of the O-C-C-C torsion, Mg-N bond, and Mg-N-C angle as shown in Figure 4.18. Thus, the target potential takes the form, (**Eq 4.4**)

$$E_{target} = E_{DFT} + \left(E_{UFF} - E_{OCCC\ tors} - E_{MgNC\ angle} - E_{MgN\ bond}\right) \qquad (\mathbf{4.4})$$

where E_{DFT} is the energy as calculated by DFT, E_{UFF} is the energy as calculated by the default forcefield and $E_{OCCC\ Tors}$, $E_{MgNC\ angle}$, $E_{MgN\ bond}$, are the energetic contributions due to the torsional rotation about the O-C-C-C dihedral, the Mg-N-C angle, and the Mg-N bond. Thus the target potential is the energy contributions of these terms at the DFT level. GROMACS uses a cosine functional form for the dihedral torsional potential which follows this form, $V_{d(\phi ijkl)} = k\phi(1 + \cos(n\phi - \phi_s))$ where k is the dihedral constant in kJ/mol, n is the multiplicity, ϕ is the equilibrium dihedral angle in degrees. The UFF default has k = 5.2335 kJ/mol, and $\phi = 180^o$. The modified potential was fitted to the target potential by a least squares fit and has k = 2.1281 kJ/mol and $\phi = 164.0899^o$. The PES as calculated by DFT and by the fitted UFF potential is

shown in Figure 4.19 and shows a reasonable fit in terms of the rotational barrier being ~ 10kcal/mol.

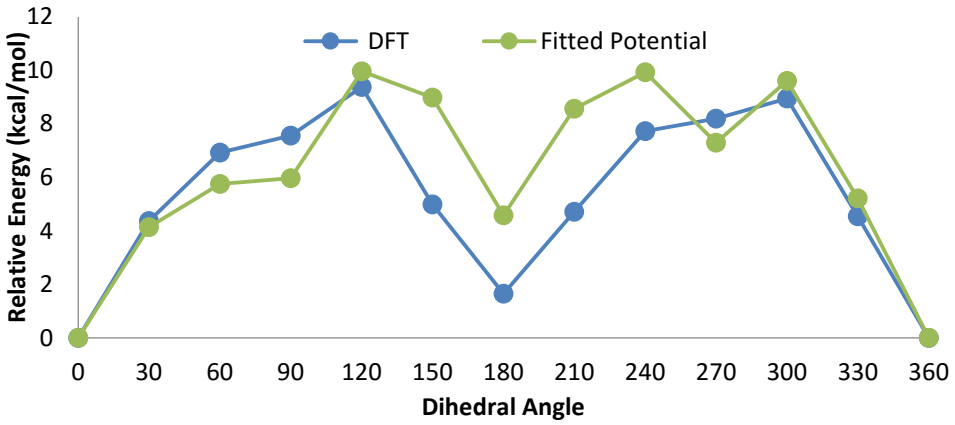


Figure 4.19: The potential energy surface of the dihedral rotation of a Mg-4PyC toy model. In blue is the potential as calculated by DFT and in green is the potential as calculated by a fitted UFF potential. The rotational barrier of ~10 kcal/mol is in great agreement.

4.2.2.3. Validation of Modified Potential

In order to gain more confidence in the modified potential, a few tests were done in order to determine its validity. The first test was to see whether or not the modified potential recreated the crystal structure under similar conditions. Thus, a 1 ns MD simulation with 0.2 ns of equilibration, 1 ns for the production run, and a time step of 0.001 ps with an NVT ensemble with the optimized crystal structure with DMF in the pores as the starting initial configuration in a 3x2x2 supercell. The temperature used was 100K to recreate experimental conditions. The forcefield parameters for DMF were taken from the work of Vasudevan et al.[45] which was developed specifically for improved prediction of bulk properties of DMF. The final configuration from the 1ns simulation matches the crystal structure almost perfectly as shown in Figure 4.20 thus showing that the modified potential can accurately reproduces the experimentally determined crystal structure at 100K when initiated from the crystal structure.

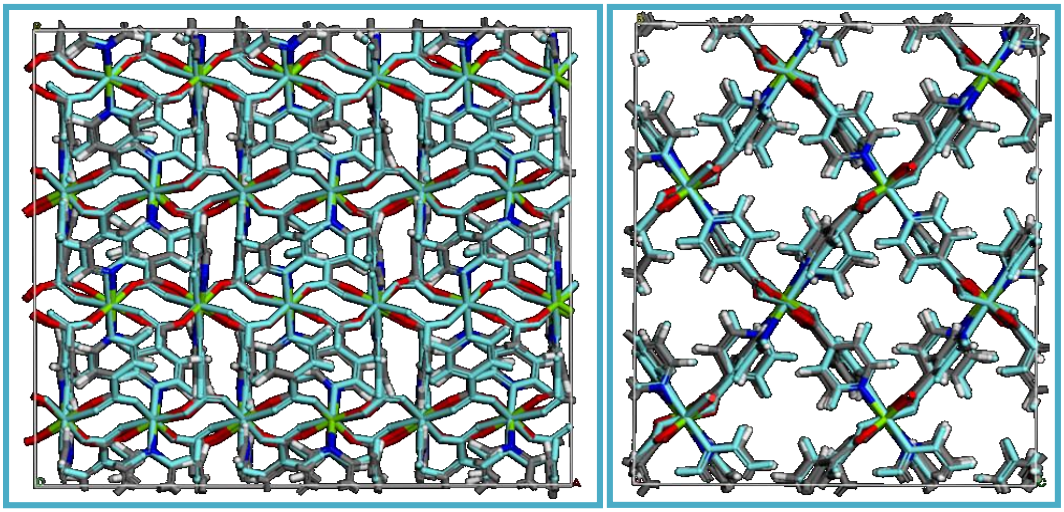


Figure 4.20. An overlay of the output structure of Mg-4PyC from a 1ns (100K) MD simulation (blue) over the crystal structure. On the left is a side view down the b-axis with the a-axis running laterally and on the left is a front view down the a-axis showing the middle of the channels. DMF molecules have been deleted from the representation for clarity.

Furthermore, an MD simulation was run where 1/3rd of the supercell was occupied with DMF and the other 2/3rd was occupied by CO_2. A 1 ns simulation at 100K with 1 milion steps (1 fs timestep) was performed. The saturation loading used was 8 DMF molecules and 32 CO_2 molecules. It was found that the CO_2 molecules can diffuse through the channels of the MOF whereas the DMF stays localized to its initial pore. Furthermore, the CO_2 molecules do not diffuse in the b or c direction and remain within the a-axis channel. No hopping from channel to channel was observed. A snapshot of the final configuration shows that the pores with DMF retained the crystal structure motif whereas the pores with CO_2 are in more of a rotational flux with some dihedral angles in the "closed" position as shown in Figure 4.21.

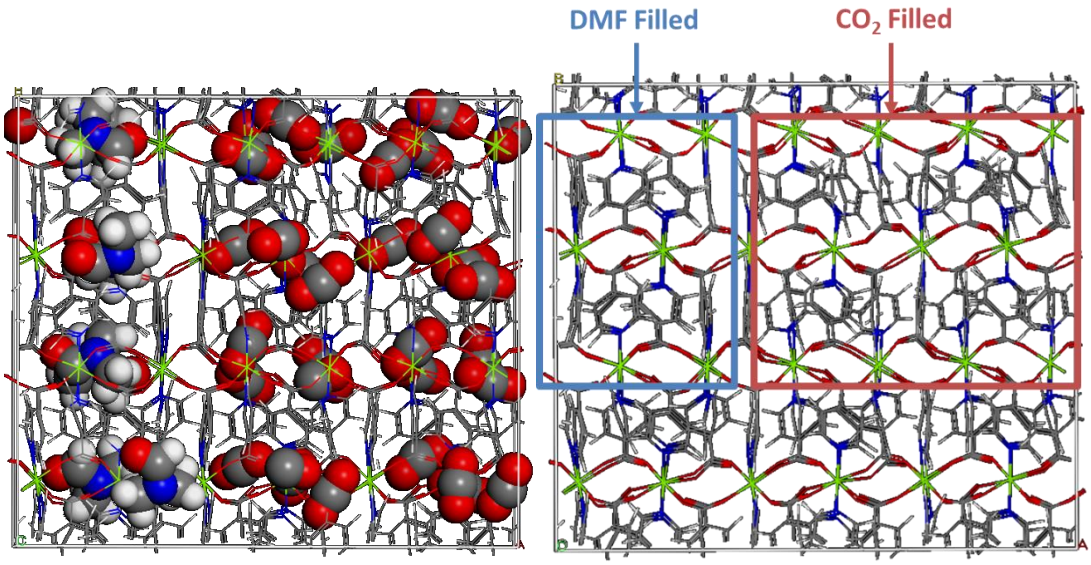


Figure 4.21: Final configuration of 100K MD simulation of mixed DMF/CO_2 loading (left). The DMF filled pores maintain a crystal structure motif whereas CO_2 filled pores are in flux (right).

Additionally, it was found that there was no lateral diffusion of DMF within the channel and no conformational shift from the crystal structure conformation to either open or closed structures. Furthermore, a dihedral distribution was calculated for this MD simulation and the peaks show dihedral angles matching that of the static crystal structure relatively well, but the dihedral angle distribution with a mixture of DMF and CO_2 show much broader peaks representing increased disorder. The dihedral distribution was calculated from recording the dihedral angle at every step of the MD simulation. A GROMACS utility script was used to calculate the dihedral angle distribution.

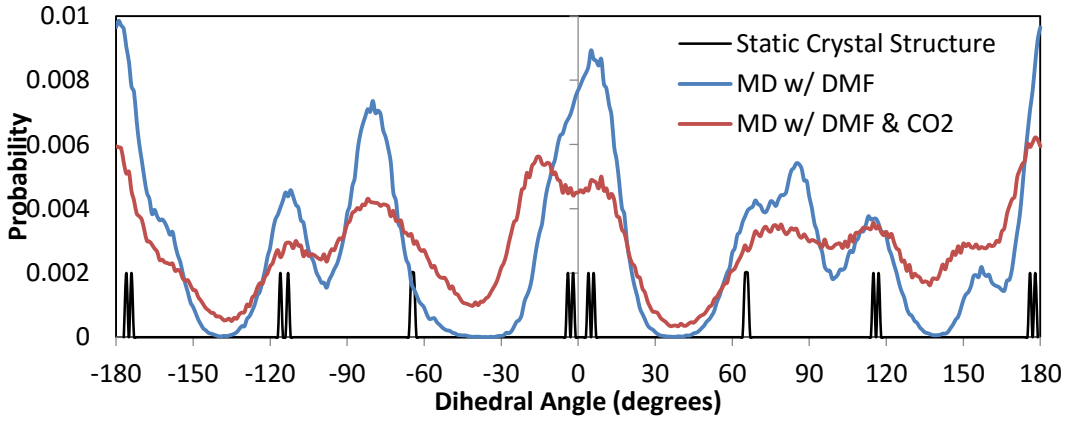


Figure 4.22: Dihedral angle distributions from 1 ns MD simulations of Mg-4PyC saturated with DMF (blue) and DMF/CO_2 mixture (red).

In order to test the modified potential further, the crystal structure was optimized using the modified potential. It was found that the optimized structure using the modified potential was in very good agreement to the experimental crystal structure and the DFT optimized structure as shown in Table 4.6. Listed in Table 4.6 are the linker dihedral angles as well as the N-Mg bond length for the various structures.

Table 4.6: Dihedral angles and N-Mg bond lengths for crystal structures optimized by the modified UFF forcefield and DFT as well as the original experimental crystal structure.

	Modified Forcefield Optimized	**DFT Optimized**	**Experimental Crystal Structure**
Dihedral 1	7.365°	-5.007°	-5.002
Dihedral 2	62.126°	65.349°	64.594
N-Mg bond 1	2.153 Å	2.169 Å	2.214
N-Mg Bond 2	2.230 Å	2.216 Å	2.206

Finally, the energetics of the MD structures obtained from the modified forcefield were compared. A 1ns MD simulation at 100k was performed on a 3x1x1 supercell using the modified forcefield with the crystal structure as the initial configuration. The initial crystal structure and final MD structure were optimized using the modified forcefield in GROMACS. A DFT single point calculation was done on these structures using VASP. The forcefield optimized output structure from 100K MD was found to be 15.09 kcal/mol higher than the crystal structure which is reasonable. Conversely the initial crystal structure and MD structure were optimized with DFT in VASP. The DFT optimized final MD structure was 17.46 kcal/mol higher than the initial crystal structure. Thus, regardless of optimization by classical forcefield or DFT, the modified potential produces structures which give consistent and reasonable relative energies when evaluated by DFT.

4.2.2.4. MD Insights on Pressure Dependent Conformational Changes

Now that a force field was developed that allowed for the rotation of an SBU, a series of MD simulations were devised to probe the gradual conformational change with respect to pressure. A series of MD simulations were performed on a 3x2x2 supercell of Mg-4PyC at 273K with a range of loading from empty to 44 CO_2 molecules per supercell which corresponds to the uptake at a pressure of 1.2 bar. All simulations were 1 ns in length and run for 1000000 steps (1 fs timestep). It was found that the linker rotation was rapid and dynamic. The empty configuration showed a final conformation with many linkers perpendicular to the a-axis channel. The accessible surface area was calculated using Materials Studio[7] and a Connolly surface area probe of 1.72 Å which corresponds to the VDW surface of CO_2 . It can be shown that when there is no CO_2 loading the linkers are free to rotate and thus the channel that runs along the a-axis is no longer connected but becomes disjointed pockets. This is a representation of Mg-4PyC in a "closed" state as shown in Figure 4.23.

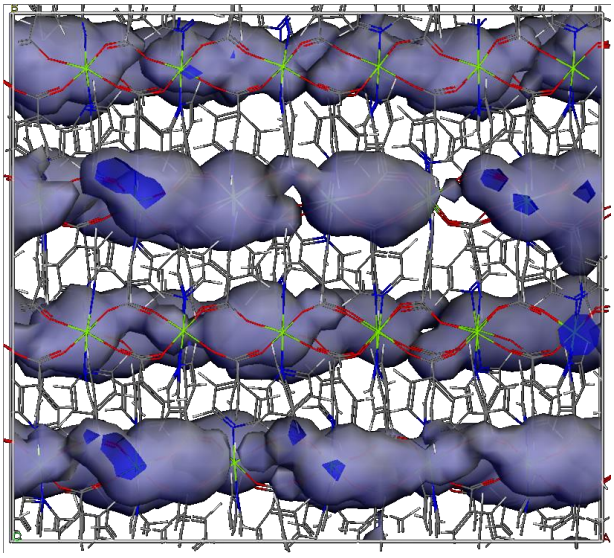


Figure 4.23: The final configuration of a 1 ns MD simulation of Mg-4PyC with no guests. The accessible surface area is represented by the blue surface and was calculated in Materials Studio.

As CO_2 loading increased, it was found that the linkers rotated in such a way to accommodate the greater number of CO_2 molecules. This was quantified by calculating the accessible surface areas of Mg-4PyC configurations from each MD simulation. At each loading a snapshot configuration was taken at 0.6 ns, 0.8 ns, and 1ns then the CO_2 molecules were deleted from the framework and the surface area was calculated for each of these structures using ZEO++[30] with a probe size of 1.72Å. The average of these surface areas were plotted against the CO_2 loading which corresponds to a specific pressure. It was found that as the number of CO_2 molecules increased the accessible surface area of area of Mg-4PyC increased as well. Interestingly, the stepped feature at ~1.8 bar is also seen when looking at the average surface area as a function of pressure as shown in Figure 4.24. The average maximum pore size was also calculated with ZEO++ and showed an increase at empty loading from 3.95 Å up to 4.73 Å at a loading of 44 CO_2 molecules corresponding to a pressure of 1.2 bar.

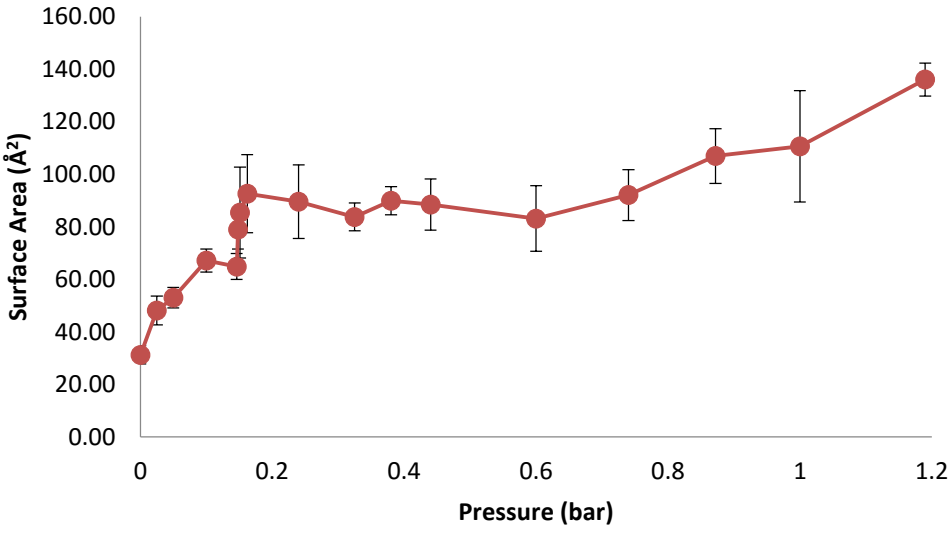


Figure 4.24: A plot of average surface area as a function of pressure. Note that this is not a simulation pressure, but rather the pressure taken from the experimental isotherm and a specified CO_2 loading. Surface areas are averages calculated from structures at the 0.6 ns, 0.8ns, and 1ns timesteps from the MD simulation.

The conformational change with respect to CO_2 loading can also be quantified by analyzing the dihedral distribution. Dihedral distribution plots were calculated for all MD simulations and it was found that as more CO_2 molecules are loaded into Mg-4PyC, the distribution starts to favour dihedral orientation that is parallel to the a-axis channel. This is shown by the increase in peaks at -180°, 0°, and 180° which is indicative of a greater presence of linkers in the "open" configuration. Plots of the dihedral distribution with 1 CO_2 (0.03 bar), 16 CO_2 (0.15 bar), 24 CO_2 (0.33 bar) and 48 CO_2 (1.25 bar) loading are shown below in Figure 4.25.

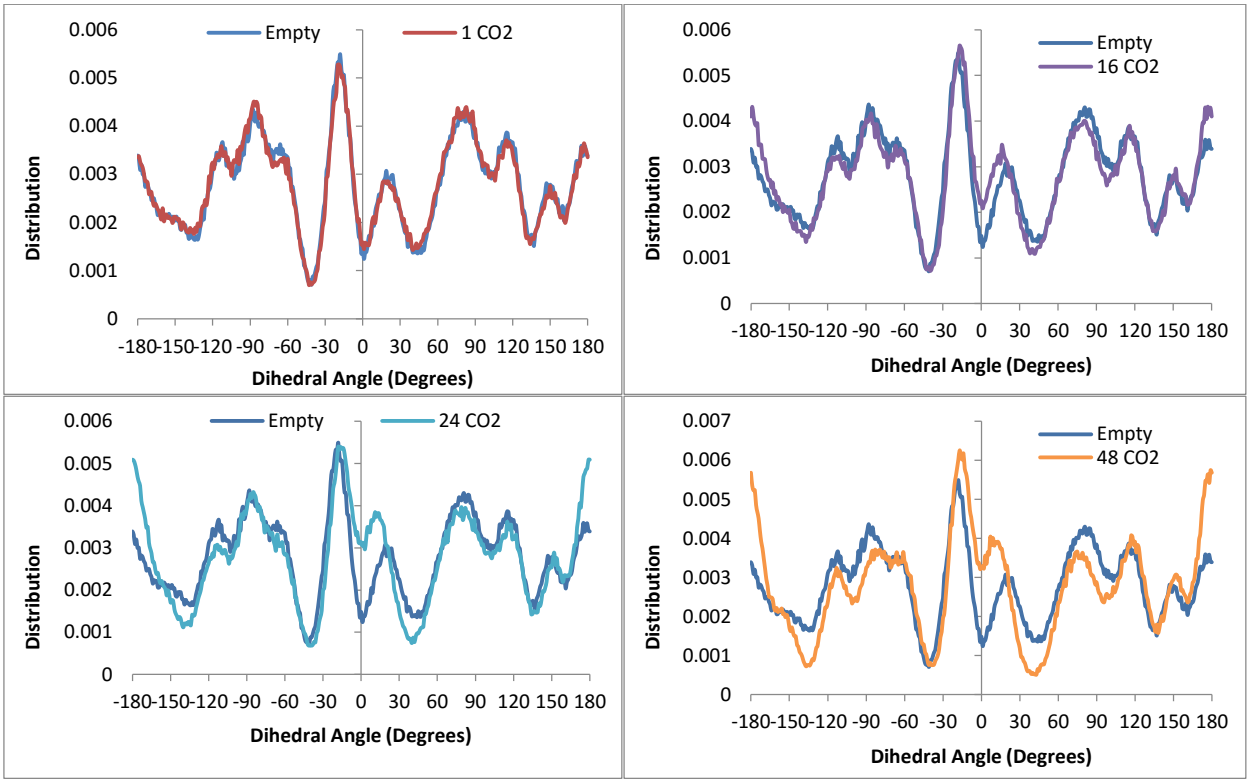


Figure 4.25: Dihedral distribution functions from 1 ns MD simulations of a 3x2x2 supercell of Mg-4PyC. Top left corresponds to a loading of 1 CO_2 per supercell and a pressure of 0.03 bar. Top right corresponds to a loading of 16CO_2 per supercell and a pressure of 0.15 bar. Bottom left corresponds to a loading of 24 CO_2 per supercell and a pressure of 0.33 bar. Bottom right corresponds to a loading of 48 CO_2 per supercell and a pressure of 1.25 bar.

Notice how the dihedral distribution of the empty MOF matches that of 1 CO_2 almost perfectly. This shows that at low pressure the number of CO_2 molecules is not sufficient to

induce a significant conformational change. To quantify the change in dihedral angles, the ratio of open (angles = -180°, 0°, 180°) to closed (dihedral angles = -90°, 90°) was calculated by dividing the value of distribution of these angles for all MD simulations. At an empty loading it was found that this ratio is 1.05, which corresponds to a 1:1 mixture of dihedral angles in open and closed positions. However, as loading increases the ratio starts to favour the open conformations and show that more organic linkers are spending their time parallel to the a-axis channel with dihedral angles of -180°, 0°, and 180°. At 1.2 bar there are twice as many linkers in an open configuration than in a closed one as shown in Table 4.7.

Table 4.7: Ratio of open to closed linker conformations as calculated from the dihedral angle distributions from MD simulations.

CO_2 Molecules per Supercell	Pressure (bar)	Ratio of Open to Closed Linker Conformations
0	0	1.05
1	0.025	1.02
2	0.05	1.08
4	0.1	1.05
8	0.146	1.22
12	0.148	1.28
16	0.151	1.45
20	0.162	1.63
22	0.24	1.55
24	0.325	1.82
28	0.38	1.79
32	0.44	1.79
34	0.5	1.94
36	0.6	1.89
38	0.74	1.89
40	0.872	2.01
42	1	2.15
44	1.19	2.28
48	1.25	2.48

Now that it has been established that CO_2 loading induces a conformational change, further efforts to prove this change is gradual with respect to pressure were done. The CO_2

uptake was evaluated via GCMC simulations for MD snapshot framework structures at different CO_2 loadings which were the same structures used to calculate the surface areas. Therefore each pressure point on the isotherm is the average of uptakes calculated from unique structures which were determined from MD simulations with corresponding CO_2 loading (Figure 4.26). At each pressure point the calculated uptake was similar for each structure taken at snapshots of 0.6 ns, 0.8 ns, and 1ns with the greatest standard deviation being only 0.215 mmol/g. This shows that the structures sampled from this time range in the MD simulation are fairly consistent with each other in terms of uptake. Furthermore, it was found that the isotherm produced from these structures matched the experimental isotherm quite well after 0.2 bar in terms of trendas shown in Figure 4.26.

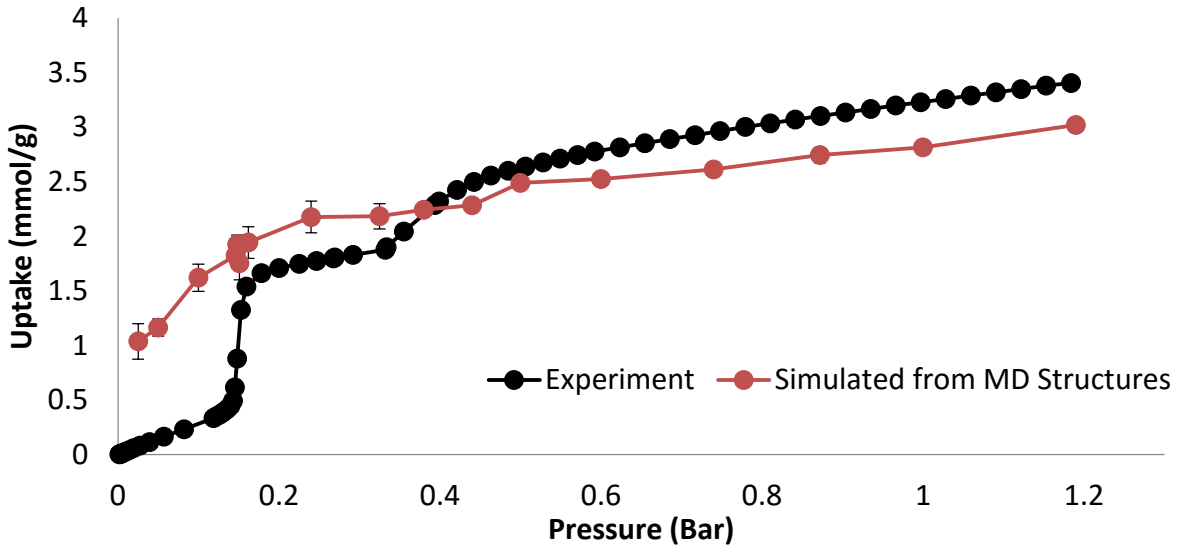


Figure 4.26: Simulated isotherm where each pressure point is an average uptake calculated from structures derived by MD simulations. Error bars represent the standard deviation calculated from structures taken at 0.6, 0.8, and 1 ns snapshots.

There is a clear gradual increase in uptake from 0.6 bar to 1.2 bar and the standard deviation of the uptakes calculated in this range is very low (0.04 mmol/g – 0.07 mmol/g). This demonstrates that as the pressure increases there is a gradual change in the conformation of

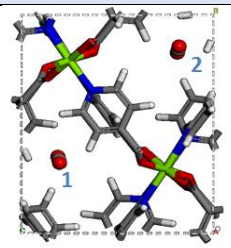
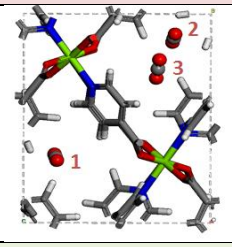
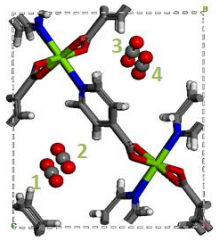
Mg-4PyC via the rotation of the linker rings which accommodate more CO_2 molecules to pass through.

So far computational simulation has been able to explain the experimental isotherm from 0.18 bar to 1.2 bar via a gradual conformational change from crystal structure to open pore structure starting around 0.35 bar. However, the initial low pressure regime is still unexplained. One hypothesis for this low pressure regime is that the Mg-4PyC exists in a closed conformation preventing the CO_2 from entering the MOF and the minimal uptake from 0 bar to 0.18 bar is due to CO_2 adsorbing on the surface.

4.2.2.5. Binding Sites and Energetics

The binding sites were calculated for the optimized crystal structure, partially open structure, and open pore structure as determined from the previous conformational search. The loading of 2, 3 and 4 CO_2 molecules per unit cell correspond to pressures of 0.325 bar, 0.75 bar, and 1.2 bar respectively. Since these are calculated from GCMC simulations, while each structure *can accommodate these CO_2 molecules, the CO_2 molecules might not be able to diffuse through the MOF*. The binding sites were determined via the lab's binding site location algorithm (ABSL) and energies by single point periodic DFT calculations via VASP[8]. The percent electrostatic contribution was determined from DL_POLY binding energies. A summary of the binding sites and energies can be shown below in table 4.8.

Table 4.8: Binding sites and energies for crystal structure, partially open structure, and open pore structure of Mg-4PyC.

	Crystal Structure			
Binding Site	**VASP Binding Energy (kcal/mol)**	**DL_POLY Binding Energy (kcal/mol)**	**% Electrostatic Contribution**	
1	-9.24	-10.99	16.90	
2	-9.20	-11.11	15.84	
	Partially Open Pore Structure			
1	-8.83	-10.93	16.00	
2	-8.24	-10.11	12.75	
3	-8.24	-10.89	14.14	
	Open Pore Structure			
1	-7.93	-8.84	-1.88	
2	-7.18	-9.42	0.21	
3	-7.72	-9.31	-1.71	
4	-7.53	-9.57	0.02	

Next, to further investigate whether the rotation of the linkers is driven by the binding, we compared the energetics of CO_2 binding versus the energetic cost of conformational change via linker rotation. It was found that the cost of rotating all of the linkers to their open pore state is about 9.7 kcal/mol. However, the total energy with four CO_2 molecules within the unit cell is about -49 kcal/mol. This number of molecules relate to the uptake at 1bar and 273K. Thus, the energy associated with binding of CO_2 can overcome the cost of conformational change which further supports the rotation occurring. A graph of relative energies of the various systems is shown in Figure 4.27.

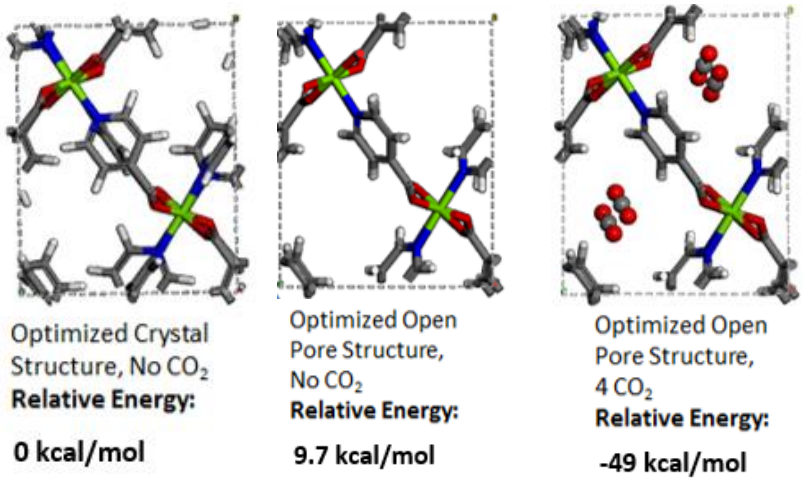


Figure 4.27: Relative energy comparing the energy of conformational change and CO_2 loading. The standard being compared to is the optimized crystal structure with no CO_2 molecules. The optimized open pore structure is 9.7 kcal/mol higher than the empty crystal structure. However, once this structure is filled with four CO_2 molecules, there is a decrease in energy to -49 kcal/mol.

In addition to a looking at the energy of the system, the cooperative binding energy was also investigated. It was found that there was a significant reduction in binding energy with the addition of CO_2 molecules into the MOF. This shows that there is significant cooperative binding occurring within the MOF which was calculated using high level DFT calculations and CO_2 positions from ABSL placements. The cooperative binding energy was calculated per molecule in a system of 2, 3, and 4 molecules within the MOF. This cooperative binding energy per molecule increased from -7.47 kcal/mol to -7.74 kcal/mol and finally -7.97 kcal/mol in a system of 2, 3 and 4 molecules respectively. A graph of cooperative binding energy with inserts of the representative systems are shown below in Figure 4.28.

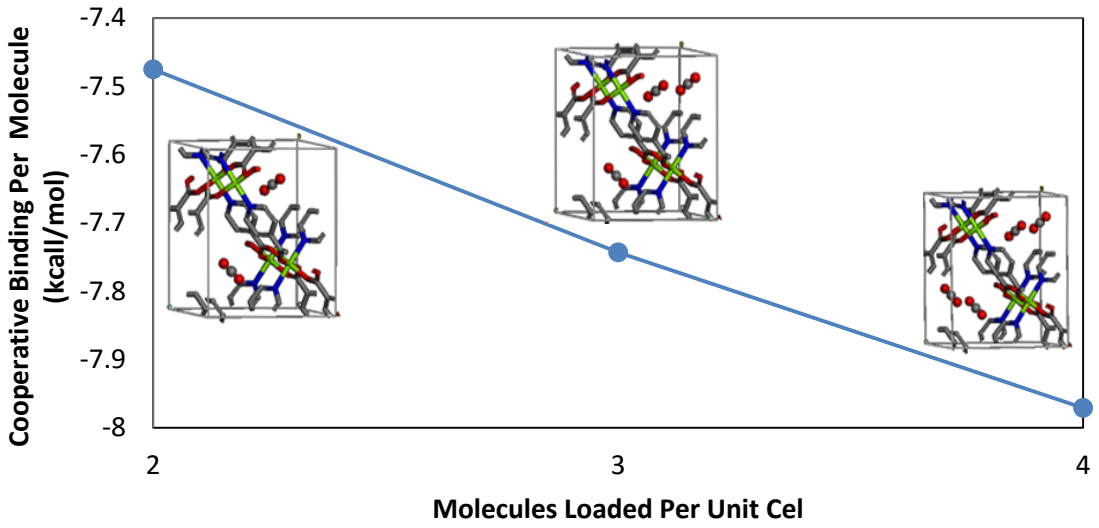


Figure 4.28. Cooperative binding energy as a function of molecules loaded in Mg-4PyC. Inserts of the various unit cells with the appropriate CO_2 loadings as are also shown.

4.2.3. Conclusions

This section presented Mg-4PyC which was a flexible dynamic gate-opening MOF which is made possible through the intelligent design of metal-ligand coordination. The CO_2 specific gate-opening in this rigid framework is unprecedented and a thorough structural analysis could provide valuable molecular level insights key to modular-design of a wide range of soft porous solids. Our simulations show a reasonable agreement to experimental CO_2 uptake from a pressure range of 0.15 to 1 bar. As the pressure increases to 0.15 bar, the increased loading of CO_2 molecules results in greater guest-host interaction. This causes the bulk Mg-4PyC to undergo a conformational change into the crystal structure conformation. This can accommodate CO_2 molecules until 0.33 bar at which point it undergoes another conformational change as evidenced by surface area calculations. Finally, at 0.33 bar there is now a gradual conformational change from the crystal structure to the open pore structure whereby the

linkers are rotate in such way to maximize the channel size but also minimize steric clash. The presence of DMF solvent as a template allows for the creation of the crystal structure conformation. MD simulations suggest the larger DMF does not diffuse laterally throughout the a-axis channel but rather stays localized to its own pore. Conversely, CO_2 is small and allows for the linkers to rotate dynamically accommodating more CO_2 molecules. Lastly, the binding energies and number of molecules involved in each gate-opening was quantified, and shown to possess sufficient energy to overcome the energy penalty involved in rotating the bonds. Once again this project was an example of how simulation was utilized to help explain novel experimental phenomena.

At low pressures, it is proposed that the CO_2 adsorbs on the surface of Mg-4PyC but this pressure is not sufficient to cause a gate opening conformational change beyond the surface of the MOF and thus the bulk is remained closed. Further investigation in this low pressure regime is currently ongoing.

4.3. References

(1) Dolomanov, O. V; Blake, A. J.; Champness, N. R.; Schröder, M. *J. Appl. Crystallogr.* **2003**, *36*, 1283.

(2) Walton, K. S.; Snurr, R. Q. *J. Am. Chem. Soc.* **2007**, *129* (27), 8552.

(3) Jean, Y. C. *Microchem. J.* **1990**, *42* (1), 72.

(4) Ravikovitch, P. I.; Neimark, A. V. *J. Phys. Chem. B* **2001**, *105* (29), 6817.

(5) Farha, O. K.; Yazaydın, A. Ö.; Eryazici, I.; Malliakas, C. D.; Hauser, B. G.; Kanatzidis, M. G.; Nguyen, S. T.; Snurr, R. Q.; Hupp, J. T. *Nat. Chem.* **2010**, *2* (11), 944.

(6) Simmons, J. M.; Wu, H.; Zhou, W.; Yildirim, T. Carbon capture in metal–organic frameworks—a comparative study. *Energy & Environmental Science*, 2011, *4*, 2177.

(7) Segall, M.; Linda, P.; Probert, M.; Pickard, C. *Accelrys San Diego, CA* **2002**.

(8) Hafner, J. Ab-initio simulations of materials using VASP: Density-functional theory and beyond. *Journal of Computational Chemistry*, 2008, *29*, 2044–2078.

(9) Kresse, G.; Hafner, J. *Phys. Rev. B* **1993**, *47* (1), 558.

(10) Kresse, G. *Phys. Rev. B* **1996**, *54* (16), 11169.

(11) Kresse, G.; Hafner, J. *Phys. Rev. B* **1994**, *49* (20), 14251.

(12) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77* (18), 3865.

(13) Blöchl, P. E. *Phys. Rev. B* **1994**, *50* (24), 17953.

(14) Kresse, G. *Phys. Rev. B* **1999**, *59* (3), 1758.

(15) Grimme, S. *J. Comput. Chem.* **2006**, *27* (15), 1787.

(16) Campañá, C.; Mussard, B.; Woo, T. K. *J. Chem. Theory Comput.* **2009**, *5* (10), 2866.

(17) Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. *J. Am. Chem. Soc.* **1992**, *114* (25), 10024.

(18) Belof, J. L.; Stern, A. C.; Space, B. *J. Chem. Theory Comput.* **2008**, *4* (8), 1332.

(19) Forrest, K. A.; Pham, T.; Hogan, A.; McLaughlin, K.; Tudor, B.; Nugent, P.; Burd, S. D.; Mullen, A.; Cioce, C. R.; Wojtas, L.; Zaworotko, M. J.; Space, B. *J. Phys. Chem. C* **2013**, *117* (34), 17687.

(20) García-Sánchez, A.; Ania, C. O.; Parra, J. B.; Dubbeldam, D.; Vlugt, T. J. H.; Krishna, R.; Calero, S. *J. Phys. Chem. C* **2009**, *113* (20), 8814.

(21) Harris, J. G.; Yung, K. H. *J. Phys. Chem.* **1995**, *99* (31), 12021.

(22) Stryjek, R.; Vera, J. H. *Can. J. Chem. Eng.* **1986**, *64* (2), 323.

(23) Vaidhyanathan, R.; Iremonger, S. S.; Shimizu, G. K. H.; Boyd, P. G.; Alavi, S.; Woo, T. K. *Science* **2010**, *330* (6004), 650.

(24) Vuong, T.; Monson, P. A. *Langmuir* **1996**, *12* (22), 5425.

(25) Vaidhyanathan, R.; Iremonger, S. S.; Shimizu, G. K. H.; Boyd, P. G.; Alavi, S.; Woo, T. K. *Science* **2010**, *330* (6004), 650.

(26) Momma, K.; Izumi, F. *J. Appl. Crystallogr.* **2008**, *41* (3), 653.

(27) Vaidhyanathan, R.; Iremonger, S. S.; Shimizu, G. K. H.; Boyd, P. G.; Alavi, S.; Woo, T. K. *Angew. Chemie - Int. Ed.* **2012**, *51* (8), 1826.

(28) Herm, Z. R.; Swisher, J. A.; Smit, B.; Krishna, R.; Long, J. R. *J. Am. Chem. Soc.* **2011**, *133* (15), 5664.

(29) Herm, Z. R.; Swisher, J. A.; Smit, B.; Krishna, R.; Long, J. R. *J. Am. Chem. Soc.* **2011**, *133* (15), 5664.

(30) Willems, T. F.; Rycroft, C. H.; Kazi, M.; Meza, J. C.; Haranczyk, M. *Microporous Mesoporous Mater.* **2012**, *149* (1), 134.

(31) Martin, R. L.; Smit, B.; Haranczyk, M. *J. Chem. Inf. Model.* **2012**, *52* (2), 308.

(32) Forrester, T. R.; Smith, W. DL_POLY ver. 2.18 (C.L.R.C. Daresbury Laboratory, Daresbury, UK, 1995).

(33) Zhang, L.; Wu, G.; Jiang, J. *J. Phys. Chem. C* **2014**, *118* (17), 8788.

(34) Salles, F.; Jobic, H.; Ghoufl, A.; Llewellyn, P. L.; Serre, C.; Bourrelly, S.; Férey, G.; Maurin, G. *Angew. Chemie - Int. Ed.* **2009**, *48* (44), 8335.

(35) Zhao, Z.; Li, Z.; Lin, Y. S. *Ind. Eng. Chem. Res.* **2009**, *48* (22), 10015.

(36) Saha, D.; Bao, Z.; Jia, F.; Deng, S. *Environ. Sci. Technol.* **2010**, *44* (5), 1820.

(37) Silva, J. A. C.; Schumann, K.; Rodrigues, A. E. Sorption and kinetics of CO2 and CH4 in binderless beads of 13X zeolite. *Microporous and Mesoporous Materials*, 2012, *158*, 219–228.

(38) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graph.* **1996**, *14* (1), 33.

(39) Schneemann, a; Bon, V.; Schwedler, I.; Senkovska, I.; Kaskel, S.; Fischer, R. a. *Chem. Soc. Rev.* **2014**, 6062.

(40) Achmann, S.; Hagen, G.; Kita, J.; Malkowsky, I. M.; Kiener, C.; Moos, R. *Sensors (Basel).* **2009**, *9* (3), 1574.

(41) Nijem, N.; Wu, H.; Canepa, P.; Marti, A.; Balkus, K. J.; Thonhauser, T.; Li, J.; Chabal, Y. J. *J. Am. Chem. Soc.* **2012**, *134* (37), 15201.

(42) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, flexible, and free. *Journal of Computational Chemistry*, 2005, *26*, 1701–1718.

(43) Ogorodnikov, V. V.; Rogovoi, Y. I. *Powder Metall. Met. Ceram.* **1995**, *33* (5-6), 327.

(44) Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard III, W. A.; Skiff, W. M. *J. Am. Chem. Soc.* **1992**, *114* (25), 10024.

(45) Vasudevan, V.; Mushrif, S. H. *J. Mol. Liq.* **2015**, *206*, 338.

5. Conclusions

This thesis serves as an example where computational simulation was successfully used to compliment and accelerate experimental discovery. Two distinct projects involving two completely different chemical problems were the subject of the previous four chapters. First, was an effort to create a model through machine learning techniques to predict based on a molecule's structure, whether that molecule would exhibit ice recrystallization inhibition activity. The second topic of ultra-microporous MOFs encompassed two projects, Ni-4PyC an ultra-microporous MOF with optimal pre-combustion CO_2 capture features and Mg-4PyC an ultra-microporous MOF with extraordinary CO_2 pressure dependent gate-opening behaviour.

In chapter three, a 3D-QSAR model was created and tested from an experimental database of 124 molecules and then used to predict IRI activity for 24 proposed compounds which were then experimentally synthesized and tested. The model creation involved multiple steps. First, a conformational search was performed on the molecules in order to establish a low energy conformation with which to correlate the activity with. From here single point quantum chemical DFT calculations were performed in order to calculate the van der Waals surface based on the total electron density and electrostatic potential of each compound. Grid independent descriptors were calculated using a maximum auto-correlation and maximum cross correlation correlograms transform of the molecular surface curvature and electrostatic potential. The descriptors were then pruned using a genetic algorithm. Finally, a partial least squares regression fit was performed in order to yield the optimal QSAR model. This model was created with 84 compounds and then tested with 40 compounds from the experimental library.

It was found that the model successfully identified 80% of the IRI active molecules with a precision of 0.8 in the calibration set and 83% of active molecules with a precision of 0.8 in the test set. Once confidence was established with a model that could classify compounds based on IRI activity, the model was tested with 24 hypothetical structures which had yet to be synthesized. Of these 24 molecules, 14 were classified as active and 10 were classified as inactive. From here 11 of the 14 predicted active molecules were created and tested for IRI activity. It was found that the model was able to successfully identify 82% of the active molecules which is remarkably consistent with the test sets. In addition to providing a route to accelerate discovery, the 3D-QSAR model was able to identify potentially the most important structural features important for activity. These interactions are between the aryl group substituents, aryl group, and hydroxyl group on the carbohydrate portion. These computational findings provide evidence which support the hypothesis that an important feature of these molecules is the interaction and balance between the hydrophobic and hydrophilic portions. This project is complete and being prepared for publication.

Chapter four can be split into two sections which describe two different projects. The first was computational analysis of Ni-4PyC. This MOF exhibited exceptional CO_2 saturation capacity at low pressures despite being an ultra-microporous MOF with pore sizes of 3.8 Å and 4.8 Å. In order to explain this behaviour simulation was utilized. Simulated isotherms via GCMC revealed two things, one is that the initial experimental sample was not fully activated, and secondly that Ni-4PyC has exceptionally high CO_2 uptakes at high pressures (10.8 mmol/g at 10 bar). It was because of these findings that our experimental collaborators were able to activate Ni-4PyC further and also pursue its characterization for pre-combustion CO_2 capture by determining its

high pressure uptake capacity (which is an expensive experiment that is normally done for standard characterization). Furthermore, the CO_2 binding sites were determined and the energetics was also examined. It was found that Ni-4PyC has a high density of binding sites which also exhibit cooperative binding between CO_2 guests. It was through simulation that the potential of Ni-4PyC for pre-combustion PSA CO_2 capture was identified and determined to be one of the highest performing materials, and the best material found to date that is also humidity stable (a necessary property for real-world application). Simulation also helped to further provide some explanation to the outstanding uptake characteristics. This work has been submitted for publication to *Science Advances* and has been approved pending revisions which are currently underway.

The second portion of chapter four concerns Mg-4PyC, an ultra-microporous MOF which exhibited CO_2 guest dependent dynamic gate-opening behaviour. In this project, computational simulation was used to first explore the conformational space of the different linker configurations. Then, an open conformation and partially open conformation was identified where the simulated isotherm matched well with experiment. Next, molecular dynamics simulations were used to investigate the dynamic ring rotation. A molecular mechanics potential for dihedral angle rotation was fit to a DFT potential surface determined from a rotational scan of dihedral angles. This modified potential was tested and was able to recreate the crystal structure with the presence of DMF molecules. An analysis of the dihedral angle distribution revealed that a conformational change was occurring with respect to the CO_2 loading. GCMC simulations from MD calculated structures showed that there was indeed a gradual change in conformation from closed to open states with respect to CO_2 loading. Binding

sites and energetics were calculated and it was found that the CO_2-framework interactions offset the energetic cost of linker rotation to the open conformation. It was determined that behaviour from 0.15 bar to 1 bar can be explained by simulation. However, simulations at the lowest pressure regime are inconclusive and an experimental crystal structure has yet to be determined likely due to disorder in the crystal. The manuscript for this project is currently in preparation.

5.1. Future Work

Currently of the three projects described in this thesis, Ni-4PyC has been submitted and tentatively accepted for publication to *Science Advances*, IRI 3D-QSAR is being prepared for submission, and Mg-4PyC is in manuscript development. While these projects have been successful, there is always room for further work. The next few sections will describe areas in these research projects which can be further studied or where there is potential for future discovery.

5.1.1. Ice Recrystallization Inhibition

Perhaps the biggest criticism of this project is also its surprising strength – the use of gas phase modelled molecules to describe aqueous dynamic activity. The 3D-QSAR model worked surprisingly well for molecules which were optimized in gas phase with no explicit or implicit solvent involved. Indeed, as mentioned previously, the conformation which is the lowest energy in the gas phase is not necessarily the "active" or most optimal conformation for IRI activity. Furthermore, as more experimental data is made available the 3D-QSAR model can be further refined to include more structures. Finally, simulation of the molecules interacting at the

barrier between the ice and quasi-liquid layer would be essential in discovering exactly how these molecules inhibit ice recrystallization. Thus, future work can be centered on the following projects.

- **Solvated Structure Models**: Performing a conformational search on the library of experimental molecules within a solvated system would help recreate the environment of activity much better. In order to do this a molecular dynamics simulation in explicit solvent or a GCMC search in implicit solvent could be used to determine the most likely structures. 3D-QSAR models have been built from solvated small molecule models before. For example, Lill et al. used an explicit solvation combined with molecular docking to develop a 3D-QSAR model to predict small molecule binding to inhibit Cytochrome P450, a protein responsible for undesired drug-drug interactions.[1]
- **Improvement of QSAR Model**: The QSAR model is only as effective as the data and structures with which it is developed from. As more compounds are synthesized and tested for IRI activity, the QSAR model can be expanded to include other types of molecules. Indeed, the model made here was very effective at phenyl-alditols especially since much of the training set had this form. However, other variations hydrophobic and hydrophilic functional groups can be combined which would require the QSAR model to be parameterized for these groups. Another route would be to make different QSAR models for different "families" of molecules. Indeed, another "family" of proposed compounds that could be incorporated into the existing model or serve new 3D-QSAR model are those of phenyl-pyranose. These are closed ring phenyl-substituted

carbohydrates based off of pyranose as the hydrophilic portion. The basic structure of one of these molecules is shown below in Figure 5.1.

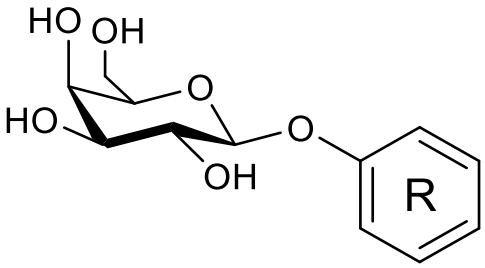


Figure 5.1: General structure of the phenyl-pyranose family of molecules. In addition to the variation due to the substitution of the phenyl ring there can be stereochemical variation in terms of axial or equatorial hydroxyl groups on the pyranose ring.

- **In Depth Simulation of Ice Recrystallization Inhibitors:** Many burning questions on the mechanics of ice recrystallization inhibition remain. Molecular dynamics simulations could be utilized to simulate a system of IRI molecules at the interface of solid ice crystal and liquid water. The best experimentally measured IRI molecules would be used in the simulation in order to elucidate the dynamic interaction of this molecule with the ice crystal surface. There have been multiple studies in literature of molecular dynamics simulations on antifreeze proteins,[2–5] but currently there have been no molecular dynamics simulations on small molecule ice crystallization inhibition.

Recently, the Ben Lab were able to quantitatively measure the rate constant of the ice recrystallization process with the presence of their small molecule inhibitor molecules.[6] In fact, they were able to determine that the increase in ice crystal grain size with respect to time can be fit to a first-order reaction model. Importantly, by binning the ice crystal size as a function of time they were able to establish a new metric from dose-response curves called IC_{50} which is the concentration of inhibitor that gives 50% antagonism. Essentially, this gives a quantitative

concentration of the IRI molecules which directly correlates to the IRI activity. The IC_{50} values from this method could serve as a better metric to measure IRI activity versus the mean grain size of the ice crystals. Thus, a promising avenue to explore is to develop QSAR models with IC_{50} values in hopes of creating a model capable of predicting explicit activity rather than a binary classifier.

5.1.2. Ni-4PyC

Ni-4PyC (**1**) represented a very interesting case of a high performance MOF with excellent CO_2/H_2 selectivity, extraordinary CO_2 saturation uptake, and optimal working conditions for pre-combustion CO_2 capture. The discovery of this material has prompted our experimental collaborators to try and find different phases of Ni-4PyC.

A new phase of Ni-4PyC (**2**) was very recently isolated and characterized by the Vaidhyanathan group. In this case dimethylformamide (DMF) was used as the solvent rather than tetrahydrofuran (THF) for phase **1**. Here the choice of solvent being used has a high degree of control on what the final structure shall be. Indeed, DMF crystallizes within the pores of Ni-4PyC-**2**. This specific phase only has one dimensional pores along the a-axis resulting in channels running down the a-axis. This topology is similar to that of Mg-4PyC. **2** has a uniform pore size distribution of 5.4 Å which is slightly larger compared to **1**'s pore sizes of 3.8 Å and 4.8 Å. The crystal structure along the a-axis of **2** is shown below in Figure 5.2

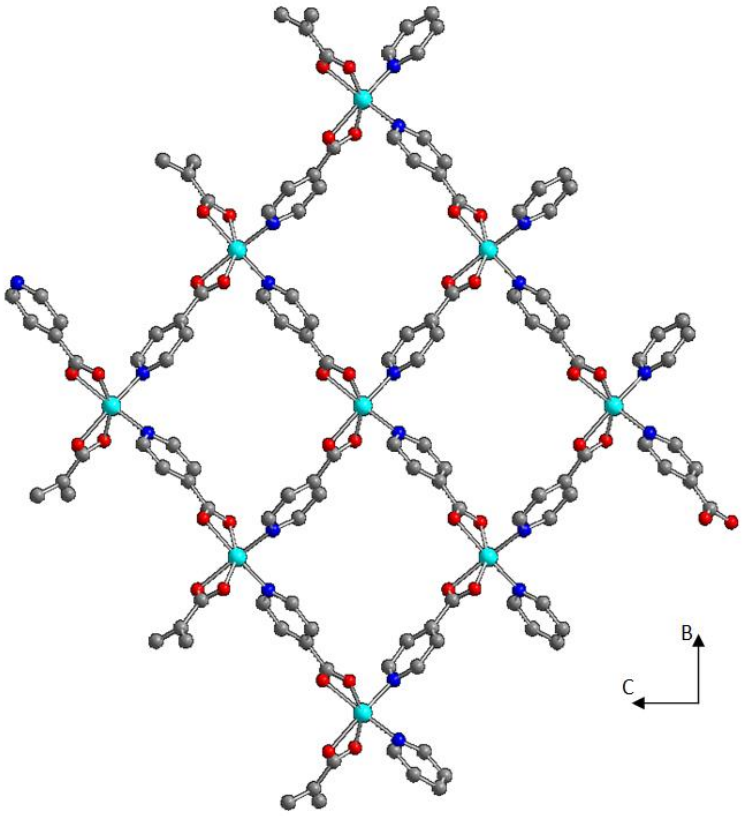


Figure 5.2: Crystal structure of Ni-4PyC-2 down the a-axis clearly showing the 1-dimensional pore. Light blue is Ni, dark blue is N, grey is C, and red is O.

CO_2 adsorption experiments were performed on **2** and an isotherm showed significantly higher capacity at room temperature of 4.2 mmol/g compared to that of **1** with a room temperature capacity of 3.6 mmol/g at 1 bar. Interestingly, **2** has a relatively lower 195K uptake (7.1 mmol/g vs. 11 mmol/g) compared to **1**. Thus a slight increase in pore size allows for a significant increase in CO_2 capacity at room temperature but the lack of a two or three dimensionally connected pore system results in lower saturation capacity. One would expect that a large pore size would result in greater saturation capacity. It is suspected that the simple 1D channel system does not lend itself to a high density of binding sites which made **1** such a high performing MOF. There exists an opportunity to probe this MOF further and analyze its

adsorption characteristics in a similar fashion as what was done with **1**. The first step would be to find an optimized crystal structure which also produced simulated isotherms which would match to experiment. Next, binding site analysis could be done to get a better understanding of the energetic binding of CO_2 to this MOF in comparison to **1**. Finally, selectivity against other gases could be calculated using binary GCMC simulations in order to analyze the feasibility of **2** for various gas separations.

5.1.3. Mg-4PyC

With regards to Mg-4PyC, the greatest impact can be made in accurately simulating the isotherm of a flexible MOF. Unfortunately, our simulations were unable to provide a sure-fire explanation for the low-pressure regime of 0 – 0.15 bar. This may be due to the current limitations of our computation simulation methods. The greatest approximation employed during the simulation of the isotherm is that the MOF is held rigid and its positions are fixed in space. This approximation works well for rigid non-flexible MOFs which are the majority of cases. However, Mg-4PyC is a very unique case and the first MOF to display gate-opening behaviour via linker rotation as a response to CO_2 pressure. As such a more accurate representation of this system is necessary.

One avenue which can be taken is to incorporate the linker movement as a step in the GCMC simulation. That is, in addition to random insertions, deletions, and movements of the guest, incremental rotation of the linker by a set torsional angle can also be included. One caveat is that this additional Monte Carlo move may require that the simulation be run for

more number of steps in order for convergence to reach. However, this will allow for the simulation to sample the conformational movement of the linker ring.

Another avenue which can be taken is to sample a predetermined set of low-energy geometric conformations as calculated using DFT. In this case, rather than allowing for only the rotation of a linker, which would sample even high energy states, we would sample pre-determined conformations. These predetermined conformations would be a sequence of geometries which provides a low-energy pathway from the closed structure partially open and finally to fully open. This can be done using the nudged elastic band method which is an optimization procedure for finding minimum energy paths.[7] This sequence of geometric configurations will have a specific calculated energy from the DFT level of theory for each specific configuration. These energies will be used to calculate the acceptance criteria ($e^{\Delta E/kt}$) where the energy of the system will now be equal to the guest-host energy as well as the energy of the MOF framework. The greatest approximation to using this scheme is that we assume the conformations obtained from a low-energy pathway are the most likely conformations to occur. However, since CO_2 does not provide strong polarization effects which would influence and alter the energy of the MOF, it is reasonable that a low-energy pathway of closed to open conformations would be likely to occur experimentally.

5.2. References

(1) Lill, M. A.; Dobler, M.; Vedani, A. *ChemMedChem* **2006**, *1* (1), 73.

(2) Nutt, D. R.; Smith, J. C. *J. Am. Chem. Soc.* **2008**, *130* (39), 13066.

(3) Kundu, S.; Roy, D. *J. Mol. Graph. Model.* **2008**, *27* (1), 88.

(4) Kundu, S.; Roy, D. *J. Mol. Graph. Model.* **2012**, *38*, 211.

(5) Nada, H.; Furukawa, Y. *J. Phys. Chem. B* **2008**, *112* (23), 7111.

(6) Abraham, S.; Keillor, K.; Capicciotti, C. J.; Perley-Robertson, G. E.; Keillor, J. W.; Ben, R. N., *Crystal Growth and Design* **2015**, accepted.

(7) Henkelman, G.; Uberuaga, B. P.; Jónsson, H. *J. Chem. Phys.* **2000**, *113* (22), 9901.