

Computational High Throughput Screening of Metal Organic Frameworks for Carbon Dioxide Capture and Storage Applications

Peter G. Boyd

Thesis submitted to the
Faculty of Graduate and Postdoctoral Studies
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Chemistry

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

Abstract

This work explores the use of computational methods to aid in the design of Metal Organic Frameworks (MOFs) for use as CO₂ scrubbers in carbon capture and storage applications. One of the main challenges in this field is in identifying important MOF design characteristics which optimize the complex interactions governing surface adsorption. We approach this in a high-throughput manner, determining properties important to CO₂ adsorption from generating and sampling a large materials search space.

The utilization of MOFs as potential carbon scrubbing agents is a recent phenomenon, as such, many of the computational tools necessary to perform high-throughput screening of MOFs and subsequent analysis are either underdeveloped or non-existent. A large portion of this work therefore involved the development of novel tools designed specifically for this task. The chapters in this work are contiguous with the goal of designing MOFs for CO₂ capture, and somewhat chronological in order and complexity, meaning as time and expertise progressed, more advanced tools were developed and utilized for the purposes of computational MOF discovery.

Initial work towards MOF design involved the detailed analysis of two experimental structures; CALF-15 and CALF-16 using classical molecular dynamics, grand canonical Monte Carlo simulations, and DFT to determine the structural features which promote CO₂ adsorption. An unprecedented level of agreement was found between theory and experiment, as we are able to capture, with simulation, the X-ray resolved binding sites of CO₂ in the confined pores of CALF-15. Molecular simulation was then used to provide a detailed breakdown of the energy contributions from nearby functional groups in both CALF-15 and CALF-16.

A large database of hypothetical MOF structures is constructed for the purposes of screening for CO₂ adsorption. The database contains 1.3 million hypothetical structures, generated with an algorithm which snaps together rigid molecular building blocks extracted from existing MOF crystal structures. The algorithm for constructing the hypothetical MOFs and the building blocks themselves were all developed in-house to form the resulting database. The topological, chemical, and physical features of these MOFs are compared to recently developed materials databases to demonstrate the larger structural and chemical space sampled by our database.

In order to rapidly and accurately describe the electrostatic interactions of CO₂ in the hypothetical database of MOFs, parameters were developed for use with the charge equilibration method. This method assigns partial charges on the framework atoms based on a set of parameters assigned to each atom type. An evolutionary algorithm was used to optimize the charge equilibration parameters on a set of 543 hypothetical MOFs such that the partial charges generated would reproduce each MOFs DFT-derived electrostatic potential. Validation of these parameters was performed by comparing the CO₂ adsorption from the charge equilibration method vs DFT-derived charges on a separate set of 693 MOFs. Our parameter set were found to reproduce DFT-derived CO₂ adsorption extremely well using only a fraction of the time, making this method ideal for rapid and accurate high-throughput MOF screening.

A database of 325,000 MOFs was then screened for CO₂ capture and storage applications. From this study we identify important binding pockets for CO₂ in MOFs using a binding site analysis tool. This tool uses a pattern recognition method to compare the 3-D configurations of thousands of pore structures surrounding strong CO₂ adsorption sites, and present common features found amongst them.

For the purposes of developing larger databases which sample a more diverse materials space, a novel MOF construction tool is developed which builds MOFs based on abstract graphs. The graph theoretical foundations of this method are discussed and several examples of MOF construction are presented to demonstrate its use. Notably, not only can it build existing MOFs with complicated geometries, but it can sample a wide range of unique structures not yet discovered by experimental means.

Contents

Table of Figures	x
Table of Tables	xiv
List of Acronyms	xv
Acknowledgements.....	xvi
1 Introduction, Motivation and Background Information.....	1
1.1 Introduction	1
1.1.1 Carbon Capture	1
1.1.2 Metal-organic Frameworks.....	6
1.1.3 Evaluating MOFs for CCS applications	9
1.1.4 Computational Rational Design.....	11
1.2 Overview and Goals	13
1.3 Summary of Chapters.....	14
1.4 Background Information	17
1.4.1 Adsorption Isotherms and CO ₂ Working Capacity.....	17
1.4.2 CO ₂ /N ₂ Selectivity	20
1.4.3 Isosteric Heat of Adsorption, q_{st}	21
1.4.4 Other Flue Gases.....	22
1.4.5 Stability and recyclability	23
1.4.6 Kinetic factors of gas adsorption	24
1.4.7 Computer Simulations of MOFs.....	25
1.4.8 Sampling an Ensemble with Monte Carlo Simulations	26
1.4.9 Computing Isotherms - the Grand Canonical Monte Carlo simulation;.....	28
1.4.10 Energy Calculations with Periodic Boundary Conditions	32

1.4.11	Force Field Parameters A. van der Waals interactions	34
1.4.12	Force Field Parameters B. Partial Atomic Charges	36
1.4.13	The Ewald Summation Method	37
1.5	References	41
2	Modelling CO ₂ adsorption in Experimental MOFs	49
2.1	Abstract	49
2.2	Introduction	50
2.3	Computational Details.....	51
2.4	CALF-15	57
2.5	CALF-16	64
2.6	Conclusions	72
2.7	References	73
3	Construction of a hypothetical database of MOFs with large chemical, physical and topological diversity.....	76
3.1	Abstract	76
3.2	Introduction	76
3.3	Methods.....	79
3.3.1	Structure Generation Basics.....	80
3.3.2	Special Bonding Parameters to Reduce the Combinatorial Search	86
3.3.3	Structural diversity in the database	89
3.3.4	Output of the structure generation program.....	92
3.3.5	Functionalization of the hypothetical MOFs	92
3.3.6	MOF database generation	93
3.4	Simulation Details	100
3.5	Results and Discussion.....	102

3.5.1	Characterization of Unfunctionalized MOFs.....	103
3.5.2	Properties of the functionalized MOF database.....	106
3.6	Conclusions	112
3.7	References	113
4	Ultra-Fast Generation of Electrostatic Potentials in MOFs	117
4.1	Abstract	117
4.2	Introduction	117
4.3	Methods.....	119
4.3.1	Charge Equilibration (QEq) Methods.....	119
4.3.2	Fitting Procedure.....	122
4.3.3	GCMC simulations	123
4.3.4	Selection of MOFs for training and validation	124
4.4	Results	126
4.5	Conclusions	135
4.6	References	136
4.7	Appendix – Derivation of the Charge Equilibration Equations	137
5	Discovery of Adsorbophores	142
5.1	Introduction	142
5.2	Methods.....	143
5.2.1	3-D histograms from GCMC simulations.....	144
5.2.2	Binding sites from 3-D histograms	147
5.2.3	Active site extraction	150
5.3	Substructure matching.....	155
5.3.1	Correspondence graphs of two molecular fragments.....	157
5.3.2	Common Subgraphs from Maximum Clique Detection	159

5.3.3	Recognizing SBUs in MOFs using Maximum Clique Detection	163
5.3.4	Finding Common Adsorbophores by subgraph mining.....	170
5.4	Results	175
5.4.1	Comparison of Top Adsorbophores Across Random Seeds.....	175
5.4.2	Analysis of Discovered Adsorbophores.....	181
5.5	Summary	185
5.6	References	185
6	MOF Design Features for Carbon Capture Technology: a Computational High Throughput Screening Study	188
6.1	Abstract	188
6.2	Introduction	188
6.3	Methods.....	192
6.3.1	Generation of a hypothetical database of MOFs.....	192
6.3.2	Screening protocol	193
6.3.3	Adsorbophore discovery	196
6.4	Computational Details.....	197
6.5	Results	199
6.5.1	CO ₂ screening	199
6.5.2	Strong CO ₂ adsorption sites	205
6.5.3	MOF Pores Lined with Conjugated Aromatic Systems.....	212
6.5.4	Ab initio optimization of the top performers	216
6.6	Conclusions	219
6.7	References	221
6.8	Appendix – High performing MOFs.....	224
7	Construction of Hypothetical MOFs using a Graph Theoretical Approach	227

7.1	Abstract	227
7.2	Introduction	227
7.3	Background	231
7.4	Methods.....	235
7.4.1	Periodic Lattice Representations of Labelled Quotient Graphs.....	236
7.4.2	The Metric Tensor.....	239
7.4.3	Constructing Non-Barycentric Net Embeddings from SBUs.	242
7.4.4	Implementation	249
7.5	Results	251
7.5.1	Fitting Different 4-c SBUs to Nets	251
7.5.2	Structures built from the RCSR	256
7.5.3	Diverse MOFs Built with the tbo Topology	258
7.5.4	Building MOFs with Unstable Nets.....	261
7.5.5	Challenging Topologies and High Surface Area MOFs	262
7.6	Conclusions	269
7.7	References	270
8	Conclusions and Future Directions	273
8.1	Summary	273
8.2	On-going and Future Work	279
8.3	References	283

Table of Figures

Figure 1.1: Graphical portrayal of the carbon capture and storage process.	3
Figure 1.2: Publications in peer-reviewed journals containing the key-word metal-organic frameworks or MOFs	6
Figure 1.3: MOFs come in all different shapes, sizes and chemistry.	8
Figure 1.4: Topological abstraction of MOF-5.	9
Figure 1.5: Generic picture of two adsorption isotherms within a nano-porous material.	19
Figure 1.6: The Metropolis Monte Carlo sampling scheme.	28
Figure 1.7: Representation of sampling the grand canonical ensemble.	30
Figure 1.8: The Lennard-Jones function.	35
Figure 1.9: Graphical depiction of the Ewald summation scheme.	38
Figure 2.1: Representation of the MOF CALF-15.	59
Figure 2.2: Agreement between experimental and simulated CO ₂ adsorption properties of CALF-15	60
Figure 2.3: The carbon-nitrogen radial distribution functions for the NH ₂ -CO ₂ guest-host interactions in CALF-15 extracted from the GCMC simulations.	62
Figure 2.4: Representation of the MOF CALF-16.	65
Figure 2.5: Experimental and simulated adsorption isotherm of CALF-16 for CO ₂ at 273 K.	66
Figure 2.6: Centre-of-mass probability-density plots of CO ₂ molecules in CALF-16.	68
Figure 2.7: Simulated CO ₂ adsorption isotherms for the amine-functionalized MOFs CALF-15 and CALF-16, and their respective calculated methyl-for-amine substituted derivatives.	70
Figure 3.1: Mechanics of generating hypothetical MOFs by snapping SBUs together.	81
Figure 3.2: Schematic of the tree which exhaustively samples every bonding possibility of each SBU.	83
Figure 3.3: Sensitivity of the parameters to the construction of hypothetical MOFs.	86
Figure 3.4: A demonstration of how the program assembles MOFs based on the pto topology.	88
Figure 3.5: Parameterizing new SBUs based on extraction of bond parameters from an experimental MOF.	90
Figure 3.6: Two topologies where the primitive cell supports multiple edge length variations.	91
Figure 3.7: The 28 metal SBUs used to construct the database of hypothetical MOFs.	95
Figure 3.8: The first 64 organic SBUs parameterized to construct the database of hypothetical MOFs.	96

Figure 3.9: Organic SBUs 64 – 128 parameterized to construct the database of hypothetical MOFs.....	97
Figure 3.10: The last 45 organic SBUs 129 – 164 parameterized to build MOFs in the database.	98
Figure 3.11: The 50 functional groups used to decorate the 21,383 unfunctionalized MOFs.	99
Figure 3.12: Pie chart of the symmetry found in the 23,935 unfunctionalized.....	104
Figure 3.13: Comparison of the topologies used to generate hMOFs in this work with another computationally generated MOF database.	105
Figure 3.14: Comparison of physical property distributions for the sets of hypothetical structure databases found in the literature.	108
Figure 3.15: Hydrophilic surface areas of the 1.3 million hypothetical MOFs.	110
Figure 3.16: Property distributions of the 1.3 million functionalized MOFs based on topology.....	112
Figure 4.1: Description of the hypothetical MOFs used for the training and validation sets.	128
Figure 4.2: Agreement between the MEPO-QEq charges and DFT-derived charges for the training set..	131
Figure 4.3: Agreement between the MEPO-QEq charges and DFT-derived charges for the validation set.....	134
Figure 5.1: A visual guide to the workflow discussed in this chapter.	144
Figure 5.2: Equitable binning of an oxygen atom in a discretized 3-D grid.	147
Figure 5.3: CO ₂ binding sites from CALF-15 seen from multiple perspectives.	150
Figure 5.4: Justification for extracting MOF atoms 5 Å from the CO ₂ binding site.....	153
Figure 5.5: Extracting an active site from a CO ₂ binding site in a MOF.....	154
Figure 5.6: Concept of a clique.	156
Figure 5.7: The correspondence between two molecular fragments.	159
Figure 5.8: The maximum clique from a correspondence graph yields the largest common substructure of two molecules.	160
Figure 5.9: A demonstration of graph coloring on two graphs.	162
Figure 5.10: Difficulty in molecule recognition in bulk systems.....	166
Figure 5.11: Effectiveness of the maximum clique detection method in finding all the SBUs and functional groups inside the test set of hypothetical MOFs.	170
Figure 5.12: Average number of adsorbophores found from 30 random seeds as a function of the total number of binding sites.....	176
Figure 5.13: Tanimoto similarity of the adsorbophores from 9558 binding sites.....	179

Figure 5.14: Plot of similarity as a function of the program termination parameter for 9558 active sites.....	180
Figure 5.15: A selection of the most representative adsorbophores found from visual inspection of the top 50 most frequent adsorbophores found from the random pairing method.	184
Figure 6.1: Color map of the 324,474 MOFs screened for working capacity and CO ₂ /N ₂ selectivity with GCMC simulations.....	201
Figure 6.2: The distribution of void volumes and surface areas of the low performing MOFs in the range of 0-0.5 mmol/g CO ₂ working capacity and 0-20 CO ₂ /N ₂ selectivity.	202
Figure 6.3: Frequency of a selection of functional groups present in the database as a function of performance metrics.	204
Figure 6.4: The difference between CO ₂ adsorption when using DFT-derived charges and the MEPO QEq parameters.	205
Figure 6.5: The top adsorbophores ranked by the number of MOFs which contain them.	207
Figure 6.6: The force field calculated energy contributions of all the CO ₂ binding sites in the three adsorbophores discussed in the text.....	210
Figure 6.7: Charge density difference plot of CO ₂ bound in the A3 adsorbophore.....	212
Figure 6.8: Performance of the 28,292 hypothetical MOFs identified as possible layered sheet-like MOFs.	214
Figure 6.9: Common SBUs used to produce the structural motifs found in high-performing MOFs.	216
Figure 6.10: Difference in CO ₂ adsorption of the top 95 performing MOFs after DFT optimization.	218
Figure 6.11: Some of the top performing MOFs optimized at the DFT level of theory.	219
Figure 6.12: Image reference for the indices and functional group short-hands provided in Table 6.3.....	226
Figure 7.1: An abstraction of the SBUs of HKUST-1 to vertices and edges to produce the net, tbo	229
Figure 7.2: The diamond (dia) net in different representations.	230
Figure 7.3: Labelled quotient graph of the dia and qtz nets.	234
Figure 7.4: A barycentric embedding of the labelled quotient graph of dia and qtz	242
Figure 7.5: An inorganic and organic SBU along with their underlying vertex representations.....	245
Figure 7.6: Augmenting the labelled quotient graph bcu to support the SBU connectivities.....	248
Figure 7.7: The graphical representation of embedding bcu to fit SBU geometries.....	249
Figure 7.8: SBUs used as input into the structure generation program and their underlying vertex representations.	253
Figure 7.9: Specific surface area of 46 MOFs assembled with nets found in the RCSR.	257

Figure 7.10: The generation of three different hypothetical structures based on the same underlying topology, tbo , using three tritopic SBUs of different symmetry.	260
Figure 7.11: Building a hypothetical MOF based on the unstable net $3(3^2,4)2$	262
Figure 7.12: The hypothetical generation of two experimental MOFs - MOF-210 and NU-110.	266

Table of Tables

Table 1.1: Mole fractions of component gases in a typical flue gas stream	10
Table 2.1: Lennard-Jones parameters for framework atoms from the UFF force field and CO ₂	52
Table 2.2: REPEAT charges used for the framework molecules of CALF-15 and its methylated derivative.	52
Table 2.3: REPEAT charges used for the framework molecules of CALF-16 and its methylated derivative.	53
Table 3.1: A description of the metal SBU geometries and the topologies they assemble into.	100
Table 4.1: Standard QEq and MOF ESP optimized QEq electronegativity and hardness parameters.	130
Table 4.2: Training set correlation coefficients for the CO ₂ uptake and HOA computed with various charge schemes when compared with those determined with DFT-derived charges	132
Table 4.3: Validation set correlation coefficients for the CO ₂ Uptake and HOA computed with various charge schemes when compared with those determined with DFT-derived charges.	133
Table 5.1: Description of the hypothetical MOFs used to validate the clique detection program.	168
Table 5.2: number of active sites contained in the adsorbophores A1 – A3 as a function of MCT	183
Table 6.1: Full list of SBUs and topologies used to build the structures in this study.	193
Table 6.2: Data for the adsorbophores shown in Figure 6.5.	209
Table 6.3: List of the post-DFT-optimized top performing MOFs ranked by their working capacities.	224
Table 7.1: Measurements of the fit of three nets to the same organic and inorganic SBUs.	255
Table 7.2: The cell vectors output from the generation program and the transformation matrices used to generate experimental-like cells.	265
Table 7.3: Cell parameters and selected physical properties of two experimental MOFs, NU-110 and MOF-210, along with purely hypothetical MOFs generated with the ntt , toz and ith-d topologies.	268

List of Acronyms

• ABSL:	Automatic Binding Site Localization
• API	Application Program Interface
• BTC	Benzene Tri-Carboxylate
• CCS	Carbon Capture and Storage
• CPMD	Carr Parrinello Molecular Dynamics (software)
• CSD	Cambridge Structural Database
• DFT	Density Functional Theory
• EORP	End of Run Pairings
• EQEq	Extended Charge Equilibration (method)
• ESP	Electrostatic Potential
• GCMC	Grand Canonical Monte Carlo
• GULP	General Utility Lattice Program (software)
• HOA	Heat of Adsorption
• IAST	Ideal Adsorbed Solution Theory
• L-J	Lennard-Jones (function)
• MC	Monte Carlo
• MCT	Maximum Clique Threshold
• MD	Molecular Dynamics
• MEA	Monoethanolamine
• MEPO	MOF Electrostatic Potential Optimized
• MOF:	Metal Organic Framework
• PAW	Projector Augmented Wave
• PSA	Pressure Swing Adsorption
• QEeq	Charge Equilibration
• QM	Quantum Mechanics
• RCSR	Reticular Chemistry Structure Resource
• RDF	Radial Distribution Function
• REPEAT	Repeating Electrostatic Potential Extracted Atomic (charge)
• RMS	Root Mean Squared
• RMSD	Root Mean Squared Deviation
• SBU	Structural (or Secondary) Building Unit
• TI	Tanimoto Index
• TPSA	Temperature/Pressure Swing Adsorption
• TSA	Temperature Swing Adsorption
• UFF	Universal Force Field
• VASP	Vienna Ab initio Simulation Program (software)
• hMOF	hypothetical MOF
• hPPN	hypothetical Porous Polymer Network
• hfMOF	hypothetical functionalized MOF
• 3-c	three-connected
• 4-c	four-connected
• 6-c	six-connected

Acknowledgements

From all that I have learned over the past nearly 6 years, most pertinent is that the journey of completing a Ph.D. in the natural sciences cannot be done alone. This thesis was possible because of the support of my supervisor, Tom Woo. When I first started this work, he had just begun his research on MOFs, his creative and ground-breaking research ideas provided me with a playground of unique, inspiring and stimulating projects to tackle. Tom strove to keep his group close outside of the research environment, organizing many events which I will look back on with nostalgia. In particular, I have many fond, albeit selective, memories of beating him at volleyball and badminton.

My sincere gratitude to the Woo lab members past and present who provided such a welcoming environment as well as thought provoking discussions. The ‘classics’; Professor Chris Rowley, Dr. James Hooper, Nick Trefiak, Mike Nohra, Arif ‘100’ Ismail, Bianca Provost. The post-docs; Drs. Saman Alavi, Carlos ‘Scarface’ Campaña, Tom ‘fapmaster’ Daff, Michael Fernandez, and Eugene Kadantsev. The ‘currents’; Mohammad Zein-Aghaji, Sean Collins, Phil Deluna, Jason Lo, Tom Burns, and Hana Dureckova. To the honorary member – the infamous Tommy Wiseau – no thought provoking discussions, but many laughs (mostly at your expense).

For more time than I care to admit, I have been as helpless as an infant. My mind in deep concentration, activities like eating, sleeping, and leisure have been governed by those close to me. The following people I owe a deep debt I can never repay; my dad and mom John and Julie, also Joe, Sio and Rick. I could not have done this work without your support. My brother Henry, sister-in-law Nicole and niece Sadie for many happy times. Sandy for your friendship spanning more years than I can count on my fingers and toes, and the memories of camping on an island of ill-repute. To my wife, Rebecca, you have helped me more than you will ever know. I only hope I can return the favour in the many happy years to come. To Aaron and Liam, thank you for

brightening my life, and reminding me that my sense of humour has not evolved past that of a two-year-old's.

1 Introduction, Motivation and Background Information

1.1 Introduction

1.1.1 Carbon Capture

The disturbing causal relationship between the increase in global temperatures and atmospheric CO₂ concentrations has become a major source of concern in the increasingly globalized world. Maintaining status quo for the next 100 years will see surface temperatures rising 2°C,¹ sea levels rising by 60 cm,² and mass extinction of aquatic life.³ This widespread threat to the Earth's ecology is arguably the biggest problem facing our generation and is the motivation behind many strategies to reduce global CO₂ production.

One of the major offenders in terms of CO₂ production are stationary energy production sources. Coal power plants represent roughly 44% of the total 31.2 Gt/year anthropogenic CO₂ emissions.⁴ Due to the availability and abundance of coal, studies project that coal will be the primary electricity source for the next 100 years,¹ failing the discovery of some cheap, abundant, environmentally sustainable energy source. At the same time, the rate of energy consumption is projected to increase exponentially, which for coal is projected to increase to 21.9 trillion kW·h in 2035 from the 2007 value of 11.8 trillion kW·h.⁵ Coal-fired power plants therefore represent a major target for CO₂ gas mitigation strategies since focused reductions in this industry could dramatically reduce global CO₂ emissions in the short to medium term. Carbon Capture and Storage (CCS) is the umbrella term for several strategies proposed to mitigate CO₂ production from point sources. The objective of CCS is to trap CO₂ on site and permanently store it in relatively pure form deep underground. It is estimated that 70% of the overall cost of CCS, including the costs associated with pressurization and underground storage of CO₂, are attributed to its capture from the combustion source.⁶ There are three major strategies currently being

investigated for capture of CO₂ from stationary point sources. One such method is to use oxy-fuel combustion.⁷ This involves purifying the air involved in combustion, such that coal is burned in a diluted CO₂/O₂ mixture. The by-products of this combustion are H₂O and CO₂ which are relatively easy to separate in a condenser. A second option is pre-combustion capture.⁸ This involves gasifying coal at high temperatures and pressures prior to combustion, producing H₂, and CO syngas. Syngas then undergoes the water-gas shift reaction to produce a mixture of CO₂/H₂ in concentrations of 15-50% v/v. The high pressures in which these reactions are performed permit an easy separation of H₂ from CO₂, leaving pure, clean burning H₂ for electricity production. Finally, a possible solution is to use post-combustion capture. This involves trapping the 3-15% v/v CO₂ produced from burning coal in air,⁹ while releasing the harmless remaining flue gas components into the atmosphere. Each of these methods involves a gas purification step. However, the conditions of purification are vastly different, each accompanied by its own set of costs and challenges which is the focus of intense research in their respective areas.^{10,11} One of the most promising methods of the three discussed is post-combustion capture because existing power plants can be retro-fit with CO₂ scrubbers downstream from the reactor, whereas the other strategies require whole new power plants to be built. For this reason, post combustion CO₂ capture is in a more advanced stage of development over the other methods, representing a majority of the pilot scale plants operating a CCS technology in the world.¹² Notably, Canada boasts the only full-scale 110 MW power plant retro-fitted with CCS technology, which was brought online in October 2014 near Estevan, Saskatchewan.

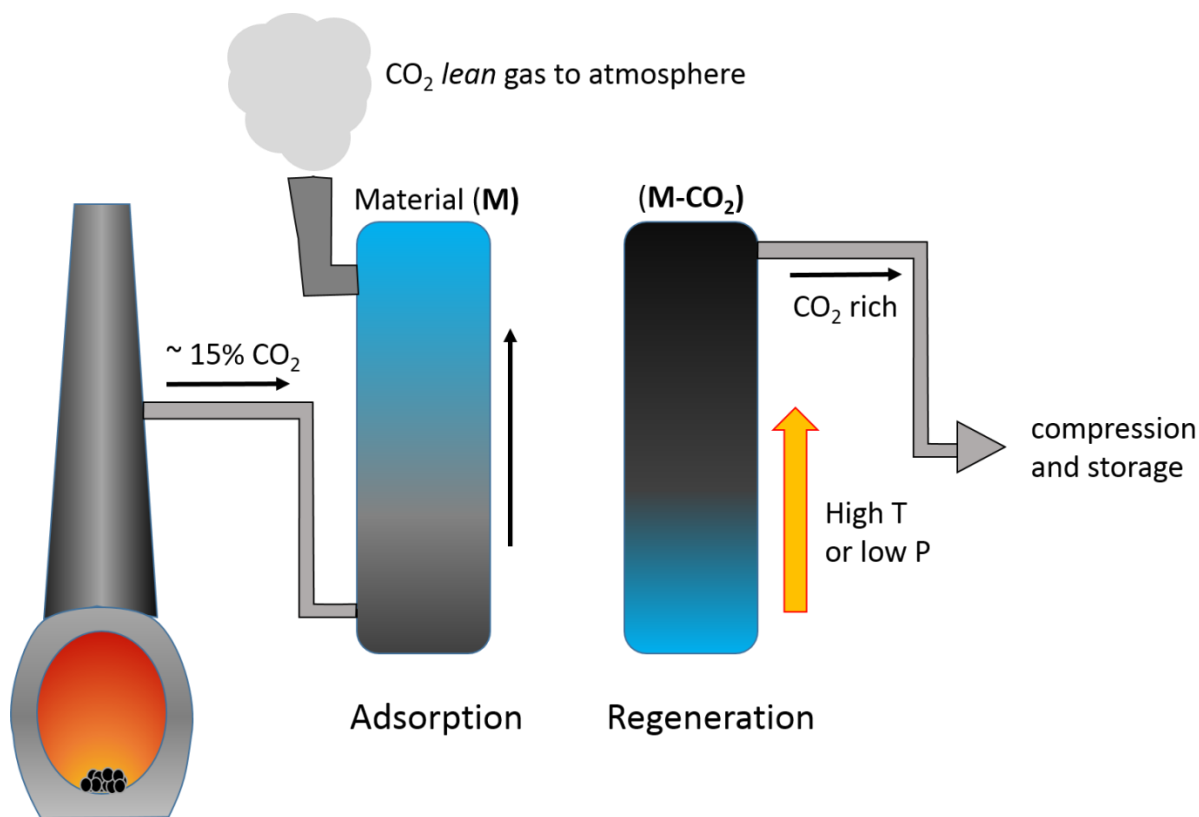


Figure 1.1: Graphical portrayal of the carbon capture and storage process. CO₂ is selectively trapped in a material M, which undergoes regeneration to isolate the pure CO₂ gas and prepare the material for further adsorption/regeneration cycles.

Figure 1.1 shows the general schematic of a post-combustion capture process. The absorber apparatus is composed of a column of sorbent material (**M**), which is subjected to an inlet feed stream of the flue gas mixture. The ideal material will selectively trap CO₂ from the gas stream, allowing the other gases (mostly N₂) to flow through the unit. The sorbent material will selectively capture CO₂ either by forming chemical bonds (absorption) or through physical non-bonded interactions (adsorption). The resulting material (**M-CO₂**) will then undergo a regeneration process through a stripper. The stripper will typically apply high temperatures or low pressures to then recover the sorbent material (**M**) for further adsorption/desorption cycles and extract the purified CO₂ for pressurization and storage. The purity and quantity of CO₂ recovered from this process will ultimately depend on the material used and conditions in which

the gas is trapped and released by the material. It is understood that nearly 70% of the cost associated with CCS can be attributed to the regeneration stage of the material in the absorption apparatus, due to the high levels of energy required to separate CO₂ from the material.¹³

Presently, technologies designed to strip CO₂ gas from flue streams are in their infancy. The reasons are two-fold: i) CO₂ is considered an inert compound, thus there are relatively few industrial applications that can use CO₂ as a feedstock commodity at the rate in which it is produced, and ii) public knowledge and acceptance of CO₂ as a greenhouse gas is a recent phenomenon, so the desire to capture CO₂ in this context is relatively new. The current 'state of the art' technology for CO₂ scrubbing involves using aqueous alkanol amines, which selectively traps CO₂ via chemical absorption.¹⁴ While this process has been widely used in other CO₂ sequestration applications, such methane purification from acid natural gas reservoirs¹⁵ (CO₂ concentrations of up to 90% v/v¹⁶), the high levels of heat required to regenerate the aqueous amine solution and break the N-C carbamate bond is considered too costly for post-combustion capture processes. It is estimated that regeneration of the most commonly used amine, monoethanolamine (MEA), consumes up to 30% of the total energy produced by the plant. In other words, one would have to burn 30% more coal in order to produce the same amount of power generated without scrubbing CO₂. While costs associated with CO₂ capture are expected for any technology, studies estimate the cost of liquid amines to be around \$45-50 USD/ton CO₂ removed^{9,17}, which is beyond the current limit to maintain a profitable energy plant. With government policies and "carbon tax" tariffs in place in certain regions, it is currently more economical to pollute and pay the tax than to capture CO₂.¹⁸ For example B.C. has one of the highest carbon taxes of \$30/tonne CO₂ produced, which is only about half the projected cost of

implementing a post-combustion capture apparatus. Therefore, there is an urgent need to develop a carbon capture technology which can lower costs to encourage wide-spread adoption of CCS.

Due to the significant cost associated with regenerating the material in the CCS scheme, there is opportunity to develop a novel material for use in the capture and release of CO₂. Many of the alternative technologies being investigated are solid, nano-porous materials which can selectively bind CO₂ via physical adsorption. There are three key advantages in using solid sorbents over aqueous amines. First their interaction strength with CO₂ is typically much smaller (20 - 40 kJ/mol vs. 90 kJ/mol for aqueous amines), making it easier to remove CO₂ from these materials in the regeneration stage. Second, the heat capacity of these materials is much lower than aqueous alkanolamines¹⁹ (< 1 vs. $\sim 4 \text{ J}\cdot\text{g}^{-1}\cdot\text{K}^{-1}$) meaning much less energy is wasted heating these materials up during the CO₂ desorption process. Finally, the interaction of CO₂ with these materials is reversible, allowing for different types of desorption processes to be optimized with respect to energy consumption. For example, one could remove CO₂ by lowering the pressure via a pressure swing adsorption (PSA) process,²⁰ or by raising the temperature via temperature swing adsorption (TSA) process, or using a combination of the two via a temperature-pressure swing adsorption (TPSA) process.^{21–23} In these processes, gas molecules are selectively and reversibly adsorbed on the surface of a material, whereupon saturation, a swing in the temperature or pressure to removes the adsorbed gas and regenerates the material. These gas separation processes are relatively mature. For example, they have been employed for decades in purifying methane gas upon extraction from underground reservoirs.²⁴ Several porous materials are being proposed as possible CO₂ scrubbers in these swing systems such as zeolites,²⁵ activated carbons,²⁶ and Metal Organic Frameworks (MOF)s.²⁷ It is said if one were to find a material which could effectively remove 4 mmol CO₂/g sorbent at each regeneration cycle and adsorb

quantities of CO₂ over N₂ at ratios > 150, then one could reduce the cost of CCS to less than \$30 per ton CO₂ removed.¹⁷ Of the materials considered for this application, MOFs show greatest promise. This can be attributed to the large range of properties these materials can exhibit due to their highly tunable nature.

1.1.2 Metal-organic Frameworks

Interest in Metal-organic Frameworks (MOF)s has exploded over the past two decades. To date, this interest has been relegated to specific fields in academic research, as these materials have yet to be implemented in an industrial setting. This is in part due to the fact that this field is relatively young, being first introduced in the early 1990's²⁸ and gaining traction in the 2000's for a wide range of applications, including gas storage,²⁹ separation,³⁰ and heterogeneous catalysis.³¹ While MOFs are the most common

term for these materials, the emergent nature of this field has resulted in many naming conventions. Zeolitic Imizazolate Frameworks (ZIFs),³² Porous Coordination Polymers (PCPs),³³ Porous Coordination Networks (PCNs),³⁴ and Microporous Coordination Polymers (MCPs)³⁵ all describe the same class of material, however for the purposes of

consistency within this work, the term MOF will be used exclusively. It should be additionally noted that this field has produced “offspring” such as covalent organic frameworks (COFs) which is an exciting class of nano-porous crystalline materials composed solely of organic moieties.³⁶ Figure 1.2 demonstrates the growth of the field over the past two decades in terms of

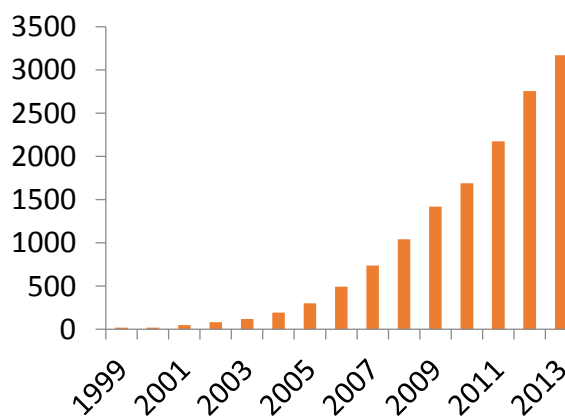


Figure 1.2: Publications in peer-reviewed journals containing the key-word metal-organic frameworks or MOFs

publications in peer reviewed journals. It is worth noting that a number of these publications are considered ‘high-impact’, being published in journals such as *science*^{37–45} and *nature*^{46–54}.

MOFs, as the name suggests, consist of isolated or clustered metal ions coordinated to rigid organic molecules. The organic moieties typically possess multiple coordination sites, allowing for polymeric-type growth in one-, two-, or three-dimensions. These materials generally form high-symmetry crystalline lattices which can be characterized by single crystal x-ray diffraction experiments (see, for example Figure 1.3). Moreover, many of these materials, upon evacuation of solvent, possess pores on the nano-scale. The unique chemistry of these materials, coupled with record breaking internal surface areas for nano-porous materials,^{39,55} provide a large playground for a variety of applications, including CCS.

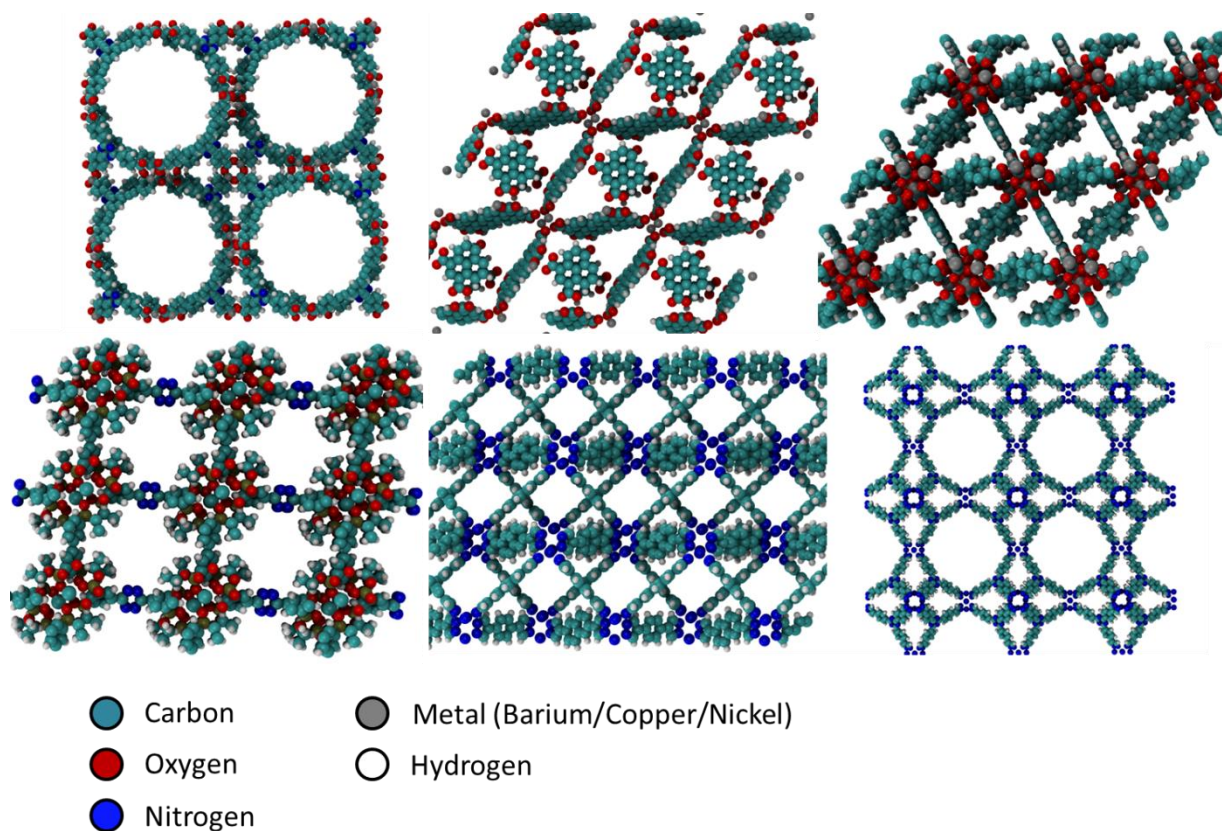


Figure 1.3: MOFs come in all different shapes, sizes and chemistry. Structures of six different MOFs with a variety of different pore sizes and shapes are shown.

For the purposes of classification, a relatively recent trend in MOF chemistry is to break the periodic framework into discrete building blocks called secondary (or structural) building units (SBUs)^{56,57} which can be seen in Figure 1.4a. SBUs consist of clusters of atoms which can easily be abstracted to simple geometric shapes based on their extended connectivity. Upon abstraction of these atomic structures to a series of vertices and edges the underlying 3-dimensional pattern, called a network topology (or net for short) of the material is revealed (Figure 1.4b). Aside from the aforementioned utility in classifying MOFs, nets provide experimental chemists with a guide for synthesizing MOFs with desired pore shapes and connectivity. The nets most common to crystalline materials have been stored in the freely available Reticular Chemistry Structure Resource (RCSR).⁵⁸

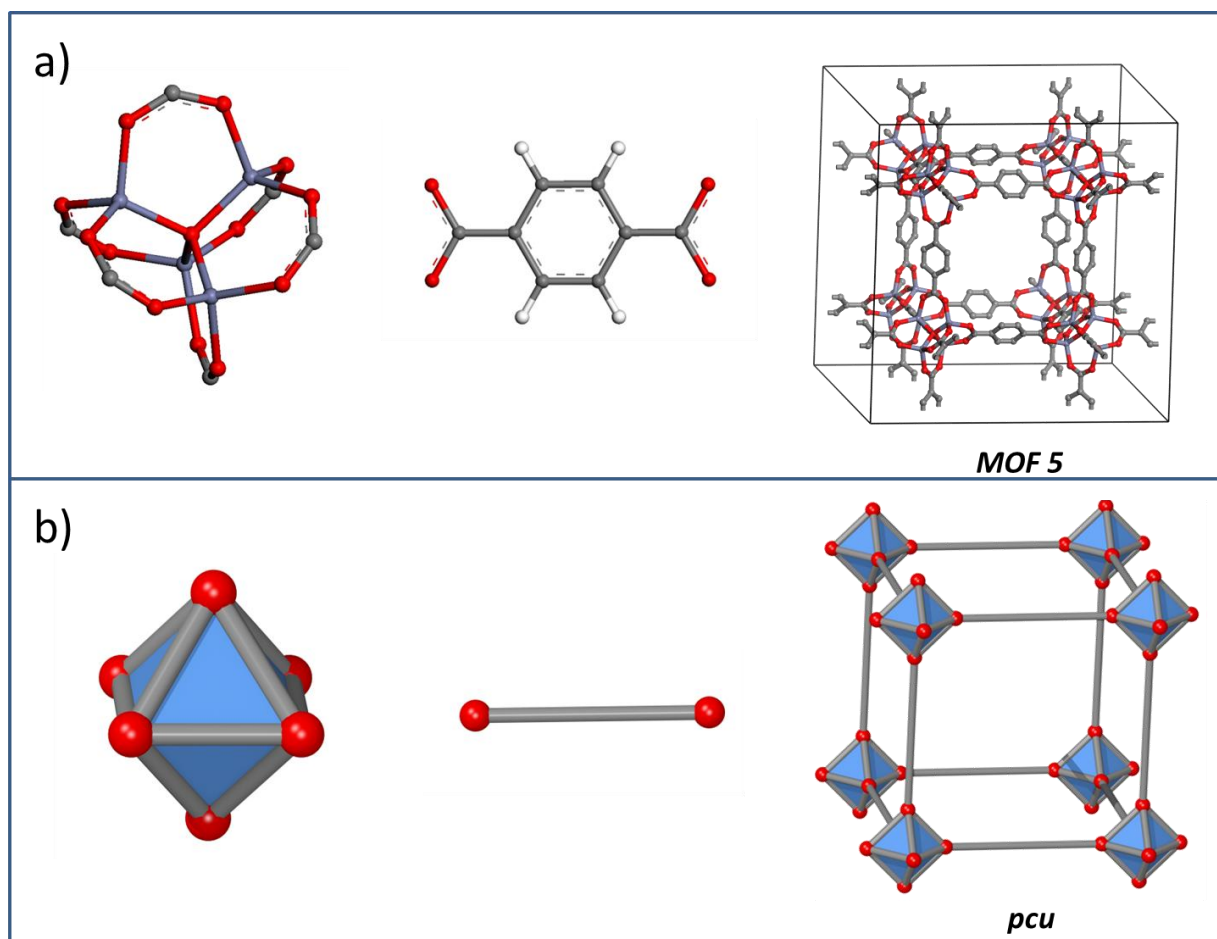


Figure 1.4: a) The SBUs of MOF-5; a Zn_4O metal SBU and a benzene dicarboxylate (BDC) organic SBU. At the far right is a representation of MOF-5 with hydrogens removed for clarity. Purple: zinc, white: hydrogen, red: oxygen, grey: carbon. b) The underlying geometries of the SBUs and their assembly into the primitive cubic (denoted by the 3-letter RCSR code, **pcu**) topology. The Zn_4O SBU is abstracted to an octahedral geometry, and BDC is abstracted to an edge bridging two octahedrons.

1.1.3 Evaluating MOFs for CCS applications

As mentioned in section 1.1.2, MOFs have not yet found an application in an industrial setting. This is despite significant interest in developing a MOF ideal for trapping CO_2 from post-combustion flue gas. In the context of adsorption of CO_2 from flue streams, a promising material must show two general traits; a large amount of CO_2 must be removed from the flue exhaust per unit weight (or volume) of the material. This is called the materials ‘working capacity’, which will ultimately depend on the CO_2 -surface interactions, as well as the thermodynamic conditions of adsorption and desorption. These values are typically measured in

mmol/g, mol/kg, or mmol/cm³ quantifying a ratio of CO₂ to material. Values for working capacity are in the range of 1-2 mmol/g for commercial zeolites⁵⁹ while MOFs which show promise as potential CCS scrubbers are close to 4 mmol/g.⁶⁰ The second criteria pertains to the adsorptive selectivity for CO₂ over other flue gases, particularly N₂ which is the most concentrated flue gas. The preference a material has for CO₂ over N₂ will ultimately determine the purity of the gas which undergoes compression and storage, where the higher the purity of CO₂, the lower the operating cost to the plant. Selectivity is measured as a unitless ratio of the amount of CO₂ adsorbed vs the amount of N₂ adsorbed at flue conditions, normalized by their respective partial pressures. Typical constituents of the flue gas are presented in Table 1.1.

Table 1.1: Mole fractions of component gases in a typical flue gas stream

Component gas	Mole fraction
CO ₂	0.14
N ₂	0.75
O ₂	0.05
H ₂ O	0.07
SO _x	220 ppm
NO _x	< 50 ppm
CO	10 ppm

Until recently, development of MOFs for CO₂ capture has been from the ‘top down’, meaning that MOFs were initially synthesized in the lab, then their CO₂ adsorption performance was evaluated in a controlled experiment.^{61–64} One particular challenge associated with this type of discovery approach is that each study of a newly synthesized MOF is somewhat stand-alone. Whether it is using different thermodynamic conditions for analyzing CO₂ adsorption, employing different techniques for activating the pores of the materials, or even having no CO₂ adsorption data at all, scanning the literature for the best possible MOF for CCS will yield very few structures with which to make a direct comparison. In addition, MOFs are time-consuming and

expensive to make. There are on the order of tens of thousands of nano-porous MOF materials^{57,65} that have been reported in the Cambridge Structural Database (CSD),⁶⁶ where only on the order of hundreds are permanently porous, which is but a fraction of the possible structures that can be assembled. There is likely an ideal configuration of interconnected organic and inorganic molecules which is optimal for CCS, however due to the time and costs associated with MOF synthesis, a more rational design approach is highly desirable.

1.1.4 Computational Rational Design

As eluded to above, reproducibility has been a problem in the MOF field, which makes comparisons challenging. For example, the reported CO₂ zero-loading isosteric heats of adsorption for the MOF HKUST-1⁶⁷ are in the vast range of -15 to -35 kJ/mol, which can be attributed to the differences in solvent removal techniques used to activate the material. In addition, discovering important aspects of the framework structure and chemistry related to CO₂ performance is nearly impossible at the experimental level. This is because all of the measurements used to evaluate CO₂ sorption are macroscopic thermodynamic variables, while the framework interactions with CO₂ are microscopic and difficult to elucidate with current experimental methods. This makes rational design of new materials incredibly difficult, and is the primary reason why a MOF with ideal properties for CCS has not yet been discovered.

In answer to this, computational methods are beginning to emerge as a powerful tool for materials discovery. Several studies demonstrate the power of using computational tools to discover important properties for adsorption. For example, Mellot-Draznieks *et al.* demonstrated the importance of introducing highly polar functional groups in the pores of the MOF MIL-53(Al) by building the hypothetically functionalized structures and computing the DFT binding energies of CO₂ near the functional groups. They were able to identify the energetic contributions of these functional groups and reveal their strong polarization effects on CO₂.⁶⁸ In

another study, the same group demonstrated that -COOH moieties lead to a profound increase in CO_2 heats of adsorption over other functional groups such as -CH_3 , -NH_2 , and -OH .⁶⁹ This was then validated experimentally by Cooper *et al.* who discovered their -COOH functionalized porous polymer network possessed a significant increase in zero loading isosteric heat of adsorption of CO_2 (33 kJ/mol vs 27 kJ/mol unfunctionalized).⁷⁰

While these computational studies demonstrate the power of simulation in directing materials discovery, they are limited to highly focused studies of particular frameworks, the results of which may not provide general enough strategy for intelligent design and synthesis. For example, the previous study demonstrates that -COOH functional groups improve CO_2 heats of adsorption, however this does not improve the overall CO_2 adsorption capacity. In fact, the -COOH groups perform the worst in this respect because their steric bulk fill up the pore space, reducing available adsorption sites for CO_2 . So then the question becomes, how can one improve both isosteric heat of adsorption *and* CO_2 adsorption capacity? To sufficiently answer this question one must sample a larger chemical and physical space encompassing MOFs. This can be accomplished by computationally screening a large series of MOFs in a high-throughput manner.

Several high throughput screening studies have emerged with the motivation of sampling a larger physical and chemical space. For example, Haldoupis *et al.* screened 500 MOFs taken from the CSD⁷¹ for CCS performance. From their screening, they discovered several materials with high potential. However, to the author's knowledge, there has been no follow up experimental studies to validate these claims, additionally, no important structural information was teased out of these materials, thus one could not say what made them high performing. A study by Wilmer *et al.* demonstrated that MOFs with optimal CCS properties required surface

areas of 1500 - 2000 m²/g and pore diameters of 5-6 Å.⁷² While this may provide some structural insight, it does not give much guidance to the experimental chemist. Moreover the screening in this study was performed on a database of over 130,000 hypothetical MOFs constructed with a relatively homogeneous set of underlying network topologies.^{73,74} One could argue that the database of structures is then skewed to a particular distribution, meaning the trends found from their study may not encompass a 'sufficiently diverse' set of data. This is an issue they mention in their work.⁷²

Materials discovery through computational high throughput screening is a relatively recent phenomena. The small breadth of experimental MOFs in existence has prompted researchers to develop algorithms which generate MOF materials *in silico* using discrete molecular building blocks.^{73,75} It is only very recently that innovation in this field has paved the way for the construction of diverse hypothetical MOF databases ideal for materials discovery,⁷⁵ including the work of these authors.

1.2 Overview and Goals

In this work, we develop a materials discovery program to build and computationally screen hundreds of thousands of MOFs for their performance in post-combustion CCS. The goals of this thesis are to identify important structural characteristics which promote CO₂ adsorption in MOFs in order to guide experimental synthesis, and identify 'hit' materials for potential synthesis and commercialization. Due to the emergent nature of this type of study, additional secondary, but essential goals of this work involve the development of the computational tools necessary to perform the screening. Of prominence is the need to develop a large and diverse database of hypothetical MOF materials in which to screen for carbon capture.

To compute adsorption data within these MOFs, we will use well-established statistical mechanical methods for computing thermodynamic properties of these systems. These calculations use classical molecular mechanics force fields, which are a coarse description of the electronic system, to compute interaction energies between the gas particles of CO₂ and the MOF framework atoms. It is vital that the parameters input into these calculations yield results that are as accurate as possible. Without experimental data to compare with, part of this work will involve establishing parameters, specifically for charge interactions, that will accurately reproduce high-level quantum mechanical data.

In summary, this body of work addresses the need for a rational design approach to materials discovery for post-combustion CO₂ capture and storage. The results of this work may enable significant advances in the cost reduction of CO₂ capture technologies, such that widespread adoption of CCS in coal fired power plants can be realized in the near future.

1.3 Summary of Chapters

The contents of this thesis are described briefly herein. In Chapter 2 we study two promising synthesized MOFs exhibiting strong CO₂ adsorption at low pressures, CALF-15 and CALF-16. These materials were functionalized with amine groups with the hypothesis that the strong CO₂ – amine interactions would contribute to strong CO₂ adsorption. Remarkably, we demonstrate an unprecedented level of agreement between experiment and computation by reproducing, not only the experimental adsorption properties, but the specific CO₂ binding sites found in CALF-15 from x-ray resolved data. This permitted a detailed computational analysis of the CO₂ binding, demonstrating the CO₂ – amine interactions can be either stabilizing or destabilizing. This chapter consists of two papers published in the journals *Science*⁴⁵ and *Angewandte Chemie: International Edition*.⁷⁶

In Chapter 3, we develop a novel tool for constructing hypothetical MOFs and use it to construct a database of MOFs. This program utilizes SBUs extracted from experimental crystal structures and ‘snaps’ them together using alignment parameters assigned to each SBUs connection sites. Effort was placed in parameterizing the connection sites of a large number SBUs so that they would effectively assemble a large number of hypothetical materials with a diverse range of topologies and chemistry. We report the construction of the largest database of MOFs currently available; over 1.3 million materials. Their physical and chemical properties are discussed in this chapter.

In anticipation of high-throughput screening hypothetical MOFs for CO₂ capture and storage applications, Chapter 4 highlights our work in developing a scheme to rapidly produce *ab initio* quality electrostatic potentials (ESPs) for each MOF. Here we developed parameters for use in the empirical charge equilibration method (QEq) of Rappé and Goddard.⁷⁷ The method is highly sensitive to differing chemical environments found in MOFs, thus it is important to train the QEq parameters on a diverse set of materials which encompass a broad distribution of MOFs. We used MOF structures developed *in silico* from the previous chapter to train the QEq parameters with, which allowed for an incredible amount of control over the chemical and physical environments used in our training set. This is the first time such a robust set of MOF materials was used to develop QEq parameters, the result of which yielded a fast, accurate method for generating near *ab initio* quality ESPs in MOFs. The resulting parameters were shown to provide high-quality ESPs on a separate set of MOFs both real and hypothetical. This work was published in the *Journal of Physical Chemistry Letters*.⁷⁸

In Chapter 5 we present the development of powerful cheminformatics tools to determine 3-dimensional (3-D) similarities between CO₂ binding sites in MOFs. The chapter first describes

the novel tools developed in-house to go from an empty MOF structure to identifying regions of strong CO₂ adsorption. Following this we present a method, analogous to those used in drug discovery programs to identify active sites in small drug molecules, which extracts CO₂ binding pockets and determines 3-D similarities amongst them. Specifically, we implemented a method based in graph theory, called maximum clique detection, to identify similar 3-D shapes amongst a set of MOF binding pockets. This is the first time such a set of tools has been implemented for a MOF materials discovery, in which one can extract atomic-level detail from screening a large dataset of materials.

Chapter 6 represents the culmination of work developed in Chapters 3, 4, and 5. Here we present the results from screening a set of 324,474 hypothetical MOFs for post-combustion CCS. From this study we identified a set of materials which demonstrate exceptional performance in selectively capturing CO₂. Using the 3-D similarity program discussed in Chapter 5, we discovered a common binding motif in these MOFs; 1-dimensional channels lined with poly-aromatic carbon sheets spaced 7 Å apart. These channels are just large enough to support a monolayer of CO₂ atoms, which are encapsulated with strong dispersion interactions. The results from this chapter describe, for the first time, detailed chemical guidelines for the design of high performing MOFs for CCS.

In Chapter 7 a novel method to build hypothetical MOFs from graph theoretical principles is presented. This proves to be much more robust in its application compared to the method developed in Chapter 3, allowing for the construction of MOFs with more diverse topologies while significantly reducing the amount of user intervention. For example, the hypothetical MOF database containing 130,000 structures developed by Wilmer et al.⁷³ was built with only 6 different network topologies. Using our new graph theoretical approach for

generating structures, we have quickly constructed a database containing over 1000 network topologies.

Finally, Chapter 8 will conclude the work and discuss future directions for this study. The remainder of Chapter 1 serves as a general methodology section, where a detailed description of the metrics used to evaluate MOFs for CCS and the basic computational tools used to model these metrics are presented.

1.4 Background Information

The remainder of this Chapter serves as a background to some of the performance metrics commonly used to evaluate MOFs for CCS, and introduces the basics of computing these metrics with molecular simulation.

1.4.1 Adsorption Isotherms and CO₂ Working Capacity

When evaluating a solid sorbent for PSA/TSA gas separation for CCS viability, there are three gas adsorption properties that are of primary importance that affect the energy efficiency of the process. These are i) CO₂ uptake capacities at the adsorption and desorption conditions, ii) selectivity for CO₂ over N₂, and iii) CO₂ isosteric heat of adsorption. These will be discussed below along with some other factors which need to be addressed, such as MOF stability and recyclability.

Adsorption isotherms are perhaps one of the most widely used techniques for evaluating a MOFs' adsorption properties. These experiments measure the equilibrium amount of a gas adsorbed by the material as a function of the partial pressure of the gas species at a fixed temperature. Figure 1.5 shows two isotherms, I₁ and I₂, where the partial pressure of the gas is plotted on the x-axis, and the quantity of gas adsorbed in the material is plotted on the y-axis. Adsorption quantities are typically in units of mmol CO₂ per gram of material (see y axis of

Figure 1.5) but they can be reported as a volume (cm^3/g) or weight percent (g/g) measurement. The shape of an isotherm reveals important information regarding gas-solid interactions. Materials which show a sharp increase in adsorption of the gas at low partial pressures, such as I_1 in Figure 1.5, exhibit strongly favourable interactions with the gas particles while isotherms resembling I_2 , or those with relatively flat curves, demonstrate a weaker interaction. In addition, other information regarding the materials physical properties, such as internal surface area or pore volume, can be computed from this data using various theoretical models.^{79,80}

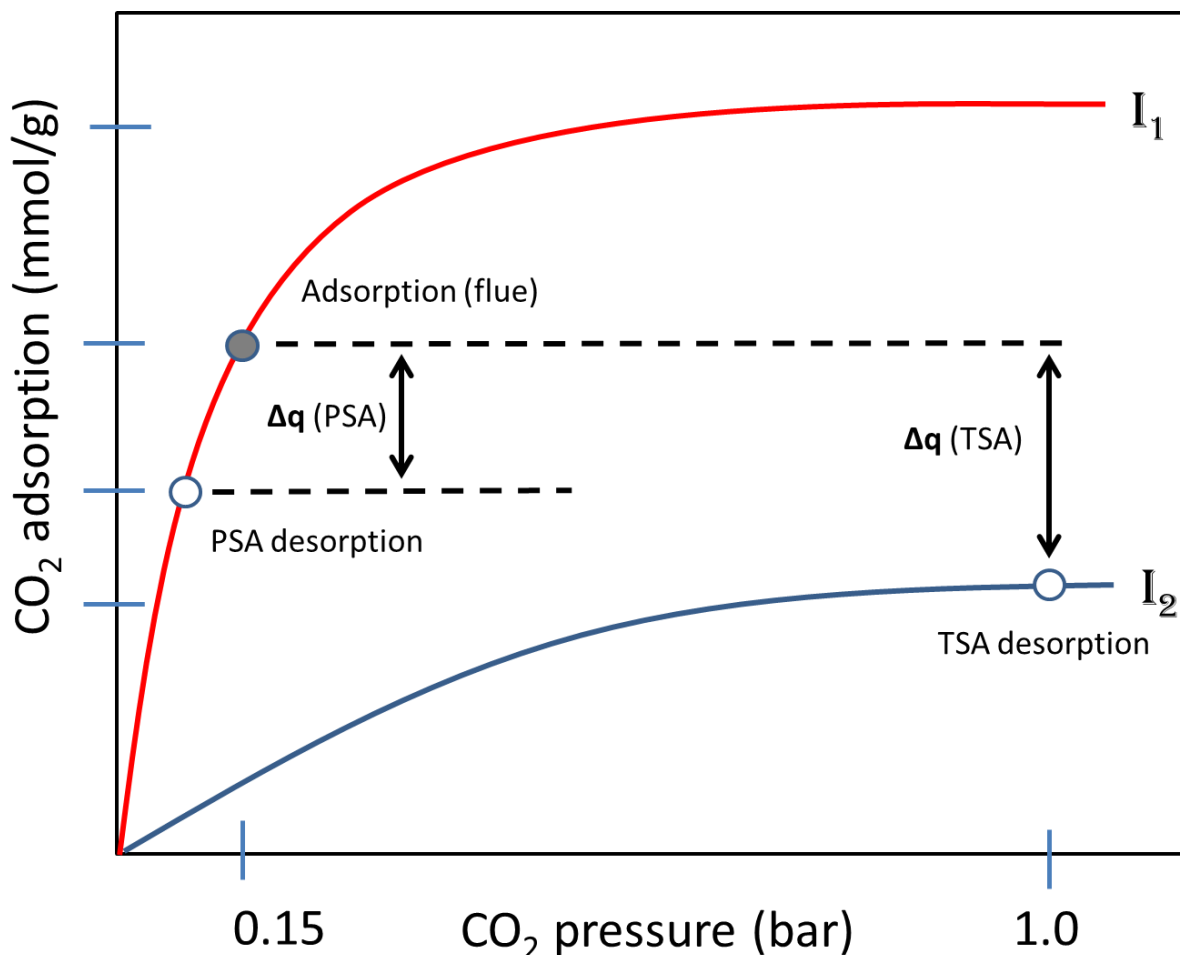


Figure 1.5: Generic picture of two adsorption isotherms within a nano-porous material. Here isotherm I_1 represents a strong adsorption regime of CO₂, where as I_2 represents a weaker adsorption. For this discussion I_2 can represent a higher temperature of the same material, or the same temperature of a different material which exhibits weaker interactions with CO₂. The closed circle represents the uptake of CO₂ from flue gas, open circles are points of desorption based on various desorption regimes such as temperature (TSA) or pressure (PSA) swing. Δq signifies the working capacity of both regimes.

Figure 1.5 demonstrates gas loading as a function of pressure. Because the partial pressure of CO₂ in flue gas is roughly 0.15 bar, materials showing promise for CO₂ capture applications demonstrate sharply rising adsorption at low partial pressures (see for example, I_1 in Figure 1.5). The affinity must not be too strong, however, as it will be much more difficult to remove CO₂ at the desorption stage. Desorption occurs at a higher temperature and/or lower pressure under the aforementioned TSA/PSA regimes, and materials are evaluated based on their CO₂ *working capacity*, which is the difference between the amount of CO₂ adsorbed on the

material at flue gas conditions and the amount of CO₂ remaining at the higher temperature/lower pressures. Here, a larger working capacity represents a greater amount of CO₂ captured at each adsorption/desorption cycle per unit mass of the material. With respect to capital costs of implementing these materials in post-combustion scrubbing units, it is preferable to minimize the quantity of material needed for capture. Thus the ideal material will maximize CO₂ working capacity at a given set of adsorption and desorption conditions.

As an example of how the working capacity is calculated from isotherms under temperature or pressure swing regimes, consider the two isotherms I_1 and I_2 in Figure 1.5 as properties of the same MOF, only at different temperatures. The open circles represent the different temperatures and pressures at which CO₂ can be desorbed. TSA processes increase the temperature of the CO₂ saturated MOF, desorbing the CO₂ at atmospheric pressure at the higher temperature isotherm (I_2). PSA involves removing CO₂ with a lower pressure, which can be seen by travelling left along the isotherm I_1 . As it is shown in the figure, it appears that the TSA process results in a higher working capacity compared with PSA, however this will ultimately depend on the isotherm shapes and temperatures or pressures used in either method. As a final comment on costs, we note that in addition to large working capacities, it is preferable to have milder conditions for desorption, as this will lower the total costs to the plant.

1.4.2 CO₂/N₂ Selectivity

A high performing sorbent material must demonstrate a strong selectivity for adsorption of CO₂ over N₂ for post-combustion CO₂ capture. This is so the purity of CO₂ recovered in the regeneration step is high, reducing compression and storage costs. With a low selectivity, one ends up compressing and storing significant amounts of N₂ rather than CO₂, which increases the costs of CCS. Measurements of selectivity are found by taking adsorption data from isotherms of CO₂ and N₂. The selectivity parameter is calculated as a ratio of the mole fractions of CO₂ (x_{CO_2})

over N_2 (x_{N_2}) adsorbed at flue gas conditions normalized by the ratio of their partial pressures (y_{CO_2}/y_{N_2}).

$$S = \frac{x_{CO_2}y_{N_2}}{x_{N_2}y_{CO_2}} \quad 1.1$$

In many cases, this selectivity parameter is determined from two separate isotherm experiments, one on CO_2 and one from N_2 . Obtaining adsorption data from two separate isotherms instead of a binary mixture ignores effects from competitive adsorption. In this case it is a conservative guess of the selectivity as CO_2 will tend to bind stronger to the surface over N_2 . This can be attributed to electrostatic interactions of the surface with the strong quadrupole moment ($13.4 \times 10^{-40} \text{ C}\cdot\text{m}^2$) exhibited by CO_2 . There are, however, theoretical methods for evaluating the adsorption of a binary mixture from two pure isotherms. The most commonly used method is the Ideal Adsorption Solution Theory (IAST)⁸¹ which assumes that the gases mix with ideal behaviour on a homogeneous surface. Despite these assumptions, for many MOFs IAST appears to predict binary adsorption quite well.^{82–88} Notable limitations include binary mixture adsorption in MOFs with flexible frameworks,⁸⁹ and those with open metal sites exhibiting strong interactions with CO_2 .⁹⁰

Because CO_2 tends to bind stronger to the surface, CO_2/N_2 selectivities in MOFs are typically greater than 1. However to obtain the target CO_2 purity of 90 - 95% when separating these two gases, the selectivity must be high, in the range of $S=150$.¹⁷

1.4.3 Isosteric Heat of Adsorption, q_{st}

The isosteric heat of adsorption is a more direct measurement of the affinity for CO_2 adsorbed on the surface of the MOF. This quantity describes the change in energy associated with an infinitesimal number of particles transferring from the gas phase to the material surface. This measurement is typically calculated using the Clausius-Clapeyron equation, which reports

the differential isosteric heat⁹¹ at a given loading of gas. It is thus a useful measurement of how strongly CO₂ will bind to the MOF, as high isosteric heats correspond to sharp increases in CO₂ loading at low partial pressures, and will also give an estimate as to how much energy will be needed to remove CO₂ at the desorption stage. It is therefore an adsorption property when evaluating the energy cost of CO₂ capture by PSA/TSA systems. To compute the isosteric heat of adsorption experimentally, it is necessary to perform isotherm experiments at several temperatures. A choice of constant CO₂ loading (N) must be made, then the q_{st} at that loading can be computed from the slope of a line fit to a plot of $\ln(p_N)$ vs $1/T$. Where P is the partial pressure at the fixed loading (N), R is the gas constant, and T is the temperature, as given by equation 1.3.

$$(\ln P)_N = \frac{q_{st}}{R} \left(\frac{1}{T} \right) + C \quad 1.2$$

Alternatively, in experimental studies it is common to fit a series of data points from several isotherms to a virial equation.⁹² This equation has no physical meaning, however the isosteric heat of adsorption at the limit of zero loading (q_{st}^0) can be reported from this fitting procedure. This quantity gives a measure of the interaction strength of the strongest adsorption sites in the material. In pressure ranges we are concerned with, both the differential and zero loading calculations will give similar results.

1.4.4 Other Flue Gases

Flue gas consists of a mixture of several species as seen in Table 1.1. While the major focus of novel materials is on its selectivity for CO₂ over N₂, other species present in flue gas such as water must be addressed. Indeed highly polar species such as water, SO_x and NO_x will competitively bind to adsorption sites much stronger than CO₂, and also have the potential to degrade the material over time due to possible nucleophilic attacks on electrophilic species in the

framework, or insertion into the metal ions.^{93,94} While there already exist industrial processes to remove these species from the flue stream,⁹⁵ the presence of a very small fraction of these gases over repeated adsorption/desorption cycles can reduce the lifetime of these materials and their ability to trap CO₂. The effects of water and acid gases on a MOF framework are difficult to quantify and have only recently been addressed.^{96,97}

1.4.5 *Stability and recyclability*

Much of the work done to validate MOFs as a CCS technology has been under controlled laboratory conditions, looking at one or two properties important to CCS which ignores some important practical considerations. That is, the implementation of MOF materials in a real TSA/PSA system fitted to a coal combustion power plant. This introduces many factors which could affect the efficacy and performance of a particular MOF. Importantly, if one were to put a MOF in a column to adsorb CO₂ downstream from a reactor, how long will it last before it would need to be replaced? This question raises concerns about each MOFs structural and thermal stability.

We define structural stability, in this case, as the ability of a MOF to maintain its shape following the evacuation of its pores. This stability is of paramount importance when a material must survive multiple adsorption/desorption cycles in which gas is constantly being adsorbed and evacuated from its void space. In many cases the crystalline structure of a MOF will possess solvent atoms in its pores, which supports and maintains the apparent porosity of the material. Upon evacuation of solvent, the material will collapse, rendering it useless as a scrubbing agent. This is more common in large-pore MOFs where the void fraction reaches values of 0.8 or higher.³⁹

For the purposes of industrial implementation of MOFs for CCS, we include here an interesting subclass of MOFs which maintain some porosity upon solvent removal, however

exhibit ‘breathing’ phenomena when exposed to various gas pressures.⁹⁸ This phenomena is due to a degree of conformational flexibility, typically exhibited by coordination of organic moieties to metal ions, which can act as a ‘hinge’ as pores become filled with CO₂.⁹⁹ These MOFs exhibit fascinating properties, including volume changes of nearly 50% upon adsorption and desorption of guest molecules,⁹⁸ however they are not ideal for CCS implementation where MOFs will be placed in a packed bed column of fixed volume. In this case, mechanical flexibility would lead to attrition, and eventual break down of the material.

Thermal stability is also an important factor in the development of this technology. In temperature swing adsorption systems, MOFs will be exposed to a range of temperatures where it must maintain framework integrity. Unfortunately, in comparison with other classes of solid materials such as zeolites, MOFs generally exhibit low thermal stability, most being unstable at temperatures above 200 °C.¹⁰⁰ The degradation of MOFs when subjected to high temperatures can be attributed to the strength of the organic ligand – metal coordination bond. There are exceptions, for example the ZIF subclass of MOFs possess strong imidazolate nitrogen bonds to Zn (II) ions, resulting in their stability in temperatures upwards of 500 °C.³²

1.4.6 Kinetic factors of gas adsorption

Many of the metrics used to evaluate MOFs relies on equilibrium thermodynamic data, such as the quantity of CO₂ adsorbed at a given temperature and pressure. However in practice, gas will flow through a fixed bed, where adsorption dynamics will play an important role in the efficiency of the material. The rate at which the component flue gases transport through a column of MOF material will be governed by their diffusion kinetics. Each gases rate of diffusion is typically quantified by a diffusion coefficient, which is proportional to its flux through a particular MOF, in units of area/time (typically m²/s). In the case of CCS technologies, the MOF should exhibit relatively high gas diffusion coefficients, to ensure that the material can

be rapidly loaded to saturation and unloaded for quick cycling in plants fitted with PSA/TSA apparatus.¹⁰¹ In addition, the difference between retention times of CO₂ and other gases such as N₂ are a direct indication of the MOFs selectivity for CO₂.^{44,51,101–106} Encouragingly one study demonstrated that two well-known MOFs (MOF-5 and MOF-177) report diffusion coefficients which are two orders of magnitude higher than those found in a commercial zeolite, 5A, ($\sim 10^{-9}$ m²/s vs 10^{-11} m²/s)¹⁰³ which is commonly used in PSA/TSA units. This is attributed to the larger pore structures of the MOFs studied, which reduces congestion of gas particles upon surface adsorption. It can therefore be understood that, because MOFs can exhibit a huge range of pore sizes, those with small pores on the order of the kinetic diameter of the flue gases (Eg. ~ 3.3 Å for CO₂) will generally exhibit poor diffusion characteristics.¹⁰⁷

Another possibility being considered in post-combustion gas separation is to exploit the differences in diffusion coefficients of individual gases via membrane separation.¹⁰⁵ In this case MOFs are synthesized as thin films or are integrated into polymer membranes, and gas mixtures are separated based on differences in kinetic diameters and adsorption affinity. In comparison to fixed bed separation, this technology is relatively young, however a handful of studies have demonstrated promising selectivities for CO₂ over N₂ using MOF membranes, despite using non-flue gas conditions.^{108–110}

1.4.7 Computer Simulations of MOFs

Recently, theoretical and computational tools have developed to the point where one can accurately simulate the adsorption properties of MOFs. Perhaps the most important weapon in a computational chemist's arsenal is the Grand Canonical Monte Carlo (GCMC) algorithm. GCMC simulations allow one to derive the values for many of the thermodynamic performance metrics mentioned in section 1.1.3 using only the empty MOF framework as input. Importantly, it relates microscopic energy interactions between guest particles (gas) and the host (MOF) to

macroscopic observed values such as the total gas uptake, as seen in an adsorption isotherm, or the isosteric heat of adsorption. This is accomplished by using computer simulations to sample a statistical mechanical ensemble (or simply, ensemble). The following subsections will discuss the concepts and computational details involved when sampling an ensemble

1.4.8 *Sampling an Ensemble with Monte Carlo Simulations*

At a given set of thermodynamic variables, such as temperature, pressure and volume, the ensemble describes all the possible configurations the atoms in a system can be in. An ensemble can be considered a large number of replicas of the system, where in each replica, the thermodynamic variables remain identical, but the momenta and positions of the atoms are different. Which thermodynamic variables the replicas are in equilibrium with will dictate the type of ensemble being sampled. As an example of an ensemble, consider a CO₂ particle diffusing through the pores of a MOF at a constant temperature. This can be considered a 'movie' of a particle where the molecule is propagated through the pore based on its velocity and the forces acting upon it from interactions with the framework. In this scenario, the number of atoms, the volume and the temperature remain constant, and this movie is said to be sampling a *canonical* ensemble, where each snapshot of the particle moving through the pore is a single state of the ensemble. If one observes the particle long enough, the average magnitude of all the velocities collected at each frame would yield the ensemble average kinetic energy of the particle within the MOF.

In this example, the CO₂ molecule is governed by the forces acting upon it, which includes a degree of correlation between each frame of the movie. Therefore, obtaining good statistically meaningful averages can require observing this particle for a very long time. Thus statistical ensembles are commonly sampled using a Monte Carlo (MC) algorithm. In MC programs, particles move, only not as a function of the forces acting upon them. In fact, there is

no time dependence at all - each particle in the system is randomly perturbed (translated and/or rotated) to sample a new state of the ensemble (see Figure 1.6). To prevent the sampling of states which are not likely going to significantly contribute to the average macroscopic properties, such as very high energy states where the CO₂ particle is extremely close to some MOF atoms, each random perturbation is evaluated using a set of rules which define what is called the Metropolis MC regime.¹¹¹ These rules, when applied to the canonical ensemble, are to accept a trial perturbation if the energy of the new state is lower than the previous ($\Delta E = E_{new} - E_{old} < 0$), or, if the energy of the new state is higher ($\Delta E > 0$) to accept the trial move based on a probability proportional to $\exp(-\Delta E/k_b T)$, which is a Boltzmann weighting of the change in energy (ΔE) of that random move, where $k_b T$ is the thermal energy of the system. This allows a particle to move to a high energy state with a probability proportional to the Boltzmann weight, preventing the system from sticking in energy wells. If the random perturbation is rejected, then the previous state is counted again towards the statistical average, thus favourably sampling lower energy states. Sampling this ensemble with a fixed loading of gas molecules yields 3-dimensional probability distributions of the gas, giving important information about where CO₂ preferentially adsorbs in the pores.

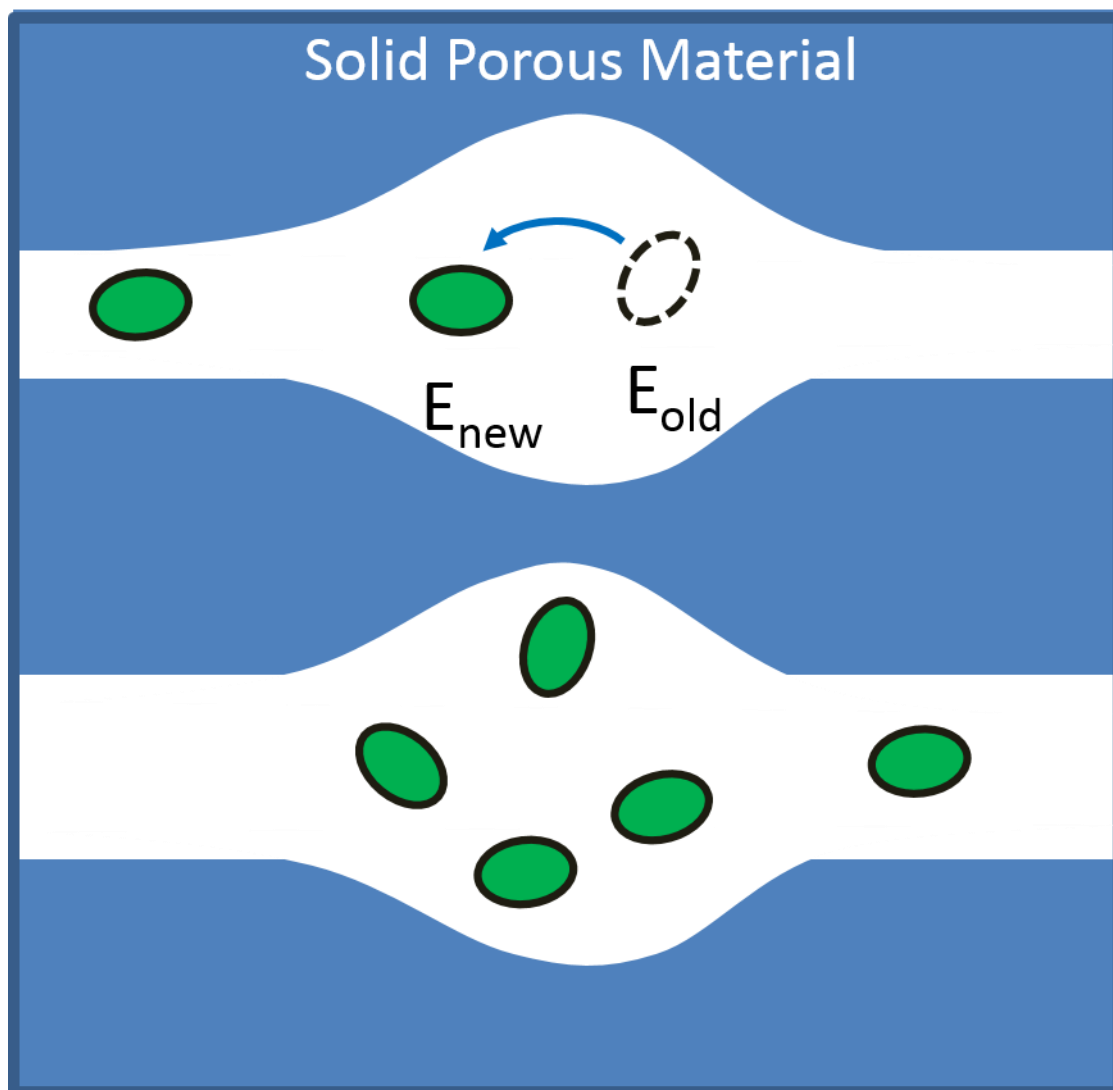


Figure 1.6: The Metropolis Monte Carlo sampling scheme. At each step a particle is selected at random to randomly displace (translation + rotation). Following displacement the energy difference is evaluated ($\Delta E = E_{\text{new}} - E_{\text{old}}$). If ΔE is negative, the move is accepted. If ΔE is positive, then it has a chance of being accepted based on a Boltzmann weighting.

1.4.9 Computing Isotherms - the Grand Canonical Monte Carlo simulation;

In the context of gas adsorption in MOFs, we want to be able to determine the quantity of gas that will adsorb within the MOF given an external pressure of the bulk phase gas and system temperature. This can be found by calculating the average number of particles in the system sampled from the *grand canonical* (GC) ensemble. In this ensemble, the chemical potential, the system volume and the temperature are fixed to specific values, while notably, the number of

particles is allowed to fluctuate. Here the chemical potential of a species is defined as the potential energy which can be absorbed or released upon a change in its quantity. Implicit in these calculations is that the gas bulk phase chemical potential at the given pressure will be equal to the chemical potential of the gas adsorbed in the pores of the MOF. In turn, the bulk phase chemical potential will be dependent on its thermodynamic state point. Determining the chemical potential of a gas in a mixture is non-trivial, however at the pressures and temperatures of flue gas, one can approximate the bulk phase as an ideal gas and the chemical potential can be determined analytically as a function of the temperature and pressure. The chemical potential for an ideal gas is expressed accordingly in Eqn. 1.3

$$\mu_{id}(p, T) = \mu_{id}^0 + RT \ln \left(\frac{P}{P_0} \right) \quad 1.3$$

Where P_0 is the total pressure of the system, μ_{id}^0 is the standard state chemical potential of the gas at $P = P_0$, and R , T , and P are the gas constant, temperature and partial pressure, respectively. Therefore by sampling the GC ensemble, and using Eqn. 1.3 to determine the bulk phase chemical potential, we can determine on average, how much of the gas will be adsorbed in the pores of a MOF at a given pressure and temperature. Metropolis Monte Carlo sampling in the GC ensemble includes translation and rotation moves identical to those performed in the canonical ensemble, and introduces two new moves; the creation and destruction of particles, with the acceptance being proportional to the chemical potential defined above. A graphical representation of this scheme is shown in Figure 1.7

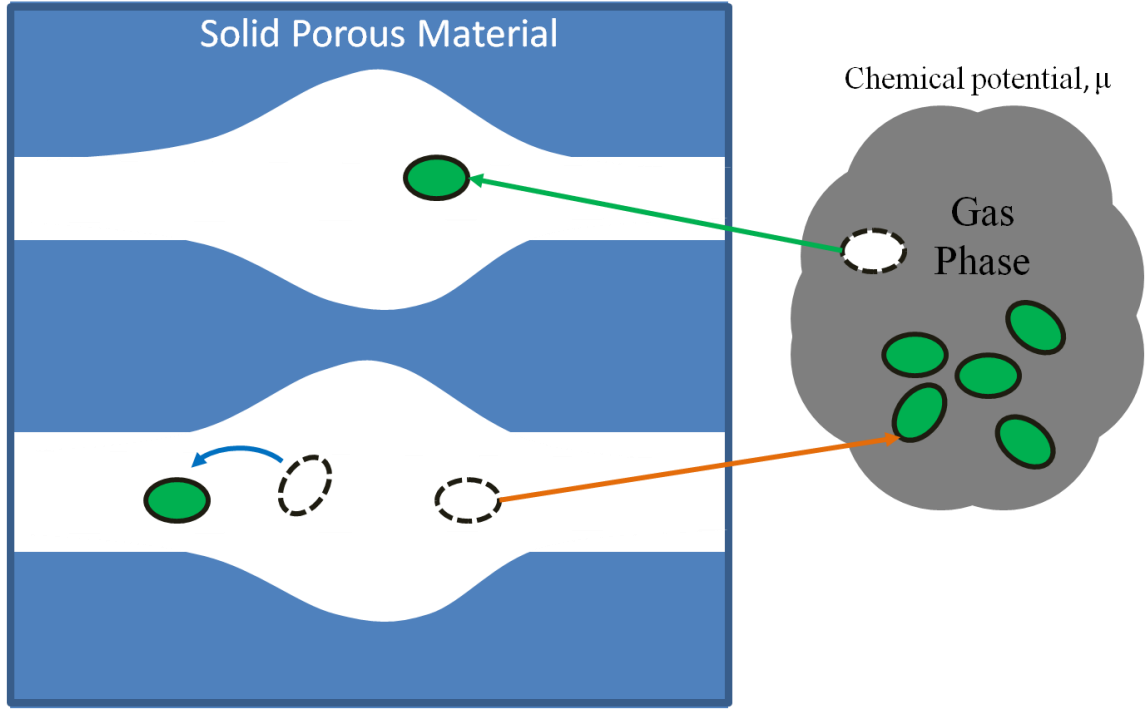


Figure 1.7: Representation of sampling the grand canonical ensemble. Particles are green. Annihilated particles are white with dashed borders. The green arrow represents the creation of a particle in the pore of the material from the gas reservoir, the orange arrow represents a particle being annihilated, and the blue arrow represents a concerted shift and rotation of a particle within the porous material. Each GCMC move is accompanied by an calculation of the change in system energy and an acceptance or rejection of the move based on Metropolis Monte Carlo sampling.

To ensure efficient sampling of the Boltzmann distribution of states, the Monte Carlo moves are accepted or rejected based on the following probabilities;

Translating/rotating a molecule already in the pore;

$$P(old \rightarrow new) = \min \left[1, \exp \left(\frac{-(E_{new} - E_{old})}{k_B T} \right) \right] \quad 1.4$$

Inserting a new gas molecule into the pore;

$$P(N \rightarrow N + 1) = \min \left[1, \frac{VP}{k_B T(N + 1)} \exp \left(\frac{-(E_{N+1} - E_N)}{k_B T} \right) \right] \quad 1.5$$

Annihilating a gas molecule from the pore;

$$P(N \rightarrow N - 1) = \min \left[1, \frac{N}{k_B T(VP)} \exp \left(\frac{-(E_{N-1} - E_N)}{k_B T} \right) \right] \quad 1.6$$

Where N is the number of gas molecules in the pores of the MOF, E is the internal energy of the states (described by subscripts), V is the total system volume, and P represents the partial pressure of the molecules in the gas phase. The equations 1.4 - 1.6 ensure that the random moves are preferentially sampling the more probable states of the Boltzmann distribution, whose contribution will dominate the average thermodynamic properties of the system (such as the ensemble average value of N which describes how much gas is adsorbed in the pores of a material at a given temperature and pressure). Despite this preferential sampling, good thermodynamic averages typically require on the order of 10^7 Monte Carlo moves. For example, each point on the isotherm defined by a temperature and pressure in Figure 1.5 are averaged values of $\langle N \rangle$ computed over 10^7 Monte Carlo steps. As such, the energy calculations performed at each step must be computationally expedient (if it took 1 second to compute the energy of a single state, then 10^7 energy calculations would require 4 months of simulation time!). Thus *ab initio* calculations are not feasible at this scale, and molecular interactions are instead calculated with molecular mechanics force fields.

Two approximations are typically made to expedite GCMC calculations, i) no intramolecular interactions are calculated; the MOF framework atoms are held fixed in their crystallographic positions and are not allowed to move during the simulation, while guest particles can move, however their bonds are rigid, and ii) intermolecular interactions are relegated to short-range steric and dispersion interactions as well as long-range Coulombic interactions. Despite these seemingly severe approximations, thermodynamic properties extracted from these calculations can be quite accurate,¹¹² and can be computed in a couple of hours on a single CPU. There are of course, exceptions. This is particularly evident when MOF frameworks are shown to be flexible,^{113,114} or there is some form of chemisorption on the

framework open metal sites. In these cases tailored molecular simulation methods are required to accurately reproduce experimental isotherms.^{115,116}

For the gas adsorption calculations reported in this thesis, an in-house GCMC algorithm was written by the author in the FORTRAN programming language, and based on the open source MD program DL_POLY.¹¹⁷ The GCMC program computes adsorption data using the rigid approximations discussed above, and includes the considerations necessary to calculate energy interactions in periodic materials, which are discussed in sections 1.4.10 and 1.4.13 below.

1.4.10 Energy Calculations with Periodic Boundary Conditions

Computer simulations of MOFs and other crystalline materials require a truncation of the infinite system to a tangible representative set of atoms. For this reason, simulations are performed on a translationally invariant set of atoms described by a unit cell (or a set of multiple adjacent unit cells, called a ‘supercell’). To avoid border effects at the edges of the simulation cell, *periodic boundary conditions* are implemented to approximate the infinite nature of the material, where particles can pass through one side of the cell only to re-appear on the opposite side, preserving any properties it possessed (forces, velocities, etc.). While this effectively truncates the infinite system to something computationally feasible, artifacts can arise from these boundary conditions which can render the models ineffective. These must be addressed ensuring some conditions are met prior to performing calculations on a periodic material.

The first condition is to impose a cutoff for particle interactions to half the cell length distance. This ensures that particles do not interact more than once with other particles, as doing so introduces periodic artifacts in the simulation. Because particles can only interact with other particles which are at most half a cell length away, this introduces another condition; minimum cell size. One must ensure that at half the cell distance, one is capturing all of the pairwise energy

interactions, and that interactions beyond this point are essentially negligible. In cases where the unit cell is too small, one can construct a supercell to ensure this condition is met. At the same time, setting the cell size too large can result in a computationally cumbersome simulation. A typical value for the half-cell cutoff is between 12 and 14 Å, which is dependent on the parameters used to compute the interaction energies between atoms.

The interaction energies in GCMC simulations are calculated separately for short range dispersion and steric interactions and long range Coulombic interactions. The short range dispersion and steric interactions can be approximated with what is called a Lennard-Jones or 12-6 potential, where ε_{ij} and σ_{ij} are parameters assigned for interactions between atom i and j , summing over all N atoms in the simulation cell,

$$E_{L-J}(r) = \frac{1}{2} \sum_i^N \sum_j^N 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \quad 1.7$$

where the $\frac{1}{2}$ factor at the front is included to avoid double counting atom pairs. This function decays rapidly with distance between atoms i and j (r_{ij}^{-6}) which does not pose an issue for the cutoffs imposed by the periodic boundary conditions. However the Coulombic interactions decay much slower (r_{ij}^{-1}) and require special considerations for the energy convergence, thus the interaction between atoms with charges q_i and q_j , summing over all n periodic images of the cell with sides of length L ,

$$E_{coul}(r) = k \frac{1}{2} \sum_n \sum_i^N \sum_j^N \frac{q_i q_j}{|r_{ij} + nL|} \quad 1.8$$

converges very slowly, which can dramatically increase the time needed to compute thermodynamically averaged properties. One solution to this convergence problem is the Ewald summation method,¹¹⁸ where charge interactions are split into a short range calculation

performed in real space which converges rapidly, and a long range calculation which converges rapidly in reciprocal space. Before going into details of the Ewald method, we describe how one obtains parameters in these molecular mechanics force fields, such that we can provide a reasonable model for the interaction energy of a system.

1.4.11 Force Field Parameters A. van der Waals interactions

In molecular modelling, both steric repulsion and attractive dispersion interactions are computed pairwise (i.e. between two distinct atoms) and combined in a single function. There are several different functions which are commonly used to obtain the interaction energy between atoms, the choice of which largely depends on the type of system being studied (gases, liquids, organics, inorganics, metallic systems, etc.) and force field parameters available.

The Lennard-Jones (L-J) function, which is shown in Figure 1.8, represents one of the most popular functions for calculating the steric and dispersion interactions, particularly for gases. This is primarily due to computational time; it is computationally cheaper to compute 6^{th} and 12^{th} powers than it is to compute exponential functions, which is common in other van der Waals functions.^{119,120} The parameters, ϵ_i and σ_i describe, for each unique atom *type*, the L-J functions well depth and y-axis intercept when interacting with an identical atom of the same *type*.

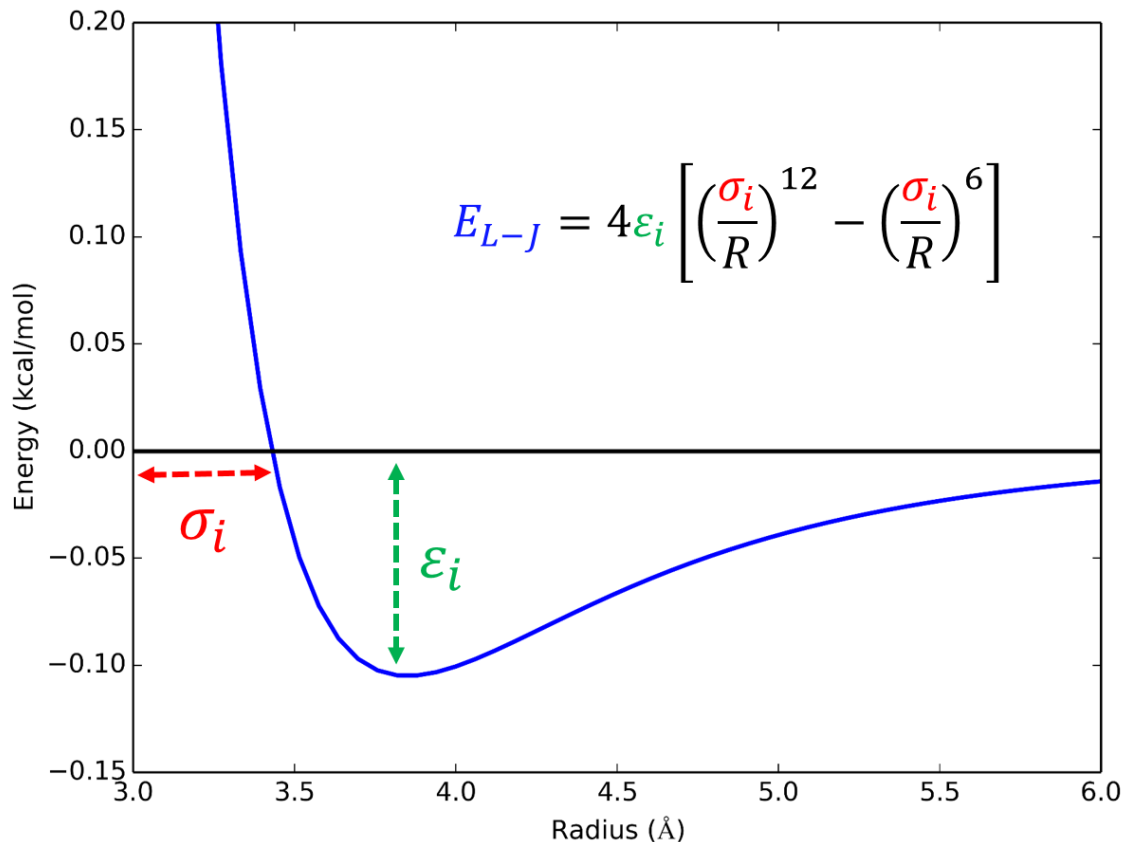


Figure 1.8: The Lennard-Jones function for $\sigma_i = 3.43 \text{ \AA}$ and $\epsilon_i = 0.105 \text{ kcal/mol}$.

Here we distinguish what the word *type* means, since it is an important concept in force field calculations. The atom type is a description of a class of atoms which are identical in their parameterization. Implicit in the force field parameters for each atom type is the electronic interactions it will have with all other atom types in the system. It is therefore important to discern separate atom types even for the same element in different electronic environments. For example, the Generalized Amber Force Field (GAFF),¹²¹ which was developed for organic systems, contains 7 different force field types for carbon, including separate types for aromatic carbon, sp^2 hybridized carbon in an aliphatic double bond, and carbonyl carbons.

Because the study of MOFs is relatively young, the development of Lennard-Jones parameters specifically for this class of materials is lacking. Of the force field parameters used,

most common are from the DREIDING force field¹²² and the Universal Force Field (UFF)¹²³ due to their coverage of a large portion of the periodic table. Despite the L-J parameters being developed with different systems in mind, they have been proven to give quite accurate results when comparing computed thermodynamic properties with MOF experimental data.¹¹² While these force fields provide ε and σ parameters for *homogeneous* atomic interactions, it is important to identify all van der Waals interactions in each system, including those between different atom types. For this purpose, the widely used Lorentz-Berthelot mixing rules are applied to all possible heterogeneous atom type pairings. This involves taking a geometric mean of the well depth parameters, $\varepsilon_{ij} = \sqrt{\varepsilon_i \varepsilon_j}$ and an arithmetic mean of the y-intercept parameter, $\sigma_{ij} = (\sigma_i + \sigma_j)/2$.

Finally, we note that the aforementioned UFF and DREIDING parameters are typically used to describe MOF atom types, however gas molecules come with their own specific sets of L-J parameters (and atomic charges). These parameters are typically fit to experimental data to reproduce some observed property or trend of the gas. For example, García-Sánchez *et al.* fit the L-J parameters for the carbon and oxygen atoms of CO₂ to reproduce its vapour-liquid coexistence curve, which was shown to provide good agreement with experimental zeolite isotherms.¹²⁴

1.4.12 Force Field Parameters B. Partial Atomic Charges

The static coulomb interaction between two atoms can be considered the electrostatic interaction between each atoms combined nuclear charge and local electron density. In molecular mechanics simulations, the electronic system is not explicitly treated, and the Coulombic interactions are typically approximated by assigning partial atomic point charges to each atom. When assigning partial atomic charges to a system, the desire is to produce an electrostatic

potential (ESP) which agrees well with the ESP from a full treatment of the electronic system. For molecules, this is done by fitting atom-centred charges such that they reproduce the electrostatic potential produced from a quantum mechanical (QM) calculation.^{125,126} However, in the case of periodic systems, fitting atomic charges to the QM ESP introduces an additional challenge; an ill-defined reference state of the ESP. This was solved by the Woo group in 2009 with introduction of the REPEAT method.¹²⁷ This method fits a set of partial atomic charges to the QM ESP of a periodic material by introducing an error functional which shifts the ESPs by their average value, such that the relative difference between the QM ESP and the ESP from point charges is minimized.

While the REPEAT method has yielded highly accurate results in a number of studies,^{76,128–131} one of its limitations is the computational cost of performing a QM calculation to obtain the ESP necessary for the fitting procedure, which can take hours for each material. As an alternative, parameterized methods such as the charge equilibration method (QEq),⁷⁷ which require only seconds to compute partial charges on atoms, are sometimes used when speed is desired.^{71,72} However, care must be used with these parameterized methods, as they can produce poor ESPs in MOFs which can result in spurious conclusions being drawn.⁷⁸

The next section discusses how these charges are used to compute the Coulombic energy in a molecular mechanics simulation with periodic boundary conditions. Specifically, how one recovers the long range nature of the Coulomb potential when faced with a short-range cutoff.

1.4.13 The Ewald Summation Method

In lieu of a full derivation of the Ewald summation equations (which are elegantly derived in reference 132) we will briefly discuss their form to demonstrate how charge interactions can be computed within a relatively small simulation cell. The main challenge associated with the electrostatic interaction between two ions is that it decays quite slowly, resulting in divergent

energy calculations as ions interact with themselves in adjacent periodic images. The solution, as proposed by Ewald is to ‘smear’ an equal, but opposite screening charge distribution around each ion, making the electrostatic potential in real space decay exponentially, which converges much quicker than the Coulomb potential. The smeared charge distribution is in the form of a Gaussian function, centred on each ion point charge. A graphical representation of the effect of a Gaussian screening charge on the real space electrostatic potential is plotted in Figure 1.9a.

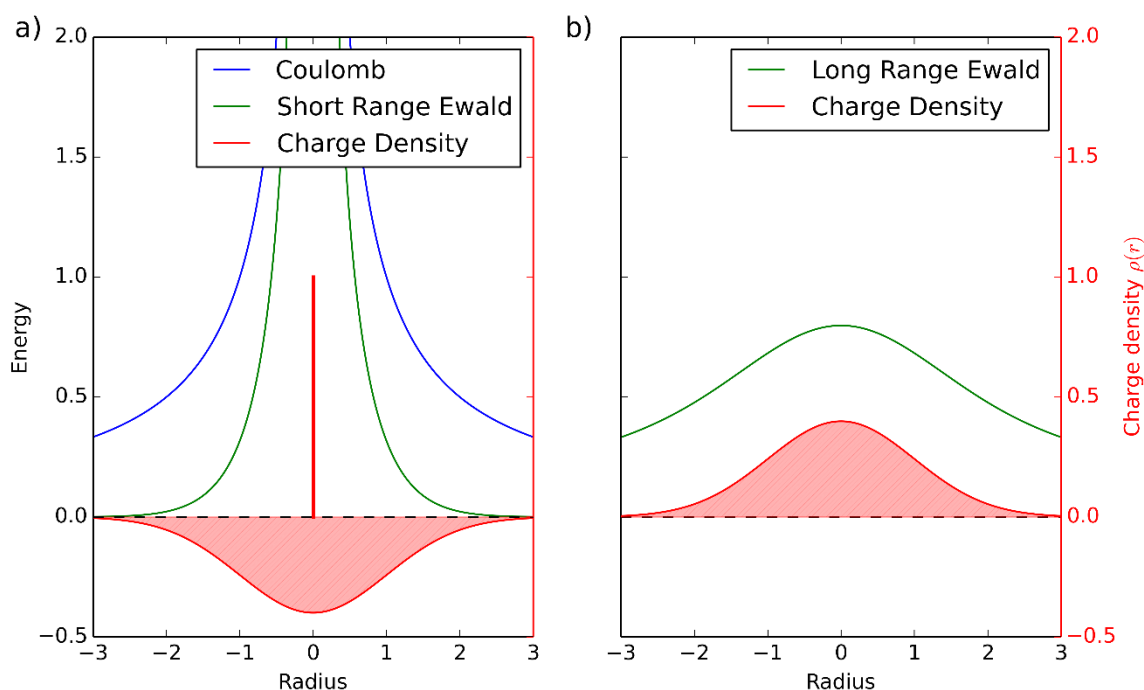


Figure 1.9: a) Effects of the real space potential energy decay when an equal but opposite Gaussian screening charge is placed at the +1 point charge. The green lines represent the electrostatic potential when the Gaussian charge is applied, the blue represents the potential without the Gaussian screening charge. b) An additional Gaussian charge is added back to restore the long range character of the electrostatic potential. This charge density is treated separately in reciprocal space.

While the Gaussian charge distributions permit the rapid convergence of the electrostatic interactions in a simulation cell, their very inclusion artificially adjusts the electrostatic potential of the system. Therefore an additional set of opposing Gaussian charge distributions are

introduced to restore the electrostatic potential to its original form, thus the charge distribution for an ion i at position $\mathbf{r}_i$ with charge q_i can be described by

$$\rho_i(\mathbf{r}) = q_i\delta(\mathbf{r} - \mathbf{r}_i) - q_iG(\mathbf{r} - \mathbf{r}_i) + q_iG(\mathbf{r} - \mathbf{r}_i) \quad 1.9$$

where $\delta(\mathbf{r} - \mathbf{r}_i)$ is a Dirac delta function which equals 1 when $\mathbf{r} = \mathbf{r}_i$ and 0 everywhere else. This form of the charge distribution allows one to split the electrostatic potential into a short and long range term. The short range term is described by the first two charge distributions in Eqn. 1.9 and are shown in red in Figure 1.9a, while the long range interactions are calculated using the last Gaussian charge term shown in red in Figure 1.9b. This allows the potential which describes this charge distribution to also be split into short and long range terms. For each ion i , the potential at point $\mathbf{r}$ becomes,

$$\phi_i(\mathbf{r}) = \phi_i^S(\mathbf{r}) + \phi_i^L(\mathbf{r}) \quad 1.10$$

Where for the short range,

$$\phi_i^S(\mathbf{r}) = \frac{q_i}{4\pi\epsilon_0} \int \frac{\delta(\mathbf{r} - \mathbf{r}') - G(\mathbf{r} - \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' \quad 1.11$$

And the long range,

$$\phi_i^L(\mathbf{r}) = \frac{q_i}{4\pi\epsilon_0} \int \frac{G(\mathbf{r} - \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' \quad 1.12$$

We note that the integrals over all space in Eqns. 1.11 and 1.12 must be introduced because we are dealing with a *continuous* charge distribution, instead of the point charges described in the original system. The short range potential described in Eqn. 1.11 converges rapidly (shown in green in Figure 1.9a) whose final form, summing over all ions j in the system is,

$$\phi^S(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} \sum_{j=1}^N \frac{q_j}{|\mathbf{r} - \mathbf{r}_j|} \operatorname{erfc}\left(\frac{|\mathbf{r} - \mathbf{r}_j|}{\sqrt{2}\sigma}\right) \quad 1.13$$

where $\text{erfc}(x)$ is the complimentary error function arising from integration of the Gaussian function and σ is the width of the Gaussian. The long range potential, however, does not converge, and the summation of $\phi_i^L(r)$ from all lattice ions is a periodic function that cannot be computed in real space. The solution is to sum the long range potential in reciprocal space by performing a Fourier transform of the Gaussian charge distributions $q_i G(r - r_i)$. The electrostatic potential of this long range charge distribution is as follows,

$$\phi^L(\mathbf{r}) = \frac{1}{V\epsilon_0} \sum_{\mathbf{k} \neq \mathbf{0}} \sum_{j=1}^N \frac{q_j}{k^2} e^{i\mathbf{k} \cdot (\mathbf{r} - \mathbf{r}_j)} e^{-\sigma^2 k^2 / 2} \quad 1.14$$

where the decay term $e^{-\sigma^2 k^2 / 2}$ ensures convergence in reciprocal space. This is summed over reciprocal lattice vectors $\mathbf{k}$, where $\phi^L(\mathbf{r})=0$ when $\mathbf{k} = \mathbf{0}$ assuming that the periodic cell is charge neutral.

When computing the electrostatic energy from the potential described by equations 1.13 and 1.14 an additional real-space correction is included in the Ewald energy summation which eliminates the effect of the Gaussian distributions in reciprocal space interacting with themselves. This is of the form;

$$E_{self} = \frac{1}{4\pi\epsilon_0} \frac{1}{\sqrt{2\pi}\sigma} \sum_{i=1}^N q_i^2 \quad 1.15$$

Thus the total electrostatic energy computed from the Ewald summation using Eqns. 1.13, 1.14, and 1.15 is,

$$E = E_S + E_L - E_{self} \quad 1.16$$

$$E = \frac{1}{2} \sum_{i=1}^N q_i \phi_S(\mathbf{r}_i) + \frac{1}{2} \sum_{i=1}^N q_i \phi_L(\mathbf{r}_i) - E_{self} \quad 1.17$$

1.5 References

- (1) Pachauri, R. K.; Allen, M. R.; Barros, V. R.; Broome, J.; Cramer, W.; Christ, R.; Church, J. a.; Clarke, L.; Dahe, Q.; Dasgupta, P.; Dubash, N. K.; Edenhofer, O.; Elgizouli, I.; Field, C. B.; Forster, P.; Friedlingstein, P.; Fuglestedt, J.; Gomez-Echeverri, L.; Hallegatte, S.; Hegerl, G.; Howden, M.; Jiang, K.; Cisneros, B. J.; Kattsov, V.; Lee, H.; Mach, K. J.; Marotzke, J.; Mastrandrea, M. D.; Meyer, L.; Minx, J.; Mulugetta, Y.; O'Brien, K.; Oppenheimer, M.; Pereira, J. J.; Pichs-Madruga, R.; Plattner, G.-K.; Pörtner, H.-O.; Power, S. B.; Preston, B.; Ravindranath, N. H.; Reisinger, A.; Riahi, K.; Rusticucci, M.; Scholes, R.; Seyboth, K.; Sokona, Y.; Stavins, R.; Stocker, T. F.; Tschakert, P.; Vuuren, D. Van; Ypersele, J.-P. Van. *IPCC Climate Change 2014: Synthesis Report*; 2014.
- (2) Nicholls, R. J.; Cazenave, A. *Science* **2010**, 328, 1517.
- (3) McLaughlin, J. F.; Hellmann, J. J.; Boggs, C. L.; Ehrlich, P. R. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, 99, 6070.
- (4) International Energy Agency. *CO2 Emissions From Fuel Combustion Highlights*; 2013.
- (5) Botzen, W. J. W.; Gowdy, J. M.; van den Bergh, J. C. J. M. *Cumulative CO2 emissions: shifting international responsibilities for climate debt*; 2008; Vol. 8.
- (6) Metz, B.; Davidson, O.; de Coninck, H.; Loos, M.; Meyer, L. *IPCC Special Report on Carbon Dioxide Capture and Storage*; 2005.
- (7) Kanniche, M.; Gros-Bonnivard, R.; Jaud, P.; Valle-Marcos, J.; Amann, J.-M.; Bouallou, C. *Appl. Therm. Eng.* **2010**, 30, 53.
- (8) Dijkstra, J. .; Jansen, D. *Energy* **2004**, 29, 1249.
- (9) Hasan, M.; Boukouvala, F. *Ind. ...* **2014**, 53, 7489.
- (10) Babu, P.; Ho, C. Y.; Kumar, R.; Linga, P. *Energy* **2014**, 70, 664.
- (11) Park, S.; Lee, S.; Lee, Y.; Lee, Y.; Seo, Y. *Int. J. Greenh. Gas Control* **2013**, 14, 193.
- (12) Folger, P. *Carbon Capture : A Technology Assessment*; 2010.
- (13) Sumida, K.; Rogow, D. L.; Mason, J. a; McDonald, T. M.; Bloch, E. D.; Herm, Z. R.; Bae, T.-H.; Long, J. R. *Chem. Rev.* **2012**, 112, 724.

- (14) Abu-Zahra, M. R. M.; Schneiders, L. H. J.; Niederer, J. P. M.; Feron, P. H. M.; Versteeg, G. F. *Int. J. Greenh. Gas Control* **2007**, *1*, 37.
- (15) Kumar, S.; Cho, J. H.; Moon, I. *Int. J. Greenh. Gas Control* **2014**, *20*, 87.
- (16) Tan, L. S.; Lau, K. K.; Bustam, M. a.; Shariff, a. M. *J. Nat. Gas Chem.* **2012**, *21*, 7.
- (17) Ho, M. T.; Allinson, G. W.; Wiley, D. E. *Ind. Eng. Chem. Res.* **2008**, *47*, 4883.
- (18) EESI. *Carbon Pricing Around The World: Fact Sheet*; Washington, DC, 2012.
- (19) Weiland, R. H.; Dingman, J. C.; Cronin, D. B. *J. Chem. Eng. Data* **1997**, *42*, 1004.
- (20) Harlick, P. J. E.; Tezel, F. H. *Microporous Mesoporous Mater.* **2004**, *76*, 71.
- (21) Wang, L.; Liu, Z.; Li, P.; Yu, J.; Rodrigues, A. E. *Chem. Eng. J.* **2012**, *197*, 151.
- (22) Su, F.; Lu, C. *Energy & Environmental Science*, 2012, *5*, 9021.
- (23) Mulgundmath, V.; Tezel, F. H. *Adsorption* **2010**, *16*, 587.
- (24) Voss, C. *Adsorption* **2005**, *11*, 527.
- (25) Akhtar, F.; Liu, Q.; Hedin, N.; Bergström, L. *Energy Environ. Sci.* **2012**, *5*, 7664.
- (26) Siriwardane, R.; Shen, M.; Fisher, E.; Poston, J. *Energy & Fuels* **2001**, 279.
- (27) Li, J.-R.; Ma, Y.; McCarthy, M. C.; Sculley, J.; Yu, J.; Jeong, H.-K.; Balbuena, P. B.; Zhou, H.-C. *Coord. Chem. Rev.* **2011**, *255*, 1791.
- (28) Hoskins, B.; Robson, R. *J. Am. Chem. ...* **1990**, *112*, 1546.
- (29) Morris, R. E.; Wheatley, P. S. *Angew. Chem. Int. Ed. Engl.* **2008**, *47*, 4966.
- (30) Li, J.-R.; Kuppler, R. J.; Zhou, H.-C. *Chem. Soc. Rev.* **2009**, *38*, 1477.
- (31) Férey, G. *Chem. Soc. Rev.* **2008**, *37*, 191.
- (32) Park, K. S.; Ni, Z.; Côté, A. P.; Choi, J. Y.; Huang, R.; Uribe-Romo, F. J.; Chae, H. K.; O'Keeffe, M.; Yaghi, O. M. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 10186.
- (33) Kitagawa, S.; Kitaura, R.; Noro, S. *Angew. Chem. Int. Ed. Engl.* **2004**, *43*, 2334.
- (34) Ma, S.; Zhou, H.-C. *J. Am. Chem. Soc.* **2006**, *128*, 11734.
- (35) Cychosz, K. A.; Wong-Foy, A. G.; Matzger, A. J. *J. Am. Chem. Soc.* **2008**, *130*, 6938.

- (36) Côté, A. P.; Benin, A. I.; Ockwig, N. W.; O'Keeffe, M.; Matzger, A. J.; Yaghi, O. M. *Science* **2005**, *310*, 1166.
- (37) Rowsell, J. L. C.; Spencer, E. C.; Eckert, J.; Howard, J. A. K.; Yaghi, O. M. *Science* (80-.). **2005**, *309*, 1350.
- (38) Deng, H.; Doonan, C. J.; Furukawa, H.; Ferreira, R. B.; Towne, J.; Knobler, C. B.; Wang, B.; Yaghi, O. M. *Science* **2010**, *327*, 846.
- (39) Furukawa, H.; Ko, N.; Go, Y. B.; Aratani, N.; Choi, S. B.; Choi, E.; Yazaydin, a O.; Snurr, R. Q.; O'Keeffe, M.; Kim, J.; Yaghi, O. M. *Science* **2010**, *329*, 424.
- (40) Férey, G.; Mellot-Draznieks, C.; Serre, C.; Millange, F.; Dutour, J.; Surblé, S.; Margiolaki, I. *Science* **2005**, *309*, 2040.
- (41) Chui, S. S. A Chemically Functionalizable Nanoporous Material [Cu₃(TMA)2(H₂O)₃]_n. *Science*, 1999, *283*, 1148–1150.
- (42) Furukawa, H.; Ko, N.; Go, Y. B.; Aratani, N.; Choi, S. B.; Choi, E.; Yazaydin, A. O.; Snurr, R. Q.; O'Keeffe, M.; Kim, J.; Yaghi, O. M. *Science* **2010**, *329*, 424.
- (43) Kitaura, R.; Kitagawa, S.; Kubota, Y.; Kobayashi, T.; Kindo, K.; Mita, Y.; Matsuo, A.; Kobayashi, M.; Chang, H.; Ozawa, T.; Suzuki, M.; Sakata, M.; Takata, M. *Science*, 2002, *298*, 2358–2361.
- (44) Banerjee, R.; Phan, A.; Wang, B.; Knobler, C.; Furukawa, H.; O'Keeffe, M.; Yaghi, O. M. *Science* **2008**, *319*, 939.
- (45) Vaidhyanathan, R.; Iremonger, S. S.; Shimizu, G. K. H.; Boyd, P. G.; Alavi, S.; Woo, T. K. *Science* **2010**, *330*, 650.
- (46) Zhao, X.; Bu, X.; Wu, T.; Zheng, S.-T.; Wang, L.; Feng, P. *Nat. Commun.* **2013**, *4*, 1.
- (47) Seo, J.; Whang, D.; Lee, H.; Jun, S.; Oh, J.; Jeon, Y.; Kim, K. *Nature* **2000**, *404*, 982.
- (48) Li, H.; Eddaoudi, M.; O'Keeffe, M.; Yaghi, O. M. *Nature* **1999**, *402*, 276.
- (49) Kawamichi, T.; Haneda, T.; Kawano, M.; Fujita, M. *Nature* **2009**, *461*, 633.
- (50) Yaghi, O. M.; O'Keeffe, M.; Ockwig, N. W.; Chae, H. K.; Eddaoudi, M.; Kim, J. *Nature* **2003**, *423*, 705.
- (51) Wang, B.; Côté, A. P.; Furukawa, H.; O'Keeffe, M.; Yaghi, O. M. *Nature* **2008**, *453*, 207.
- (52) Matsuda, R.; Kitaura, R.; Kitagawa, S.; Kubota, Y.; Belosludov, R. V.; Kobayashi, T. C.; Sakamoto, H.; Chiba, T.; Takata, M.; Kawazoe, Y.; Mita, Y. *Nature* **2005**, *436*, 238.

- (53) Chae, H. K.; Siberio-Pérez, D. Y.; Kim, J.; Go, Y.; Eddaoudi, M.; Matzger, A. J.; O'Keeffe, M.; Yaghi, O. M. *Nature* **2004**, 427, 523.
- (54) Nugent, P.; Belmabkhout, Y.; Burd, S. D.; Cairns, A. J.; Luebke, R.; Forrest, K.; Pham, T.; Ma, S.; Space, B.; Wojtas, L.; Eddaoudi, M.; Zaworotko, M. J. *Nature* **2013**, 495, 80.
- (55) Farha, O. K.; Eryazici, I.; Jeong, N. C.; Hauser, B. G.; Wilmer, C. E.; Sarjeant, A. a; Snurr, R. Q.; Nguyen, S. T.; Yazaydin, a Ö.; Hupp, J. T. *J. Am. Chem. Soc.* **2012**, 134, 15016.
- (56) Yaghi, O. M.; O'Keeffe, M.; Ockwig, N. W.; Chae, H. K.; Eddaoudi, M.; Kim, J. *Nature* **2003**, 423, 705.
- (57) Ockwig, N. W.; Delgado-Friedrichs, O.; O'Keeffe, M.; Yaghi, O. M. *Acc. Chem. Res.* **2005**, 38, 176.
- (58) O'Keeffe, M.; Peskov, M. A.; Ramsden, S. J.; Yaghi, O. M. *Acc. Chem. Res.* **2008**, 41, 1782.
- (59) Wang, Y.; LeVan, M. D. *J. Chem. Eng. Data* **2009**, 54, 2839.
- (60) Mason, J. A.; Sumida, K.; Herm, Z. R.; Krishna, R.; Long, J. R. *Energy Environ. Sci.* **2011**.
- (61) Zhu, Y.; Wang, Y.-M.; Zhao, S.-Y.; Liu, P.; Wei, C.; Wu, Y.-L.; Xia, C.-K.; Xie, J.-M. *Inorg. Chem.* **2014**, 53, 7692.
- (62) Wang, J.; Luo, J.; Zhao, J.; Li, D.-S.; Li, G.; Huo, Q.; Liu, Y. *Cryst. Growth Des.* **2014**, 14, 2375.
- (63) Niu, Z.; Fang, S.; Ma, J.-G.; Zhang, X.-P.; Cheng, P. *Chem. Commun. (Camb)*. **2014**, 50, 7797.
- (64) Ye, S.; Jiang, X.; Ruan, L.-W.; Liu, B.; Wang, Y.-M.; Zhu, J.-F.; Qiu, L.-G. *Microporous Mesoporous Mater.* **2013**, 179, 191.
- (65) Van Heest, T.; Teich-McGoldrick, S. L.; Greathouse, J. a.; Allendorf, M. D.; Sholl, D. S. *J. Phys. Chem. C* **2012**, 116, 13183.
- (66) Allen, F. H. *Acta Crystallogr. Sect. B Struct. Sci.* **2002**, 58, 380.
- (67) Chui, S. S. Y.; Lo, S. M. F.; Charmant, J. P. H.; Orpen, A. G.; Williams, I. D. *Science (80-)*. **1999**, 283, 1148.
- (68) Torrisi, A.; Bell, R. G.; Mellot-Draznieks, C. *Microporous Mesoporous Mater.* **2013**, 168, 225.
- (69) Torrisi, A.; Bell, R.; Mellot-Draznieks, C. *Cryst. Growth Des.* **2010**, 10, 2839.
- (70) Dawson, R.; Adams, D. J.; Cooper, A. I. *Chem. Sci.* **2011**, 2, 1173.
- (71) Haldoupis, E.; Nair, S.; Sholl, D. S. *J. Am. Chem. Soc.* **2012**, 134, 4313.
- (72) Wilmer, C. E.; Farha, O. K.; Bae, Y.-S.; Hupp, J. T.; Snurr, R. Q. *Energy Environ. Sci.* **2012**, 5, 9849.

- (73) Wilmer, C. E.; Leaf, M.; Lee, C. Y.; Farha, O. K.; Hauser, B. G.; Hupp, J. T.; Snurr, R. Q. *Nat. Chem.* **2012**, *4*, 83.
- (74) Sikora, B. J.; Winnegar, R.; Proserpio, D. M.; Snurr, R. Q. *Microporous Mesoporous Mater.* **2014**, *186*, 207.
- (75) Martin, R. L.; Haranczyk, M. *Cryst. Growth Des.* **2014**, *14*, 2431.
- (76) Vaidhyanathan, R.; Iremonger, S. S.; Shimizu, G. K. H.; Boyd, P. G.; Alavi, S.; Woo, T. K. *Angew. Chem. Int. Ed. Engl.* **2012**, *51*, 1826.
- (77) Rappe, A. K.; Goddard, W. A. *J. Phys. Chem.* **1991**, *95*, 3358.
- (78) Kadantsev, E. S.; Boyd, P. G.; Daff, T. D.; Woo, T. K. *J. Phys. Chem. Lett.* **2013**, *4*, 3056.
- (79) Brunauer, S.; Emmett, P. H.; Teller, E. *J. Am. Chem. Soc.* **1938**, *60*, 309.
- (80) Brunauer, S.; Deming, L. S.; Deming, W. E.; Teller, E. *J. Am. Chem. Soc.* **1940**, *62*, 1723.
- (81) Myers, a L.; Prausnitz, J. M. *AIChE J.* **1965**, *11*, 121.
- (82) Simmons, J. M.; Wu, H.; Zhou, W.; Yildirim, T. *Energy & Environmental Science*, **2011**, *4*, 2177.
- (83) Bae, Y.-S.; Farha, O. K.; Hupp, J. T.; Snurr, R. Q. *Journal of Materials Chemistry*, **2009**, *19*, 2131.
- (84) Zheng, B.; Bai, J.; Duan, J.; Wojtas, L.; Zaworotko, M. J. *J. Am. Chem. Soc.* **2011**, *133*, 748.
- (85) Keskin, S.; Liu, J.; Johnson, J. K.; Sholl, D. S. *Langmuir* **2008**, *24*, 8254.
- (86) Keskin, S. *Ind. Eng. Chem. Res.* **2010**, *49*, 11689.
- (87) Liu, B.; Yang, Q.; Xue, C.; Zhong, C.; Chen, B.; Smit, B. *J. Phys. Chem. C* **2008**, *112*, 9854.
- (88) Liu, B.; Smit, B. *Langmuir* **2009**, *25*, 5918.
- (89) Coudert, F.-X. *Phys. Chem. Chem. Phys.* **2010**, *12*, 10904.
- (90) Krishna, R.; van Baten, J. M. *Phys. Chem. Chem. Phys.* **2011**, *13*, 10593.
- (91) Pan, H.; Ritter, J.; Balbuena, P. *Langmuir* **1998**, *14*, 6323.
- (92) Czepirski, L.; Jagiełło, J. *Chem. Eng. Sci.* **1989**, *44*, 797.
- (93) Low, J. J.; Benin, A. I.; Jakubczak, P.; Abrahamian, J. F.; Faheem, S. a; Willis, R. R. *J. Am. Chem. Soc.* **2009**, *131*, 15834.

- (94) Schröck, K.; Schröder, F.; Heyden, M.; Fischer, R. A.; Havenith, M. *Phys. Chem. Chem. Phys.* **2008**, *10*, 4732.
- (95) Hansen, B. B.; Kiil, S. *Ind. Eng. Chem. Res.* **2012**, *51*, 10100.
- (96) Uthor, U. T. D. a; Tan, S. K.; Nijem, N.; Gao, Y.; Chabal, Y. J. *CrystEngComm* **2014**, *17*, 247.
- (97) Sun, W.; Lin, L.-C.; Peng, X.; Smit, B. *AIChE J.* **2014**, *60*, 2314.
- (98) Serre, C.; Millange, F.; Thouvenot, C.; Noguès, M.; Marsolier, G.; Louër, D.; Férey, G. *J. Am. Chem. Soc.* **2002**, *124*, 13519.
- (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*, 15763.
- (100) Li, J.-R.; Sculley, J.; Zhou, H.-C. *Chem. Rev.* **2012**, *112*, 869.
- (101) Rios, R. B.; Correia, L. S.; Bastos-Neto, M.; Torres, a. E. B.; Hatimondi, S. a.; Ribeiro, A. M.; Rodrigues, A. E.; Cavalcante, C. L.; de Azevedo, D. C. S. *Adsorption* **2014**, *20*, 945.
- (102) Chen, D. L.; Shang, H.; Zhu, W.; Krishna, R. *Chem. Eng. Sci.* **2014**, *117*, 407.
- (103) Saha, D.; Bao, Z.; Jia, F.; Deng, S. *Environ. Sci. Technol.* **2010**, *44*, 1820.
- (104) Salles, F.; Jobic, H.; Devic, T.; Llewellyn, P. L.; Serre, C.; Férey, G.; Maurin, G. *ACS Nano* **2010**, *4*, 143.
- (105) Keskin, S.; van Heest, T. M.; Sholl, D. S. *ChemSusChem*, **2010**, *3*, 879–891.
- (106) Banerjee, R.; Furukawa, H.; Britt, D.; Knobler, C.; O’Keeffe, M.; Yaghi, O. M. *J. Am. Chem. Soc.* **2009**, *131*, 3875.
- (107) D’Alessandro, D. M.; Smit, B.; Long, J. R. *Angewandte Chemie - International Edition*, **2010**, *49*, 6058–6082.
- (108) Adams, R.; Carson, C.; Ward, J.; Tannenbaum, R.; Koros, W. *Microporous Mesoporous Mater.* **2010**, *131*, 13.
- (109) Car, A.; Stropnik, C.; Peinemann, K. V. *Desalination* **2006**, *200*, 424.
- (110) Perez, E. V.; Balkus, K. J.; Ferraris, J. P.; Musselman, I. H. *J. Memb. Sci.* **2009**, *328*, 165.
- (111) Fernandez, M.; Boyd, P. G.; Daff, T. D.; Aghaji, M. Z.; Tom, K. *J Phys Chem Lett* **2014**, *Submitted*, 1087.
- (112) Düren, T.; Bae, Y.-S.; Snurr, R. Q. *Chem. Soc. Rev.* **2009**, *38*, 1237.

- (113) Loiseau, T.; Serre, C.; Huguenard, C.; Fink, G.; Taulelle, F.; Henry, M.; Bataille, T.; Férey, G. *Chemistry* **2004**, *10*, 1373.
- (114) Schneemann, a; Bon, V.; Schwedler, I.; Senkovska, I.; Kaskel, S.; Fischer, R. a. *Chem. Soc. Rev.* **2014**, *43*, 6062.
- (115) Bureekaew, S.; Amirjalayer, S.; Tafipolsky, M.; Spickermann, C.; Roy, T. K.; Schmid, R. *Phys. Status Solidi* **2013**, *250*, 1128.
- (116) Borycz, J.; Lin, L.-C.; Bloch, E. D.; Kim, J.; Dzubak, A. L.; Maurice, R.; Semrouni, D.; Lee, K.; Smit, B.; Gagliardi, L. *J. Phys. Chem. C* **2014**, *118*, 12230.
- (117) Todorov, I. T.; Smith, W.; Trachenko, K.; Dove, M. T. *J. Mater. Chem.* **2006**, *16*, 1911.
- (118) Ewald, P. P. *Ann. Phys.* **1921**, *369*, 253.
- (119) Buckingham, R. A. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **1938**, *168*, 264–283.
- (120) Morse, P. M. *Phys. Rev.* **1929**, *34*, 57.
- (121) Wang, J. M.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. a; Case, D. a. *J. Comput. Chem.* **2004**, *25*, 1157.
- (122) Mayo, S.; Olafson, B.; Goddard, W. *J. Phys. ...* **1990**, *94*, 8897.
- (123) Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. *J. Am. Chem. Soc.* **1992**, *114*, 10024.
- (124) 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*, 8814.
- (125) Breneman, C. M.; Wiberg, K. B. *J. Comput. Chem.* **1990**, *11*, 361.
- (126) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. ...* **1993**, *97*, 10269.
- (127) Campañá, C.; Mussard, B.; Woo, T. K. *J. Chem. Theory Comput.* **2009**, *5*, 2866.
- (128) Vaidhyanathan, R.; Iremonger, S. S.; Shimizu, G. K. H.; Boyd, P. G.; Alavi, S.; Woo, T. K. *Science* **2010**, *330*, 650.
- (129) Iremonger, S. S.; Liang, J.; Vaidhyanathan, R.; Martens, I.; Shimizu, G. K. H.; Daff, T. D.; Aghaji, M. Z.; Yeganegi, S.; Woo, T. K. *J. Am. Chem. Soc.* **2011**, *133*, 20048.
- (130) Fischer, M.; Bell, R. G. *J. Phys. Chem. C* **2013**, *117*, 24446.
- (131) Chen, D.-L.; Stern, A. C.; Space, B.; Johnson, J. K. *J. Phys. Chem. A* **2010**, *114*, 10225.

- (132) Frenkel, D.; Smit, B. *Understanding Molecular Simulation*; 2nd ed.; Academic Press: New York, 2002.

2 Modelling CO₂ adsorption in Experimental MOFs

This chapter discusses calculations performed on two MOFs, CALF-15 and CALF-16[†] detailed in sections 2.4 and 2.5. These studies were performed in collaboration with the Shimizu group at the University of Calgary, and were published in part in *Vaidhyanathan, R.; Iremonger, S. S.; Shimizu, G. K. H.; Boyd, P. G.; Alavi, S.; Woo, T. K. Science* **2010**, 330, 650, and *Vaidhyanathan, R.; Iremonger, S. S.; Shimizu, G. K. H.; Boyd, P. G.; Alavi, S.; Woo, T. K. Angew. Chem. Int. Ed. Engl.* **2012**, 51, 1826, respectively. In these studies, my role was to perform all of the computations and the analysis of those results, which included the detailed binding site analysis, GCMC simulations, and DFT calculations to elucidate the behaviour of CO₂ adsorption in both MOFs. It should be noted that for the first paper (*Science* **2010**, 330, 650), the CALF-15 compound had already been published in reference 1, so much of the novelty of this work was in the detailed analysis of the binding sites.

2.1 Abstract

Understanding the molecular details of CO₂-sorbent interactions is critical for the design of better carbon-capture systems. Here we report joint computational and experimental studies of two high-performing, amine functionalized MOFs; CALF-15 and CALF-16. The crystallographic resolution of CO₂ molecules and their binding domains in CALF-15 permitted an investigation into the nature of binding in these sites. The modeling apportioned specific binding interactions for each CO₂ molecule in their crystallographic binding sites, including substantial cooperative binding effects among the guest molecules. We also demonstrate an unprecedented level of validation between theory and experiment, where the modelled gas sorption isotherms, high heat of adsorption, and CO₂ lattice positions showed high agreement on all three fronts. While CO₂ binding sites could not be resolved in CALF-16, using GCMC simulations and *ab initio* molecular dynamics simulations we were nevertheless able to identify regions of strong CO₂ adsorption, which cooperative effects play a significant role. The validation of the capacity of such simulations to accurately model molecular-scale binding bodes well for the theory-aided development of amine-based CO₂ sorbents. The analysis shows that the

[†] The names of these MOFs were allocated after both studies were published in 2010 (CALF-15) and 2012 (CALF-16). In the original articles, they are referred to by their abbreviated chemical formulae ([Zn₂(Atz)₂(Ox)] and [Zn₃(Atz)₃(PO₄)] respectively).

combination of appropriate pore size, strongly interacting amine functional groups, and the cooperative binding of CO₂ guest molecules is responsible for the low-pressure binding and large uptake of CO₂.

2.2 Introduction

The currently employed CO₂ capture method involves alkanolamine-based solvents that act as CO₂ scrubbers^{2,3} by chemisorptive formation of N-C bonded carbamate species (bonding energies are typically 100 kJ/mol). Regeneration of the amine requires cleavage of this covalent bond by heating (at 100°C to 150°C) to release CO₂. Major drawbacks of this process include the corrosive nature and volatility of the amines, their occasional decomposition, and most prominently, the high energy cost of their regeneration.^{2,3} Merging the inherent sorptive behavior of MOFs with less basic amines offers a route to easy-on/easy-off materials. Less-basic amines would favor physisorption over chemisorption of CO₂, thus greatly reducing the energy of regeneration. This prospect has prompted research on many amine-functionalized solid materials that on the whole demonstrates that amines can enhance CO₂ uptake.^{4–10} Despite this conclusion, experimental insights at a molecular level on the nature of the NH₂···CO₂ interaction are lacking.¹¹ The crystallinity of MOFs enables diffraction experiments to study structure at a molecular level. Beyond characterization of the framework, in exceptional cases, x-ray or neutron diffraction can allow direct visualization of gases within pores^{12–18} to elucidate specific binding interactions and enable better sorbent design. Locating gas molecules in a MOF is challenging, but the systems typically display strong confinement effects on the guest molecules and/or specific sites of strong interaction (such as bare metal sites) that can serve as excellent models for understanding the interactions of gases in all porous systems.

The following describes detailed studies on two amine-functionalized MOFs, CALF-15 and CALF-16 via a combination of crystallographic and computational methods where the nature of CO₂ binding is revealed. In a crystal structure of CALF-15 loaded with CO₂ molecules, the CO₂ binding sites are readily identified, even from room temperature diffraction data. The characteristics of CO₂ uptake in CALF-15, including isotherm, heat of adsorption, and location of CO₂ molecules, are modeled with high accuracy via a combination of classical grand canonical Monte Carlo (GCMC) simulations, molecular dynamics (MD) simulations, and periodic density functional theory (DFT) calculations. The modeling enables the partitioning of CO₂ binding in CALF-15 into components based both on neighboring groups and the nature of interaction (electrostatics/dispersion). We then apply these computational methods to CALF-16 which has CO₂ adsorption properties at 273K and ΔH_{ads} (~32 kJ/mol at zero loading) that are surprisingly lower than that of CALF-15 (ΔH_{ads} =40 kJ/mol), despite its increased pore density of amine groups. The results from classical and *ab initio* simulations reveal that the nature of CO₂ binding is not only dependent on amine functionalization, but also the spatial distribution of these functional groups in the pores, where more does not necessarily mean better performance.

2.3 Computational Details

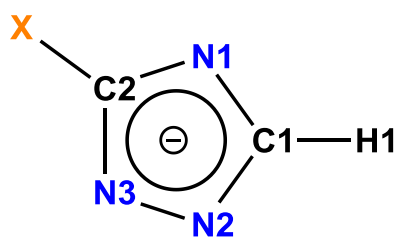
Periodic density functional theory (DFT) calculations used to derive the partial atomic charges, to calculate binding energies, and to perform *ab initio* molecular dynamics were performed with the CPMD software package. The gradient-corrected PBE functional,¹⁹ and the norm-conserving pseudopotentials of Goedecker²⁰ were utilized. The Kohn-Sham orbitals were expanded in plane waves with an energy cutoff of 80 Ry. The empirical dispersion corrections of Grimme²¹ was used for both the energy and force evaluations. Here a scaling factor of 0.75 as parameterized by Grimme specifically for the PBE functional was used. All DFT calculations

used a single unit cell of both CALF-15 and CALF-16 with Γ -point sampling of the Brillouin zone. For the *ab initio* molecular dynamics, the Car-Parrinello method²² was used with a fictitious electronic mass of 400 a.u and a MD timestep of 0.12 fs. A 35 ps NVE simulation is reported which was first equilibrated for 10 ps with a Nosé-Hoover thermostat at 273 K.

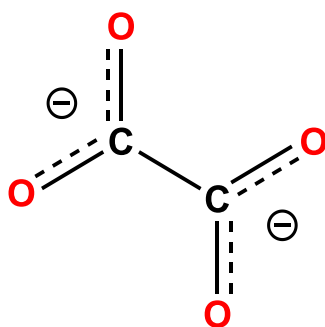
Table 2.1: Lennard-Jones parameters for framework atoms from the UFF force field²³ and CO₂ guest molecules.

Atom	ϵ / kcal/mol	σ / Å
C	0.1050	3.431
O	0.0600	3.118
N	0.0690	3.261
P	0.3050	3.695
Zn	0.1240	2.460
O (CO ₂)	0.0595	2.745
C (CO ₂)	0.1702	3.017

Table 2.2: REPEAT atomic charges used for the framework molecules of CALF-15 and its methylated derivative.



Triazolate



Oxalate

Atom	Charge (e)	
	<i>Aminated</i> (X=NH ₂)	<i>Methylated</i> (X=CH ₃)
<i>Triazolate</i>		
N1	-0.2644	-0.3125
N2	-0.2495	-0.1696
C1	+0.0233	+0.0450
H1	+0.1706	+0.1081
N3	-0.3277	-0.2247
C2	+0.5952	+0.5065
X	-0.8806	-0.7190
H(-X)	+0.3793	+0.2074
<i>Zn</i>	+0.7625	+0.5869
<i>Oxalate</i>		
C	+0.4399	+0.4080
O	-0.4906	-0.4254

Table 2.3: REPEAT atomic charges used for the framework molecules of CALF-16 and its methylated derivative.

Atom	Charge (e)	
<i>triazolate (labels as above)</i>	<i>Aminated (X=NH₂)</i>	<i>Methylated (X=CH₃)</i>
N1	-0.5862, -0.2341, -0.3649	-0.5284, -0.1532, -0.3669
N2	-0.0797, -0.2204, -0.5072	+0.0110, -0.3561, -0.3967
C1	+0.3116, -0.0529, -0.0412	+0.5817, +0.5944, +0.6885
H1	+0.1038, +0.1439, +0.1295	+0.1191, +0.0662, +0.0240
N3	-0.6336, -0.2147, -0.0924	-0.5591, -0.3370, -0.2621
C2	+0.7122, +0.4410, +0.7099	+0.2480, +0.0314, +0.1450
X	-0.8716, -0.7699, -0.8576	-0.8535, -0.7533, -0.8276
H(-X)	+0.3645, +0.3496, +0.3854	+0.2387, +0.2000, +0.2497
Zn	+0.6738, +0.8118, +0.9730	+0.7198, +0.7677, +0.9347
<i>phosphate (PO₄³⁻)</i>		
P	+1.5183	+1.2560
O	-0.7968	-0.6817
O	-0.7994	-0.7427
O	-0.7888	-0.7116
O	-0.8490	-0.7220

Classical molecular dynamics and grand canonical Monte Carlo (GCMC) simulations were performed with the force field parameters provided in Table 2.1 and Table 2.2 and described below. Since the MOF framework is frozen in GCMC simulations, only non-bonded interaction parameters were required for the framework. Two-body, additive, atom centered Lennard-Jones and Coulomb potentials were used for this purpose, as is common practice. The Lennard-Jones parameters for the framework atoms were taken from the Universal Force Field (UFF)²³ and are given in Table 2.1. Electrostatic potential (ESP) derived charges for used based on our in-house developed REPEAT method²⁴ which allows for ESP charges to be determined from periodic DFT calculations. Unrestrained REPEAT charges were derived from planewave periodic DFT calculations as described above, using the experimental x-ray structures where the hydrogen atoms of CALF-15 and CALF-16 have been optimized. Van der Waals radii used to

define the valid grid points for the REPEAT fit were taken from the UFF force field. For the methyl substituted frameworks of CALF-15 and CALF-16, the geometries of the methyl groups and all hydrogen atoms were optimized at the DFT level. Van der Waals radii used to define the valid grid points for the REPEAT fit were taken from the UFF force field. Charges used for the classical simulations are given in Table 2.2 and Table 2.3. Charges were assigned to unique atoms within the cell. In the case of CALF-16 and methyl functionalized CALF-16, the crystallographic unit cell contains three unique aminotriazolate groups, three unique zinc ions, and one unique phosphate group. This is reflected by the unique charges assigned to them in Table 2.3.

The non-bonded parameters of CO₂ were taken from García-Sánchez *et al.*²⁵ who optimized them from a self-consistent fitting of the experimental adsorption isotherm data in a large class of zeolite frameworks with the predictions of Gibbs ensemble Monte Carlo calculations on these systems. The Lennard-Jones parameters for CO₂ are given in Table 2.1. The bond length (1.149 Å) and partial charges on CO₂ atoms (C = +0.6512e, O = -0.3256e) in this force field are from the Harris and Yung potential.²⁶

Grand Canonical Monte Carlo simulations were performed with an in-house developed code. At each external gas pressure, a total of 10⁷ Monte Carlo steps were generated after initial equilibration. In the present simulations, only pressures less than 1 bar are simulated. Under these conditions, we use the ideal gas model to define the chemical potential in the Monte Carlo guest insertion and deletion criteria. The Monte Carlo routine used equal probabilities for the processes of guest displacement, insertion, and deletion. This approach was applied to study zeolitic imidazolate MOFs and reproduced the results of other GCMC simulations published in the literature.²⁷ In the GCMC simulations of CALF-15 and CALF-16, the MOF framework is

frozen to the x-ray structure for the non-hydrogen atoms. The hydrogen atom positions in the MOF framework were optimized at the DFT level as specified above. 2×2×2 unit cell replicas of the unit cell was used for the GCMC simulations of CALF-15 while a 3×3×3 unit cell was used for CALF-16. The same cell sizes were used for GCMC simulations of their methyl substituted derivatives.

Classical MD simulations were performed with the DL_POLY package,²⁸ using the force fields described above. The equations of motion for the guest motions were integrated with a time step of 1.0 fs using the Verlet leapfrog algorithm and a Nosé-Hoover thermostat relaxation time of 0.1 ps. A 100 ps thermostated equilibration period was found to be sufficient to obtain converged simulation energies. Long-range electrostatic interactions were calculated using the Ewald summation method with a precision of 1×10^{-6} and all intermolecular interactions in the simulation box were calculated within a cutoff distance of $R_{\text{cutoff}} = 12.0 \text{ \AA}$. Different loading numbers of CO₂ molecules were placed in the pores of the MOF frameworks to analyze the binding enthalpies of the guests. The binding enthalpies for the MOFs with $n\text{CO}_2$ molecules per unit cell is defined as,

$$E_{\text{binding}} = E(\text{MOF} + n\text{CO}_2) - E(\text{MOF}) - nE(\text{guests}) + RT \quad 2.1$$

where potential energies for the MOF + $n\text{CO}_2$ guest simulations are used for $E(\text{MOF} + n\text{CO}_2)$. If the CO₂ gas is assumed as ideal, the $E(\text{guests})$ term in Eqn. 2.1 is zero. The potential energy for the empty frozen framework $E(\text{MOF})$ term for the frozen framework is determined from a simulation of the systems with a single non-interacting gas molecule in the system. Although the MOF framework is frozen in the molecular dynamics calculations, there are still interatomic van der Waals and electrostatic interactions between the atoms that constitute the MOF which contribute to the absolute value of the system-wide configurational (potential) energy. This

potential energy is considered a baseline value, but it is not zero. This baseline energy cannot be calculated directly by the DL_POLY code as a simulation of the frozen framework with no mobile atoms leads to error. To determine this baseline energy we added a single non-interacting atom, designated as a “ghost” to the simulation cell. The potential energy content of this reference simulation is the potential energy of the framework, $E(\text{MOF})$.

The $\text{CO}_2\text{-CO}_2$ interaction energies from the DFT calculations were partitioned into dispersion and electrostatic components. In the DFT calculations, the dispersion interactions are accounted for by an empirical pair-potentials developed by Grimme²¹ that are similar to the attractive component of a Lennard-Jones potential. As a result, the dispersion contribution to the interaction energy can be easily separated from the total. Further, if we assume that there is no charge transfer or orbital interaction between the molecules, we can consider the interaction energy from the DFT calculation without the dispersion correction to be due to electrostatics. This is a good approximation given the intermolecular CO_2 distances. Binding energies calculated at the DFT level were performed with full relaxation of the CO_2 and framework geometries, but with the lattice vectors fixed to the experimental values.

The amine- CO_2 electrostatic interaction was estimated using a point charge model and the partial atomic charges used in the GCMC simulations. These charges are electrostatic potential fitted charges extracted from the periodic DFT calculations of the empty framework of CALF-15 using the REPEAT method.²⁴ For the reported amine- CO_2 electrostatic interaction energies, the DFT optimized geometries were used. To show that the classical force field reproduces the interaction energies calculated by DFT, we have calculated the total CO_2 -host interaction energy using the aforementioned force field using the DFT optimized geometry of a single CO_2 molecule in site I of CALF-15. This gives a total binding energy for $\text{CO}_2\text{-I}$ of 42.6

kJ/mol which agrees well with the calculated DFT binding energy of 39.5 kJ/mol. Moreover, the fraction of the CO₂-I binding energy that can be attributed to electrostatics calculated with the classical force field is also in good agreement with the DFT calculations. Specifically, 19% of the CO₂-I binding energy can be attributed to electrostatic interactions with the classical force field, which is in good agreement to the 18% found in the DFT calculations.

To calculate the ensemble averaged CO₂ interaction energies, the positions of the CO₂ molecules from the GCMC simulation were extracted and placed in an equivalent simulation cell but without the framework. The interaction energy of these molecules was then calculated with the classical force field described above and averaged.

2.4 CALF-15

The previously reported¹ MOF CALF-15 containing Zn(II) ions coordinated to 3-amino-1,2,4-triazole (Atz) and oxalate (ox) moieties [Zn₂(Atz)₂(ox)]. The crystal structure of CALF-15 is shown in Figure 2.1B. Zn triazolate layers (Figure 2.1A) are bridged by oxalate groups to form the 3D periodic structure. This creates channels with a cross section of $3.5 \times 4.0 \text{ \AA}$ along the *a*-axis (Figure 2.1B). These channels are cross-linked by smaller channels ($3.9 \times 2.1 \text{ \AA}$) along the *b*-axis and *c*-axis ($3.0 \times 1.6 \text{ \AA}$) to produce cube-shaped spaces of approximate dimensions $4.0 \times 4.0 \times 4.0 \text{ \AA}^3$. These are accessible via a 2D channel in the *ab*-plane. Notably, the free amine groups protrude into the pores giving a zig-zag arrangement to the largest pore down the *a*-axis. The x-ray resolved binding sites of CO₂ are reported in Figure 2.1C where two unique molecules are depicted with atoms C100, O100, and O101 (binding site I) and C200, O200, and O201 (binding site II). The experimental isotherm measurements shown in Figure 2.2A demonstrate high CO₂ uptake at low pressures with a high initial heat of adsorption (ΔH_{ads}) of ~40 kJ/mol. After manifesting the expected trend of decreased ΔH_{ads} with increased coverage, the material

showed a subsequent increase in ΔH_{ads} , which remained over 35 kJ/mol during continuous gas exposure, suggesting a cooperativity effect in the binding mechanism.

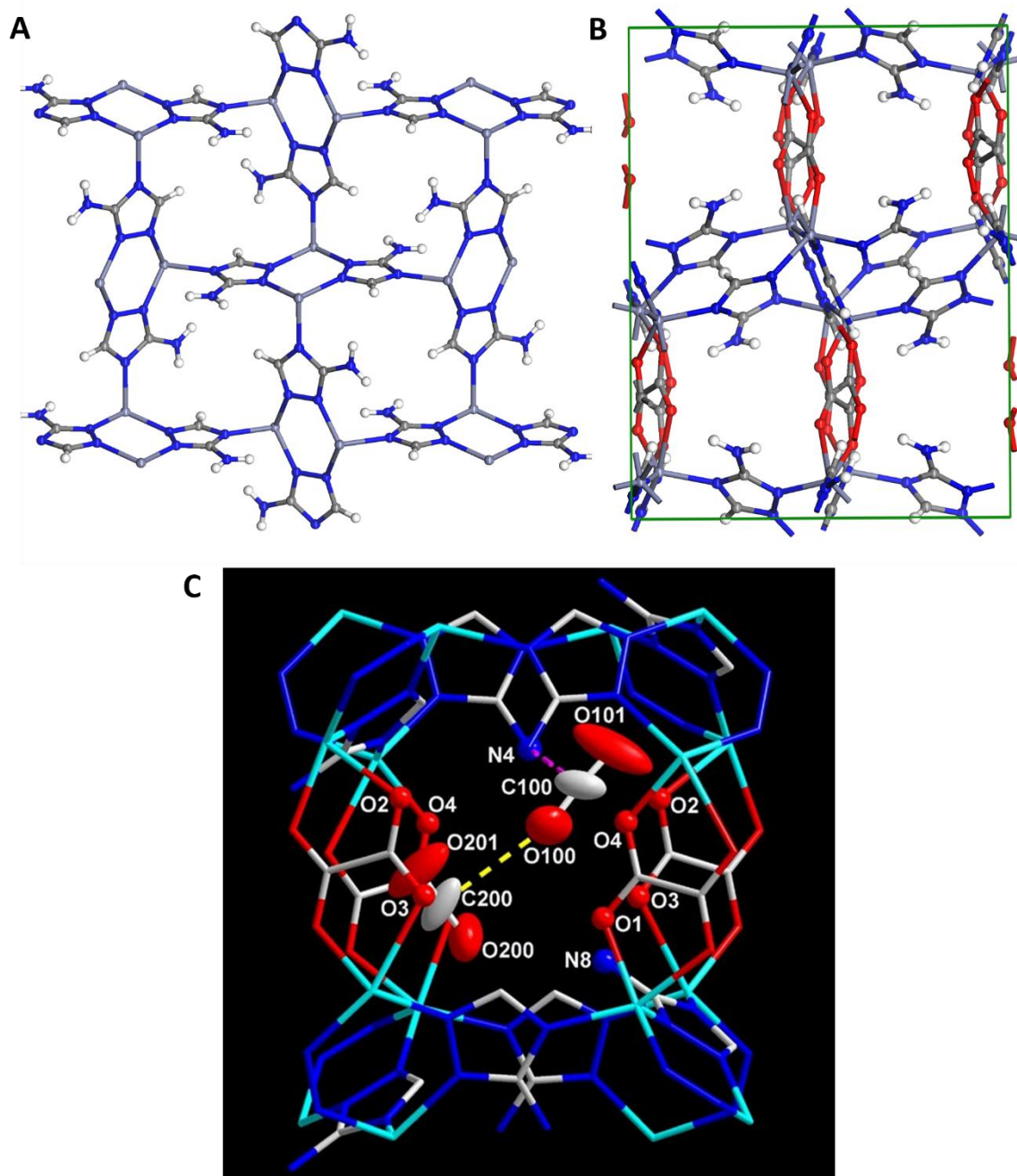


Figure 2.1: A) Zn-aminotriazolate layers in CALF-15. The layer is made up of Zn₂ dimers in 90° orientations with respect to each other. B) The full structure of CALF-15 where oxalate units bridge the Zn-aminotriazolate layers to form the 3D structure. C) X-ray resolved structure of CALF-15·(CO₂)_{1.3} showing two crystallographically independent CO₂ molecules trapped in a pore. The purple dotted bond between the C_{CO2} and N_{amine} indicates a close interaction with the amine nitrogen, spaced 3.16 Å apart, while the dotted yellow bond representing CO₂···CO₂ interactions is spaced 3.02 Å from carbon to oxygen.

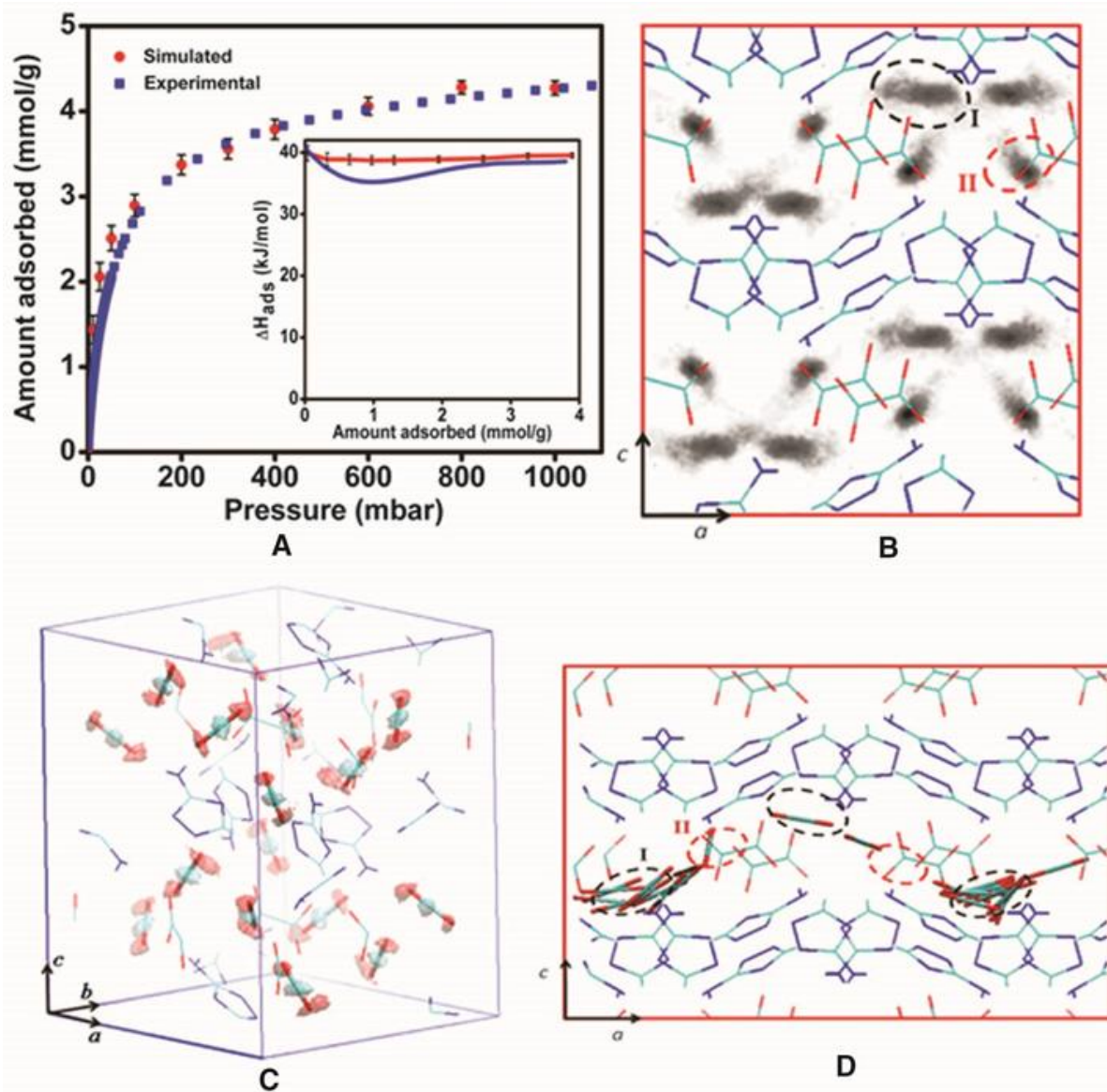


Figure 2.2: A) Comparison of the simulated (red) and experimental (blue) CO₂ gas adsorption isotherm of CALF-15 at 273 K. The inset shows a comparison of experimental and simulated heats of adsorption as a function of guest loading. B) Center-of-mass probability density plots of CO₂ molecules in CALF-15 from a GCMC simulation at 850 mbar and 273 K. Probabilities along the b axis are summed and projected onto the ac plane. C) Comparison of the location of the CO₂ binding sites in CALF-15 obtained from x-ray analysis and those predicted by simulation. Probability isosurface plots of the CO₂ O atoms are shown as transparent red surfaces and those of the C atoms as transparent cyan surfaces. The 16 experimentally determined CO₂ molecule positions are shown as tubes. An isosurface value of 0.15 Bohr⁻³ is used. D) Trace of a single CO₂ molecule during a MD simulation of CALF-15 with five CO₂ molecules per unit cell or a loading of 1.6 mmol/g. Shown are 23 successive snapshots, separated by 62.5 ps. Other CO₂ molecules present in the simulation are not shown. In (B) and (D), the approximate locations of binding sites I and II in a single channel are shown with dotted ellipses (site I, black; site II, red). In (B) and (C), a single unit cell of CALF-15 is shown, whereas in (D) a 2×1×1 cell is depicted. In (B) to (D), the MOF framework is shown as lines and the CO₂ molecules are shown as tubes (C, light blue; O, red; N, dark blue; Zn, not shown).

To investigate the nature of the CO₂ interactions with CALF-15 and the cooperative guest-guest binding effects, we used a combination of dispersion-corrected²¹ periodic DFT calculations and classical GCMC simulations and MD simulations²⁹, in which the partial charge parameters were derived from the periodic DFT calculations by means of the REPEAT method.²⁴ Figure 2.2A displays the excellent agreement between the experimental and GCMC simulated CO₂ adsorption isotherms of CALF-15 at 273 K. The parameters associated with the CO₂-CALF-15 intermolecular potential were not adjusted to obtain this quality of fit. Figure 2.2B shows a center-of-mass probability density plot from a GCMC simulation of CALF-15 (at 850 mbar and 273 K), where darker regions reveal a greater probability of finding CO₂. The binding sites are well localized, and the symmetry of the probability clouds suggests two distinct binding sites, corroborating the crystallography. To compare the CO₂ binding sites determined computationally and crystallographically, the experimental CO₂ positions are superimposed on three-dimensional isosurfaces of the probabilities for C and O in Figure 2.2C. The remarkable agreement between the simulated and experimental CO₂ positions suggests that GCMC simulations that are often used to study gas adsorption in MOFs^{29,30} not only reproduce adsorption isotherms but can also accurately reproduce specific binding sites. The nature of the CO₂ binding was further studied with dispersion-corrected periodic DFT calculations. With all 16 binding sites in the unit cell occupied, the resulting fully optimized geometry is in excellent agreement with the x-ray structure, including the relevant CO₂-amine and CO₂-oxalate distances (Figure 2.1C). The exception to this was the geometries of both CO₂ molecules, which, unlike their geometry in the x-ray analysis, were optimized to linear configurations, revealing that there is minimal geometric distortion from interaction with the framework.

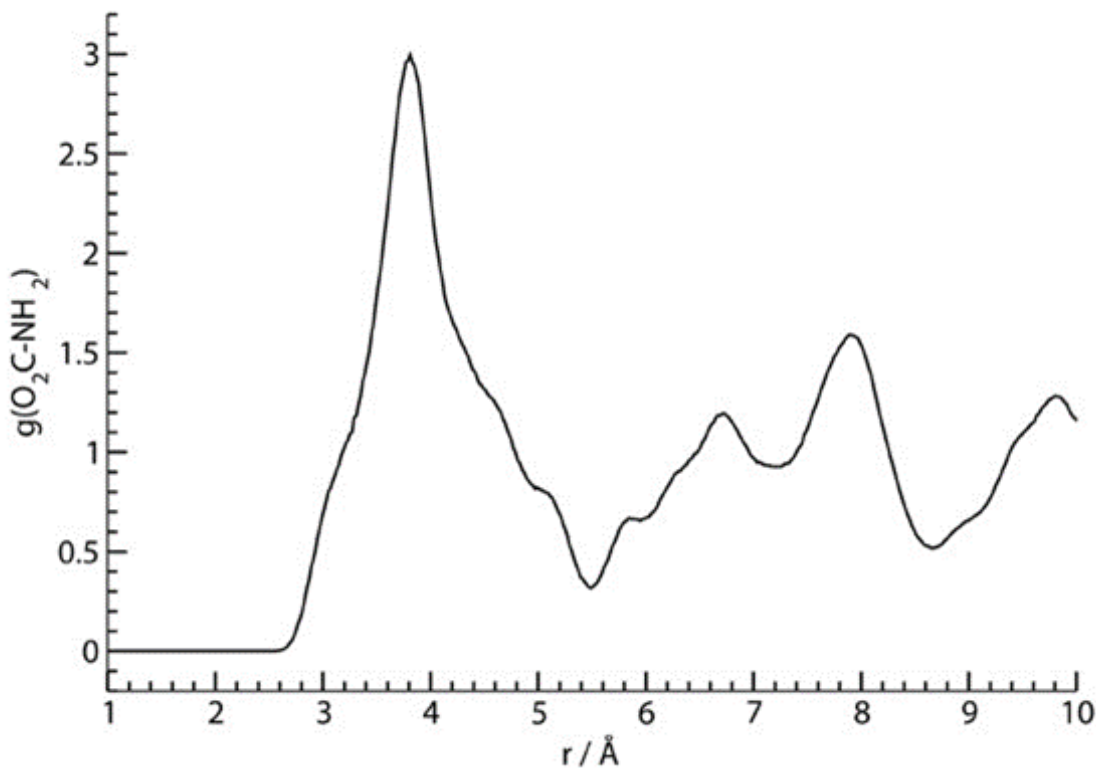


Figure 2.3: The carbon-nitrogen radial distribution functions for the $\text{NH}_2\text{-CO}_2$ guest-host interactions in CALF-15 extracted from the GCMC simulations at 273 K at 0.85 bar.

Ensemble averaged geometric details can be obtained from radial distribution plots (RDF) extracted from the GCMC simulations. These results are also in agreement with the crystallographic results. The first peak of the $\text{N}_{\text{amine}}\text{-C}_{\text{CO}_2}$ RDF given in Figure 2.3 shows a shoulder at ~ 3.2 Å, in agreement with the $\text{N}_{\text{amine}}\text{-C}_{\text{CO}_2}$ distance of 3.16 Å found crystallographically for $\text{CO}_2\text{-I}$. The first large peak at ~ 3.9 Å corresponds to the $\text{CO}_2\text{-I}$ which is positioned 3.93 Å and 3.94 Å from a second and third amine in the pore, and $\text{CO}_2\text{-II}$ whose closest amine is positioned 3.90 Å away. Although specific binding sites were located experimentally, the CO_2 binding is expected to be dynamic in nature. This prospect has been investigated via classical MD simulations, which show that, in a 1.4-ns time span, the CO_2 molecule can hop between several of the binding sites I and II (Figure 2.2D).

The CO₂ binding energies were calculated at various occupancies at the DFT level. When CALF-15 is empty, CO₂-I has a binding energy of 39.6 kJ/mol, which is in good agreement with the experimental zero-loading ΔH_{ads} (40.8 ± 0.8 kJ/mol). We see a strong cooperative enhancement of CO₂ binding that increases with loading. Specifically, the binding energy of CO₂-II increases by 4.6 kJ/mol to 37.0 kJ/mol when an adjacent site I is occupied. When CALF-15 is fully occupied less one binding site, the binding energy for CO₂-II increases to 38.1 kJ/mol, whereas that of CO₂-I is 44.2 kJ/mol. Further insight into the nature of the CO₂ binding can be gained by partitioning the total binding energy. This analysis reveals that 82% of the binding energy of site I is due to dispersion, whereas 18% results from electrostatics. This contrasts with site II, where the binding energy is almost entirely (99%) due to dispersion interactions. These observed DFT calculated binding energies as a function of CO₂ loading can be used to qualitatively explain the trend in the experimental ΔH_{ads} measured in Figure 2.2A (inset). At zero loading, the binding enthalpy was measured to be 40.8 kJ/mol but then drops to a value of approximately 35 kJ/mol at a loading of 1.0 mmol/g. It then slowly increases to a value of 38.6 kJ/mol at maximum loading. At very low loading, CO₂-II is not thermally accessible with a Boltzmann population of less than 4% at 273 K compared to that of CO₂-I. Thus, the measured binding enthalpy is due mostly to binding in site-I which we have calculated to be approximately 40 kJ/mol. However, as CO₂-I begins to be occupied, CO₂-II sites become more thermally accessible due to the strong cooperative effect between adjacent binding sites. As a result, since CO₂-II is weaker the average binding energy decreases from that of the zero-loading value. As the number of CO₂-II sites become increasingly populated, stronger cooperativity effects increase the ΔH_{ads} which is confirmed at the DFT level.

The cooperative binding in CALF-15 can be attributed to a combination of dispersion and electrostatic interactions between CO₂ molecules, which can be quantified by using DFT to evaluate the interaction energy between two CO₂ molecules at sites I and II in a vacuum. This calculation gives a CO₂-CO₂ interaction energy of 3.9 kJ/mol, which accounts for most of the observed 4.6 kJ/mol binding enhancement. Of the 3.9 kJ/mol interaction between CO₂ molecules, 66% can be attributed to dispersion and 34% can be assigned to electrostatics. The same analysis, with the MOF fully loaded, attributes 61% of the interaction to dispersion and 39% to electrostatics, suggesting similar cooperativity at higher loadings.

2.5 CALF-16

The experimental data reported in Section 2.4 permitted an unprecedented level of validation between computational methods and experiment. Herein we apply these methods to understanding CO₂ uptake in another MOF, CALF-16 consisting of Zn(II) ions coordinated to 3-amino-1,2,4-triazole (Atz) and trianionic phosphate groups (PO₄) [Zn₃(Atz)₃(PO₄)], that intuitively should give better CO₂ capture properties. In comparison to CALF-15, only two-thirds of the number of trianionic phosphate groups are required to charge compensate [Zn(Atz)]⁺ layers, so larger, amine-lined pores were anticipated and observed. Despite this, the CO₂ uptake (at 273 K) and heat of adsorption do not exceed those of CALF-15. The computational methods provide crucial insight to understanding these phenomena and demonstrate the wide spread applicability of such techniques to ascertain binding details in MOFs not directly accessible by experiment. Although the role of the amine functionalities in CALF-16 is surprisingly diminished, the cooperative interactions between CO₂ molecules are found to augment overall binding by over 7 kJmol⁻¹, a significant result for CO₂ capture in any porous material.

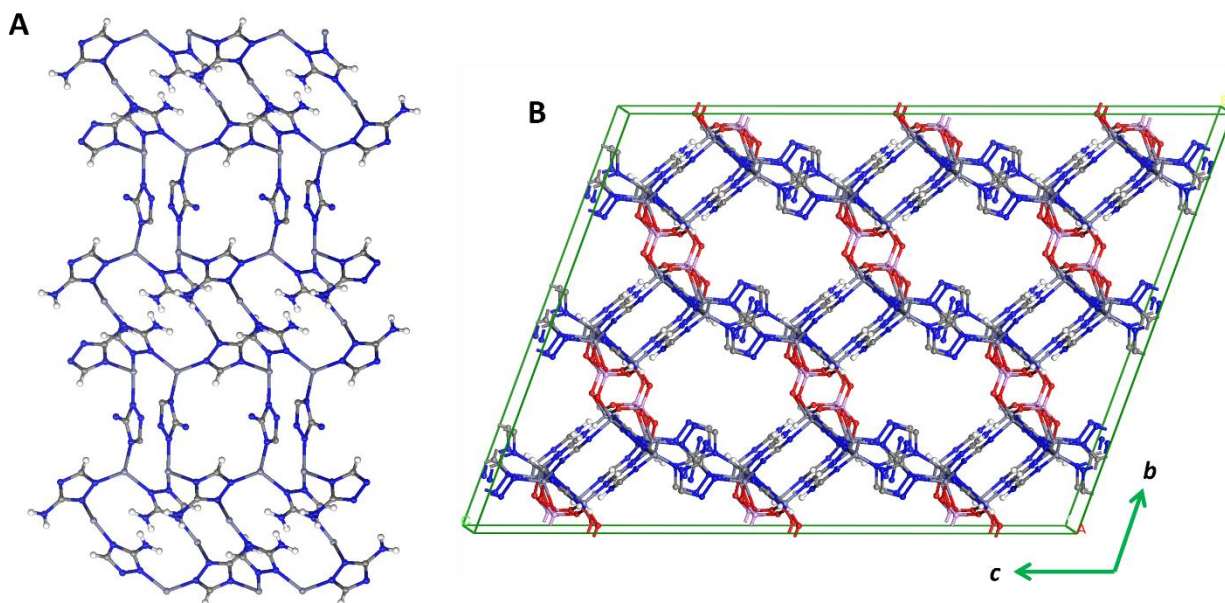


Figure 2.4: Structure of CALF-16. Ball-and-stick representations showing the A) zinc–aminotriazolate layer and B) the structure, complete with bridging phosphate groups showing the juxtapositioning of the Atz ligands. C gray, N dark blue, O red, P purple, Zn pale blue.

CALF-16 is made up of cationic Zn–Atz layers pillared by PO_4 anions to form a 3D porous network (Figure 2.4). The $\text{Zn}(\text{Atz})$ layers lie in the ac plane and contain three independent Zn ions and Atz ligands. No amine groups coordinate to Zn ions; ligation is exclusively through triazole nitrogen atoms. Pillaring of these layers by the phosphate ions results in a 3D network of pores (accounting for van der Waals radii: $a=4.40 \times 6.55$, $b=2.45 \times 2.78$ $\text{\AA}^2$; $[011]$ direction $= 2.66 \times 2.56$ $\text{\AA}^2$). Experimental and simulated CO_2 thermodynamic data are reported in Figure 2.5. The heat of adsorption for CO_2 , calculated using a Virial model using the 263 and 273 K adsorption isotherms was 32 kJ/mol at zero loading (inset Figure 2.5). This value is higher than most non-amine modified MOFs but significantly lower than the 40.8 kJ/mol observed in CALF-15.

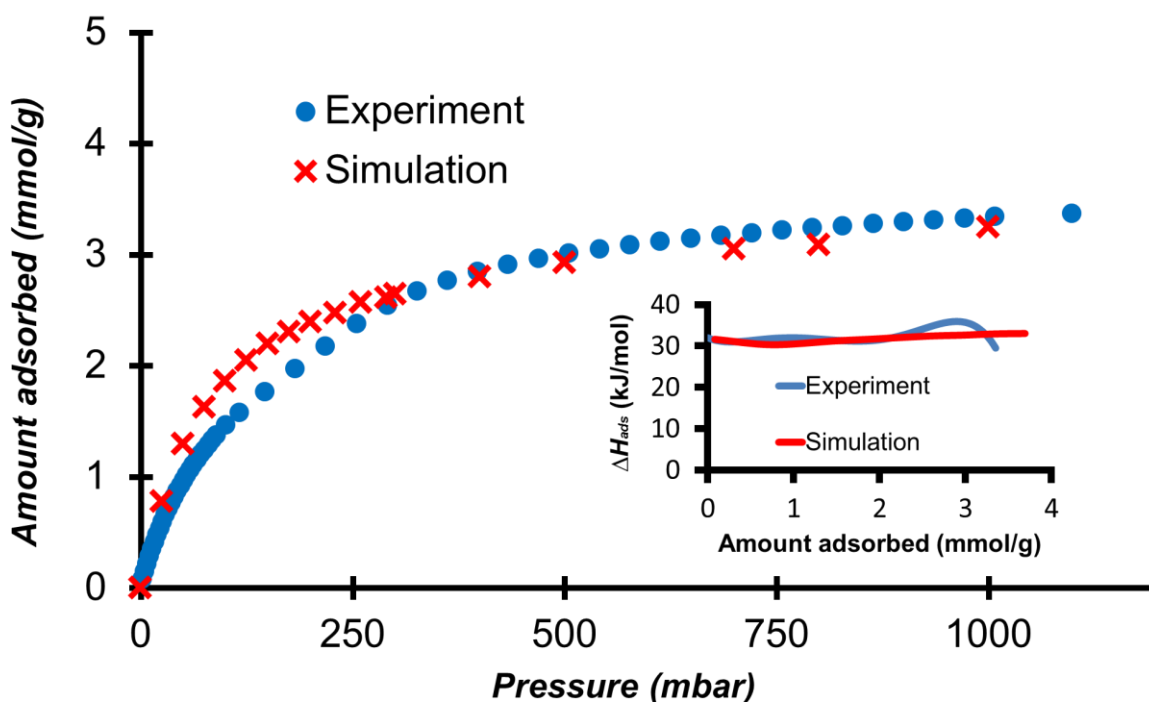


Figure 2.5: Sorption isotherm of CALF-16 for CO₂ at 273 K (experiment, ○; simulated, ×). Inset: Comparison of the simulated (determined from MD simulations at 273 K) and experimental (obtained from isotherms at 263 and 273 K) CO₂ enthalpy of adsorption of CALF-16 as a function of guest loading.

The CO₂ uptake is considerably higher than observed in other reported Atz MOFs,³¹ but not as high as would have been expected in comparison to CALF-15. The pores in CALF-16 are larger than those in CALF-15, and given that CALF-16 retains the available amine groups to enhance framework CO₂ interactions, higher CO₂ uptake was expected. Attempts to observe CO₂ crystallographically in CALF-16 were not successful. A key structural feature extracted from the XRD data of the pure phase CALF-16 was the buckling or staggered conformation of the ZnAtz layers (see Figure 2.4). The ZnAtz layer in CALF-16 is corrugated, leading to juxtapositioning of adjacent Atz molecules in an antiparallel fashion. This results in the amines of CALF-16 not protruding significantly into the pores. Quantitative analysis of the lower CO₂ uptake and ΔH_{ads} in CALF-16 was provided computationally.

The CO₂ uptake of CALF-16, including isotherm, ΔH_{ads} , and location of CO₂ molecules, are modeled by a combination of classical grand canonical Monte Carlo (GCMC) simulations, molecular dynamics (MD) simulations, and periodic density functional theory (DFT) calculations.^{24,29,30,32} The inset in Figure 2.5 compares experimental and simulated CO₂ adsorption isotherms for CALF-16 at 273 K. The simulated binding enthalpies shown in Figure 2.5 are calculated from the difference in the average potential energy resulting from 500 ps MD simulations. The overall agreement is excellent, but the simulated isotherm predicts slightly higher uptake than observed experimentally at low pressure. Figure 2.6a,b show center-of-mass probability-density plots of CO₂ resulting from a GCMC simulation of CALF-16 at 850 mbar and 273 K. The binding is dispersed in two regions, denoted α and β in Figure 2.6. The α regions are roughly parallel to the *ac* plane and located near the phosphate groups, while the β sites are roughly in the *bc* plane (the ZnAtz layer). The probability plots reveal that CO₂ molecules are not strongly localized, corroborating that CO₂ could not be located crystallographically in CALF-16 as compared to CALF-15.

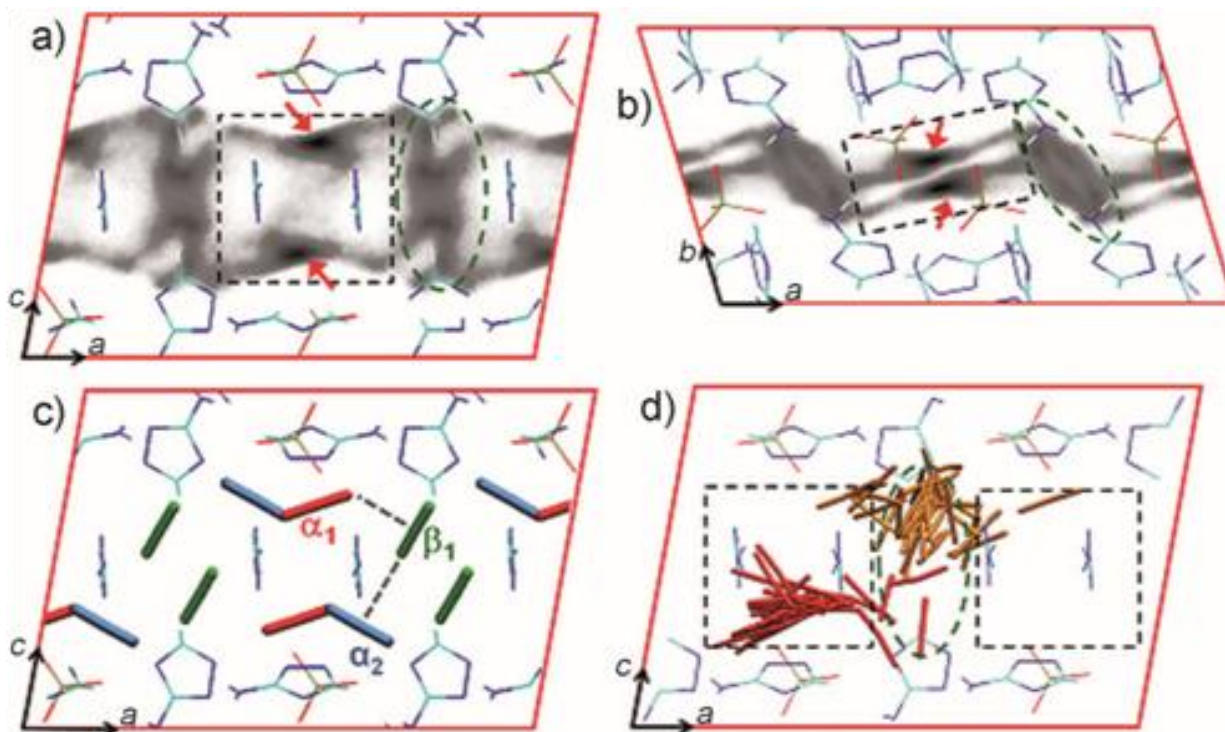


Figure 2.6: Centre-of-mass probability-density plots of CO₂ molecules in CALF-16 at 273 K and 850 mbar pressure. Black dashed boxes: α region; green circles: β region. Shown are probability densities that are a) projected onto the ac plane, and b) projected onto the ab plane. c) Selected CO₂ binding-site geometries optimized at the DFT level. Symmetry-equivalent CO₂ molecules are represented in the same color. d) Trace of two CO₂ molecules (red and orange) during a 35 ps ab initio MD simulation of CALF-16 at 273 K with a loading of four CO₂ molecules per unit cell (the other two CO₂ molecules are not shown). Thirty snapshots, separated by 1.2 ps, are depicted. For (a)–(d), a $2 \times 1 \times 1$ representation of the unit cell is shown that is shifted by 0.5 in the a direction in (d).

To locate the binding sites, CO₂ positions from the high probability regions were extracted from the GCMC simulations and optimized with dispersion corrected periodic DFT calculations. Three of these sites are given in Figure 2.6c, where symmetry-equivalent sites are color-coded. The strongest binding site, α_1 , matches the region of highest probability from the GCMC simulations (red arrows in Figure 2.6a). Site α_1 was determined to have an empty framework binding enthalpy of 30.6 kJ/mol calculated with DFT, in good agreement with the experimental zero loading ΔH_{ads} of 32 kJ/mol, but lower than the 39.6 kJ/mol calculated for the strongest binding site of CALF-15.

As the design premise of using the Atz ligands is that the amine groups enhance the CO₂ uptake, a widely accepted hypothesis, the amine–CO₂ binding was examined in more detail. Site α_1 has three amines in proximity with H_{amine}–O_{CO2} distances of 2.66, 2.83, and 3.26 Å. Using partial atomic charges derived from DFT calculations,²⁴ the electrostatic interactions between the amines with the CO₂ in a given binding site can be estimated. Interestingly, for α_1 , this amine–CO₂ electrostatic interaction is found to be only -0.44 kJ/mol. For comparison, in CALF-15, the amine–CO₂ distances in the strongest binding site are longer (H_{amine}–O_{CO2} = 2.72, 3.10, 3.68 Å), yet the amine–CO₂ electrostatic interaction is considerably more stabilizing at 5.4 kJ/mol. This is explained considering that, for α_1 in CALF-16, the CO₂–amine electrostatic interaction was attractive for two of the nearby amines (-0.02 and -2.39 kJ/mol) but repulsive for the other (+1.57 kJ/mol). For comparison, in CALF-15, all CO₂–amine interactions were attractive.

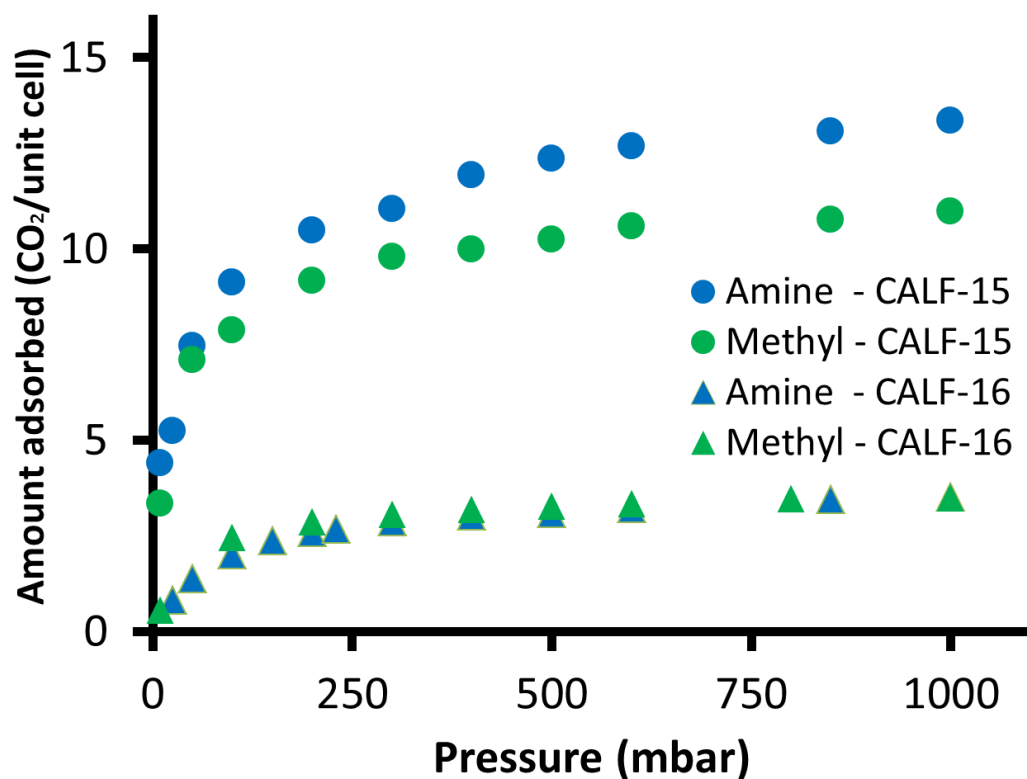


Figure 2.7: Simulated CO₂ adsorption isotherms for the amine-functionalized MOFs CALF-15 (blue ○), CALF-16 (blue Δ), and their respective calculated methyl-for-amine substituted derivatives (green) at 273 K.

The analysis of the amine–CO₂ binding suggests that the densely grouped amines in CALF-16 interfere with each other’s ability to bind CO₂. It is important to note that this is only for a single binding site in CALF-16, and the binding sites are not as localized as in CALF-15. Nevertheless, the results suggest that the role of the amines in CO₂ binding in CALF-16 is significantly diminished compared to CALF-15. To test this, we have simulated the isotherms of CALF-15 and 16 replacing the amine groups with methyl groups in GCMC calculations. Methyl groups are isoelectronic with primary amine groups and so should have similar dispersion interactions, and are similar in size. However, as the two groups have different electron donating abilities, the resulting charge distributions should be quite distinct. Figure 2.7 shows the effect on CO₂ uptake of substituting the amines with methyl groups. In CALF-16, there is negligible difference in uptake upon substitution, consistent with the notion that the amines do not

significantly contribute to the CO₂ binding. On the other hand, the same substitution gives a substantive decrease in CALF-15 where CO₂ uptake decreases by about 20% over the pressure range examined.

Cooperative effects between CO₂ molecules have been recognized as contributing significantly to the overall heat of adsorption of CO₂,^{33,34} particularly by Snurr *et al.*,³³ and key to interpreting adsorption isotherm features. These effects were found to be significant in CALF-15 and so were examined in CALF-16 by studying the DFT-optimized binding sites. Cooperative binding is evinced in the strongest binding site in the β region of CALF-16 (Figure 2.6c, labeled β_1). With an empty framework, the binding energy of β_1 is 29.0 kJ/mol at the DFT level. This energy increases to 32.0 kJ/mol when an adjacent α_1 site is occupied, implied by one of the dashed lines in Figure 2.6c. In examining the output configurations of the GCMC simulations, we found that an interesting triad of CO₂ molecules can form involving α_1 , β_1 , and α_2 (blue in Figure 2.6c). Within an empty framework, α_2 has a binding energy of 26.9 kJ/mol. However, as an $\alpha_1/\alpha_2/\beta_1$ triad, the average binding energy of the triad is 31.3 kJ/mol per CO₂ molecule. This is 7.4 kJ/mol more than the sum of the empty pore binding energies of α_1 , α_2 , and β_1 . Adjacent α_1 and α_2 sites are mutually exclusive in that both cannot be occupied by CO₂ at the same time, shown as blue and red overlap in Figure 2.6c. Thus, it was thought that a CO₂ molecule in site α_1 (the most stable site) might occasionally slip into an adjacent α_2 , to benefit from the favorable $\alpha_1/\alpha_2/\beta_1$ triad interactions. The stabilization imparted by an appropriately oriented T-shaped dimer of CO₂ molecules discussed in the previous section for CALF-15 was estimated to be 3.9–4.6 kJ/mol. The value of 7.4 kJ/mol for a triad can lead one to postulate that appropriately oriented higher aggregates (T-shapes can further assemble into pinwheel tetrads or even infinite herring-bone arrays) will demonstrate pronounced cooperativity and enhanced heats of

adsorption for CO₂. To examine the general mobility of CO₂ molecules in the pores, MD simulations of CALF-16 at 273 K were performed.

As the unit cell of CALF-16 is small, *ab initio* MD simulations were performed on a single unit cell with four CO₂ molecules at the same DFT-D level of theory used to evaluate the binding energies. Figure 2.6d shows two of four positions of CO₂ molecules resulting from a 35 ps MD simulation. The red CO₂ was initially in the α_1 site, whereas orange CO₂ was initially in the β region. Thirty successive snapshots, separated by 1.2 ps, are depicted. Figure 2.6d shows that the red CO₂ is generally localized to the α_1 site and the snapshots map a region similar to that in the GCMC probability distributions in Figure 2.6a. The trajectory shows that the CO₂ does slip into the α_2 site and even into the β region during the short simulation. When CO₂ slips into the α_2 site, snapshots of the MD simulation (not shown) indeed show the $\alpha_1/\alpha_2/\beta_1$ triad forming.

2.6 Conclusions

Several key insights obtained from the detailed analysis of the binding interactions in CALF-15 and CALF-16 have implications for the design of future materials used for the physisorption of CO₂. The relatively high binding energies observed in each material are dominated by dispersion interactions. Because CO₂ has a substantial quadrupole moment, there is opportunity to further increase the binding energies through appropriately designed binding sites that maximize the electrostatic interactions with CO₂. This is still a challenging task, however, as the results from analysing representative binding sites from CALF-16 suggest that the 3D arrangement of functional groups (which were shown to be favourable in CALF-15) have subtle effects that can hinder CO₂ binding.

The importance of the cooperative guest binding to the uptake of CO₂ in both MOFs represents another key insight. In particular, the proper mutual orientation and high density of binding sites can be used as a strategy to increase CO₂ binding energies. Further, we believe that these strategies can be incorporated into materials with larger pores, in order to increase the overall CO₂ uptake capacity and binding energies, thereby improving the uptake properties in the important low-partial-pressure regime. In the broadest sense, via detailed modeling, this study lays the groundwork for the design of easy-on/easy-off physisorptive materials for CO₂ capture, designed from the characteristics of the guest molecules outward rather than from the host framework inward.

2.7 References

- (1) Vaidhyanathan, R.; Iremonger, S. S.; Dawson, K. W.; Shimizu, G. K. H. *Chem. Commun. (Camb)*. **2009**, 5230.
- (2) Rochelle, G. T. *Science* **2009**, 325, 1652.
- (3) Arstad, B.; Blom, R.; Swang, O. *J. Phys. Chem. A* **2007**, 111, 1222.
- (4) Belmabkhout, Y.; Sayari, A. *Adsorption* **2009**, 15, 318.
- (5) Kim, S.; Ida, J.; Gulians, V. V.; Lin, J. Y. S. *J. Phys. Chem. B* **2005**, 109, 6287.
- (6) Kim, S.-N.; Son, W.-J.; Choi, J.-S.; Ahn, W.-S. *Microporous Mesoporous Mater.* **2008**, 115, 497.
- (7) Wei, J.; Shi, J.; Pan, H.; Zhao, W.; Ye, Q.; Shi, Y. *Microporous Mesoporous Mater.* **2008**, 116, 394.
- (8) Demessence, A.; D'Alessandro, D. M.; Foo, M. L.; Long, J. R. *J. Am. Chem. Soc.* **2009**, 131, 8784.
- (9) An, J.; Geib, S. J.; Rosi, N. L. *J. Am. Chem. Soc.* **2010**, 132, 38.
- (10) Arstad, B.; Fjellvåg, H.; Kongshaug, K. O.; Swang, O.; Blom, R. *Adsorption* **2008**, 14, 755.

- (11) Villiers, C.; Dognon, J.-P.; Pollet, R.; Thuéry, P.; Ephritikhine, M. *Angew. Chem. Int. Ed. Engl.* **2010**, *49*, 3465.
- (12) Kitaura, R.; Kitagawa, S.; Kubota, Y.; Kobayashi, T. C.; Kindo, K.; Mita, Y.; Matsuo, A.; Kobayashi, M.; Chang, H.-C.; Ozawa, T. C.; Suzuki, M.; Sakata, M.; Takata, M. *Science* **2002**, *298*, 2358.
- (13) Matsuda, R.; Kitaura, R.; Kitagawa, S.; Kubota, Y.; Belosludov, R. V.; Kobayashi, T. C.; Sakamoto, H.; Chiba, T.; Takata, M.; Kawazoe, Y.; Mita, Y. *Nature* **2005**, *436*, 238.
- (14) Takamizawa, S.; Nataka, E.; Akatsuka, T.; Miyake, R.; Kakizaki, Y.; Takeuchi, H.; Maruta, G.; Takeda, S. *J. Am. Chem. Soc.* **2010**, *132*, 3783.
- (15) Zhang, J.-P.; Chen, X.-M. *J. Am. Chem. Soc.* **2009**, *131*, 5516.
- (16) Dietzel, P. D. C.; Johnsen, R. E.; Fjellvåg, H.; Bordiga, S.; Groppo, E.; Chavan, S.; Blom, R. *Chem. Commun. (Camb)*. **2008**, 5125.
- (17) Wu, H.; Simmons, J. M.; Srinivas, G.; Zhou, W.; Yildirim, T. *J. Phys. Chem. Lett.* **2010**, 1946.
- (18) Peterson, V. K.; Liu, Y.; Brown, C. M.; Kepert, C. J. *J. Am. Chem. Soc.* **2006**, *128*, 15578.
- (19) Perdew, J.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865.
- (20) Goedecker, S.; Teter, M.; Hutter, J. *Phys. Rev. B* **1996**, *54*, 1703.
- (21) Grimme, S. *J. Comput. Chem.* **2006**, *27*, 1787.
- (22) Car, R.; Parrinello, M. *Phys. Rev. Lett.* **1985**, *55*, 2471.
- (23) Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. *J. Am. Chem. Soc.* **1992**, *114*, 10024.
- (24) Campañá, C.; Mussard, B.; Woo, T. K. *J. Chem. Theory Comput.* **2009**, *5*, 2866.
- (25) 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*, 8814.
- (26) Harris, J. G.; Yung, K. H. *J. Phys. Chem.* **1995**, *99*, 12021.
- (27) Sirjoosingh, A.; Alavi, S.; Woo, T. K. *J. Phys. Chem. C* **2010**, *114*, 2171.
- (28) Forrester, T. R.; Smith, W. *DL_POLY ver. 2.18 (C.L.R.C. Daresbury Lab. Daresbury, UK, 1995)*.

- (29) Düren, T.; Bae, Y.-S.; Snurr, R. Q. *Chem. Soc. Rev.* **2009**, *38*, 1237.
- (30) Tafipolsky, M.; Amirjalayer, S.; Schmid, R. *Microporous Mesoporous Mater.* **2010**, *129*, 304.
- (31) Park, H.; Krigsfeld, G.; Teat, S. J.; Parise, J. B. *Cryst. Growth Des.* **2007**, *7*, 1343.
- (32) Smit, B.; Maesen, T. L. M. *Chem. Rev.* **2008**, *108*, 4125.
- (33) Walton, K. S.; Millward, A. R.; Dubbeldam, D.; Frost, H.; Low, J. J.; Yaghi, O. M.; Snurr, R. Q. *J. Am. Chem. Soc.* **2008**, *130*, 406.
- (34) Kauffman, K. L.; Culp, J. T.; Goodman, A.; Matranga, C. *J. Phys. Chem. C* **2011**, *115*, 1857.

3 Construction of a hypothetical database of MOFs with large chemical, physical and topological diversity

3.1 Abstract

Recently, databases of computationally generated hypothetical materials have sprung up to tackle the problem of identifying candidate materials for a wide range of applications. When conducting a discovery-based screening study, it is important to include as much diversity in the dataset as possible. In the case of metal organic frameworks (MOFs) this can include a number of descriptors such as porosity, chemical structure, topology, and surface area. In this study, a database of hypothetical MOFs was generated for the purposes of screening MOFs for a variety of applications. Using a method of building MOFs based on rigid topological parameters, we report the generation of close to 24,000 MOFs using 28 metal structural building units (SBUs) and 164 organic SBUs based on 20 topologies. These structures were functionalized with 50 different functional groups to produce over 1.3 million hypothetical MOFs. All of these structures were optimized at the force field level, and their physical and chemical diversity was evaluated with respect to other hypothetical materials databases.

3.2 Introduction

It has long been a goal in the field of MOF chemistry to design materials with specific purpose, pre-designing pore shape and chemistry for optimum performance in a desired application. The importance of rational MOF design has been demonstrated in several recent experimental studies. For example the MOF which currently holds the record BET surface area¹ of 7140 m²/g was obtained by increasing the molecular surface area of an organic SBU used in a MOF with relatively high surface area.² The MOF UiO-66³ was tuned towards more hydrophobic pores by adding methyl groups to the organic SBU.⁴ Interestingly, the shape of its

inorganic cluster has biomimetic properties which can catalyze the decomposition of toxic nerve agents, as was shown recently.⁵ While these studies demonstrate the power of rational design, they are limited to highly focused optimizations on a limited set of materials. They do not consider the nearly infinite chemical space searchable to find the most hydrophobic MOF, or the highest achievable surface area. This is an understandable limitation, considering the astronomical time and costs required to effectively sample these possibilities experimentally.

Due to this vast search space of SBU diversity, identifying the ideal MOF for any property remains a serious challenge. This is not only limited to the chemical space sampled by SBUs, but also how they assemble into periodic structures. Their relative orientations in the lattice will determine the pore shape and chemistry of the MOF, and ultimately, its functional properties. As a testament to the importance of this design factor, the field of *reticular chemistry* has been established, dedicated to understanding and classifying crystalline materials into their underlying network topology or net.^{6,7} As of this writing, there are over 2000 experimentally realized nets a MOF could assemble into, which have been stored in the Reticular Chemistry Structure Resource (RCSR)⁸ (<http://rcsr.anu.edu.au>). Thus the search space for rational design of MOFs is enormous, when considering the vast possible SBU chemistry with the possible underlying nets MOFs can achieve.

Recently, databases of hypothetical porous materials have surfaced for the purposes of sampling this large search space. The first database of hypothetical MOFs contains over 137,000 structures.⁹ These MOFs were constructed with a recursive-based algorithm, which assembled them by starting with a single SBU and iteratively attaching new SBUs to un-bonded connection sites until a periodic structure is found. This was a pioneering work and established computational methods as a powerful tool for rational MOF design. Later analysis revealed that

the structures built with the algorithm sampled only 6 network topologies.¹⁰ This can be attributed to the nature of the algorithm, which 'snaps' together rigid SBUs based on pre-defined orientation parameters. These parameters must be adjusted by hand to assemble new MOFs with different topologies. As such, the range and diversity of chemical and topological space sampled by this program is limited to the number of SBUs someone has parameterized for MOF assembly. The hypothetical MOFs in this database were constructed using 5 inorganic and 97 organic SBUs. This produced an estimated 3,000 unfunctionalized structures, which were decorated with 13 functional groups to produce the 137,000 structures in the database. One of the short-comings of these structures is that they were not relaxed to their energetic minimum, an important consideration when screening these materials for the purposes of identifying synthetic candidates. This resulted in unphysical structures where atoms were too close together; ultimately affecting many of the structural properties reported from these materials, such as their atomic density and void fraction. Further, there was no bonding information reported in these structures, making identification of bonded atoms impossible in some cases, and optimization difficult at the force field level.

Martin *et al.* constructed a small set of hypothetical MOFs.¹¹ In their study, organic ligands were selected from a database of commercially available rigid di-carboxylate organic compounds. These linear organic compounds were assembled into MOFs with the **pcu** topology by bonding to a 6-coordinate (6-c) Zn_4O inorganic SBU, resulting in a small database of 111 hypothetical MOFs. Importantly, the authors detailed some important considerations when screening hypothetical structures. First, check for structural similarities such that no redundant calculations are performed, secondly, optimize the structures to obtain more physically realistic structures, and finally remove those which introduce significant distortion in the optimized

structure. The same authors have also constructed a hypothetical database of close to 18,000 porous polymer networks, which consist of tetrahedral-like SBUs coordinated fit into the **dia** topology.¹²

With the diverse range of possible applications for MOFs, it is desirable to have access to a larger topological and chemical space, such that one can readily identify promising candidate materials. While it is currently unclear what diversity measures are important for MOF applications, it is likely that broadening the chemical and topological space will increase the breadth of applications the database can be screened for. This work details the construction of our hypothetical MOF database, which is built using a similar algorithm as presented by Wilmer *et al.*⁹ Care has been placed in including a considerably larger chemical and topological space where we report the generation of 21,383 un-functionalized structures from a diverse set of 28 metal building units, 164 organic building units and 20 topologies. These have been functionalized with 50 common functional groups to produce over 1.3 million structures. This is by far the largest topological, and chemically diverse database of hypothetical MOFs generated to date.

3.3 Methods

In the following section we discuss our implementation of the algorithm used to generate MOFs in the database. The program's core structure is similar to the recursive method introduced by Wilmer *et al.*⁹ where the underlying network topology of the generated MOFs are predetermined by hand-parameterized bonding vectors assigned to the SBUs used to build it. The method is described in detail, in part for transparency but also to explain why some topologies in the database are sampled more than others.

3.3.1 Structure Generation Basics

To construct an individual MOF in the program, a selection of SBUs are taken from a hand-built library of molecules extracted from existing MOF crystal structures. These representative SBUs are deemed the *building set* for the MOF. The description of each SBU consists of the 3D arrangement of the molecules atoms, complete with their bonding topology, the connection sites where the SBU can form bonds to other SBUs, and two orthogonal vectors expanding from the connection site describing the alignment of the SBU in the periodic structure. These vector parameters, seen as red and blue lines extending from the SBUs in Figure 3.1, are set carefully by hand for each SBU to reproduce the underlying network topology of an experimental MOF. While these parameters are extracted from the bonds of an existing crystal structure, in principle they can be extracted from any 3-dimensional periodic net (from the RCSR⁸ for example) so long as the nodes in the net are of the same connectivity and similar geometry as the desired SBUs.

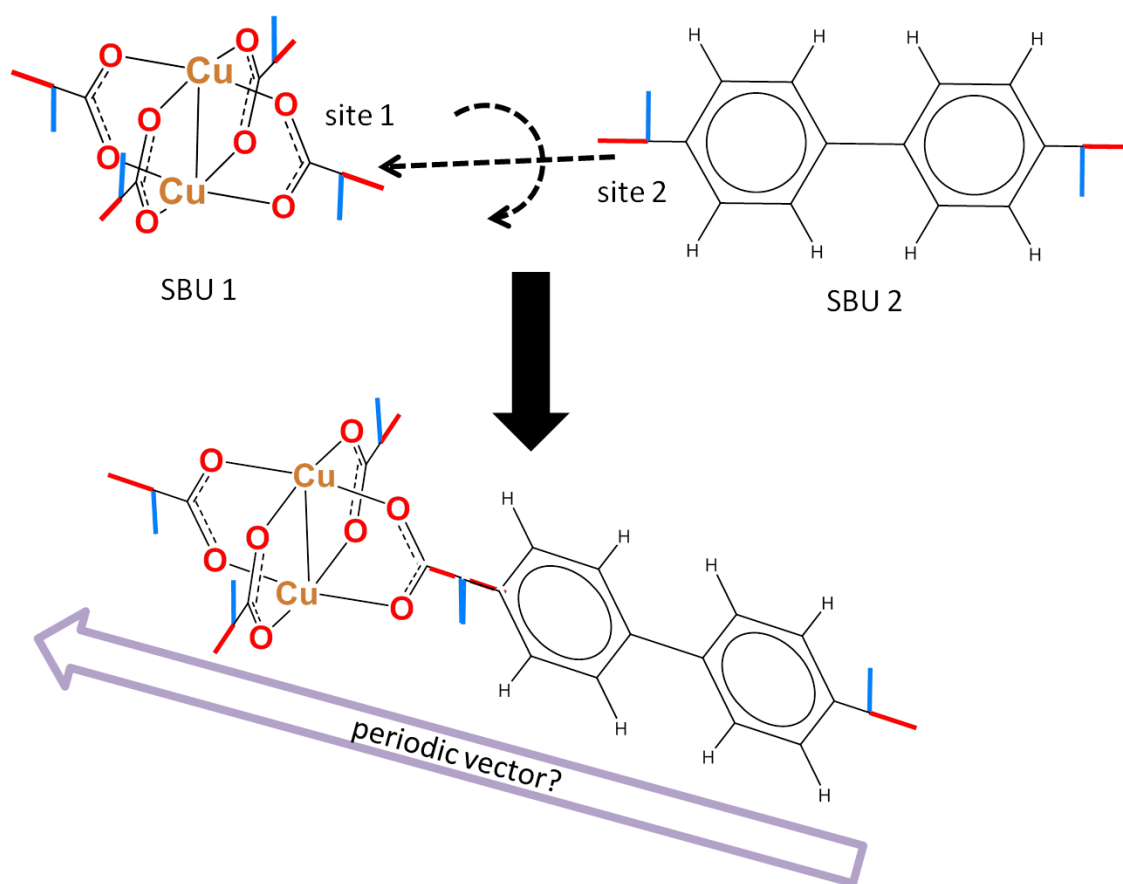


Figure 3.1: Demonstrating the translation and rotation operations performed each time SBUs are joined together in the algorithm. The picture depicts biphenyl (top right) being oriented to bond with the Cu paddlewheel SBU (top left). The blue and red vectors extending from each SBU depict the bonding parameters, and will ultimately determine the final topology of the hypothetical MOF structure. After the bond has been made, each unattached connection site is scanned for possible bonds, either locally or via a periodic translation vector. Each SBU can possess more than one set of these bonding parameters to build MOFs with different underlying topologies.

The first SBU is randomly selected from the building set and is placed at the origin to ‘seed’ the MOF. Each connection site on the seed SBU can be bonded to any of the connection sites on any of the SBUs from the building set. Thus the number of possible bonds between SBUs and connection sites creates a combinatorial explosion of deterministic trial moves, where each new placement of an SBU yields a new set of possible connections to other SBUs. Of course, sampling all of these possibilities is intractable and so the algorithm will descend branch-first into this search tree, where each ‘branch’ is a specific set of instructions to bond SBUs together. A graphical schematic of the tree is shown in Figure 3.2 for two SBUs, the Cu

paddlewheel which has 4 connection sites, and the 1,3,5-benzene tricarboxylate (BTC) SBU which has 3 connection sites. A branch-first sampling of this tree would sample the outer left nodes first, then slowly move to the right by incrementing the bottom node. The depth of the tree is set to a user-defined limit, so that a single branch of the tree does not encompass an attempt to connect an infinite number of SBUs together. This number depends on the network topology of the desired MOF, which is determined at the beginning when selecting the SBUs in the building set. For example, the vector parameters for both SBUs in Figure 3.2 were designed to build the **tbo** topology. This topology supports 14 SBUs in its primitive cell, so the search tree will have 14 levels of nodes. The search tree is exhaustive up to the maximum number of SBUs, attempting every possible bonding combination until a structure is made.

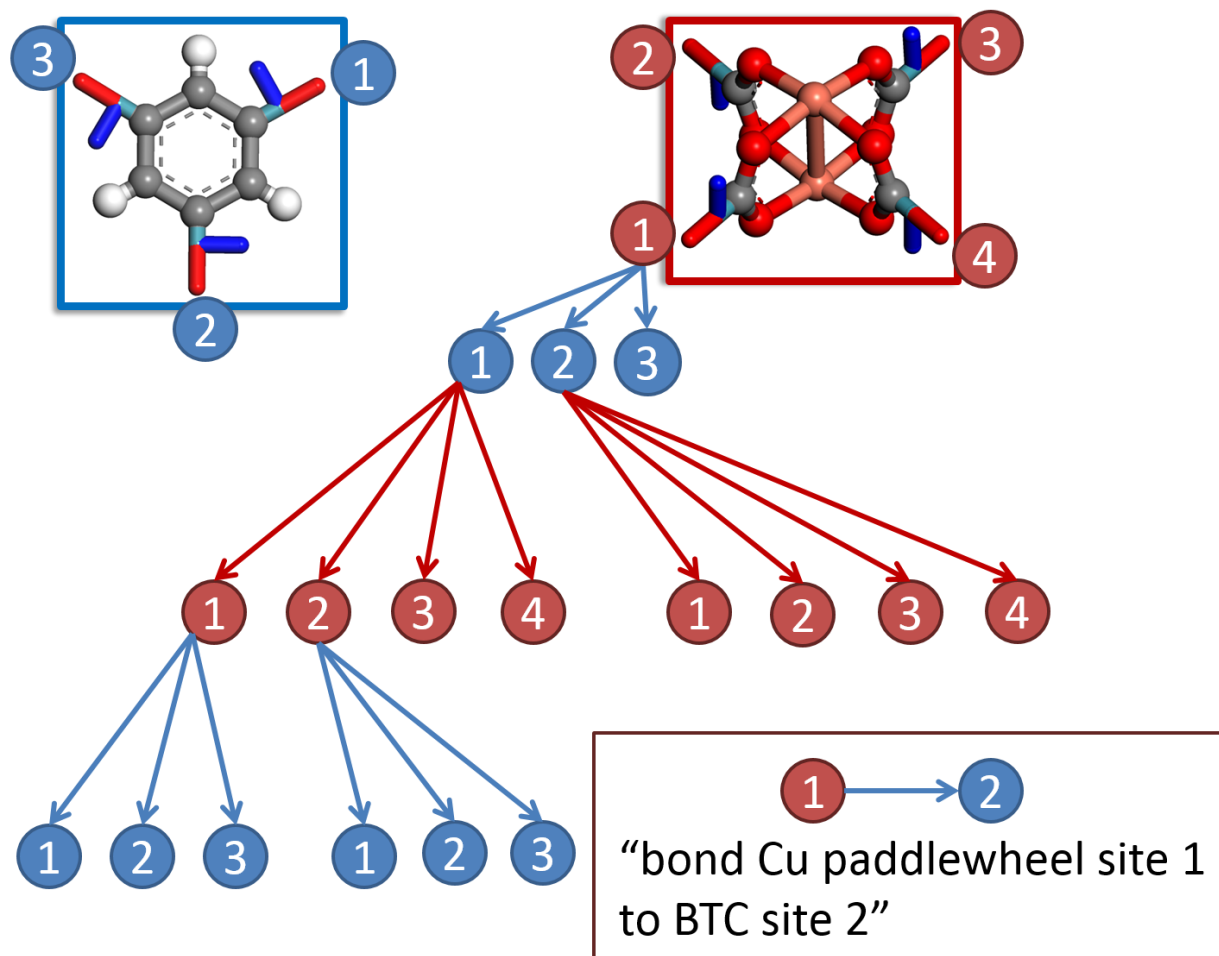


Figure 3.2: Schematic of the tree which exhaustively samples every bonding possibility. Every red node represents a Cu paddlewheel SBU connection site, and every blue node represents a BTC connection site. The maximum depth of the tree will depend on the desired topology of the MOF.

At each step the program receives an instruction from the search tree, e.g. “*bond SBU 1 connection site 1 with a new SBU 2 connection site 2*”. This direction is depicted graphically in Figure 3.1 where a biphenyl SBU (new SBU 2) is connected to a copper paddlewheel SBU (SBU 1) at their connection sites (site 2 and site 1, respectively). After their vector parameters are aligned, the connection between SBUs is formed by translating SBU 2 such that its connection site overlaps with SBU 1. Once the program has performed the instruction, all freely available connection sites (not already connected to an SBU) in the growing structure are scanned for additional possible connections between existing SBUs. These include non-local periodic bonds,

as well as local bonds if two sites are within 1 Å. Non-local bonds are found between two SBUs when their bond parameters are anti-parallel, pointing away from each other, and are farther than 1 Å. If these conditions are satisfied, then there exists a possible periodic vector between these connection sites, describing a translationally invariant direction of the hypothetical crystal. To realize this periodicity, three additional conditions must be met; i) the periodic vector formed from joining these two SBUs is not parallel with existing periodic vectors ii) if this is the third periodic vector, it is not in a plane with the first two, and iii) there are less than 3 periodic vectors already defined in the material. If these tests are passed, the vector between the two connection sites is added to the periodic boundaries of the structure, and the two SBUs are bonded together. Following the completion of a single step in the branch directive, as well as the additional bonding checks discussed above, atomistic overlap is evaluated. Bonded atoms notwithstanding, two atoms are considered to be in close contact if the distance between them is less than the sum of their van der Waals radii multiplied by a scaling factor of 0.4. If overlap is found, then the bond step is ignored and the algorithm moves on to the next branch in the sampling tree. It should be noted that only organic and inorganic SBUs are allowed to bond to each other, unless an SBU has connection sites with special bonding flags, which will be discussed below.

The attempt to build a MOF from a specific building set will terminate if one of two events occur. The first is if a structure is made. This is determined if all of the connection sites in the structure are bonded, three periodic vectors are found, and all of the SBUs in the building set are represented at least once in the hypothetical MOF. Upon building this structure, no other attempts to make new MOFs with these SBUs are tried. This is so that we can guarantee that no two MOFs in the database were built with the same SBUs. The second possibility is if an upper

limit of branches is attempted without generating a MOF. This is user-defined, and was set to 100,000 branches in this work.

As a general comment, the success or failure of each MOF building attempt is heavily dependent on the SBU vector parameters. If these parameters are offset by just a few degrees from their crystallographic determination, the SBUs will collide during a build attempt and no MOF will be made. An example is presented in Figure 3.3 where the program is told to construct a hypothetical version of HKUST-1 using two SBUs; the copper paddlewheel and the tritopic benzene SBU using two sets of parameters for the organic SBU. One set of parameters were extracted from the crystal structure, shown in Figure 3.3a where the vectors for the organic SBU are raised by about 4° from the plane of the benzene ring. Figure 3.3b shows that these parameters effectively build the MOF without problems, where the characteristic 10-membered cluster of nodes of the underlying **tbo** topology is captured (in green). However if one were to construct a MOF where the parameters are co-planar with the aromatic ring, such as those described as ‘flat’ in Figure 3.3a, the underlying topology is not effectively captured and collisions occur during assembly. Figure 3.3c demonstrates this, where lines in red indicate parts where the **tbo** topology should overlap. Such small deviations in the parameters create catastrophic changes in the overall structure, demonstrating the ‘butterfly effect’ is applicable here; small changes in the starting conditions (such as the 4° shift of the orientation vectors) generate large deviations later on. It is therefore vital to the success in building a structure that these parameters are carefully tested and validated before proceeding to generate new hypothetical structures.

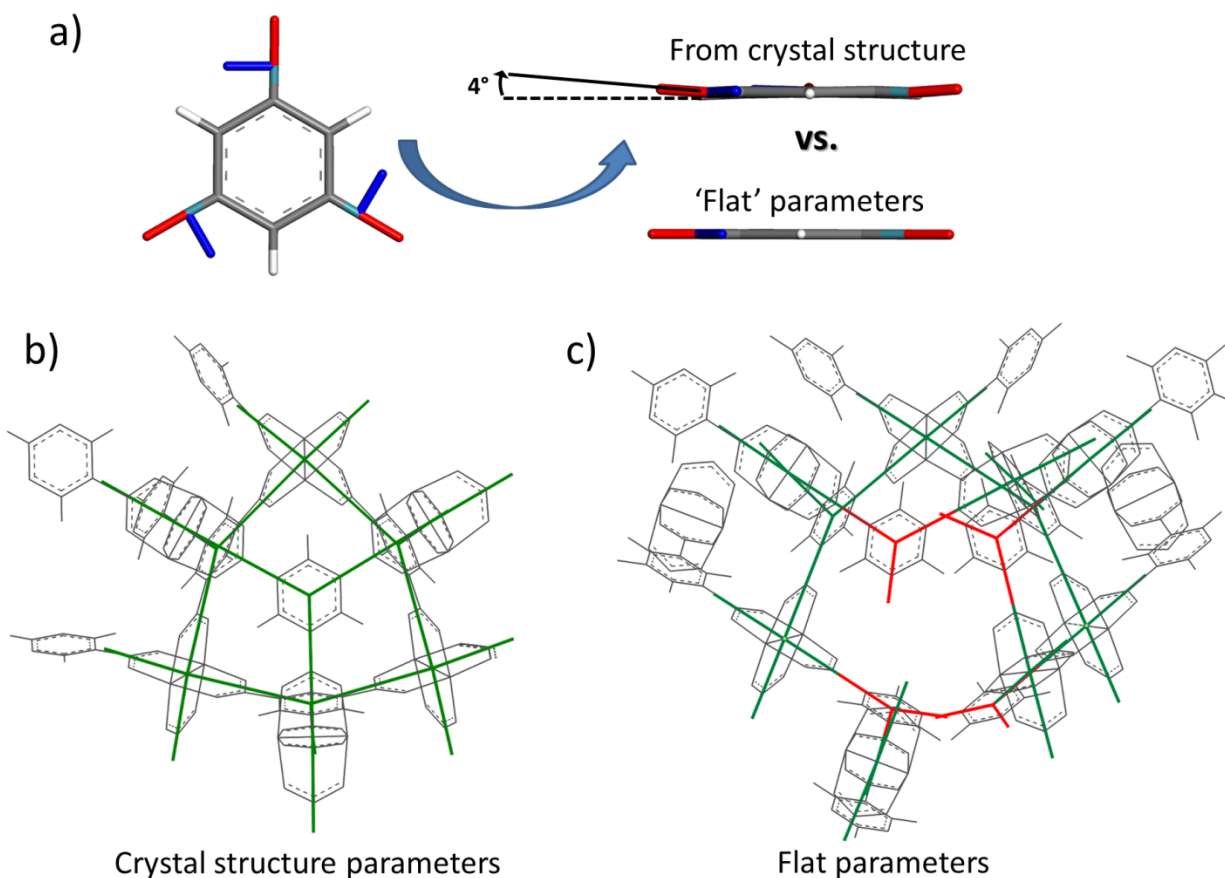


Figure 3.3: Demonstration of the sensitivity of the parameters to the construction of hypothetical MOFs. a) two sets of parameters constructed one using bonding from the crystal structure, and the other 'flat' ones set by hand to be co-planar with the benzene ring. b) snapshot of the structure generation code assembling HKUST-1 with the crystal structure parameters, the underlying network shown in green. c) snapshot of the code using the 'flat' parameters. Lines in red indicate disagreement with the *tbo* topology.

3.3.2 Special Bonding Parameters to Reduce the Combinatorial Search

The purpose of implementing this brute force sampling method is to ensure that every configuration could be attempted in the case that a single trial yields a bad structure (overlapping atoms, SBUs with the wrong orientation, etc.) or to try all possible configurations when multiple organic/inorganic SBUs are being used to build a single MOF. In many cases, however, this exhaustive search tree does not need to be fully realized. The extent of sampling depends on the symmetry of the SBUs, as well as the desired underlying topology of the final MOF. For example, the MOF HKUST-1's primitive unit cell consists of 6 copper paddlewheel SBUs and 8

tritopic benzene SBUs. The topology that describes the connectivity of these SBUs is a (3,4)-c network **tbo**. If the copper paddlewheel SBU is initially seeded, it can form 64 possible bonds with 4 benzene SBUs. The subsequent inserted tritopic benzene and copper paddlewheel SBUs can form 16 and 27 bond combinations, respectively. For the structure generation program to build this MOF, it would need to select a single branch from a tree with an upper limit of $64 \times (27)^5 \times (16)^8 = 4.0 \times 10^{18}$ possible bonding instructions. In the case of assembling HKUST-1, *all* of these bonding combinations yield the exact same structure. This is because the SBUs and the underlying topology of the MOF are highly symmetric, in fact, there is only one unique SBU-SBU connection in the entire structure with space group symmetry *FM-3M*. To reduce the number of searches the program must make, the point group of each SBU was found to flag symmetry equivalent connection sites. As the program proceeds, it keeps track of the symmetry equivalent connection sites, and the search tree is adjusted to ignore multiple redundant connection attempts.

In contrast, SBUs assembled using a topology with lower symmetry will generally join with other SBUs at very specific bonding sites to create a MOF, and sampling all possible bonding will consume a needless amount of CPU time. As an example, consider the net with the RCSR label **pto**, which is a (3,4)-c net like **tbo**, only with lower symmetry. MOFs possessing this net contain 3-c SBUs connected to planar 4-c SBUs with a 55° out of plane torsion. The presence of a mirror plane which cuts right through the 4-c vertex in the highest symmetry space group of **pto** (*PM-3N*) yields 3-c vertices with *chiral* mirror images. Thus, to generate hypothetical MOFs with the **pto** topology in our algorithm, both the 3-c SBU and its mirror image were explicitly parameterized. Moreover, constraints were introduced to reduce the number of bond combinations tried by the program before a successful one is found. Figure 3.4

shows these constraints where the gold colored atoms were only allowed to connect to other gold colored atoms. The MOF on the bottom right of Figure 3.4 shows the hypothetical version of MOF-143 assembled in our algorithm¹³ with the chiral organic SBUs colored grey and gold.

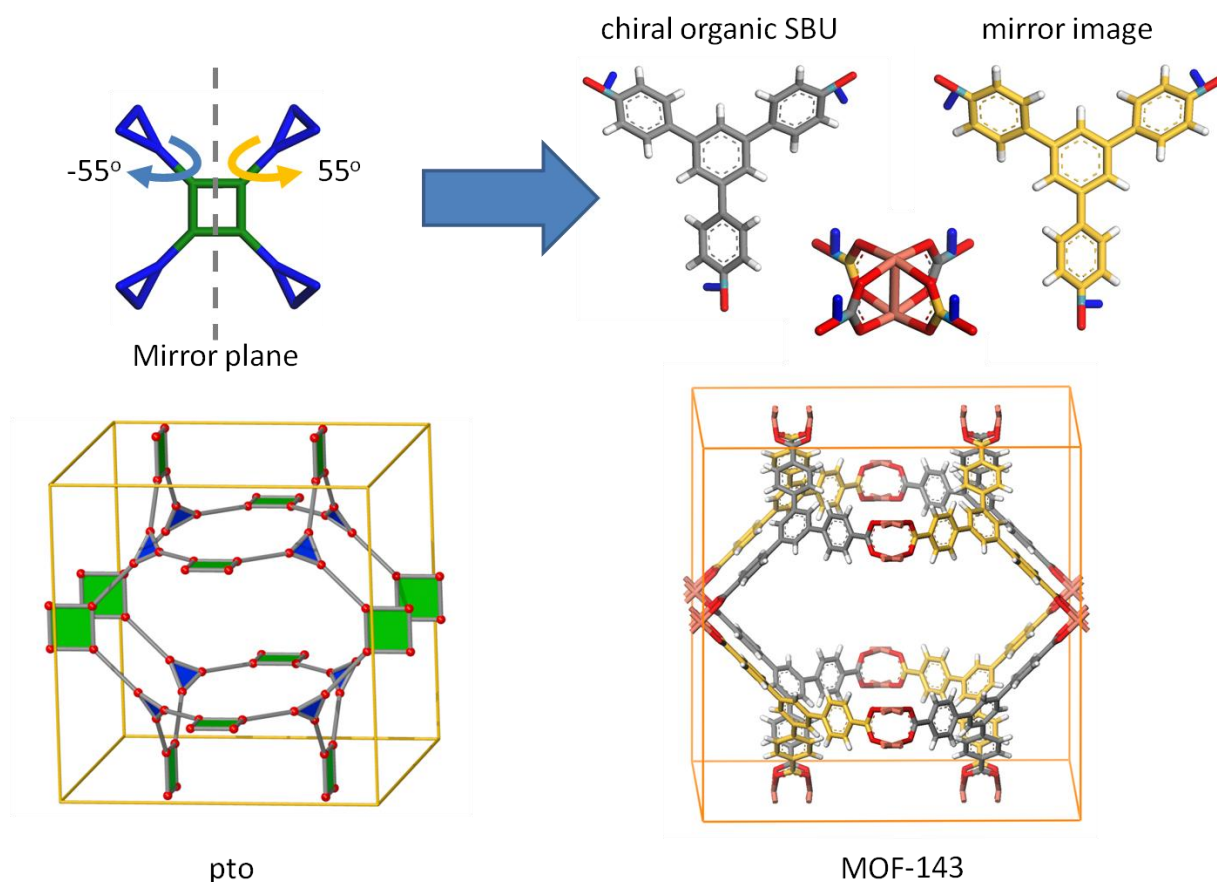


Figure 3.4: A demonstration of how the program assembles MOFs based on the (augmented) **pto** topology, pictured bottom left. Top left: the geometry of the 4-c vertex (shown as a green square) connected to four 3-c vertices (blue triangles). The mirror plane cutting through the middle of the square yields 3-c vertices with chiral connectivity. To build MOF-143, seen bottom right, the algorithm requires both chiral images of the 3-c SBUs, and appropriate bonding flags to ensure the mirror images are preserved in the final structure.

These bonding constraints are not only limited to ensuring the correct bonding to chiral molecules. Several inorganic SBUs in the database are rod like, meaning they extend infinitely in one dimension. The SBU of these rod-like structures is a discrete ‘cut’ of the infinite chain contained within the unit cell of the original MOF. These SBUs possess connection sites which are constrained to only connect with themselves, so that the extended rod-like shape can be re-

created in the hypothetical MOFs. Additionally, nitrogen terminated organic SBUs were constrained to only connect to open metal sites of inorganic SBUs to ensure correct chemistry in the final structure.

3.3.3 *Structural diversity in the database*

The goal of this work was to construct a large number of hypothetical materials with a diverse range of pore sizes, surface area, and pore chemistry so that it could be screened for various applications. Up until now we have shown how existing experimental MOFs with different underlying topologies could be constructed with the program. Now we will discuss how new SBUs were parameterized to build a particular topology. As mentioned previously, the program is highly sensitive to the vectors describing the connection sites on each SBU, such that small deviations in the orientation of these vectors could result in atom collisions and MOF building failure. New SBUs were parameterized by isotropically expanding or contracting the bond parameters extracted from experimental structures, while ensuring that the relative orientations of the vectors remained constant. This limits the connective geometries of the new SBUs to possess the same point group symmetry as the SBU extracted from the crystal structure. A visual representation of this process is shown in Figure 3.5, where the isotropic expansion of the 3-c SBU parameters yields a new hypothetical MOF seen in the top right, while extending only *one* of the connection sites of the 3-c SBU results in MOF building failure.

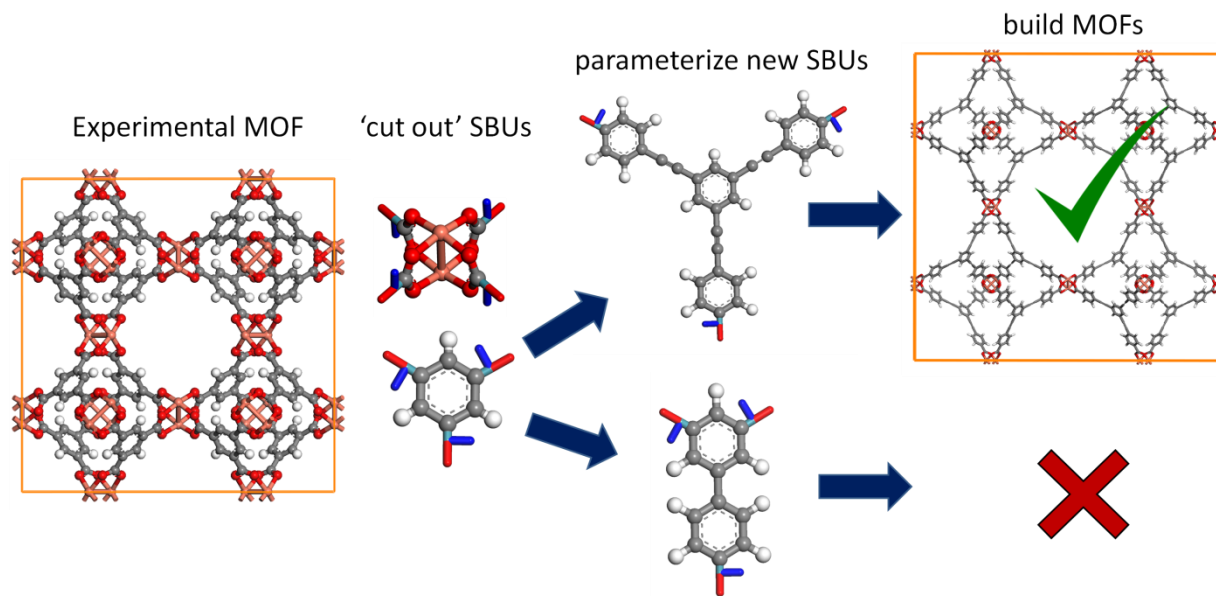


Figure 3.5: Demonstration of parameterizing new SBUs based on extraction of bond parameters from an experimental MOF. Only concerted expansions and contractions of the bonding parameters would permit the construction of new hypothetical MOFs, so that the process on the bottom, which consists of an asymmetric parameterization of a biphenyl SBU resulted in no new hypothetical MOF.

As with the database of Wilmer et al.,⁹ we strove to build MOFs with one inorganic SBU and up to two organic SBUs. In most cases, using two or more organic SBUs yields no new MOFs, as repeated attempts to assemble structures containing multiple organic SBUs would result in atom collisions. However some network topologies permit the use of multiple SBUs without problems. Exploiting these topologies effectively increased the number of hypothetical MOFs by iterating over all possible combinations of one inorganic SBU and up to two organic SBUs, but skewed the topology distribution in the database in favor of those topologies. For example, Figure 3.6a shows the primitive cubic net (**pcu**) can support up to three linear organic SBUs of different length. This is achievable because the three lattice vectors are determined solely by three linearly independent edges incident upon the 6-c vertex. In another, slightly more complex example, the net **tbo** can support up to two 3-c SBUs. This is because **tbo** can be abstracted to isolated clusters of edges and vertices, where the relative edge lengths of these

clusters can be altered with respect to one another without adjustment to the SBU geometries. To achieve this type of SBU diversity in crystallographic topologies which do not support this edge length variation would require a non-straightforward adjustment of the periodic boundaries. Expanding the primitive cell to accommodate this was not included in the building of the current database.

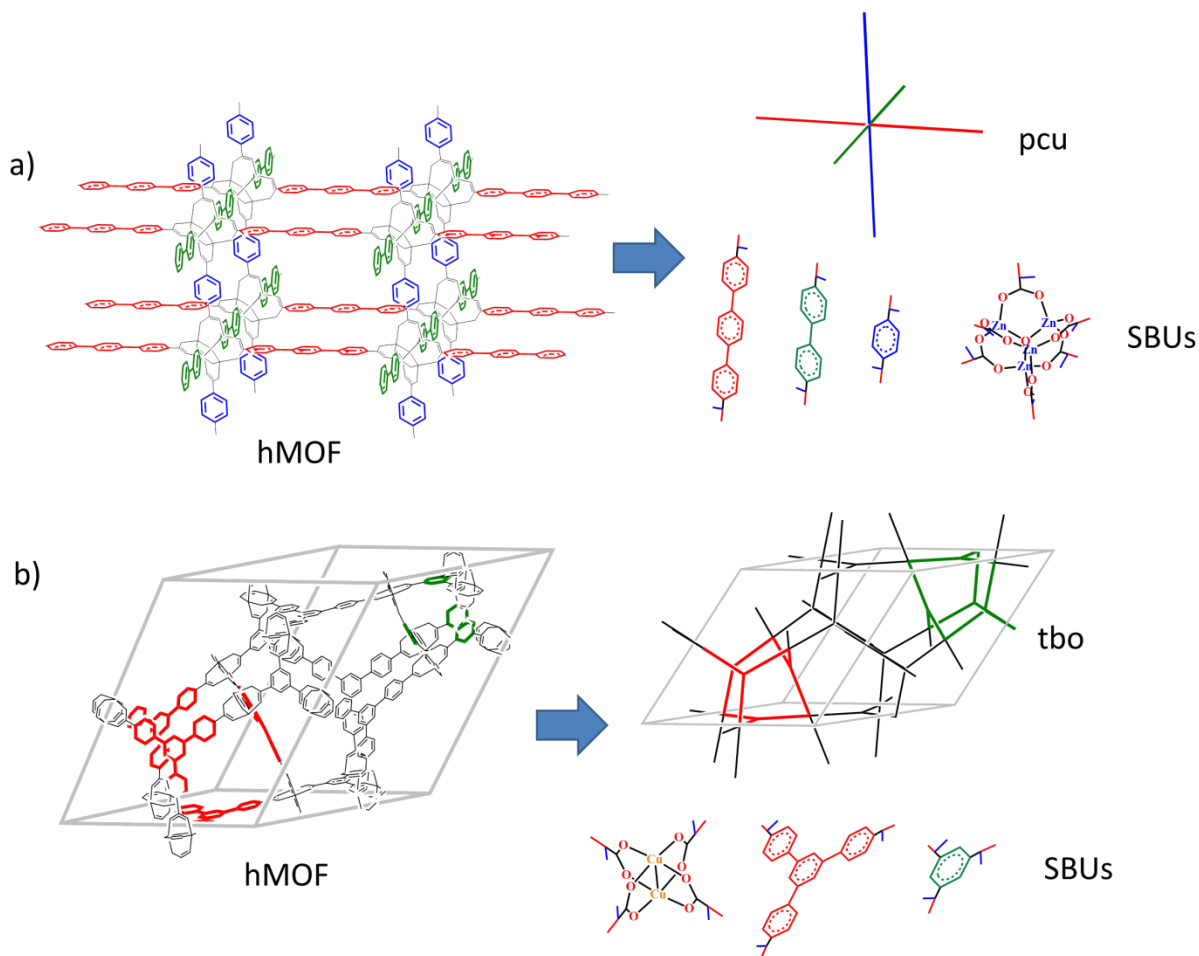


Figure 3.6: Two topologies where the primitive cell supports multiple edge length variations. a) the **pcu** topology can support three edges of different length without altering the geometry of the nodes. An example of a MOF generated by the program is shown on the left, where organic SBUs of different length are represented by the colors red, blue and green. These are attached to the 6-c Zn_4O metal SBU to form the hypothetical MOF b) the primitive cell of **tbo** can support up to two organic SBUs of different length. The red and green nodes and edges represent isolated clusters which can be expanded or contracted without affecting the node geometries of the net.

3.3.4 *Output of the structure generation program*

While minimally, all that is needed to describe a periodic structure are the lattice parameters and the atoms in the unit cell, upon generation of a hypothetical structure, the program includes some extra features necessary for the next steps in generating the MOF database, specifically functionalization and structural optimization.

Because the algorithm snaps together rigid SBUs, there exists the possibility that the atoms are not in their energy minima positions, thus bond topologies and force field atom types were included in each MOF so they could be optimized at the classical force field level. Prior to generating structures, each SBU was described with a bonding table, complete with atom and bond types consistent with the UFF¹⁴ parameterization. This was done at the SBU level so that upon construction of a new hypothetical MOF, the program knows exactly which atoms are to be bonded with a particular bond type. This strategy poses an advantage over determining the bonding table after the MOF is generated, in that close atoms are not incorrectly bonded based on distance measurements, and are thus optimized away from each other when subjected to a force field optimization. The atom and bond typing for each SBU in the database was found using the python API of openbabel (pybel).¹⁵

In anticipation of functionalizing the hydrogen atoms of the MOFs organic SBUs to further enhance the chemistry and pore geometry of the database, the space group and asymmetric atoms in each hypothetical MOFs were identified. These were computed using the python API of the spglib program (spglib.sourceforge.net).¹⁶

3.3.5 *Functionalization of the hypothetical MOFs*

Functionalization of the hypothetical structures was carried out as a separate procedure in order to increase the chemical diversity of the pores within the structure. The symmetry of the crystal was used to identify hydrogen atoms that were symmetrically equivalent so that

equivalent sites would be functionalized to retain as much of the parent structure's symmetry as possible. Each individual functionalized structure is generated using the following procedure; A single functional group is chosen to decorate a MOF from a library of 50 common functional groups.¹⁷ These functional groups were pre-oriented to a preferential configuration and alignment with respect to a benzene ring, where the energy minimization is performed using the UFF forcefield in the openbabel package.¹⁵ The asymmetric hydrogens of a hypothetical MOF were randomly chosen for functionalization, where a single functional group was randomly selected from the set of 50 to replace the H's. Each functional group is initially inserted based on its preferential geometry and orientation, after which an overlap check is carried out. Here, if any atoms fall within the van der Waals radius of the existing atoms, the functional group is rotated about its bond by 30° until there is either no overlap, or the rotation completes a 360° turn and the functionalization is aborted. If the procedure is successful the final structure is identified by a unique string: 'functional group'@'unique site'. This ensures that if the random functionalization method will not try to produce the same functionalization for a MOF, and that the resulting hMOF database contains unique structures.

3.3.6 MOF database generation

In total, a set of 28 inorganic and 164 organic SBUs were parameterized. The resulting MOF structures were characterized into one of 20 different underlying topologies. The following outlines the steps performed in the algorithm to construct the database of hypothetical MOFs:

1. *Select one inorganic and two organic SBUs as the building set to construct a MOF.* The program generates all possible combinations of organic and inorganic SBUs and iterates over each. Only those SBU combinations which have been parameterized for the same topology are combined, so the program isn't assembling SBUs together 'blindly'. There is thus a high chance a structure will be assembled from a given building set.

2. *Build the hypothetical MOF.* Given a building set from the previous step, the algorithm for snapping together SBUs is run. This is described in detail above, so a brief description is mentioned here. SBUs are joined together iteratively, starting with an initial 'seed' SBU. Each time a new SBU is placed, possible periodic vectors or other local bonds are searched, and atomic collisions are detected. The program will terminate once an upper limit of attempts was made, or a periodic structure with no atomistic collisions was found.
3. *Compute the symmetry and bonding table.* If a structure is generated in the previous step, the bonding table (with appropriate bonds attached between SBUs) is constructed, and space group symmetry is identified.
4. *Functionalize the structure.* For each functional group in a library of 50 groups, generate up to four functionalized derivatives of the symmetrized structure.
5. *Optimize the structures.* Once the functional groups have been added, and the bond tables updated, each structure is optimized at the molecular mechanics level. The computational details of this optimization are discussed below.

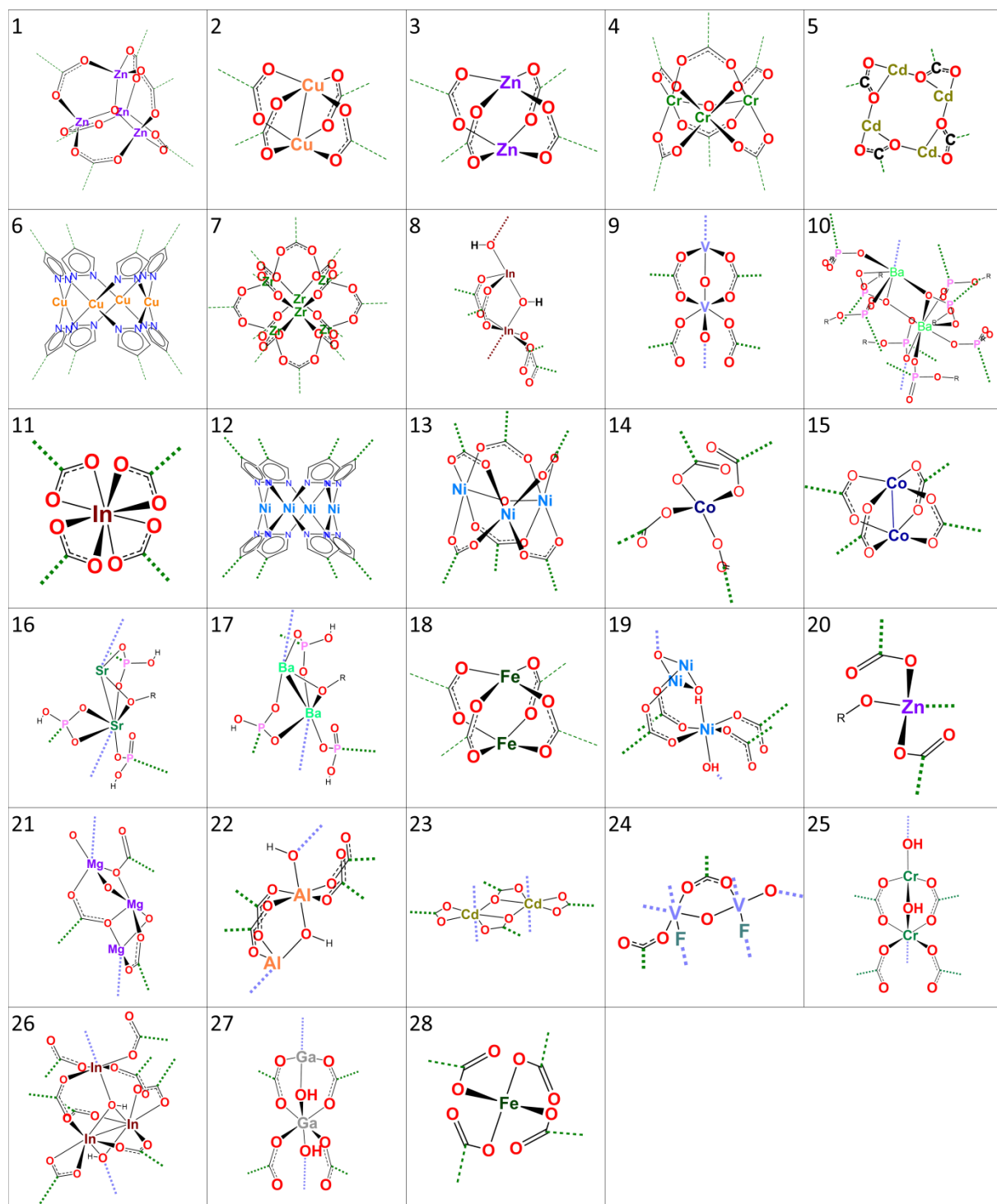


Figure 3.7: The 28 metal SBUs used to construct the database of hypothetical MOFs.

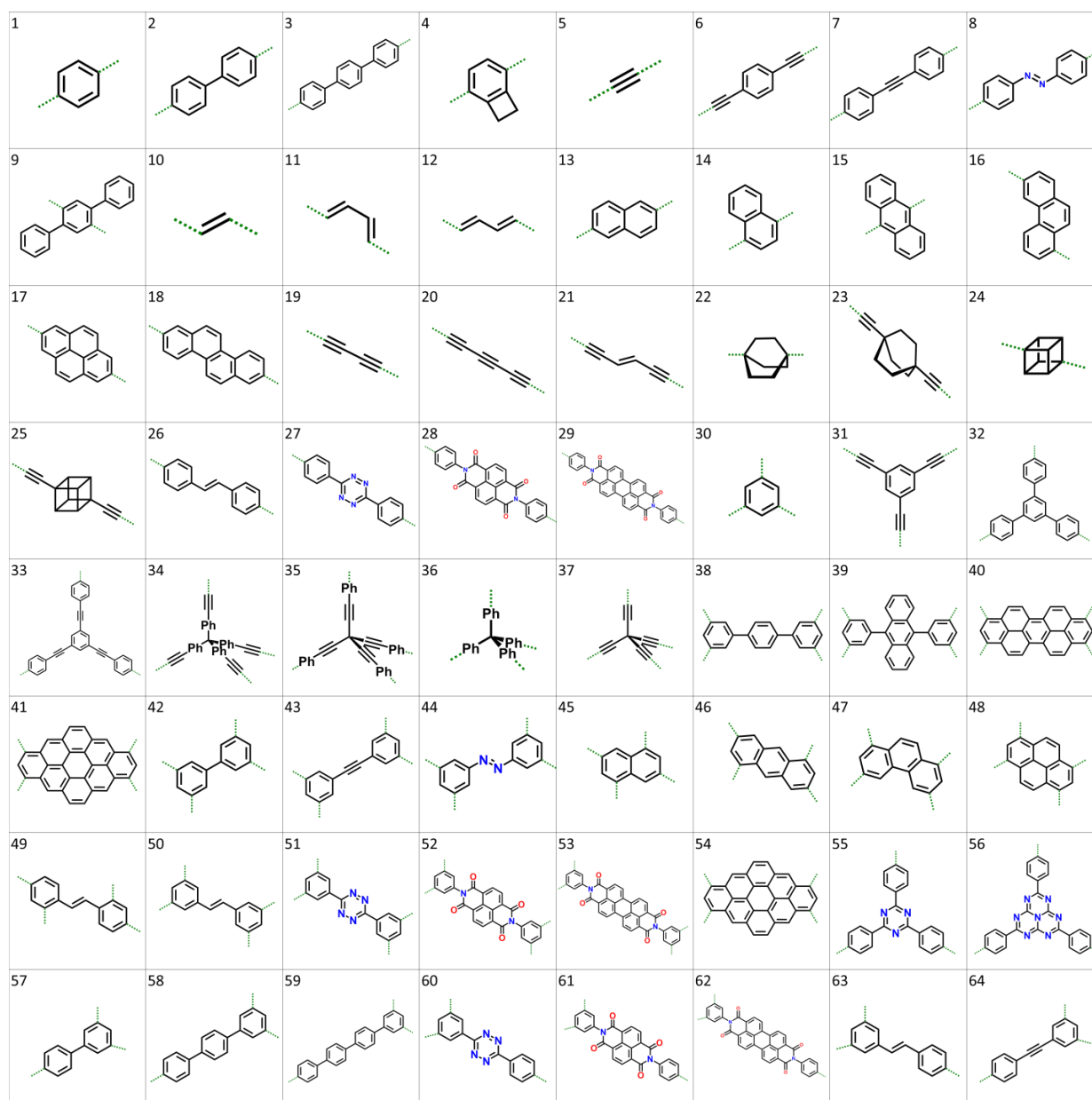


Figure 3.8: The first 64 organic SBUs parameterized to construct the database of hypothetical MOFs.

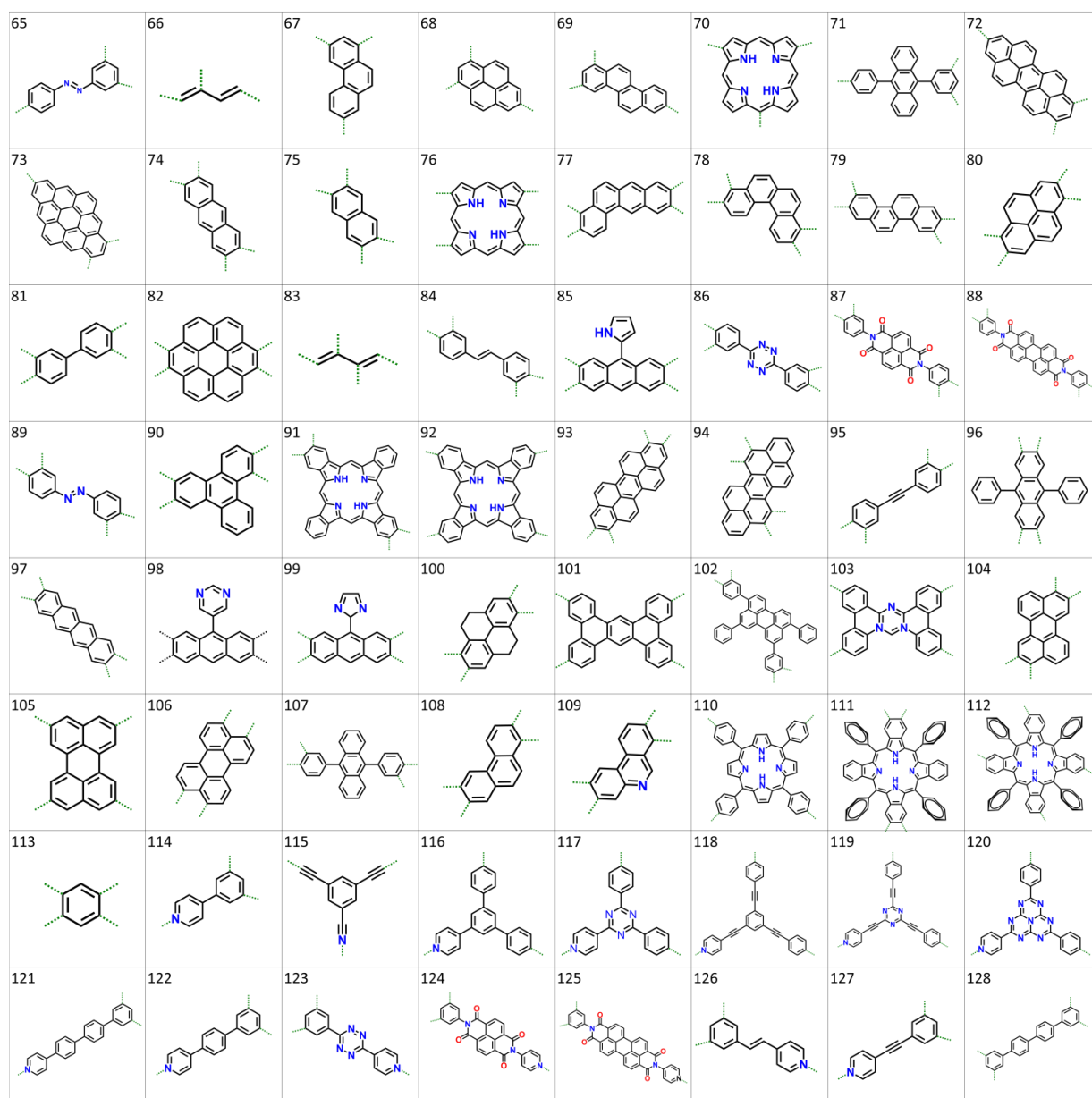


Figure 3.9: organic SBUs 64 – 128 parameterized to construct the database of hypothetical MOFs.

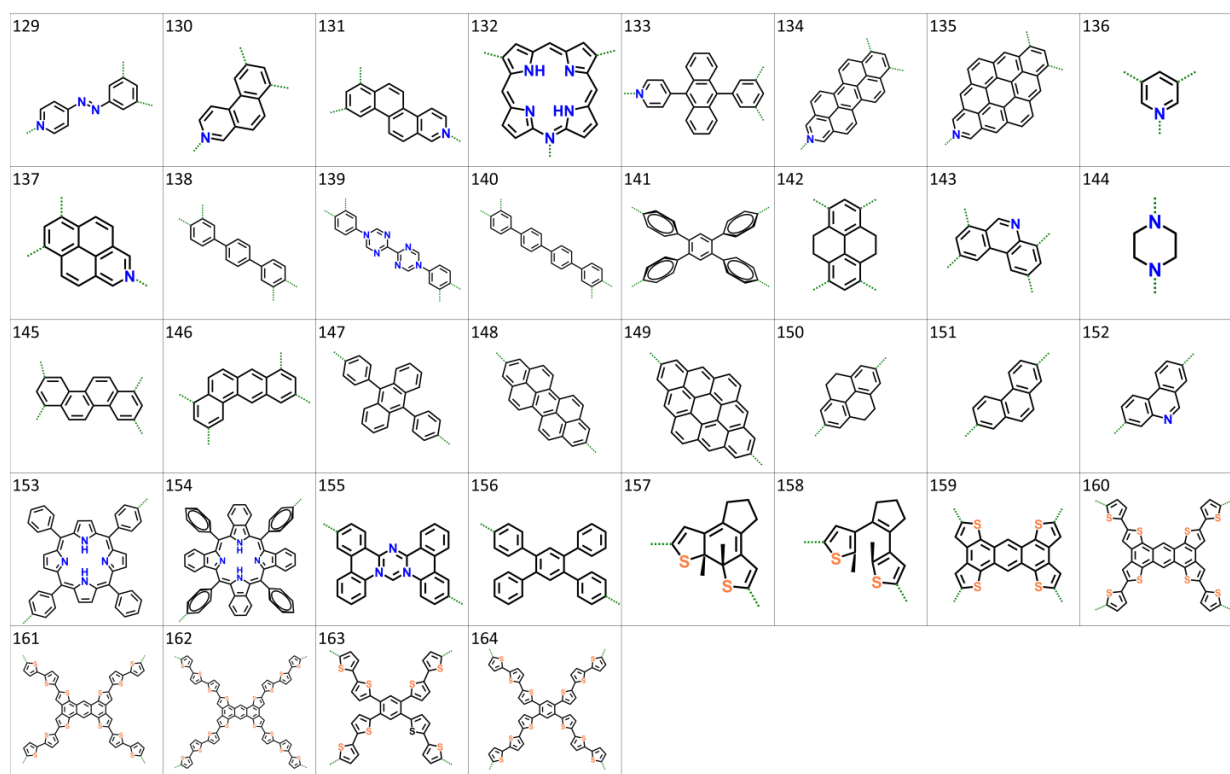


Figure 3.10: The last 45 organic SBUs 129 – 164 parameterized to build MOFs in the hypothetical MOF database.

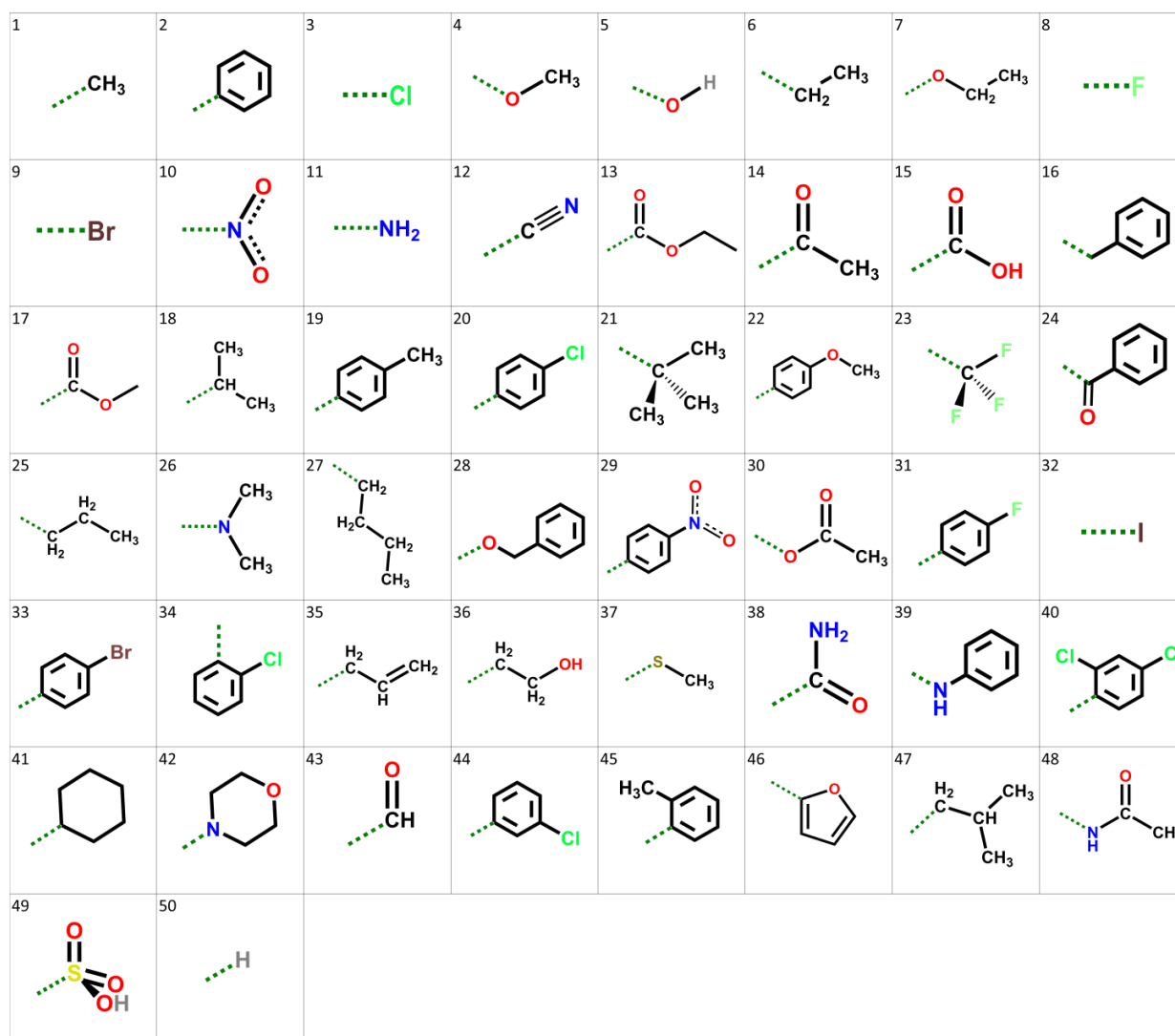


Figure 3.11: The 50 functional groups used to decorate the 21,383 unfunctionalized MOFs.

In this study, the metal SBUs are closely tied to specific topologies, and were extracted from existing crystal structures. Many of the new structures generated with the program arose from parameterizing different organic SBUs to fit into a particular topology supported by the metal SBUs. The metal SBUs (indexed by their order in Figure 3.7) and their topologies are reported in Table 3.1, along with the references where the crystal structure of the experimental MOF had been elucidated.

Table 3.1: A description of the metal SBU geometries and the topologies they assemble into. Also provided is the reference to the experimental work for each metal and the bond parameters used for each topology.

Metal SBU Geometry	Metal (index)	topology
octahedral	Zn(1) ¹⁸ , , Cd(23) ¹⁹	pcu ¹⁸
square planar	Cu(2) ²⁰ , Zn(3) ²¹ , Co(15) ²² , Fe(18) ²³	pcu ²¹ , tbo ²⁰ , nbo ²⁴ , pto ¹³ , fof ²⁵ , rht ²⁶ , pts ²⁷ , fsc ²⁸
square pyramidal	Ni(13) ²⁹ , Cr(4) ³⁰	acs ³⁰
Cube	Cd(5) ³¹ , Mn(6) ³² , Ni(12) ³³	flu ³¹ , the ³³ , bcu [†]
cuboctahedral	Zr(7) ³	ubt ³
tetrahedral	In(11) ³⁴ , Co(14) ³⁵ , Fe(28) ³⁶	dia ³⁴ , pts ²⁷
trigonal prism	Zn(20) ³⁷	utj ³⁷
rod-like	V(9,24) ^{38,39} , In(8,26) ^{40,41} , Ba(10,17) ^{42,43} , Sr(16) ⁴³ , Ni(19) ⁴⁴ , Mg(21) ⁴⁵ , Al(22) ⁴⁶ , Ga(27) ⁴⁷ , Cr(25) ⁴⁸	sra ³⁸ , fri ^{41,42} , tna ⁴³ , irl ⁴⁴ , etb ⁴⁵

†the **bcu** topology parameters were made by hand without experimental support.

3.4 Simulation Details

Force field optimizations were carried out in the GROMACS package,⁴⁹ which requires the user to produce a full description of the force field topology, including terms for bonding and non-bonding interactions, angle harmonics, and proper and improper torsions. Recently, Garberoglio⁵⁰ introduced a program, OBGMX, to automate the process to generate a topology and assign all the UFF¹⁴ parameters and force field functions to fully describe a periodic system for GROMACS, using only the atomic positions of a MOF. Moreover, Garberoglio demonstrated that this implementation yielded reasonable optimization results, showing the atomic root-mean-squared deviation (RMSD) for a set of 37 experimental periodic structures (including zeolitic imidazolate frameworks (ZIFs), covalent organic frameworks (COFs) and MOFs) deviates by at most 1.5 Å from their crystallographic positions, but in most cases the RMSD was less than 0.7 Å.⁵⁰ Since our SBU library includes metal atoms in geometries that have not been parameterized in the UFF, in this work we used a modified version of the OBGMX package, where the bond, angle, and torsion terms on the atoms and bonds adjacent to metal ions were adjusted to be

harmonic potentials with equilibrium geometries taken from the input structure and force constants increased by an order of magnitude to keep the geometry around the metals rigid throughout optimization. We observed that without this correction metal ions would adopt the geometric configurations parameterized by the UFF, resulting in odd structural deformities. For example, the Cu^{2+} ions in the copper paddlewheel SBU would change from a square planar geometry to a tetrahedral one, creating structural deformations in the adjacent organic SBUs.

The optimization in GROMACS was performed in three steps. First, the atomic positions within the structure are optimized, using fixed cell parameters and the L-BFGS optimizer, to reduce all forces to less than $10 \text{ kJ mol}^{-1} \text{ nm}^{-1}$. Following this, quenched molecular dynamics (MD) is run allowing the cell parameters to relax for 2000 steps. After the MD, the L-BFGS optimization is repeated, although in most cases the forces have not increased during the MD. The purpose of implementing this three-step optimization procedure is to accommodate fluctuations of the cell shape and volume as the atoms relax to their energetic minima. In GROMACS, one is not permitted to simultaneously relax the atomic positions and the lattice parameters, as the objective function to be minimized (the total energy) becomes ill-defined when the lattice parameters fluctuate. The consequence of this is that minimizations can become 'stuck' before the desired tolerance is reached, resulting in very long simulation times where the system may not reach the desired energetic minimum.

Atomic point charges were assigned to the atoms of each MOF for the purposes of computing hydrophilic surface areas. Atomic charges were computed using the charge equilibration method of Rappé and Goddard⁵¹ and the MEPO parameters for MOFs detailed in a previous study.⁵² Parameters for atoms which were not included in the MEPO set were taken from UFF.¹⁴ The hydrophilic surface areas were calculated by sampling a random distribution of

points on the van der Waals surface of each atom which could be reached by a spherical probe. Those points which corresponded to atoms with charge magnitudes of 0.3 or higher were considered 'hydrophilic' points. The hydrophilic surface area fraction was then determined by taking the ratio of hydrophilic points to total surface area points. For the hydrophilic surface calculation we used a probe of radius 1.4 Å to represent a water molecule.

Physical properties associated with each MOF were computed with the zeo++⁵³ software using a probe radius of 1.82 Å. Surface areas were computed with 10,000 points on each atom and void volumes were calculated using 100,000 points in the unit cell. The van der Waals radii for each MOF atom was taken from the UFF¹⁴ parameters.

Network topologies of all of the base (unfunctionalized) structures were found using the TOPOS 4.0 software.⁵⁴ Each structure was initially simplified by using the cluster method within the Automatic Description of Structure (ADS) subprogram, which breaks the structure into molecules based on enclosed rings. The fragmented MOF was then simplified, again using ADS, by converting closed rings of atoms into vertices connected to other rings by edges. All 0-, 1-, and 2-coordinated vertices were then removed to produce the underlying nets. These nets were compared to the TOPOS topological database (TTD) provided in TOPOS to yield classifications for each hypothetical MOF.

3.5 Results and Discussion

Using our structure generation algorithm we constructed a total of 1.3 million unique hypothetical functionalized MOF structures. In what follows we consider two separate sets of these structures. The *hMOF* (hypothetical MOF) database consists of 21,383 unfunctionalized framework structures. The analysis of these structures alone can provide insight on structural features such as the topology and symmetry when compared with other hypothetical structure

databases. The complete *hfMOF* (hypothetical functionalized MOF) database is of more interest where the pore chemistry can be tuned to specific applications.

3.5.1 Characterization of Unfunctionalized MOFs

Figure 3.12 demonstrates the diversity of the space groups present in the database of unfunctionalized hypothetical structures. The large number of low-symmetry *P1* structures can be attributed to the choice of building MOFs using two organic SBUs and one metal SBU. The inclusion of multiple SBUs introduces a large amount of diversity, but it also reduces the overall symmetry of the crystal due to the introduction of non-equivalent atoms in the unit cell. As mentioned in Section 3.3.3 there are a number of nets which can support multiple organic SBUs of different lengths. These represented 97% of the *P1* and 99% of the *P-1* space group structures and were **pcu**, **pts**, **tbo**, **nbo**, **fsc**, **sra**, **frl**, **dia**, **etb**, and **irl**. The remaining 3% of structures with *P1* symmetry were MOFs with underlying nets which did not support multiple SBUs of the differing length; however they possessed a combination of two organic SBUs of the same length but different chemistry, introducing non-equivalent atoms in positions of otherwise high symmetry.

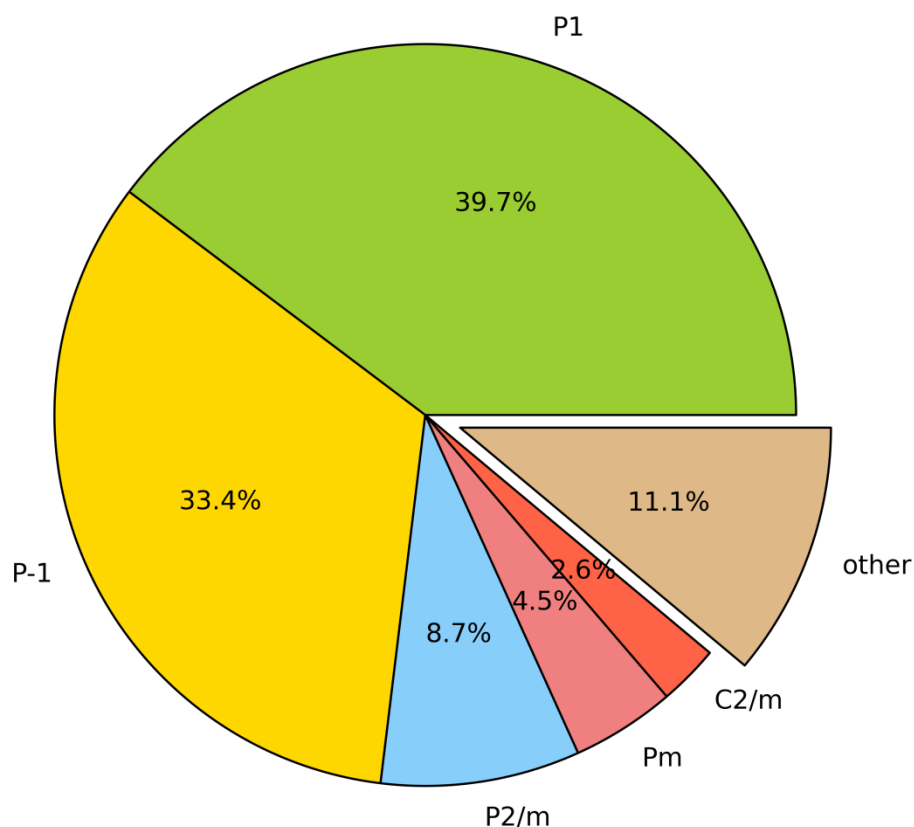


Figure 3.12: Pie chart of the symmetry found in the 23,935 unfunctionalized MOFs in the hypothetical hMOF database. There are 47 space groups found in the 'other' 11% of the database.

It was recently shown that a MOFs net topology can alter not only the pore shape, but also the range of achievable properties.⁵⁵ Thus, as a measure of diversity, it remains to be seen how the topologies used to generate structures in this study compare with those from the database reported from Wilmer *et al.*⁹ The program TOPOS was used to derive the topology for each unfunctionalized, non-catenated structure in their database, which totaled 2518 hypothetical MOFs. This was compared with the topologies of the 21,383 base structures in our database. Topological classification can be made difficult when there exists many different ways to break a periodic structure into nodes and edges. Thus, when a discrepancy arose between the network

topology reported by TOPOS and the related topology reported by either the original work, or by the review by O’Keeffe and Yaghi⁷ the topological classifications from literature were used.

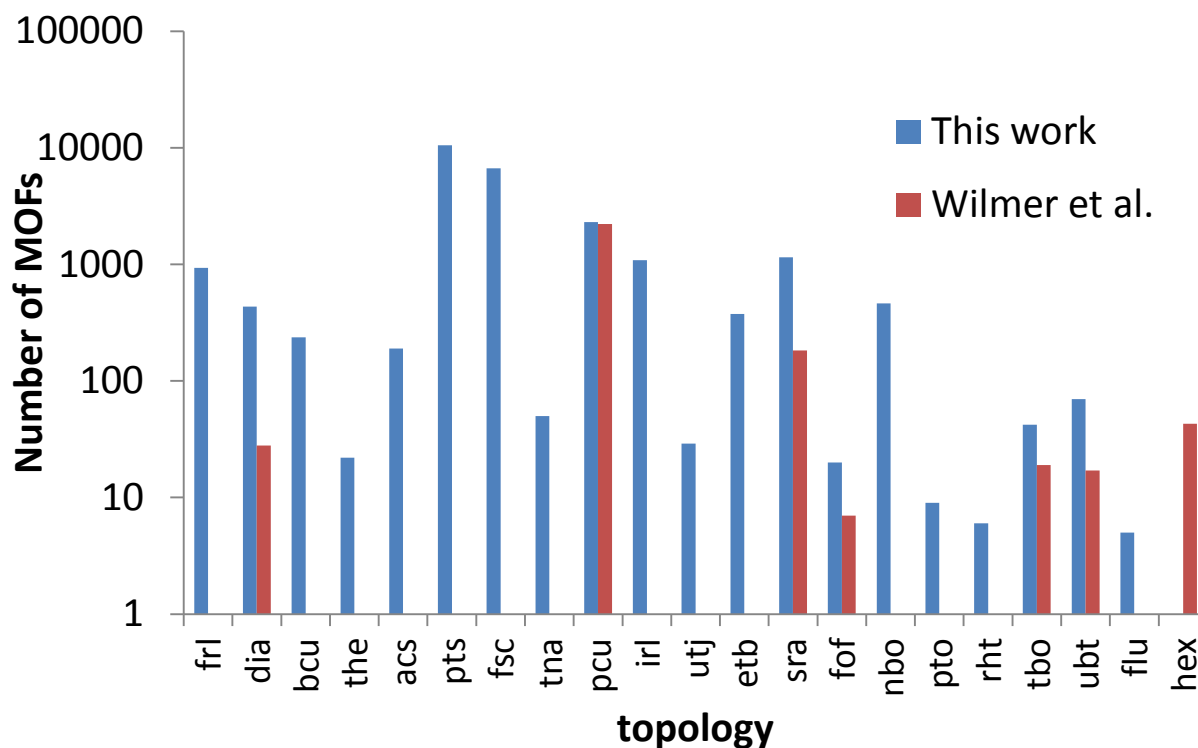


Figure 3.13: Comparison of the topologies used to generate hMOFs in this work, and those in the database provided by ref 9.

It can be seen from Figure 3.13 that the overall quantity of base structures and number of underlying topologies is much larger in the current work. It also demonstrates how the database is represented in terms of topology. For example it is quite evident that a large portion of the database is constituted of the nets **pts** and **fsc**. These topologies represent 44%, and 28% of the MOF database, respectively. On the other hand the number of MOFs for a given topology can be quite small, for example, the topology of **rht** contains only 6 MOFs, this can be increased by including more SBUs parameterized for that particular topology. We note again here that a large reason why some topologies were over-represented in this database was due to the number of organic SBU combinations available to those nets which could support multiple organic SBUs. The **pts** and **fsc** nets, for example, can support multiple organic SBUs, while the **rht** net cannot.

3.5.2 Properties of the functionalized MOF database

The unfunctionalized structures characterized in the previous section were decorated with 50 functional groups using the procedure discussed in Section 3.3.5. Reported in Figure 3.14 is a selection of physical properties computed from the hypothetical functionalized MOF (*hfMOF*) database, shown in green. For comparison, the hypothetical databases of Wilmer *et al.*⁹ and Martin *et al.*¹² are also presented as blue and red histograms, respectively. The normalized frequencies are used so that the databases can be related by the breadth of their properties, not the absolute number. It is interesting to compare the difference in distributions of gravimetric and volumetric surface areas across all the hypothetical databases. We note a sharp spike in the distribution of gravimetric surface area of our *hfMOF* database around 2500 m²/g, as seen in green in Figure 3.14a. We attribute this spike to the MOFs assembled under the **pts** topology, whose distribution is strongly localized in this region (see Figure 3.16b). It is apparent from this distribution that the metal SBU does not contribute significantly to the total gravimetric surface area of these MOFs. These **pts**-based MOFs were assembled with 3 different inorganic SBUs, containing cobalt (59 g/mol), iron (56 g/mol) and indium ions (115 g/mol). A similar spike in the volumetric surface area can be seen in Figure 3.14b which can be seen around 2000 m²/cm³. This peak can be attributed to the limited size of the SBUs we parameterized for MOF generation. These were limited to a range of molecular surface areas centered around 524 Å² ± 219 Å² and an average maximum SBU span (measured between the farthest connection points of each SBU) of 10.2 ± 4.6 Å, which, when assembled into a large number of structures, will limit the breadth of achievable volumetric surface areas. The reason for this focused distribution of SBUs was for the purposes of synthesizability. We elected to build MOFs with SBUs of similar size and chemistry as those commonly reported in experimental MOFs.

The distribution of physical properties of the MOFs in our database more closely resembles that of the MOF database of Wilmer et al.⁹ which is due to the use of similarly-sized SBUs. It is interesting to see that the hPPN database, containing structures of entirely **dia** topology, show such a broad diversity of the presented properties, which can be attributed to the broader range of organic SBU sizes used in their database. Their organic SBUs were obtained from a library of synthesized molecules. However, it remains to be seen if they can react to form extended polymer networks with reported pore diameters of up to 7.6 nm.¹²

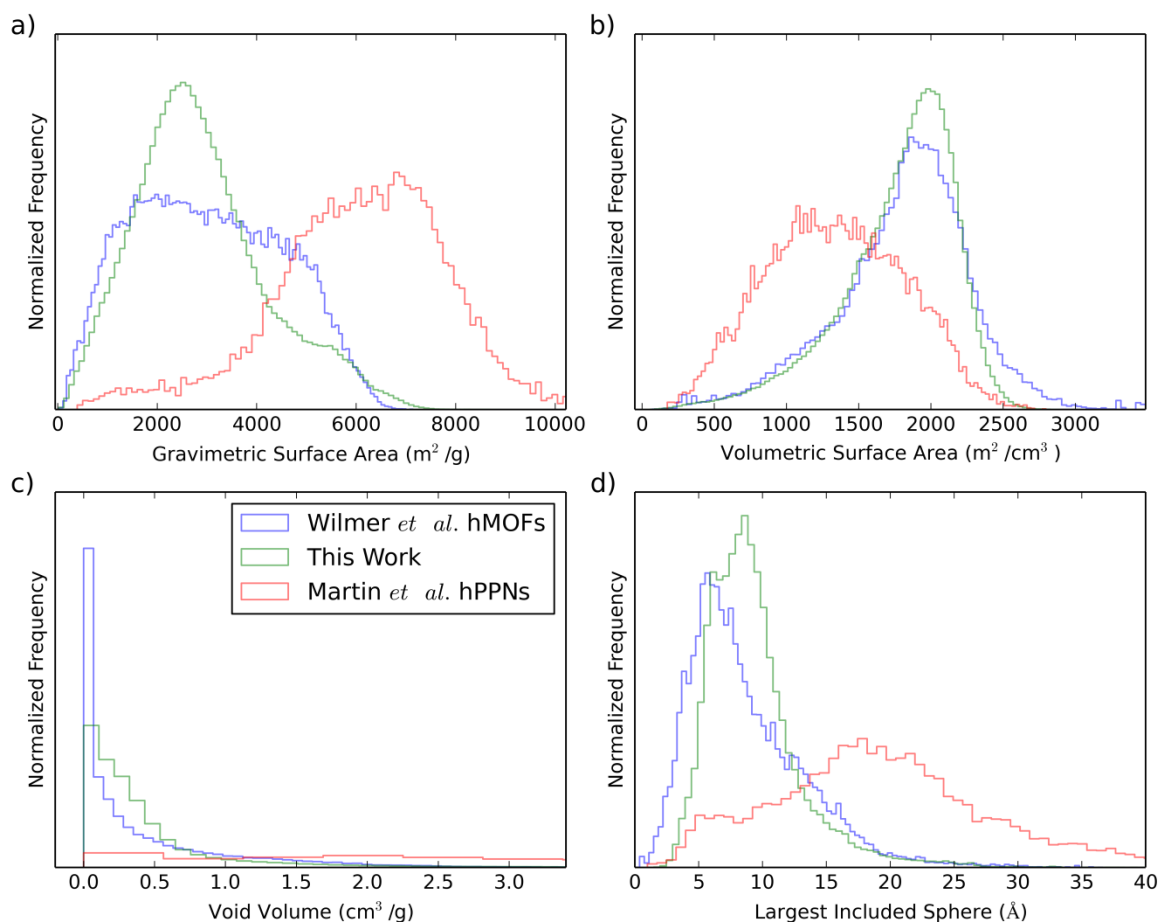


Figure 3.14: Comparison of physical property distributions for the sets of hypothetical structure databases discussed in the text, including the hypothetical MOFs of Wilmer *et al.*⁹ (blue), the hypothetical porous polymer networks (hPPNs) of Martin *et al.*¹² (red) and the hMOFs of this work (green). a) gravimetric surface area, b) volumetric surface area, c) void volume, d) maximum free pore diameter. All properties were calculated with a probe size of 1.82 Å, with the exception of the largest included sphere (which is a measure of the largest probe that can fit into a MOF).

Recently there have been a number of studies demonstrating the correlation between physical properties of porous materials with performance in CO₂ capturing processes, CH₄ storage, and Xe/Kr separation. It was found that a pore limiting diameter of 4.5-4.8 Å was ideal for separating noble gases from N₂ in a study that screened 16 experimental MOFs with pore limiting diameters ranging from 3.4 – 7.8 Å.⁵⁶ Wilmer *et al.* screened their database of MOFs to find gravimetric surface areas of approximately 1500 m²/g provided the best working capacity of CO₂ in a carbon capture process from flue gas.⁵⁷ It can be seen from Figure 3.14 that our

database contains MOFs in a wide range of physical properties in these ideal regions, showing potential for screening in a variety of applications. However, these structure-property relationships are a relatively simple look at a complex pore environment, where, depending on the application, activity will depend highly on local chemical features in each MOF. Complex models have been developed to use more chemical features of materials to predict their properties,⁵⁸ however the complexity of the models can make it difficult to extract physical meaning from them, therefore, instead of a full characterization of the materials in this work, we show chemical diversity using hydrophilic surface area as an indicator. Figure 3.15a reports the hydrophilic surface area distribution in the database, which is closely related to the exposure of the surface to highly polarized atoms. We see a large distribution of hydrophilic surface areas near 0 with a tail which reaches 3000 m²/g. The hump located near 500 m²/g is primarily contributed by the MOFs present in the **dia**, **acs**, and **pts** topologies. This can be seen by the hydrophilic surface area distribution reported for these topologies in Figure 3.16a, which shows 50% of the values reported for **dia** MOFs are between 400 and 800 m²/g, while the same distribution for **acs** and **pts** MOFs are seen between 0 and 600 m²/g.

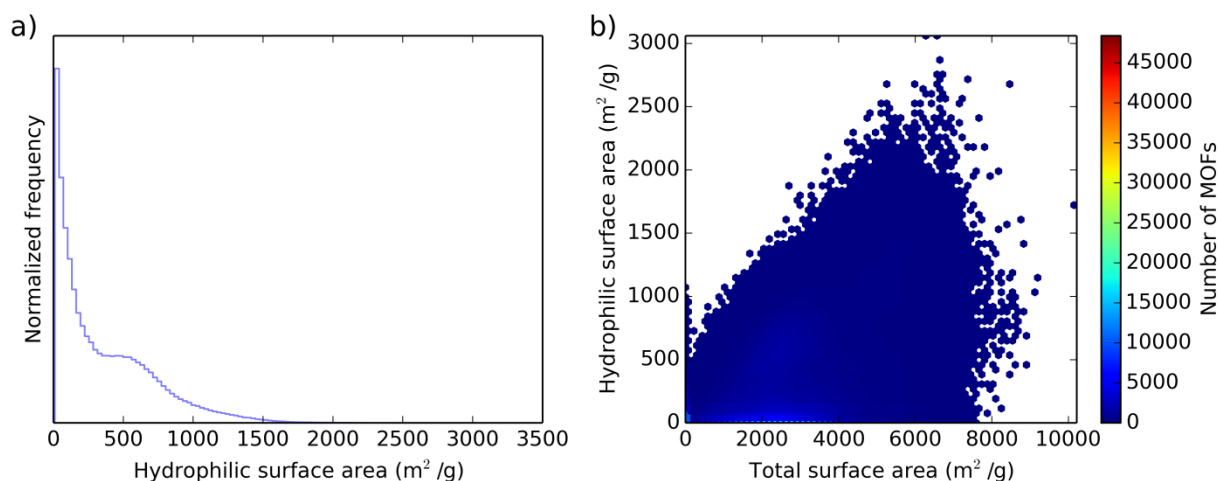


Figure 3.15: a) Distribution of hydrophilic surface areas in the hfMOF database. b) plot of hydrophilic vs. total gravimetric surface area.

The hydrophilic surface area has a very low correlation with all of the physical properties reported, including the related total gravimetric surface area as seen in Figure 3.15b, with Spearman and Pearson correlations of 0.28 and 0.4, respectively. Thus looking at only a selection of properties (particularly macroscopic physical properties) may not give a full picture of relevant features which affect structure-property relationships. It should be noted that the hydrophilic surface area relies on the calculation of partial atomic charges on the atoms, which for the case of 1.8 million MOFs, required a parameterized charge equilibration method.⁵² Due to the parameterized nature of the charge derivation method, this property can be a source of error. Regardless, it is encouraging that hydrophilic surface area was found to be uncorrelated with the reported physical properties, as it demonstrates this measure of chemical diversity samples a diverse range of structures. This can be useful for such applications where chemistry will play a dominant role in the activity of the material, such as CO₂ capture and storage, where the affinity of CO₂ with the surface of a MOF will be determined largely by its electrostatic interaction with the significant quadrupole moment in CO₂.

Figure 3.16 gives a description of how each topology contributed to the diversity of the hfMOF database. It is worth noting that topologies **nbo** and **pcu** are assembled using many of the same SBUs, however their property distributions are centered on different values, with **nbo** showing larger median values over all reported properties. This is likely due to the naturally larger pore sizes adopted with **nbo**. For example, two MOFs assembled with the **pcu** and **nbo** topologies using the Cu paddlewheel and benzene dicarboxylate SBU (as well as a pyrazine SBU to bridge two Cu paddlewheels in the **pcu** topology) yield void fractions of 0.19 and 0.48, respectively. This implies that, when functionalizing these structures with the 50 functional groups reported in Figure 3.11, **nbo** can support a more diverse range of functional groups, as there is less chance of atomic collisions due to its more spacious pores. The introduction of functional groups played a large role in the property distributions reported in Figure 3.16, showing an overall reduction in the median values for each topology by roughly 18%, while standard deviations increased by 210% compared to the unfunctionalized structures (not shown). It is notable that the **dia** topology supports such a diverse range of properties, despite sharing many of the same linear organic SBUs found in structures with the **pcu** and **nbo** topologies. It is therefore clear that topology can affect many of the physical and chemical properties of the MOF pores, which confirms what is discussed elsewhere.⁵⁵

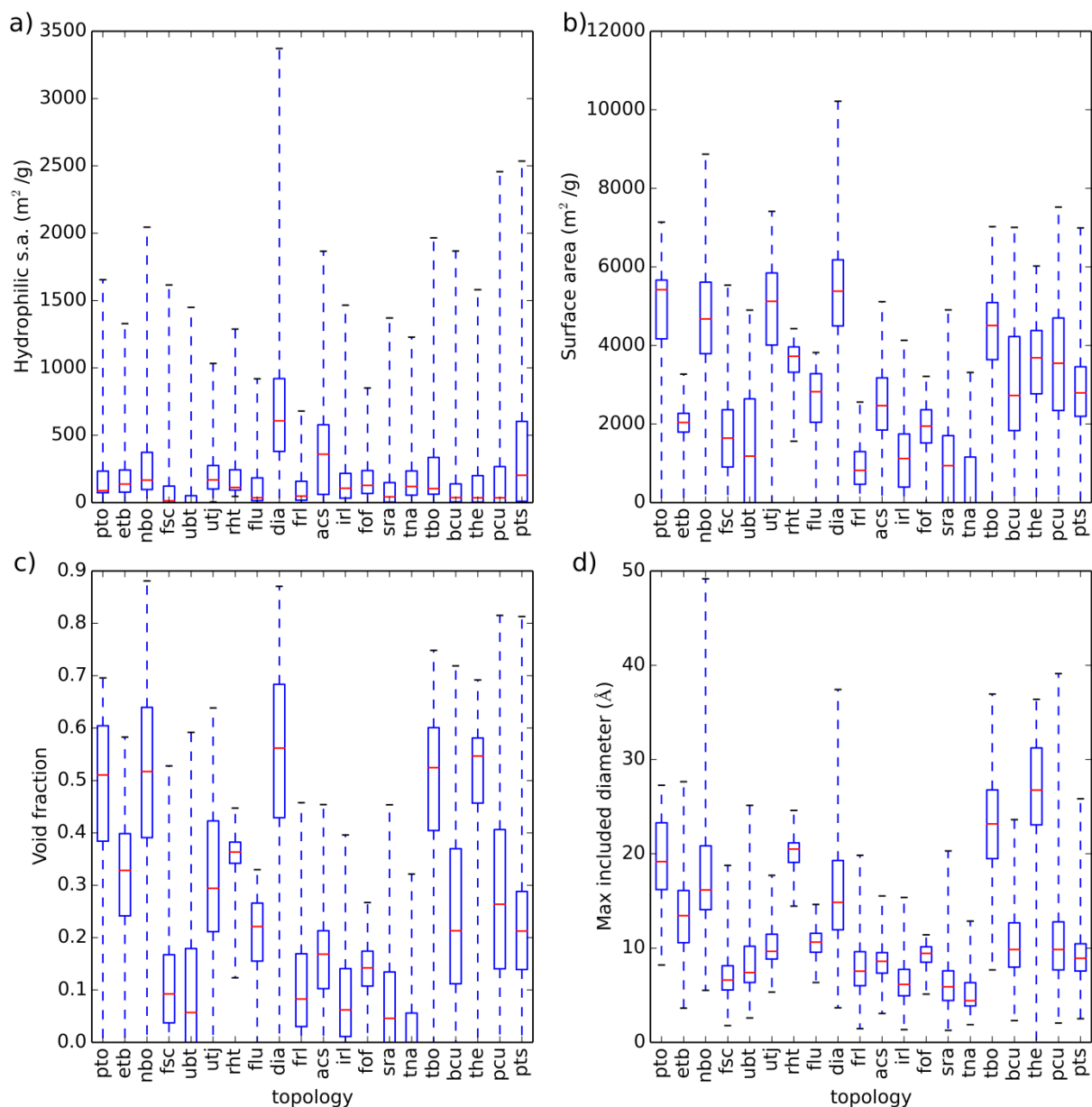


Figure 3.16: Property distributions of the hfMOF database based on topology. a) hydrophilic surface area, b) gravimetric total surface area, c) void fraction, and d) max free probe diameter. The red bar represents the median value, and the blue box shows where 50% of the data lies. The dotted line tails above and below the boxes represent the min and max values found for each property.

3.6 Conclusions

A database of hypothetical MOF structures was generated for the purposes of screening a wide variety of applications. To provide as robust a database as possible, structures were built with a diverse range of topological, chemical and physical properties. The 21,383

unfunctionalized structures were constructed using an algorithm which joins SBUs together based on orientation parameters fit to a particular network topology. The symmetry was computed for each hypothetical MOF, which was then used to decorate them from a selection of 50 functional groups. Each MOF generated with the program was optimized at the molecular mechanics level to reduce the strain imposed by the algorithm when joining rigid SBUs together with relatively few degrees of freedom. Reported is an expanded database of 1.3 million structures with more chemical and topological diversity than the previously reported materials databases provided by Wilmer *et al.*⁹ and Martin *et al.*¹²

3.7 References

- (1) Farha, O. K.; Eryazici, I.; Jeong, N. C.; Hauser, B. G.; Wilmer, C. E.; Sarjeant, A. a; Snurr, R. Q.; Nguyen, S. T.; Yazaydin, a Ö.; Hupp, J. T. *J. Am. Chem. Soc.* **2012**, *134*, 15016.
- (2) Farha, O. K.; Yazaydin, 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*, 944.
- (3) Cavka, J. H.; Jakobsen, S.; Olsbye, U.; Guillou, N.; Lamberti, C.; Bordiga, S.; Lillerud, K. P. *J. Am. Chem. Soc.* **2008**, *130*, 13850.
- (4) Jasuja, H.; Zang, J.; Sholl, D. S.; Walton, K. S. *J. Phys. Chem. C* **2012**, *116*, 23526.
- (5) Katz, M. J.; Mondloch, J. E.; Totten, R. K.; Park, J. K.; Nguyen, S. T.; Farha, O. K.; Hupp, J. T. *Angew. Chem. Int. Ed. Engl.* **2014**, *53*, 497.
- (6) Ockwig, N. W.; Delgado-Friedrichs, O.; O’Keeffe, M.; Yaghi, O. M. *Acc. Chem. Res.* **2005**, *38*, 176.
- (7) O’Keeffe, M.; Yaghi, O. M. *Chem. Rev.* **2012**, *112*, 675.
- (8) O’Keeffe, M.; Peskov, M. A.; Ramsden, S. J.; Yaghi, O. M. *Acc. Chem. Res.* **2008**, *41*, 1782.
- (9) Wilmer, C. E.; Leaf, M.; Lee, C. Y.; Farha, O. K.; Hauser, B. G.; Hupp, J. T.; Snurr, R. Q. *Nat. Chem.* **2012**, *4*, 83.
- (10) Sikora, B. J.; Winnegar, R.; Proserpio, D. M.; Snurr, R. Q. *Microporous Mesoporous Mater.* **2014**, *186*, 207.
- (11) Martin, R. L.; Lin, L.-C.; Jariwala, K.; Smit, B.; Haranczyk, M. *J. Phys. Chem. C* **2013**, *117*, 12159.

- (12) Martin, R. L.; Simon, C. M.; Smit, B.; Haranczyk, M. *J. Am. Chem. Soc.* **2014**, *136*, 5006.
- (13) Furukawa, H.; Go, Y. B.; Ko, N.; Park, Y. K.; Uribe-Romo, F. J.; Kim, J.; O'Keeffe, M.; Yaghi, O. M. *Inorg. Chem.* **2011**, *50*, 9147.
- (14) Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. *J. Am. Chem. Soc.* **1992**, *114*, 10024.
- (15) O'Boyle, N. M.; Banck, M.; James, C. a; Morley, C.; Vandermeersch, T.; Hutchison, G. R. *J. Cheminform.* **2011**, *3*, 33.
- (16) Togo, A. spglib <http://spglib.sourceforge.net/>.
- (17) Ertl, P. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 374.
- (18) Eddaoudi, M.; Kim, J.; Rosi, N. L.; Vodak, D. T.; Wachter, J.; O'Keeffe, M.; Yaghi, O. M. *Science (80-.)*. **2002**, *295*, 469.
- (19) Zhang, J.; Wojtas, L.; Larsen, R. W.; Eddaoudi, M.; Zaworotko, M. J. *J. Am. Chem. Soc.* **2009**, *131*, 17040.
- (20) Chui, S. S. Y.; Lo, S. M. F.; Charmant, J. P. H.; Orpen, A. G.; Williams, I. D. *Science (80-.)*. **1999**, *283*, 1148.
- (21) Ma, B.-Q.; Mulfort, K. L.; Hupp, J. T. *Inorg. Chem.* **2005**, *44*, 4912.
- (22) Zheng, B.; Bai, J.; Duan, J.; Wojtas, L.; Zaworotko, M. J. *J. Am. Chem. Soc.* **2011**, *133*, 748.
- (23) Li, Y.-W.; Zhao, J.-P.; Wang, L.-F.; Bu, X.-H. *CrystEngComm* **2011**, *13*, 6002.
- (24) Eddaoudi, M.; Kim, J.; O'Keeffe, M.; Yaghi, O. M. *J. Am. Chem. Soc.* **2002**, *124*, 376.
- (25) Chen, B.; Ockwig, N. W.; Millward, A. R.; Contreras, D. S.; Yaghi, O. M. *Angew. Chem. Int. Ed. Engl.* **2005**, *44*, 4745.
- (26) Farha, O. K.; Yazaydin, 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*, 944.
- (27) Chen, B.; Eddaoudi, M.; Reineke, T. M.; Kampf, J. W.; O'Keeffe, M.; Yaghi, O. M. *J. Am. Chem. Soc.* **2000**, *122*, 11559.
- (28) Bi, M.; Li, G.; Hua, J.; Liu, Y.; Liu, X.; Hu, Y.; Shi, Z.; Feng, S. *Cryst. Growth Des.* **2007**, *7*, 2066.
- (29) Vuong, G.-T.; Pham, M.-H.; Do, T.-O. *CrystEngComm* **2013**, *15*, 9694.
- (30) Férey, G.; Mellot-Draznieks, C.; Serre, C.; Millange, F.; Dutour, J.; Surblé, S.; Margiolaki, I. *Science* **2005**, *309*, 2040.

- (31) Chun, H.; Kim, D.; Dybtsev, D. N.; Kim, K. *Angew. Chem. Int. Ed. Engl.* **2004**, *43*, 971.
- (32) Dincă, M.; Dailly, A.; Liu, Y.; Brown, C. M.; Neumann, D. A.; Long, J. R. *J. Am. Chem. Soc.* **2006**, *128*, 16876.
- (33) Colombo, V.; Galli, S.; Choi, H. J.; Han, G. D.; Maspero, A.; Palmisano, G.; Masciocchi, N.; Long, J. R. *Chem. Sci.* **2011**, *2*, 1311.
- (34) Yang, S.; Lin, X.; Lewis, W.; Suyetin, M.; Bichoutskaia, E.; Parker, J. E.; Tang, C. C.; Allan, D. R.; Rizkallah, P. J.; Hubberstey, P.; Champness, N. R.; Mark Thomas, K.; Blake, A. J.; Schröder, M. *Nat. Mater.* **2012**, *11*, 1.
- (35) Köferstein, R.; Robl, C. *Zeitschrift für Anorg. und Allg. Chemie* **2003**, *629*, 2186.
- (36) Fateeva, A.; Devautour-Vinot, S.; Heymans, N.; Devic, T.; Grenèche, J.-M.; Wuttke, S.; Miller, S.; Lago, A.; Serre, C.; de Weireld, G.; Maurin, G.; Vimont, A.; Férey, G. *Chem. Mater.* **2011**, *23*, 4641.
- (37) Chun, H.; Seo, J. *Inorg. Chem.* **2009**, *48*, 9980.
- (38) Barthelet, K.; Marrot, J.; Riou, D.; Férey, G. *Angew. Chemie Int. Ed.* **2002**, *41*, 281.
- (39) Barthelet, K.; Adil, K.; Millange, F.; Serre, C.; Riou, D.; Férey, G. *J. Mater. Chem.* **2003**, *13*, 2208.
- (40) Volkringer, C.; Loiseau, T.; Férey, G. *Solid State Sci.* **2009**, *11*, 29.
- (41) Lin, Z.; Jiang, F.; Yuan, D.; Chen, L.; Zhou, Y.; Hong, M. *Eur. J. Inorg. Chem.* **2005**, *2005*, 1927.
- (42) Taylor, J. M.; Vaidhyanathan, R.; Iremonger, S. S.; Shimizu, G. K. H. *J. Am. Chem. Soc.* **2012**, *134*, 14338.
- (43) Vaidhyanathan, R.; Mahmoudkhani, A. H.; Shimizu, G. K. *Can. J. Chem.* **2009**, *87*, 247.
- (44) Kurmoo, M.; Kumagai, H.; Akita-Tanaka, M.; Inoue, K.; Takagi, S. *Inorg. Chem.* **2006**, *45*, 1627.
- (45) 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*, 1018.
- (46) Alaerts, L.; Maes, M.; Giebel, L.; Jacobs, P. A.; Martens, J. A.; Denayer, J. F. M.; Kirschhock, C. E. A.; De Vos, D. E. *J. Am. Chem. Soc.* **2008**, *130*, 14170.
- (47) Volkringer, C.; Meddouri, M.; Loiseau, T.; Guillou, N.; Marrot, J.; Férey, G.; Haouas, M.; Taulelle, F.; Audebrand, N.; Latroche, M. *Inorg. Chem.* **2008**, *47*, 11892.
- (48) Guillou, N.; Millange, F.; Walton, R. I. *Chem. Commun. (Camb)*. **2011**, *47*, 713.
- (49) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435.

- (50) Garberoglio, G. J. *Comput. Chem.* **2012**, 33, 2204.
- (51) Rappe, A. K.; Goddard, W. A. *J. Phys. Chem.* **1991**, 95, 3358.
- (52) Kadantsev, E. S.; Boyd, P. G.; Daff, T. D.; Woo, T. K. *J. Phys. Chem. Lett.* **2013**, 4, 3056.
- (53) Willems, T. F.; Rycroft, C. H.; Kazi, M.; Meza, J. C.; Haranczyk, M. *Microporous Mesoporous Mater.* **2012**, 149, 134.
- (54) Blatov, V. a. *Struct. Chem.* **2012**, 23, 955.
- (55) Martin, R.; Haranczyk, M. *J. Chem. Theory Comput.* **2013**, 9, 2816.
- (56) Parkes, M. V.; Staiger, C. L.; Perry, J. J.; Allendorf, M. D.; Greathouse, J. a. *Phys. Chem. Chem. Phys.* **2013**, 15, 9093.
- (57) Wilmer, C. E.; Farha, O. K.; Bae, Y.-S.; Hupp, J. T.; Snurr, R. Q. *Energy Environ. Sci.* **2012**, 5, 9849.
- (58) Fernandez, M.; Boyd, P. G.; Daff, T. D.; Aghaji, M. Z.; Woo, T. K. *J. Phys. Chem. Lett.* **2014**, 5, 3056.

4 Ultra-Fast Generation of Electrostatic Potentials in MOFs

The contents of this chapter were published in part in Kadantsev, E. S.; Boyd, P. G.; Daff, T. D.; Woo, T. K. *J. Phys. Chem. Lett.* **2013**, 4, 3056. My role in this work was to develop the training and validation sets, compute the adsorption for each MOF with all the different charge generation schemes, and evaluate the adsorption correlations between them.

4.1 Abstract

The charge equilibration (QEq) method has been parametrized to reproduce the ab-initio-derived electrostatic potential in a large and diverse training set of 543 metal organic frameworks (MOFs) containing the most popular Zn, Cu, and V structural building units (SBUs), 52 different organic carboxylate- and nitrogen-capped SBUs, and the 17 functional groups. The MOF electrostatic-potential-optimized charge scheme, or MEPO-QEq, was validated by evaluating the CO₂ uptake and heats of adsorption in the 543 member training set and a non-overlapping 693 member validation set. Compared with the results obtained from ab-initio-derived charges, the MEPO-QEq charges give Pearson (linear) and Spearman (rank-order) correlation coefficients of >0.97 for these two sets. MEPO-QEq enables near-ab-initio quality non-bonded electrostatic interactions to be evaluated using the fast QEq method for fast and accurate virtual high-throughput screening of gas-adsorption properties in MOFs.

4.2 Introduction

When evaluating the CO₂ adsorption in MOFs with computer simulations, the guest molecule-framework interaction energies can be evaluated millions of times to determine the adsorption properties. As a result, force fields are used where the van der Waals and partial atomic charge parameters are critical. While the van der Waals (i.e., Lennard-Jones) parameters are generally transferable, the partial atomic charges typically need to be evaluated for each

individual MOF. Here electrostatic potential fitted charges that are derived from an ab initio quantum mechanical calculation are generally considered to be the most appropriate. These charges can be derived from density functional theory (DFT)^{1,2} calculations on the individual MOF building units, or more appropriately, from periodic DFT calculations of the whole MOF using the REPEAT³ or related.⁴⁻⁶ However, such computation-demanding calculations cannot be practically performed when screening hundreds of thousands of MOFs with diverse building units and functional groups.

One approach for quick derivation of atomic charges is the Rappé-Goddard charge equilibration method (QEq).⁷ In this method, the hardness and electronegativity parameters for each atom type need to be defined and an energy expression is minimized to yield partial atomic charges. Sholl and coworkers⁸ used QEq charges to screen 500 real MOFs for their gas adsorption properties. On the basis of a comparison of the CO₂ uptake from an ab initio treatment of four Zn-containing MOFs, it was concluded that the QEq method with the standard parameters is capable of capturing electrostatic interactions with a reasonable accuracy. In another recent work,⁹ Wilmer *et al.* proposed an “extended” QEq method, which we will abbreviate EQEq, where the parameters of the metal cations are adjusted to reflect their charge state, while the parameters of other elements are the same as those in the standard QEq method. The EQEq method was tested on a set of 12 MOFs, and the CO₂ uptake derived from this method was compared with the uptake obtained using the standard QEq charges, DFT-derived charges, and experimental CO₂ uptakes. While the EQEq results compared well with those of the DFT-derived charges, the results from the standard QEq method for copper containing MOFs showed large deviations. Using the EQEq method, Snurr and coworkers have screened a database of ~130 000 hypothetical MOFs for CO₂ uptake and recommended functionalization

with F and Cl groups to promote high CO₂ uptake.¹⁰ The question with respect to the applicability of QEq, its extensions, and parameterizations for describing electrostatics in MOFs is clearly of great urgency considering that the method is being applied to screen databases of hundreds of thousands of MOFs while the comparison with *ab initio* electrostatics is limited to only a few tens of MOFs. In this work, we present a set of QEq parameters that have been trained to reproduce the DFT-derived electrostatic potential in a large and diverse training set of 543 hypothetical MOFs. The parameterization was validated by comparing the CO₂ uptakes and heats of adsorption (HOA) on this training set and another set of 693 hypothetical MOFs, which were not part of the training set. Our parameterization, which we term MEPO-QEq (MOF electrostatic potential optimized), is deemed to be robust and allows one to employ the fast QEq charge generation scheme to describe electrostatic interactions in MOFs close to quantum-mechanical level of theory. In turn, this enables fast and accurate virtual high-throughput discovery of novel materials for CO₂ capture.

4.3 Methods

4.3.1 Charge Equilibration (QEq) Methods

The QEq equations postulate that the charge of a system is spatially dependent on the atomic positions, and will re-distribute to minimize the total system energy. In classical simulations, the system charge is described by point charges assigned to each atom, whose fluctuations are governed by fundamental atomic parameters (electronegativity and hardness for each atomic species). Specifically, it states that The N_{ATOM} atomic charges can be obtained by minimizing the following energy equation with respect to each atomic charge Q_i .

$$E(Q_1, Q_2, \dots, Q_N) = \sum_{i=1}^N \left(E_{i0} + Q_i \chi_i + \frac{1}{2} Q_i^2 J_{ii} \right) + \frac{1}{2} \sum_{i=1}^N \sum_{j \neq i}^N Q_i Q_j J_{ij} \quad 4.1$$

$$-\lambda \left(\sum_{k=1}^N Q_k - Q_T \right)$$

where E_{i0} is the energy of atom i at the charge neutral state, χ_i and J_{ii} are the electronegativity and hardness parameters for atom i . J_{ij} represents the coulomb interaction between two atoms i and j . The final term which includes the lagrange multiplier (λ) arises from asserting that the sum of individual atomic charges is equal to the total system charge Q_T . This expression of the energy as a function of charge is generated from a Taylor expansion of the energy around the charges, a brief discussion on the derivation of the equations and the generation of atomic charges from these equations is presented in the appendix to this chapter (including considerations for periodic systems). The QEq equations can be written in the matrix form as

$$\mathbf{G} \begin{pmatrix} \mathbf{Q} \\ \lambda \end{pmatrix} = \begin{pmatrix} -\mathbf{X} \\ 0 \end{pmatrix}, \quad \mathbf{G} = \begin{pmatrix} \mathbf{J} & \mathbf{1} \\ \mathbf{1}^\dagger & 0 \end{pmatrix} \quad 4.2$$

where the charges are represented as a $N_{ATOM} \times 1$ column vector $\mathbf{Q} = (Q_1, Q_2, \dots)^\dagger$, λ is the Lagrange multiplier that stems from the charge conservation condition, $\mathbf{X} = (\chi_1, \chi_2, \dots)^\dagger$ is the electronegativity column vector and $\mathbf{J}$ is the interaction matrix

In the standard QEq method with periodic boundary conditions, the interaction matrix element between two atoms is given by the lattice summation of the electrostatic interaction between distributed unit charges ρ_A , ρ_B or, in the case of the self-interaction, by the idempotential

$$J_{AB} = \sum_{\vec{r}} (1 - \delta_{A,B} \delta_{\vec{r},0}) \int \int \frac{\rho_A(\vec{r}_1; \vec{R}_A) \rho_B(\vec{r}_2; \vec{R}_B + \vec{T})}{|\vec{r}_1 - \vec{r}_2|} d\vec{r}_1 d\vec{r}_2 + \delta_{A,B} J_{AA}^0 \quad 4.3$$

where the individual unit cells are referenced by lattice vectors $\vec{T}$, the atoms within the unit cell are referenced by vectors $\vec{R}_A$, $\vec{R}_B$, and the unit charges are described by a single normalized s-type Slater orbital (STO). For an atom A with an outer valence orbital belonging to nth shell, the

STO's is $\rho_A(\vec{r}; \vec{R}_A) = N_n |\vec{r} - \vec{R}_A|^{n-1} \exp\{-\xi_A |\vec{r} - \vec{R}_A|\}$. The decay parameter ξ_A is tabulated, for example, in the Universal Force Field for each atom's type and is inversely proportional to the atom's covalent radius. Throughout this work, we employ the decay parameters which were taken from the OpenBabel software.¹¹ The same decay parameters were used by Sholl *et al.* to screen 500 real MOFs for CO₂ uptake.⁸ Finally, we mention that in the standard QEq method, the QEq equations have to be solved self-consistently if hydrogen is present in the system. The self-consistency is introduced due to the fact that the hydrogen idempotential is charge-dependent.⁷

We have implemented the standard QEq method in conjunction with periodic boundary conditions in a stand-alone software package. This package was developed in C programming language. In our calculations, we use a simple linear mixing scheme with a mixing parameter of 0.3 and terminate the self-consistent procedure if the change in the charges per atom is less than 10^{-5} . Once the QEq equations are solved and the charges are determined, one can obtain the electrostatic potential at any point in the solid through Ewald summations.

In the case of "extended" QEq method (EQEq)⁹, we use implementation by the authors of Ref. 9 obtained from the Supporting Information. The EQEq method differs from the standard QEq in 3 respects: 1) The QEq energy expression is modified by considering Taylor series expansion around charge-positive metal cations; 2) The self-consistency for hydrogen is dropped and the Coulomb pair interactions are screened by a finite dielectric constant; 3) The form of distributed charges is taken as 1s Gaussians instead of Slater type orbitals (STOs). The two ad-hoc parameters required for the EQEq method were set to the values recommended in Ref 9. The dielectric constant was taken to be $\epsilon_R = 1.67$ and the electron affinity for hydrogen was taken as $I_0 = -2.0$ eV. A 5x5x5 supercell for the reciprocal and real space was used to perform Ewald sums. The charge centers assigned to each metal were as follows: Cu = +2, Zn = +2, V = +4.

The ionization potentials and electron affinities for the Taylor series expansion were the experimentally derived values centered on the above ionic species.

4.3.2 Fitting Procedure

The MEPO-QEq parameters were developed by varying the atomic electronegativity and hardness in the standard QEq scheme to minimize the "distance" between QEq and *ab initio* derived quantities. The decay parameters (exponents of STOs) are not varied. Our QEq parameterization (MEPO-QEq) was trained to reproduce *ab initio* electrostatic potentials in a large and diverse set of MOFs. The first step in development of a new parameterization is the generation of the "target" data. In our case, for a given set of MOFs, we perform self-consistent-field DFT^{1,2} calculations of electrostatic potential with PBE exchange-correlation functional^{12,13} and PAW *ab initio* pseudopotentials as implemented in the VASP package.^{14–16} A plane wave cutoff of 500 eV was used. The energy tolerance of 10^{-5} eV was used to terminate the self-consistent cycle. The reciprocal space was sampled using Γ -point only. Then, for each MOF, the monopole representation of this *ab initio* electrostatic potential is derived using the REPEAT method.³ The set of all *ab initio* charges and MOF geometries constitute the input ("target") data.

For each MOF, our code accepts the structure's geometry, *ab initio* derived charges ($\{q^{DFT}\}$), and QEq parameters (electronegativity and hardness for each atomic type) as an input and calculates new charges using QEq formalism: $\{q^{QEq}\}$. For each MOF we then calculate the difference between *ab initio* and QEq charges: $\{\Delta q = q^{QEq} - q^{DFT}\}$ and construct the electrostatic potential due to this charge difference: $V(\vec{r})[\Delta q]$. If the norm of this potential is small, then the given QEq parameterization is deemed as a good description of the *ab initio* potential. In practice, the electrostatic potential due to the charge difference is computed on the real space grid made up of N points with space discretization step of $h = 0.25$ Bohr. The points at which the electrostatic potential is computed lie within 3 Å outside of the van der Waals radii of

the framework atoms. The calculation of the electrostatic potential is a time-consuming task as it involves an Ewald sum for each grid point. To overcome this computational bottleneck, our code employs shared memory multi-processing. As the electrostatic potential in systems with translational invariance is defined up to an additive constant, the potential is further modified to remove this gauge dependence,

$$V(\vec{r}_i)[\Delta q] \longrightarrow V(\vec{r}_i)[\Delta q] - \sum V(\vec{r}_j)[\Delta q]/N \quad 4.4$$

The average (per grid point) value of the gauge-modified electrostatic potential due to the *ab initio* and QEq charge difference constitutes the value of the "fitness" for the MOF and QEq parameterization under consideration. The small value of the fitness function corresponds to small potential difference or, in other words, the fitness function is small when *ab initio* and QEq charges yield the same electrostatic potential. The total fitness (per training set of MOFs) is the sum of fitness function values computed for individual MOFs. The QEq parameterization is obtained by minimizing fitness for the whole training set. We employ a number of schemes to optimize the total fitness including full scans for subsets of MOFs, global optimization with custom-built genetic algorithm, and, as a final step, the steepest descent method. In the case of the gradient-based optimization, the derivatives of the fitness function with respect to the hardness and electronegativity parameters is computed using 2-point central difference scheme with discretization step of 0.1 eV.

4.3.3 GCMC simulations

The details of the GCMC simulations used to calculate the uptake and HOAs of CO₂ performed in this work are as follows. We perform 10⁷ GCMC production steps following an initial 5×10⁶ equilibration GCMC steps. The sampling of configurational energies in the GCMC simulations was carried out with an in-house developed program using a classical force field.

The non-covalent interactions were described by the Lennard-Jones pair potential with parameters taken from the Universal Force Field (UFF)¹⁷ in conjunction with the Lorentz-Berthelot mixing rules for the cross-terms. The non-covalent parameters for CO₂ gas guest molecule were taken from García-Sánchez *et al.*¹⁸ A 12.5 Å cutoff distance with an isotropic tail correction was used in evaluation of the Lennard-Jones potential. The guest CO₂ gas molecule had a fixed CO bond length of 1.16 Å and partial charges of +0.645 and -0.3325 on carbon and oxygen, respectively. The GCMC simulations were performed at temperature T=298K and 0.15 bar CO₂ pressure.

4.3.4 Selection of MOFs for training and validation

At the outset of this study, it was unclear how much geometric and chemical diversity was required in the training set of MOF structures such that the χ and $\frac{1}{2}J$ parameters could broadly reproduce DFT-derived electrostatic potentials. For example, how many, and what kind of nitrogen-containing MOFs must be included in the training set to produce a set of parameters which would encompass nitrogens in heteroaromatic compounds, amines, amides, and nitro compounds, whose surrounding chemical environments would be quite different across the hypothetical MOFs constructed in chapter 3. An initial training set of 100 MOFs, including functional groups where nitrogen was only represented by $-\text{C}\equiv\text{N}$ and $-\text{NHMe}$ species, suggested that a more all-encompassing set of MOFs was needed. More specifically, the QEq parameters produced from this training set using the fitting method described in section 4.3.2 yielded poor agreement on a set of 300 MOFs containing more functional group diversity (including $-\text{NH}_2$ and $-\text{NO}_2$ groups). Recognizing that our training set must therefore better represent properties of the entire database, a cherry-picking algorithm was written to extract sets of MOFs from the hypothetical database based on desired criteria. The criteria we deemed appropriate were to ensure *i*) SBU and functional group diversity, *ii*) charge sensitivity, and *iii*) an even range of

computed CO₂ uptake at the thermodynamic conditions of post-combustion flue (298K and 0.15 bar CO₂ partial pressure). A description of how these points were evaluated is described below.

Each of the structures generated with the program detailed in chapter 2 were given an identifier which provides a concise description of the SBUs used to assemble the MOF and their 3-D connectivity. Specifically, each MOF was labelled as ‘str_m[**i**]_o[**j**]_o[**k**]_f[**m**][**t**’, where **i** is the metal SBU index, **j** and **k** are organic SBU indices, **m** is the index of the functional group used to decorate the structure, and **t** is the MOFs underlying topology. When selecting a set of structures from the database according to criteria *i*) above, the indices **i**, **j**, **k** and **m** were scanned to ensure that no one index was over-represented in the set. For example, there are a total of 19 functional groups in the database, thus each functional group was represented in no more than 5% of the MOF structures selected.

To ensure the fitting procedure converged on a solution in a timely fashion, we constrained the optimization of QEq parameters on a set of only a few hundred MOFs. Thus a ‘pool’ of roughly 300 MOFs was generated for each metal SBU from the algorithm using the first selection criteria. This study focused on SBUs with 3 different metals; those containing Cu, Zn and V. It was found that charge assignment on these metals were quite sensitive to the χ and $\frac{1}{2}J$ parameters (while others, such as Zr, were not), thus a total of 900 MOFs were generated as a training pool, only a subset of which would be selected for parameterization training. The pool of 900 structures were used to sample the criteria *ii*) and *iii*) above, which required time-intensive DFT calculations and GCMC simulations to evaluate. A MOF was deemed ‘charge sensitive’ if it demonstrated a significantly different CO₂ uptake at flue conditions when the MOF atomic charges were set to zero as compared with DFT-derived charges. The validation set of MOFs were generated using an identical approach, while also ensuring that none of the MOFs used for

the training set were included. From the constraints imposed by criteria *i)* and *ii)*, we found that, while the database of MOFs in general sampled an even distribution of CO₂ adsorption, the final training and validation sets suffered from a weighted distribution towards lower CO₂ adsorption.

4.4 Results

The MOF structures used in the training and validation sets were generated in a manner similar to the geometric approach developed by Wilmer *et al.*,¹⁹ except that steric collisions in the structures were relaxed by geometry optimization using the UFF force field. Our training set, containing 543 non-interpenetrating MOFs, was carefully selected to be diverse in structural building units (SBUs) and functional groups as well as to span a range of CO₂ uptake values. A 693 member validation set was also constructed in a similar manner as the training set but with different combinations of SBUs and functional groups. Figure 4.1a shows a list of the organic SBUs used to construct the MOFs in the training and validation sets. The dotted bonds represent either a carboxylate- or a N-capped moiety that binds to the inorganic SBU, as depicted in the lower right-hand corner of Figure 4.1a. Figure 4.1b shows the distribution of the organic SBUs (carboxylate- or N-capped combined) in the two sets. Four of the most popular Zn, Cu, and V inorganic SBUs – Zn₄O, Zn₂ (paddlewheel), Cu₂ (paddlewheel) and V₂O₂ – are nearly equally represented in both the training and validation sets. The neutral N-capped organic SBUs were used only to fully saturate the metal in the Zn and Cu paddle wheel SBUs. The following functional groups were included in the two sets: –NH₂, –NHMe, –Me, –Et, –Pr, –Ph, –Cl, –F, –Br, –I, –OH, –OEt, –OPr, –COOH, –CHO, –CN, and –NO₂. MOFs in the two sets were selected using the selection criteria mentioned in section 4.3.4. Because of the strong importance placed on SBU diversity, the distribution of CO₂ uptake values, as shown in Figure 4.1c, is skewed to lower uptakes. Although the application of standard force fields to the CO₂ binding in

MOFs with open metal sites without special modification is questionable (due to possible covalent interactions), we have included 38 open metal site paddlewheel MOFs in the training set.

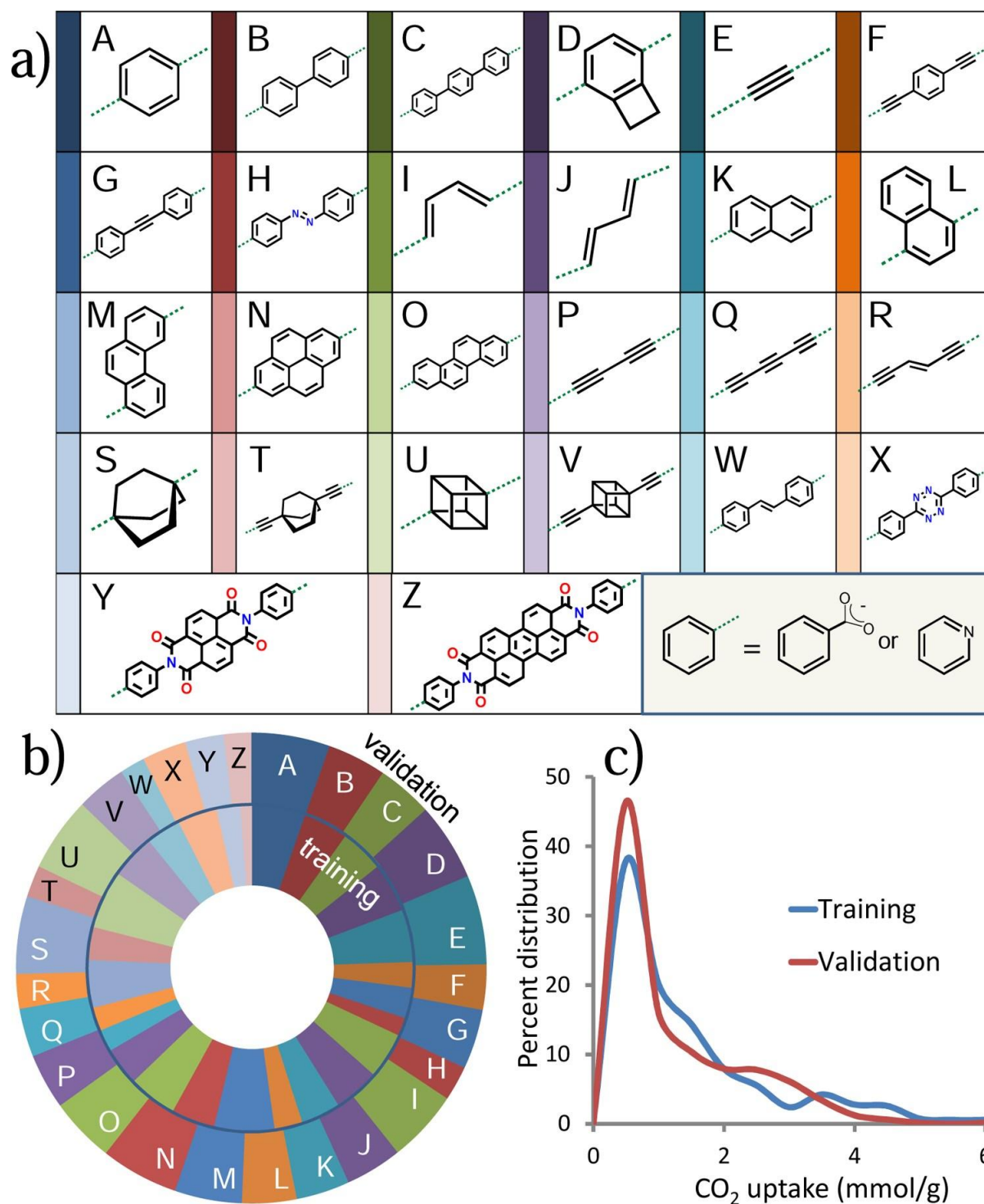


Figure 4.1: (a) List of base organic SBUs used to construct training and validation sets of hypothetical MOFs. The dotted bonds in the structures represent either a carboxylate- or N-capped moiety that binds to the inorganic SBU, as shown in the lower right-hand corner of panel a. (b) Circle charts showing the distribution of organic SBUs (carboxylate- and N-capped combined) in the training and validation sets. (c) Plot showing the distribution of MOFs as a function of the CO_2 uptake (at 0.15 bar, 298 K) obtained with DFT-derived charges for the training and validation sets.

The QEq method defines electronegativity, χ , and hardness, $\frac{1}{2}J$, parameters for each atom type. For the elements found in our MOF sets, the standard parameters taken from the UFF implementation in the OpenBabel package¹¹ are given in Table 4.1. To fit the parameters, periodic DFT calculations using the VASP package^{14–16} and the PBE functional^{12,13} were performed on all MOFs of the training set. The QEq parameters were then optimized to reproduce the gauge modified³ DFT-derived electrostatic potential on grid points lying outside of the van der Waals radii of atoms in each MOF. A custom genetic algorithm was used to simultaneously fit all parameters so as to minimize the difference between the DFT electrostatic potential and the QEq electrostatic potential in a least-squares manner. Following the genetic algorithm optimization, a numeric gradient-based optimizer was used to fine-tune the parameters. The QEq parameters for all elements were optimized, with the exception of hydrogen, which was fixed to the original values. The hydrogen parameters were fixed to enable uniqueness of the parameterization because multiple sets of parameters can give rise to the same QEq charges.²⁰ The resulting optimized parameters are given in Table 4.1. Although the original QEq method was not designed to reproduce the electrostatic potential, it is notable that the ESP-optimized parameters do not deviate more than 30% from the standard parameters for C, N, O, Zn, Cu, and V, while hardness parameters for the halogens deviate by 47% or more.

Table 4.1: Standard QEq and MOF Electrostatic Potential Optimized QEq Electronegativity and Hardness parameters.

element	standard parameters ^a		MEPO-QEq (this work)	
	χ (eV)	$\frac{1}{2}J$ (eV)	χ (eV)	$\frac{1}{2}J$ (eV)
H	4.528	6.945	4.528	6.945
V	3.650	3.410	4.093	4.217
Cu	4.200	4.220	5.429	3.468
Zn	5.106	4.285	3.701	4.463
C	5.343	5.063	5.431	5.857
N	6.899	5.880	6.688	6.622
O	8.741	6.682	8.714	8.568
F	10.874	7.474	6.416	11.131
Cl	8.564	4.946	5.821	7.273
Br	7.790	4.425	5.692	8.760
I	6.822	3.762	5.431	5.720

^aParameters taken from OpenBabel¹¹ software.

During the course of this study, we discovered that the standard QEq parameters for Cu differ in commonly used software packages. The values given in Table 4.1 are taken from OpenBabel¹¹ and were used by Sholl and coworkers to evaluate the CO₂ uptake in 500 experimental MOFs.⁸ The Cu parameters in the popular Materials Studio and GULP²¹ packages ($\chi = 3.729$ eV and $\frac{1}{2}J = 2.501$ eV) differ significantly from the OpenBabel parameters. Although the origin of this difference is unclear, we note that the Materials Studio parameters for Cu seem out of line when considering the trends in the periodic table. Additionally, we found the OpenBabel Cu parameters generally gave CO₂ uptakes that were in better agreement with those calculated using DFT-derived charges, as compared with those obtained with the Materials Studio Cu parameters.

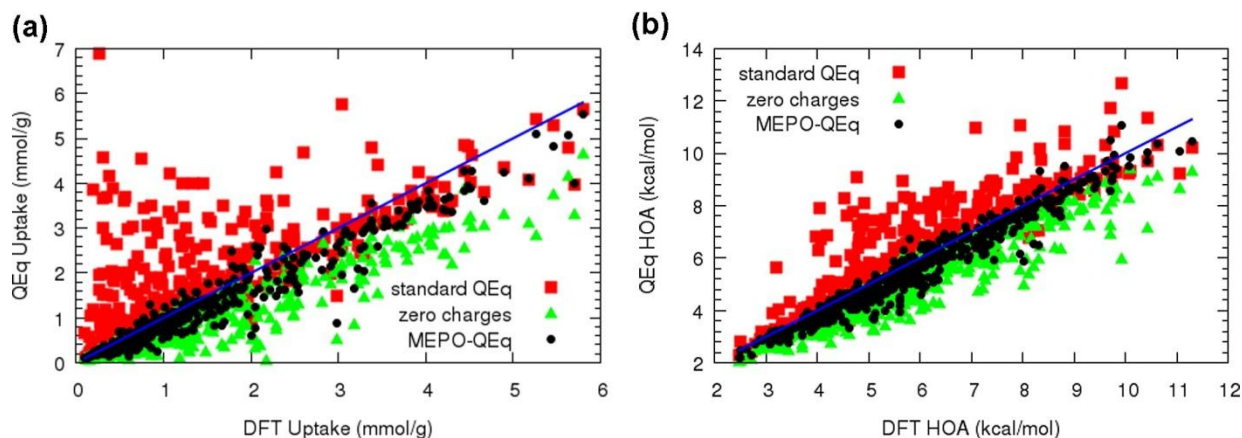


Figure 4.2: (a) CO₂ gas uptake and (b) heats of adsorption at 0.15 bar, 298 K for the training set of MOFs obtained from the GCMC simulations using standard QEq charges, using no charges, and MEPO-QEq charges compared with those obtained using DFT-derived charges. The line of perfect correlation is shown in dark blue.

CO₂ gas uptake and HOA determined from grand canonical Monte Carlo (GCMC) simulations²² were used to evaluate the performance of charge schemes. Figure 4.2 compares the CO₂ uptake and HOA results determined using DFT-derived charges to those obtained using the standard QEq parameters, the MEPO-QEq parameters, and where all charges have been set to zero. The corresponding correlation coefficients using the different charge schemes with respect to the DFT-derived charges are provided in Table 4.2. The results reveal that the MEPO-QEq charges are able to reproduce the CO₂ uptake and HOAs determined with DFT-derived charges very well over a diverse set of MOFs and over a broad range of uptake capacities. The correlation coefficients are exceptionally high (0.97), with the Pearson coefficient measuring the linear correlation in value and the Spearman coefficient measuring the rank-order correlation.

Table 4.2: Correlation Coefficients for the CO₂ Uptake and HOA Computed with Standard QEq Charges, Zero Charges, and MEPO-QEq Charges When Compared with Those Determined with DFT-Derived Charges^a

charge scheme	correlation coefficients				error in uptake (mmol/g)	
	Pearson		Spearman		max	RMS
	uptake	HOA	uptake	HOA		
standard QEq	0.78	0.89	0.83	0.88	+6.61	0.87
zero charges	0.95	0.94	0.93	0.95	-2.44	0.63
MEPO-QEq	0.97	0.97	0.97	0.97	-2.10	0.34

^aRMS and maximum errors are reported for the CO₂ uptake. The calculations are carried out for the training set of 543 hypothetical MOFs.

Interestingly, the standard QEq charges do not perform as well as using zero charges, as shown by the correlation coefficients in Table 4.2. Using zero charges results in an underestimation of the CO₂ uptake and HOAs, while the standard QEq charges often result in the large overestimation of these properties. The root-mean-square (RMS) and maximum error in the CO₂ uptake are two to three times larger with the standard QEq charges as compared to with the MEPO-QEq charges. The standard QEq charges give an overestimation in the CO₂ uptake of greater than 1 mmol/g for 73 of the MOFs in the 543 member test set. A closer inspection reveals that 75% of these 73 MOFs contain halogen functional groups, with F- and Cl-functionalized MOFs possessing the most severe errors and greatest overestimation of uptake when compared with the DFT-derived charges.

A more stringent test of the MEPO-QEq charge scheme is to evaluate the CO₂ adsorption properties in MOFs that are not part of the training set. We have derived REPEAT charges from periodic DFT calculations of 693 MOFs of the “validation” set and determined the CO₂ uptake and HOA with these charges. A comparison of the CO₂ adsorption properties determined from the MEPO-QEq and the DFT-derived charges is given in Figure 4.3 and Table 4.3. As seen with

the training set, there is excellent agreement in the uptakes and HOAs, with correlation coefficients of 0.98 in all cases. The maximum and RMS errors in the uptake with the validation set are comparable but slightly smaller than those seen with the training set. The slope ($m = 0.82$) of the best fit line (not shown) does indicate that the MEPO-QEq method tends to underestimate the CO₂ uptake.

Table 4.3: Correlation Coefficients for the CO₂ Uptake and HOA Computed with Standard QEq Charges, Zero Charges, and MEPO-QEq Charges When Compared with Those Determined with DFT-Derived Charges^b

charge scheme	correlation coefficients				error in uptake (mmol/g)	
	Pearson		Spearman		max	RMS
	uptake	HOA	uptake	HOA		
standard QEq	0.79	0.92	0.91	0.93	+4.76	0.80
zero charges	0.92	0.96	0.96	0.97	-4.92	0.61
MEPO-QEq	0.98	0.98	0.98	0.98	-1.77	0.31
EQEq	0.87	0.89	0.91	0.88	+3.23	0.54

^aRMS and maximum errors are reported for the CO₂ uptake. The calculations are carried out for the validation set of 693 hypothetical MOFs.

As observed with the training set, the CO₂ uptake and HOAs are sometimes greatly overestimated with the standard QEq charges. Again, the MOFs with halogen functional groups give the largest errors. We also compare the EQEq method where the metals are treated differently than in the standard QEq method. Although there is an improvement over the standard QEq method, the method also shows a large overestimation in the uptakes and HOAs for Cl- and F-functionalized MOFs. This may not be surprising because the halogen parameters in the standard QEq and EQEq methods are identical.

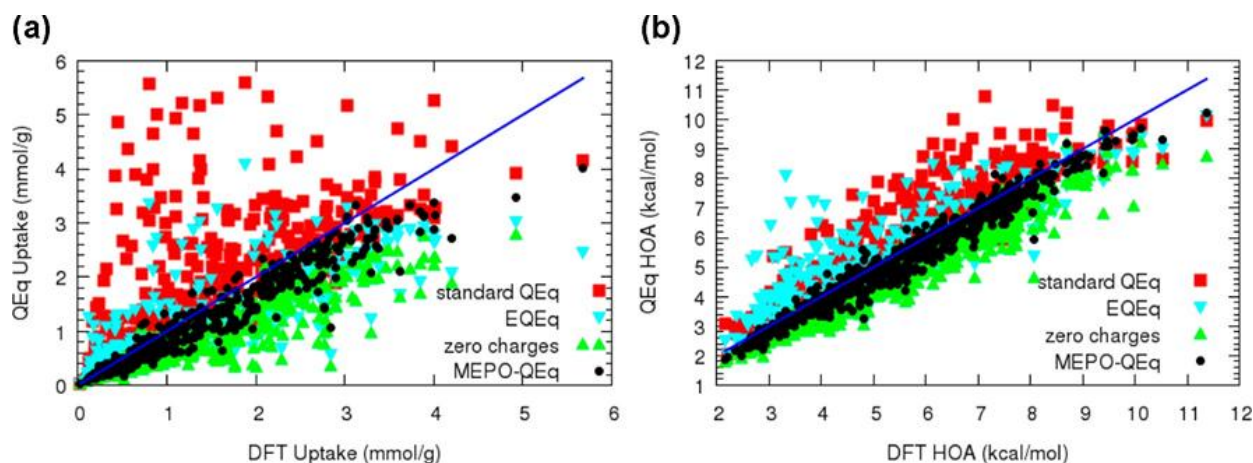


Figure 4.3: (a) CO₂ gas uptake and (b) HOA at 0.15 bar, 298 K for the validation set of MOFs obtained from the GCMC simulations using standard QEq charges, zero charges, MEPO-QEq charges, and EQEq charges as compared with those obtained using DFT-derived charges. The line of perfect correlation is shown in dark blue.

Although the two evaluation sets contain the most popular Zn, Cu, and V inorganic SBUs, the question of parameter transferability is of interest. To examine this question, we have evaluated the CO₂ uptake of 11 Zn-zeolitic imidzolate frameworks (ZIFs) and 16 V MOFs based on the MIL100 inorganic SBU combined with different organic SBUs. Neither the inorganic or organic SBUs in ZIFs are found in our training or validation sets. Compared with results obtained with DFT-derived charges, MEPO-QEq still performs well with this ZIF validation set with a Pearson correlation of 0.96 and RMS error of 0.16 mmol/g. Although these results are good, it is notable that for this set there is not a large spread in the uptake and that using zero charges gives an equally high correlation coefficient of 0.94, albeit with a significantly larger RMS error of 0.36 mmol/g. The MIL100 inorganic SBU has open metal sites, and while open-metal-site Cu and Zn SBUs were part of the training set, no open-metal-site V centers were used to train the model. For this set of MOFs, the MEPO-QEq charges do not perform as well, with a Pearson correlation coefficient of 0.92 and a large RMS error of 0.86 mmol/g. As one might expect, the results suggest that caution should be used in applying the MEPO-QEq method to SBUs, particularly inorganic SBUs, that are not part of the validation set.

This begs the question as to whether it is possible to define universal QEq parameters for elements that span a range of oxidation states or hybridizations. Although answering that question is beyond the scope of this work, one can overcome this by introducing more than one atom type for such elements.

4.5 Conclusions

The MEPO-QEq provides a robust method for rapidly generating partial atomic charges to reproduce the electrostatics in a diverse set of MOFs. The parameterization was validated by examining the CO₂ uptake and HOAs in over 1200 MOFs constructed from 52 different organic SBUs, 4 of the most popular Zn, Cu, and V inorganic SBUs, and 17 functional groups. The results were found to be in excellent agreement with those determined with DFT-derived charges, having linear and rank-order correlation coefficients of 0.97 or better. It is notable that the MEPO-QEq scheme was not directly fit to reproduce the CO₂ uptake and HOAs, rather it was fit to reproduce the electrostatic potential of MOFs in the training set. Therefore, the exceptional agreement in the uptake and HOAs suggests that the MEPO-QEq charges should perform similarly well for other properties that depend on the electrostatic potential of the MOF and for evaluating the uptake of polar guest molecules other than CO₂. Caution should always be exercised when applying the QEq method to study MOFs with inorganic SBUs and functional groups that are not part of its training set, particularly those containing new elements. Finally, we have investigated the accuracy of the standard QEq parameters and “extended” QEq (EQEq) methods and found that they generally perform well, considering that they were not developed to reproduce the electrostatic potential. However, we also find that the application of standard QEq and EQEq methods can result in a severe overestimation of the CO₂ uptake in F and Cl

functionalized MOFs. The latter emphasizes the importance of careful validation of the QEq parameters.

4.6 References

- (1) Hohenberg, P.; Kohn, W. *Phys. Rev.* **1964**, *136*, B864.
- (2) Kohn, W.; Sham, L. J. *Phys. Rev.* **1965**, *140*.
- (3) Campañá, C.; Mussard, B.; Woo, T. K. *J. Chem. Theory Comput.* **2009**, *5*, 2866.
- (4) Manz, T. A.; Sholl, D. S. *J. Chem. Theory Comput.* **2010**, *6*, 2455.
- (5) Chen, D.-L.; Stern, A. C.; Space, B.; Johnson, J. K. *J. Phys. Chem. A* **2010**, *114*, 10225.
- (6) Xu, Q.; Zhong, C. *J. Phys. Chem. C* **2010**, *114*, 5035.
- (7) Rappe, A. K.; Goddard, W. A. *J. Phys. Chem.* **1991**, *95*, 3358.
- (8) Haldoupis, E.; Nair, S.; Sholl, D. S. *J. Am. Chem. Soc.* **2012**, *134*, 4313.
- (9) Wilmer, C. E.; Kim, K. C.; Snurr, R. Q. *J. Phys. Chem. Lett.* **2012**, *3*, 2506.
- (10) Wilmer, C. E.; Farha, O. K.; Bae, Y.-S.; Hupp, J. T.; Snurr, R. Q. *Energy Environ. Sci.* **2012**, *5*, 9849.
- (11) O'Boyle, N. M.; Banck, M.; James, C. a; Morley, C.; Vandermeersch, T.; Hutchison, G. R. *J. Cheminform.* **2011**, *3*, 33.
- (12) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1997**, *78*, 1396.
- (13) Perdew, J.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865.
- (14) Kresse, G.; Furthmüller, J. *Phys. Rev. B* **1996**, *54*, 11169.
- (15) Kresse, G.; Hafner, J. *Phys. Rev. B* **1993**, *47*, 558.
- (16) Kresse, G.; Hafner, J. Ab initio molecular-dynamics simulation of the liquid-metal–amorphous-semiconductor transition in germanium. *Physical Review B*, 1994, *49*, 14251–14269.
- (17) Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. *J. Am. Chem. Soc.* **1992**, *114*, 10024.

- (18) 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*, 8814.
- (19) Wilmer, C. E.; Leaf, M.; Lee, C. Y.; Farha, O. K.; Hauser, B. G.; Hupp, J. T.; Snurr, R. Q. *Nat. Chem.* **2012**, *4*, 83.
- (20) Verstraelen, T.; Bultinck, P.; Van Speybroeck, V.; Ayers, P. W.; Van Neck, D.; Waroquier, M. *J. Chem. Theory Comput.* **2011**, *7*, 1750.
- (21) Gale, J. J. *Chem. Soc., Faraday Trans.* **1997**, *93*, 629.
- (22) Düren, T.; Bae, Y.-S.; Snurr, R. Q. *Chem. Soc. Rev.* **2009**, *38*, 1237.
- (23) Wells, B. A.; De Bruin-Dickason, C.; Chaffee, A. L. *J. Phys. Chem. C* **2015**, *119*, 456.

4.7 Appendix – Derivation of the Charge Equilibration Equations

In classical molecular dynamics simulations, it is typical to approximate electrostatic energy using point charges on each atom, whereupon affixing a single representative partial charge on each atom may not accurately describe the energy of a molecule as it evolves through time. The development of the charge equilibration (QEq) method was originally designed to adjust charges on the fly in classical molecular dynamics simulations. This is based on the idea that charge has a tendency to ‘flow’ much like clouds of electrons redistribute to a ground state when a molecule’s conformation changes.

An atom, A, possesses an energy E_A which can be expressed as a function of its partial charge, Q_A ; $E_A(Q_A)$. This energy expression can be expanded by an infinite Taylor series around a central charge value, Q_0 :

$$E_A(Q_A) = E_{A0} + \Delta Q \left(\frac{\partial E_A}{\partial Q_A} \right) + \frac{1}{2} \Delta Q^2 \left(\frac{\partial^2 E_A}{\partial Q_A^2} \right) + \cdots + \frac{1}{n!} \Delta Q^n \left(\frac{\partial^n E_A}{\partial Q_A^n} \right), \quad n \rightarrow \infty \quad 4.5$$

$$\Delta Q = (Q_A - Q_0)$$

In most charge equilibration schemes, two approximations are typically made; the infinite sum is cut off at the second order term, and (with a couple of exceptions^{9,23}) the charge is expanded around the neutral atom, thus $Q_0 = 0$. This simplifies the above equation to;

$$E_A(Q_A) = E_{A0} + Q_A \left(\frac{\partial E_A}{\partial Q_A} \right) + \frac{1}{2} Q_A^2 \left(\frac{\partial^2 E_A}{\partial Q_A^2} \right) \quad 4.6$$

The QEq expression for the energy of a single atom, A as a function of charge, where E_{A0} is the energy of the atom when $Q_A = 0$. The partial derivative terms in Eq. 4.6 evaluate to the electronegativity (χ_A) and hardness (J_{AA}) (or idempotential) parameters optimized and discussed in the main text. The Mulliken definition of electronegativity (χ_A) is the average difference between the first ionization energy and the electron affinity of the atom, while an atoms chemical hardness (J_{AA}) is defined as the sum of the atoms first ionization energy and electron affinity.

The ionization energy can be found by determining the energy expression in Eq. 4.6 when an electron is removed from the atom, resulting in a +1 charge;

$$I = E_A(+1) - E_{A0} = \left(\frac{\partial E_A}{\partial Q_A} \right) + \frac{1}{2} \left(\frac{\partial^2 E_A}{\partial Q_A^2} \right) \quad 4.7$$

and the electron affinity can be found similarly by adding an electron, resulting in a -1 charge;

$$A = E_A(-1) - E_{A0} = - \left(\frac{\partial E_A}{\partial Q_A} \right) + \frac{1}{2} \left(\frac{\partial^2 E_A}{\partial Q_A^2} \right) \quad 4.8$$

The partial derivative expressions can then be isolated by adding and subtracting Eqs 4.7 and 4.8 yielding

$$\left(\frac{\partial E_A}{\partial Q_A} \right) = \frac{I - A}{2} = \chi_A \quad 4.9$$

which is the mulliken definition for electronegativity, and

$$\left(\frac{\partial^2 E_A}{\partial Q_A^2} \right) = I + A = J_{AA} \quad 4.10$$

which is the definition of chemical hardness. This can be considered the coulombic repulsion of two electrons occupying the same orbital. Thus an expression of the energy for a single atom as a function of charge can be defined by substituting Eqs 4.9 and 4.10 into 4.7;

$$E_A(Q_A) = E_{A0} + Q_A\chi_A + \frac{1}{2}Q_A^2J_{AA} \quad 4.11$$

The energy expression for a system of N atoms (in a molecule or periodic solid) can then be expressed as a sum of the individual atom energies from Eq. 4.11.

$$E(Q_1, Q_2, \dots, Q_N) = \sum_{i=1}^N \left(E_{i0} + Q_i\chi_i + \frac{1}{2}Q_i^2J_{ii} \right) + \frac{1}{2} \sum_{L=0}^N \sum_{i=1}^N \sum_{j \neq i*}^N Q_i Q_j J_{Lij} \quad 4.12$$

The first summation goes over all individual atomic energies, and the last triple summation spans the interatomic coulombic potential J_{Lij} over all the periodic images, L (in the case of periodic systems), where the expression $j \neq i*$ holds only when the atoms are in the same image, $j = i$ otherwise. We note that the last term gives a unique expression for the molecular energy as a function of the inter-atomic distances, which was the desire at the outset of this derivation.

We add a constraint to ensure that the total charge sums to Q_T (which is zero in periodic simulations).

$$\sum_{i=1}^N Q_i = Q_T \quad 4.13$$

This is added as a constraint to the total energy expression in Eqn 4.12 using a lagrangian multiplier, λ

$$E(Q_1, Q_2, \dots, Q_N) = \sum_{i=1}^N \left(E_{i0} + Q_i \chi_i + \frac{1}{2} Q_i^2 J_{ii} \right) + \frac{1}{2} \sum_{L=0}^N \sum_{i=1}^N \sum_{j \neq i^*}^N Q_i Q_j J_{Lij} - \lambda \left(\sum_{k=1}^N Q_k - Q_T \right) \quad 4.14$$

We are now left to define how these charges will fluctuate to minimize the total energy described in Eq. 4.14. This is done by taking the partial derivative of the total energy with respect to all of the charges. For atom A, we define its chemical potential $-\mu_A$ as the derivative of the total energy expression with respect to its charge, Q_A . We say that at an energetic minimum these derivatives are at a stationary point, therefore are all equal to zero

$$-\mu_A = \frac{\partial E(Q_1, \dots, Q_N)}{\partial Q_A} = \chi_A + Q_A J_{AA} + \frac{1}{2} \sum_{L=0}^N \sum_{i \neq A^*}^N Q_i J_{LiA} - \lambda = 0 \quad 4.15$$

Where, again, the final summation is over all indices where $i \neq A$ only if the atoms are in the same periodic image, $i = A$ otherwise. A set of N equations can then be defined by rearranging Eqn 4.15,

$$\sum_{L=0}^N \sum_{k=1}^N J_{Lik} Q_k - \lambda = -\chi_i, \quad i = 1, \dots, N \quad 4.16$$

Which, including Eqn 4.13, can be set up as a series of N+1 equations with N+1 unknowns;

$$\left. \begin{array}{l} \sum_{L=0}^N \sum_{k=1}^N J_{Lik} Q_k - \lambda = -\chi_i \\ \sum_{k=1}^N Q_k - Q_T = 0 \end{array} \right\} \begin{array}{l} i = 1, \dots, N \\ N + 1 \end{array}$$

The above equations can be expressed in matrix form as shown in Eqn. 4.2 in the main text;

$$\begin{pmatrix} \mathbf{J} & \mathbf{1} \\ \mathbf{1}^\dagger & 0 \end{pmatrix} \begin{pmatrix} \mathbf{Q} \\ \lambda \end{pmatrix} = \begin{pmatrix} -\mathbf{\chi} \\ 0 \end{pmatrix}$$

We note that the $\mathbf{J}$ terms in decay very slowly ($\propto 1/R$), which results in divergent behaviour for periodic systems with unit cells on the order of tens of Ångströms. The solution is to use an Ewald summation, which splits the coulombic interaction into real-space short range interactions and long range interactions which are summed in reciprocal space. An Ewald summation contains the coulombic interaction potentials between all the periodic images in the simulation, which are in turn influenced by the set of charges $\mathbf{Q}$. Thus the above matrix equation must be solved self-consistently.

5 Discovery of Adsorbophores

5.1 Introduction

The motivation behind developing a substructure matching program in this thesis was to identify common CO₂ binding ‘pockets’ found in strongly adsorbing MOFs. This methodology is integral in drug discovery programs, where substructure matching programs are usually the initial step in identifying lead compounds for further study.¹ While there is a tremendous amount of interest in developing these matching algorithms in the lucrative pharmaceutical field (many of which have licensing fees), no such methods have been developed for finding common structural elements in MOFs. Thus for our purposes a program was developed from scratch.

The term *pharmacophore* was officially defined in 1998 by the International Union of Pure and Applied Chemistry – “the ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target structure and to target (or to block) its biological response”. Structural features in MOFs with optimal interactions with adsorbed CO₂ can be considered conceptually similar to this definition, we therefore take this opportunity to coin the term *adsorbophore* to represent these features. The challenge is to then mine large datasets to discover these adsorbophores. In virtual screening drug programs a pharmacophore is typically known *a priori*, through some known binding interaction with a target protein. Potential lead drug molecules are then identified by performing a substructure search for the pharmacophore in databases of millions of organic compounds. In the case of adsorbophores, our initial hypothesis is more general; there is likely some common pore features in high performing MOFs which form strong interactions with CO₂. We do not know what these common features are, so the program was developed with this motivation in

mind. This chapter will deliberate the methods and algorithms developed to identify common CO₂ binding sites in MOFs.

5.2 *Methods*

The following sections will discuss the data mining methods implemented, all in house, to progress from knowing only the MOF structures to identifying common structural features which strongly adsorb CO₂. A graphical representation of the workflow progressing from empty MOF structures to adsorbophores is presented in Figure 5.1. Briefly, for each MOF, a GCMC simulation is performed to obtain 3-D probability distributions in the pores as seen in Figure 5.1b. The Automated Binding Site Locator (ABSL) program is used to identify CO₂ binding sites from the probability distribution, an example of binding sites found from the probability distribution is shown in Figure 5.1 c. This algorithm is briefly discussed in Section 5.2.2. The local MOF atoms are cut out from each binding site, whereupon common atoms are identified with other MOF binding sites (Figure 5.1d and 5.1e) using a substructure matching program discussed in Section 5.3.

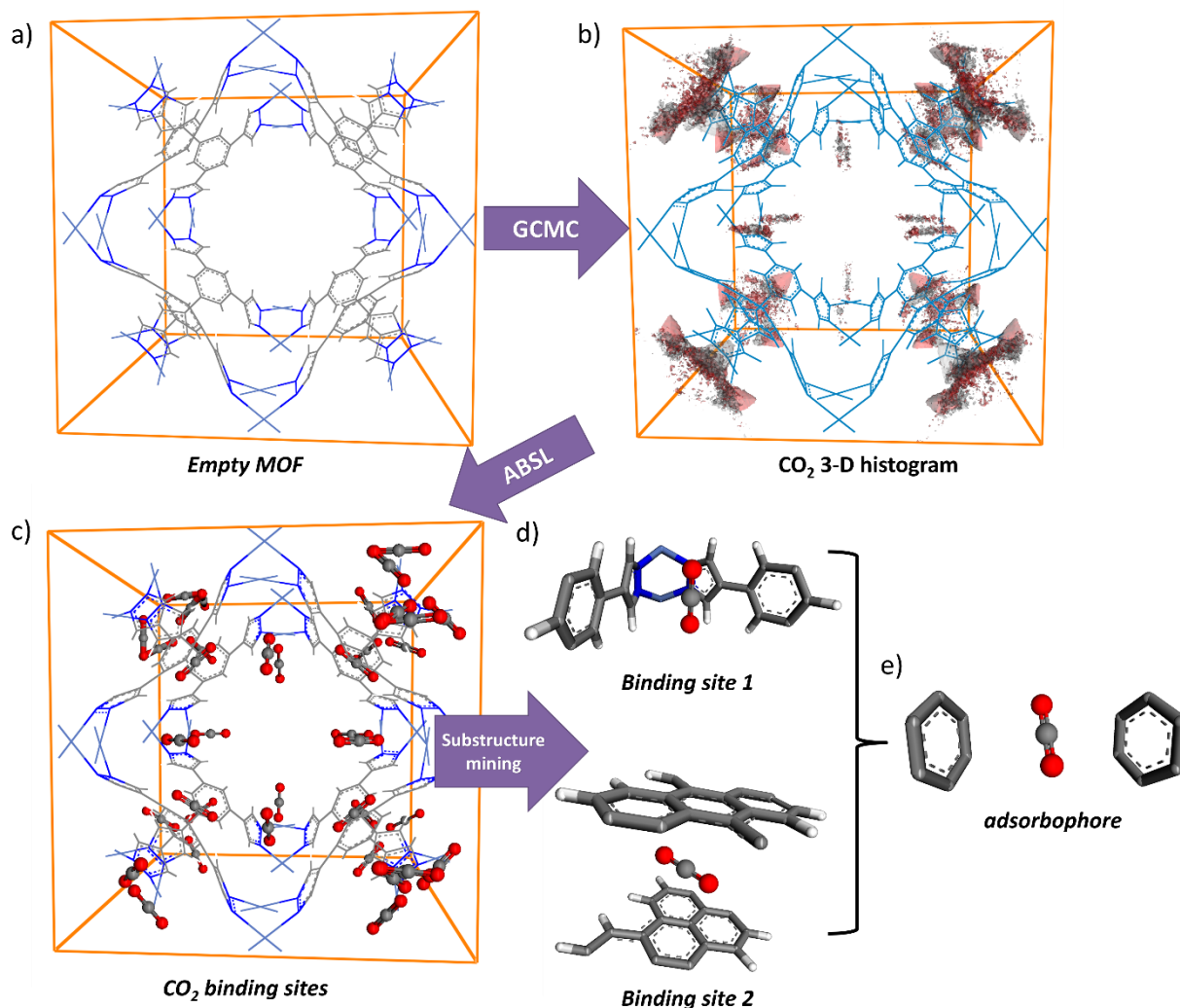


Figure 5.1: A visual guide to the workflow discussed in this chapter. Starting with a) empty MOF frameworks, GCMC simulations are performed to obtain b) 3-D probability distributions of CO₂ carbon and oxygens in the MOF pores (represented as transparent grey and red clouds, respectively). The c) maxima found in these probability distributions give representative CO₂ binding sites in each MOF using the ABSL program. d) Local MOF atoms are extracted from each binding site and common atom distributions are found using a 3-D substructure matching algorithm. The most frequent substructures scanned over many MOFs and binding sites are the e) adsorbophores.

5.2.1 3-D histograms from GCMC simulations.

3-D histograms of CO₂ distributions were taken from grand canonical Monte Carlo (GCMC) simulations on MOFs using a temperature and CO₂ gas pressure consistent with the thermodynamic state point of flue gas (298K and 0.15 bar CO₂). The histograms were computed by storing, or *binning*, the positions of the adsorbed CO₂ particles at each post-equilibration GCMC step in a 3-dimensional real-space grid. The grid spanned the unit cell of the MOF,

possessing the same three dimensional periodic vectors, $(\vec{a}_1, \vec{a}_2, \vec{a}_3)$. The periodic vectors were respectively discretized into N_1 , N_2 , and N_3 grid points such that the grid spacings (voxels) were maintained a size of roughly $|\vec{n}_i| = 0.1 \text{ \AA}$ per side ($\vec{n}_i = \vec{a}_i/N_i, i = 1,2,3$). The spacing was set to this value to ensure a high resolution of the resulting probability distribution. Each point in the grid was initially assigned a value of 0, and were incremented with fractional values according to the rules of so-called ‘equitable binning’. This method was implemented to smoothly distribute positional data on a discretized grid. As an example consider the oxygen atom resting between 8 grid points in Figure 5.2. To store the position of this atom on the grid, the distance of this position to all 8 grid points is determined, and each point is assigned a fraction based on the atom’s relative distance to that point. For a Cartesian position of an atom, $\vec{r}$ in the cell, the surrounding grid points can be found by determining its fractional position relative to the cell vectors. Since the Cartesian position can be expressed as a sum of its cell vector components;

$$\vec{r} = m_1 \cdot \vec{a}_1 + m_2 \cdot \vec{a}_2 + m_3 \cdot \vec{a}_3 \quad 5.1$$

We can define its fractional position in the cell direction $\vec{a}_i$ as

$$m_i = \frac{\vec{r} \times \vec{a}_j \cdot \vec{a}_k}{\vec{a}_i \times \vec{a}_j \cdot \vec{a}_k} \quad 5.2$$

Without loss of generality, indices j and k represent the other cell vectors and the operators $\times$ and $\cdot$ represent the cross and dot products of three dimensional vectors, respectively. From Eqn. 5.2 we can determine the grid points surrounding the atom in the $\vec{a}_i$ direction as

$$\begin{aligned} g_i^{upper} &= [N_i \cdot m_i] \\ g_i^{lower} &= \lfloor N_i \cdot m_i \rfloor \end{aligned} \quad 5.3$$

Where the ceiling operator is represented as upper capped brackets (eg. $[0.5] = 1$) and the floor operator is represented by lower capped brackets (eg. $\lfloor 0.5 \rfloor = 0$) which round to the nearest larger and smaller integer value, respectively. Thus Eqn 5.3 gives an upper and lower grid point

encompassing the atom along each cell direction $\vec{a}_i$. Moreover, we can identify how much of each grid point the atom should occupy based on the distance of that atom to the grid point.

$$p(g_i) = 1 - |\vec{a}_i| \cdot \left| m_i - \frac{g_i}{N_i} \right| / |\vec{n}_i| \quad 5.4$$

Eqn 5.4 defines the partitioning in one cell direction, which says that the closer an atom is to the grid point, the larger will be the grid points occupancy. To obtain the overall three dimensional partitioning in cell directions i,j, and k, we apply Eqn. 5.4 in each direction for every surrounding grid point of the atom. Equitable binning is obtained via

$$p(g_{ijk}) = p(g_i) \cdot p(g_j) \cdot p(g_k) \quad 5.5$$

By assigning weights to the 8 surrounding grid points of an atom using Eqn. 5.5, the exact position of the atom can be recovered from the values of the surrounding grid points. The smoothness of this distribution is important for the filtering method discussed below to effectively identify maxima in the probability distribution.

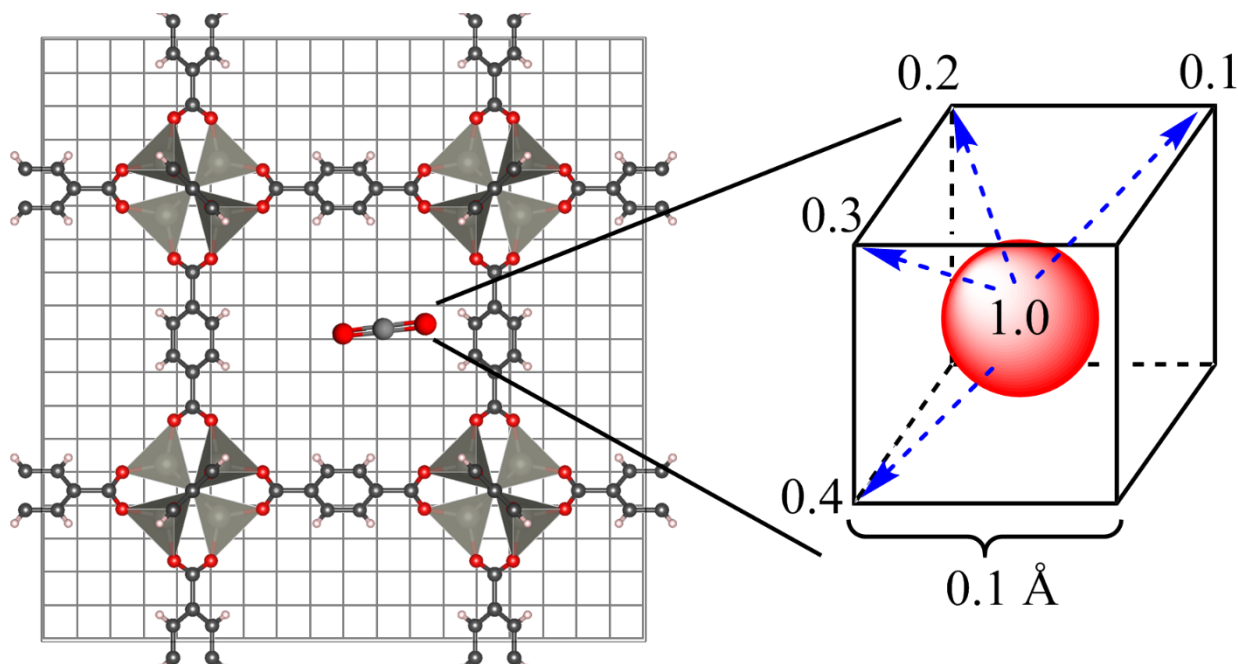


Figure 5.2: Equitable binning of an oxygen atom (red sphere) in a discretized 3-D grid. The oxygen's occupancy (1.0) is split between all of the surrounding grid points.

Following a GCMC simulation, the resulting grids containing the absolute number of times each CO₂ was counted in space were normalized by i) the number of GCMC steps counted and ii) the volume occupied by a single voxel. This normalization ensures that the probability distribution is transferrable across different simulations and cell volumes. Separate distributions were constructed for both the C and O atoms. This was so that later we could resolve CO₂ orientations when finding representative binding sites.

5.2.2 Binding sites from 3-D histograms

The 3-D histogram data contains averaged values of CO₂ distributions over the span of a GCMC simulation which, by its construction, samples the free energy of the system. It can be said that maximum values found in the histogram are the regions of highest residence of CO₂. These maxima correspond to wells in the free energy surface of the system, and are thus considered *free energy* binding sites of CO₂ in the MOF pores. The task of identifying these maxima, and assigning full molecules of CO₂ corresponding to these maxima in an automated

fashion is briefly described here. This method is largely the unpublished work of a colleague, Dr. Eugene Kadantsev, however we find it necessary to outline the method for the purposes of transparently progressing to the substructure search program, which takes as input, the binding sites located with this algorithm.

Discovering maximum values in a 3-D histogram seems relatively straight forward. One only needs to iterate over each grid point and return a list of the highest recorded values as locations of binding sites. This can be made short work by a computer, which can scan 10^7 grid points in a MOF 3-D histogram in a manner of seconds. However, in practice it is not that simple, as the landscape of the CO₂ distributions can be quite rough. The roughness stems from Monte Carlo sampling, where the movement of molecules is governed by a Markov chain of random orientations and translations. The sampling noise in the data will result in thousands of maxima in the probability distribution of a typical MOF unit cell, most of which do not correspond to CO₂ binding sites.² Thus, the primary machinery used in the ABSL program involves smoothing the histogram data, such that noise due to the GCMC sampling is filtered and fewer maxima are found, while still maintaining the integrity of the probability distribution.

The data smoothing is performed via a Fourier transformation of the probability distribution. In this method, the probability data, $p(\vec{r})$, obtained on the 3-D grid ($\vec{r} = m_1\vec{a}_1 + m_2\vec{a}_2 + m_3\vec{a}_3$) is converted to coefficients in reciprocal space, $p(\vec{G})$, via the transformation;

$$p(\vec{G}) = \frac{1}{V} \int_V p(\vec{r}) \exp\{-i\vec{G}\vec{r}\} d\vec{r} \quad 5.6$$

where $\vec{G}$ are vectors in the reciprocal space of the unit cell, and V is the unit cell volume. This effectively expresses the probability data as a series of continuous plane wave functions ($\exp(-i\vec{G}\vec{r})$) whose periodicity is the same as that of the reciprocal unit cell. The $\vec{G}$ vectors can be considered frequencies of the plane wave oscillations, where high frequencies in reciprocal

space correspond to sharp oscillations (or noise) in the real space data. We can therefore easily smooth out noisy data by removing the high frequencies in reciprocal space with a cutoff threshold, G_{cut} , such that,

$$p_s(\vec{G}) = \begin{cases} p(\vec{G}), & \text{for } |\vec{G}| \leq G_{cut} \\ 0, & \text{for } |\vec{G}| > G_{cut} \end{cases} \quad 5.7$$

The new, smoother function, $p_s(\vec{G})$, can then be converted back to the real space grid via

$$p_s(\vec{r}) = \int p_s(\vec{G}) \exp\{-i\vec{G}\vec{r}\} d\vec{G} \quad 5.8$$

The maxima were then discovered by iterating over the new smoothed data and identifying the values in $p_s(\vec{r})$ which were greater than a threshold parameter p_{cut} and larger than its surrounding 26 grid points. In the discussion that follows, the choice of G_{cut} was set to 3 bohr⁻¹ which corresponds to the spatial resolution of the C-O bond length, and p_{cut} was set to 0.025. These values were determined after successful identification of all the CO₂ binding sites from a MOF, CALF-15, which were resolved experimentally from x-ray crystallography.²

For the purposes of producing representative CO₂ binding sites, the maxima were found and combined from two histogram grids, one which stored the average positions of carbon atoms, and separately, oxygens from CO₂. Difficulties arose from identifying full CO₂ molecules from this data; in some cases strong localization would occur on one atom of the molecule while the remaining atoms would have more rotational degrees of freedom resulting in a spreading or smearing of the probability distribution. This meant that maxima within the smeared data would be difficult to identify, and one could distinguish only fragments of a CO₂. Full molecules were formed only if carbon and oxygens could be found near each other in a linear arrangement. In cases where only one oxygen can be found in proximity to a carbon, an additional oxygen was

added with a typical CO bond length and angle of 180° from the existing C-O atoms to complete the molecule.

As an example, the ABSL program was run on a 3-D histogram of CO₂ in the MOF CALF-15. Here the histogram was generated from a GCMC simulation using the thermodynamic conditions used to resolve the experimental CO₂ binding sites from x-ray crystallography. While in chapter 2, we identified these binding sites from the probability distribution by hand, here the ABSL program was used to automatically assign CO₂ binding sites, with excellent agreement, as identified in Figure 5.3.

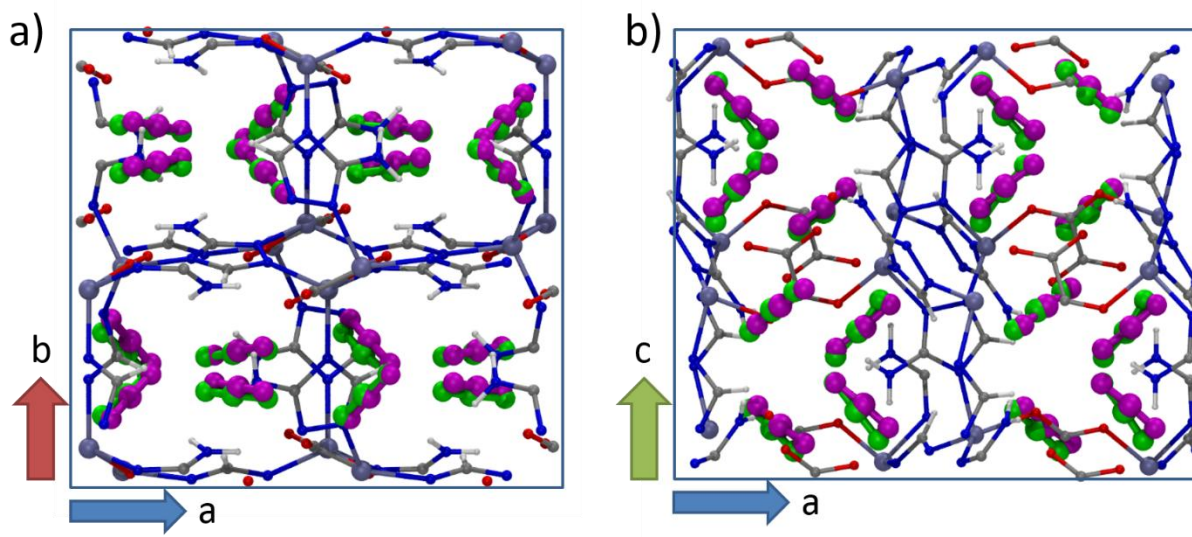


Figure 5.3: CO₂ binding sites from CALF-15 seen down the a) c-axis and b) b-axis. The ABSL determined CO₂ binding sites are shown in green and the crystallographic binding sites are shown in pink.

5.2.3 Active site extraction

With binding sites identified in the MOF pores, the next task is to extract the local MOF chemical information from each binding site, termed an *active site*. This was performed by extracting spherical regions of radius R surrounding each CO₂ binding site. The choice of R was a value that would capture a majority of the local (short-range) electrostatic and van der Waals interaction energy from the surrounding MOF, while being small enough to preserve only the

intimate surroundings of the CO₂ molecule. Herein we use a value of $R = 5 \text{ \AA}$, which was determined as reasonable from the functions used in the GCMC program to compute the CO₂-framework interaction energy. For electrostatic energy, the Ewald summation technique³ was used, which conveniently splits the electrostatic interactions into a short and long range term. The short range energy term involves screening the point charges with an equal but opposite charge distributed over a Gaussian function. The resulting real space summation over all atomic charges N is shown in Eqn 5.9 as

$$E_{Sr} = \left(\frac{1}{4\pi\epsilon_0}\right) \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \frac{q_i q_j}{\vec{r}_{ij}} \operatorname{erfc}\left(\frac{\vec{r}_{ij}}{\sqrt{2}\sigma}\right), i \neq j \quad 5.9$$

where $\vec{r}_{ij}$ is the distance between point charges q_i and q_j , ϵ_0 is the permittivity of free space, erfc is the complimentary error function arising from integration of the Gaussian smeared charge and σ controls the width of the Gaussian curve. The σ term is determined at simulation time, however is typically set to $2.78 \text{ \AA}$. The largest magnitude partial atomic charges found within a MOF is roughly 1.5 e , however this value is usually much lower (between 0 and 1). The short range electrostatic interaction energy between this value and the carbon atom from CO₂ (with a charge of $+0.645 \text{ e}$) is shown as the blue curve in Figure 5.4: a. One can see that the majority of the interaction energy is found before the $5 \text{ \AA}$ cutoff, shown as a red vertical line. Please note, the electrostatic energy shown in Figure 5.4: a is a repulsive energy between two like charged ions. This was plotted to show the magnitude of the energy only, as oppositely charged atoms would see this function inverted over the origin. Also note, the long range electrostatic energy term is not discussed here. This involves a summation over Gaussian charges in reciprocal space which is MOF dependent, as vectors are needed for the periodic simulation cell, as well as charges assigned to each atom. While this value can fluctuate depending on the MOF configuration and

number of CO₂ particles, the average absolute value of the long range portion of the electrostatic interactions for a CO₂ molecule is on the order of 0.1 kcal/mol.

The other term contributing to the short range interaction energy of CO₂ with the framework is the van der Waals, or instantaneous dipole interaction term. This term is typically much shorter range than the electrostatic interactions, and is calculated using a Lennard-Jones function

$$E_{VDW} = \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{\tilde{r}_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{\tilde{r}_{ij}} \right)^6 \right], i \neq j \quad 5.10$$

The values ε_i and σ_i are Lennard-Jones parameters which determine the interaction strength and the range of the function respectively, the values of which are assigned to different atom types. Carbon, for example has values of $\varepsilon_C = 0.105$ kcal/mol and $\sigma_C = 3.43$ Å, as taken from the UFF force field.⁴ The green plot in Figure 5.4: a shows the van der Waals interaction energy between two carbon atoms, which decays rapidly before 5 Å. Integrating both the electrostatic and van der Waals functions after the 5 Å cutoff shows that, with the extreme case, we are missing at most 0.04 kcal/mol in dispersion interactions (green area) and 0.2 kcal/mol in short range electrostatic interactions per atom (blue area).

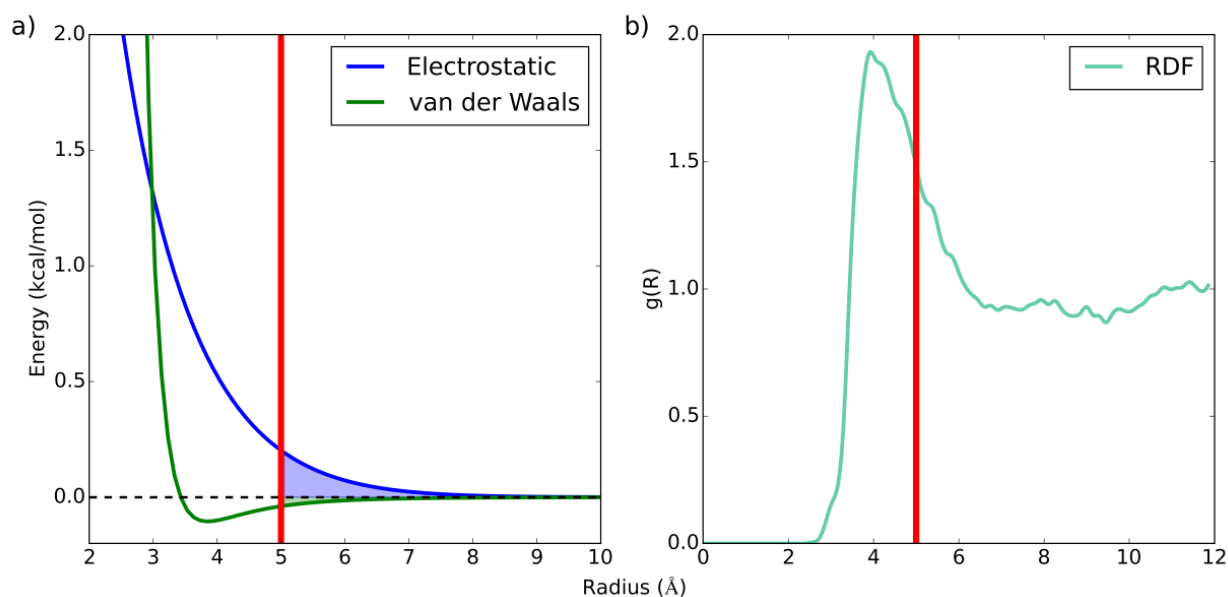


Figure 5.4: a) Short range interaction energies of a fictitious pair of carbon atoms with charges +0.645 and +1.5. The dashed black line indicates an interaction energy of 0 kcal/mol b) the radial distribution plot of all MOF atoms from the C and O positions of 4183 CO₂ binding sites. In both plots, the bold red line indicates the cutoff radius, R, used to extract active sites from MOFs.

While 5 Å appears to capture a majority of the short-range interaction energy, there is also the question of making this radius larger to capture more framework atoms in the active site description. This will guarantee a larger portion of the short-range energy is captured, however it also introduces the possibility of including too many atoms in the description of the active site. The goal of identifying the local pore characteristics was to include the closest layer of atoms surrounding the CO₂ binding site. We hypothesize that these atoms will provide the majority of pore geometric features and energetic interactions necessary to describe CO₂ binding in that location. Including atoms in regions outside of the first neighbour shell could introduce the possibility of discovering ‘uninteresting’ similarities between binding sites when comparing them with the clique detection method described in detail below, as the method is designed to recognize geometric similarities between fragments regardless of the CO₂ location. Figure 5.4b represents a pair distribution function which plots the normalized distribution of all atoms around

the carbon and oxygens of CO₂ found in 4183 binding sites from hypothetical MOFs generated in chapter 3. The 5 Å radius captures the majority of the first peak, confirming that the first neighbours will be included when active sites are extracted. We note that the peak extends to roughly 7 Å indicating that there is an increased probability of finding atoms at this distance from the CO₂ binding sites, however considering the "highest possible" interaction energies plotted in Figure 5.4a are negligible at this distance, the correlation is likely unrelated to CO₂ binding. Beyond this point, the pair distribution function shows a uniform average density of atoms, indicating that increasing the radii beyond this point would result in capturing atoms which are completely uncorrelated to CO₂ binding sites.

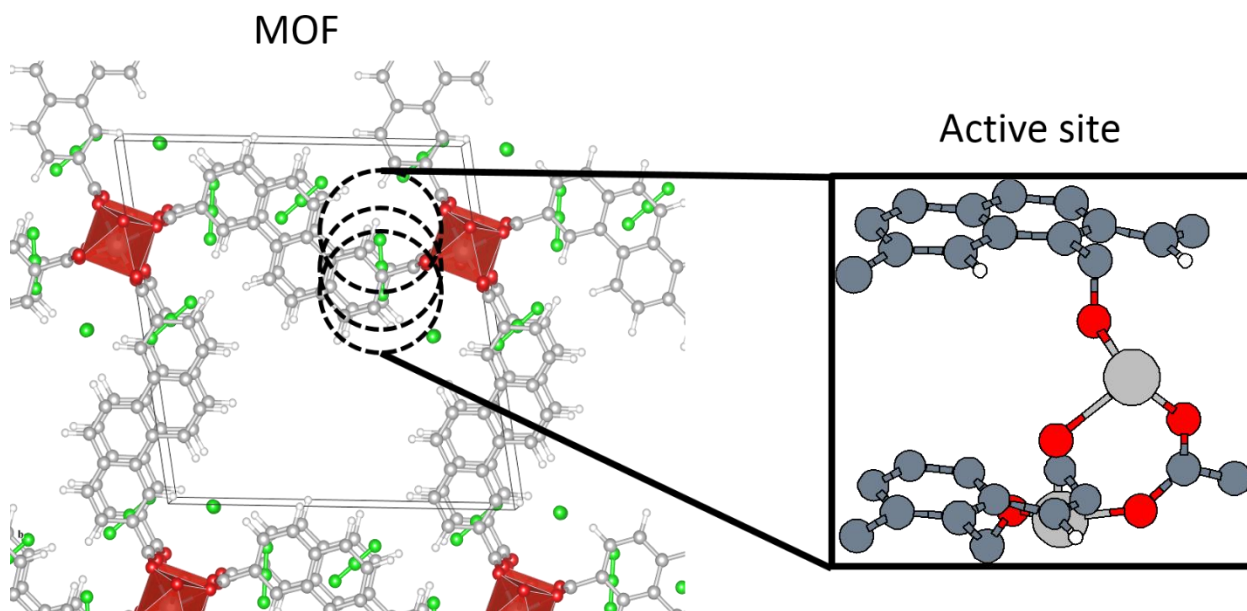


Figure 5.5: Extracting an active site from a CO₂ binding site in a MOF. The MOF atoms are grey, white and red with red tetrahedrons representing the vanadium inorganic clusters, the CO₂ binding sites are represented in green. A radius of 5.0 Å around each CO₂ atom is used to cut out the active site seen on the right (CO₂ excluded).

A physical description of extracting the active site from a CO₂ binding site is represented in Figure 5.5. The extracted MOF atoms retain their charges and Lennard-Jones parameters for the purposes of quantifying the energy contributions of the adsorbophores to CO₂ binding, discussed later.

5.3 Substructure matching

For the purposes of discovering adsorbophores, we implemented a pattern recognition program using a concept from graph theory called maximum clique detection. A maximum clique, as will be described below, represents the maximum number of common atoms in the same geometric arrangement, regardless of orientation, between two molecular fragments. Maximum clique detection is a familiar concept in bioinformatics as it is frequently used in discovering common fragments in small organic molecules.⁵ However to our knowledge, such an approach has never been applied to periodic metal organic frameworks before. For this reason, some definitions will be presented before proceeding to the discussion of the algorithm.

A labelled graph G can be represented by the set $G(V, E, L, \mathcal{L})$ where V is the graphs set of vertices, E the set of edges, L , the set of labels, and $\mathcal{L}$, the mapping of labels in L to vertices and edges in V and E . Labels add extra meaning to the vertices and edges of a graph depending on the associated domain. For example, in the case of a graph describing a molecule, vertex labels can be the element names, and edge labels the bond types connecting vertices (single, double, aromatic, etc.). Here we state that the labels exist as simply a list of elements and bonds until the mapping $\mathcal{L}$ assigns 'carbon' to the intended vertex, or 'aromatic' to the intended edge. Two graphs, G_1 and G_2 can be considered isomorphic if there exists a bijective map, or one to one relationship, f , such that $f: V_1 \rightarrow V_2$ and for all $v \in V$, $\mathcal{L}_1(v) = \mathcal{L}_2(f(v))$ and $f: E_1 \rightarrow E_2$ and for all $e \in E$, $\mathcal{L}_1(e) = \mathcal{L}_2(f(e))$. In other words, every vertex, edge and associated label in G_1 can be matched with only one labelled vertex or edge in G_2 (and vice-versa).

A subgraph, or substructure as we have called it (for the purposes of equating a graph to a set of atoms), is a graph g which is isomorphic to a subset of vertices and edges in G . A common subgraph between two graphs G_1 and G_2 consists of subgraphs $g_1 \in G_1$ and $g_2 \in G_2$

such that g_1 and g_2 are isomorphic. A maximum common subgraph consists of the common subgraphs g_1 and g_2 which possess the largest possible number of vertices from G_1 and G_2 , respectively.

A clique is a fully connected graph G , which is to say each pair of vertices $u, v \in V$, are connected via an edge in $e \in E$. This concept is shown in Figure 5.6, where the graph G_1 is a clique since vertices A, B, and C are interconnected. G_2 is not due to the missing edge between vertex A and D.

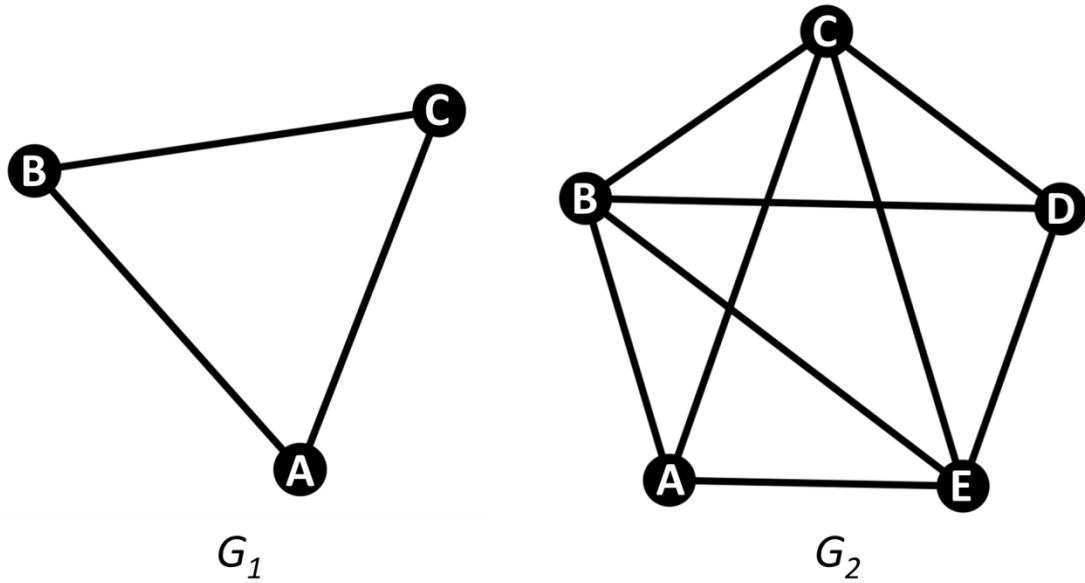


Figure 5.6: Two graphs, G_1 is a clique, G_2 is not.

A correspondence graph $G_C(V_C, E_C)$ is a graph containing pairs of vertices from graphs $G_1(V_1, E_1, L_1, \mathcal{L}_1)$ and $G_2(V_2, E_2, L_2, \mathcal{L}_2)$ such that for $v \in V_C$, $v = (v_1, v_2) \in V_1 \times V_2$. The Cartesian product $(V_1 \times V_2)$ of the vertex sets ensures that every vertex from the first graph is paired with every vertex from the second. Here we add an additional requirement; namely, the Cartesian product of two vertex sets contains only those pairs which possess the same label, i.e. $(v_1, v_2) \in V_1 \times V_2$ iff $\mathcal{L}_1(v_1) = \mathcal{L}_2(v_2)$. Vertices of the correspondence graph $u, v \in V_C$ are connected via an edge in E_C if the edge labels are equal between edges of the original graphs

$\mathcal{L}_1(u_1, v_1) = \mathcal{L}_2(u_2, v_2)$ where $u_1, v_1 \in V_1$ and $u_2, v_2 \in V_2$. The correspondence graph is a measure of the similarity between two graphs. Similarity (based on labelling) is represented in the correspondence graph by edges, where the more edges in the correspondence graph signifies a greater degree of similarity between the original graphs.

5.3.1 Correspondence graphs of two molecular fragments

The initial step in discovering common atomic distributions between two active sites is to represent them as graphs. Here we say that the graph representation of an active site possesses a vertex set V which is a general description of the atoms and an edge set E , which by intuition may be thought of as the bonds between the atoms, but here we define the edges differently; every vertex (or atom) is connected to every other vertex by an edge, regardless of their original bond topology. We define labels for the vertices, which in our case are set to the chemical elements. In principle, the labels could be any value associated with atomic sites (eg. charge, electronegativity, hybridization, etc.) in this study however, we used the chemical elements as labels for the purposes of identifying chemical fragments which could serve as design targets. Labels for the edges are the pairwise distances between the atoms. This labelling is crucial when defining the geometric similarity between two active sites. Thus if two carbon atoms represented as vertices v_1 and v_2 are 4 Å away, the edge label connecting their respective vertices would be $\mathcal{L}(v_1, v_2) = 4$.

The correspondence graph (G_C) is constructed for the sake of measuring the similarity of two molecular fragments, where each vertex in G_C represents an atom from the first and second fragment, so long as they possess the same labels. For example a vertex in the correspondence graph $v = (v_1, v_2) \in V_1 \times V_2$ between two fragment graphs $G_1(V_1, E_1, L_1, \mathcal{L}_1)$ and $G_2(V_2, E_2, L_2, \mathcal{L}_2)$ consists of a carbon atom from molecule 1 ($\mathcal{L}_1(v_1) = \text{'C'}$) and a carbon atom

from molecule 2 ($\mathcal{L}_2(v_2) = \text{'C'}$). Edges were assigned between vertices in the correspondence graph by performing a comparison of edge labels from the original active sites. Specifically, two vertices in the correspondence graph, $u = (u_1, u_2) \in V_1 \times V_2$ and $v = (v_1, v_2) \in V_1 \times V_2$, possess an edge $e \in E_C$ if the edge labels between their constituent vertices are the same. That is to say, an edge in the correspondence graph exists if the label of the edge $e_1 = (u_1, v_1) \in E_1$ is equal to the label of the edge $e_2 = (u_2, v_2) \in E_2$. We can therefore say there exists an edge in the correspondence graph if the atoms from both the molecular fragments possess the same distance. This condition is depicted graphically in Figure 5.7. In practice, a tolerance is applied to the evaluation of edge label equality to ensure that slight structural deviations are accommodated when matching substructures.

Even with the introduction of a distance tolerance, our definition of correspondence and graph ‘similarity’ is very specific to rigid conformations. For example two aliphatic C_5 chains would not be recognized as the same if one was linear and the other distorted by some dihedral rotations. This introduces some difficulties if one were interested in matching fragments with significant conformational degrees of freedom, such as in drug molecules. In which case, clique detection methods are typically applied to thousands of possible conformations of each drug molecule. Here, however, the rigidity of the MOF and the pocket surrounding CO_2 binding sites permits us to ignore these considerations.

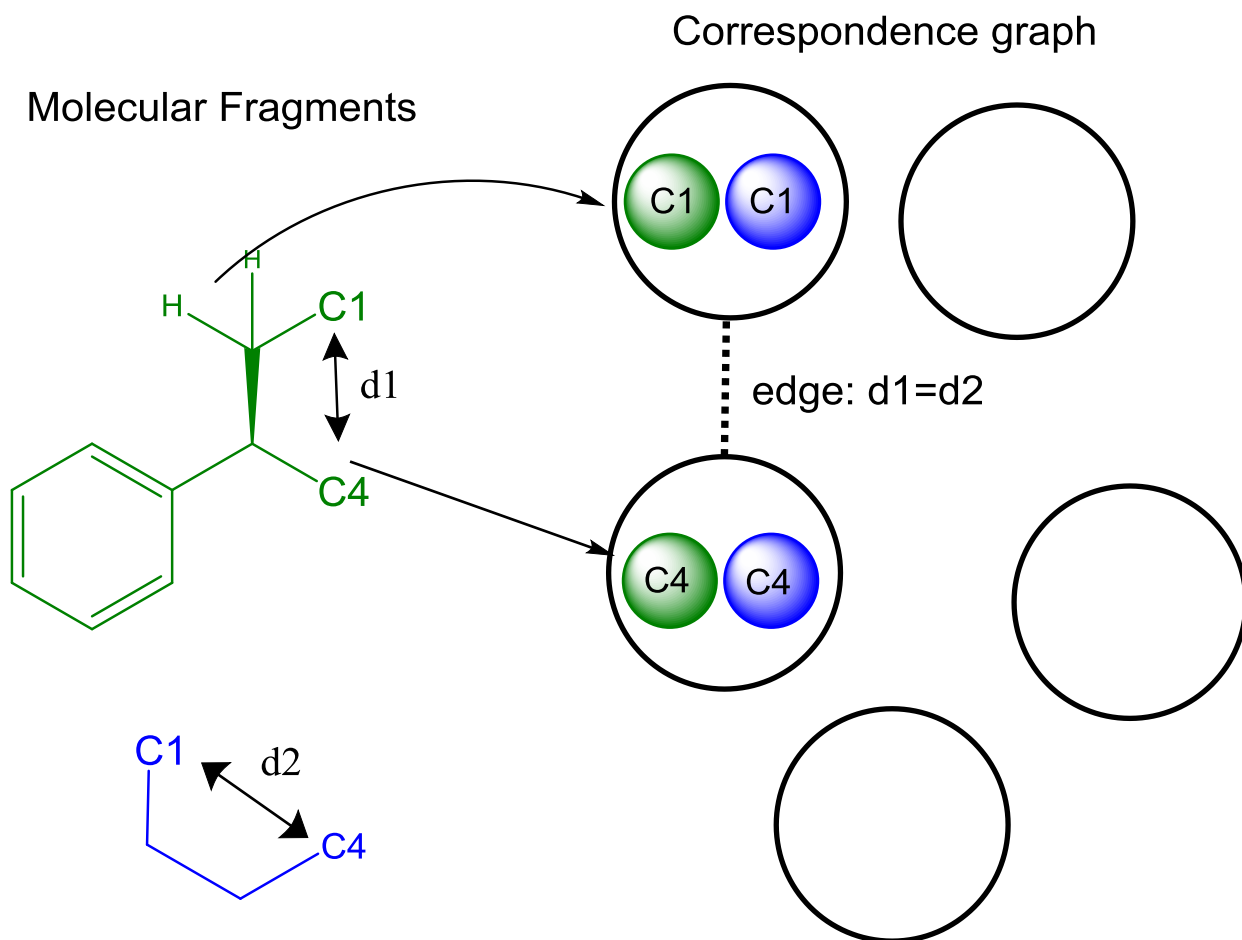


Figure 5.7: The correspondence graph is composed of pairs of atoms from each molecular fragment, depicted as green and blue molecules on the left. An edge exists between the two vertices of a correspondence graph if the distances between the atoms in the original molecular fragments are the same within a specified tolerance.

5.3.2 Common Subgraphs from Maximum Clique Detection

It was shown that the maximum clique from a correspondence graph determines the maximum common subgraph between two graphs.⁶ A graphical depiction of identifying the maximum clique from a correspondence graph is shown in Figure 5.8 which shows that the common set of atoms between the two molecular fragments is a bent C₄ chain. The correspondence graph shows the pairs of atoms which correspond to the C1-C4 atoms in both fragments are all interconnected (a clique, shown with red edges), due to their similar pairwise distances in the original fragments.

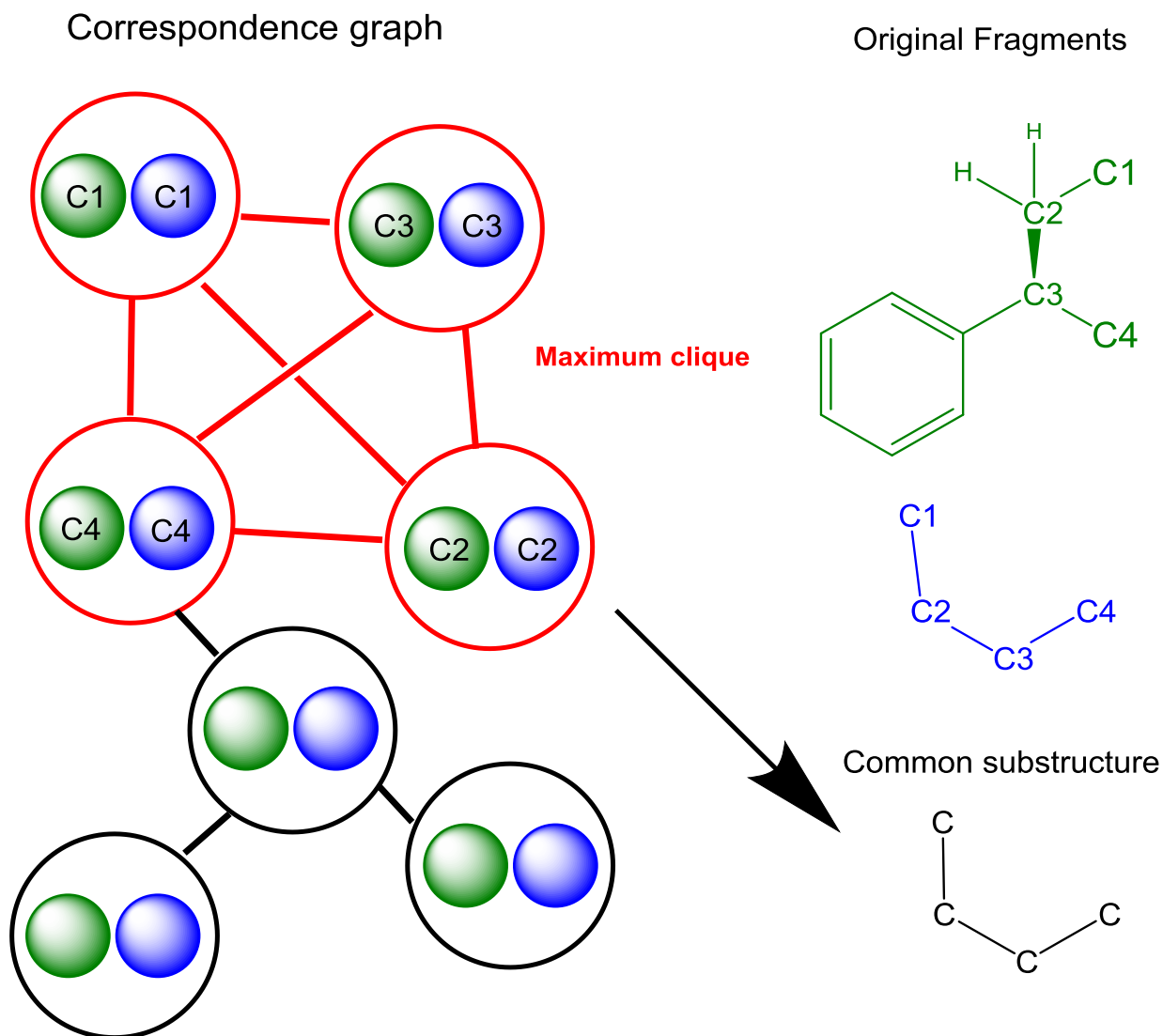


Figure 5.8: The maximum clique (red nodes and edges) from a correspondence graph yields the largest common substructure of two molecules.

Discovery of maximum cliques is a Non-deterministic Polynomial-time(NP)-hard problem. This means that, despite its name, the time required to discover the maximum clique in a graph will grow exponentially (not polynomially) with the number of vertices and edges in the correspondence graph.⁷ The nature of this problem lies in the search procedure of finding a maximum clique when presented with a graph of arbitrary size. A basic algorithm for finding the maximum clique is discussed verbosely here to highlight the complexity. Upon initiation the program must decide upon which vertex to commence the maximum clique search. For this

example we say the decision is arbitrary, however improved search methods order the vertices based on the number of edges incident upon it. The program then lists all the neighbours of the starting vertex (i.e. vertices which are connected to it by an edge). These neighbours could all potentially form a maximal clique with the starting vertex. The algorithm then travels to a vertex in the neighbour list, while keeping track of the neighbours from the first vertex to determine if they are also neighbours of the current vertex. Neighbours of the current vertex which are not common with the initial vertex are removed from the search. The algorithm then travels to vertices in the common neighbours as before, until there are no more common neighbours to traverse. At this time, the program has found a *maximal* clique. That is, a clique which is not part of a bigger clique. This may not be the maximum clique, which could be a larger sequence of vertices elsewhere in the graph. Thus, once the program cannot proceed any further with the current maximal clique, it must backtrack out of the current set of vertices to search for other cliques in the graph in the same fashion. Upon identifying all of the maximal cliques in the graph, the largest of these will be the maximum clique.

Improvements can be made upon the efficiency of searching for the maximum clique by making non-arbitrary decisions on which vertices to traverse during the search. This involves what are called branch-and-bound methods⁸ for searching for the maximum clique. In these algorithms, checks are made to evaluate if the current branch being searched (maximal clique) will surpass the largest maximal clique already found. If this is not the case then the current search is discarded and the algorithm backtracks out of the branch. Each clique being searched is evaluated for its upper bound based on estimation. The estimate is based on graph coloring, which is performed prior to searching for the maximum clique.

Graph coloring can be defined as a label assignment (separate from the atomic labelling) to the vertices of a graph such that no two adjacent vertices possess the same label (or color). If there exists an edge $e = (u, v) \in E$ for two vertices $u, v \in V$ then the graph coloring $\mathcal{L}$ ensures that $\mathcal{L}(u) \neq \mathcal{L}(v)$. It was shown that the upper bounds for a maximum clique is equal to the number of colors needed to describe a graph.⁹ A specific example of graph coloring and how it demonstrates an upper bound for maximum cliques is shown in Figure 5.9. The graph in Figure 5.9a requires 4 colors, and its maximum clique is 3 vertices, while the graph in Figure 5.9b has all 6 vertices interconnected with edges (it is a clique), and requires 6 colors to describe the vertices. A note on the terminology; the word color is used here traditionally, but in computer algorithms ‘colors’ are usually represented as integer values.

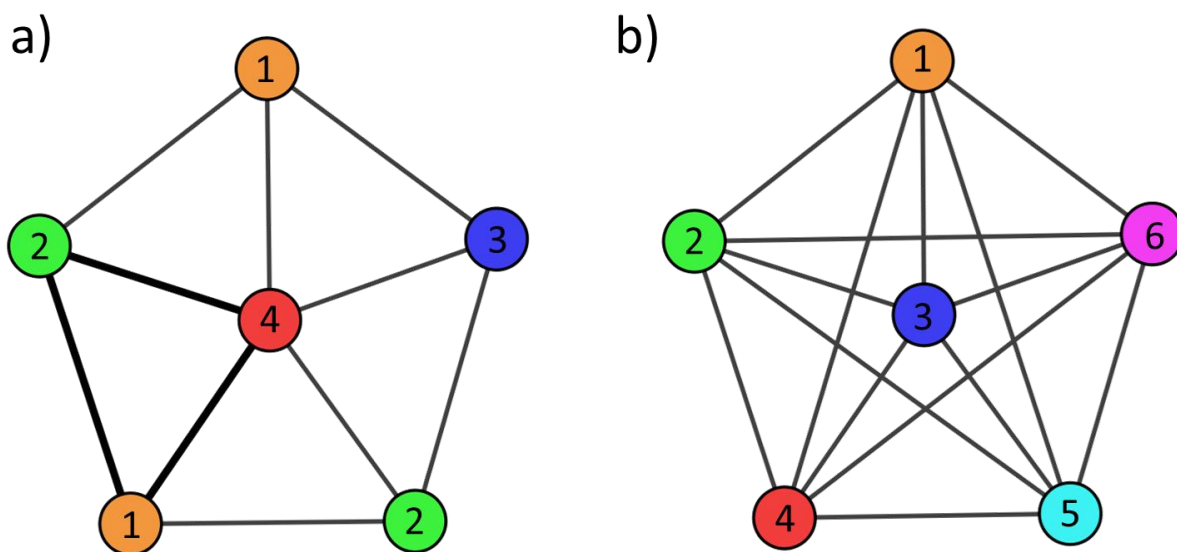


Figure 5.9: A demonstration of graph coloring on two graphs. a) a graph with 6 vertices containing 4 colors. The maximum clique of this graph is 3, shown in bold edges. b) A fully connected graph with 6 vertices and 6 colors.

The algorithm used to find maximum cliques in our program was adopted from the work of Konc and Janežič where they demonstrate improved efficiency by performing a coloring of

the graph, after which the maximum clique is searched using a branch-and-bound method.¹⁰ For a full description of the work, the reader is referred to reference 10. Very briefly, the algorithm works as follows, the set of vertices in the graph are initially ordered based on their color. In practice, the color classes are assigned an integer value '1', '2', '3', etc. The coloring commences with '1' which is assigned to the vertex of the highest degree (which have the potential to form the largest cliques) and continually adds new colors if the vertices visited in the graph are connected to any of the vertices in the existing color classes. This way, vertices which are assigned to color classes which are higher integer values are likely involved in larger maximal cliques. The ordering of vertices with the same color allows the program to efficiently traverse the possible maximum cliques at the outer-most branch (recall, vertices with the same color cannot be connected to each other, and therefore are involved in different maximal cliques). As the program samples new vertices, a decision is made whether to backtrack out of the current vertex if the sum of its assigned color and the size of the current clique is smaller than the largest maximal clique already discovered. This method was shown to roughly halve the compute times of discovering the maximum clique in a number of large graphs with the number of vertices ranging between 100 – 1000, as compared with an algorithm which does not color the graph prior to clique detection.¹¹ Ultimately, the time improvement depends on the size of the graph and the number of edges in the graph, where the program showed the best improved performance with more edge dense graphs.¹⁰

5.3.3 *Recognizing SBUs in MOFs using Maximum Clique Detection*

Before proceeding to the algorithm which was designed to detect frequent common substructures (adsorbophores), as an initial test, a series of hypothetical MOFs were selected to determine if maximum clique detection could be applied to identify all their constituent SBUs and functional groups. This was accomplished by performing repeated maximum clique

detections in a correspondence graph generated by pairing known molecules (SBUs and functional groups) with the entire set of MOF atoms.

SBUs or functional groups were considered ‘found’ when a maximum clique was discovered that was the same size as the fragment being searched. These atoms were removed from the MOF and a new correspondence graph (less those atoms found in the maximum clique) was created so that further maximum cliques could be detected. The program would continue searching for the SBU or functional group within the MOF in this fashion until the size of the discovered maximum cliques dropped below the size of the fragment being searched, after which time it would stop searching for that fragment in the MOF and move on to search for the next SBU or functional group.

The order in which fragments are searched in MOFs is quite important in this algorithm. In some cases smaller SBUs were substructures of larger SBUs, as an example consider a benzene SBU and a larger polyaromatic SBU, containing multiple conjugated benzene rings. Searching for the smaller SBUs first and finding them in the larger SBUs eliminates these atoms from future maximum clique searches in the MOF, so upon completion of finding all the small SBUs in the MOF, the program will find no large SBUs and the program yields an improper description of the MOFs molecular fragments. Thus prior to maximum clique detection, the SBUs and functional groups were organized hierarchically based on their size, and searched within the MOF sequentially from largest to smallest.

Ultimately, the success or failure of finding a SBU or functional group within a MOF will depend on the distance tolerance set when constructing the correspondence graph. There will be some disagreement between the atomic configuration of the (typically highly symmetric) SBU and its configuration within a MOF. This is because the atomic positions were relaxed to their

minimum energy for each MOF, and as such the atomic positions are subject to the unique strain presented within each MOF structure. A tolerance in the atomic distances must therefore be accounted for in order to correctly identify SBUs within MOFs. However setting this tolerance too low, such that it recognizes atom fragments in different molecules as the same despite length differences on the order of inter-atomic bonds, will yield poor molecular fragment recognition.

In what follows, we evaluated the performance of the clique detection method by finding all the SBUs within a diverse set of MOFs assembled with different SBUs, underlying connectivity, and functional groups. In addition, we evaluate which tolerance level is sufficient to recognize all the SBUs and functional groups in each MOF, which will later be used for discovering adsorbophores.

The effectiveness of the clique detection program was determined if it could identify all the chemical fragments (functional groups and SBUs) within the unit cell of the material. The clique detection program failed if any atoms remained within the unit cell without being associated with a molecular fragment. As such, a problem presents itself when searching for molecular fragments within periodic materials described by a unit cell. The program will not recognize molecules in MOFs which cross the border into one of the image cells, as their relative distances will not coincide with the reference molecule. A simple two dimensional example of this is presented in Figure 5.10 where the benzene molecule in the periodic system is split between two images. In this case the distances between atoms in the unit cell will be artificially long, and the clique detection algorithm will not recognize it as a benzene molecule. Thus the clique detection must be performed on an even larger system than the unit cell, which we chose to be a $2 \times 2 \times 2$ supercell. This effectively extends the repeat unit of the MOF out by 0.5 cells in each direction, allowing for recognition of molecules at the border of the unit cell. Of course, this

generates fragmented molecules at the border of the supercell, however we ignore these when considering successful identification of all the molecules making up a MOF.

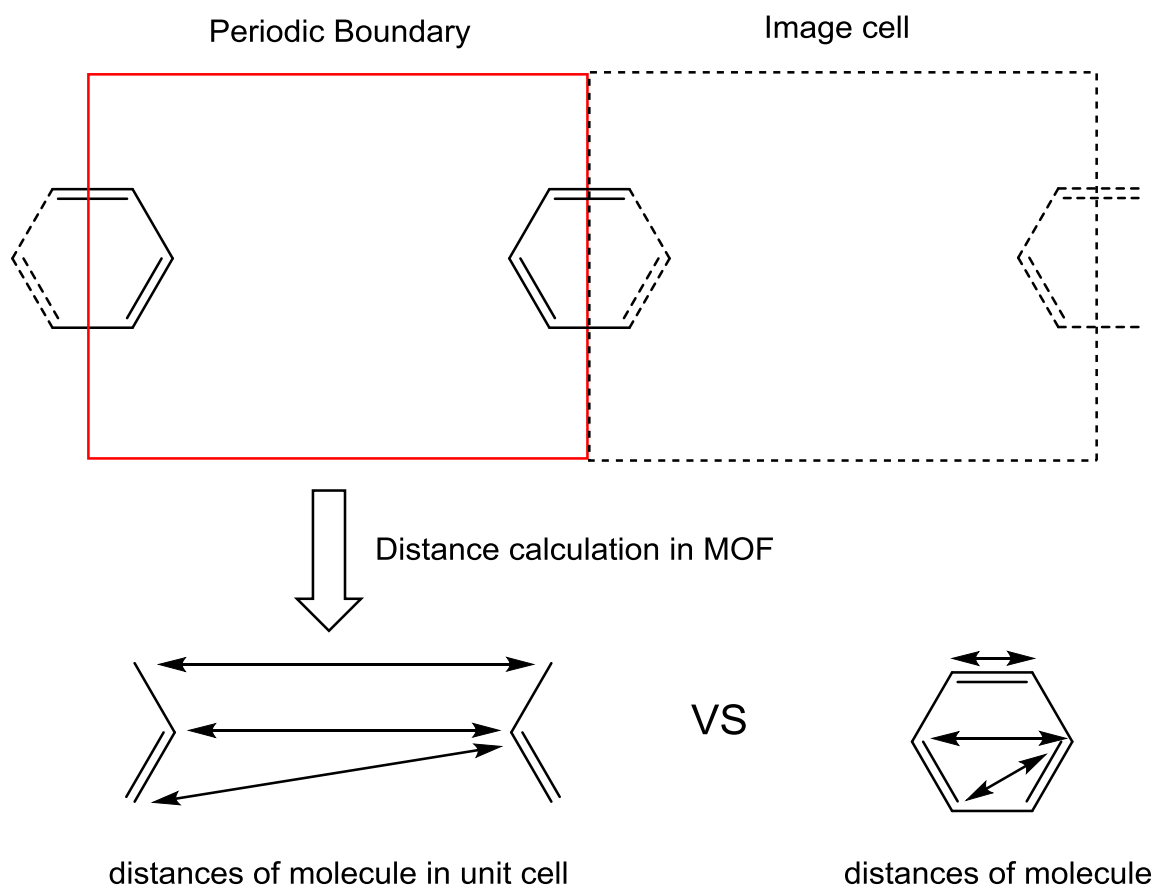


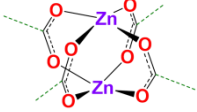
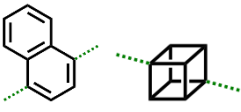
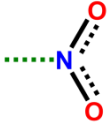
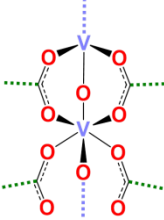
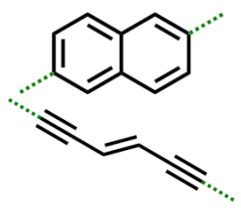
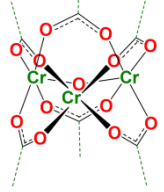
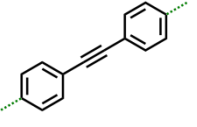
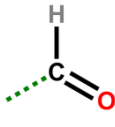
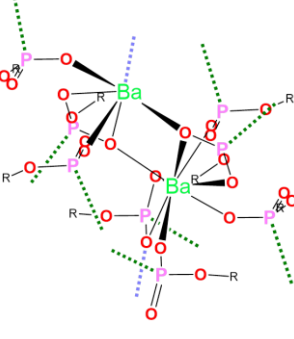
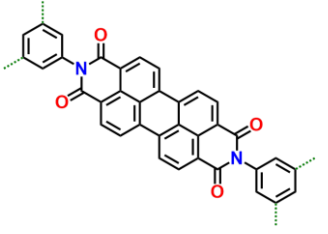
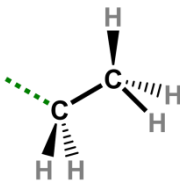
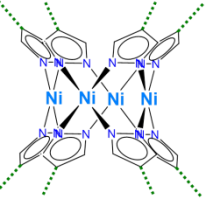
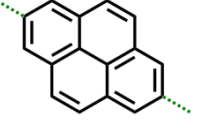
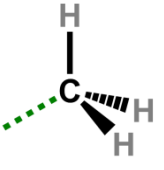
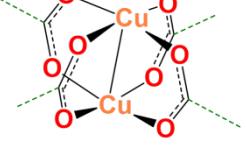
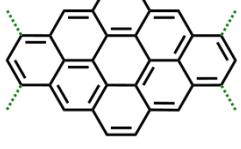
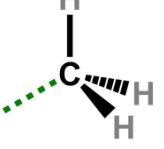
Figure 5.10: Molecule recognition in bulk systems is made difficult by the presence of molecules at the borders of repeating units.

A set of MOFs were chosen that spanned a diverse range of inorganic and organic SBUs and functional groups, to ensure broad applicability of the clique detection algorithm. Eight MOFs were analyzed, which are described by their SBUs and topologies in Table 5.1. The clique detection algorithm was tested on each of these MOFs by searching for the SBUs and functional groups described in the table. While some of the MOFs possess the same SBUs, their underlying connectivities (topology) are different, yielding unit cells of different sizes and shapes. For example MOFs 6 and 7 contain the same Cu paddlewheel SBU, however their topologies are different, MOF 7 consists of a total of 15 SBUs in its unit cell, compared with the 4 SBUs used

to build MOF 6. The size of the MOF will ultimately affect the time required and difficulty in finding the molecules as the correspondence graphs with molecules grows.

While, in general, MOFs are constructed with very rigid SBUs to maintain structural integrity, in some cases functional groups with a large degree of conformational freedom can be found which can yield interesting properties (such as methoxy functional groups that provide a gating mechanism in gas adsorption¹²). The extra degrees of freedom these groups add to crystalline structures provide a challenge to the structure recognition program, which must account for possible angle and bond deviations by relaxing the user-defined distance tolerance value. Both MOFs 4 and 8 possess ethyl groups as a challenge for the clique detection method. The tolerance value could be altered on-the-fly for atoms recognized to have some conformational flexibility, however this was not implemented in the current version of the code.

Table 5.1: Description of the hypothetical MOFs used to validate the clique detection program. The metal and organic SBUs, the functional group and the topology used to construct the MOFs are listed in the table.

MOF number	metal SBU	organic SBU(s)	functional group	topology
1				pcu
2				sra
3				acs
4				eta
5				bcu
6				fof

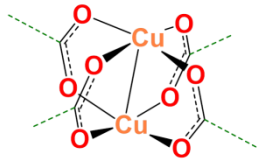
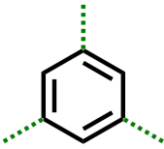
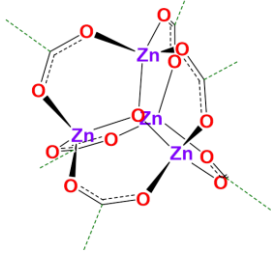
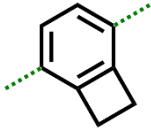
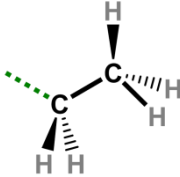
7				tbo
8				pcu

Figure 5.11 demonstrates the dependence of the success of the clique detection algorithm on the tolerance parameter. Here, if one sets the tolerance too low, deviations in the MOF structures will prevent the discovery of the SBUs and functional groups. Atoms in the unit cell of the MOF remain unassigned to SBUs or functional groups, represented in the graph as a less than 100% assignment of atoms. All of the MOF atoms could be assigned to their appropriate molecular fragment using a distance tolerance of 0.4 Å. This value is therefore a conservative distance tolerance one can set to successfully find molecules in the database of hypothetical MOFs, and is what we set when discovering common adsorbophores in the following section.

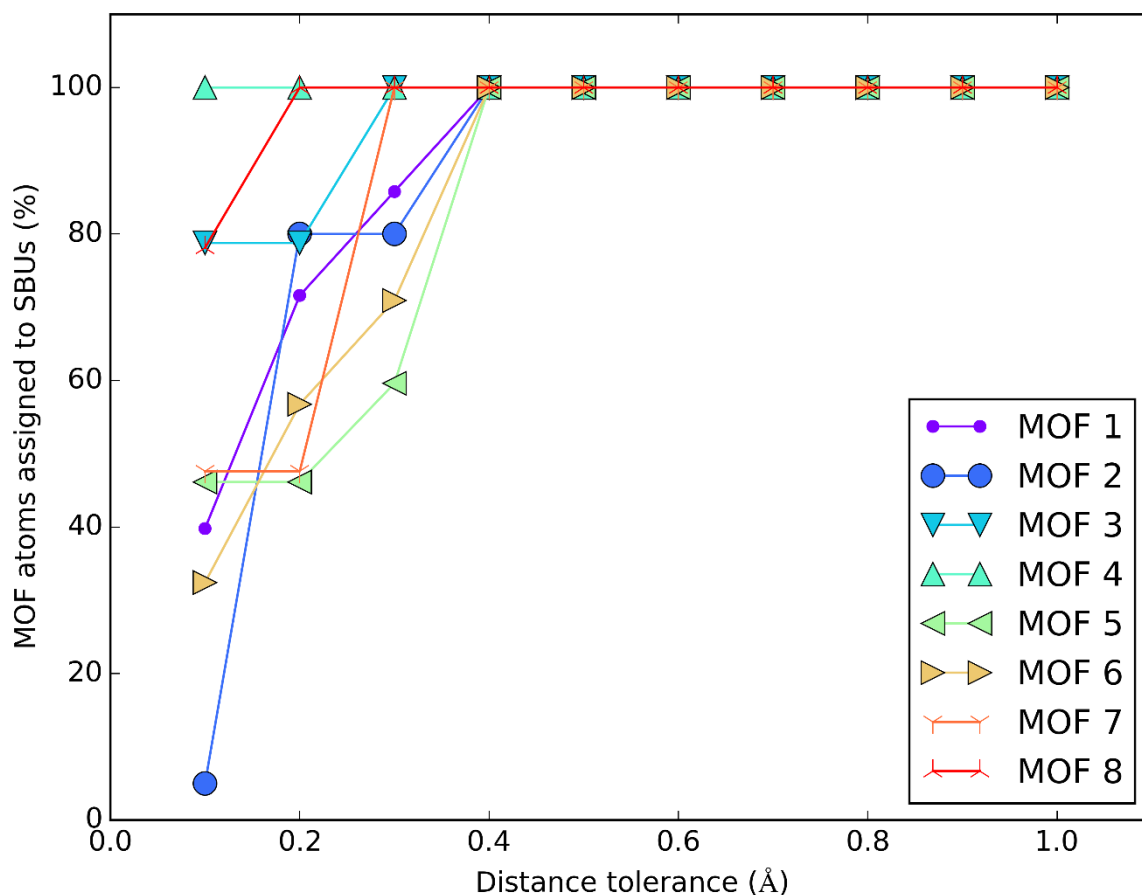


Figure 5.11: Effectiveness of the maximum clique detection method in finding all the SBUs and functional groups inside the test set of hypothetical MOFs. The numbering in the legend corresponds to the MOF numbering provided in Table 5.1. Plotted is the percentage of atoms in unit cell found belonging to an SBU or functional group as a function of the distance tolerance parameter.

5.3.4 Finding Common Adsorbophores by subgraph mining

The main purpose of the maximum clique detection algorithm is to compare a large number of active sites to find common CO₂ binding site structural motifs. In the previous section it was demonstrated that the maximum clique detection algorithm could identify SBUs and functional groups within MOF structures by first, knowing what SBUs were contained in the MOF and second, constructing correspondence graphs by pairing representative SBU molecules with the MOF. The challenge underlying the discovery of common adsorbophores is that we have no *a priori* knowledge of what we are looking for, meaning one must compare all the collected

active sites to discover what features are important for CO₂ adsorption in what is called frequent subgraph mining. Robust discovery of the common adsorbophores requires repeated, exhaustive pairwise comparisons of common atoms found from the maximum clique algorithm which, in short time becomes computationally intractable. Performing exhaustive comparisons of all the active sites also suffers from the assumption that they all possess the same subgraph of atoms that contribute to the CO₂ binding, while in reality many possible adsorbophores could (and do) exist.

In practice the exhaustive search for frequent subgraphs must be sacrificed for computational efficiency. This is recognized in drug discovery programs, as various methods have been implemented to identify common pharmacophores from a set of active ligands with various levels of approximation and heuristics.^{13–15} These searches are complicated by the fact that drugs are typically small molecules with several degrees of conformational freedom. Thus finding frequent pharmacophores requires sampling on the order of thousands of conformations of a single molecule. One example of finding frequent pharmacophores involves ‘growing’ cliques¹³; starting with all possible cliques with two vertices which, if found to be present in a significant number of molecules, is grown by one vertex, and so on until the frequent clique reaches a user-defined maximum. To significantly reduce computational time (and it could be argued, to improve the power and predictability of the pharmacophore model) atomistic graphs are often reduced to more general or coarse-grained molecular descriptions called *features*, where benzene rings are converted to a single ‘aromatic’ vertex and hydroxyl groups are converted to ‘hydrogen donor/acceptor’ vertices. This has the advantage that the size of the correspondence graphs will be significantly reduced, thus finding the maximum cliques will take less time, as well, one can limit the size of the pharmacophore searched to a very small number

of nodes (typically 3-5) while still yielding informative chemical properties of protein binding. In many cases, the common substructures found in these searches serve as a means to an end; they are used as hypotheses for developing a model of activity for the protein in question.

In this study, we took a different approach to frequent subgraph mining. No prior robust study has ever been performed to identify CO₂ binding features in MOF pores, so our knowledge of what makes a good MOF binding pocket for CO₂ is considerably limited, and reducing active sites to a set of features would be too presumptuous. Thus we based the search on fully atomistic molecular fragments in lieu of coarse-graining the systems. Frequent adsorbophores were identified by finding common substructures of randomly paired active sites. Starting with the population of N active sites found, each was randomly paired together to produce the set of approximately $N/2$ common subgraphs using the maximum clique detection algorithm detailed above. These common subgraphs are subsequently paired and their common atoms were found until no new subgraphs could be discovered. The resulting frequent subgraphs are then ranked by the number of active sites they contain.

The success of the random pairing method is based on the assumption that the maximum common subgraph obtained from an arbitrary number of molecules will not depend on the order in which the molecules are paired. This is of course, an approximation which ignores the fact that the order may have some inherent effect on the final adsorbophores. As a thought experiment, consider two active sites, AS1 and AS2, which when paired together, possess several isolated common subgraphs. The maximum clique algorithm will find the largest of these to be carried forward for the next random pairing. Now we introduce a third active site, AS3, which possesses a common subgraph with the other two active sites, *but not the largest one*. If this third site is paired with the maximum clique found from the first two active sites, it may only find a smaller

fragment which is common amongst the three, as opposed to the largest possible subgraph. This order bias can yield fragments which have unintentionally discarded atomic features potentially important for activity.

The order dependency in which the subgraphs are paired can be minimized by 1) reducing the size of the active sites extracted from each MOF; the fewer atoms that exist in an active site decreases the chances of multiple isolated common subgraphs being present upon pairing them and 2) increasing the minimum allowable size of the common subgraph, or minimum clique threshold (MCT); this is a user-defined variable describing a cutoff value of the minimum number of atoms found between two molecular fragments. With respect to the first option, the size of the active site will be determined by the radius from the CO₂ molecule used to extract the MOF atoms. While this value can be tuned to improve the random sampling method, we do not adjust the radius in this study. We deemed a 5.0 Å radius discussed in section 5.2.3 as large enough to capture a majority of the local non-bonded interactions between MOF atoms while still limiting the size to the local environment. We are therefore left to discover which value of the minimum allowable clique size will yield robust and interesting adsorbophores. Increasing the value of this threshold serves two purposes; it ignores cliques which may not be physically or chemically interesting. Consider, for example finding only 3 carbon atoms common from two active sites with sizes greater than 50 atoms. There is likely more information than the three carbon atoms which contribute to CO₂ binding, thus setting a clique threshold greater than 3 will prevent this information loss. This also ties into the second advantage, in that by increasing this value we can align the random pairing method closer with the approximation that pairing order will not overly affect the resulting frequent adsorbophores. By excluding cliques of less than a certain size, we are biasing the frequent subgraph mining to only larger subgraphs.

From our previous thought experiment, let us add the constraint that the maximum clique AS3 possesses with AS1 and AS2 is less than the cutoff value. Thus no matter what order these sites are paired in, the maximum clique between AS1 and AS2 will always be found, while AS3 will remain to search for larger cliques with other active sites. As with any parameterization, a balance must be achieved; setting this cutoff very high will improve the random pairing method, but may be too large to combine common atomic arrangements of potential interest, while setting it too small will allow more active sites to be successfully paired, but the resulting adsorbophores would be reduced to a chemically unintuitive size.

In using the random pairing method to mine frequent subgraphs, there must be an endpoint in which the program no longer tries to pair molecular structures together, and yields a list of the most frequent adsorbophores found. Here we say that the sampling is done when, after performing a random pairing of molecular fragments, no new cliques could be found. We note that in this case some active sites may not be paired, thus another parameter is introduced which ensures K attempts to randomly pair subgraphs together after no new subgraphs were found, we call this parameter the end-of-run pairings (EORP). In the following section, we discuss the effects of varying the parameters of the minimum allowable clique size and the number of end-of-run random pairings to determine how different the resulting adsorbophores are.

In one pharmacophore-based study, a comparison was performed between relatively compute intensive, state-of-the-art subgraph mining techniques which involve generating and comparing various cheminformatics descriptors for molecules,^{16–18} and a random pairing method much like the one described above, where common substructures were found from random pairing of molecules. They demonstrated that the pharmacophores found via random pairing of molecules yields a more robust model for activity in inhibiting growth of human tumor cell lines

when compared with these other methods.¹⁹ It is believed that the random pairing method reduces the redundancy of the mined subgraphs, thus producing more interesting and diverse features for analysis. This also has the advantage of being roughly 400% faster than the ‘state-of-the-art’ methods.²⁰

5.4 Results

Our main concern with the discovery of adsorbophores is to ensure that our frequent subgraph mining methodology discussed in section 5.3.4 will yield robust and transferrable results. In the remaining part of this chapter, we will evaluate the effectiveness of the program across several sets of CO₂ binding sites taken from our database of hypothetical MOFs. We took MOFs from the database assembled with a diverse range of SBUs and functional groups to ensure that we are sampling a representative set of binding sites. We detail several tests which validate the method while issuing a word of caution about using it in systems with more chemical and geometric diversity.

5.4.1 Comparison of Top Adsorbophores Across Random Seeds

The results of the substructure mining technique will be used directly as potential MOF design criteria for CO₂ capture, thus it is important that the top adsorbophores obtained from the program (which are ranked by how many active sites they include) are reproducible and not an artifact of the random pairing method. We therefore compared the adsorbophore results adjusting different values for the parameters discussed in section 5.3.4, namely the minimum clique threshold (MCT) and the end of run pairings (EORP). At each parameter value, we ran the frequent subgraph mining program 30 times (using 30 different random seeds) to obtain the average deviation between runs. We note that the number of adsorbophores found will depend on the number of active sites input into the program, and, as Figure 5.12 shows, there is a linear

relationship between the number of adsorbophores found and the initial number of active sites. This agrees well with the results found in reference 19. From Figure 5.12 we also note that increasing the MCT increases the total number of adsorbophores found, which is expected since this introduces more discrimination in the common atoms found between the active sites.

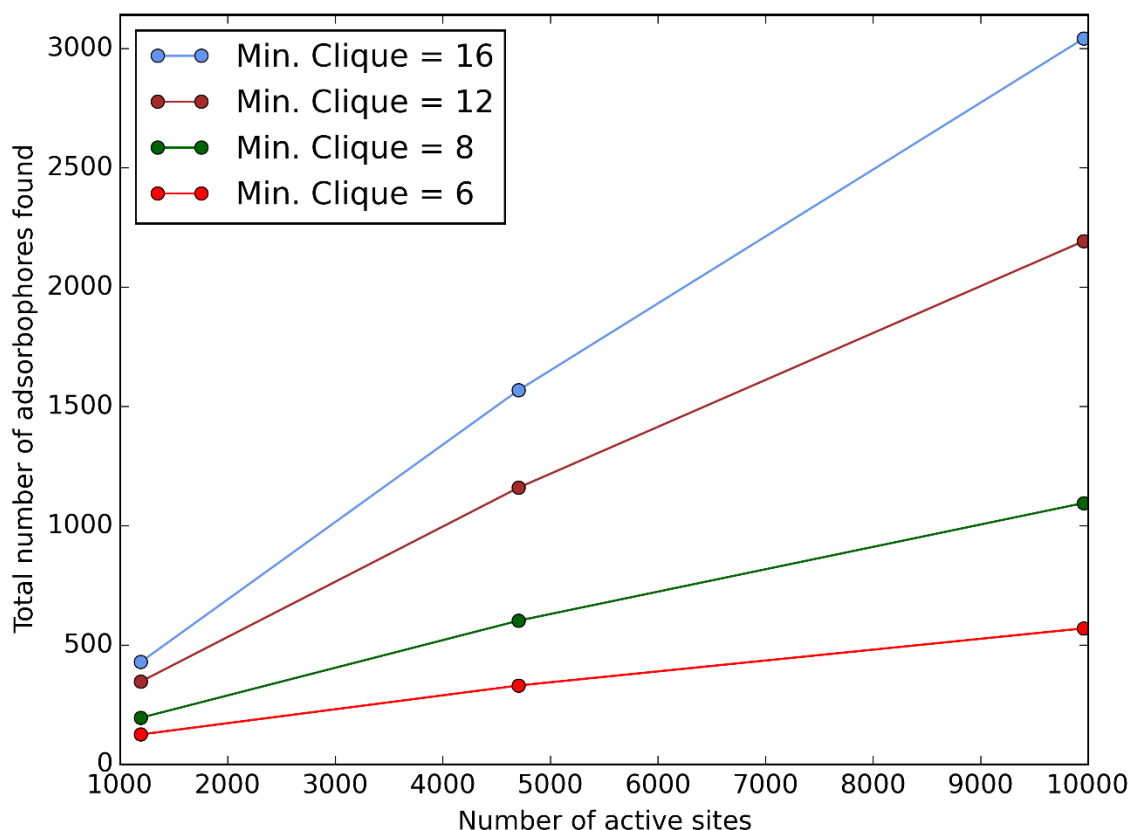


Figure 5.12: Average number of adsorbophores found from 30 random seeds (y axis) as a function of the total number of binding sites input into the common substructure search algorithm (x axis). Each color represents a different minimum clique cutoff. Lines are used to guide the eye.

For the work conducted in this thesis, we focused on the top 50 adsorbophores as molecular features for MOF design criteria, thus the comparisons made in the following studies consider only the top 50 adsorbophores found from each run. Recognizing that the ranks of the top 50 adsorbophores will certainly fluctuate based on the different random seed used, a Hungarian algorithm was implemented to compare the results from two separate runs. The

Hungarian algorithm matches adsorbophores from each set to maximize the total similarity between them (not necessarily maximizing the similarity between each adsorbophore pair). Input into the Hungarian algorithm is the similarity computed between each possible pair of adsorbophores from two different runs. In this study similarity was computed in two ways, first was geometric; each adsorbophore was compared with the other adsorbophores by finding the maximum common subgraph between the two species using the clique detection algorithm, with the same distance tolerance used to identify the adsorbophores (0.4 Å). Here, the larger is the clique found between two adsorbophores, the more similar they are. The second similarity measure was based on which active sites were used to generate the adsorbophore. Each active site was labelled with the hypothetical MOF name it was extracted from and a number corresponding to the order its corresponding CO₂ binding site was listed in the ABSL program. ABSL ranks the binding sites in a MOF based on the intensity of the associated maxima in the carbon 3-D histogram. A typical active site would be labeled ‘hypothetical_MOF.1’ which would describe the atoms surrounding the binding site found with the highest maximum in the 3-D histogram of the MOF named ‘hypothetical_MOF’ (NB: this is a placeholder name. In the hypothetical MOF database, each MOF is named after the organic SBUs, inorganic SBU, functional group, and topology used to build it). The labels of the active sites associated with a substructure were retained throughout the common substructure search algorithm, such that each resulting adsorbophore could be described by the list of active sites combined to form it. In both the geometric and labelling comparisons, the similarity was measured using the Tanimoto index;

$$TI = \frac{|C|}{|A1| + |A2| - |C|} \quad 5.11$$

where C are the common features found between adsorbophores A1 and A2. Completely dissimilar adsorbophores have a TI=0, and approaches 1 as the adsorbophores become more

similar. Each point on the following graphs represent an average similarity value computed across 30 random seeds, where the maximum possible similarity was computed between each combination of random seeds using the Hungarian algorithm.

Figure 5.13 shows the two similarity measures of the top 50 adsorbophores as a function of the MCT for 9558 active sites. It is interesting to notice the drop in geometric similarity from around 0.85 to 0.71 as the MCT passes from 12 to 20. It seems that, in the materials studied here, there is a natural limit in the number of common atoms in each active site which corresponds to between 12 – 14 atoms. As the MCT is increased above this number, the program is capturing molecular fragments which may not necessarily be transferrable to adsorbophores from other runs. We note that a geometric similarity of 0.71 between adsorbophores with sizes of 20 atoms corresponds to an average of 14 similar atoms, which agrees well with the aforementioned limit.

It is striking that the similarity of the active site labels can be more than 0.2 lower than the geometric similarity. This implies that the random pairing method will not consistently pair the same active sites together to form the top 50 adsorbophores, however there is an inherent geometric similarity between the active sites which is still being captured. Perhaps somewhat counterintuitive is the decrease in geometric similarity while the label similarity rises with increasing MCT. The pairing of active sites is more exclusive, suggesting an increasing tendency for pairing the same active sites in different runs, however it also implies that the most frequent adsorbophores contain fewer active sites, which artificially raises the labelling TI while the atomic difference between adsorbophores lowers the geometric TI. Also captured in Figure 5.13 is the relationship between the time required to compute the adsorbophores and the MCT (red plot), which increases considerably if one further increases the value of this parameter.

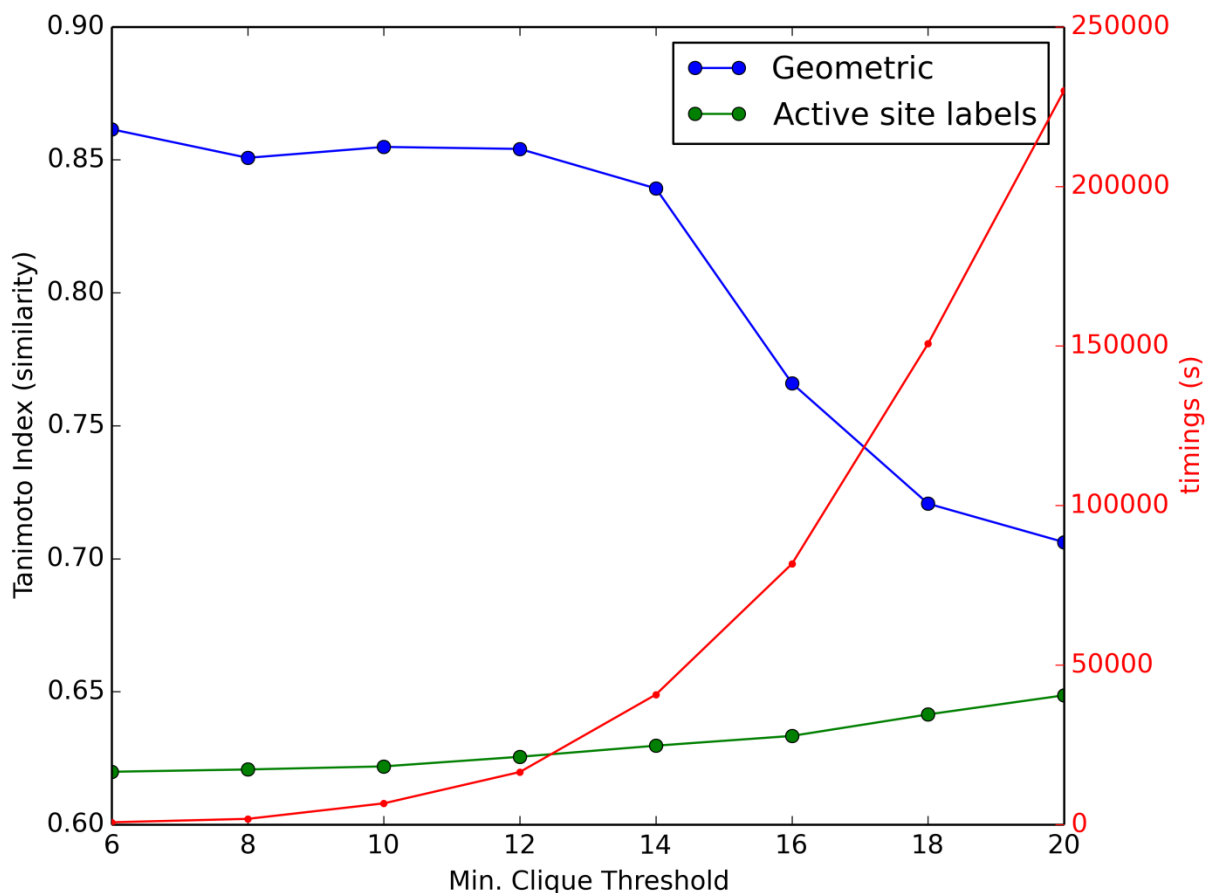


Figure 5.13: Similarity of the adsorbophores from 9558 binding sites. Left axis shows the average Tanimoto similarity computed between 30 runs. The red plot corresponding to the right axis gives the average time required to obtain the adsorbophores.

The termination criteria defined in section 5.3.4 was implemented to ensure that at the end of the run, adsorbophores which have common atoms will be amalgamated to increase their ranking and reduce the redundancy in the results. However as Figure 5.14 demonstrates, there is little improvement in the overall results of the top 50 adsorbophores when increasing the number of end-of-run random pairings from 0 – 900. This suggests that the most frequent adsorbophores are found during the initial random pairing phase. It is interesting to note that the *total* number of adsorbophores drops dramatically by about 1000 when the EORP number increases from 0 to 100 (seen in red in Figure 5.14). This suggests that many of the lower ranked adsorbophores are

poorly sampled and benefit from increasing the termination parameter (note this parameter scales linearly with time).

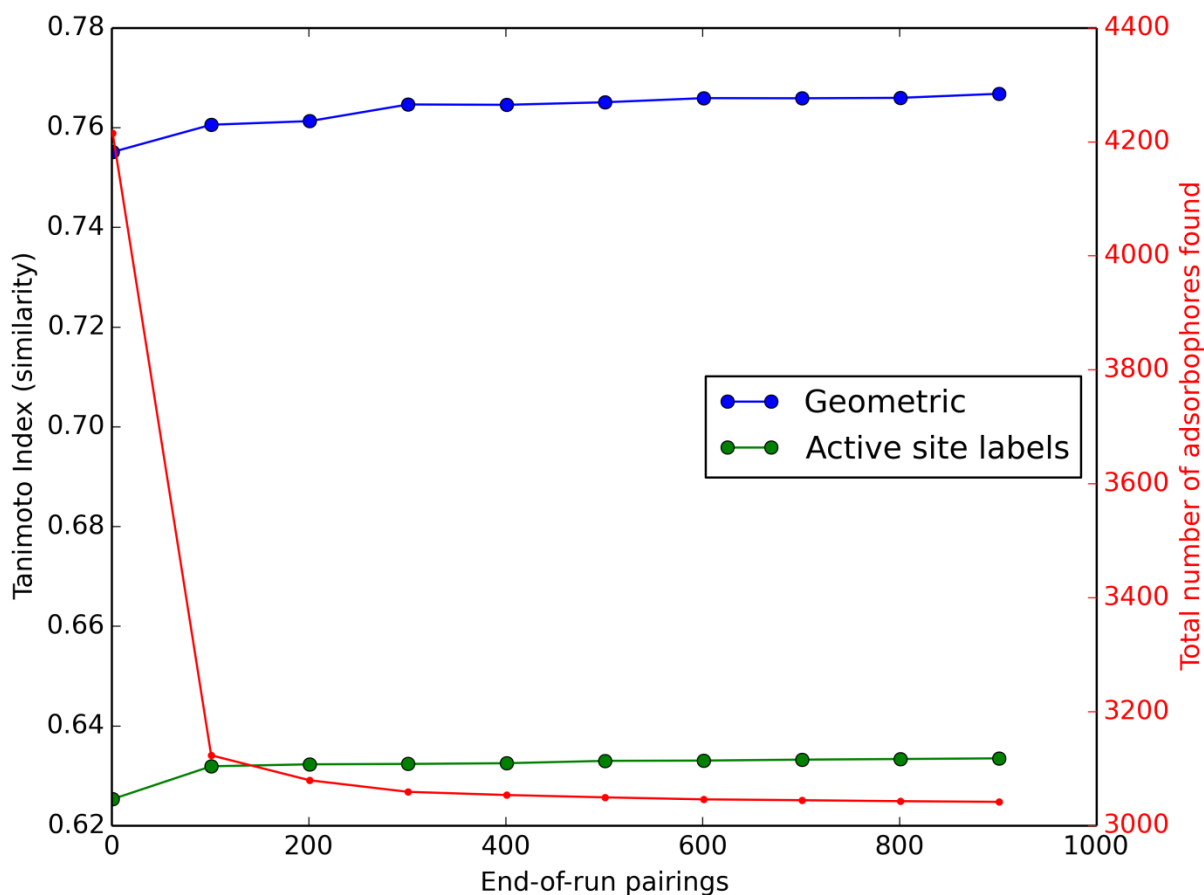


Figure 5.14: Plot of similarity as a function of the program termination parameter (EORP) for 9558 active sites and using a minimum clique threshold of 16. The termination parameter is defined as the number of consecutive random pairings performed in which no new cliques have been found.

We found that upon inspecting the top ranked adsorbophores from each run, the molecules appear to be very similar which agree well with geometric similarities reported in Figure 5.13 and Figure 5.14. From the results presented in this section it is clear that there are very specific binding sites common to many of the high performing MOFs in our hypothetical database. CO₂ has a tendency to bind close to conjugated aromatic systems, as well as near metal-coordinated oxygen species. The random pairing method was able to capture these interesting sites because of their prevalence in the database. However, the random pairing

method is by no means exhaustive. As demonstrated by the lower *label* similarity index (0.6-0.65), the program does not consistently pair the same active sites to yield the same adsorbophores in different runs. This implies that it will not be able to effectively identify interesting, yet less frequent adsorbophores that may nevertheless be good design criteria for MOFs. In addition, as the database diversity grows in terms of topology, SBUs and functional groups, this will likely reduce the efficacy of the random pairing method. It is therefore left for future work to investigate more advanced methods for identifying common CO₂ binding sites.

5.4.2 Analysis of Discovered Adsorbophores

The question then remains as to what are some of the most common collections of atoms which favour CO₂ binding, and do these common features actually provide an energetically strong CO₂ binding pocket, or are they present due to some inherent chemical and geometric similarity between all MOFs in the training set. For this purpose, the top 50 most frequent adsorbophores were visually inspected across multiple random seeds and MCT values to establish common features from a total of 17,710 active sites. These active sites were obtained from the 8325 top performing materials after screening a database of 324,474 hypothetical MOFs for CO₂ capture and storage, the details of which are discussed in chapter 6. To reveal the energetic contributions made by the adsorbophore atoms to CO₂ binding, the interaction energy was determined of CO₂ with just the adsorbophore atoms extracted from each of the original MOFs. These were computed in a force field calculation, which breaks the energy down into pairwise non-bonded coulombic and dispersion interactions, using the original partial atomic charges (calculated with the DFT-derived REPEAT method²¹) and Lennard-Jones parameters assigned to those atoms. Figure 5.15 shows the 3 most common adsorbophores found from the runs inspected with different MCT values. Table 5.2 complements the image with the number of active sites included in the representative adsorbophores. It is clear from the data in Table 5.2

that using a MCT of 6 includes a degree of overlap between the discovered adsorbophores, for example, each set of 6 atoms describing A1 and A2 encompass 105% of the active sites! Thus an MCT value of 6 is likely not capturing the interesting local properties that contribute to CO₂ binding. This is further confirmed by the energy contours plotted in Figure 5.15 a) and d) where the adsorbophore atoms appear to capture at most half of the total average binding energy (~-8.8 kcal/mol). Interestingly, the Cr ions in A3 appear to contribute a large portion of the binding energy as seen by the intense red color indicating ~10 kcal/mol energy in Figure 5.15 g). Looking at the energy distributions for the adsorbophores with MCT=12 (Figure 5.15 b,e,h) it appears that a majority of the average binding energy is captured by these atoms, while still capturing a large portion of the active sites (31%, 32% and 21% for A1, A2, and A3 respectively). As one progresses to MCT=20 the adsorbophores appear to be over-defined; they contain extraneous atoms which, by chemical intuition and location from the region of CO₂ binding, would not contribute significantly to the CO₂ binding site. This is particularly clear for A3 pictured in Figure 5.15 i), which contains hydrogen and carbon atoms extending out from the Cr SBUs. This over-defined nature of the adsorbophores is also exemplified by the significant reduction in the number of active sites which contain the adsorbophores, as demonstrated in the last column of Table 5.2. These results agree with the similarity reductions observed when increasing the MCT from 12-20 and comparing the results across multiple random seeds as discussed above in section 5.4.1.

Table 5.2: number of active sites contained in the adsorbophores A1 – A3 as a function of the MCT parameter. In brackets is the average binding energy of CO₂ in the active sites in kcal/mol.

Adsorbophore	MCT=6	MCT=12	MCT=20
A1	9805 (-8.7)	5409 (-9.1)	1336 (-9.1)
A2	9031 (-8.8)	5595 (-9.2)	2024 (-9.5)
A3	5226 (-9.2)	3739 (-9.4)	1074 (-9.4)

From the results shown in this section we can conclude that the random pairing method is indeed capturing chemically interesting adsorbophores, particularly when using a minimum clique threshold of 12. This appears to be the ‘middle ground’ between poorly defined adsorbophores at lower MCT values which do not contain enough atoms to describe the binding energy of the CO₂ site, and over-defined adsorbophores at higher values of the MCT, which contain extra atoms that reduce the number of active sites it can include.

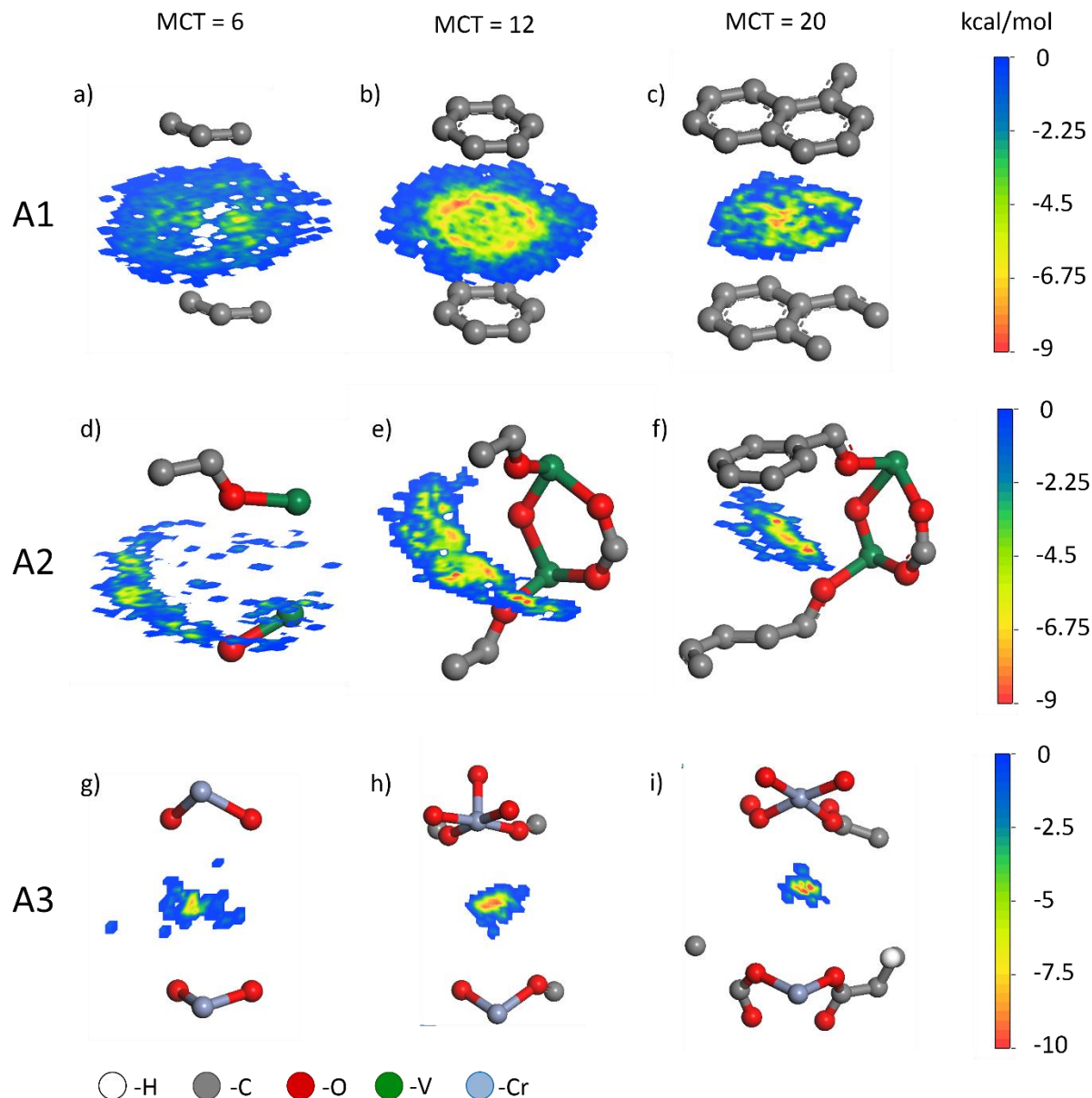


Figure 5.15: A selection of the most representative adsorbophores found from visual inspection of the top 50 most frequent adsorbophores found from the random pairing method. There are three major trends in the molecular fragments which can be seen going down the columns of the image, labelled **A1**, **A2** and **A3**. The resolution of the fragments increases as one goes across each row (MCT is increased). Pictured in each adsorbophore is a representative contour map of the energy produced from CO₂ binding with the adsorbophore atoms from each original CO₂ binding site. a-c) **A1**: planar aromatic systems where CO₂ binds in between, d-f) **A2**: CO₂ binding near the bridging oxygen of a pillared vanadium SBU, g-i) **A3**: CO₂ bound between open-metal Cr SBUs.

5.5 Summary

Presented in this chapter are the programs implemented to identify common CO₂ binding features in MOFs. The first step in this methodology was to identify the active sites in each MOF, or the region surrounding a strong CO₂ binding site. The binding sites were identified with the ABSL program, which identifies maxima from CO₂ probability distributions computed from GCMC simulations. Active sites were identified by extracting the MOF atoms regions surrounding these binding sites. A common substructure search algorithm was implemented, which uses maximum clique detection to identify common atoms between pairs of active sites. This was used to identify frequent substructures (adsorbophores) from a large set of active sites using a method that performs sequential random pairings of common substructures. This method was tested and validated using a set of active sites from our database of hypothetical MOFs. The adsorbophores found with the random pairing method demonstrate a large presence amongst the active sites, and at moderate cutoff values for the maximum clique (no less than 12 atoms), demonstrate a considerable energetic contribution to binding CO₂. However, the results from comparing similarities across multiple runs with different random seeds suggest that the random pairing method may have some limitations. Across 30 different runs, there is a clear disparity between the geometric similarity of the top 50 adsorbophores (~0.85) and the active sites combined to form those adsorbophores (~0.65). This suggests that upon increasing the chemical diversity of the database, a more advanced frequent substructure mining technique should be used. This is left for future work and discussed further in Chapter 8.

5.6 References

1. G. Sliwoski, S. Kothiwale, J. Meiler, and E. W. Lowe, *Pharmacol. Rev.*, 2014, **66**, 334–95.

2. R. Vaidhyanathan, S. S. Iremonger, G. K. H. Shimizu, P. G. Boyd, S. Alavi, and T. K. Woo, *Science*, 2010, **330**, 650–3.
3. P. P. Ewald, *Ann. Phys.*, 1921, **369**, 253–287.
4. A. K. Rappe, C. J. Casewit, K. S. Colwell, W. A. Goddard, and W. M. Skiff, *J. Am. Chem. Soc.*, 1992, **114**, 10024–10035.
5. J. W. Raymond and P. Willett, *J. Comput. Aided. Mol. Des.*, 2002, **16**, 521–533.
6. H. G. Barrow and R. M. Burstall, *Inf. Process. Lett.*, 1976, **4**, 83–84.
7. M. Troyer and U.-J. Wiese, *Phys. Rev. Lett.*, 2005, **94**, 170201.
8. H. A. Land and A. G. Doig, *Econometrica*, 1960, **28**, 497–520.
9. E. Balas and J. Xue, *Algorithmica*, 1996, **15**, 397–412.
10. J. Konc and D. Janezic, *MATCH Commun. Math. Comput. Chem.*, 2007, **58**, 569–590.
11. D. Kumlander, *Commun. Comput. Inf. Sci.*, 2008, **14**, 165–174.
12. S. S. Iremonger, J. Liang, R. Vaidhyanathan, I. Martens, G. K. H. Shimizu, T. D. Daff, M. Z. Aghaji, S. Yeganegi, and T. K. Woo, *J. Am. Chem. Soc.*, 2011, **133**, 20048–51.
13. Y. Podolyan and G. Karypis, *J. Chem. Inf. Model.*, 2009, **49**, 13–21.
14. S. L. Dixon, A. M. Smondyrev, E. H. Knoll, S. N. Rao, D. E. Shaw, and R. a Friesner, *J. Comput. Aided. Mol. Des.*, 2006, **20**, 647–71.
15. O. Dror, D. Schneidman-Duhovny, Y. Inbar, R. Nussinov, and H. J. Wolfson, *J. Chem. Inf. Model.*, 2009, **49**, 2333–43.
16. B. Bringmann, A. Zimmermann, L. De Raedt, and S. Nijssen, in *PKDD*, 2006, vol. 4213, pp. 55–66.
17. N. Wale, I. a. Watson, and G. Karypis, *Knowl. Inf. Syst.*, 2007, **14**, 347–375.
18. P. Willett, *Drug Discov. Today*, 2006, **11**, 1046–53.
19. L. Schietgat, F. Costa, J. Ramon, and L. De Raedt, *Mach. Learn.*, 2010, **83**, 137–161.
20. L. Schietgat, F. Costa, J. Ramon, and L. De Raedt, *In Proc. At SRL-MLG-ILP, Leven, Belgium*, 2009.

21. C. Campañá, B. Mussard, and T. K. Woo, *J. Chem. Theory Comput.*, 2009, **5**, 2866–2878.

6 MOF Design Features for Carbon Capture Technology: a Computational High Throughput Screening Study

6.1 *Abstract*

The interest in developing metal organic frameworks (MOFs) as materials for CO₂ capture has recently lead to a number of computational high throughput screening studies. The goal of these studies is to discover important structure property relationships as a guide for rational design. Previous works have focused primarily on macroscopic physical properties such as specific surface area and void fraction when attempting to correlate structure properties with CO₂ adsorption performance. In this work we show important pore chemistry from screening 325,000 hypothetical MOFs for their CO₂ working capacity and CO₂/N₂ selectivity. Using a program to identify strong CO₂ binding sites from our isotherm simulations and find common MOF features from the surrounding environment of each site, we discover some important trends amongst the top performers. Some of the most common binding environments are one-dimensional channels, where CO₂ intercalates between planar aromatic systems spaced by roughly 7 Å. This was further validated by showing a number of high performing materials possessing one-dimensional sheets of conjugated aromatic systems, large enough to support a monolayer of CO₂.

6.2 *Introduction*

The abundance of coal and our dependency on burning it as a source of energy has made it a primary target for mitigation of CO₂ emissions.¹ The exhaust (flue gas) from coal-fired power plants typically contains 10-15% CO₂, mixed with 75% N₂, ~10% H₂O, as well as small fractions of other gases. The challenge associated with mitigation strategies is to separate and purify CO₂ from the flue mixture while keeping costs to a minimum. Fortunately gas separation

technologies from other industrial applications can be adopted for this purpose.² One such technology, involving selective physical adsorption of CO₂, is being adopted in a number of pilot-scale low carbon emission power-plants.^{3–7} In this process, CO₂ is adsorbed in a solid porous material which is then regenerated by increasing the temperature or decreasing pressure in what are called temperature or pressure swing adsorption (TSA/PSA) systems.

While this process has demonstrated ability in purifying CO₂ from the flue gas, the total costs of these CO₂ purification systems have been a subject of extensive study. Estimated costs can fluctuate from \$20 – 70 USD/ton CO₂ captured depending on the process conditions and what is included in the economic model.^{8–10} Despite this disagreement, there is consensus that minimizing the amount of material needed to capture CO₂ and maximizing the purity of the recovered CO₂-rich gas will minimize process costs. Thus there is considerable interest in finding the optimal material for use in a TSA/PSA system based on its working capacity, which is the quantity of CO₂ recovered per unit mass of the material, and its selectivity for CO₂ over N₂, which is measured as the molar ratio of CO₂ adsorbed over N₂ normalized by their partial pressures. A previous cost study based on a PSA apparatus suggests that the optimal material should possess a working capacity that is twice as high and CO₂/N₂ selectivity that is three times higher than zeolite 13X, a material that is currently used on an industrial scale to scrub CO₂ from natural gas using PSA technology.¹⁰

Recognizing the potential of tuning MOFs for this purpose, several computational high-throughput screening studies on nano-porous materials have recently been realized. The goal of these high-throughput screening studies is two-fold, to identify materials which show potential for the specified application, as well as to identify correlations between material properties and activity. Haldoupis *et al.*¹¹ computationally screened 500 experimentally realized MOFs from the

Cambridge Structure Database (CSD)¹² for their performance in CO₂ capture. They were able to identify two hits from their screening, which can be targets for experimental validation and testing. Missing from their analysis was the chemical reasoning for the high performance of these structures, which could be used to guide the synthesis of improved materials. An alternative approach is to generate hypothetical databases of materials, which can cover a wider range of possible structures giving more robust trends between structure and sorption performance, at the expense of reducing the synthetic viability of the structures. These studies have proven fruitful in the past. Consider the work of Wilmer *et al.*¹³ who pioneered the first screening study on a database of 138,000 hypothetical MOFs for methane storage. In that work they identified a material from their database which performed well, which, unbeknownst to them at the time, had been previously synthesized (NOTT-107¹⁴). Despite this, their discovery demonstrated a synthesizable material with high performance in methane storage, validating that hypothetical MOF screening could provide targets for synthesis. In a follow-up study¹⁵, the same database was screened for CO₂ capture applications (including post-combustion capture and storage). While no MOFs were synthesized in this work, their study provided ‘bounds’ of MOF physical properties (such as surface area and void fraction) which could yield high performance. While instructive, the results do not provide insight at the molecular level as to what pore environments will govern strong CO₂ sorption. While their study did suggest the use of halogens as important for CO₂ adsorption, this was an artificial conclusion due to the force field parameters used in their study, as was revealed in our work discussed in Chapter 4. The work of Lin *et al.*¹⁶ utilized a database of hypothetical zeolite materials and zeolitic imidazolate frameworks (ZIF)s to screen for CO₂ separation from post-combustion exhaust of coal power plants. They were able to identify an important feature not found in experimental zeolites,

namely large 1-dimensional channels ideal for CO₂ diffusion and adsorption. This discovery underlines the importance of sampling a large diverse range of materials, going beyond what has been currently experimentally realised.

Interestingly, there is a surprising lack of chemical insight provided within these studies which, for intelligent MOF design, is a critical issue that must be addressed. Our group recently demonstrated that correlating low pressure CO₂ adsorption to MOF structures based on only their geometric properties results in a poor predictive model¹⁷ further necessitating a need for more detailed chemical analysis. In contrast, numerous high throughput screening studies of potential drug molecules include what is called pharmacophore matching.¹⁸ In these screening programs, molecules which show strong binding to proteins are characterized by 3-dimensional arrangements of, for example, aromatic rings, hydrophilic groups, and hydrogen donors/acceptors. Extensive molecule databases are then mined for these characteristics (pharmacophores) for the purposes of discovering novel drugs with good target activity. In this work we hypothesize that there are common pore shapes local to strong CO₂ sorption sites which can be characterized as the MOF analogue of a pharmacophore, we call an *adsorbophore*. Discovery of adsorbophores from screening a database of MOFs will yield important design criteria for MOFs tailored for implementation in a post-combustion carbon capture apparatus.

In what follows, we describe the screening of 324,474 hypothetical MOFs from our own hypothetical database for CO₂ capture performance. To obtain robust results, care has been taken to produce structures with reasonable atom-atom distances and electrostatic environments within these structures which match, as close as possible, to *ab initio* derived calculations. From our screening, we have identified over 8,000 materials with working capacities and selectivities that surpass zeolite 13X, 63 of which satisfy the target conditions set by Ho *et al.*¹⁰ Further, we

implemented a 3D matching algorithm designed specifically to discover adsorbophores in MOFs. Using this, we analyze the local MOF structure surrounding CO₂ binding sites in our high-performing materials to reveal three important adsorbophores; the most common one being CO₂ sorption sites intercalated between aromatic planes, with perpendicular distances of roughly 7 Å. This is validated by identifying a high frequency (40%) of structures in the top performing MOFs possessing this structural motif.

6.3 Methods

6.3.1 Generation of a hypothetical database of MOFs

The hypothetical MOF structures in our database were constructed with the program discussed in chapter 3. Briefly, MOFs are built using rigid secondary building units (SBU)s, which are snapped together using bond-aligning parameters obtained from existing MOF crystal structures. In this study, the MOFs were assembled using a set of SBUs whose chemistry was found to be consistent with the MEPO QEq parameters developed in chapter 4 such that charged interactions in the MOFs could be computed accurately and quickly. This consisted of 7 inorganic SBUs and 87 organic SBUs, forming 11 topologies. The full list of building units used in this study are provided in Table 6.1, where the SBU and functional group indices can be traced back to the database images provided in chapter 3. It should be noted that the organic SBU indices span from 1-59, however the first 28 of these SBUs were both carboxylate and nitrogen-capped so they could form bonds to open-metal site metal SBUs, yielding the total 87 SBUs reported above. The hypothetical MOFs were built with one inorganic SBU and up to two organic SBUs. MOFs have been experimentally reported with even larger numbers of SBUs,¹⁹ however we felt that limiting the number of SBUs used to assemble each MOF improved the synthetic viability of the structures. Following structure generation, symmetry elements were

found for each structure using a program which computes space group symmetry.²⁰ This program finds symmetry using the method introduced by Grosse-Kunstleve,²¹ where the possible symmetry operations are applied from a lookup table of a specified space group. A MOF is considered symmetric under a specific space group if the symmetry operations successfully map atoms to identical ones in the cell given a specified tolerance. In our work we used a tolerance of 0.4 Å. The asymmetric unit of each MOF was then used as a basis for functionalization. This method was described previously in chapter 3.

Following functionalization, each MOF underwent structural optimization of both the atoms and the lattice. For this purpose, each SBU is detailed with its own connectivity table, which identifies not only which atoms are connected, but also the types of bonds between them. Atoms within SBUs are described by their Universal Force Field (UFF)²² type, so that, together with the bond topologies, each MOF can be fully described by a force field. The atom and bond force-field types for each SBU are automatically generated using the python API (pybel)²³ for openbabel²⁴.

Table 6.1: Full list of SBUs and topologies used to build the structures in this study. The indices correspond to the figures provided in chapter 3.

	Indices from database in chapter 3
Metal SBUs (Figure 3.6)	1, 2, 3, 4, 9, 10, 12
Organic SBUs (Figure 3.7)	1-59
Functional Groups (Figure 3.10)	1-12, 15, 25-27, 32, 43, 49
Topologies	pcu, nbo, tbo, fof, rht, acs, the, pts, sra, frl, pto

6.3.2 Screening protocol

In the post-combustion flue gas separation process, gas flows through a fixed-bed column containing the solid sorbent material, whereupon saturation undergoes a regeneration stage in

which purified CO₂ is desorbed from the material for transportation to a permanent storage facility. Various conditions can be realized at the adsorption and regeneration steps, and the optimal conditions will vary for different materials. For our screening purposes we sampled fixed adsorption and desorption conditions, such that different materials' performances could be directly compared at a given state point. While this may not serve as a direct measure of parasitic load for implementation into a power plant, we can compare some key indicators of relative cost, namely CO₂ selectivity over N₂ and CO₂ working capacity. In this study, we have considered the flue exhaust as having a temperature of 298 K and a total pressure of 1 bar, where the ratio of CO₂ to N₂ is fixed at 0.15:0.85. These conditions have been used in controlled experiments evaluating solid sorbents for performance in post-combustion CO₂ capture.^{4,25-27} Importantly, the thermodynamic conditions are within a feasible range found from several pilot plant studies using various carbon capture technologies.^{28,29} For example, in a pilot plant study using zeolite 13X as the solid sorbent material; the flue gas reaches a temperature of between 293 - 303 K before it reaches the CO₂ adsorber, which is preceded by desulfurization and dehumidification of the flue stream.^{5,30}

For regeneration, a temperature/pressure swing (TPSA) system was simulated using a temperature of 363 K and vacuum pressure of 0.1 bar for desorption. This condition was chosen to agree with a range of suggested desorption conditions from a study on a pilot plant in Yokosuka, Kanagawa, Japan. They demonstrated that a TPSA system could reduce parasitic load by 11% over a pressure swing adsorption (PSA) system alone.³ In addition, these conditions were shown to achieve a high purity of CO₂ and capture efficiency from a laboratory-scale study using zeolite 13X as the sorbent material.³¹

Prior to gas sorption screening, each structure within the hypothetical MOF database underwent structural optimization using a molecular mechanics force field as implemented in the GULP package³². While previous databases of MOFs³³ and porous polymer networks (PPNs)³⁴ have been optimized at the PM6 level, optimizing with a force field poses a subtle advantage over *ab initio* or semi-empirical methods at this stage of screening. This is because the atomic bonding in each MOF is defined *during* the structure assembly stage, with bonding tables described for each SBU. Thus the atoms in the final structure will relax to a low energy state which is governed by intended bonding. If a structure initially is quite strained with close-atoms, computing bonds after the structure is assembled, or using higher-level calculations to relax the ionic and electronic system will likely yield undesirable side-effects due to high initial forces and artificial bond formation. Both atoms and cell shape were relaxed at this step, using a conjugate gradient algorithm until the root mean square gradient per atom reached below 0.3 eV/Å, after which a rational function optimization³⁵ was used to converge the norm of the forces to below 0.001 eV/Å. In cases where convergence criteria were not met, the structure was not carried forward to the initial gas adsorption screening step.

As an initial screening step, the structures were subjected to a GCMC calculation of a binary mixture of CO₂ and N₂ consisting of 2000 equilibration cycles and 2000 production cycles. A single cycle consists of N Monte Carlo steps, where N is the number of guest molecules present in the system. The top structures whose adsorption of CO₂ was higher than 2 mmol/g of material obtained from this screening were then subjected to a more refined calculation where charges on the framework atoms were computed using the REPEAT methodology.³⁶ The REPEAT method solves a linear problem of fitting a set of atomic point charges to the framework atoms to match the electrostatic potential from a single-point DFT

calculation. Both adsorption and desorption conditions were calculated at this new step, using 2×10^5 equilibration and 10^7 production steps.

Finally, a selection of top performing materials with working capacities above 4 mmol/g underwent geometry optimization at the DFT level, the adsorption and desorption conditions were re-evaluated using 5×10^5 equilibration and 10^7 production steps.

6.3.3 Adsorbophore discovery

We have previously introduced the concept of an adsorbophore in chapter 5 as a molecular fragment from a MOF which shows activity with respect to strong adsorption of a CO₂ molecule. To discover adsorbophores in this study, we employ the tools developed in that chapter which extract regions of strong CO₂ sorption from the MOFs, and compare them to find common structural features. Specifically, adsorbophores are found using the random pairing method to pair MOF fragments and maximum clique detection method³⁷ to determine their common 3D substructure. The minimum clique threshold (MCT), defined as the smallest allowable common substructure, was set to 14 atoms in this study. This was for the purpose of identifying adsorbophores with enough atoms to extract some chemical meaning, while at the same time not excluding important common substructures. To ensure an end point of the search, the program stopped randomly pairing fragments after 200 consecutive pairings failed to find a single new substructure. The work flow associated with adsorbophore discovery is detailed below:

- From a GCMC simulation of CO₂ at 298K and 0.15 bar partial pressure, the CO₂ positions from every post-equilibration Monte Carlo step was stored on a 3D grid with a resolution of $0.1 \times 0.1 \times 0.1 \text{ \AA}^3$. This histogram was obtained separately for both the C atom and O atoms of CO₂. This gives the probability regions of CO₂ sorption from a sampling of the free energy of the system.

- Representative CO₂ binding sites in each MOF were found using a program which filters the high frequency 'noise' from the 3D histograms and isolates maxima. Molecules of CO₂ are then identified by evaluating the proximity and geometry of C and O atoms found from the maxima. These molecules represent CO₂ *binding sites*. The interaction energy of this CO₂ with the framework is then computed using the force field parameters described in the computational details section.
- A radius of 5 Å surrounding the CO₂ atoms of each binding site was used to collect MOF framework atoms. The framework atoms found in this radius were considered the local region of the framework which supports CO₂ adsorption, henceforth called *active sites*.
- Starting with 106,680 active sites from the top 8325 MOFs with computed CO₂ uptakes above 2 mmol/g, frequent adsorbophores were discovered by performing repeated random pairings of molecular fragments, where at each pairing, common 3D structures were discovered using the maximum clique algorithm. Upon completion, the common substructures, or *adsorbophores*, were ranked based on how many original binding sites it encompassed.
- Because the random pairing method does not scan every possible common substructure between the active sites, the most common adsorbophores found from this method were then searched for in the original 106,680 active sites using the same clique detection algorithm mentioned above. This way, for each adsorbophore we identified the total number of active sites that possessed it as a substructure.

6.4 Computational Details

Gas uptake properties were found using a self-authored GCMC algorithm which treats the framework as fixed and guests rigid. All energetic interactions between atoms are computed

using a Lennard-Jones potential for dispersion and steric repulsion terms and the fixed partial charge approximation for electrostatic interactions. A 12.5 Å cutoff was used for the long-range interactions in the hypothetical materials, and a supercell was created if the cut-off was larger than half the width of the smallest orthogonal distance between repeat units. An isotropic long-range correction term was applied to the Lennard-Jones function to ensure convergence of the energy term to zero as the distance between two atoms approaches infinity. Charges were assigned to the atoms in the MOF framework using the charge equilibration method developed by Rappe and Goddard³⁸, using parameters designed specifically to re-produce the electrostatic potentials derived from an *ab-initio* calculation on a diverse training set of MOFs from the hypothetical database. The details of these parameters were reported previously in chapter 4. The Lennard-Jones parameters assigned to framework atoms were taken from the Universal Force Field.²² Lorentz-Berthelot mixing rules applied to heterogeneous atom interactions.

The CO₂ guest molecule was modelled using parameters obtained from García-Sánchez *et al.*³⁹ which were fit to experimental adsorption isotherm data of zeolites. The N₂ parameters were obtained from the TraPPE model.⁴⁰ Each phase of the screening encompassed a different set of Monte Carlo equilibration and production steps, which will be described below.

Several of the high performing MOFs were subjected to higher-level calculations to confirm properties reported from the initial screening. In this study the DFT calculations were carried out in the Vienna Ab Initio Simulation Package^{41,42} (VASP) using the projector augmented wave (PAW) method^{43,44} and the PBE functional^{45,46} with a planewave cut-off of 500 eV. In all cases the gamma point was used to sample the Brillouin zone. In cases where the lattice parameters and atoms were allowed to relax, a conjugate gradient algorithm was used. The lattice configuration was considered converged once all forces were smaller than 2×10^{-4} eV/Å.

Dispersion corrections of Grimme (D2) were used to include the attractive van der Waals contributions to the total energy.

To investigate the DFT binding energy and the possibility of charge transfer between CO₂ and the adsorbophores, representative MOFs possessing the adsorbophore were selected and DFT calculations were performed where the MOF framework was held fixed and the CO₂ atoms were relaxed to their energy minima. A single point energy calculation of the CO₂ atoms were then computed in an empty box of the same dimensions as the MOF unit cell, and separately, a single point calculation of the MOF less the CO₂ atoms was computed. Binding site energies were then calculated using the following equation,

$$E_{binding\ site} = E_{CO_2+MOF} - E_{MOF} - E_{CO_2} \quad 6.1$$

The charge density difference of the MOF+CO₂ was calculated by subtracting the charge densities of the isolated MOF and CO₂. Bader charge decomposition was computed for each MOF with and without the CO₂ using the program distributed by the Henkelman group,⁴⁷ which partitions the charge density of a VASP CHGCAR onto particular atoms based on zero flux surfaces between atoms.

6.5 Results

6.5.1 CO₂ screening

The results from the screening are shown in Figure 6.1 plotted as CO₂ working capacity vs CO₂/N₂ selectivity for the database of close to 325,000 MOFs. One interesting feature of this graph is the density of MOFs localized in a small region of working capacities less than 0.5 mmol/g and selectivity less than 20. With such a highly localized region of performance within the database, the question then arises as to how this is related to physical diversity within these MOFs. Interestingly the surface areas and void fractions of the materials in this region of high

concentration are surprisingly diverse. They possess a large distribution from 195 to 8085 m²/g in surface area and 0.01 to 9.26 cm³/g in void volume. The distribution of these physical properties for the narrow low range of performance of 0-0.5 mmol/g working capacity and 0-20 selectivity is shown in Figure 6.2. While there appears to be a narrow streak of low void volumes in this distribution, this is an artefact of the physical property distribution of the entire set of MOFs screened, which is qualitatively very similar to this distribution. Thus we can say that while the MOF database is skewed towards these void volumes, the presence of low performing MOFs with void volumes outside of this narrow range suggests that it cannot be considered a design factor for CO₂ capture performance. It is clear from this study that there is no straightforward relationship between the materials geometric properties and CO₂ sorption performance. This confirms what was recently discussed elsewhere.^{15,17}

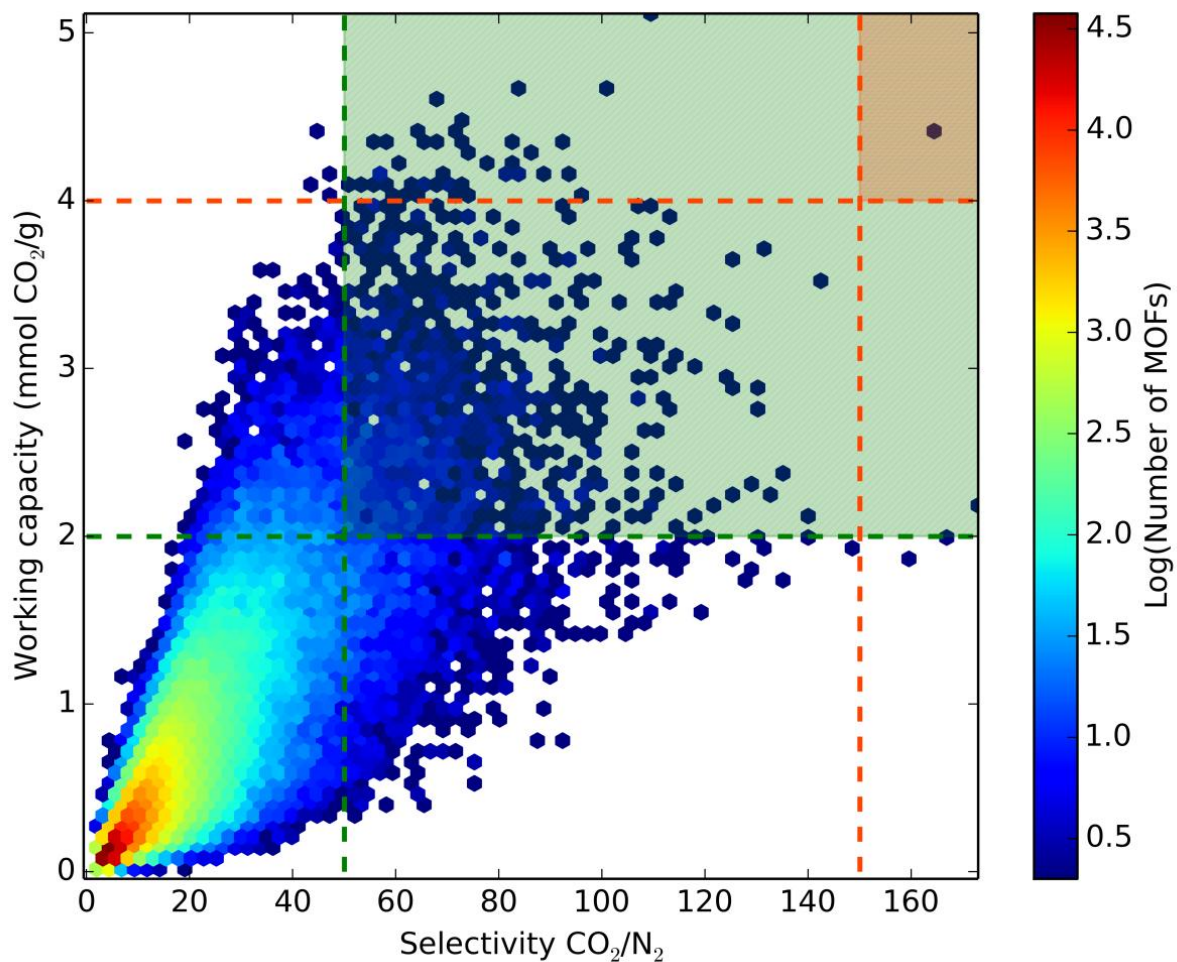


Figure 6.1: Color map of the 324,474 MOFs screened for working capacity and CO₂/N₂ selectivity with GCMC simulations. The green zone represents the region of MOFs which possess working capacities and selectivities above zeolite 13X, the region in red represents materials which would significantly reduce the cost of CO₂ capture from flue streams according to ref 10.

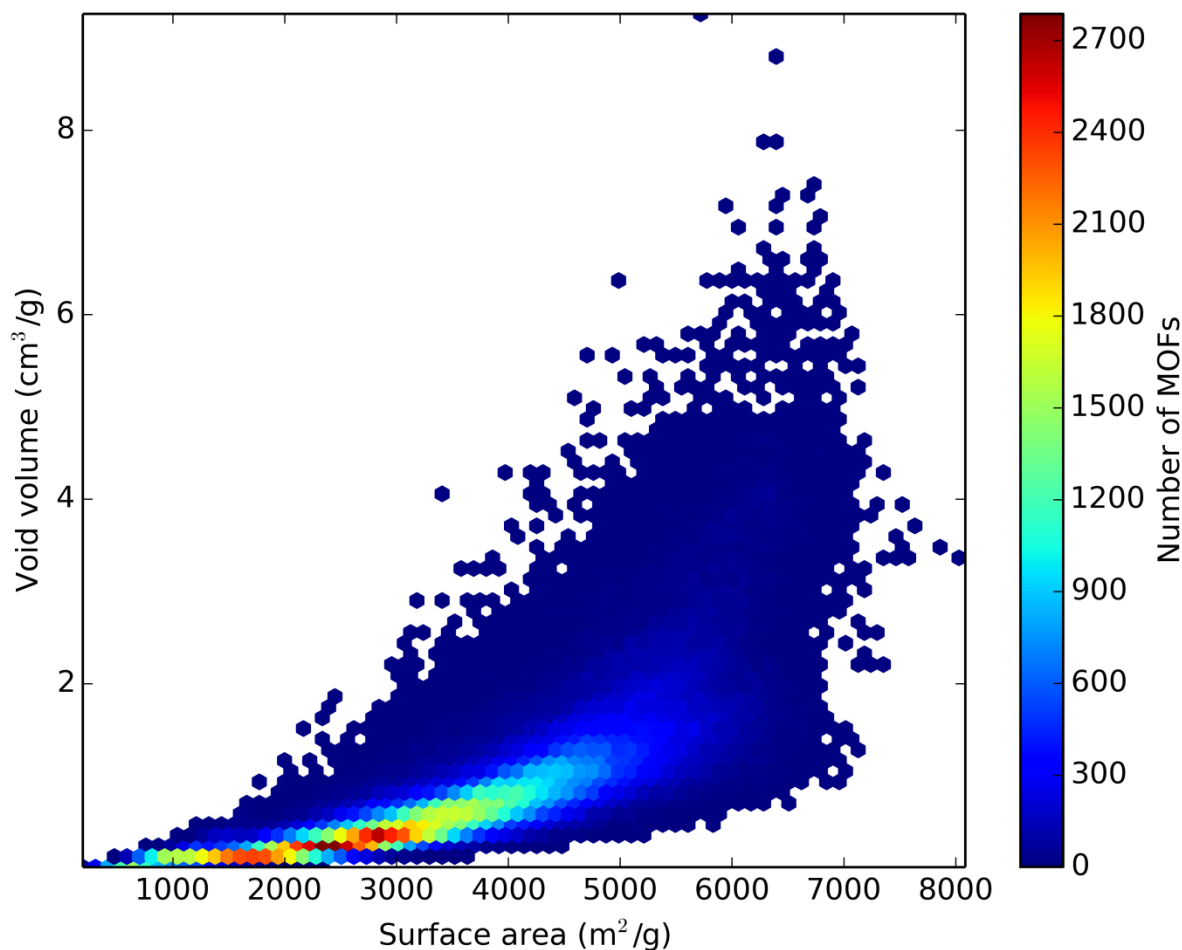


Figure 6.2: The distribution of void volumes and surface areas of the low performing MOFs in the range of 0-0.5 mmol/g CO₂ working capacity and 0-20 CO₂/N₂ selectivity.

Of interest is to elucidate important chemistry related to strong CO₂ adsorption. One of the purposes of decorating base hypothetical MOF structures with functional groups was to identify a possible connection between functionalization and performance. What we see from Figure 6.3 is that, in fact no functional group stands out as a clear winner. Moreover, it appears that unfunctionalized MOFs appear in higher frequencies amongst the top performers than does any MOF decorated with a functional group. This is in contradiction with the study of Wilmer *et al.* who found that F and Cl functional groups can enhance the CO₂ capture performance.¹⁵ We discuss two possible reasons for this. First, we have screened at different desorption conditions,

using a TPSA system instead of the PSA system proposed in their study. It is known that each material will have an optimal set of desorption conditions, whose relationship is not linear with the temperature and pressure. So ranking these structures will be slightly different when comparing them at different sorption conditions. Second, the strength of CO₂ adsorption is governed by the way we compute non-bonded interactions. For CO₂ a significant portion of the binding energy is governed by the electrostatic potential produced by the MOF. This is dependent on the atomic charges assigned to each atom, which in our study was generated by the QEq method with specific parameters tailored for MOF atoms (MEPO).⁴⁸ It is therefore possible that the differing electrostatic environments generated by our model would cause this discrepancy. Moreover, our previous work⁴⁸ highlights a significant issue with the halogen parameters used in the study by Wilmer *et al.*¹⁵ It was shown that the partial charges assigned to halogen atoms were typically much larger than those found from DFT derived charges, yielding over-estimations in CO₂ adsorption. To ensure that our high performing materials were accurately modelled, we compare the CO₂ adsorption using the charges derived from MEPO parameters with that of the *ab initio* derived REPEAT charges in Figure 6.4, where it is demonstrated that the MEPO QEq parameters will, on average, underestimate CO₂ adsorption by approximately 1 mmol/g. This demonstrates that we are conservatively underestimating performance in our initial screening, and further validates the functional group performance reported in Figure 6.3.

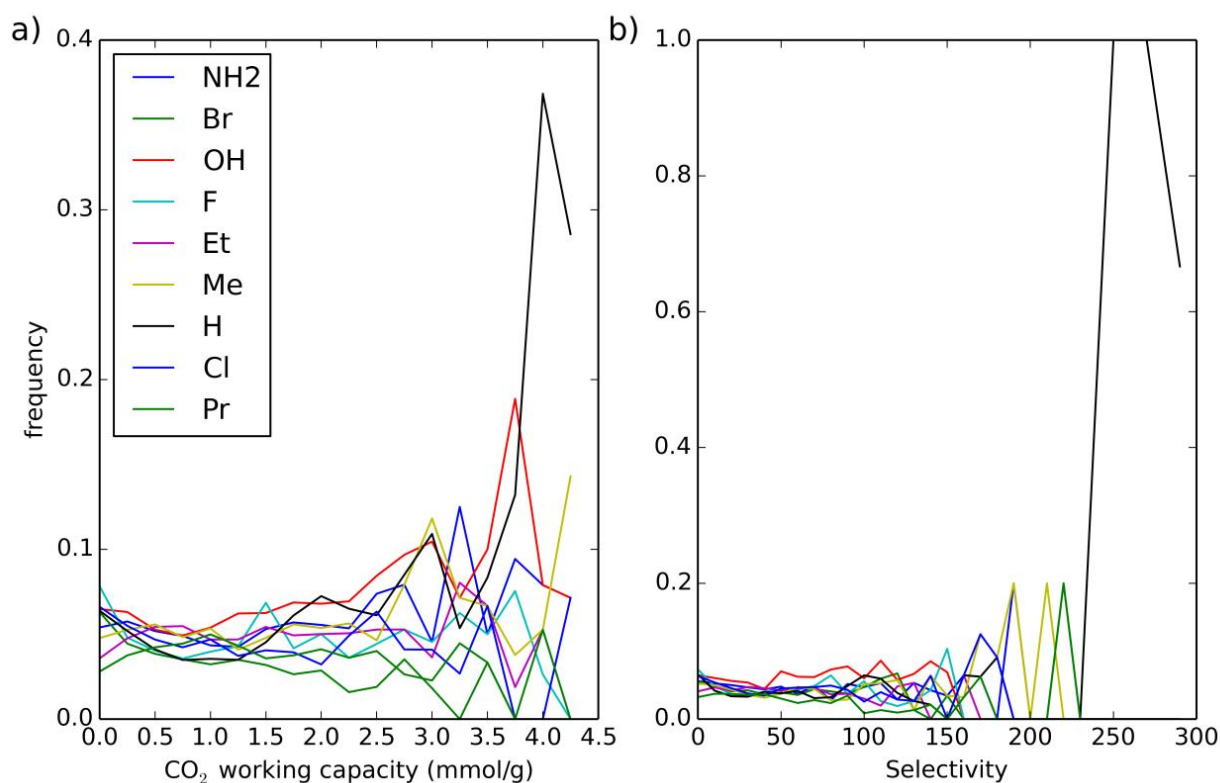


Figure 6.3: Frequency of a selection of functional groups present in the database as a function of performance metrics. a) working capacity, b) selectivity.

From these results we have taken the top 8325 MOFs within our database to perform detailed calculations with. These MOFs have CO₂ adsorption greater than 2 mmol/g at 298K and 0.15 bar pressure, which surpasses that of zeolite 13X.⁴⁹

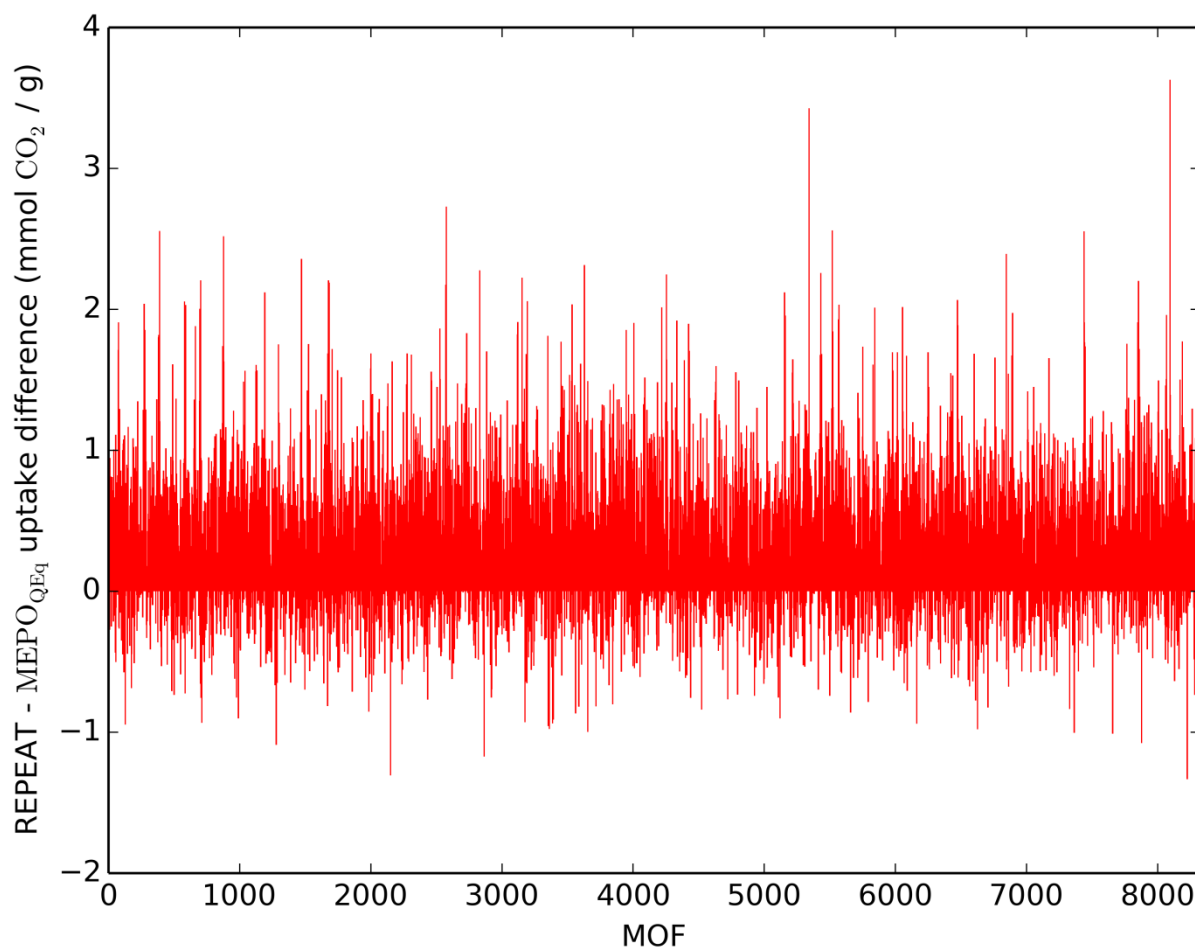


Figure 6.4: The difference between CO₂ adsorption when using DFT-derived charges and the MEPO QEq parameters. Positive values indicate the MEPO charges are under-estimating CO₂ adsorption. CO₂ uptakes were evaluated at 298K and 0.15 bar partial pressure.

6.5.2 Strong CO₂ adsorption sites

This section discusses the discovery of important binding regions (adsorbophores) of CO₂ in the top 8325 MOFs in our database, shown in Figure 6.5. These adsorbophores were discovered using the substructure search algorithm discussed in the methods section, which matches the atoms of two sites in two different MOFs based on relative atom distances. The tolerance used for matching atoms was set to 0.4 Å. This was to accommodate the possible structural deviations resulting from optimizing the MOF geometries, yet small enough to distinguish between fragments with different bonding. As such, the adsorbophores shown in

Figure 6.5 are a single representation of the atoms from many active sites in MOFs. To determine how much the atomic positions deviated between active sites, as well as to create the CO₂ binding site distributions shown in the figure, each active site containing the adsorbophore, and its CO₂ binding site, was oriented to minimize the root mean squared deviation (RMSD) with the representative adsorbophore. The average RMSD atoms, along with the average total binding energy of CO₂ in the pocket described by the adsorbophore, and the total number of high performing MOFs containing the adsorbophore are reported in Table 6.2.

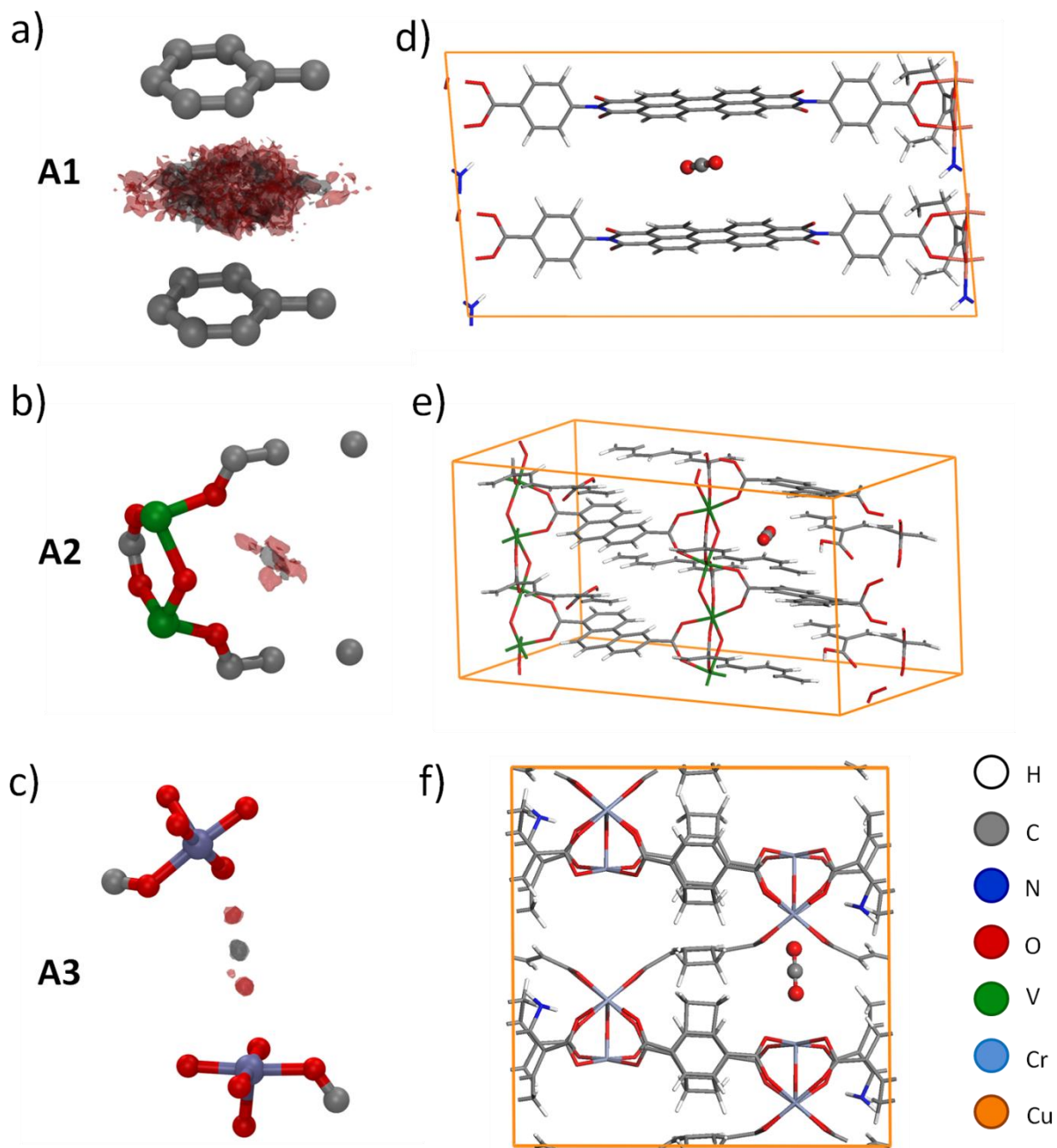


Figure 6.5: a-c) The top adsorbophores ranked by the number of MOFs which contain them. Transparent grey and red clouds represent the distributions of carbon and oxygen atoms from the CO₂ binding sites in the original MOFs containing the adsorbophore. d-f) A CO₂ binding site is highlighted for each adsorbophore. To highlight the binding site, the CO₂ is represented as a ball-and-stick model, while the MOF framework is represented as sticks. d) A representative MOF from the hypothetical database containing the **A1** adsorbophore. e) a MOF containing the **A2** adsorbophore, f) a MOF containing the **A3** adsorbophore.

Figure 6.5a shows the adsorbophore (**A1**) which encompasses the most diverse range of MOFs in our database. It describes two parallel aromatic rings with spacings of approximately 7 Å. While the distribution of CO₂ binding sites is sparse (as can be seen by the breadth of the CO₂ probability clouds in between the rings) the average binding site energy of a CO₂ molecule in this cloud was 8.9 kcal/mol, indicating a strong yet delocalized binding location for CO₂. Figure 6.5b represents a collection of CO₂ binding sites local to V^{IV} inorganic SBUs (**A2**). All of the hypothetical MOFs with this binding site consist of the rod like V^{IV} SBU bonded to linear 2-connected (2-c) organic SBUs, yielding an underlying connectivity described by the **sra** topology. The CO₂ distribution surrounds the bridging oxygen between V^{IV} ions. The average binding energy reported for this CO₂ distribution is 9.2 kcal/mol where 18% of this energy is electrostatic. This electrostatic contribution is likely due to the charge interaction of the partially positive carbon of CO₂ with the bridging oxygens. The lone carbon atoms both above and below the region of CO₂ binding correspond to organic SBUs extending outward from the vanadium pillar. Figure 6.5e demonstrates this configuration in a hypothetical MOF containing the **A2** adsorbophore, where above and below the representative binding site are aromatic carbon SBUs. The final adsorbophore (**A3**) shown in Figure 6.5c also describes binding near a specific inorganic SBU. In this case a highly localized and oriented distribution of CO₂ binding sites is found between two Cr SBUs, where clearly, the negative poles on the quadrupole moment of CO₂ are oriented close to two Cr ions. It is clear that this quadrupolar interaction is a huge contributing factor of the average binding site energy of 10 kcal/mol, as 48% of this energy is electrostatic. The SBU associated with the Cr ions is 6-c and is connected to 2-c linear organic SBUs to form periodic structures with the **acs** topology. Interestingly, due to the nature of this topology, the adsorbophore was only found in MOFs where the organic SBUs had lengths in the

range of 4.1 – 4.3 Å. This permitted the spacing between Cr SBUs to be the correct length to fit a CO₂ lengthwise between them. We note that the open metal sites on the Cr ions are likely going to be occupied by solvent or water molecules in experimental studies. However, it does not invalidate this as an interesting site of adsorption for CO₂, and is not the first instance where MOFs with unsaturated metal ions were shown as a site for strong CO₂ binding.⁵⁰

Table 6.2: Data for the adsorbophores shown in Figure 6.5.

Adsorbophore	RMSD (Å)	Average Total Binding Energy [†] (kcal/mol)	Number of high performing MOFs
A1	0.75	8.9 ± 1.0	2249 (27%)
A2	0.69	9.2 ± 0.9	2077 (25%)
A3	0.85	10.2 ± 0.9	777 (10%)

[†]The total binding energy includes the interaction of the CO₂ binding site with the entire MOF, not just the local environment described by the adsorbophore.

Because these adsorbophores characterize a common *local* environment around strong CO₂ sorption sites, it is important to stress that the extended region also contributes to the CO₂ - MOF interaction energy. To examine the importance of the local region to the CO₂ binding strength for each adsorbophore, we computed their isolated interaction energy with CO₂. These were performed as single point energy calculations using the same force field parameters from the GCMC simulations. The adsorbophore atom distributions, their charges, and the CO₂ binding sites were extracted from each original MOF to determine how much of the energy is contributed solely by the atoms of the adsorbophore. A qualitative description of the binding energy contributions from these adsorbophores are shown in Figure 6.6, where a 2-dimensional contour plot is shown in the region of the most populated CO₂ distribution (and binding energy). It is shown in Figure 6.6a that the energy contributions of the atoms in **A1** can contribute close to 96% of the average binding site energy, particularly when the CO₂ binding site is localized

directly between the aromatic rings. In adsorbophore **A2** the strongest energetic contribution from this fragment is when CO₂ is near the vanadium bridging oxygen and directly between the two extending arms of carbon which can be seen in red in the contour plot in Figure 6.6b. Here the ‘hottest’ region shows the adsorbophore contributes roughly 88% of the total binding site energy. The distributions of CO₂ binding sites outside of this red region will have binding energy contributions from organic SBUs which extend from this adsorbophore. These may complement the **A1** site, since the carbon-to-carbon distance in the **A2** representation is 7.2 Å, a similar distance between aromatic planes in the **A1** site. Nevertheless, these atoms are not captured in **A2** because their chemical diversity prevented them from being found as part of the common substructure. Adsorbophore **A3** contains a highly localized distribution of CO₂, and the energetic contribution from this fragment completely describes the binding energy found in the MOF.

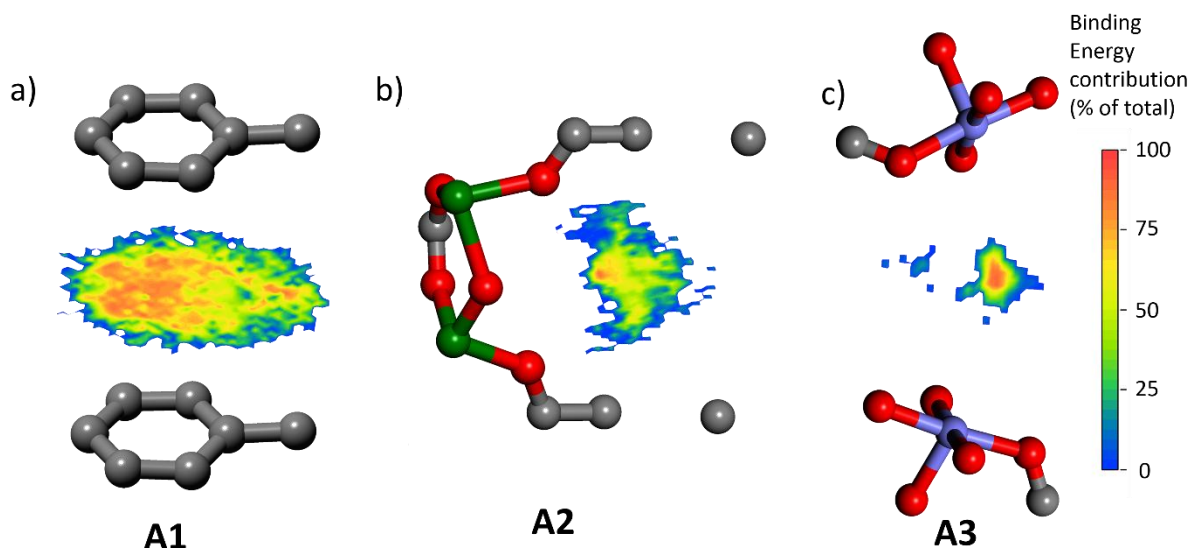


Figure 6.6: A two dimensional slice representation of the force-field calculated energy contributions of all the CO₂ binding sites in the three adsorbophores discussed in the text. The location of the slices were chosen to encompass a large portion of the CO₂ binding sites, and to capture the regions of strongest CO₂ interaction. The heat map shows the percent contribution of the adsorbophore atoms to the total CO₂ binding site energy calculated in each MOF.

Adsorbophore **A3** represents CO₂ adsorption close to open metal sites, where there is a possibility of polarization or charge transfer from the CO₂ oxygens to the Cr ions. Our classical simulations cannot characterize this type of interaction, it is therefore worthwhile investigating the electronic effects between the CO₂ and Cr ions in the **A3** adsorbophore with *ab initio* methods. For this purpose, we performed periodic DFT calculations on a representative MOF containing the **A3** binding site (shown in Figure 6.5f). After decomposing the charge density to each atom based on the method of Bader,⁵¹ we observe a very small charge difference (0.03 e more negative) on the Cr ion 2.6 Å from the CO₂ oxygen. In comparison, the Bader charges for the other two adsorbophores, which are not near any coordinatively unsaturated metal sites and therefore will have no charge transfer interactions, differed by a maximum of -0.008 e for **A1** and -0.010 e for **A2** respectively. We plot the *ab initio* charge density difference for the **A3** site in Figure 6.7, which demonstrates that CO₂ binding in this site induces a polarization of the electron density in both the CO₂ and the MOF. Here it is shown that the electron density on CO₂ rearranges to make the carbon more electron rich, and the closest CO₂ oxygen to the open metal Cr site becomes slightly electron poor, while the Cr ion gains electron density. Electron polarization interactions represent a relatively weak energy (compared with charge transfer), thus much of the interaction strength of **A3** should be captured by the force field calculations. This is indeed the case, as the difference between the force field and DFT binding energies for **A3** is only 0.65 kcal/mol.

The DFT binding energies of the representative adsorbophores were calculated to be -7.6, -8.8, and -8.85 kcal/mol for **A1**, **A2**, and **A3**, respectively. These are in good agreement with the force field binding site energies which were -8.8 kcal/mol for **A1**, -9.8 kcal/mol for **A2**, and -9.5 kcal/mol for **A3**. As stated above, it is likely that water will competitively bind to the open metal

Cr sites in **A3**. To confirm this, CO₂ was replaced with H₂O in the **A3** binding site and optimized to its lowest energy state, keeping the framework fixed. The resulting binding energy of H₂O was -18.9 kcal/mol, more than twice as strong as the binding energy calculated for CO₂, confirming that the presence of water in the system would likely eliminate this adsorbophore as a potentially strong CO₂ binding site.

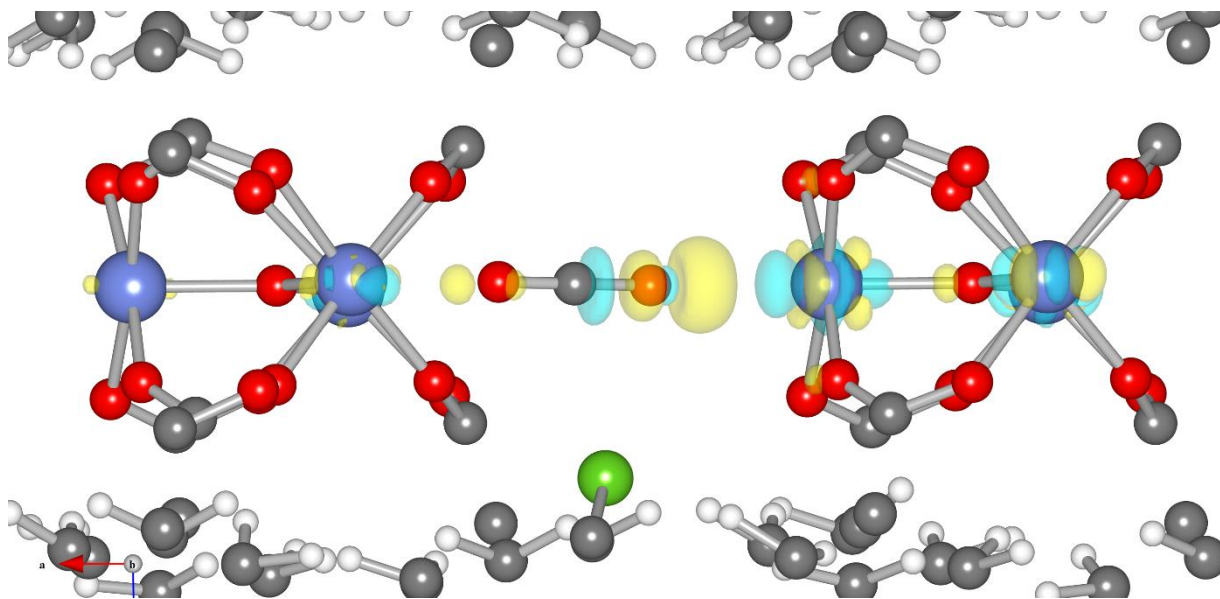


Figure 6.7: Charge density difference plot of CO₂ (centre) bound in the **A3** adsorbophore in the hypothetical MOF found in Figure 6.5f. The isosurface value is 0.01 e/Å³ positive differences are shown in yellow while negative differences are shown in blue. The bonds between the Cr SBUs and the rest of the framework were removed for clarity.

6.5.3 MOF Pores Lined with Conjugated Aromatic Systems

The top adsorbophore in Figure 6.5a (**A1**) represents an interesting site for further exploration. We identified the MOFs which contained this common binding site and discovered a striking common feature amongst these materials; the presence of one dimensional channels of *layered sheets* consisting of poly-aromatic molecules separated by a perpendicular distance of roughly 7 Å from aromatic plane to aromatic plane. As shown in Figure 6.9, this feature is seen across different topologies and inorganic SBUs within the database, suggesting that the important factor in the performance of these MOFs is the arrangement of these aromatic organic

SBU. This distance between two aromatic carbon sheets is ideal for adsorption of a monolayer of CO₂. To the author's knowledge, no such claim has been previously made from a study regarding CO₂ sorption in MOFs.

Some questions immediately arise from this result. The most glaring one is how many of these structures are actually not high performing? In other words, have we identified a false-positive in relating these structures to high CO₂ capture performance? Other questions also arise; would we be able to identify these structures based on physical properties alone, or does screening MOFs for CO₂ capture necessitate a closer look at the chemical nature of these structures? To answer the former question we looked at the distribution of MOFs assembled using metal and organic SBUs which had the potential to provide the ideal spacing and chemistry to support this layered sheet configuration. We identified a total of 28,292 structures containing these SBUs which is 8.7% of the total database. Of these, 2,854 are flagged as 'high performing' materials which have an uptake greater than 2 mmol/g, which constitutes 40% of all materials above this threshold. The remaining 25,438 were 'low performing' (8% of all low performers). They are thus over-represented in the high performing materials, but one cannot ignore the large quantity of these materials under the performance threshold. Recall that there is a measure of structural redundancy in the MOF database, in that a set of base structures have been decorated with functional groups at symmetry equivalent hydrogen positions. Upon functionalization, these MOFs then undergo a geometry optimization step which will ultimately yield pores with different shapes and sizes, thus two MOFs built with the same SBUs, but with different functional groups can yield very different pore structures and can give rise to significant performance differences. Another important factor when comparing across this set of MOFs is that functionalizing these materials will change their unit weight, skewing the *gravimetric* uptake

data towards preference for lighter MOFs. While gravimetric uptake may ultimately be important design criteria for implementation into an adsorption column, comparing these MOFs based on their *pore volumetric* uptake, which specifies the amount of CO₂ adsorbed per unit volume of available pore space, will be more consistent with the claim that these layered aromatic sheets are important for performance. Figure 6.8a shows the pore volumetric working capacity of CO₂ as a function of the maximum free pore diameter. A large portion of these materials pack CO₂ quite densely, particularly in the pore diameter range of 4-5 Å. Assuming a gaseous CO₂ concentration of 0.045 mmol/cm³ at 298K and 1 bar, 5145 of these MOFs recover CO₂ at a density 200 times greater than this value, 80 of these materials recover CO₂ at a density surpassing the solid phase of CO₂ (~35.5 mmol/cm³).

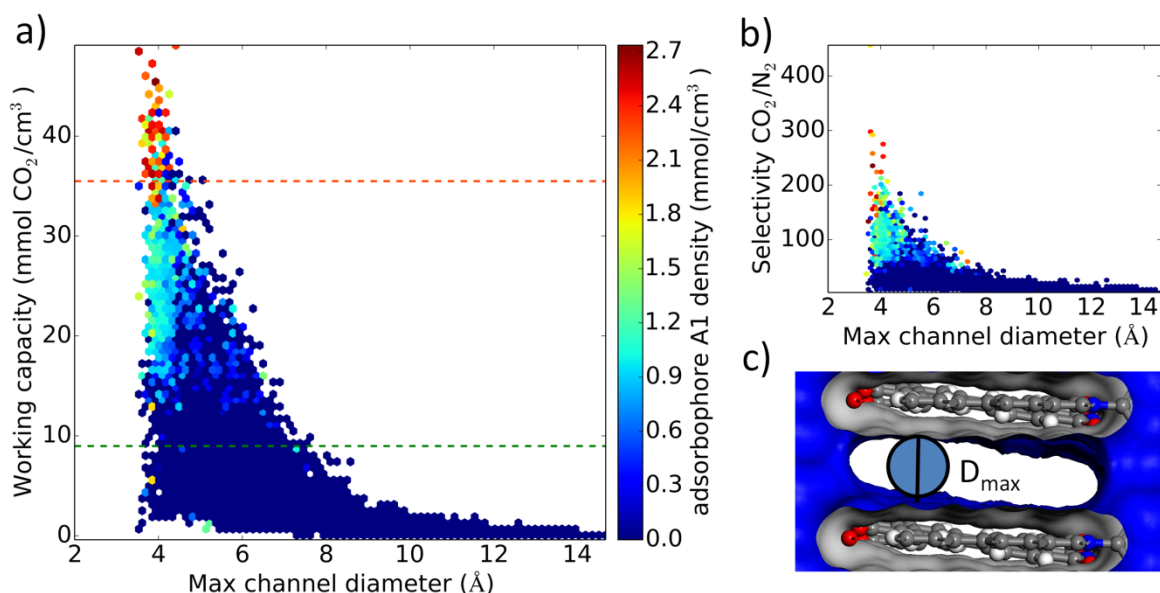


Figure 6.8: Volumetric performance of the 28,292 hypothetical MOFs identified as possible layered sheet-like MOFs. a) Working capacity of CO₂ in units of mmols per unit free volume of MOF vs. the maximum free pore diameter. The green dashed line represents CO₂ density 200 times greater than gaseous CO₂ at 298K and 1 bar. The orange dashed line represents the density of solid CO₂ at 1 bar and 194.6 K b) selectivity of these MOFs, c) graphical representation of how the x-axis is measured in the plots a) and b). D_{max} = 4 Å for the pore described in the picture. The color bar quantifies the number of **A1** adsorbophores found in each MOF per unit volume. A detailed description of how this is computed is discussed in the text.

The maximum pore diameter calculation measures the maximum spherical diameter that can be included between the van der Waals radii of the MOF atoms. Thus we can associate the peak in Figure 6.8a at 4 Å with a spacing of 7.4 Å between aromatic planes, as the van der Waals radii for aromatic carbons was set to 1.7 Å. This result agrees well with the spacings in **A1** in Figure 6.5a, however, the sizeable distribution of MOFs with poor performance at this pore diameter indicates that it is a necessary but not sufficient condition for strong CO₂ adsorption. We note that the density of **A1** found within these MOFs correlates extremely well with the working capacity and selectivity (shown in Figure 6.8a and b respectively). Instances of adsorbophore **A1** were found within each MOF using the maximum clique detection algorithm discussed in chapter 5. The number of adsorbophores found in each MOF were converted to a molar value and normalized by the volume of the unit cell to yield the z axis values for each plots in Figure 6.8. Figure 6.9 shows two of these high performing materials and their gravimetric working capacities. These structures possess the roughly 7 Å spacing between planar conjugated SBUs.

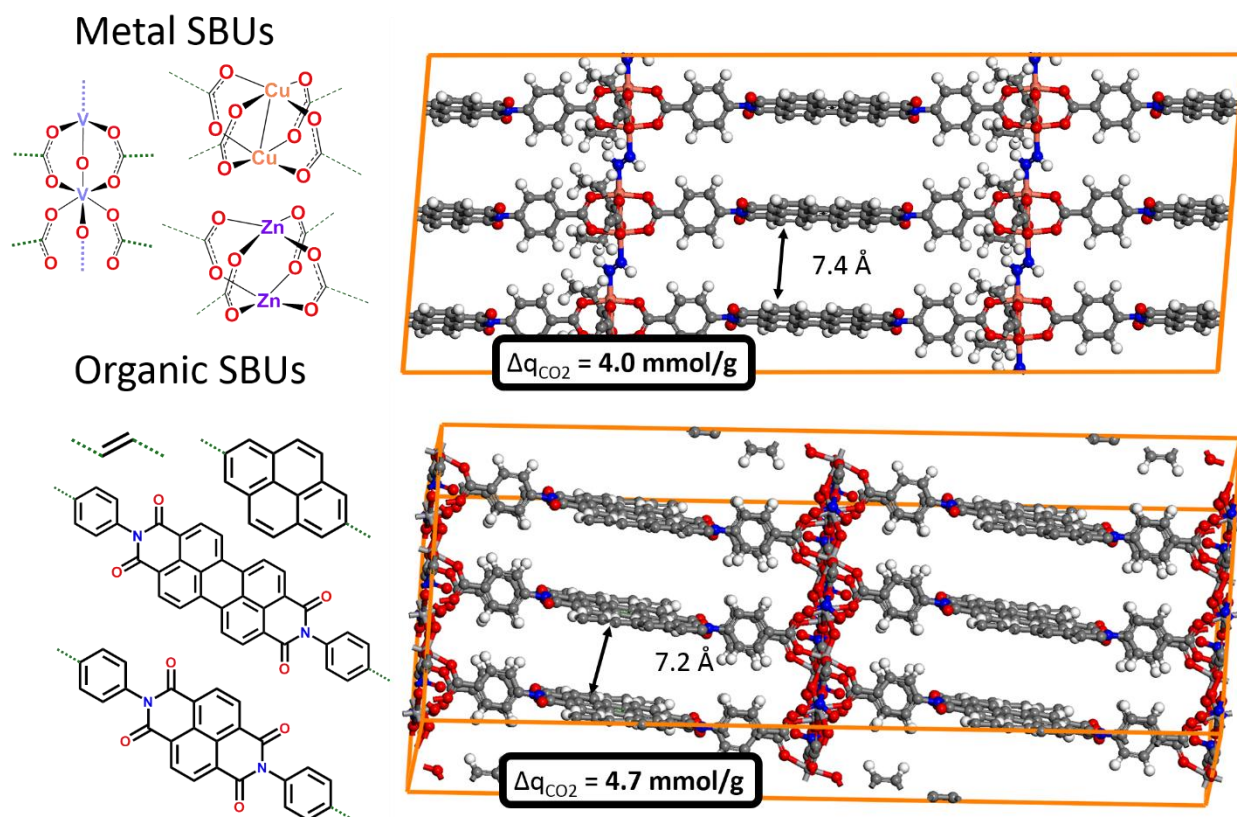


Figure 6.9: Common SBUs used (left) to produce the structural motifs (right) common in high-performing MOFs from the database. Inset within the MOFs are their working capacities (Δq_{CO_2}) for the adsorption and desorption conditions mentioned in the screening protocol. The topology of the first system (top) is **pcu** in which the large conjugated organic SBUs are coordinated to a hexacoordinate copper paddlewheel. Below, the organic SBUs are connected to 1-dimensional chains of vanadium oxide yielding the **sra** topology.

6.5.4 *Ab initio* optimization of the top performers

The 95 top performing materials, which we considered as those with CO_2 working capacities greater than 4 mmol/g were further scrutinized to determine their stability after DFT optimization. We classified the structures into one of three types; *class 1*; MOFs with 1-D sheet-like channels spaced by 7 Å with planar aromatic SBUs. This included MOFs with characteristics of both adsorbophores **A1** and **A2** described above, *class 2*; MOFs assembled with the Cr^{III} SBU in the **acs** topology (containing adsorbophore **A3**), and *class 3*; MOFs with paddlewheels (both Cu and Zn) connected to 4-c organic SBUs forming the **fof** topology. These MOFs produced 1D tube-like channels surrounded by aromatic SBUs.

Figure 6.10 has been partitioned into three plots to demonstrate how each class of structure performed after DFT optimization. It is clear that the DFT and force field optimizations for the first two classes yielded significantly different results, as seen in Figure 6.10a and b. For both classes, the reason for poor performance in the DFT optimized structures is connected to the disappearance of the characteristic adsorbophores discussed in Figure 6.5. The 1D channel characteristic of class 1 is disrupted by distortions in the aromatic alignment, however some structures ‘survived’ the DFT optimization. In this case, functional groups had a subtle effect on the final performance; they could either aid or hinder the stability of the 1D channel. As an example, two MOFs with the same base structure but decorated with different functional groups yielded drastically different post-DFT optimized uptakes. One maintained a high flue gas CO₂ uptake of 4 mmol/g, the pore shape is preserved by bulky bromine functional groups preventing significant twisting of neighbouring aromatic systems (shown in Figure 6.11a). In contrast, the other MOFs CO₂ adsorption drops from 5.3 to 1.9 mmol/g. This was due to significant twisting in the DFT optimized structure, eliminating the ideal pore space seen in these top performing MOFs.

The CO₂ uptake of MOFs in class 2 systematically dropped following optimization at the DFT level as can be seen in Figure 6.10b. This drop was not due to significant distortions in the molecular structure, unlike those in class 1, however the Cr SBUs were optimized slightly away from each other. This eliminated the strong electrostatic binding pocket characterized by **A3**. This is also confirmed by the isosteric heat of adsorption of CO₂, which drops from 10 kcal/mol in the force field optimized structure to 6 kcal/mol in the DFT optimized structure. Finally, structures classified in class 3 (shown in Figure 6.10c) demonstrated the greatest stability after DFT optimization, in both CO₂ adsorption and geometry. These structures did not lose their

strong affinity for CO₂, which appears to be most prominent in a small aperture between larger tubular channels. The channels of one of these high performing structures is shown in Figure 6.11b, the radius of this aperture is roughly 7 Å which agrees well with the spacings of the first adsorbophore for flat binding pockets. The representative binding sites are shown as green CO₂ molecules in Figure 6.11b, many of these sites are surrounded on three sides by large polycyclic hydrocarbons.

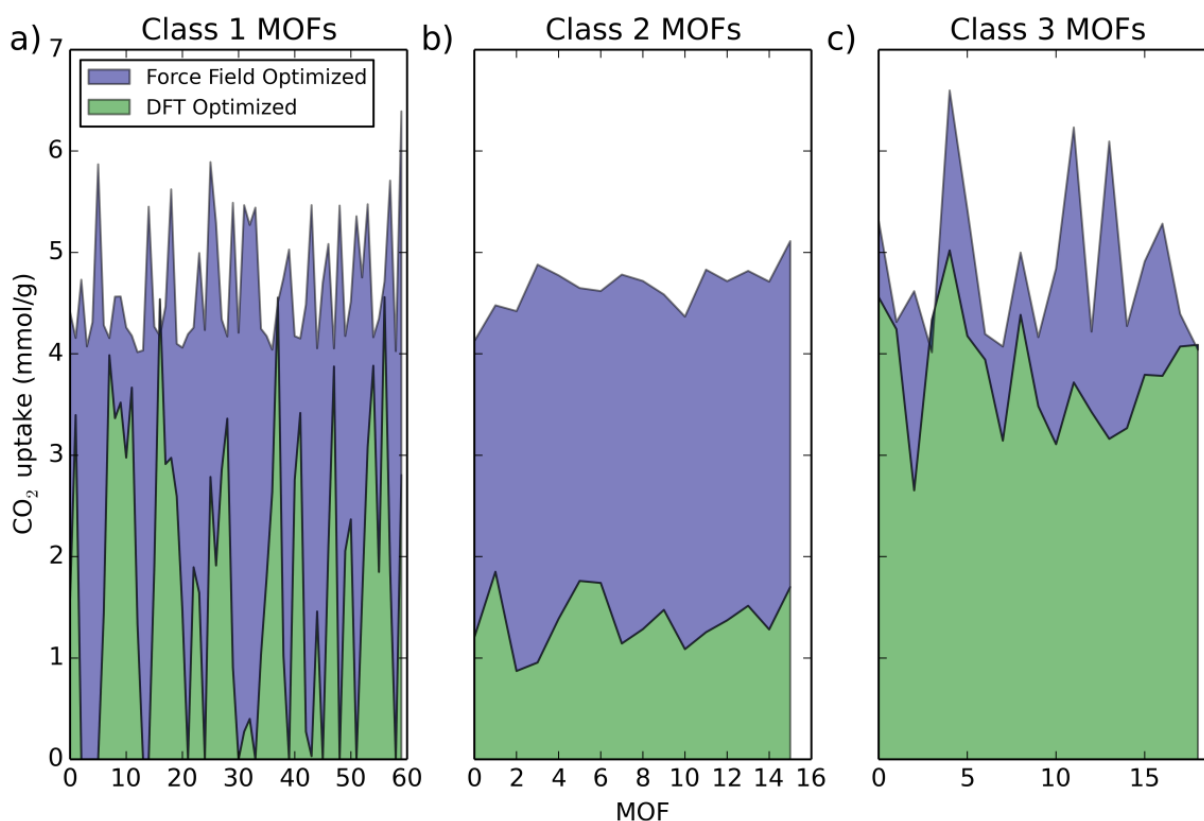


Figure 6.10: CO₂ adsorption comparison of the top 95 performing structures based on the optimization performed on the structures. The CO₂ adsorption was measured at 298 K and 0.15 bar. The plots has been separated into three different classes of MOF: a) contains those MOFs which have the 1-D aromatic-lined pores with spacings of 7 Å, b) shows those MOFs which contain Cr^{III} SBUs in the **acs** topology, and c) shows MOFs with both Zn and Cu paddlewheels bonded to 4-c organic SBUs in the **fof** topology.

While class 3 MOFs showed the most stability after DFT optimization, their selectivity for CO₂ over N₂ was also the weakest. They averaged a selectivity of 70, which is less than half of the target value. Indeed the materials we discovered which show the most promise are those of

class 1, whose selectivity averaged 150, but could reach as high as 487. One of the most promising structures is reported in Figure 6.11a, whose CO₂/N₂ selectivity is 157 while still maintaining a good working capacity of 3.5 mmol/g after DFT optimization.

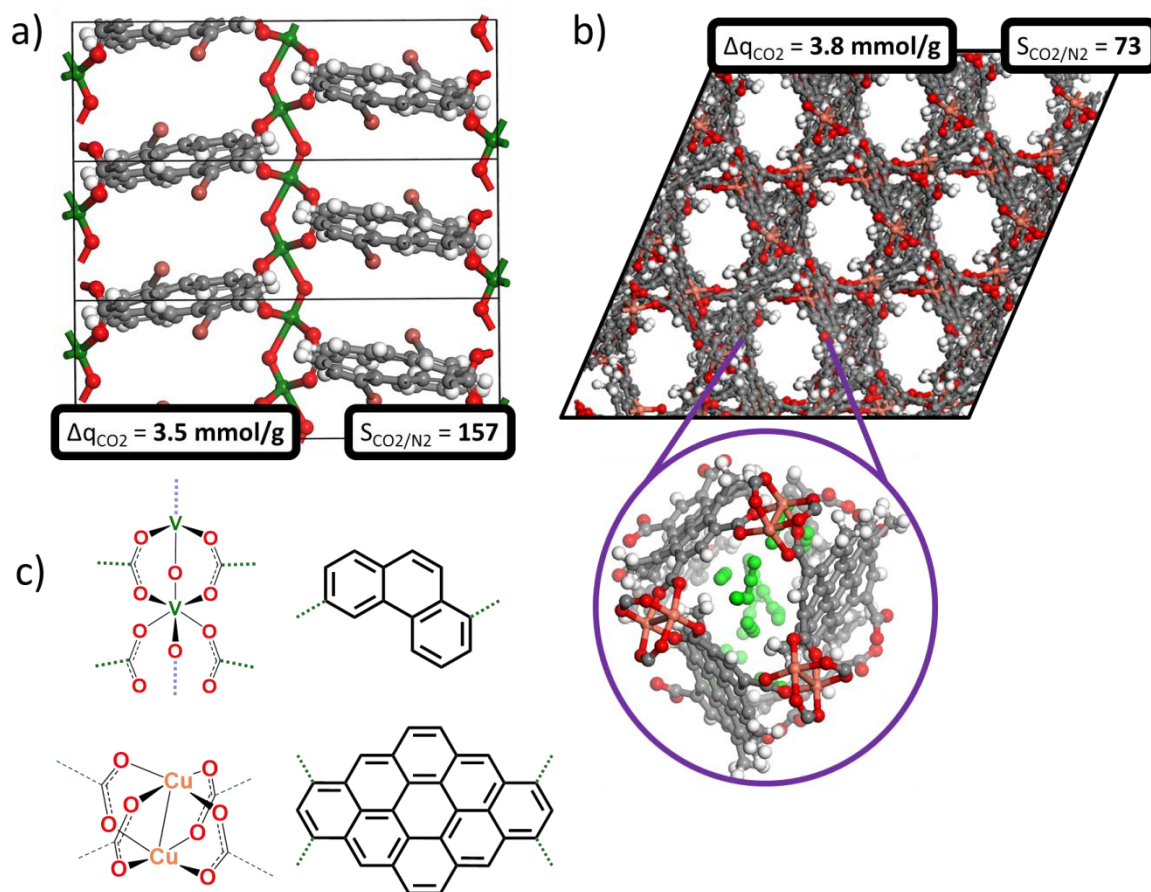


Figure 6.11: Some of the top performing MOFs optimized at the DFT level of theory. a) V^{IV}O rod SBU bonded to a dicarboxylated phenanthrene organic SBU functionalized with bromine (seen in brown). b) View down the 1-D pore of a high performing MOF with the **fof** topology. The binding sites are represented as green molecules. c) The SBUs used to construct these MOFs. The vanadium chain and the phenanthrene SBU were used to construct the MOF in a), while the MOF in b) was constructed with the Cu paddlewheel and the large dibenzocoronene SBU.

6.6 Conclusions

In this work we have conducted a screening study on 325,000 hypothetical MOFs with the goal of discovering similar structural characteristics which aid in CO₂ sorption performance. The hypothetical MOFs were generated using a method which snaps together rigid SBU building

blocks and were constructed from a set of 7 inorganic, 87 organic and 18 functional groups, forming 11 topologies. From the initial screen, we discovered over 8000 MOFs with adsorption performance competitive with zeolite 13X. These structures were examined in more detail with the intent of finding common chemistry within the pores to suggest as synthetic guides, as well as to identify some possible structures for targeted synthesis. The top structures with CO₂ working capacities above 4 mmol/g were optimized at the DFT level to determine the stability of the performance indicators. In general, the post-optimized performance decreased, however for some, their working capacity and selectivity remained quite high. In total 10 of the top 95 structures maintained their working capacities above 4 mmol/g while 48 maintained a working capacity above 3 mmol/g. A list of these high performing MOFs is provided in the appendix to this chapter. In this list each MOF is broken down into its SBUs and functional groups, and is accompanied by their post-DFT-optimized working capacities and selectivities.

For the first time, this screening study has identified molecular-level trends related to high performing MOFs for CO₂ adsorption. Here we developed a tool to identify common strong CO₂ binding motifs using a substructure matching algorithm. These motifs, termed adsorbophores, show a considerable binding energy with CO₂ (8-10 kcal/mol) which can be almost completely described by the atoms from the substructure. The discovered adsorbophores aided in the separation of the top 95 MOFs into one of three classes. Namely MOFs with 1-D pores of layered aromatic sheets separated by roughly 7 Å, MOFs assembled in the **acs** topology using 6-c Cr SBUs whose open metal sites form a strong interaction with CO₂, and MOFs assembled in the **fof** topology with cylindrical channels possessing 7 Å wide apertures.

There are several aspects of the first two adsorbophores **A1** and **A2** which make them exciting targets for synthetic design. They lack open metal sites, a feature which has plagued

experimental high-performing MOFs. Open metal sites show strong interactions with CO₂, however they are known to be unstable when exposed to water, which competitively binds to open metal sites and promotes decomposition of the material. The reported average binding energies for **A1** and **A2** correspond to low heats of adsorption, suggesting that the presence of these adsorbophores will not increase the cost of regenerating these materials while at the same time providing strong interactions with CO₂. In addition, the two adsorbophores represent local structures in MOFs which are not overly confined, meaning they are accessible to CO₂ while at the same time not preventing further diffusion of gas into the MOF. This factor is important when considering the kinetics of rapid loading and unloading of these materials during TPSA cycles.

6.7 References

- (1) Espinal, L.; Morreale, B. D. *MRS Bull.* **2012**, 37, 431.
- (2) Rubin, E. S.; Mantripragada, H.; Marks, A.; Versteeg, P.; Kitchin, J. *Prog. Energy Combust. Sci.* **2012**, 38, 630.
- (3) Ishibashi, M.; Ota, H.; Akutsu, N.; Umeda, S.; Tajika, M.; Izumi, J.; Yasutake, A.; Kabata, T.; Kageyama, Y. *Energy Convers. Manag.* **1996**, 37, 929.
- (4) Liu, Z.; Wang, L.; Kong, X.; Li, P.; Yu, J.; Rodrigues, A. E. *Ind. Eng. Chem. Res.* **2012**, 51, 7355.
- (5) Wang, L.; Yang, Y.; Shen, W.; Kong, X.; Li, P.; Yu, J.; Rodrigues, A. E. *Chem. Eng. Sci.* **2013**, 101, 615.
- (6) Takamura, Y.; Narita, S.; Aoki, J.; Hironaka, S.; Uchida, S. *Sep. Purif. Technol.* **2001**, 24, 519.
- (7) Zhang, J.; Webley, P. A. *Environ. Sci. Technol.* **2008**, 42, 563.
- (8) Su, F.; Lu, C. CO₂ capture from gas stream by zeolite 13X using a dual-column temperature/vacuum swing adsorption. *Energy & Environmental Science*, 2012, 5, 9021.

- (9) Radosz, M.; Hu, X.; Krutkramelis, K.; Shen, Y. *Ind. Eng. Chem. Res.* **2008**, *47*, 3783.
- (10) Ho, M. T.; Allinson, G. W.; Wiley, D. E. *Ind. Eng. Chem. Res.* **2008**, *47*, 4883.
- (11) Haldoupis, E.; Nair, S.; Sholl, D. S. *J. Am. Chem. Soc.* **2012**, *134*, 4313.
- (12) Allen, F. H. *Acta Crystallogr. Sect. B Struct. Sci.* **2002**, *58*, 380.
- (13) Wilmer, C. E.; Leaf, M.; Lee, C. Y.; Farha, O. K.; Hauser, B. G.; Hupp, J. T.; Snurr, R. Q. *Nat. Chem.* **2012**, *4*, 83.
- (14) Lin, X.; Telepeni, I.; Blake, A. J.; Dailly, A.; Brown, C. M.; Simmons, J. M.; Zoppi, M.; Walker, G. S.; Thomas, K. M.; Mays, T. J.; Hubberstey, P.; Champness, N. R.; Schröder, M. *J. Am. Chem. Soc.* **2009**, *131*, 2159.
- (15) Wilmer, C. E.; Farha, O. K.; Bae, Y.-S.; Hupp, J. T.; Snurr, R. Q. *Energy Environ. Sci.* **2012**, *5*, 9849.
- (16) Lin, L.-C.; Berger, A. H.; Martin, R. L.; Kim, J.; Swisher, J. a; Jariwala, K.; Rycroft, C. H.; Bhowan, A. S.; Deem, M. W.; Haranczyk, M.; Smit, B. *Nat. Mater.* **2012**, *11*, 633.
- (17) Fernandez, M.; Trefiak, N. R.; Woo, T. K. *J. Phys. Chem. C* **2013**, *117*, 14095.
- (18) Leach, A. R.; Gillet, V. J.; Lewis, R. a; Taylor, R. *J. Med. Chem.* **2010**, *53*, 539.
- (19) Deng, H.; Doonan, C. J.; Furukawa, H.; Ferreira, R. B.; Towne, J.; Knobler, C. B.; Wang, B.; Yaghi, O. M. *Science* **2010**, *327*, 846.
- (20) Togo, A. spglib (<http://spglib.sourceforge.net>), 2012.
- (21) Grosse-Kunstleve, R. W. *Acta Crystallogr. Sect. A Found. Crystallogr.* **1999**, *55*, 383.
- (22) Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. *J. Am. Chem. Soc.* **1992**, *114*, 10024.
- (23) O'Boyle, N. M.; Morley, C.; Hutchison, G. R. *Chem. Cent. J.* **2008**, *2*, 5.
- (24) O'Boyle, N. M.; Banck, M.; James, C. a; Morley, C.; Vandermeersch, T.; Hutchison, G. R. *J. Cheminform.* **2011**, *3*, 33.
- (25) Serna-Guerrero, R.; Belmabkhout, Y.; Sayari, A. *Chem. Eng. J.* **2010**, *158*, 513.
- (26) Tlili, N.; Grévilot, G.; Vallières, C. *Int. J. Greenh. Gas Control* **2009**, *3*, 519.
- (27) Wang, L.; Liu, Z.; Li, P.; Wang, J.; Yu, J. *Adsorption* **2012**, *18*, 445.

- (28) Artanto, Y.; Jansen, J.; Pearson, P.; Puxty, G.; Cottrell, A.; Meuleman, E.; Feron, P. *Int. J. Greenh. Gas Control* **2014**, *20*, 189.
- (29) Yu, H.; Morgan, S.; Allport, A.; Cottrell, A.; Do, T.; McGregor, J.; Wardhaugh, L.; Feron, P. *Chem. Eng. Res. Des.* **2011**, *89*, 1204.
- (30) Wang, L.; Yang, Y.; Shen, W.; Kong, X.; Li, P.; Yu, J.; Rodrigues, A. E. *Ind. Eng. Chem. Res.* **2013**, *52*, 7947.
- (31) Wang, L.; Liu, Z.; Li, P.; Yu, J.; Rodrigues, A. E. *Chem. Eng. J.* **2012**, *197*, 151.
- (32) Gale, J. D.; Rohl, A. L. *Mol. Simul.* **2003**, *29*, 291.
- (33) Martin, R. L.; Lin, L.-C.; Jariwala, K.; Smit, B.; Haranczyk, M. *J. Phys. Chem. C* **2013**, *117*, 12159.
- (34) Martin, R. L.; Simon, C. M.; Smit, B.; Haranczyk, M. *J. Am. Chem. Soc.* **2014**, *136*, 5006.
- (35) Banerjee, A.; Adams, N.; Simons, J.; Shepard, R. *J. Phys. Chem.* **1985**, *89*, 52.
- (36) Campaña, C.; Mussard, B.; Woo, T. K. *J. Chem. Theory Comput.* **2009**, *5*, 2866.
- (37) Konc, J.; Janezic, D. *MATCH Commun. Math. Comput. Chem.* **2007**, *58*, 569.
- (38) Rappe, A. K.; Goddard, W. A. *J. Phys. Chem.* **1991**, *95*, 3358.
- (39) 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*, 8814.
- (40) Potoff, J. J.; Siepmann, J. I. *AIChE J.* **2001**, *47*, 1676.
- (41) Hafner, J. *J. Comput. Chem.* **2008**, *29*, 2044.
- (42) Kresse, G.; Furthmüller, J. *Phys. Rev. B* **1996**, *54*, 11169.
- (43) Kresse, G.; Joubert, D. *Phys. Rev. B* **1999**, *59*, 1758.
- (44) Blöchl, P. *Phys. Rev. B* **1994**, *50*, 17953.
- (45) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1997**, *78*, 1396.
- (46) Perdew, J.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865.
- (47) Tang, W.; Sanville, E.; Henkelman, G. *J. Phys. Condens. Matter* **2009**, *21*, 084204.

- (48) Kadantsev, E. S.; Boyd, P. G.; Daff, T. D.; Woo, T. K. *J. Phys. Chem. Lett.* **2013**, 4, 3056.
- (49) Wang, Y.; LeVan, M. D. *J. Chem. Eng. Data* **2009**, 54, 2839.
- (50) Mason, J. A.; Sumida, K.; Herm, Z. R.; Krishna, R.; Long, J. R. *Energy Environ. Sci.* **2011**.
- (51) Bader, R. F. W. *Atoms in Molecules - A Quantum Theory*; Oxford University Press: Oxford, 1994.

6.8 Appendix – High performing MOFs

Table 6.3: List of the post-DFT-optimized top performing MOFs, ranked by their working capacities. Each MOF in the table is described by the functional group and SBUs which were used to build them.

Metal SBU	Organic SBU #1	Organic SBU #2†	Functional group	Working capacity (mmol/g)	CO ₂ /N ₂ selectivity
2	10	29	OH	4.37	132
2	10	29	Me	4.33	84
2	10	29	HCO	4.22	116
2	10	29	OMe	4.16	205
2	10	29	CN	4.14	77
3	41		H	4.09	72
3	40		HCO	4.06	76
2	10	29	OH	4.04	56
3	41		NH ₂	4.01	134
2	10	17	Et	3.99	98
2	10	29	NH ₂	3.97	84
9	10	29	OMe	3.95	139
2	10	29	H	3.91	152
2	10	29	NH ₂	3.89	207
2	10	29	Me	3.89	171
2	10	29	OH	3.85	79
2	10	29	OMe	3.84	132

2	10	29	SO3H	3.84	141
2	10	29	Cl	3.80	155
9	4	18	CN	3.75	306
3	40		Me	3.72	71
2	10	29	NH2	3.68	281
3	10	29	OH	3.64	81
3	5	27	NHMe	3.62	94
3	41		NH2	3.59	64
3	40		Et	3.56	64
3	40		F	3.40	51
9	10	29	OH	3.39	148
3	5	17	SO3H	3.36	96
3	41		OH	3.36	64
9	10	29	NO2	3.33	121
2	10	28	Et	3.32	66
3	41		CN	3.30	44
2	10	29	CN	3.26	163
3	41		CN	3.20	53
3	40		NH2	3.16	77
3	41		F	3.12	73
3	40		Cl	3.12	44
2	10	29	Et	3.10	87
3	41		OH	3.09	48
2	10	17	NH2	3.09	185
2	10	29	HCO	3.09	238
3	41		Cl	3.07	60
2	5	29	OMe	3.06	37
9	10	29	CN	3.03	116
9	11	16	NH2	3.03	230

†in some cases only one type of SBU was used to build the structure. In which case the Organic SBU #2 column was left blank.

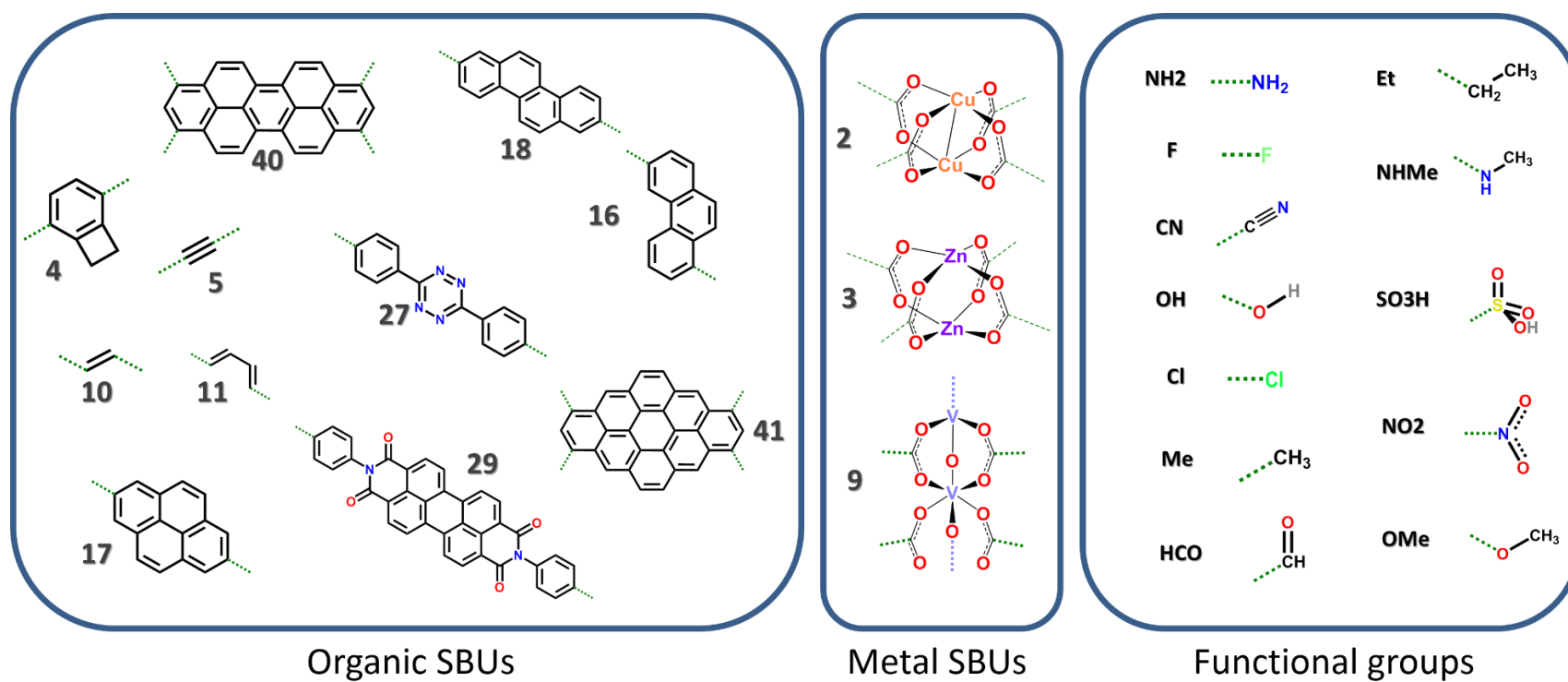


Figure 6.12: Image reference for the indices and functional group short-hands provided in Table 6.3

7 Construction of Hypothetical MOFs using a Graph Theoretical Approach

7.1 Abstract

There have been several recent methods developed to generate hypothetical metal-organic framework (MOF) materials for the purposes of structure prediction and high throughput screening applications. Here we present a method for generating 3-periodic crystalline materials which takes as input secondary building units (SBU)s, as well as the underlying connectivity of a net in the form of a labelled quotient graph. The hypothetical materials are constructed based on the principle that commencing from a labelled quotient graph one can construct a net embedding in Euclidean space with an infinite number of different representations. Thus, crystalline structures can be realized by manipulating a net's embedding such that vertices of the net match the geometries of the desired SBUs. Examples are presented showing the application of this method in the context of generating hypothetical MOFs where utilizing the most common realization of a net, called the barycentric representation, may not produce a viable structure.

7.2 Introduction

The method for generating hypothetical MOFs introduced in Chapter 3 demonstrates some of the challenges associated with arranging molecules together to form three-dimensional (3D) periodic structures. In that work, alignment vectors were applied to each SBU, which were then rigidly snapped together to form new materials. While this method enabled the construction of over a million hypothetical MOFs, parameterizing these SBUs requires a significant understanding of how 3D crystalline materials organize. Moreover, introducing geometric diversity in this database involved a tremendous amount of effort, where hand-built parameters were constructed for each SBU and new framework topology. This was recently exemplified in a study where a database of 137,000 hypothetical MOFs developed by Wilmer *et al.*¹, using a

similar method detailed in Chapter 3, was shown to have MOFs in a very limited range of 6 topologies.² Thus, expanding the database in terms of chemical and topological diversity is a significant undertaking and a large bottle-neck for future screening studies. As a result, there is an impetus develop a program which can forego the extensive parameterization required to build these structures, making the generation of new structures much easier, less time-consuming, and more automated. One possibility is to align SBUs in space using pre-defined periodic templates as guides, in which SBUs are abstracted to single vertices, connected to other vertices by edges. This abstraction of chemical information to a collection of vertices and edges is not new, in fact it was introduced in the 70's and 80's as a way to classify crystalline materials by Wells³ and O'Keeffe⁴.

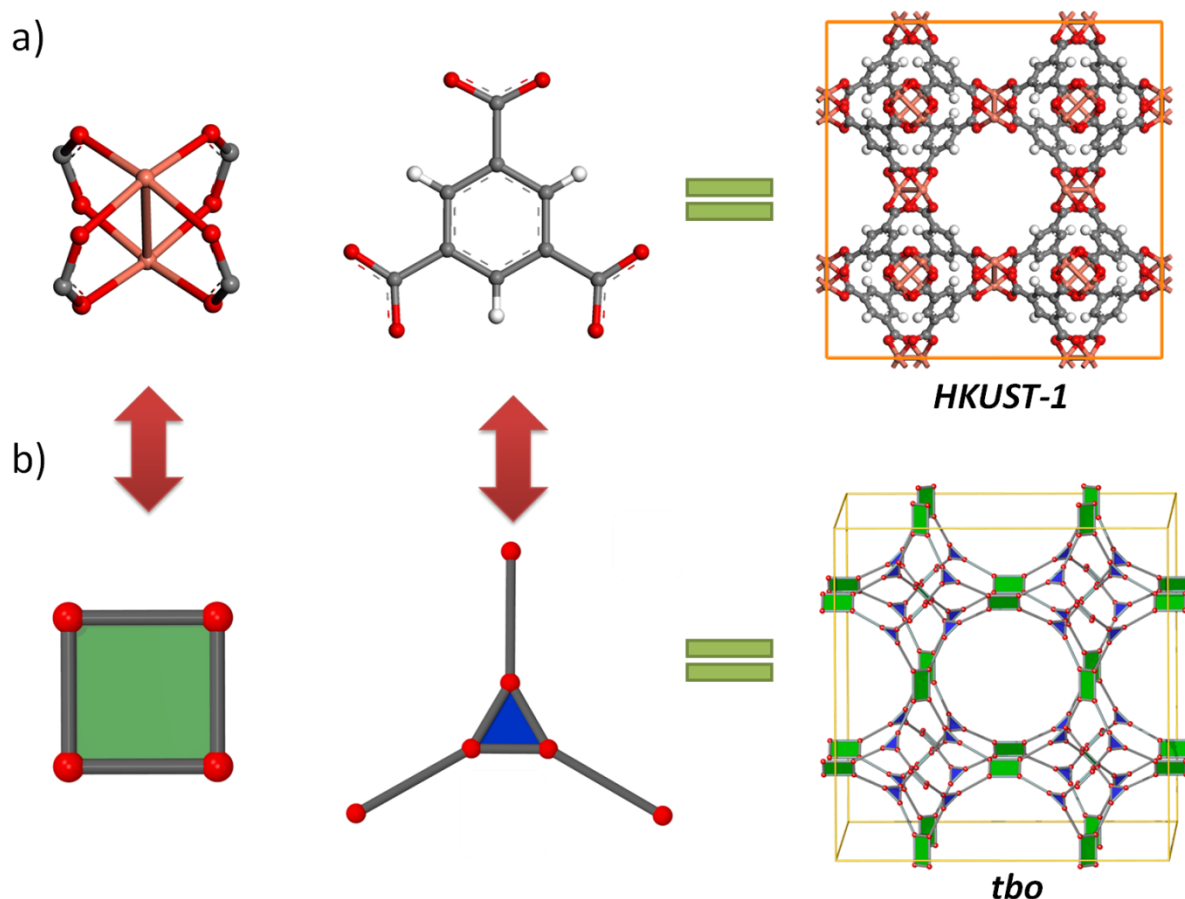


Figure 7.1: An abstraction of a) the SBUs of HKUST-1 to b) vertices and edges to produce the net, *tbo*.

These collections of vertices and edges are special types of graphs called nets. An example of a net is shown in Figure 7.1b, with edges as grey lines and vertices as red balls. The green squares and blue triangles are used to guide the eye to their geometric placement in the net *tbo*. It should be stressed that nets do not specify a particular geometry, rather they describe an underlying connectivity between vertices. As such a single net can encompass a number of different representations in Euclidean space, making the concept useful as a classification tool for crystalline materials.⁵ The number of possible unique nets, i.e. nets which cannot be superimposed on each other by some minor adjustment of edges and vertices, is a subject of mathematical enumeration and is not upper bounded. However, databases of nets have sprung up such as the Reticular Chemistry Structure Resource (RCSR), which contain many of the nets

important in classifying crystal structures.⁶ These nets are typically referred to with a three-letter label, in some cases a short form of a well known crystal structure which it can be classified with. The **twisted boracite (tbo)** net, for example, contains vertices of degree 3 and 4, and is commonly used to classify the well-known MOF, HKUST-1.⁷ In other cases the label describes a well-known crystal system, such as the **primitive cubic (pcu)** net, which is famously associated with MOF-5.⁸ On its surface, the RCSR database provides three dimensional representations of these nets. The vertices and edges of each net are placed to obtain the maximum possible symmetry achievable, which was shown to be when vertices are at the centre of its edge-connected neighbours.⁹ This is known as the nets *equilibrium* or *barycentric* placement. Figure 7.2 demonstrates this for the diamond net, where a) shows the barycentric form and b) shows a distortion in the node geometry, which lowers the maximum achievable symmetry.

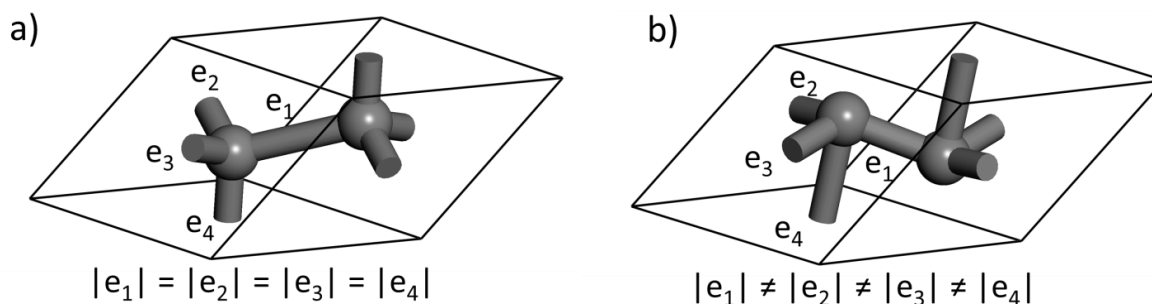


Figure 7.2: The diamond (**dia**) net in it's a) barycentric representation where all edge lengths are equal, and b) a non-barycentric representation. The different edge lengths result in different geometries of the nodes in the net.

Eon formally demonstrated that one can obtain a unique representation of a net with deviations from its barycentric placement using a series of distortion vectors called colattice vectors.¹⁰ These vectors, along with a net's lattice vectors can fully describe a three-dimensional embedding. Herein we report a graph theoretical approach for obtaining hypothetical MOFs from underlying nets. Our algorithm requires the same information as described in the method of Martin and Haranczyk,¹¹ that is to say, the user must input a set of SBUs with arbitrarily defined

connection points and a desired net topology. However our method utilizes the underlying net, not the barycentric representation, to obtain hypothetical MOF structures. This maximizes the number of achievable hypothetical MOF structures from a given set of SBUs.

The following sections are outlined here to aid the reader. Section 7.3 will introduce some of the graph theoretic terminology used in the remainder of this text, including definitions of nets and labelled quotient graphs which are fundamental concepts used in the construction of hypothetical MOFs in this work. Section 7.4 will discuss how a dimensionless labelled quotient graph is embedded in Euclidean space, initially in its barycentric form, and then in a form which supports SBUs of arbitrary geometry. In section 7.5, we will demonstrate the method in a number of applications; 7.5.1 demonstrates the challenges of embedding nets to fit SBUs with varying geometries, while 7.5.2 demonstrates the wide variety of structures one can obtain from using only two SBUs and labelled quotient graphs from the RCSR. In 7.5.3 we build MOFs using the **tbo** topology, using SBUs with different geometries and symmetries. Section 7.5.4 shows an example of building a MOF using the labelled quotient graph of an unstable net, which possess vertex collisions in its barycentric form. In 7.5.5 we construct hypothetical versions of recently discovered high-surface area MOFs which poses a challenge to current structure generation methods. Finally, we conclude and summarize this work in section 7.6.

7.3 Background

As is the case in several studies presenting graph theory in context of studying the topologies of crystal structures,^{10,12–15} these authors find it necessary to define some basic graph theoretical terms to aid the reader in navigating the remainder of this work.

- A **graph**, $G(V, E)$, consists of a set of vertices V and edges E . Each edge, $e \in E$ is terminated by two vertices $u, v \in V$ or one vertex $v \in V$ in the case of loops. If the graph has

a finite number of vertices in the set V , then the graph is called **finite**. If each vertex $u \in V$ contains a finite number of edges incident upon it, the vertices are **locally finite**.

- A **walk** in a graph $G(V, E)$ is a traversal of edges $\{e_1, e_2, \dots, e_n\} \in E$ from u to v in V .
- A **path** between two vertices $u, v \in V$ is a walk over a sequence of vertices and edges in $G(V, E)$ such that no edge is visited twice.
- **Cycles** are paths which the starting and ending vertex is the same, however all other edges and vertices visited are unique.
- **Connected graphs** are graphs which there exists a path between any two vertices $u, v \in V$. All graphs discussed herein are connected graphs.
- **Co-cycles** are sets of edges in E , which upon removal, will fragment a connected graph. Co-cycles are also called **cuts** of the graph.
- **Trees** are connected graphs but contain no cycles.
- **Automorphisms** on a graph are permutations of the graphs edges and vertices which preserve incidence. They can be likened to symmetry operations in crystalline structures, whose operation (rotation, reflection, translation, or a combination thereof) preserves atomic structure.
- An **isomorphism** between two graphs G and G' describes a one-to-one correspondence of their vertex and edge sets. There exists a mapping which sends each vertex and edge in G to a vertex and edge in G' .
- The **degree** of a graph with locally finite vertices is the number of edges incident upon each vertex. If a graph is of degree four, then all vertices possess four neighbours.
- A **net**, as described by Delgado-Friedrichs and O'Keeffe¹², is a 'special kind of graph'. Nets are connected infinite graphs with locally finite vertices. Moreover, nets are simple, meaning

they possess no edge loops or multiple edges connecting two vertices. Nets are dimensionless, being fully described by an infinite set of vertices and edges. A net is considered n -periodic if there exists a group of automorphisms which act freely on the net (leave no vertex fixed) with rank n . In other words, the net has n unique translational symmetry elements. A discussion of these automorphisms, called the translational group, can be found in Eon.¹⁰ In that work, he demonstrates the method of embedding nets in the general n -periodic case, however in most of the cases we will discuss, the 3-periodic nets obtained from the RCSR⁶ will be used. Thus, for the sake of simplicity we will focus our discussion on 3-dimensional embeddings of 3-periodic nets, and the term *net* and *3-periodic net* will be synonymous.

It is this concept of a net which is frequently used in studies to classify a particular material. For example, the isorecticular series of MOFs introduced by Eddaoudi *et al.*¹⁶ can all be classified into the net **pcu**. This is accomplished by abstracting the Zn_4O cluster as a single 6-connected vertex and the linear organic dicarboxylate molecules as edges between vertices. The definition of the net above infers an important concept; the vertex geometries and spatial distances between neighbouring vertices are irrelevant when classifying MOFs into a net. For example in Figure 7.1, the MOF HKUST-1 is abstracted to net **tbo**. However the net does not define a particular set of edge lengths and vertex geometries, it merely describes *how* the vertices connect to each other. Topological analysis programs such as Systre⁹ and TOPOS¹⁷ perform analyses on the incidence structure of the MOF's underlying vertices and edges to determine isomorphism with a known net, ignoring the spatial geometries that arise from chemical bonding information. Thus it can be said there are an infinite number of possible 3-dimensional representations, or embeddings, of a net. Here we refer to embeddings as what Delgado-

Friedrichs and O'Keeffe¹² call faithful embeddings, that is, the net vertices are assigned to nodes in Euclidean space, and the edges assigned to lines connecting nodes such that there are no line crossings nor overlapping nodes.

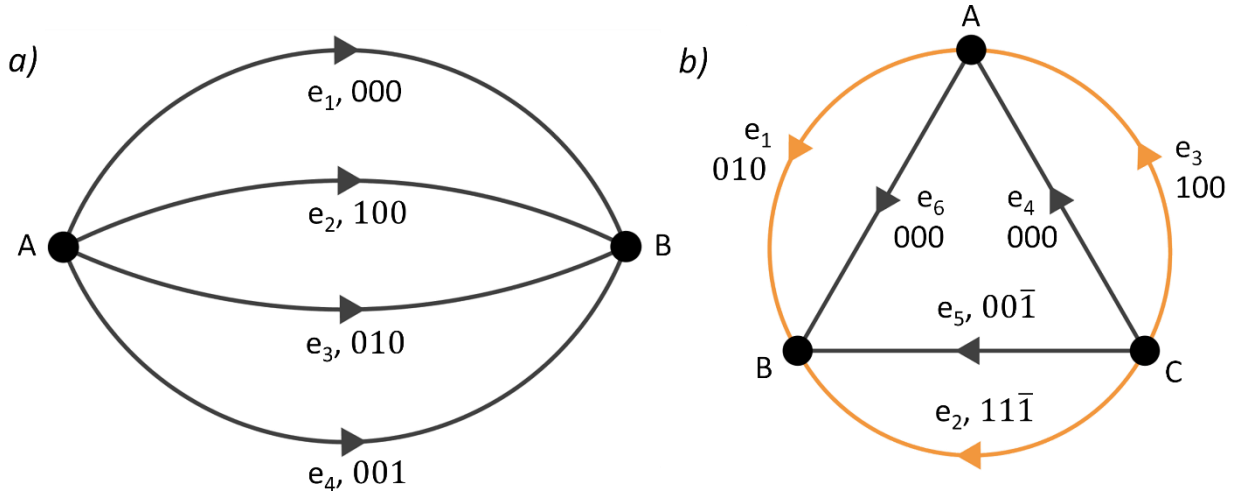


Figure 7.3: a) The labelled quotient graph of **dia**, b) The labelled quotient graph of **qtz**. By convention the arrows on each edge indicate the positive orientation of the arc.

We will work with a finite graph called the **labelled quotient graph**, which contains all the translationally equivalent vertices and edges of a 3-periodic net. This is conceptually analogous to the unit cell of a crystalline material. To ensure a one-to-one relationship between quotient graphs and their nets, it is necessary to assign labels and directions to the edges of the quotient graph.¹³ Labels represent the translational periodicity of edges, with assignments in three dimensional integer space ($\mathbb{Z}^3$). For example in Figure 7.3a, the edge labelled e_2 is associated with the label '100' which indicates that e_2 connects vertex A to vertex B via by crossing the periodic vector 100. In many cases (particularly with very large quotient graphs), there is some redundancy in the edge labels. For example Figure 7.3b shows two edges, e_4 and e_6 which possess the '000' label. This indicates that the two edges do not span a periodic vector, they can be found in the unit cell. Since 3D vectors possess an end point and an origin by definition, each edge in the labelled quotient graph must be assigned a "+" or "-" direction to

ensure a one-to-one relationship between the edge incidence structure and the direction of a vector embedded in Euclidean space. By convention, arrows placed on edges point in the "+" direction (all the edges in Figure 7.3 possess arrows pointing in the "+" direction). We will differentiate an edge from a directed (or oriented) edge by calling edges with orientation, **arcs**. An orientation of the graph is a set of directions assigned to the edges of the graph. For example the orientation of the graph in Figure 7.3a describes the four edges as arcs, whose positive directions point towards vertex B from vertex A. Like the nets they describe, quotient graphs do not have a particular n-dimensional representation; they are purely topological constructs. For example a quotient graph of the 3-periodic diamond net can be fully described by the set of directed edges from vertex A to vertex B: $(e_1, 000)$, $(e_2, 100)$, $(e_3, 010)$, $(e_4, 001)$. A 2-dimensional representation of the quotient graph of **dia** is provided in Figure 7.3a.

Due to their utility in describing the topology of real crystal structures, many studies of crystalline structures work with the labelled quotient graph.^{9,13,18–25} Moreover, they are the starting point of mathematically enumerating the infinite number of possible 3-D structures.¹³ While the RCSR provides 3-D representations of nets in their database, it is also accompanied by the associated labelled quotient graph.⁶ In this study we use the labelled quotient graph as a starting point to assemble new hypothetical structures for the reasons that it is a flexible and robust way of describing a net, and it is straight-forward to assemble novel hypothetical materials as new labelled quotient graphs (and therefore nets) are enumerated.

7.4 Methods

To describe how generating hypothetical MOFs is accomplished, we must first demonstrate the method of producing a net representation from a labelled quotient graph. The ultimate goal of this method is to send the dimensionless arcs of a labelled quotient graph to three-dimensional (3-D) vectors in a periodic lattice. These so-called ‘embeddings’ are

constrained based on the conditions that the vertices and arcs maintain their incidence relationships, the arcs are mapped to straight lines (3-D vectors), and the periodicity indicated by the arc labels are preserved. This is done by mapping the arcs of a labelled quotient graph to two sets of 3-dimensional vectors called the lattice and co-lattice vectors. For the sake of clarity, we will say vertices and arcs (or edges) are dimensionless elements of the labelled quotient graph, and their images in Euclidean space are nodes and lines, respectively. The following will outline the method of mapping vertices and edges to nodes and lines, and mention how this was implemented in our structure generation code.

7.4.1 Periodic Lattice Representations of Labelled Quotient Graphs

The first step to building a 3D representation of a net is to embed the arcs of a labelled quotient graph as lines, expressed as fractions of the translational symmetry elements (100), (010), and (001). Arcs from the labelled quotient graph with m arcs and n vertices can be expressed in the form of a basis of cycles and co-cycles of the graph.

We define a set of $m - n + 1$ cycle vectors, $\mathcal{B}_c$, as a generating set of cycles in the edge space of the graph. Recall, a cycle is a path starting and ending with the same vertex. We express this cycle as a summation over the arcs which lead back to the original vertex. The arc orientation (in the + or - direction) must be considered such that they always point in the direction travelled in the path. This is done by traversing an arc as $+e$ if the arc is pointing towards the next vertex in the path, or 'flipping' it as $-e$ if the arc is pointing in the opposite direction. For example, we define the orange arcs in the labelled quotient graph of **qtz** in Figure 7.3b as the cycle δ . An expression of δ can be found, starting and ending at vertex C, by traversing e_3 and e_1 in the positive direction and reversing the direction of e_2 such that the cycle vector is the summation of the edges $\delta = (e_3 + e_1 - e_2)$. The set of lattice vectors (L) are generated by the cycles of $\mathcal{B}_c$ by summing the oriented labels of the cycle arcs. For example the

lattice vector associated with the cycle δ would be a summation of the labels of $+e_3$, $+e_1$, and $-e_2$ giving $100 + 010 + \overline{111} = 001$. Conceptually, this cycle mapping can be described by a walk of the net starting at vertex C, traversing e_3 to vertex A, traversing e_1 to vertex B, and finally traversing $-e_2$ to arrive at an image of vertex C in the periodic direction 001.

The co-cycle basis of the labelled quotient graph, $\mathcal{B}_s$, can be generated by the sum of all outward oriented arcs from the first $n - 1$ vertices, reversing the arc direction in the same manner as above if necessary. For example the co-cycle space of the labelled quotient graph of **qtz** can be found by taking the outward oriented arcs from the vertices A and B, yielding $\mathcal{B}_s = (e_1 - e_3 - e_4 + e_6), (-e_1 - e_2 - e_5 - e_6)$. The arcs associated with the final vertex, C, is included within this basis. Each co-cycle basis vector is mapped to an arbitrary vector, called a co-lattice vector (L^*) in $\mathbb{R}^3$. This co-lattice vector can be considered a deviation of an embedded vertex (node) from the centre of mass of its neighbouring nodes. Eon¹⁰ demonstrated that if one maps the cycle basis to the lattice vectors and the co-cycle basis to a set of zero vectors ($L^* = 0$), one obtains the unique barycentric representation of a net, a representation seen in the RCSR⁶ as well as publications discussing nets in the context of MOF classification.^{5,26} The method of mapping the two graphs shown in Figure 7.3 to their barycentric representations will be demonstrated below. Importantly it was shown that uniqueness of the net embedding can be extended to all representations where the co-cycle basis is not mapped to zero vectors ($L^* \neq 0$), which is the general strategy employed in this study to obtain net embeddings supporting the geometries of a set of SBUs.

Once the bases of the cycle and co-cycle spaces are constructed, we now map the arcs to images in Euclidean space. This is accomplished by constructing a matrix, B^* , in the edge space. The first $m - n + 1$ rows contain the cycles in $\mathcal{B}_c$, the remaining $n - 1$ rows contain each co-

cycle in $\mathcal{B}_s$. The cycle and co-cycle vectors are represented in the edge space such that elements in each row contains +1 if the arc contributes in the positive direction, -1 if the arc is reversed, and 0 if the arc does not contribute to the vector. This type of arc representation was shown to be non-singular of rank n ,²⁷ indicating that it is a change-of-basis representation of the arcs of the labelled quotient graph. We will map these cycle and co-cycle vectors to a particular lattice direction using a $m \times 3$ matrix, which we will call the lattice-co-lattice matrix (α). This is constructed by setting the first $m - n + 1$ rows as the 3-dimensional lattice vectors associated with the cycles and the last $n - 1$ rows as the co-lattice vectors. We define a unique embedding of the positive orientations of arcs, Ω^+ , in Euclidean space such that

$$B^* \Omega^+ = \alpha \quad 7.1$$

which can be rewritten to solve for the embedding (lines) of the positive orientations of the arcs:

$$\Omega^+ = B^{*-1} \alpha \quad 7.2$$

In the case of the labelled quotient graph for the **dia** net, we construct the B^* matrix using the 3 cycles $(-e_1 + e_2)$, $(-e_1 + e_3)$, $(-e_1 + e_4)$ as a basis for $\mathcal{B}_c$, and one co-cycle, summing the outgoing edges from the vertex A, $(e_1 + e_2 + e_3 + e_4)$ as a basis for $\mathcal{B}_s$. The image of these vectors, represented in the matrix α_{dia} , has its first 3 rows as the sum of the edge labels associated with the 3 cycle vectors, and the final row as the zero vector, which as stated above, will yield the barycentric placement of the vertices.

$$B_{\text{dia}}^* = \begin{pmatrix} -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}, \alpha_{\text{dia}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

Then invoking Eqn 7.2, we obtain a representation of the positive orientations of the arcs in the lattice space.

$$\Omega_{\text{dia}}^+ = \begin{pmatrix} -1/4 & -1/4 & -1/4 & 1/4 \\ 3/4 & -1/4 & -1/4 & 1/4 \\ -1/4 & 3/4 & -1/4 & 1/4 \\ -1/4 & -1/4 & 3/4 & 1/4 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

$$= \begin{pmatrix} -1/4 & -1/4 & -1/4 \\ 3/4 & -1/4 & -1/4 \\ -1/4 & 3/4 & -1/4 \\ -1/4 & -1/4 & 3/4 \end{pmatrix}$$

Each row in the matrix Ω^+ defines the fractional coordinates of a line in the lattice space of the net. With all of the lines defined, a lattice representation of the nodes and lines can be realized by placing the node corresponding to vertex A in any fractional lattice position, say (5/8, 5/8, 5/8). The vertex B can be found by adding the first line in Ω^+ to our initial placement of A. Thus B will have the lattice coordinates (3/8, 3/8, 3/8).

7.4.2 The Metric Tensor

We have defined the arcs as lines and the vertices as nodes in the lattice space. To construct a representation of a net in Cartesian space for the purposes of building hypothetical MOFs, we must construct a representation of the lattice in $\mathbb{R}^3$. The metric tensor, Z , thus completes this description, representing a mapping from the lattice coordinate system defined by the integer vectors (100), (010), and (001) to some other basis in three dimensional space. It is characterized by the dot products of the lattice basis vectors,

$$Z = \begin{pmatrix} \mathbf{a} \cdot \mathbf{a} & \mathbf{a} \cdot \mathbf{b} & \mathbf{a} \cdot \mathbf{c} \\ \mathbf{b} \cdot \mathbf{a} & \mathbf{b} \cdot \mathbf{b} & \mathbf{b} \cdot \mathbf{c} \\ \mathbf{c} \cdot \mathbf{a} & \mathbf{c} \cdot \mathbf{b} & \mathbf{c} \cdot \mathbf{c} \end{pmatrix} \quad 7.3$$

One can assign arbitrary values to the vectors $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$ to yield isomorphic representations of the net in $\mathbb{R}^3$, however, it was shown¹⁰ that a metric tensor of the net can be constructed from the labelled quotient graph, implicitly assuming a barycentric placement of the vertices and arcs in $\mathbb{R}^3$. Let $\mathcal{L}_A$ be a $3 \times m$ matrix containing the set of cycle vectors which form a lattice basis of the

3-periodic net. The metric tensor can be expressed by creating an orthogonal projection in the edge space, $\mathcal{P}$, whose kernel is a basis of the edge space with labels that sum to (000) such that,

$$Z = \mathcal{L}_A \cdot \mathcal{P} \cdot \mathcal{L}_A^t \quad 7.4$$

The projection matrix $\mathcal{P}$ ensures that upon constructing the metric tensor from the edge space, the cycles whose edge labels sum to zero are indeed projected to the null vector. The kernel, K , of the projection contains a basis of the cycles whose edge labels sum to (000) as well as the co-cycle vectors. The orthogonal projection is then constructed according to Godsil and Royle:²⁷

$$\mathcal{P} = I - K^t \cdot (K \cdot K^t)^{-1} \cdot K \quad 7.5$$

In cases of quotient graphs whose 3-periodic net is minimal (their cycle basis is of rank 3), the projection is reduced to the identity, I_m . For example, the **dia** net has 3 vectors as a basis of its cycle space, as described above and is thus minimal in 3-periodic representations. We can therefore construct its metric tensor (Z) using the 3 cycles forming the basis, $\mathcal{B}_c$ and Eqn. 7.4.

$$\begin{aligned} Z_{\text{dia}} &= \begin{pmatrix} -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{pmatrix} \cdot I_4 \cdot \begin{pmatrix} -1 & -1 & -1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix} \end{aligned}$$

This indicates 3 lattice vectors of equal length, $(\sqrt{2})$, and angles between vectors of 60° which describes the primitive cell of **dia**. For nets not minimal in 3-periodic representations such as **qtz**, which has a cycle basis of rank 4, K is constructed as follows. A lattice basis is constructed in the cycle space, in the **qtz** graph, $\mathcal{L}_A$ is constructed using the cycles $L_1 = (e_3 - e_4)$, $L_2 = (e_1 - e_6)$, and $L_3 = (e_4 + e_6 + e_5)$ which correspond to the linearly independent lattice

directions (100), (010), and (001). The lattice vectors obtained from the remaining cycles which form the basis of the cycle space, $\mathcal{B}_c$, can then be expressed as a linear combination of the lattice basis. In the case of **qtz**, a single remaining cycle $L_4 = (e_2 - e_5)$, whose sum of edge labels is $(\overline{110})$ can be expressed as a linear combination of the first two lattice vectors, $L_4 = -L_1 - L_2$. Thus one can construct a cycle vector whose edge labels sum to (000) by summing the above cycles $\mathcal{D} = L_4 + L_1 + L_2$. The cycle, expressed in the edge space, $(e_1 + e_2 + e_3 - e_4 - e_5 - e_6)$ along with the two co-cycle vectors, complete the kernel of the projection.

$$K_{\text{qtz}} = \begin{pmatrix} 1 & 1 & 1 & -1 & -1 & -1 \\ 1 & 0 & -1 & -1 & 0 & 1 \\ -1 & 1 & 0 & 0 & 1 & -1 \end{pmatrix}$$

Where the first row represents the cycle $\mathcal{D}$, and the last two rows are the co-cycle vectors, representing outward oriented arcs incident on vertices A and B. Using equation (3) the projection is then,

$$\mathcal{P}_{\text{qtz}} = \begin{pmatrix} 1/2 & 0 & 0 & -1/3 & -1/3 & 1/6 \\ 0 & 1/2 & 0 & -1/3 & 1/6 & -1/3 \\ 0 & 0 & 1/2 & 1/6 & -1/3 & -1/3 \\ -1/3 & -1/3 & 1/6 & 1/2 & 0 & 0 \\ -1/3 & 1/6 & -1/3 & 0 & 1/2 & 0 \\ 1/6 & -1/3 & -1/3 & 0 & 0 & 1/2 \end{pmatrix}$$

We are assuming a barycentric placement of the vertices by including the co-cycle vectors in the kernel of the projection. Using the projection matrix $\mathcal{P}$ above and Eqn. 7.4, the metric tensor for the **qtz** net is

$$Z_{\text{qtz}} = \begin{pmatrix} 4/3 & -2/3 & 0 \\ -2/3 & 4/3 & 0 \\ 0 & 0 & 3/2 \end{pmatrix}$$

Using the metric tensor and the lattice representations of the lines (Ω^+) a cartesian barycentric representation of these nets can be defined. The Cartesian representations for the

labelled quotient graphs for **dia** and **qtz** are shown in Figure 7.4 where the **a** lattice vector is placed along the x-axis and the **b** vector is fixed in the x-y plane.

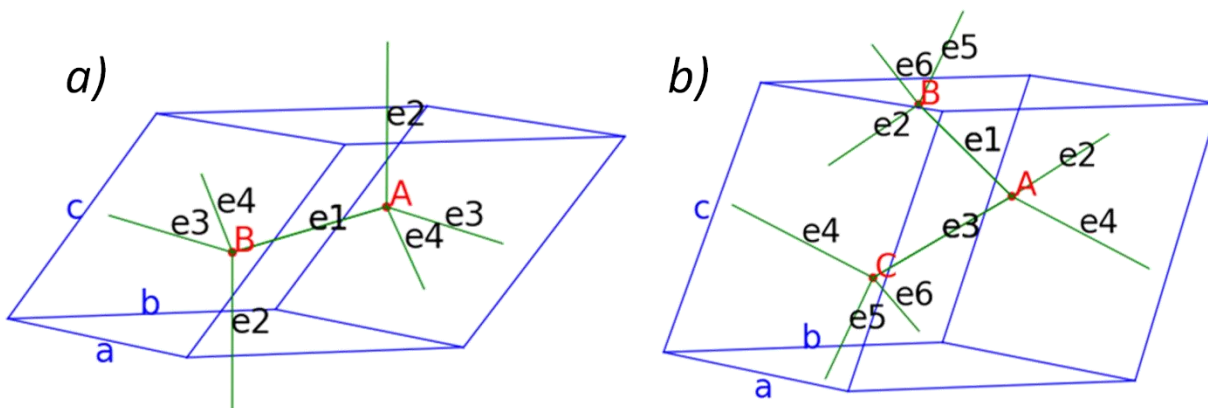


Figure 7.4: A barycentric embedding of the labelled quotient graph of **dia**, b) A barycentric embedding of the labelled quotient graph of **qtz**. Each was constructed with the algorithm.

This section has demonstrated the method in which one can obtain a representation of a labelled quotient graph as a net in Cartesian space. We generated the barycentric embeddings of the nets **dia** and **qtz** could be found from knowing the lattice (L) vectors and setting the co-lattice (L^*) vectors equal to zero. However we are interested in using these nets as structure directing, subject to the local geometries governed by the chemistry in SBUs. This will require adjustments of the metric tensor and deviations of co-lattice vectors outside of their barycentric values. Subsequent sections will show how to obtain non-barycentric representations from labelled quotient graphs, such that local geometries of embedded nets match those of desired SBUs.

7.4.3 Constructing Non-Barycentric Net Embeddings from SBUs.

In the previous section it was shown how to construct a barycentric embedding by setting the co-lattice vectors to zero. Here, we will set them as variables which will depend on the SBU geometries. Thus we re-define the co-lattice vectors with the set of variables X ;

$$L^*(X) = \{(x_i^1, x_i^2, x_i^3) | 1 \leq i \leq n - 1\} \quad 7.6$$

Recall the co-lattice vectors are included in the last $n-1$ entries of the lattice-co-lattice matrix, α . For example, consider the simple case of the **dia** net discussed in the previous section. There is a single co-lattice vector in the lattice-co-lattice matrix, which we define with $x_1 = (x_1^1, x_1^2, x_1^3)$, giving

$$\alpha_{\text{dia}}(X) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ x_1^1 & x_1^2 & x_1^3 \end{pmatrix}$$

This affords a unique embedding of the net in the lattice space. From Eqn. 7.2,

$$\Omega^+(X) = B^{*-1}\alpha(X) \quad 7.7$$

Because the SBU geometries will dictate the final shape of the unit cell, a second set of variables are defined for the unit cell $P = (a, b, c, \alpha, \beta, \gamma)$ which can be represented in the metric tensor;

$$Z(P) = \begin{pmatrix} a^2 & a \cdot b \cdot \cos(\gamma) & a \cdot c \cdot \cos(\beta) \\ a \cdot b \cdot \cos(\gamma) & b^2 & b \cdot c \cdot \cos(\alpha) \\ a \cdot c \cdot \cos(\beta) & b \cdot c \cdot \cos(\alpha) & c^2 \end{pmatrix}$$

Where a , b and c are the lengths of the lattice vectors $(\vec{a}, \vec{b}, \vec{c})$, and α is the angle between $\vec{b}$ and $\vec{c}$, β is the angle between $\vec{a}$ and $\vec{c}$, and γ is the angle between $\vec{a}$ and $\vec{b}$.

This set of variables X and P will permit the embedding of a net which supports the geometries of a given set of SBUs. To demonstrate, we use an example using the two SBUs shown in Figure 7.5, an octatopic Cd^{2+} inorganic SBU and a ditopic biphenyl organic SBU. We will construct a hypothetical MOF with these two SBUs using the net **bcu**, which contains a single vertex of degree 8 connected to itself by four loops. An initial step is to assign SBUs to specific vertices within the labelled quotient graph. This can be accomplished by matching vertex and SBU coordination numbers. In the case of **bcu**, this task is straight forward; there is a single node with eight edges incident upon it. Thus the 8-coordinate Cd^{2+} SBU in Figure 7.5a

will be assigned to the vertex. In order to account for particular bonding geometries of SBUs, the labelled quotient graphs are augmented with additional vertices and edges. Specifically, additional vertices of degree 2 are added in the edges between each original vertex. The extra edge between these vertices represents the bond made between adjacent SBUs to account for bond orientations which cannot be represented by a single edge originating from the SBUs centre of mass. An example of this can be seen in Figure 7.5a for the Cd^{2+} inorganic SBU. The underlying geometry of the SBU and its connection sites can best be described by a vertex of degree 8 located at the SBU centre of mass, and 8 vertices of degree two which describe the bonding orientation of the carboxylate moieties.

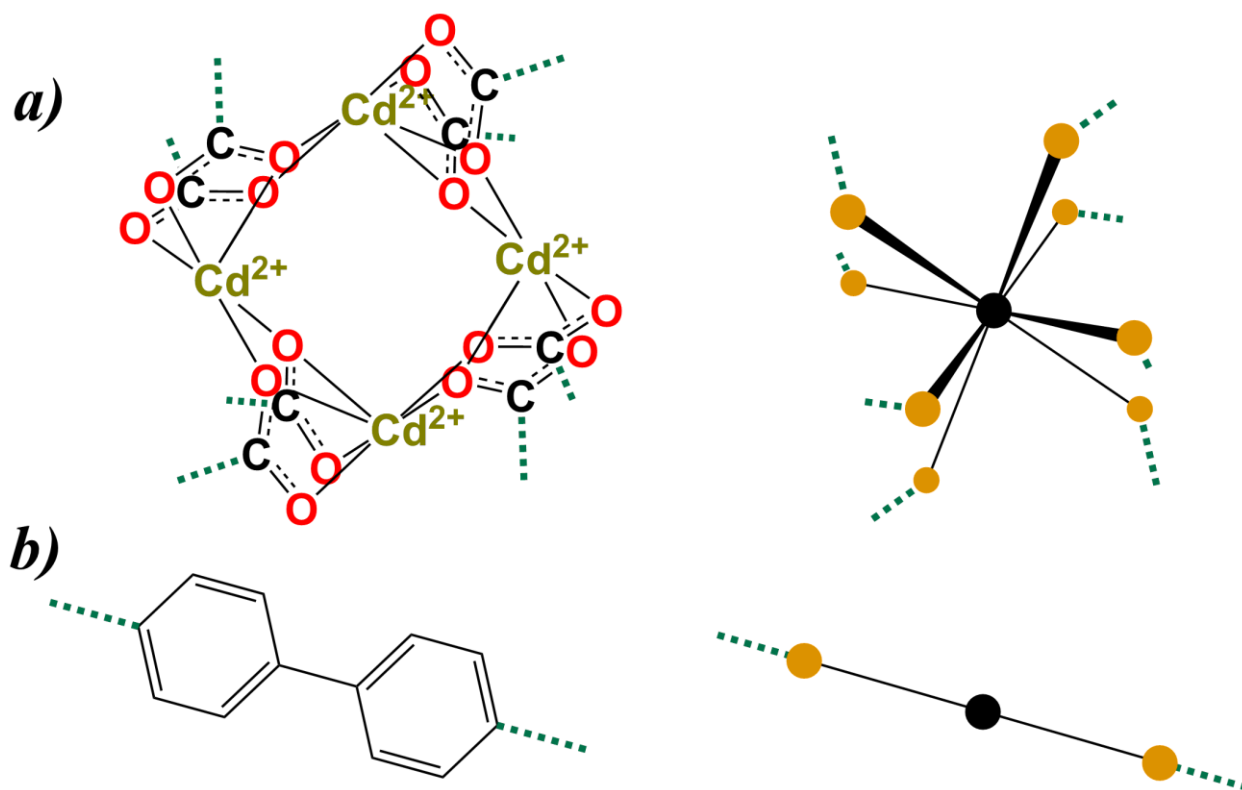


Figure 7.5: a) An octatopic inorganic SBU along with a b) linear organic SBU and their underlying vertex representations. Note that due to the bond orientations of the carboxylate groups to the Cd^{2+} ions, the underlying representation is best described by 9 nodes and 16 lines; A single node at the centre of mass of the SBU, and 8 nodes located at its connection sites.

It should be noted that vertices of degree 2 are typically not found in crystallographic nets, as they can be substituted for a single edge between vertices of a higher degree. However, in the case of structure generation, we find it necessary to include these vertices to ensure a good geometric match with the SBUs. For example, all of the zeolite topologies possess quotient graphs with vertices of degree 4, which represent the tetrahedral Si^{4+} ions within the cell. The divalent oxygens, which are typically bent from linearity with a 145° angle are reduced to edges between tetravalent vertices. To construct a zeolite crystal from the quotient graph, these divalent vertices would need to be re-inserted to describe a bent oxygen between the tetrahedral Si atoms. By inserting these vertices of degree 2, we do not alter the topology of the underlying graph. To ensure this, any new edges are assigned a label of '000' such that the sum of edge labels in the

cycle basis yields the same value as the cycle in the un-augmented form. A specific example is shown in Figure 7.6b. The labelled quotient graph of **bcu** is augmented with several new vertices and edges to account for the linear organic SBU and the bond between the inorganic and organic SBUs.

With each SBUs underlying geometry assigned to specific vertices and edges of the augmented labelled quotient graph, the geometry of its net embedding can now be compared to the SBU geometries. Each SBUs underlying geometry, consisting of n vectors, can be represented by the set of inner products

$$S_{SBU} = \{\mathbf{v}_p \cdot \mathbf{v}_q | (1 \leq p \leq n), (q \leq p)\} \quad 7.8$$

This includes every dot product between each vector describing the SBU's underlying geometry. The Cd SBU's underlying geometry (shown in Figure 7.5a right) can be represented by the set of inner products of the $n=16$ vectors, pointing away from the vertex at the centre of mass of the SBU. S_{SBU} is therefore a set of 136 dot products between these 16 vectors.

Likewise, the geometry of a net embedding can be described by a set of inner products of its lines. Given a specific Euclidean embedding of a net defined by a set of variables X and P , one can construct the inner product matrix of its lines using Eqn. 7.9,

$$g = \Omega^+(X) \cdot Z(P) \cdot [\Omega^+(X)]^t \quad 7.9$$

which is a symmetric matrix whose diagonal entries are the squared lengths of the lines, $g_{ii} = \|\vec{e}_i\|^2$ and off-diagonal entries can be considered the angular component between two different lines, $\vec{e}_i$ and $\vec{e}_j$; $g_{ij} = \|\vec{e}_i\| \|\vec{e}_j\| \cos\theta$. This affords a representation of the line geometries without the need to represent the vectors in a Cartesian coordinate system. The only elements needed to generate g are the matrices defining an embedding of the net in Euclidean space,

namely the cycle-co-cycle matrix B^* , the lattice-co-lattice matrix $\alpha(X)$, and the metric tensor, $Z(P)$.

We can now construct an objective function in which to optimize. We define elements g_{ij} of the matrix g as belonging to a particular SBU if the edges e_i and e_j in the labelled quotient graph were assigned to that SBU. For example, in the case of the **bcu** topology, the Cd^{2+} SBU is assigned all of the arcs connected to the central vertex A in Figure 7.6b as well as the adjacent arcs representing the SBU's connection sites. Suppose the n edge indices assigned to a particular SBU are found in the set $a = (a_1, \dots, a_k, \dots, a_n)$, the elements of g associated with an SBU would be,

$$N_{SBU} = \{g_{a_r a_s} | (1 \leq r \leq n), (s \leq r)\} \quad 7.10$$

In the case of the Cd^{2+} SBU, there are 16 arcs assigned to it in the augmented quotient graph of **bcu**. Thus $n = 16$ and the set $N_{\text{Cd}^{2+}}$ runs over 136 elements of the net embedding's inner product matrix.

Because the geometry of the SBUs can be defined using Eqn. 7.8, and we know the elements of the inner product matrix, g , associated with each SBU, a fitting metric can be constructed, summing over all the K elements in the associated sets from Eqns. 7.8 and 7.10.

$$M_{SBU} = \sum_{i=1}^K (N_i - S_i)^2 \quad 7.11$$

It should be noted that when constructing the geometry of the reduced SBU representation in Eqn. 7.8, the orientations of the vectors must match the positive orientation of the associated arc in the labelled quotient graph. For example, the arc e_9 in Figure 7.6b points toward the vertex A. Thus the vector associated with e_9 in the underlying SBU representation of the Cd^{2+} SBU must be oriented into the SBUs centre of mass (which was assigned to the vertex A).

We formally define an objective function - for all V SBUs assigned to vertices and edges of the labelled quotient graph, we minimize,

$$O = \sum_{j=1}^V M_j \quad 7.12$$

which is the sum squared differences between the inner products of the vectors describing the underlying SBU geometries and their associated lines in the embedded net. This objective function is non-linear with respect to the variables X and P , and thus requires an optimization scheme to solve for each net and set of SBUs. At each step of the optimization, we construct an embedding of the net by adjusting the two sets of variables, the set X defining the co-lattice vectors, $L^*(X)$, and the set P which define the metric tensor $Z(P)$. Upon optimizing the objective function, the resulting net embedding defined by the variables can be used as a template to orient the SBUs to their assigned nodes, yielding a periodic MOF structure.

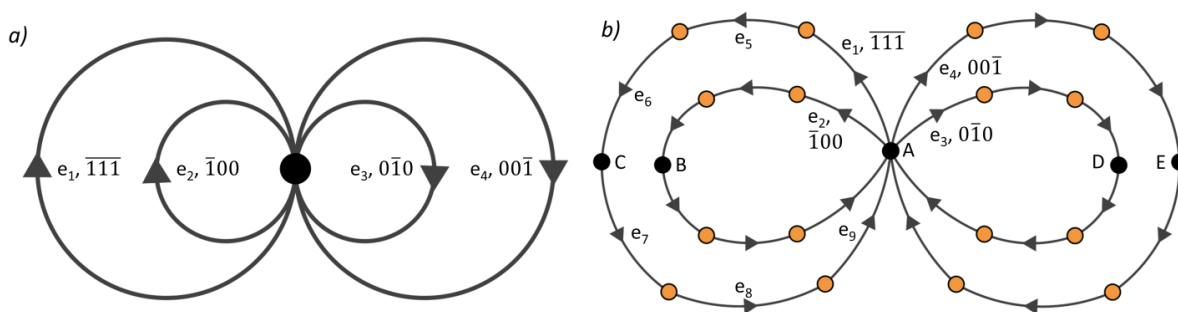


Figure 7.6: Augmenting the labelled quotient graph to support the SBU connectivities. a) The original quotient graph of **bcu** and b) its augmented form. The cycle spanning e_1 in the original graph (loop) is now a cycle spanning $e_1 + e_5 + e_6 + e_7 + e_8 + e_9$ in the augmented graph, both cycles edge labels sum to $\overline{111}$. The orange vertices represent connection points between adjacent SBUs and the black vertices represent the centre of mass of the SBUs. The additional edges included in the augmented graph are all given a label of 000.

Figure 7.7 shows a graphical progression from the barycentric embedding of the net **bcu** as taken from the RCSR to the embedding which fits the SBUs from Figure 7.5. The barycentric representation of **bcu** is shown in Figure 7.7a, the augmented form with lines and nodes added to accommodate SBU connection sites is seen in Figure 7.7b. This is followed by an optimization

of the X and P parameters of the net in Figure 7.7c, such that its geometry best accommodates the SBUs. Finally, the SBUs are oriented onto their associated vertices to construct the final hypothetical MOF in Figure 7.7d.

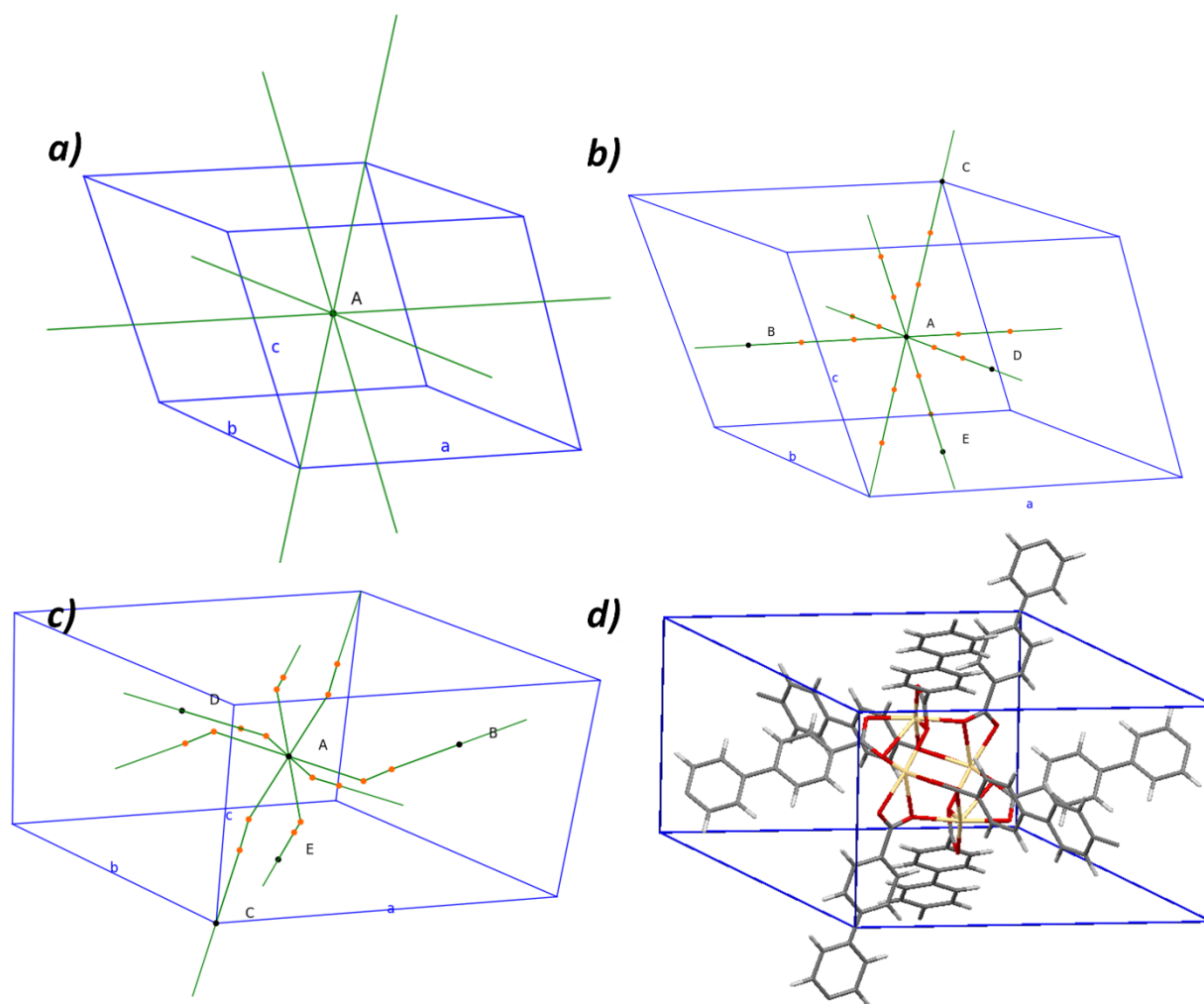


Figure 7.7: The graphical representation of progressing from the a) barycentric representation of the net **bcu** to b) an augmented form of the net with orange vertices representing SBU connection sites, black vertices representing the centre of mass of the SBUs. c) The **bcu** net is adjusted into a non-barycentric form to match the geometry of the SBUs in Figure 7.5. The SBUs are super-imposed on the net vertices to form the final hypothetical MOF structure.

7.4.4 Implementation

The MOF building program was written primarily in python, with parts in C++ to interface with the NLOPT optimization code.²⁸ The program reads in each nets' labelled quotient graph information as a single string, using the format discussed by Delgado-Friedrichs and

O’Keeffe.¹² The string describes all of the arcs in the graph; each arc is formatted ‘v1 v2 L1 L2 L3’ where v1 is the vertex origin of the arc, v2 is the terminal vertex of the arc, and L1 L2 L3 constitutes the label of the arc (e.g. 001). From this description of the labelled quotient graph, we construct the nets metric tensor, Z , and α , and B^* matrices from its cycle ($\mathcal{B}_c$) and co-cycle ($\mathcal{B}_s$) bases. The co-cycle bases can be constructed by summing the outward orienting arcs from the first $n-1$ vertices. The cycle basis is constructed from a minimum spanning tree algorithm^{29–32} from the SAGE mathematical package.³³ The minimum spanning tree of a graph is the smallest number of edges which can form a walk through all vertices. Each edge *not* in the minimum spanning tree can be added to it to produce a cycle, and the set of cycles formed this way from all of the edges not in the minimum spanning tree form a basis of the cycle space.

SBUs are chosen if they match the degree of the net, based on the number of connection sites they possess. Upon matching SBU with nets of the same degree, the labelled quotient graph is augmented to introduce more vertices and edges so they can be matched to each SBU’s centre of mass and connection sites. If linear two-connected SBU are used, they will be accommodated by inserting 3 divalent vertices into a single edge from the labelled quotient graph; a central vertex for the SBU’s centre of mass, and two vertices for its connection sites. The edges split in this manner are typically between vertices assigned to opposing SBU type. For example an edge joining two inorganic SBU will be split to support a linear organic SBU.

SBU and their connection sites are assigned to specific vertices and arcs in the labelled quotient graph using the barycentric embedding of the net for comparison. Here every possible attempt is made to assign the lines of a node to the connection sites of an SBU. The final assignment is chosen by selecting the one with the lowest root mean squared score of the difference between their (normalized) geometries. It is recognized that one can leverage

symmetry elements found in the labelled quotient graph (automorphisms) to assign arcs and vertices to SBUs (assuming the point group symmetry of the SBU is known). This method is currently not implemented, but could be implemented in the future.

With the vertices and arcs assigned to specific SBUs and connection sites, the net embedding is optimized to the geometries of the SBUs using the objective function defined in Eqn. 7.12. Because this is a non-linear optimization problem, the best general approach is to perform a global optimization of the variables. However, in many cases the SBU geometries do not deviate significantly from the barycentric representation of the net, thus in this study a local optimization was used in the form of the L-BFGS algorithm^{34,35} as implemented in the NLOPT code²⁸. At the first step of optimization, the program uses the variables X and P corresponding to the barycentric representation. That is, $X = \{x_i = (0,0,0) | 1 \leq i \leq n - 1\}$ and the values of the cell parameters are extracted from the barycentric derivation of the metric tensor using Eqn. 7.5. We say that the optimization has converged when the objective function described in Eqn. 7.12 does not differ by 10^{-5} in consecutive optimization steps.

Following optimization, the SBUs are oriented onto their assigned nodes to complete the hypothetical MOF construction. This orientation consists of a translation of the SBU's centre of mass to the central node, and a rotation to least squares fit the connection sites to the appropriate nodes.³⁶ At this stage, an atom collision check is performed to ensure that no atoms are closer than the sum of their van der Waals radii multiplied by a user-defined tolerance factor. In this work, the tolerance factor was set to 0.4.

7.5 Results

7.5.1 Fitting Different 4-c SBUs to Nets

As an initial demonstration of the methodology, we will construct periodic structures of different topologies using two four coordinate (4-c) SBUs shown in Figure 7.8a and b. The

copper paddlewheel can be considered a square planar SBU with angles of 90° between adjacent edges. The organic SBU can be considered tetrahedral-like, which has four angles of 116.5° and two angles of 96.1° . The 3 topologies taken from the RCSR were chosen for instructive purposes; the two nets **dia** and **qtz** whose labelled quotient graphs are shown in Figure 7.3, and the net **pts**. The first two nets have only tetrahedral-like vertices in their barycentric form, thus it represents a challenging case to fit the square planar Cu paddlewheel to these nets, as it will require significant distortion of the net. The net **pts** contains both square planar and tetrahedral like vertices in its barycentric representation, thus it will be less challenging to fit the net to these SBUs. Moreover, this net is known to support geometries of the first two SBUs described in Figure 7.8 and is thus likely to give the best fit to the SBUs when optimizing the net geometry.^{37,38} The labelled quotient graphs of **dia** and **pts** are classified as bipartite, which means each graph's vertices can be separated into separate sets based on their connectivity. This makes it easy for SBU placement, as the inorganic SBUs can be placed in one set of vertices and the organic SBUs can be placed in the other set, ensuring that organic and inorganic SBUs connect only to each other. However the **qtz** net produces a challenge for the choice of SBU placements. It is not bipartite and so we must relax the rule that organic and inorganic SBUs only join to each other and not themselves. In this case only organic SBUs were placed on adjacent vertices. This effectively creates a larger, more complicated organic SBU in the hypothetical MOF.

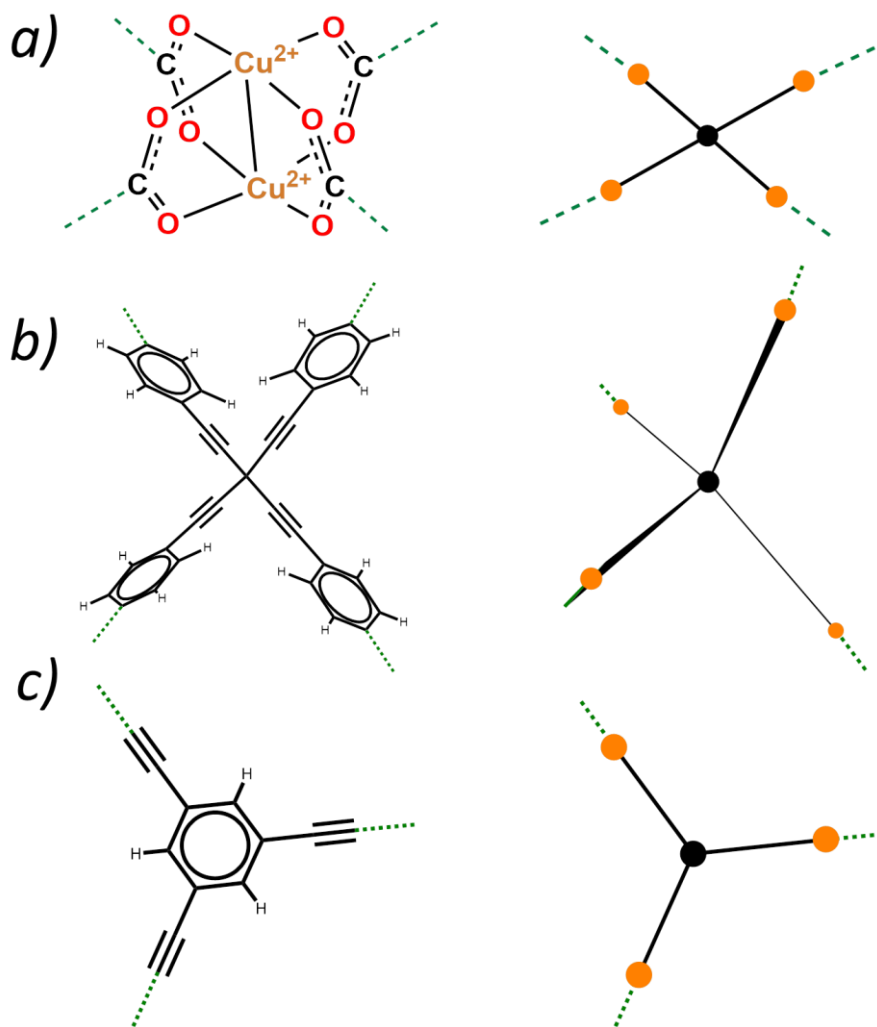


Figure 7.8: SBUs used as input into the structure generation program and their underlying vertex representations. a) A 4-c square planar copper paddlewheel SBU, b) a 4-c tetrahedral organic SBU and c) 3-c trigonal planar organic SBU.

Fitting of these nets to SBUs is not always perfect for a number of reasons. The first being the program attempts to match a net to a set of rigid SBU geometries, which, due to the constraints of their angles, may not support a particular periodic connectivity. A discussion of these types of geometries can be found in ref 39. Another reason is due to the constraints set on the optimization parameters. For example, when the program attempts to fit a net to SBU geometries, the lattice vector angles are not permitted to deviate lower than 60° and greater than 120° . For these reasons, a ‘fitness’ of the embedded net is calculated when the convergence

criteria of the program have been met. The fitness is computed as the average deviation of the net geometry, in terms of its line lengths (line length deviation shown in the first row of Table 7.1) and angles (angle deviation shown in the second row of Table 7.1) compared with underlying geometry of the rigid SBUs the program fits the net to. Upon evaluating the net's fitness, one can use discretion on whether to keep or discard the structure. In some cases the fit can be poor, but produce a reasonable structure after relaxing the atoms and bonds by optimization with a molecular mechanics force field or *ab initio* calculation. Table 7.1 reports how well each net embedding fits to the rigid SBUs. While the averages of line length deviations were relatively small across all topologies, the standard deviations appear to be the lowest for the **pts** net. For both the **dia** and **qtz** nets there is considerable angle deviations, which can be attributed to attempting to fit SBUs of square planar geometry to nets which support solely tetrahedral-like vertices.

To demonstrate how 'unhappy' these materials are after being generated from the program, the stress tensors were computed on the unit cell of each structure from a single point DFT calculation in VASP.^{40–43} Comparison of the stress tensors show the **pts** net provides the best fit of the SBUs. It is clear that the significant distortions required to fit the **dia** and **qtz** nets to the square planar Cu paddlewheel SBU has resulted in some unphysical stresses on the system. This agrees qualitatively with the angle and edge length deviations reported by the program.

Table 7.1: Measurements of the fit of three nets to the organic and inorganic SBUs shown in Figure 7.8a and b, respectively. The stress values are computed from a single-point energy calculation in VASP. No external potential was applied to obtain these values, they arise from the SBU bonding.

<i>topology</i>	<i>dia</i>	<i>qtz</i>	<i>pts</i>
<i>line length deviation</i> (Å)	1.1 ± 1.2	1.4 ± 1.4	-0.8 ± 0.7
<i>angle deviation</i> (degrees)	12.1 ± 17.7	12.2 ± 14.7	1.2 ± 4.0
<i>Stress XX (kbar)</i>	-15.6	-19.0	0.9
<i>Stress YY (kbar)</i>	-28.2	-20.3	0.9
<i>Stress ZZ (kbar)</i>	-24.2	-23.9	1.8
<i>Stress XY (kbar)</i>	2.9	-3.1	0.0
<i>Stress YZ (kbar)</i>	5.5	9.9	0.0
<i>Stress ZX (kbar)</i>	-5.4	-5.4	0.0

While not investigated here, it is possible that expanding the labelled quotient graph (or reducing the translational symmetry) so that the ‘unit cell’ of the net contains more vertices and edges to support more SBUs may reduce the stress values shown in Table 7.1. For example, 3 of the 4 arcs in the labelled quotient graph of **dia** correspond to the three lattice vectors 100, 010, and 001. Since these three arcs are assigned to a square planar paddlewheel, there is considerable pressure to make these lattice vectors coplanar. This can be seen by the non-zero stress values in the XY YZ and ZX values for **dia**. By increasing the number of SBUs in the unit cell, we can effectively distribute this stress over a larger number of bonds and degrees of freedom. We note that the DFT-optimized **dia** structure does indeed have a very flat two-dimensional appearance; while all the stress tensors become 0, two of the three lattice angles optimize to less than 60°. Similar arguments can be made for the **qtz** net whose lattice vectors are determined by a sum of only two or three arcs. It is notable that the stress values in the XX, YY, and ZZ directions are

negative for both the **dia** and **qtz** MOFs, indicating a drive towards volume compression of the unit cells. The connectivity of these structures indicates that at least two connection sites of the tetrahedral SBU in Figure 7.8b are connected to two sites of the same Cu paddlewheel. While the tetrahedral orientation of the organic SBU is provided by a single sp^3 hybridized carbon, the square planar angles of the Cu paddlewheel are much more rigid, being supported by two Cu^{2+} ions coordinated to conjugated carboxylate moieties. Therefore the compression likely arises from the forced planarity of the tetrahedral SBUs.

7.5.2 Structures built from the RCSR

To demonstrate the breadth of structures the program is able to generate, the (4-c) copper paddlewheel and a (3-c) SBU shown in Figure 7.8 a and c, respectively were used to generate 46 MOFs unique in their topology. Note, the number of nets with mixed degree (3,4) in the RCSR is higher than this value, however we eliminated those generated structures containing atomic collisions. Figure 7.9 plots the range of surface areas obtained from the 46 structures generated with the program. Many of the topologies shown along the x-axis of Figure 7.9 are not bipartite, this means that in some of the MOFs Cu paddlewheels are bonded to Cu paddlewheels, and in others the 3-c SBU is bonded to other 3-c SBUs. All of these structures were optimized using a molecular mechanics force field prior to computing their surface areas.

It is interesting to see such a diversity of surface areas in Figure 7.9 achieved from building MOFs with only two types of SBUs. As a rough comparison, the database of over 130,000 MOFs constructed by Wilmer *et al.* possesses a similar distribution of gravimetric surface areas (from approximately 0 – 8000 m^2/g).⁴⁴ However the structures in that database were assembled with 102 building blocks¹, and consisted of only 6 topologies.² This method therefore has the potential to construct a much more diverse database of structures holding the number of building blocks constant.

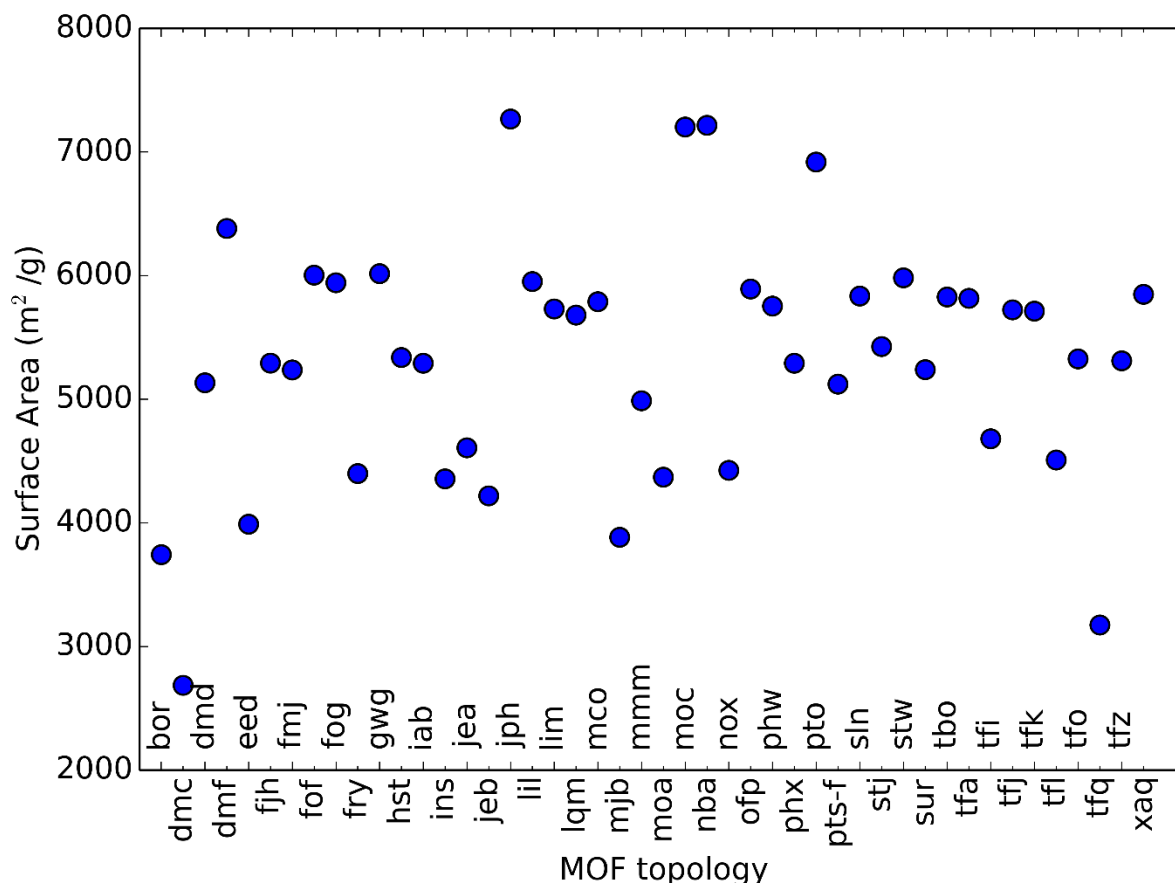


Figure 7.9: Specific surface area of 46 MOFs assembled from topologies found in the RCSR using the 4-c Cu paddlewheel and 3-c SBU shown in Figure 7.8 a and c respectively.

By design, the program forms MOFs with specific topologies by performing the appropriate net isomorphism to best match the SBU geometries. Thus it can be said that the topologies of the hypothetical MOFs generated with this program will be the same as the underlying net initially used as structure-directing. To test this, these structures were subjected to the TOPOS program,¹⁷ which abstracts chemical systems to their underlying nets and classifies them against the TOPOS Topological Database (TTD). This is a standard package commonly used to identify the topology of crystalline materials.⁵ As expected the TOPOS program provided the same topology for each hypothetical MOF structure as was used to build it.

7.5.3 Diverse MOFs Built with the *tbo* Topology

Delgado-Friedrichs and O’Keeffe demonstrated that the highest possible symmetry of a net can be achieved when the vertices are in their barycentric placement.⁹ The space groups reported for each net in the RCSR database are indeed those of the maximal achievable symmetry group of the net. However, there are many cases where the barycentric representation is not appropriate. As an example, consider the bipartite net **tbo**, which is associated with the MOF HKUST-1.⁷ The ideal space group of the net **tbo** is Fm-3m, which supports 3-c vertices with site symmetry 3m (C_{3v} in Shoenflies notation), and 4-c vertices of mmm symmetry. Now suppose one wanted to construct a hypothetical material using SBUs of significantly lower point group symmetry than those supported by the idealized space group. Consider the tritopic 1,3,5-benzenetricarboxylate (BTC) SBU used to build HKUST-1 shown in Figure 7.10a. The point group symmetry of this molecule is D_{3h} as there are three dihedral mirror planes present in the molecule as well as a 3-fold rotation axis. The SBU is supported by the ideal symmetry operations of the net as there is a 3-fold rotation axis at the 3-c vertex. If, however, we were to instead construct a material using the **tbo** topology with the molecule biphenyl-3,4',5-tricarboxylate (BPTC), we have effectively reduced the point group symmetry of the tritopic SBU to C_{2v} . It is important to stress here that the symmetry is not only of the atomic structure, but also of its underlying connectivity. If one were to abstract the biphenyl SBU seen in Figure 7.10c to vertices and edges, there would be one long edge along the principle axis of rotation, and two small edges pointing in the direction of the carboxylates located on the same benzene. Any symmetry operation supporting a three-fold rotation must therefore be scrapped in the unit cell. To test the limits of generating structures in this way, we also build a structure using a tritopic SBU with no rotational symmetry in the molecule or the underlying connectivity. This was accomplished by shifting the lone carboxylate group of BPTC over one hydrogen to form

biphenyl-3,3',5-tricarboxylate seen in Figure 7.10e. In our current implementation, no symmetry elements are applied when constructing the hypothetical materials, so it does not pose a challenge to construct these materials. While no symmetry operations were utilized in our program to build the structure, we computed the resulting symmetry of the hypothetical MOFs using a symmetry detection algorithm⁴⁵ which finds the appropriate space group of a crystal system by finding equivalent atoms using a set of symmetry operations from a lookup table. Equivalent atoms were found with a tolerance of 0.4 Å.

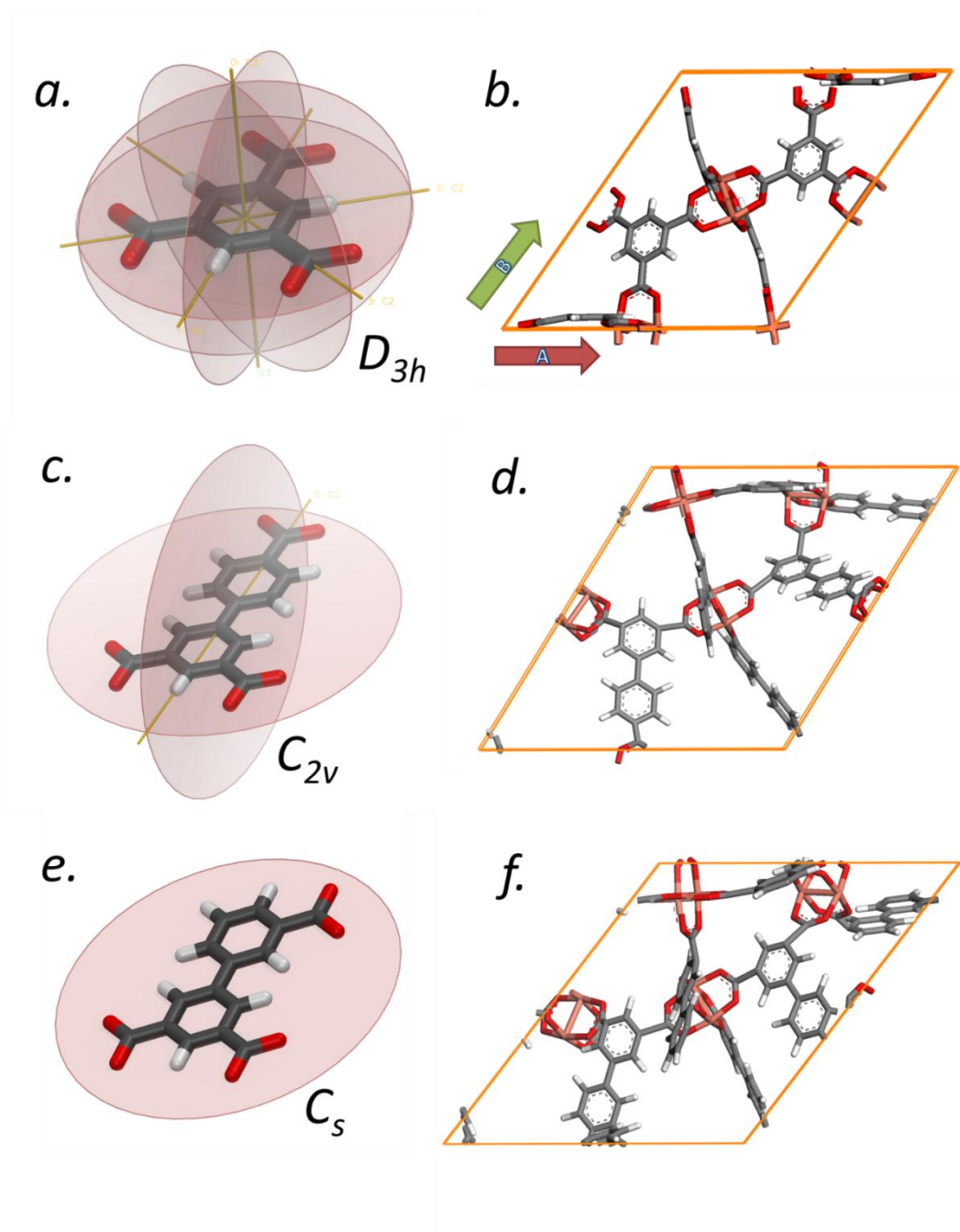


Figure 7.10: The generation of three different hypothetical structures based on the same underlying topology, **tbo**, using three different tritopic SBUs of different symmetry. a) The highly symmetric benzene-1,3,5-tricarboxylate (BTC) SBU and b) the primitive cell of the hypothetical MOF generated by the program, viewing down the c axis. c) The lower symmetry biphenyl-3,4',5-tricarboxylate and d) its hypothetical structure. e) the low symmetry biphenyl-3,3',5-tricarboxylate and f) its hypothetical structure. The point groups of the SBUs are displayed with the mirror planes as purple circles and rotation axes as yellow rods.

Figure 7.10b shows the generation a hypothetical version of HKUST-1 MOF. The structure was generated with a negligible line length deviation and an average angle deviation of 2.0 ± 3.3 degrees between the SBUs and the final embedded net. In addition, the symmetry detected in the generated structure was the ideal symmetry space group found in the barycentric representation of the net, Fm-3m. For the structure built with the SBU with C_{2v} symmetry in Figure 7.10d the resulting net embedding found from our algorithm, when compared with the SBU geometries, was reported to possess an average line length deviation of 0.3 ± 0.4 Å and angle deviation of 6.6 ± 10 degrees. The reported crystal system was found to be monoclinic C2, which possesses two-fold screw axis and a two-fold rotation axis. Both symmetry operations support the symmetry of the C_{2v} SBU. The structure in Figure 7.10f built from the low symmetry SBU (Figure 7.10e) was found to have P1 symmetry, as expected. The line length and angle deviations reported from generating this material was 0.1 ± 0.4 Å and 7 ± 10 degrees, respectively. Crystalline structures with this level of strain and low symmetry will likely never be experimentally realized, however generation of this material demonstrates the robustness of the structure generation code. The displayed structures have not been optimized in any form, they are a direct output of the structure generation program.

7.5.4 Building MOFs with Unstable Nets

The advantage this program possesses over other methods is that it can be extended beyond the nets found in the RCSR in their barycentric representation. As an example, we will build a structure based on an unstable net $3(3^2,4)2$ which was taken from ref 25. Unstable nets, so-called because they possess intersecting vertices in their barycentric form, have been shown as important as descriptions of crystallographic structures. Recently Delgado-Friedrichs *et al.* discussed several crystalline materials which could be classified into nets with collisions in their barycentric placement.²⁰ Here we assemble a MOF with the copper paddlewheel and a tritopic

organic SBU shown in Figure 7.8a and c, respectively. The labelled quotient graph was manually interpreted into a string of integer values describing the edges and labels of the quotient graph, consistent with both the input required from the program, and the format discussed by Delgado-Friedrichs and O’Keeffe.⁹ The final hypothetical MOF built from this net is shown in Figure 7.11. It should be noted that the embedding of this net produced some bending within the structure, which can be expected when fitting rigid SBUs to an underlying graph. Optimization of these materials can yield structures with more improved bond angles and atom distances. The program TOPOS confirmed the topology of this MOF as $3(3^2,4)2$.

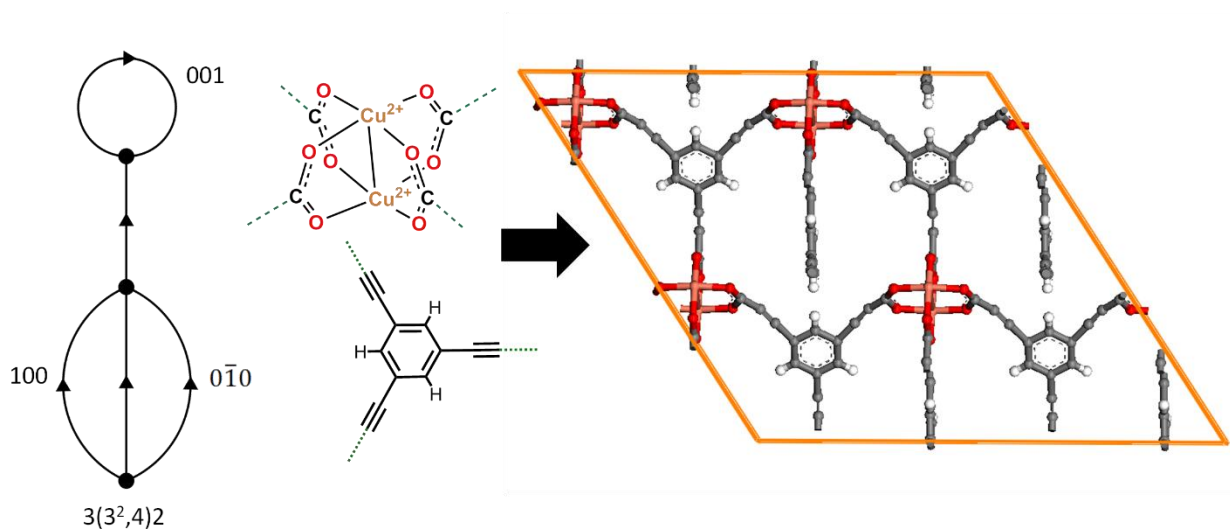


Figure 7.11: Building a hypothetical MOF based on the unstable net $3(3^2,4)2$. Left: the labelled quotient graph, middle: the SBUs used to build the structure, right: the hypothetical MOF.

7.5.5 Challenging Topologies and High Surface Area MOFs

Recently MOF-210 was reported with the **toz** topology, which possessed a record-breaking Brunauer-Emmet-Teller (BET) surface area of 6240 m²/g.⁴⁶ Building this MOF represents a challenge for current structure generation algorithms for several reasons. The MOF consists of 3 unique SBUs of different geometry; a hexatopic Zn₄O inorganic SBU found in other MOFs such as MOF-5,¹⁶ a 3-c organic SBU, 4,4',4''-(benzene-1,3,5-triyl-tris(ethyne-2,1-

diyl))tribenzoate (BTE), and biphenyl-4,4'-dicarboxylate (BPDC), which is a linear 2-c SBU. The rhombohedral primitive cell of this material is very large, possessing cell dimensions of 71 Å and containing a total of 18 Zn₄O inorganic SBU clusters, 18 BPDC SBUs, and 12 BTE SBUs. These factors alone would pose a particular challenge for the recursion method used in chapter 3 and by Wilmer *et al.*¹ which, in short, initiates a structure by placing a single SBU in space, and samples every possible combination of SBU bonds until a structure is made. To assemble the hypothetical version of MOF-210 with the recursion-based method, one would need to sample a very large number of possibilities before achieving the correct structure. Assuming that a 6-coordinate Zn₄O SBU is initially placed, then each of the 6 connection sites would sample the 3 bonds from the BTE organic SBU and the 2 bonds from BPDC, yielding $6^5=7776$ possible bonding combinations for the first SBU alone. There are a total of 48 SBUs in the primitive cell of this MOF, so the sampling quickly becomes unfeasible. In addition, the coordination environments of the Zn₄O SBUs are different. Each is coordinated to 4 tritopic BTE SBUs and 2 linear BPDC SBUs, however some Zn₄O SBUs coordinate BPDC at *cis* positions, others at the *trans* position. Thus, attempting to build this MOF would require an intractable amount of sampling to obtain the correct SBU bond combinations. The only way to obtain this MOF with the recursion method is to customize the code to recognize the numerous coordination environments of each SBU, and hand-parameterize each specifically for that particular position in the MOF. This task would allow us to assemble a limited number of novel hypothetical MOFs using only SBUs of similar geometry.

These problems can be alleviated with our proposed algorithm. Using the labelled quotient graph of the **toz** net as structure directing, we can identify where each SBU must be placed and what it must bond with to build the structure. All 6-c vertices are assigned to the

Zn₄O SBU and 3-c vertices are assigned to the BTE SBU. In cases where 6-c vertices in the graph are incident upon other 6-c vertices, we augment the edge between these vertices with an additional 2-c vertex, upon which we insert a linear ditopic BPDC SBU. To construct the hypothetical version of MOF-210, the organic SBUs were built in a conventional molecule builder. Unlike the SBUs found in the experimental crystal structure of MOF-210, these SBUs were built entirely planar, with typical C-C and C-H bond lengths and angles for aromatic and alkyne systems. The inorganic SBU was extracted from MOF-5¹⁶ which contains the same Zn₄O cluster, only connected in a different topology to different organic SBUs. Following generation, the raw structure produced from the algorithm possessed cell dimensions which did not agree with the experimental structure, and so to obtain the agreement in cell parameters shown in the first two columns of Table 7.3, the unit cell vectors were re-defined. In this process, we effectively generated a supercell from a linear combination of the original vectors to yield dimensions which more-or-less agreed with the experimental cell shape. These were linear combinations of the original cell vectors, as such the re-defined structures are in no way different from the output from the structure generation code, only their period was re-defined. The transformation matrices used to re-define the hypothetical versions of MOF-210 and NU-110 are shown in Table 7.2. The resulting larger hypothetical MOF was optimized with a molecular mechanics force field to yield the values reported in Table 7.3 as well as the agreement in the superposition of atoms of the hypothetical (blue) and experimental (red) structures shown in Figure 7.12b. Adjusting the cell dimensions in this manner does not change the inherent chemical and physical characteristics of the periodic structure. The presented values for surface area and pore volume in Table 7.3 were identical both before and after the cell parameters were

adjusted, and are in excellent agreement with the computed values for the experimental crystal structure.

Table 7.2: The cell vectors output from the generation program and the transformation matrices used to generate experimental-like cells.

	MOF-210	NU-110
original cell vectors (hypothetical)	$\begin{bmatrix} 51.4 & 0.0 & 0.0 \\ -25.7 & 44.5 & 0.0 \\ -0.1 & 29.9 & 63.8 \end{bmatrix}$	$\begin{bmatrix} 48.9 & 0.0 & 0.0 \\ -21.6 & 46.0 & 0.0 \\ -3.5 & -26.6 & 43.4 \end{bmatrix}$
transformation matrix	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & -2 & 3 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}$
transformed vectors (hypothetical supercell)	$\begin{bmatrix} 51.4 & 0.0 & 0.0 \\ -25.7 & 44.5 & 0.0 \\ -0.3 & 0.7 & 191.4 \end{bmatrix}$	$\begin{bmatrix} 48.9 & 0.0 & 0.0 \\ 27.3 & 46.0 & 0.0 \\ 23.8 & 19.4 & 43.4 \end{bmatrix}$
experimental crystal vectors	$\begin{bmatrix} 50.7 & 0.0 & 0.0 \\ -25.4 & 43.9 & 0.0 \\ 0.0 & 0.0 & 194.2 \end{bmatrix}$	$\begin{bmatrix} 48.6 & 0.0 & 0.0 \\ 24.3 & 42.1 & 0.0 \\ 24.3 & 14.0 & 39.7 \end{bmatrix}$

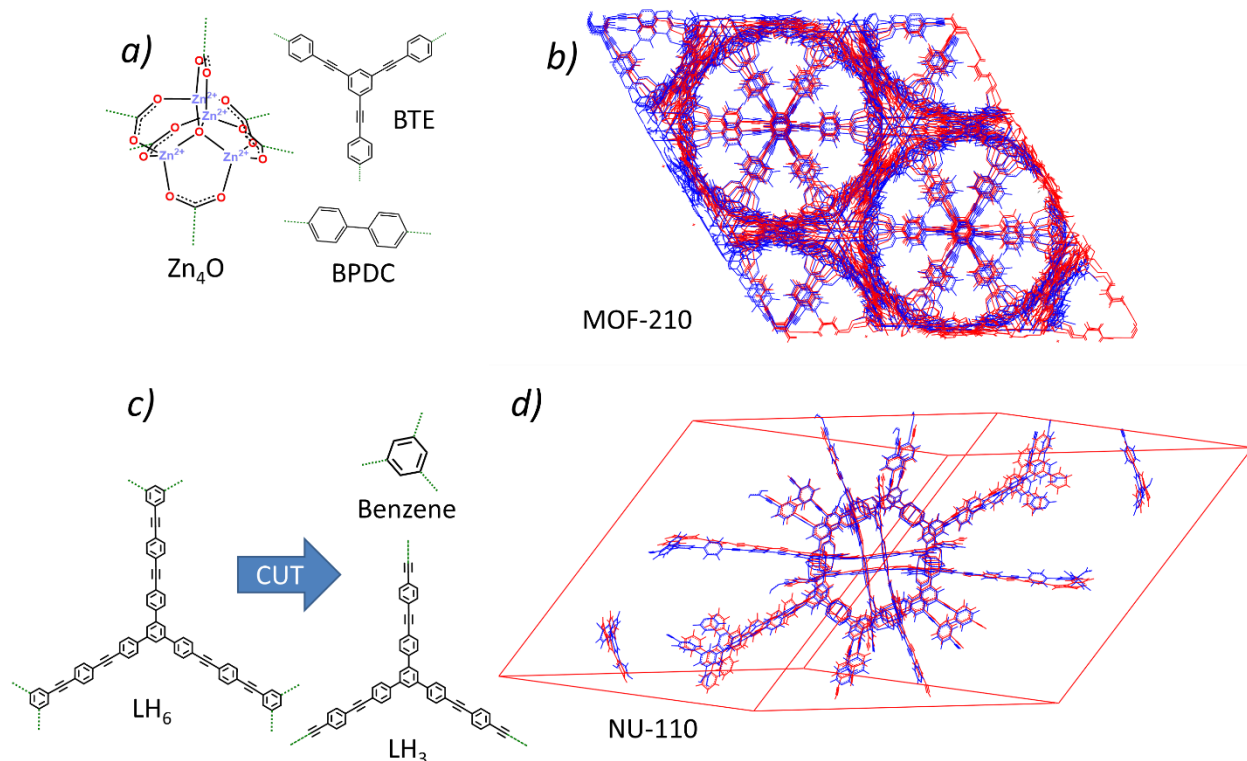


Figure 7.12: The hypothetical generation of two high surface area experimental MOFs. a) The SBUs used to assemble the hypothetical version of MOF-210, including the inorganic Zn_4O cluster, 4,4',4''-(benzene-1,3,5-triyl-tris(ethyne-2,1-diyl))tribenzoate (BTE) and biphenyl-4,4'-dicarboxylate (BPDC). b) Direct comparison of the hypothetical (blue) and experimental (red) MOF-210 structures. c) the organic SBUs used to assemble the hypothetical version of NU-110. The original 6-c LH_6 ligand (carboxylate moieties replaced with dotted lines) was fragmented into a 3-c benzene SBU, and 3-c LH_3 containing the remaining arms of the LH_6 ligand.⁴⁷ d) Superposition of the hypothetical (blue) and experimental (red) structures of NU-110.

Farha *et al.*⁴⁷ recently demonstrated a method for improving the surface area of an existing MOF, NU-100, based on the **ntt** topology. By increasing the length of the organic SBU arms with extra benzene and acetylene moieties, they were able to generate NU-110, which surpassed MOF-210 as the MOF with record breaking gravimetric surface area.⁴⁷ The organic SBU of NU-110 is similar to the 3-c BTE SBU in that it contains a central aromatic ring with three arms extending outward with D_{3h} symmetry. The differences being that the length of these arms are approximately four times bigger than those of BTE, and at the terminus of these arms two carboxylate moieties are found, yielding a 6-c SBU (the authors called LH_6 ⁴⁷) instead of the 3-c SBU of BTE. It was shown⁴⁷ that the computed physical properties of the hypothetical

structure generated with the recursion method were in good agreement with the experimentally realized version of this MOF, and so as a curiosity, in this work we assembled the hypothetical version of NU-110, as well as new MOFs based on the tritopic version of the LH_6 SBU. We constructed the hypothetical version of NU-110 using the labelled quotient graph of **ntt**, which possesses vertices of degree 3 and 4. This required an alteration of the SBUs used to assemble the MOF; the 4-c Cu paddlewheel, and a fragmentation of the 6-c LH_6 organic ligand into large (LH_3) and small (benzene) 3-c SBUs. One challenge in constructing the hypothetical NU-110 structure was to ensure the 3-c organic SBUs were placed at specific nodes in the **ntt** net such that they combine to form the large (6-c) SBU in the final structure. Knowing that only the small 3-c benzene SBU form bonds with copper paddlewheels in the crystal structure, we imposed a constraint such that only benzene SBUs were assigned to 3-c vertices adjacent to 4-c vertices in the labelled quotient graph. The remaining 3-c vertices were then assigned LH_3 SBUs. Comparing the computed physical properties for both experimental and hypothetical NU-110 in columns 3 and 4 of Table 7.3 shows a slight over prediction with the hypothetical MOF. We note that there is a slight distortion in the SBUs when the hypothetical and experimental MOFs of NU-110 were superposed as seen in Figure 7.12d, which is likely the cause of the larger values in the reported properties for hypothetical NU-110.

Table 7.3: Cell parameters and selected physical properties of two experimental MOFs, NU-110 and MOF-210, along with purely hypothetical MOFs generated with the **ntt**, **toz** and **ith-d** topologies.

MOF	MOF-210		NU-110		MOF-toz hypothetical	MOF-ith-d hypothetical
	experimental	hypothetical	experimental	hypothetical		
a (Å)	50.7	51.4	48.6	49.6	103.7	69.2
b (Å)	50.7	51.4	48.6	49.7	98	69.3
c (Å)	194.3	191.6	48.6	49.6	461.7	69.3
α (deg)	90	90	60	60	54	89.6
β (deg)	90	90	60	60	105	93
γ (deg)	120	120	60	60	118	90
calculated surface area (m ² /g) ^a	5760	5720	6258	6624	8737	8462
calculated pore volume (cm ³ /g) ^b	2.66	2.69	3.04	3.24	15.71	13.94
experimental BET surface area (m ² /g)	6580 ^c		7140 ^b			
experimental pore volume (cm ³ /g)	3.60 ^c		4.40 ^b			

The final two columns of Table 7.3 demonstrate how the algorithm can predict new MOFs with incredibly high surface areas and void volumes. Both the **toz** and **ith-d** hypothetical MOFs were constructed using the exact same SBUs, one of which is a 3-c adaptation of the LH₆ SBU found in NU-110. The other SBUs were the Zn₄O inorganic SBU seen in Figure 7.12a and a 2-c organic SBU large enough to support the extended 3-c SBUs in both the **toz** and **ith-d** topologies. Hypothetical MOF-toz was found to have a 1.4-fold increase in calculated surface area over that of the NU-110 MOF. In addition the pore volume is larger by a factor of approximately 5. Hypothetical MOF-ith-d also has potential to break the gravimetric surface area record, however the computed properties are slightly lower than that of the SBUs assembled in the **toz** topology.

Finally, we note that when this project was in the process of being written-up, Martin and Haranczyk reported an algorithm which assembles SBUs together using the barycentric placement of nets obtained from the RCSR as 'blueprints'.¹¹ In such manner, SBUs with the correct connectivities are aligned to vertices within the net representation. Following this orientation, either the ideal symmetry operations are applied to the SBUs to attempt to form the

structure, or the SBUs are connected together iteratively until a unit cell is formed. The method reported in our work captures the same philosophy as that proposed by Martin and Haranczyk; that is to use a description of a net as a template for assembling MOFs. However in our work we use the nets' labelled quotient graph to assemble our structures instead of its 3-dimensional barycentric representation. We can therefore assemble materials based on nets which are not provided in the RCSR, for example. As well, because their reported method does not attempt to fit nets to given SBU geometries, we conjecture that their method will successfully build structures with only a limited number of SBUs of 'ideal' symmetry. SBUs with geometries which deviate significantly from the barycentric representation of a net will likely result in significant atomic collisions when attempting to assemble MOFs.

7.6 Conclusions

In this work we have developed a rigorous and robust method to generate MOF structures starting from any topology. Here we improve on the MOF-assembly method presented in chapter 3 (and similarly by Wilmer *et al.*¹) by leveraging theorems and proofs developed in topological graph theory¹⁰ to build three-dimensional structures with distortions and deviations that arise from MOF chemistry. This method provides more automation to the procedure of generating structures, reducing the time and effort required to parameterize each SBU for MOF assembly. Moreover, the method demonstrated here can easily build MOFs with a much more diverse range of topologies than could be captured with the method described in chapter 3, where the underlying topology is encoded in the parameters assigned to the connection sites of each SBU. In our method, much like the method of Martin and Haranczyk,¹¹ one only needs the SBUs, an indication of where they will form bonds with other SBUs, and a net describing the interconnectivity of the SBUs within a periodic lattice. The major difference between the method proposed here and that of ref 11 is we propose using the abstract definition of the net over the

barycentric realization, which allows us to manipulate the net to match SBU geometries prior to assembly. This program can be used for several applications, including the advancement of materials for specific properties such as the surface area and void volume, or to be used to generate a large database of topologically diverse structures for high throughput screening purposes.

7.7 References

- (1) Wilmer, C. E.; Leaf, M.; Lee, C. Y.; Farha, O. K.; Hauser, B. G.; Hupp, J. T.; Snurr, R. Q. *Nat. Chem.* **2012**, *4*, 83.
- (2) Sikora, B. J.; Winnegar, R.; Proserpio, D. M.; Snurr, R. Q. *Microporous Mesoporous Mater.* **2014**, *186*, 207.
- (3) Wells, A. F. *Three Dimensional Nets and Polyhedra*; John Wiley & Sons Ltd.: New York, 1977.
- (4) O’Keeffe, M.; Hyde, B. G. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **1980**, *295*, 553.
- (5) Alexandrov, E. V.; Blatov, V. A.; Kochetkov, A. V.; Proserpio, D. M. *CrystEngComm* **2011**, *13*, 3947.
- (6) O’Keeffe, M.; Peskov, M. A.; Ramsden, S. J.; Yaghi, O. M. *Acc. Chem. Res.* **2008**, *41*, 1782.
- (7) Chui, S. S. Y.; Lo, S. M. F.; Charmant, J. P. H.; Orpen, A. G.; Williams, I. D. *Science (80-.)*. **1999**, *283*, 1148.
- (8) Li, H.; Eddaoudi, M.; O’Keeffe, M.; Yaghi, O. M. *Nature* **1999**, *402*, 276.
- (9) Delgado-Friedrichs, O.; O’Keeffe, M. *Acta Crystallogr. Sect. A Found. Crystallogr.* **2003**, *59*, 351.
- (10) Eon, J.-G. *Acta Crystallogr. A.* **2011**, *67*, 68.
- (11) Martin, R. L.; Haranczyk, M. *Cryst. Growth Des.* **2014**, *14*, 2431.
- (12) Delgado-Friedrichs, O.; O’Keeffe, M. *J. Solid State Chem.* **2005**, *178*, 2480.
- (13) Chung, S. J.; Hahn, T.; Klee, W. E. *Acta Crystallogr. Sect. A Found. Crystallogr.* **1984**, *40*, 42.
- (14) Eon, J.-G. *J. Solid State Chem.* **1998**, *138*, 55.
- (15) O’Keeffe, M.; Eddaoudi, M.; Li, H.; Reineke, T.; Yaghi, O. M. *J. Solid State Chem.* **2000**, *152*, 3.

- (16) Eddaoudi, M.; Kim, J.; Rosi, N. L.; Vodak, D. T.; Wachter, J.; O’Keeffe, M.; Yaghi, O. M. *Science* (80-.). **2002**, 295, 469.
- (17) Blatov, V. A.; Shevchenko, A. P.; Proserpio, D. M. *Cryst. Growth Des.* **2014**, 14, 3576.
- (18) Delgado-Friedrichs, O.; Foster, M. D.; O’Keeffe, M.; Proserpio, D. M.; Treacy, M. M. J.; Yaghi, O. M. *J. Solid State Chem.* **2005**, 178, 2533.
- (19) Delgado-Friedrichs, O.; O’Keeffe, M.; Yaghi, O. M. *Phys. Chem. Chem. Phys.* **2007**, 9, 1035.
- (20) Delgado-Friedrichs, O.; Hyde, S. T.; Mun, S. W.; O’Keeffe, M.; Proserpio, D. M. *Acta Crystallogr. A.* **2013**, 69, 535.
- (21) Ockwig, N. W.; Delgado-Friedrichs, O.; O’Keeffe, M.; Yaghi, O. M. *Acc. Chem. Res.* **2005**, 38, 176.
- (22) Bonneau, C.; Delgado-Friedrichs, O.; O’Keeffe, M.; Yaghi, O. M. *Acta Crystallogr. A.* **2004**, 60, 517.
- (23) Delgado-Friedrichs, O. *Discrete Comput. Geom.* **2005**, 33, 67.
- (24) Klee, W. E. *Cryst. Res. Technol.* **2004**, 39, 959.
- (25) Beukemann, A.; Klee, W. E. *Zeitschrift für Krist.* **1992**, 201, 37.
- (26) O’Keeffe, M.; Yaghi, O. M. *Chem. Rev.* **2012**, 112, 675.
- (27) Godsil, C.; Royle, G. *Algebraic Graph Theory*; 1st ed.; Springer: New York, 2001.
- (28) Johnson, S. G. The NLOpt nonlinear-optimization package.
- (29) Cormen, T. H.; Leiserson, C. E.; Rivest, R. L.; Stein, C. *Introduction to Algorithms.*; 2nd ed.; MIT Press, 2001.
- (30) Goodrich, M. T.; Tamassia, R. *Data Structures and Algorithms in Java*; 2nd ed.; John Wiley & Sons Ltd., 2001.
- (31) Joyner, D.; Nguyen, M. Van; Cohen, N. *Algorithmic Graph Theory*; 2010.
- (32) Sahni, S. *Data Structures, Algorithms, and Applications in Java*; McGraw-Hill, 2000.
- (33) Stein, W. A. *sage Dev. team*, <http://www.sagemath.org>. **2013**.
- (34) Nocedal, J. *Math. Comput.* **1980**, 35, 773.
- (35) Liu, D. C.; Nocedal, J. *Math. Program.* **1989**, 45, 503.
- (36) Arun, K. S.; Huang, T. S.; Blostein, S. D. *IEEE Trans. Pattern Anal. Mach. Intell.* **1987**, PAMI-9, 698.

- (37) Chen, B.; Eddaoudi, M.; Reineke, T. M.; Kampf, J. W.; O’Keeffe, M.; Yaghi, O. M. *J. Am. Chem. Soc.* **2000**, *122*, 11559.
- (38) Kim, J.; Chen, B.; Reineke, T. M.; Li, H.; Eddaoudi, M.; Moler, D. B.; O’Keeffe, M.; Yaghi, O. M. *J. Am. Chem. Soc.* **2001**, *123*, 8239.
- (39) Cook, T. R.; Zheng, Y.-R.; Stang, P. J. *Chem. Rev.* **2012**.
- (40) Kresse, G.; Hafner, J. Ab initio molecular-dynamics simulation of the liquid-metal–amorphous-semiconductor transition in germanium. *Physical Review B*, 1994, *49*, 14251–14269.
- (41) Kresse, G.; Hafner, J. *Phys. Rev. B* **1993**, *47*, 558.
- (42) Kresse, G.; Furthmüller, J. *Comput. Mater. Sci.* **1996**, *6*, 15.
- (43) Kresse, G.; Furthmüller, J. *Phys. Rev. B* **1996**, *54*, 11169.
- (44) Wilmer, C. E.; Farha, O. K.; Bae, Y.-S.; Hupp, J. T.; Snurr, R. Q. *Energy Environ. Sci.* **2012**, *5*, 9849.
- (45) Togo, A. spglib (<http://spglib.sourceforge.net>), 2012.
- (46) Furukawa, H.; Ko, N.; Go, Y. B.; Aratani, N.; Choi, S. B.; Choi, E.; Yazaydin, a O.; Snurr, R. Q.; O’Keeffe, M.; Kim, J.; Yaghi, O. M. *Science* **2010**, *329*, 424.
- (47) Farha, O. K.; Eryazici, I.; Jeong, N. C.; Hauser, B. G.; Wilmer, C. E.; Sarjeant, A. a; Snurr, R. Q.; Nguyen, S. T.; Yazaydin, a Ö.; Hupp, J. T. *J. Am. Chem. Soc.* **2012**, *134*, 15016.

8 Conclusions and Future Directions

8.1 Summary

All of the work performed in this thesis was driven by the general theme of aiding the design of MOFs optimized to selectively adsorb CO₂ from post-combustion flue. We approached this in a systematic manner, developing and employing computational tools to identify important design criteria from existing MOF structures, at the same time validating that our models can provide a high degree of accuracy, then broadening our search by developing a large-scale screening approach to identify more robust attributes which contribute to high performance.

We highlight the challenges associated with MOF design in Chapter 2, where we investigate two MOFs synthesized with similar design principles, yet differ significantly in their CO₂ adsorption capacity. Both MOFs, CALF-15 and CALF-16 were functionalized with amine groups, which are known for their affinity for CO₂. In CALF-15, amine groups appear to contribute significantly to CO₂ adsorption, where the DFT calculated interaction energy of the x-ray resolved CO₂ molecules adsorbed near amines is quite high (44.2 kJ/mol). In contrast we found that the amine groups in CALF-16, which are higher in density than that of CALF-15, had little effect on the CO₂ adsorption. This was found after replacing the amine functional groups with isoelectronic methyls, and showing no effect on the performance of CALF-16 while reducing CO₂ uptake in CALF-15 by 20%. Further, a breakdown of the pairwise interaction energy found that some of the amine groups had a slight de-stabilizing effect on CO₂ adsorption in CALF-16, which can be attributed to the increased density of these groups in its pores. The results from these two MOFs reveal that the relationship between the functional groups, pore size and pore chemistry for optimal CO₂ adsorption is a complex one and there is a need for more detailed studies. However, it also highlights the excellent agreement achievable between

computation and experiment. For example, not only were we able to accurately calculate the macroscopic thermodynamic properties of CALF-15 by re-producing its experimental CO₂ adsorption isotherm at 273 K, but by superposing the x-ray resolved CO₂ binding sites with 3-D probability plots from GCMC simulations, we demonstrate an unprecedented level of microscopic agreement as well. This serves to validate parameterized energetic calculations used in this study which, importantly, used DFT-derived atomic charges (using the REPEAT method¹) to compute the electrostatic interactions with CO₂.

At the commencement of this thesis, all computational CO₂ adsorption studies, including those performed in Chapter 2, were performed on previously-synthesized structures.² While these are informative, their scope is too narrow to extrapolate general design principles from. Indeed, one would have to account for the nearly infinite number of possible pore shapes, sizes, and chemistry achievable in MOFs before suggesting robust structural features optimal for CO₂ adsorption. The tremendous power and creative potential of computational programming allows one to tackle such lofty challenges. In Chapter 3, we demonstrate that by using a library of SBUs as building blocks, one can rapidly assemble large databases of new hypothetical MOFs for high throughput screening studies. The algorithm implemented in this chapter, which is similar in design to that of Wilmer *et al.*,³ assembles structures whose connectivity and shape (topology) are entirely governed by a set of alignment parameters assigned to each SBU, which are taken from existing MOF crystal structures. One can therefore argue that the construction of these hypothetical materials contain a degree of synthetic viability, which is important in terms of predicting novel structures as well as to provide MOF pores with shapes and chemistry that is amenable to real structures. Chapter 3 reports the construction of 1.3 million hypothetical MOFs, using 28 inorganic SBUs, 173 organic SBUs and 50 functional groups assembled under 21

different topologies. At the time of this writing, this is the largest known hypothetical MOF database, both in terms of sheer size, and breadth of topologies. The caveat to the method described in this chapter is that algorithm requires an expert-level understanding of how these MOFs assemble, making it difficult to parameterize and build new MOFs to further expand the database. A more advanced method, yet easier to use, is discussed in Chapter 7.

With such a large range of structures available, this raises an important issue as to how to accurately screen them in a cost effective way. Determining adsorption at a single state point with a GCMC simulation requires, on average, 2-3 hours on a single CPU. Multiply this by 1.3 million structures and the amount of computing power needed becomes intractable. In addition, computing the electrostatic energy of CO₂ – MOF interactions in GCMC simulations requires an accurate description of the electrostatic potential (ESP). In this case, the ESP from DFT – derived REPEAT charges assigned to each atom in the framework provides a great degree of accuracy as mentioned above in the studies of CALF-15 and 16. This can take on the order of hours to compute for each MOF on 8 or more CPUs, which adds a large bottle-neck to screening the database. There are more approximate methods for assigning atomic charges which take much less time, however these methods require a number of parameters assigned to each atom type in the system. One such approximation is the charge equilibration method (QEq) first described by Rappé and Goddard,⁴ which assigns charge to each atom using two parameters for each element; the atomic hardness, J_A and electronegativity χ_A . Unfortunately, the original authors developed parameters that were not designed specifically for MOFs, or to reproduce the electrostatic potential. In Chapter 3 we discuss the pitfalls of using these Universal Force Field (UFF) parameters and the QEq method to generate atomic charges in MOFs. Specifically, we found that the UFF parameters over-estimated CO₂ uptake by at most 6.6 mmol/g, which is a

significant deviation and would result in many false positives when screening these materials. Thus in Chapter 3 we detail the method of training the J_A and χ_A parameters on a set of 543 MOFs by fitting the resulting QEq ESP to that of the DFT-derived charges. We called this set of J_A and χ_A terms the MEPO parameters, and they were found to give the best agreement in CO₂ adsorption at flue conditions when compared with the adsorption values computed with REPEAT charges, yielding Pearson and Spearman correlation coefficients of 0.97. This was further validated on a set of 693 MOFs *not* included in the training set, where we report Pearson and Spearman correlations in CO₂ uptake of 0.98. Moreover, the method underestimates adsorption by at most 2.1 mmol/g indicating that in the worst case, we are conservatively screening the database with these parameters.

It is important to note that the training set of hypothetical MOFs used to develop the MEPO parameters encompass a subset of the MOFs in the database, and as such, can only be used to accurately screen that portion. Specifically, MOFs with a range of organic SBUs with Cu paddlewheels, Zn paddlewheels, Zn₄O clusters, and V₂O₂ rods. Thus, in Chapter 6 we detail the screening of a subset of the hypothetical MOF database (324,474 MOFs) which we found were agreeable with the MEPO parameters. The generation of a set of parameters transferrable across the entire database is left as future work.

One goal in screening the nearly 325,000 MOFs for their performance in capturing CO₂ was to identify some high performing materials which could serve as synthetic targets. We found in Chapter 6 several MOF types which yield promising results, recovering CO₂ at greater than 4 mmol/g of material and having a selectivity for CO₂ over N₂ of 150. These values were considered targets for reducing the cost of CO₂ capture to below \$20/tonne CO₂ captured.⁵ Importantly these materials did not possess open metal sites, and less prone to water poisoning.

We found that MOFs with large aromatic sheets separated by roughly 7 Å provide ideal binding sites for CO₂, in addition, MOFs with the **fof** topology containing narrow circular apertures of roughly 7 Å diameter were found to provide beneficial CO₂ capture properties.

Another goal related to CO₂ screening in Chapter 6 was to establish pore features which provide ideal binding pockets for CO₂. We call these features adsorbophores, the identification of which required developing several cheminformatics tools which are discussed in Chapter 5. In this chapter we briefly discuss how strong CO₂ binding sites are identified in MOFs using the ABSL package. Adsorbophores were discovered from this data by extracting the active site (pore atoms) around each binding site, and using a substructure mining algorithm to identify common atomic properties amongst them. Because this method performs random pairings to identify frequent substructures, we test its robustness by varying the random seed across 30 separate runs. The similarity found between the runs suggests that the most common adsorbophores are always ranked amongst the top 50, however the active sites used to generate the top ranked adsorbophores can be quite dissimilar. This suggests that there are a relatively common set of binding sites amongst the MOFs, and that increasing the diversity of the SBUs and topologies used to build the MOFs could possibly reduce the efficacy of this method.

In Chapter 6 the substructure mining method was used to identify 3 adsorbophores from the top performing 8325 MOF structures discovered from the CO₂ screening. These adsorbophores describe CO₂ binding i) in between aromatic rings spaced 7 Å apart, ii) adjacent to a bridging oxygen in an oxo-vanadium chain, and iii) wedged between two open metal site Cr ions on separate Cr₃O SBUs. The discovery of these common traits amongst a large range of high performing materials represents a significant step forward, not only in designing MOFs for

CO₂ capture, but more generally, in the field of materials science by developing these powerful cheminformatics tools and demonstrating their use.

Finally in Chapter 7, a new method for constructing hypothetical MOFs is developed. This method builds hypothetical MOFs by manipulating nets (which describe the underlying 3-D connectivity of MOFs) such that their nodes match the geometry of SBUs. The SBUs are then superposed on the nodes to produce new hypothetical MOFs. By using the net as structure-directing, this method does away with the difficulties associated with parameterizing SBUs for generating MOFs described in Chapter 2. We demonstrate in this chapter, the rapid construction of close to 50 unique MOFs with different topologies using only two SBUs. This is more than twice the number of topologies in the entire 1.3 million MOFs made previously. It is anticipated that the structures generated from this method will yield a much larger range of pore structures and chemistries, whereupon screening them for CO₂ capture will push the boundaries of what is currently known about CO₂ adsorption in MOFs. This method provides an opportunity for many possible future studies, some of which are discussed in the next section.

Much of the work performed in these chapters required the development of computational tools to advance the materials discovery program. Here we summarize the major software that was developed as part of this thesis work. Vital to the work presented in Chapters 2, 4, 5, and 6 is the evaluation of CO₂ adsorption within porous materials. This was accomplished using a GCMC program written in the FORTRAN 90 programming language, with routines for computing the interaction energies adopted from the open source DL_POLY molecular dynamics program.⁶ Described in Chapters 3 and 7 are two MOF structure generation algorithms, where the underlying mechanics of each are quite different. The algorithm developed in Chapter 3 involved building MOFs by seeding an SBU and iteratively 'snapping' more SBUs

to available connection sites, while in Chapter 7 SBUs are placed on a blueprint (net) which is adjusted to match the SBU geometries. Python 2.7 was primarily used for these programs, with an effort to design the code in an object oriented fashion. In Chapter 5 we present the program which identifies common adsorption sites, or adsorbophores, in MOFs. Here much of the program was written in the Python 2.7 programming language, with a portion written in C++ when computational speed was required, such as for the discovery of maximum cliques. While this summary is not an exhaustive list of the programs written over the course of this work, it serves to demonstrate the emergent nature of high throughput screening in materials science, and the lack of existing software designed for this purpose. Finally, we mention the tireless work of a colleague, Dr. Tom Daff, who authored an automated tool called the Fully Automated Adsorption Analysis in Porous Solids (FA³PS), which interfaces many programs vital for the work performed in Chapters 4 and 6, including the GCMC program of this author.

8.2 On-going and Future Work

We commence the discussion of future work in this thesis topic with a general question that arises from the work previously performed; have we sampled a broad enough range of MOF structures by way of topologies and chemistry to provide a robust set of MOF design principles? The answer to this question requires further development of the methods detailed in Chapters 4, 5 and 7 of this work. In Chapter 7 we disclose a novel method for generating hypothetical MOFs with an unprecedented range of topologies. This paves the way for a new database of structures which will sample a broader spectrum of pore shapes, sizes and chemistry ultimately leading to the discovery of features from a more robust range of data. This, in turn, may have significant implications for the ultra-fast method for constructing accurate ESPs with the MEPO parameters discussed in Chapter 4. Charge assignment in the QEq method is spatially dependent, meaning

the parameters which were fit to a set of MOFs previously in Chapter 5 may not be applicable to new spatial distributions of atoms found with new topologies and SBUs. If the MEPO parameters are shown to produce poor ESPs, new parameters must be constructed by developing a training/validation set encompassing a wider range of structures than was done in Chapter 4. We note that this work is currently under way in our lab, which will support the entire 1.3 million MOF database developed in Chapter 3. The new parameters are being trained using the split charge method, which provides a more accurate fit of the quantum mechanical ESP compared with the QEq method (but also requires more parameters per atom).⁷

In Chapter 5 we noted that, because the geometric similarities of the final adsorbophores were high, yet the active sites combined to make those adsorbophores were quite dissimilar across different runs, the random pairing method used to mine for frequent substructures may not robustly identify adsorbophores when the diversity of the database of structures increases. This is where the adoption of more involved methods^{8–11} may be necessary to successfully search for adsorbophores. One possibility is to ‘grow’ cliques, starting with sets of only two atoms which are found to be frequent amongst the active sites and then growing the size of the common atoms by one until the frequency of that clique drops below a user-defined threshold.⁸ This has the advantage that it is much more inclusive compared to the random pairing method, as it scans every active site for the presence of each clique. One of the major challenges associated with this method is to reduce the redundancy of the common substructures being mined, which can waste considerable amounts of compute time as it searches for the same adsorbophores. While the methods and results from these algorithms are presented in scientific literature, the source code is typically protected by copyright and access to which require licensing fees. Thus initial steps in

this direction will require implementing our own version of the code borrowing some of the functions from the random pairing algorithm as a starting point.

From screening the set of 325,000 structures in Chapter 6, we have discovered strong CO₂ binding sites in a range of different pockets, thus we anticipate that screening the larger database using the more advanced methods discussed above will yield further MOF design principles and targets for synthesis.

In Chapter 6 we used the substructure mining method to identify three prominent adsorbophores found from high performing MOFs. These were presented as potential design criteria for MOFs with high CO₂ adsorption. However in drug discovery studies, the pharmacophores found from substructure mining are typically used as input to develop robust models of drug activity using cheminformatics methods. In these studies, the model is developed using quantitative structure activity relationships (QSAR), which develop a function of the known activities of a training set of drug molecules using the presence or absence of each pharmacophore. The resulting model can then be used to rapidly predict activity in new drug-like molecules, and is widely used to provide an accurate prediction of drug activity. In the past, the lab has demonstrated that QSAR methods can be successfully performed on the database of MOFs to rapidly identify high performing materials.¹² However the model was only a binary classifier that would only classify a MOF as high or low performing (greater/less than 1 mmol/g) at flue conditions. This method used property-weighted radial distribution functions to describe each MOF,¹² and it is possible that introducing a description of strong CO₂ binding pockets as descriptors in the model will improve the prediction of CO₂ adsorption in these materials.

The method described in Chapter 7 represents a powerful new way to build hypothetical MOFs, however during assembly, no symmetry elements of the net representations were

considered. This can ultimately yield hypothetical MOFs with low symmetry, as SBUs with potentially lower point group symmetry are assigned to high symmetry nodes in a somewhat arbitrary fashion. Specifically, SBUs are assigned to minimize a root mean squared deviation between the SBU geometry and the nodes edges. Including symmetry considerations in the construction of new MOFs represents an important step forward in the design of new structures, as in general, crystalline materials with high symmetry are often observed in nature. Moreover, in the crystal structure prediction field, space group symmetries are typically leveraged when attempting to predict bulk and molecular crystal structures.^{13,14} Fortunately, in the work of Eon¹⁵, it is demonstrated that any 3-D net can be represented by a maximum symmetry group, and distortions to the net can be performed by controlled reductions in the space group symmetry. Future work in improving hypothetical structure prediction will require implementing these space group considerations, such that SBUs are assigned to nodes in the nets by matching point group symmetries, where allowable reductions in symmetry are accounted for based on SBU chemical and geometric features. In reference 15, each example demonstrating a reduction in space group symmetry in a net was predicated by a detailed explanation as to which symmetry operations had to be discarded to support the desired net distortions. Thus one of the major obstacles associated with this work will be to implement a systematic determination of the maximum acceptable space group symmetry of a net which supports the geometries of a set of SBUs.

Finally, citing the question posed at the beginning of this section, the most powerful justification of this work would be the successful synthesis and validation of the materials we suggest as high performing for CO₂ capture and storage. Recently, we have been collaborating with the Murugesu group at the University of Ottawa to synthesize one of the high performing

materials. This material includes large polyaromatic organic SBUs, which are, in general quite difficult to dissolve in reaction media. As such, one possible future direction of this work will be to investigate MOFs with more soluble aromatic systems which are decorated with polar functional groups.

8.3 References

- (1) Campañá, C.; Mussard, B.; Woo, T. K. *J. Chem. Theory Comput.* **2009**, *5*, 2866.
- (2) Düren, T.; Bae, Y.-S.; Snurr, R. Q. *Chem. Soc. Rev.* **2009**, *38*, 1237.
- (3) Wilmer, C. E.; Leaf, M.; Lee, C. Y.; Farha, O. K.; Hauser, B. G.; Hupp, J. T.; Snurr, R. Q. *Nat. Chem.* **2012**, *4*, 83.
- (4) Rappe, A. K.; Goddard, W. A. *J. Phys. Chem.* **1991**, *95*, 3358.
- (5) Ho, M. T.; Allinson, G. W.; Wiley, D. E. *Ind. Eng. Chem. Res.* **2008**, *47*, 4883.
- (6) Forrester, T. R.; Smith, W. *DL_POLY ver. 2.18 (C.L.R.C. Daresbury Lab. Daresbury, UK, 1995)*.
- (7) Nistor, R. a; Polihronov, J. G.; Müser, M. H.; Mosey, N. J. *J. Chem. Phys.* **2006**, *125*, 094108.
- (8) Podolyan, Y.; Karypis, G. *J. Chem. Inf. Model.* **2009**, *49*, 13.
- (9) Jiang, C.; Coenen, F.; Zito, M. *Knowl. Eng. Rev.* **2012**, *28*, 75.
- (10) Dror, O.; Schneidman-Duhovny, D.; Inbar, Y.; Nussinov, R.; Wolfson, H. J. *J. Chem. Inf. Model.* **2009**, *49*, 2333.
- (11) Yan, X. Y. X.; Han, J. H. J. In *2002 IEEE International Conference on Data Mining, 2002. Proceedings.*; IEEE Comput. Soc, 2002; pp. 721–724.
- (12) Fernandez, M.; Boyd, P. G.; Daff, T. D.; Aghaji, M. Z.; Woo, T. K. *J. Phys. Chem. Lett.* **2014**, *5*, 3056.
- (13) Caputo, R.; Kupczak, A.; Sikora, W.; Tekin, A. *Phys. Chem. Chem. Phys.* **2013**, *15*, 1471.

- (14) *Modern Methods of Crystal Structure Prediction*; Oganov, A. R., Ed.; Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, Germany, 2010.
- (15) Eon, J.-G. *Acta Crystallogr. A* **2011**, 67, 68.