

Large Scale Computational Screening of Metal Organic Framework Materials for Natural Gas Purification

Mohammad Zein Aghaji

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

© Mohammad Zein Aghaji, Ottawa, Canada, 2017

Contents

Abstract.....	vi
Table of Figures.....	ix
Table of Tables.....	xvii
List of Acronyms.....	xix
Acknowledgements.....	xx
Contributions to Original Research.....	xxi
1 Introduction	1
1.1 Natural Gas in Canada.....	2
1.2 Natural Gas Processing	3
1.3 PSA-based CO ₂ removal process	5
1.4 Metal Organic Framework Materials.....	7
1.4.1 MOFs for CO ₂ separation from natural gas.....	10
1.4.2 Computational approach in MOFs discovery.....	12
1.5 Thesis Goals and Outlines	15
1.6 References	18
2 Background Information	24
2.1 Criteria for MOF performance	24
2.1.1 CO ₂ working capacity	24
2.1.2 CO ₂ /CH ₄ selectivity.....	25
2.1.3 Heat of adsorption	26
2.1.4 MOF stability.....	27
2.1.5 Kinetics of gas adsorption in MOF	29
2.2 Computational tools in MOF study	30
2.2.1 Grand Canonical Ensemble	31
2.2.2 Grand Canonical Monte Carlo Simulation.....	33
2.2.2.1 Periodic boundary conditions	38
2.2.2.2 Force Field.....	39
2.2.2.3 Short-range van der Waals Interaction Potential	40
2.2.2.4 Long-range Electrostatic Interaction Potential.....	42
2.2.2.5 Charge Equilibrium methods.....	46
2.2.3 Quantitative Structure–Property Relationship (QSPR)	48

2.2.3.1	Descriptors.....	48
2.2.3.2	QSAR modeling Methods	50
2.2.3.3	Regression QSAR modelling methods.....	51
2.2.3.4	Classification QSAR modelling methods.....	52
2.2.3.5	Validation of QSAR models	60
2.3	References	61
3	Computational modelling of CO ₂ adsorption on Phosphonate Monoesters-base MOF	64
3.1	Introduction	64
3.2	Computational details.....	65
3.3	Result & Discussion	68
3.4	Conclusion.....	75
3.5	References	76
4	Investigation of MIL-47 and its functionalized derivatives on Natural gas purification: A Computational Study	79
4.1	Background	79
4.2	Computational Methods.....	84
4.2.1	Optimization and Generation of functionalized MIL- 47 materials	84
4.2.2	Force Fields	85
4.2.3	GCMC simulations.....	86
4.3	Result and Discussion.....	89
4.3.1	Adsorption of pure components.....	89
4.3.2	Adsorption of Binary Mixtures.....	90
4.3.3	Comparison of performance of functionalized MIL-47 materials with other materials.....	96
4.4	Conclusion.....	97
4.5	References	97
5	Water Adsorption in Metal Organic Frameworks and the Effect of Water on Natural Gas Purification in MOFs: A Computational Study.....	101
5.1	Background	101
5.2	Systems and Computational Details	104
5.3	Results and discussion	111
5.3.1	Water adsorption.....	111
5.3.2	The effect of water on natural gas purification in MOFs.....	120
5.4	Conclusion.....	121

5.5	References	122
6	Large-Scale Computational Screening of MOFs for Natural gas purification	125
6.1	Introduction	125
6.2	Methods.....	129
6.2.1	Hypothetical MOF Database	129
6.2.2	Screening Protocol	130
6.2.3	Binding Site Analysis	134
6.2.4	Discovery of Common Binding Site Motifs	135
6.2.5	GCMC Simulation Details	137
6.3	Results & Discussion	139
6.3.1	CO ₂ -CH ₄ Screening.....	140
6.3.2	Structure-property relationships	141
6.3.3	Adsorbophore Analysis	148
6.3.4	Conjugated Aromatic-Based High Performing MOFs.....	153
6.3.5	Ab initio geometry optimization of the top performers.....	157
6.3.6	The Effect of Water on the Performance of Top High-performing MOFs	161
6.4	Conclusion.....	164
6.5	References	166
6.6	Appendix - Hypothetical MOF Database.....	167
7	Quantitative Structure-Property Relationship Models for Recognizing Metal Organic Frameworks (MOFs) with High CO ₂ working Capacity and CO ₂ /CH ₄ Selectivity for Methane Purification	170
7.1.	Introduction	171
7.2	Data set and computational methods	173
7.2.1.	Hypothetical MOF dataset and GCMC simulations.....	173
7.2.2.	Machine learning classification models	176
7.2.	Result and Discussion.....	177
7.3.	Conclusions	193
7.4.	References	194
8	Investigation of Atomic Property-Weighted Radial Distribution Function and Voronoi Hologram Descriptors in Machine Learning Recognition of High Performing Metal Organic Frameworks for Natural Gas Purification	196
8.1	Introduction	196
8.2	Methods.....	200
8.2.1	Hypothetical MOF dataset	200

8.2.2	Grand Canonical Monte-Carlo (GCMC) simulations	201
8.2.3	Descriptors	202
8.2.4	Genetic algorithm-based descriptor selection.....	207
8.2.5	Support Vector Machine (SVM) models	212
8.3	Results and Discussion	212
8.3.1	Evaluating MOFs for NG Purification	212
8.3.2	Training of Support Vector Machine Classifiers.....	213
8.3.3	Predictive Capability of the Classifiers	216
8.4	Conclusion.....	222
8.5	References	222
9	MOFFinder: A Novel Web-based Tool for Aiding in the Discovery of New Metal Organic Framework Materials	225
9.1	Introduction	225
9.2	Functionality	227
9.3	Graphical User Interface	233
9.4	Data Collection.....	236
9.5	Implementation	237
9.6	Conclusion.....	241
9.7	References	242
10	Conclusion and Future Direction	245
10.1	Future directions.....	251
10.2	References	254

Abstract

An immediate reduction in global CO₂ emissions could be accomplished by replacing coal- or oil-based energy sources with purified natural gas. The most important process involved in natural gas purification is the separation of CO₂ from CH₄, where Pressure Swing Adsorption (PSA) technology on porous materials has emerged as a less energy demanding technology.

Among porous materials which are used or could potentially be used in PSA, Metal Organic Frameworks (MOFs) have attracted particular interest owing to their record-breaking surface areas, high-porosity, and high tunability. However, the discovery of optimal MOFs for use in adsorption-based CO₂ separation processes is remarkably challenging, as millions of MOFs can potentially be constructed from virtually limitless combinations of inorganic and organic secondary building units. To overcome this combinatorial problem, this thesis aims to (1) identify important design features of MOFs for CO₂/CH₄ separation through the investigation of currently existing MOFs as well as the high throughput computational screening of a large database of MOFs, and to (2) develop efficient computational tools for aiding the discovery of new MOF materials.

To validate the computational methods and models used in this thesis, the first work of this thesis presents the computational modeling of CO₂ adsorption on an experimental CuBDPMe MOF using grand canonical Monte Carlo simulations and density functional theory. The simulated CO₂ adsorption isotherms are in good agreement with experiment, which confirms the accuracy of the models used in our simulations throughout this thesis. The second work of this thesis investigates the performance of an experimental MIL-47 MOF and its seven functionalized derivatives in the context of natural gas purification, and compares their

performance with that of other well-known MOFs and commercially used adsorbents. The computational results show that introducing polar non-bulky functional groups on MIL-47 leads to an enhancement in its performance, and the comparison suggests that MIL-47-NO₂ could be a possible candidate as a solid sorbent for natural gas purification. This study is followed by the compactional study of water effects on natural gas purification using MOFs, as traces of water is present in natural gas under pipeline specifications. From the study, it is found that water has a marginal effect on natural gas purification in hydrophobic MOFs under pipeline specifications.

Following the aforementioned studies, a database of 324,500 hypothetical MOFs is screened for their performance in natural gas purification using the general protocol defined in this thesis. From the study, we identify 'hit' materials for targeted synthesis, and investigate the structure-property relationships with the intent of finding important MOF design features relevant to natural gas purification. We show that *layered sheets* consisting of poly-aromatic molecules separated by a perpendicular distance of roughly 7 Å are an important structural-chemical feature that leads to strong adsorption of CO₂.

Following the screening study, we develop efficient computational tools for the recognition of high-performing MOFs for methane purification using Machine Learning techniques. A training set of 32,500 MOF structures was used to calibrate support vector machines (SVMs) classifiers that incorporate simple geometrical features including pore size, void fraction and surface area. The SVM machine learning classifiers can be used as a filtering tool when screening large databases. The SVM classifiers were tested on ~290,000 MOFs that were not part of the training set and could correctly identify up to 70% of high-performing MOFs while only flagging a fraction of the MOFs for more rigorous screening. As a complement

to this study, we present ML classifier models for CO₂/CH₄ separation parameters that incorporate separately the Voronoi hologram and AP-RDF descriptors, and we compare their performance with the classifiers composed of simple geometrical descriptors. From the comparison, it is found that including AP-RDF and Voronoi hologram descriptors into the classifiers improves the performance of classifiers by 20% in capturing high-performing MOFs.

Finally, from the screening data, we develop a novel cheminformatics tool, MOFFinder, for aiding in the discovery of new MOFs for CO₂ scrubbing from natural gas. It has a user-friendly graphical interface to promote easy exploration of over 300,000 hypothetical MOFs. It enables synthetic chemists to find MOFs of interest by searching the database for Secondary Building Units (SBUs), geometric features, functional groups and adsorption properties. MOFFinder provides, for the first time the substructure/similarity query of porous materials for users and is publicly available on titan.chem.uottawa.ca/moffinger.

Table of Figures

Figure 1.1: Simplified schematic representation of natural gas processing.....	3
Figure 1.2: Graphical representation of PSA process for removing CO ₂ from natural gas. CO ₂ is selectively trapped in a solid material, which undergoes regeneration to release CO ₂ and make the materials useable for further adsorption/regeneration cycles.	7
Figure 1.3: a) MOF-5 from its inorganic Zn ₄ O complex and terephthalic acid linker. Green: zinc, white: hydrogen, red: oxygen, grey: carbon. b) MOF-10 from its Zn ₄ O complex and p-terphenyl-4,4'-dicarboxylate linker. c) pcu topology of MOF-10 and MOF-5 where the Zn ₄ O complex is abstracted to the octahedral geometry and the benzene-based linkers abstracted to an edge bridging two octahedrons. ...	9
Figure 2.1: Adsorption isotherm of CO ₂ in a MOF. The close circle represents the uptake of CO ₂ at the CO ₂ partial pressure relevant to natural gas processing, while the open one represents the uptake under desorption condition. Δq represent the CO ₂ working capacity.....	25
Figure 2.2: The basic moves in a GCMC simulation: a) A CO ₂ guest molecule is displaced by a random translation and rotation. b) A molecule of CO ₂ is deleted (to the right) and inserted (to the left).	37
Figure 2.3: A behaviour of functions $1/r$ (red curve), $\text{erf}(r)r$ (blue curve), and $\text{erfc}(r)r$ (black curve) against a distance between two atoms, r	44
Figure 2.4: a) A representation of the electron density, ρ_1 and ρ_2 , of two interacting molecules. The surface around each molecule shown is an isosurface of the total electron density. The colors reflect the electrostatic potential due to the nuclear charges and electron density of the molecule at that isosurface point. Red regions represent regions of negative electrostatic potential, while blue regions represent regions of positive electrostatic potential. b) Partial atomic charges (in units of electron charge) commonly used to approximate the electrostatic interaction between molecules.	45
Figure 2.5: An example of a Decision Tree (DT) model with three descriptors A, B, and C. Each branch of the tree represents a threshold level given for the value of a descriptor (X for descriptor A, Z for descriptor C, and Y for descriptor B), where the samples with values higher than the level threshold are branched to the right, while samples with lower values are branched to the left. The high-performing node is highlighted in grey. The scores in the lowest branches are averaged QSPR scores over all samples in that branch.....	53
Figure 2.6: An example of a linearly separable two-class data with possible separating lines. Lines in light-blue color are assigned as “bad” lines for characterizing the two data classes within SVM’s concept, while the line in dark-blue color is an optimal line which SVM attempts to find.....	54

Figure 2.7: An example of noisy data, where SVM finds an optimal hyperplane (blue line) by maximizing margin (m) and allowing some points to be misclassified or in the margin (e.g., X_i and X_j). The gray area represents the margin.	56
Figure 2.8: A representation of mapping descriptor vectors from an input space onto a feature space where data is linearly separable	58
Figure 3.1: Ground state spin-density localization from DFT calculations of CuBDPMe at an isosurface value of 0.5e. Up (purple) and down (green) unpaired spins are localized on the copper atoms.	66
Figure 3.2: a) The dianionic ligand, BDPR, where R = Me. b) hSingle crystal structure of one column of CuBDPMe, showing the bidentate coordination of the phosphonate monomethylester and the 4-coordinate geometry of Cu. c) crystal structure with solvent accessible voids (1.2 Å probe in gold). In c) one channel is shown in a space filling representation.	68
Figure 3.3: Probability density plots of the CO ₂ oxygen atom in a unit cell of CuBDPMe at 1 bar, 273 K from simulation. In a) probabilities along the c-axis are projected onto the a-b plane while in b) probabilities along the b-axis are projected onto the b-c plane. A green CO ₂ molecule is shown in a) for reference.....	71
Figure 3.4: Comparison of experimental and simulated CO ₂ adsorption isotherms at 273 K. a) The simulated isotherms were calculated with the X-ray structure, which was solvent loaded, and that where this structure's cell parameters were adjusted to that determined upon solvent removal and loaded with CO ₂ at 1.2 bar. b) The simulated isotherms were calculated with the X-ray structure, where the geometry of the methoxy groups have been altered from the X-ray structure in silico.....	73
Figure 3.5: Change in the DFT calculated energy for variation of the torsion angle to methyl tethers.....	74
Figure 3.6: Isosteric heat of adsorption of CuBDPMe with lattice parameters scaled for CO ₂ loading at 1.2 bar, as a function of guest loading determined from GCMC simulations at 273 K with the methyl group rotated by $\phi = 0^\circ, 2^\circ$ and 4°	75
Figure 4.1: Working capacity for CO ₂ in MOFs (blue) which is the difference between the amount of CO ₂ adsorbed at 6 bar and the amount of CO ₂ adsorbed at 1 bar. Non-MOF materials (red) are included for comparison.	81
Figure 4.2: Crystal structure of MIL-47 (right) its inorganic structural building unit (top-left) and organic structural building unit (bottom-left)	82
Figure 4.3: a) Adsorption isotherms of CO ₂ (red circle: simulation, red cross: experiment) and CH ₄ (black circle; simulation, black cross: experiment) in MIL-47. b) Adsorption isotherm of CO ₂ (red circle: simulation, red cross: experiment) in MIL-47-OCH ₃	90

Figure 4.4: CO₂ adsorption isotherms from GCMC simulations of a 1:9 CO₂-CH₄ mixture on MIL-47 (light blue), MIL-47-OMe (light gray), MIL-47-Br (dark green), MIL-47-COH (red), MIL-47-Cl (dark gray), MIL-47-COOH (dark blue), MIL-47-NO₂ (brown), MIL-47-OH (black), and MIL-47-NH₂ (light green) at 303 K. 91

Figure 4.5: CO₂ working capacity obtained from GCMC simulations of CO₂-CH₄ mixture on MIL-47 and its functionalized derivatives using a pressure swing of 10 to 1 bar 92

Figure 4.6: Simulated selectivities for CO₂ over CH₄ from their 1:9 gas mixture at 303 K in MIL-47 and its functionalized derivatives, as a function of the total pressure. 93

Figure 4.7: Calculated isosteric heats of adsorption at zero loading of pure CO₂ (blue bar) and CH₄ (red bar) gases in MIL-47 and its functionalized derivatives at 303 K..... 94

Figure 4.8: a) three CO₂ binding sites and one CH₄ binding site (black circle) are obtained from the GCMC simulation of a CO₂-CH₄ mixture at a total pressure of 10 bar in a single pore of the MIL-47 where the carboxyl O and the μ -2-O atoms interact with CO₂ molecules. b) four CO₂ binding sites are obtained from the GCMC simulation of a CO₂-CH₄ mixture at a total pressure of 10 bar in a single pore of the MIL-47-NO₂ where O atoms from NO₂ groups interact with CO₂ molecules. Green: V, Red: O, Gray: C, and Blue: N 95

Figure 4.9: CO₂/CH₄ selectivities and CO₂ working capacity for a 1:9 CO₂-CH₄ mixture at conditions relevant to natural gas purification in a variety of MOFs and two non-MOF materials. The chosen conditions correspond to an adsorption pressure of 10 bar, a desorption pressure of 1 bar, and temperature of 300 K..... 96

Figure 5.1: Atomic structure of (a) ZIF-8, (b) MIL-47, (c) Al(OH)(1,4-naphthalenedicarboxylate), (d) UiO-66, (e) Zn-DMOF-A, (f) Zn-DMOF-TM, (g) Cu-BTC, (h) CAU-OH-10, and (i) MOF-806. N, blue; C, gray; Cu, orange; O, red; V, green; Al, purple; Zn, yellow; and Zr, cyan. 106

Figure 5.2: Atom labelling of (a) ZIF-8, (b) MIL-47, (c) UiO-66, (d) Al(OH)(1,4 naphthalenedicarboxylate), (e,) Zn-DMOF-A, (F) Zn-DMOF-TM, (g) Cu-BTC, (h) CAU-OH-10, and (i) MOF-806 for the partial atomic charges given in Table 5.2..... 108

Figure 5.3: Definition of the geometric parameters used in the water models: (a) SPC/E and TIP3P, (b) TIP4P, TIP4P/Ew, and (c) TIP5P, TIP5P/Ew 110

Figure 5.4: Simulated and experimental adsorption isotherms for water in (a) ZIF-8, (b) MIL-47, (c) Al(OH) (1,4 NDC), (d) UiO-66, (e) Zn-DABCO-ADC, (f) Zn-DABCO-TM. (Experimental Data taken from: ZIF-8,³⁸ MIL-47,²⁰ AL(OH) (1,4 NDC),³⁹ UiO-66,⁴⁰ Zn-DBCO-ADC,⁴¹ Zn-DABCO-TM⁴¹). 113

Figure 5.5: (a) the number of guest molecules (TIP4P-Ew water molecules) in MIL-47 as a function of GCMC steps. (b-e) Snapshots of water adsorption during GCMC simulation: (b) 40 millions steps, (c) 80 millions steps, (d) 200 millions steps, and (e) saturation point. 115

Figure 5.6: Histogram of the number of hydrogen bonds per water molecule formed in MIL-47 (blue bar) and bulk (red bar). The distribution for bulk water is obtained through the simulation of TIP4P water molecules in a box with the same size as the MIL-47.....	116
Figure 5.7: Simulated and experimental adsorption isotherms for water in (a) CAU-OH—10 and (b) MOF-806. (Experimental Data taken from: CAU-OH-10 ⁴⁵ and MOF-806 ⁴⁶)	118
Figure 5.8: Snapshot showing TIP4P-Ew water molecules adsorbed in CAU-OH-10 at (a) $p/p_0=0.17$ and (b) $p/p_0=0.45$. Green region in (b) represents water molecules located next to OH groups.....	118
Figure 5.9: Simulated and experimental adsorption isotherms for (a) Cu-BTC without attached water, (b) Cu-BTC with attached water molecules, and (c) Cu-BTC with attached force field-type water molecules. (Experimental Data taken from: ref 47)......	119
Figure 5.10: Selectivities of CO ₂ in 1:9 mixtures of CO ₂ /CH ₄ with (cross) and without (circle) water as a function of pressure at 303 K.....	120
Figure 6.1: Flowchart of screening and analysing of 324,000 MOFs.....	133
Figure 6.2: a) GCMC-derived 3-D probability distribution of CO ₂ carbon (blue cloud) and oxygen (red cloud) in the Cu-BTC MOF and b) representative CO ₂ binding sites are identified in the Cu-BTC using the ABSL program.....	135
Figure 6.3: A) Examples of two binding sites and B) the correspondence graph of the two binding sites. C) The largest common substructure of two binding sites is obtained from the maximum cliques shown within the red circle in B.	137
Figure 6.4: A color map of the 324,474 MOFs screened for working capacity and CO ₂ /CH ₄ selectivity with GCMC simulations. A pink shaded zone in the upper right quadrant represents MOFs whose selectivity and working capacity are both comparable to those of zeolite 13x.	141
Figure 6.5: Scatter plots of CO ₂ working capacity vs CO ₂ /CH ₄ selectivity for 324,000 hypothetical MOFs color-mapped according to (a) void fraction determined by helium probe atom, (b) surface area and (c) max-pore diameter. The rectangular orange zone in the upper right hand corner of all three figures represents the region of MOFs which possess working capacities and selectivities above zeolite 13X. .	143
Figure 6.6: Distribution of CO ₂ /CH ₄ selectivity of MOFs with the pore size above 10 (blue) and the pore size in the range 4-6 Å (red).	144
Figure 6.7: Scatter plots of CO ₂ working capacity vs selectivity CO ₂ /CH ₄ of 324,000 hypothetical MOFs with a colour map of the CO ₂ heat adsorption (Q _{st})	145
Figure 6.8: Frequency of selected functional groups present in the database as a function of selectivity for MOFs with working capacity above 2 mmol/g.....	146

Figure 6.9: Averaged weighted radial distribution functions $g(r)$ of CO_2 binding sites for selected functional groups found in over 3990 high-performing MOFs.....	147
Figure 6.10: CO_2 gas uptake and at 1 bar, 298 K for high-performing MOFs obtained from the GCMC simulations using MEPO-QEQ charges as compared with those obtained using REPEAT charges. The line of perfect correlation is shown in red.....	148
Figure 6.11: a,c) Two most common binding site motifs identified with a maximum clique analysis of the high performing MOFs. The distribution of the CO_2 molecules within the binding sites are shown as the isosurfaces (red for oxygen and gray for carbon). b) and d) show examples of each binding site motif within a representative hypothetical MOF. Gray: carbon, purple: vanadium, red: oxygen, blue: nitrogen, white; hydrogen.....	149
Figure 6.12: a) Isosurface plots of the probability distribution of CH_4 in a high performing MOF containing S1 binding motif at isovalue 0.01; front view (right) and side view (left). b) Isosurface plots of the probability distribution of CO_2 in the same MOF at isovalue 0.03; front view (right) and side view (left).	151
Figure 6.13: Three examples of high-performing MOFs containing 1-D sheet-like channels spaced by 7 Å with planar poly-aromatic organic SBU: a) A MOF with sra topology which is assembled of the conjugated organic SBUs coordinated to 1-dimensional chains of vanadium oxide, and b & c) MOFs with pcu topology in which the large poly-aromatic organic SBUs are connected to hexacoordinate copper and zinc paddlewheels.....	154
Figure 6.14: CO_2/CH_4 selectivity of 31,180 MOFs identified as possible layered sheet-like MOFs vs. the maximum free pore diameter with a color mapped S1 density.....	156
Figure 6.15: CO_2/CH_4 selectivity of ~31,000 MOFs identified as possible layered sheet-like MOFs vs. the working capacity with a color mapped S1 density.	157
Figure 6.16: a) The optimal alignment of poly aromatic organic SBUs separated by ~7Å in MM-optimized structure of a high-performing layered sheet MOF b) The same MOF following optimization at the DFT level showing distortions in the aromatic sheet.....	161
Figure 6.17: a) CO_2/CH_4 selectivity and b) CO_2 working capacity of the 14 top-performing MOFs in the presence and absence of water. c) shows the SBUs codes in the labels given in A).	163
Figure 6.18: The structure of one of the top candidates for CO_2/CH_4 gas separations for natural gas purification that has the highest computed selectivity in humid conditions. It is composed of pillared V SBU bonded to a Pyrene organic SBU functionalized with propyl groups.....	164
Figure 6.19: Structural Building Units (SBUs) and functional groups used to generate the ~350,000 hypothetical MOFs from the Northwestern University database.	169

Figure 7.1: Examples of the inorganic (a), organic (b) SBUs used to generate the 324,500 hypothetical MOFs in our database. 174

Figure 7.2: Distribution of CO₂/CH₄ selectivity and CO₂ working capacity of entire database. Black dashed lines represent the ‘high-performing’ thresholds used in the QSPR models, while the blue dashed line shows the selectivity and working capacity of zeolite 13x that is currently in use. 177

Figure 7.3: Interaction scatter plots of the max. pore diameter, helium void fraction and surface area of 324,500 MOFs in the calibration data set with color mapped (a) CO₂/CH₄ selectivity and (b) CO₂ working capacity. Each hexbin contains more than one MOFs and the color represents the average (a) CO₂/CH₄ selectivity or and (b) CO₂ working capacity of MOFs. The colored dots correspond to the experimental working capacity and selectivity of published MOFs. (Cu-BTC³¹, Zn-Atz-oxalate³², MOF-5³³, Carborane-MOF³⁴, Mixed-ligand MOF³⁵, UMCM-1³⁶, bio-MOF-11³⁷, MOF-177³⁸, MIL-47³⁹). 179

Figure 7.4: Calibration receiver operating curve (ROC) plots that show how well the decision tree (DT) models label MOFs as “positive” or “negative” depending on whether the CO₂ working capacity is greater or less than 2 mmol/g (a) and greater or less than 4 mmol/g (b); and CO₂/CH₄ selectivity greater or less than 5 (c) and greater or less than 10 (d). The MOFs are labelled correctly most of the time, as indicated by the high AUC value. The dashed line represents the curve that would be obtained by random labelling of the MOFs as positive or negative. 182

Figure 7.5: Calibration receiver operating curve (ROC) plots that show how well the support vector machine (SVM) models label MOFs as “positive” or “negative” depending on whether the CO₂ working capacity is greater or less than 2 mmol/g (a) and greater or less than 4 mmol/g (b); and CO₂/CH₄ selectivity greater or less than 5 (c) and 10 (d). The MOFs are labelled correctly most of the time, as indicated by the high AUC value. The dashed line represents the curve that would be obtained by random labelling of the MOFs as positive or negative. 183

Figure 7.6: Decision tree (DT) models of the CO₂ working capacity higher than 2 mmol/g (a) and 4 mmol/g (b) of the 32,450 MOFs in the calibration set. Each branch of the tree represents a threshold level given for the value of a property, where the MOFs with values higher than the level threshold branched to the right, while MOFs with lower values branched to the left. High-performing nodes are highlighted in gray. The scores in the lowest branches are averaged QSPR scores over all the MOFs in that branch. 186

Figure 7.7: Decision tree (DT) models of the CO₂/CH₄ selectivity higher than 5 (a) and 10 (b) of the 32,450 MOFs in the calibration set. Each branch of the tree represents a threshold level given for the value of a property, where the MOFs with values higher than the level threshold branched to the right, while MOFs with lower values branched to the left. High-performing nodes are highlighted in gray. The scores in the lowest branches are averaged QSPR scores over all the MOFs in that branch. 187

Figure 7.8: Number of MOFs identified by the SVM classifiers as having a working capacity greater than 2 mmol/g (a) or greater than 4 mmol/g (b) and CO₂/CH₄ selectivity greater than 5 (c) or greater than 10 (d)

from the test set of 292,050 MOFs. Three different sensitivity cut-offs are reported for each. The dashed line indicates the total number of actual high-performing MOFs (all of the true positives), whereas the number in parenthesis gives the percentage of the total test set that the QSPR classifier identified as high-performing. 189

Figure 7.9: Distribution of the calculated CO₂ working capacity and (b) CO₂/CH₄ selectivity of all true positives in the test set using the classifiers of working capacity >2 mmol/g and selectivity>5, compared to the distribution of true positives recovered by the QSPR classifiers at different cut-offs..... 191

Figure 8.1: Averaged electronegativity weighted AP-RDF plot of two MOFs, a) MIL-47 and b) Cu-BTC.. 203

Figure 8.2: Example of Voronoi decomposition of two-dimensional space containing n sites, (red circle) into n cells such that each cell consists of all points closer to its site than to any other. 205

Figure 8.3: a) MIL-47 (purple: V, red: O, cyan: C, and white: H) and its 3D Voronoi network represented by edges (orange lines) and nodes (orange points). b) The CO₂ accessible Voronoi networks are highlighted in purple within the Voronoi network of MIL-47. For clarity of presentation, nodes are represented by spheres of equal, small radii..... 206

Figure 8.4: Voronoi hologram of the MOFs Cu-BTC (a) and MIL-47 (b). Features are assigned a color on the basis of their frequency of occurrence..... 207

Figure 8.5: A schematic representation of the steps involved in a generalized genetic algorithm optimization. 208

Figure 8.6: Schematic of the 1-cut mating process 210

Figure 8.7: Schematic of the 2-cut mating process 211

Figure 8.8: Schematic of (a) flip mutation and (b) swap mutation..... 211

Figure 8.9: Distribution of CO₂/CH₄ selectivity and CO₂ working capacity of entire database. Black dashed lines represent the ‘high-performing’ thresholds used in the current QSPR models, while the blue dashed line shows the selectivity and working capacity of zeolite 13X that is currently in use 213

Figure 8.10: Sensitivity (blue bar) and specificity (red bar) values at maximum G-mean score of the all classifiers for a) CO₂/CH₄ selectivity and b) CO₂ working capacity. The gray bars indicate the ratio of MOFs flagged as high-performing by the classifier to the total number of MOFs in the database. 217

Figure 8.11: Sensitivity of the classifiers for recognizing MOFs with high performance for CO₂/CH₄ selectivity categorized according to inorganic SBUs..... 219

Figure 8.12: Sensitivity of the different classifiers for recognizing high-performing MOFs for CO₂/CH₄ selectivity categorized according to functional groups. 220

Figure 9.1: The query for vanadium-based MOFs with the CO ₂ /CH ₄ selectivity and CO ₂ working capacity above 15 and 4 mmol/g, respectively.....	228
Figure 9.2: A resulting first page for the query of vanadium based-MOFs with selectivity and working capacity higher than those of Zeolite 13X.	229
Figure 9.3: An example of a view page of a hit resulting from structure type search	230
Figure 9.4: The query for MOFs which have surface area of at least 4000 m ² g ⁻¹ and contain naphthalene with CH ₃ CH ₂ and OCH ₃ functional groups as part of their ligands.....	232
Figure 9.5: Examples of two hits resulting from the query for MOFs with a hydrophobic nature and a surface area higher than 4000 m ² /g. Both hits have CO ₂ uptakes higher than Zeolite 13X (6.4 and 5.1 vs 4.2 mmol/g)	233
Figure 9.6: The interface in MOFFinder. The 'Structure type' search is in the upper left including five options namely Functional Groups, Organic SBU, Inorganic SBU, Metal and General. The 'Structure query' field is below the 'Structure type' to write a query as text. The 'Sort by' option can be used to order the search results by sorting the numerical properties based on the user's request. The minimum and maximum options of adsorption and geometric properties are on the upper right. The 'Organic SBU SMILES' search which is on the lower left is responsible for similarity, substructure and exact search. The 'Sketch' button invokes a pop-up the chemical structure editor.	234
Figure 9.7: a) A graphical representation of the 'Structure type' search where MOFs can be filtered by adsorption properties, functional groups, and inorganic and organic SBUs. b) 'Organic SBU SMILES' search where functionalized organic SBUs SMILES can be filtered by 'Exact', 'Substructure', and 'Similarity' search, and then the resulting SMILES are related to their parent MOFs via FKs.	236
Figure 9.8: The database layer of MOFFinder consists of three main tables. All of the structure names with their geometric features, adsorption properties and structural representations of SBUs are stored in the MOFSEARCH_MOF table using their related fields. The SMILES representations of unique ligands are stored in the MOFSEARCH_UNI_LIGAND_BINGO table. The MOFSEARCH_MOF_LIGAND table provides the relation between MOFSEARCH_MOF and MOFSEARCH_UNI_LIGAND_BINGO tables. Primary/foreign key constraints are shown with solid arrows. PK=Primary Key, FK= Foreign Key.....	241

Table of Tables

Table 1.1: Pounds of Air Pollutants Produced Per Billion Btu of Energy	1
Table 1.2: Commercialized PSA process for CO ₂ separation from CH ₄	6
Table 2.1: Examples of kernel functions used in support vector machine.....	59
Table 3.1: Lennard-Jones parameters from the UFF.	67
Table 3.2: REPEAT partial atomic charges for the X-ray structure of CuBDPMe.....	67
Table 4.1: Lennard-Jones parameters from the UFF.	86
Table 4.2: REPEAT partial atomic charges (q/e) for the X-ray structure of (1) MIL-47, (2) MIL-47-OH, (3) MIL-47-NH ₂ , (4) MIL_47-OME, (5) MIL-47-CL, (6) MIL-47-Br, (7) MIL-NH ₂ , and (8) MIL_47-COOH	88
Table 5.1: The UFF Lennard-Jones parameters for atoms in the MOFs studied in this work	107
Table 5.2: Partial atomic charges (e) of ZIF-8, MIL-47, UiO-66, Al(OH)-MOF, Zn-DMO-TM, Zn-DMOF-A, and Cu-BTC.....	109
Table 5.3: Force field parameters and geometries of SPC/E, TIP3P, TIP4P, TIP4P/Ew, TIP5P, and TIP5P/Ew	110
Table 6.1: CO ₂ working capacities (WC) and CO ₂ /CH ₄ selectivities (S) of the top high-performing MOFs optimized at the MM and DFT level. Each MOF in the table is described by the functional group and SBUs which were used to build them.	158
Table 7.1: LJ parameters for methane, CO ₂ and framework atoms in all hypothetical MOFs.	175
Table 7.2: Area under the curve (AUC) values of the receiver operator curve (ROC) plots of the recognition of high-performing MOFs for CO ₂ /CH ₄ separations using DT and SVM models of pore size, void fraction and surface area.	181
Table 7.3: Number of the top 1000 MOFs from the test set that the classifiers of working capacity >2 mmol/g and selectivity>5 correctly identifies with different cut-offs. The percentage in parenthesis is the fraction of the test set that would require GCMC simulations.....	193
Table 8.1: Tabulated atomic properties used to weight the radial distribution function (RDF) descriptors. Each property value is normalized by the values for the carbon atom.....	204
Table 8.2: G-mean, sensitivity, and specificity values for all of the SVM classifiers for CO ₂ /CH ₄ selectivity at 10 bars, and CO ₂ working capacity. Adsorption and desorption pressures are 10 and 1 bars, respectively, and the temperature is 298 K.....	215

Table 8.3: Area under the curve (AUC) values of the receiver operator curve (ROC) plots of the recognition of high-performing MOFs for CO ₂ /CH ₄ separations.	216
Table 8.4: Number of the top 1000 MOFs from the test set that the classifiers of working capacity and selectivity correctly identifies with different cut-offs. The percentage in parenthesis is the fraction of the test set that would require GCMC simulations.....	221
Table 9.1: MOFFinder components with their functionality	237

List of Acronyms

• ABSL	Automatic Binding Site Localization
• API	Application Program Interface
• BTC	Benzene Tri-Carboxylate
• 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
• CCS	Carbon Capture and Storage
• CPMD	Carr Parrinello Molecular Dynamics (software)
• MEA	Monoethanolamine
• MEPO	MOF Electrostatic Potential Optimized
• MOF	Metal Organic Framework
• PAW	Projector Augmented Wave
• PSA	Pressure Swing Adsorption
• QEq	Charge Equilibration
• QM	Quantum Mechanics
• RDF	Radial Distribution Function
• CSD	Cambridge Structural Database
• DFT	Density Functional Theory
• L-J	Lennard-Jones (function)
• MC	Monte Carlo
• MCT	Maximum Clique Threshold
• MD	Molecular Dynamics
• REPEAT	Repeating Electrostatic Potential Extracted Atomic (charge)
• SBU	Structural (or Secondary) Building Unit
• TSA	Temperature Swing Adsorption
• UFF	Universal Force Field
• VASP	Vienna Ab initio Simulation Program (software)
• NG	Natural Gas

Acknowledgements

The most important years of my life have been spent in the Woo lab since I joined the Woo group in 2010 as a PhD student from Iran. I'm feeling particularly grateful for the lessons I have learned during my time here. I believe that my journey to this thesis could not have been completed without the support of my supervisor, Tom Woo. From the start to the end of this journey, his creative research plans provided me with project ideas in both method development and applied fields, which has enabled me to grow and improved my skills. Tom is also interested in group activities outside of the research environment, and has organized many events, such as movie nights and potluck parties, which I look back on with nostalgia.

During my time spent in the Woo lab, I had the opportunity to work with knowledgeable group members who supported me in my research and took part in leisure activities with. Because of this, it is particularly important for me to thank all of those people. These people are James Hooper, Nick Trefiak, Mike Nohra, Arif Ismail, Bianca Provost, Carlos Campañá, Tom Daff, Michael Fernandez, Eugene Kadantsev, Sean Collins, Phil Deluna, Jason Lo, Tom Burns, Hana Dureckova, and Seaid Yeganegi. I would like to extend my sincere thanks to Saman Alavi for many great discussions we had throughout this work. Also, I would like to give special thanks to Hana Dureckova for giving comments on this thesis.

I would like to thank Pegah Peyman for her valuable support during this work. Finally, special thanks must go out to my parents, Fariba and Naser, for supporting me to do this work. I have constantly in my mind all the sacrifices they have made to provide a better life for me and my little brother. I only hope I can return the favour in the many happy years to come.

Contributions to Original Research

To the best of my knowledge, I hereby declare that the research presented is original work generated during my Ph.D. studies under the supervision of Prof. Tom Woo, in the Department of Chemistry and Biomolecular Sciences at the University of Ottawa. Chapters 3 and 7 were published, the software program developed in Chapter 9 is publicly available at titan.chem.uottawa.ca/moffinger, and Chapters 6 and 8 are under preparation for peer reviewed scientific journals.

I solely generated all simulation data, performed analysis to develop the computational results, and developed computational tools while Prof. Tom Woo contributed to the direction and editing of all manuscripts. Other researchers who contributed to this work are Prof. George Shimizu, Dr. Tom Daff, Dr. Peter Boyd, and Dr. Michael Fernandez. Prof. George Shimizu contributed to this work by providing the X-ray diffraction structure of the MOF studied in Chapter 2. Dr. Tom Daff contributed to this work by writing the GCMC program with which some of the results of this thesis were generated. Dr. Peter Boyd contributed to this work by developing the hypothetical MOFs database and binding site motifs discovery program in Chapter 6. Dr. Michael Fernandez contributed to this work by providing the Python Machine Learning module in Chapter 7.

1 Introduction

Many scientific studies show that rising emissions of greenhouse gases, such as CO₂, from coal- or oil-based energy sources, have caused climate change (i.e., global warming) and air pollution,^{1,2} which can result in unavoidably devastating, dangerous outcomes in the future such as the rising of global sea levels up to seven meters³, massive loss of aquatic life⁴, and global water shortages⁵. Therefore, an immediate reduction of the greenhouse gas emissions is essential to prevent these catastrophic effects on the earth. As a near-term strategy, using natural gas as a primary clean energy source has been proposed.⁶ Natural gas, one of the cleanest fossil fuels, is mainly composed of methane and is available in large quantities. It presents many environmental advantages over petroleum and coal, as shown in Table 1.1.⁷ For

example, for a given amount of energy production, it produces about half the amount of

Table 1.1: Pounds of Air Pollutants Produced Per Billion Btu of Energy

Pollutant	Natural Gas	Oil	Coal
Carbon dioxide	117,000	164,000	208,000
Carbon monoxide	40	33	208
Nitrogen oxides	92	448	457
Sulfur dioxide	0.6	1,112	2,774

CO₂ and a third of the amount of nitrous oxides compared to coal and other petroleum fuels. For this reason, the David Suzuki foundation has advocated the use of natural gas as a ‘bridging fuel’ until we can meet our intense energy needs with renewable, zero-emissions energy sources.⁸ In addition, other factors such as novel transport technology, the abundance of natural gas reservoirs, and the lower overall costs all result in the increase of natural gas demand. The International Energy Agency (IEA) estimates a 37.8 % increase in natural gas consumption between 2008 and 2035, as part of the 33.2% increase in total energy consumption under its own New Policies Scenario.⁹ Recently, the U.S. Energy Information

Administration (EIA), forecasted a greater increase in the demand for natural gas from 2013 to 2040 compared to coal (13% vs 5%), and predicted a low contribution for the renewables and nuclear energy in the increase of energy consumption in 2040 (< %18 of total energy consumption).¹⁰

1.1 Natural Gas in Canada

Natural gas, a high-profile energy source in Canada, has been produced more than any other form of energy source in the last two decades. Specifically, in 2008, 40% of primary energy produced in Canada was natural gas, which provided \$33 billion for Canadian producers through export to the U.S. by pipeline, indicating the significant role of natural gas in the Canadian market. By the year 2020, the natural gas demand is expected to increase by 18% in Canada, as part of the 13% increase in Canada's total energy demand.⁸ Although Canada has an extremely large number of natural gas reservoirs, which could supply more than the world's future demands, much of these reservoirs are in shale gases or other so-called unconventional gas reservoirs. In particular, the National Energy Board of Canada estimates that Canada is sitting on about 1000 trillion cubic feet of shale gas¹¹, which amounts to about one sixth of the World's current proven natural gas reserves. Thus, shale gas and other unconventional natural gas sources are receiving significant attention from government agencies and energy companies. Aside from the difficulties associated with extracting shale gas, another problem with unconventional gas reservoirs is that they contain variable amounts of contaminants such as water, light paraffins, aromatics, nitrogen, sulphur compound and carbon dioxide. The CO₂ concentrations in these reservoirs can reach above 15 vol%, which is high compared to ~2% in conventional reservoirs. The high concentrations of CO₂ and other contaminants not only

reduce combustion efficiency and increase transportation costs, they also cause pipeline corrosion. Therefore, before the gas is delivered in pipelines, each contaminant must be removed to its standard level as defined by pipeline specifications. To meet these specifications, natural gas extracted from shale gas usually undergoes some type of treatment or purification.

1.2 Natural Gas Processing

Because the composition of natural gas varies from reservoir to reservoir, there are a variety of processes for natural gas purification. Figure 1.1 shows a simplified representation of natural gas processing, where some steps may be integrated into one unit, performed in a different order, or omitted. The summary of each unit function is provided below.

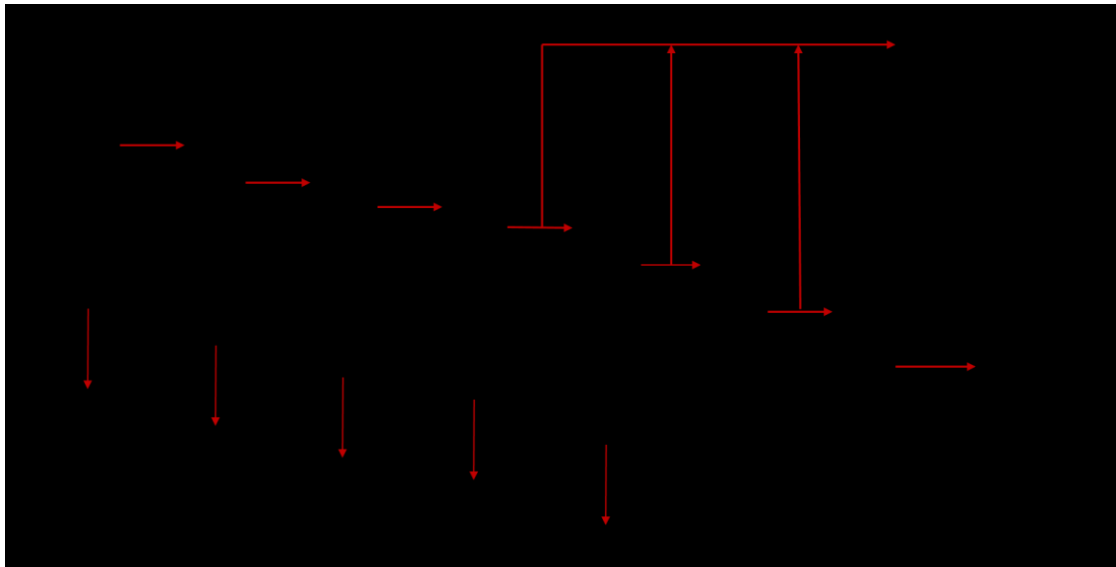


Figure 1.1: Simplified schematic representation of natural gas processing.

In the condensate separator unit, the higher hydrocarbons (two to eight carbons) are recovered from natural gas usually through condensation process, where the temperature of gas is reduced through refrigeration units which frequently use vapor-compression cycles with

propane as the refrigerant. In the dehydration unit, the water vapor content of natural gas is reduced to a concentration no greater than 112 mg of water per standard of 1 m³ of gas, which is a typical water pipeline specification across Canada.¹² This reduction is often provided through an absorption process, where water is absorbed by a glycol-based liquid system, such as triethylene glycol. In the desulfurization unit, the hydrogen sulfide is removed from natural gas through a chemical sorption process, where the iron oxide sorbents are widely used. In this process, the iron oxide sorbents remove hydrogen sulfide by forming iron sulfide, and the saturated sorbents (i.e., no longer able to react with hydrogen sulfide) are regenerated by oxidizing the iron sulfides with air, forming iron oxide and elemental sulfur. In the CO₂ removal unit, the amine-based absorption process is primarily used to remove CO₂ from natural gas, and recently the adsorption process, discussed in the next section, with the use of zeolites is also employed. The amine-based absorption process involves forming carbamate salts through the reaction of CO₂ with an amine solvent such as methyldiethanolamine, and regenerating the whole solution at near boiling point. In the optional nitrogen extraction unit, nitrogen is removed from natural gas based on the difference in boiling points of nitrogen and methane. This involves the cryogenic distillation of a gas that has been purified to very low levels of CO₂, water, and other components that could form a solid at low temperatures. In the optional demethanizer unit, a gas that has been preconditioned to low levels of CO₂, water vapor and nitrogen undergoes a cryogenic process to separate methane from the remaining higher hydrocarbons. In this process, the temperature of gas is lowered to -120 °F to condense higher hydrocarbons, while maintaining methane in its gaseous form. In the optional fractionation unit, the higher hydrocarbons stream which has been removed from the natural gas stream is

separated into individual hydrocarbons based on the varying boiling points of individual hydrocarbons. As such, heating units raise the temperature of the stream at different stages, and in each stage, one hydrocarbon is liquefied and injected into a specific holding tank.

All above-mentioned units have been engineered to reduce both energy and monetary cost. However, while the amine-based CO₂ removal process have been widely used in other CO₂ sequestration applications¹³, such as CO₂ separation from hydrogen since the 1930's, it is energetically costly and accounts for a significant portion of the monetary costs. For example, it is estimated that the annual amine treating cost for CO₂ removal from 100 MMscfd (Million standard cubic feet of gas per day) of natural gas containing 10% CO₂ is \$20MM US (\$66US/ton CO₂), which is 40% of the cost for the total process.¹⁴ Such high energetic costs are attributed to the high level of heat required to regenerate a saturated amine solution and to break the N-C carbamate bond. In contrast, pressure swing adsorption (PSA) technology designed based on the adsorption-process is an efficient and lower-cost technology (compared to absorption-based technology) due to its lower energy demands and its operation conditions which are well-suited to the natural gas conditions.

1.3 PSA-based CO₂ removal process

PSA technology for removing CO₂ from natural gas is designed based on the use of porous solid materials which can selectively bind CO₂ via physical adsorption. The strength of CO₂ interaction with solid sorbants is typically smaller than that of aqueous amines (20 - 40 kJ/mol vs. 90 kJ/mol for aqueous amines) allowing for easier regeneration of these materials. Furthermore, while aqueous amines need to be boiled in the regeneration step which requires high levels of heat due to its high heat capacity, the regeneration of solid sorbants in the PSA

system occurs at ambient temperature and pressure. A general schematic of a PSA-based process for CO₂ removal from natural gas is provided in Figure 1.2. In this process, a gas mixture is fed into a column containing the solid material at high pressure and room temperature (typically 3-20 bar and 298-303 °K). CO₂ is selectively adsorbed onto the solid material and subsequently the methane-rich gas is emitted from the column. Once the column is not able to adsorb more CO₂, the saturated column undergoes a regeneration step through decreasing the feed pressure to a near-ambient pressure. In this step, CO₂ is desorbed from the solid material, and the column will be ready for use in further adsorption/desorption cycles. Listed in Table 1.2 are examples of commercially available PSA-based systems for CO₂ recovery from natural gas.

Table 1.2: Commercialized PSA process for CO₂ separation from CH₄

	Vendor/Licensors	Adsorbent	Refs
M3100 Xebec Rotary-valve	Xebec	Metal-based	¹⁵
CO ₂ Sponge	IACX Energy	Activated Carbon	¹⁶
Molecular Gate	Guild Associate	Titanosilicate	¹⁷
UOP MOLSIV	UOP	Zeolite	¹⁸

Most materials commonly used in these systems as CO₂ scrubbers are zeolites and activated carbons. These materials have a strong affinity for CO₂ over CH₄, but the amount of CO₂ recovered per unit mass of these materials at each regeneration cycle, defined as working capacity, is relatively low compared to other porous materials. For example, zeolite 13x, the mostly commonly used zeolite for this purpose, has a CO₂ working capacity of 1.86 mmol/g¹⁹, which is only about half of the maximum working capacity recently reported for porous materials.²⁰ Since the working capacity of a sorbent dictates the amount of sorbent required, using a material with high CO₂ working capacity leads to an overall reduction in the cost of CO₂ capture. More specifically, Wiley and co-workers²¹ in their cost analysis of the PSA process for a CO₂ separation application revealed that using a material which could recover twice as much

CO₂ as Zeolite 13X at regeneration step and adsorb three times as much CO₂ relative to other gases as Zeolite 13X would reduce the cost of CO₂ capture up to 30%. Motivated by these findings, several types of porous materials have been proposed to replace traditional adsorbents in PSA application for CO₂ separation from natural gas.²² Among the proposed materials, Metal Organic Frameworks (MOFs)²³ show great potential due to their exceptionally high porosities and chemically tunable structures.

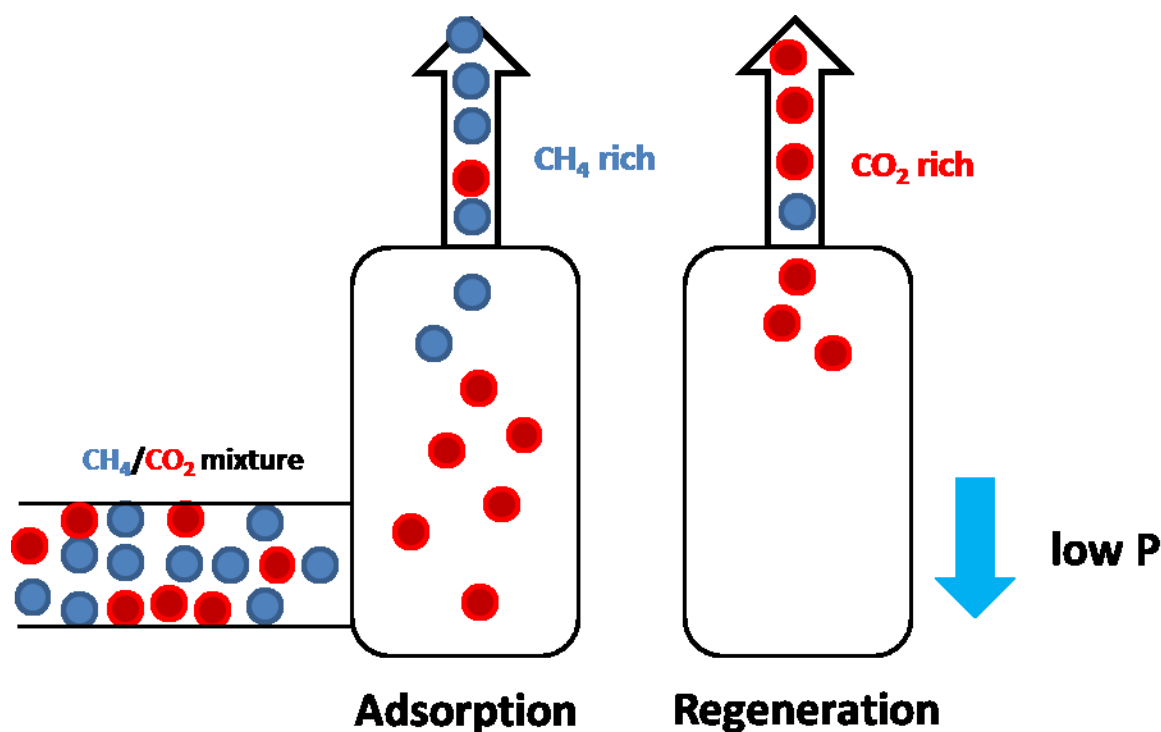


Figure 1.2: Graphical representation of PSA process for removing CO₂ from natural gas. CO₂ is selectively trapped in a solid material, which undergoes regeneration to release CO₂ and make the materials useable for further adsorption/regeneration cycles.

1.4 Metal Organic Framework Materials

MOFs, which are porous crystalline materials composed of metal clusters joined by organic linker molecules, have invigorated the porous material community since the late 90's when Yaghi et al.²⁴ discovered MOF-5, now a widely studied MOF, and used reticular chemistry

as a tool for rational design of new MOFs.^{25,26} To date, tens of thousands of these materials have been synthesized, and many avenues of applications have been pursued, such as gas storage and separation,^{27,28} drug delivery,²⁹ and heterogeneous catalysis.^{30–32} However, despite showing great promise, these materials are still in transition period from research laboratories to industrial settings. While the term “MOFs” which is adopted by IUPAC is a popular term for these materials, due to the nature of its development, many other conventional names have also emerged to describe this type of material. The most commonly used term after “MOFs” is Porous Coordination Networks (PCNs),³³ followed by Porous Coordination Polymers (PCPs).³⁴ Other names include Zeolitic Imidazolate Frameworks (ZIFs),³⁵ Microporous Coordination Polymers (MCPs),³⁶ and Metal Peptide Frameworks (MAFs).³⁷ However, in the reminder of this thesis, we will use the term “MOFs”. In terms of publication, a large numbers of papers on MOF research has been published in “high-impact” journals such as Science,^{9,25,38–43} Nature,^{24,26,27,44–46} and Chemical Reviews^{31,47–54} in a relatively short period of time.

MOFs are made of metal ions coordinated to organic molecules that can form three-dimensional structures. The most common metal ions are Cu^{+1} , Cu^{+2} , Zn^{+2} , Ni^{+3} , V^{+2} and Al^{+3} while the most common organic molecules are bidentate and tridentate carboxylics,^{55,56} and azoles.^{57,58} These materials form high symmetric crystalline lattices which can be characterized by X-ray crystallography experiments. Examples of several MOFs are given in Figures 1.3a and 1.3b. Analogies are often made between MOFs and zeolites because both are crystalline, nanoporous materials. However, MOFs offer much greater opportunities for tunability, as these materials have a larger variability in terms of metal ions and organic linkers compared to traditional zeolites that are composed solely of aluminum, oxygen, and silicon.⁵⁹ Also, novel

zeolite synthesis is much more difficult than novel MOF synthesis. Whereas less than 200 different zeolite structures have been synthesized to date,⁶⁰ thousands of MOFs have been synthesized in a much shorter time period.³⁹ Moreover, some MOFs have extremely large internal surface areas (greater than 6000 m²/g) and exceptional porosity with up to 90% free volume.⁶¹ Owing to these unique properties, MOFs have emerged as an ideal candidate for a variety of applications, including CO₂ separation from natural gas.

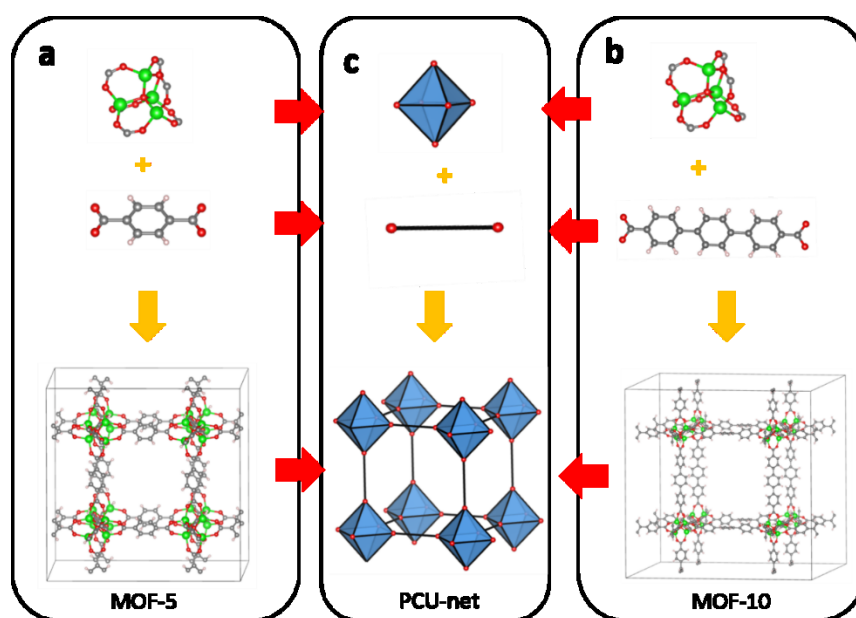


Figure 1.3: a) MOF-5 from its inorganic Zn₄O complex and terephthalic acid linker. Green: zinc, white: hydrogen, red: oxygen, grey: carbon. b) MOF-10 from its Zn₄O complex and p-terphenyl-4,4'-dicarboxylate linker. c) pcu topology of MOF-10 and MOF-5 where the Zn₄O complex is abstracted to the octahedral geometry and the benzene-based linkers abstracted to an edge bridging two octahedrons.

The rational design of MOFs can be attributed to the terms of reticular synthesis.⁶² The idea behind reticular synthesis is the self-assembly of rigid modular building blocks into nets through strong bonds. Defined in this fashion, the periodic structure of MOFs can be partitioned into organic and inorganic structural building units (SBUs), which can be seen in Figure 1.3. Using a simplified representation of the clusters of atoms in each SBU as a simple

geometric shape in edges and vertices format based on their extended connectivity, an underlying network topology (net) can be formed. It is worthy to note that these are the nets and SBUs that guide experimental chemists in synthesizing MOFs with a targeted pore size, surface area, and pore chemistry with a specific topology. Perhaps the most striking example of this is IRMOF-n (1 to 16) series in which the pore chemistry was modified through the functionalization with organic groups $-\text{Br}$, $-\text{NH}_2$, $-\text{OC}_3\text{H}_7$, $-\text{OC}_5\text{H}_{11}$, $-\text{C}_2\text{H}_4$, and $-\text{C}_4\text{H}_4$ in a controlled fashion where topology and connectivity had all remained the same between the series.²⁵ One member of these series exhibits a relatively high methane uptake, $155 \text{ V}^3(\text{STP})/\text{V}$, which had been a world record up until 2010. In a similar work, the pore size of IRMOF-74-n series increased from 20 to 90 Å through increasing the size of organic linkers without changing the topology.⁶³ Surprisingly, some MOFs of these series have extremely large pores through which large bio-molecules such as vitamin B12, metal-organic polyhedron-18, myoglobin and green fluorescent protein (GFP) can pass. Beside these examples, there are many other examples in the literature which indeed demonstrate the possibility of tuning the features of MOFs for a specific application.^{64–69}

1.4.1 MOFs for CO₂ separation from natural gas

As mentioned in section 1.4, MOFs have not yet been implemented in industrial settings, which is in contrast to intense interest in developing MOFs for CO₂ separation from natural gas. For this application, an ideal material must have a high ‘working capacity’ because it has been shown that a material with high working capacity will have lower regeneration energy, leading to an overall reduction in the cost.⁴⁶ Working capacity is the amount of CO₂ recovered from a single adsorption/desorption cycle per unit weight (or volume) of material, typically expressed

in mmol/g. The working capacity not only depends on the nature of the pore in the material, but also on adsorption and desorption conditions. Beside the high CO₂ working capacity, an ideal material must be highly selective for CO₂ over CH₄, which is the largest component in natural gas. This arises from the fact that the higher the CO₂ selectivity, the higher the purity of recovered gas, leading to reduced operating cost.^{21,70} Selectivity is measured as the ratio of the amount of CO₂ adsorbed vs the amount of CH₄ adsorbed at natural gas conditions, normalized by their partial pressures (more details provided in Chapter 2).

A large number of MOFs have been synthesized and investigated for their performance in CO₂ separation from natural gas in controlled experiments.^{71–79} However, these experiments were performed somewhat in ‘stand-alone’ fashion, meaning that different thermodynamic conditions for a sorption process or/and different techniques for activating the MOFs have been used, which could affect the MOFs performance for CO₂/CH₄ separation. Indeed, a MOF can present different sorption properties in two experiments that are not identical. For example, the CO₂ uptake for MOF-5 at 1 bar and 298 K has been reported to be in the range of 1.12 to 1.98 mmol/g. The variation is primarily attributed to the activation of the MOF – in other words different methods used to remove solvent from the MOF during activation.⁸⁰ Nonetheless, a scan of the literature for an ideal MOF for this application has resulted in very few structures, making it difficult to determine if one of these structures is an ideal candidate for natural gas purification. Although the Cambridge Structural Database⁸¹ (CSD) has thousands of experimentally realised MOFs,⁸² these account for only a small fraction of the millions of MOFs that can potentially be made.⁸³ Out of these unrealised MOFs, there are likely ideal configurations with appropriate organic and inorganic molecules that are optimal for CO₂

separation from natural gas. However, due to the cost and time required to synthesize a single MOF, let alone millions of possible MOFs, experimentally determining high-performing MOFs for separation CO₂ from natural gas is far from ideal. In addition, the microscopic features of the MOF (e.g., guest-host interaction) that provide the observable macroscopic properties are incredibly difficult to elucidate through experimental methods alone. Indeed, important physical and chemical features of MOFs that are related to sorption properties are difficult to address using the experimental methods, which results in the lack of rational design principles for MOFs. As a result, the use of rational design in the development of new MOFs is very challenging from experiment alone, and thus, a MOF with ideal properties for separation CO₂ from natural gas has not yet been discovered.

1.4.2 Computational approach in MOFs discovery

As mentioned above, owing to our current lack of rational design principles, it would be very difficult and time consuming to experimentally determine high-performing MOFs for CO₂ separation from natural gas. To accelerate this process, computational methods are emerging as a powerful tool in the MOF community.^{47,84} Several studies demonstrate the potency of computational tools in the discovery of important features of MOFs for CO₂-related gas adsorption. Yang et al.⁸⁵ investigated the effect of functional groups in UiO-66 for CO₂ upgrading from natural gas by building hypothetical functionalized UiO-66 materials and calculating their adsorption properties using molecular simulations. They found that the functionalization with -COOH group leads to higher selectivity compared to UiO-66 and other functionalized MOFs of this family, which was later confirmed by Biswas et al.⁸⁶ who synthesised a series of functionalized UiO-66 materials and found that -COOH functionalized

UiO-66 exhibits the highest selectivity among others with different functional groups. In another study, using DFT calculations, Torrisi et al.⁸⁷ demonstrated that functionalization with -COOH group leads to an enhancement in CO₂ heat of adsorption at zero loading compared to other functional groups such as -CH₃, -NH₂, and -OH. This was then seen by Xiang et al.⁸⁸ and Cooper et al.⁸⁹ who experimentally demonstrated that their MOFs functionalized with -COOH group have the highest CO₂ heat of adsorption at zero loading over other functional groups, namely CH₃, -NH₂, and -OH.

While these computational studies indisputably demonstrate the important role of computational methods in material discovery processes, they have focused on a limited number of specific MOFs where the results may not be generalized for the rational design of MOFs. For example, the aforementioned studies demonstrate that the -COOH group increases CO₂ selectivity or CO₂ heat of adsorption, but it was also found from these studies and other studies the -COOH group not only does not increase CO₂ working capacity, but reduces working capacity due to its steric effects reducing adsorption sites for CO₂. Therefore, to find how to improve both CO₂ selectivity and working capacity, one needs to sample a larger physical and chemical space of MOFs. This can be successfully achieved by computationally screening larger sets of MOFs in a high-throughput fashion.

Until now, several computational screening studies have been performed on large sets of MOFs with aims of identifying high performing materials or identifying structure-property relationships for gas-related applications. One example of this is the work done by Watanabe et al.⁹⁰ who screened >1000 MOFs taken from CSD for CO₂ capture and identified four high-performing materials. Another example is the screening of 500 MOFs taken from the CSD by

Haldoupis *et al.*⁹¹ for CO₂ capture, where they were able to recognize several high-performing materials. Additional examples include the screening of >20,000 MOFs taken from the CSD by Goldsmith *et al.*⁹² for hydrogen storage, the screening of a large set of ZIF materials by Smit *et al.*⁹³ for CO₂ capture, and the screening of a large set of hypothetical MOFs by Wilmer *et al.*^{44,94} for four CO₂-related gas separation applications and methane storage. However, despite valuable achievements in these studies, no experimental validations, except for the case of methane storage, have been made so far. Also, most of these studies have focused on CO₂ capture from flue gas or high pressure methane storage, and far less attention has been paid to separate CO₂ from natural gas under natural gas conditionings. Moreover, in the case of screening of hypothetical MOFs by Wilmer *et al.*,⁹⁴ the results (e.g., the optimal range of surface area (1500 - 2000 m²/g) for CO₂ capture) may not represent the required features for CO₂ capture, as they mention in their work that their hypothetical MOFs database may not be sufficiently diverse. As a result, the data in their work may be skewed towards a particular distribution. More importantly, it has been shown by Woo and coworkers⁹⁵ the methods used in their study had resulted in overestimating the performance of MOFs containing fluorine and chlorine functional groups.

Generating massive libraries of hypothetical materials using modular building-block approach have resulted in hundreds of thousands of hypothetical MOFs structures so far^{44,93,96} which are readily available for high-throughput (HT) computational screening. Recent high-throughput computational screening on these structures for gas adsorption properties have not only shown the capability of this approach in material discovery but also have pointed to a direction of new screening studies for researchers. Furthermore, it has paved the way for

developing new computational tools that enable screening the millions of MOF structures that can be easily constructed computationally.⁸³

1.5 Thesis Goals and Outlines

In this work, we develop efficient computational tools to aid in the discovery of new MOF materials for methane purification, and computationally screen hundreds of thousands of hypothetical MOFs for their performance in natural gas purification. The goals of this thesis will be to identify important aspects relevant for methane purification within MOFs in order to guide experimental synthesis, and identify 'hit' materials for targeted synthesis. To perform screening of large database of MOFs, we define the protocol including which simulation methods and parameters will most effectively predict the adsorption and selectivity properties important for methane purification. Following the screening, we will analyse the screening data to find the structural and chemical features that provide the best performance.

To evaluate MOFs' performance for natural gas purification, we simulate the adsorption of a CO₂/CH₄ mixture in MOFs under PSA conditions for NG purification using established statistical mechanical methods. These simulations use classical molecular mechanics force fields, which do not consider electrons but rather treat atoms as basic units for description of the system, to compute interaction energies between the gases and the MOF framework atoms. As a complement to our work, we also evaluate the effect of water on MOFs performance for natural gas purification. Thus, one chapter of this thesis will be allocated to study water adsorption on MOFs and its effect on CO₂/CH₄ separations within MOFs.

The thesis is organized as follows. In Chapter 2, we present a detailed description of the criteria used to evaluate MOFs for natural gas purification and the computational methods used

in this thesis. The computational methods discussed include Grand Canonical monte Carlo simulation for determining gas uptake in MOFs, the charge equilibrium method for charge calculation, and finally Quantitative Structure–Property Relationship approaches used to develop models.

In Chapter 3, we study CO₂ adsorption on the first reported Phosphonate Monoesters-base MOF, CuBDPMe. This MOF that was made up of Cu(1,4-benzenediphosphonate bis(monomethyl ester)) exhibits modest CO₂ capacity with a high isosteric heat of adsorption of 45 kJ·mol⁻¹. The characteristics of CO₂ uptake, including the adsorption isotherms, heats of adsorption, and binding sites, are modelled using well-established simulation methods. Then, we will show that the CO₂ uptake is extremely sensitive both to the flexing of the structure and to the orientation of the alkyl tether. This work was published in the Journal of the American Chemical Society.⁹⁷ In Chapter 4, we study MIL-47, a promising candidate for use in natural gas purification, and its functionalised derivatives (i.e., MIL-47-X, X = -Cl, -Br, -CH₃, -CF₃, -OH, -OCH₃) for their performance on natural gas purification under PSA conditions. CO₂ working capacity and CO₂/CH₄ selectivity relevant for methane purification are computed using molecular simulations. Using simulation data, we will investigate the effect of functional groups on this application in MIL-47-X. Finally, we compare the performance of these materials with that of other materials that are already reported in literature.

In Chapter 5, we present a comprehensive simulation study of water adsorption in a variety of MOFs and examine the effect of water on their performance for natural gas purification. The MOFs chosen are MIL-47, ZIF-8, Al(OH)(1,4-naphthalenedicarboxylate), UIO-66, Zn-DMOF-A, Zn-DMOF-TM, and Cu-BTC. First, we evaluate the accuracy of seven

“standard” models, including SPC/E, TIP3P, TIP3P/Ew, TIP4P, TIP4P/Ew, TIP5P, and TIP5P/Ew, by calculating pure water isotherms and comparing to experimental data for each of these MOFs. Afterwards, we investigate the adsorption of the binary mixture of CO₂-CH₄ on these MOFs in the presence of water under natural gas conditions. We show that a trace of water has a marginal effect on performance of hydrophobic MOFs for natural gas purification.

In Chapter 6, we present a computational screening study of 324,500 hypothetical MOFs for their performance in natural gas purification at the typical condition of natural gas purification with PSA systems. We will define the protocol we use in our screening. From screening data, we will identify important physical and chemical features that promote both high CO₂ working capacity and high CO₂/CH₄ selectivity in MOFs. Finally, we will identify several “hit” materials for targeted synthesis and commercialization.

In Chapter 7, we develop machine learning quantitative structure–property relationships (QSPR) models of the CO₂ working capacity and CO₂/CH₄ selectivity at conditions typical of natural gas purification with PSA systems using the purely geometric properties of pore size, void fraction and surface area. The models are trained on the computed adsorption properties of 32,450 MOFs using well-established cheminformatics tools, and subsequently the accuracy of these models are tested on a set of ~290,000 diverse MOF structures. We will demonstrate that the models allow for the accurate identification of high-performing MOFs, thereby allowing for rapid pre-screening of materials in order to significantly reduce the number of computationally expensive molecular simulations. This work was published in the European Journal of Inorganic Chemistry (invited article for the special issue on MOFs).⁹⁸

In Chapter 8, we present machine learning QSPR models for CO₂/CH₄ separation parameters (i.e., CO₂ working capacity and CO₂/CH₄ selectivity) that incorporate separately the Voronoi hologram and atomic property weighted radial distribution function (AP-RDF) descriptors. We will show how a genetic algorithm is implemented to efficiently select the best subset of hologram descriptors for use in QSPR model for recognizing high-performing MOFs for methane purification. Subsequently, the performance of the best subset of hologram descriptors and AP-RDF in QSPR models will be compared to simple geometrical features, namely pore size, void fraction and surface area.

In Chapter 9, we develop a novel, cheminformatics tool, MOFFinder, for aiding in the discovery of new MOFs for CO₂ scrubbing from natural gas. MOFFinder is a web-based interface to the group's databases of MOFs that provides searching capabilities based on the Secondary Building Units, Geometric Features, Functional Groups and Adsorption Properties. Currently the MOFFinder provides fast access (in seconds) to a set ~300,000 hypothetical metal organic frameworks, but can be expanded to the group's more current data base of more than 2 million MOF structures. MOFFinder provides, for the first time the substructure/similarity query of porous materials for users and is publicly available on titan.chem.uottawa.ca/moffinger.

Finally, in Chapter 10 conclusions will be provided for each part of the thesis work conducted, and future directions will be discussed. The reminder of this Chapter presents a detailed description of the criteria used to evaluate MOFs for natural gas purification and the computational tools used in this thesis.

1.6 References

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

- 106 (6), 1704.
- (2) Liu, Y.; Ye, Q.; Shen, M.; Shi, J.; Chen, J.; Pan, H.; Shi, Y. *Carbon dioxide capture by functionalized solid amine sorbents with simulated flue gas conditions*; 2011; Vol. 45.
 - (3) Nicholls, R. J.; Cazenave, A. *Science* **2010**, 328 (5985), 1517.
 - (4) McLaughlin, J. F.; Hellmann, J. J.; Boggs, C. L.; Ehrlich, P. R. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, 99 (9), 6070.
 - (5) Wheeler, T.; Braun, J. Von. **2013**, 288 (2012), 285.
 - (6) Liang, F.; Ryvak, M.; Sayeed, S.; Zhao, N. *Chem. Cent. J.* **2012**, 6 (Suppl 1), S4.
 - (7) Outlook, A. E.; Protocol, K.; Markets, E.; Activity, E.; States, U. *Energy Inf. Adm. Nat. Gas 1998 Issues Trends* **2012**, 49.
 - (8) DSF; PI. *Is natural gas a climate change solution for Canada?*; David Suzuki fundaton: Vancouver, 2011.
 - (9) Zielinska-Lipiec, A.; Czyrska-Filemonowicz, A.; Ule, B.; Nagode, A.; Tsukamoto, S.; Tabuch, M.; Liu, Y.; Shirane, T.; Tabuchi, M.; Abe, F.; Kondo, M.; Kubo, K.; Albert, S.; Sonderegger, B.; Sklenicka, V.; Kudrman, J.; Kucharová, K.; Míková, R.; Schüller, H.; Hagn, L.; Woitscheck, A.; Schlacher, C.; Béal, C.; Sommitsch, C.; Toda, H.; Mayr, P.; Samuel, E. I.; Choudhary, B. K.; Palaparti, D. P. R.; Mathew, M. D.; Robertson, D. G.; Shibli, A.; Plesiutchnig, E.; Beal, C.; Paul, S.; Zeiler, G.; Mitsche, S.; Pešička, J.; Kužel, R.; Dronhofer, A.; Eggeler, G.; Panait, C. G. C. G.; Zielińska-Lipiec, A.; Koziel, T.; Gourgues-Lorenzon, A.-F. a.-F.; Bendick, W.; Fuchsmann, A.; Besson, J.; Norström, L.; Morito, S.; Yoshida, H.; Maki, T.; Huang, X.; Tanaka, H.; Konishi, R.; Furuhaara, T.; Middleton, C.; Brear, J.; Munson, R.; Viswanathan, R.; Mendez, F.; Radis, R.; Masuyama, F.; Maruyama, K.; Sawada, K.; Koike, J.; Mannan, S. L.; Laha, K.; Lqwhjudwlqj, V.; Ghirupdwlrq, P.; Phfkdlvpv, D. Q. G. G.; Letofsky, E.; Kunz, L.; Lukas, P.; Krauss, G.; Yin, F.; Kloc, L.; Sklenicka, V.; Kitahara, H.; Ueji, R.; Tsuji, N.; Minamino, Y.; Kimura, K.; Kern, T. U.; Wieghardt, K.; Kircher, H.; Hyde, T. H.; Becker, A. A.; Sun, W.; Yaghi, A.; Williams, J. A.; Concari, S.; Holzer, I.; Kozeschnik, E.; Cerjak, H.; Francis, J. A.; Mazur, W.; Bhadeshia, H. K. D. H.; Egner, H.; Egner, W.; Ramteke, A.; Coleman, M.; Chew, B.; Peter, G.; Burblies, A.; Hald, J.; Jefferey, C.; Rantala, J.; DeWitte, M.; Mohrmann, R.; Dimmler, G.; Weinert, P.; Bürgel, R.; Jürgen Maier, H.; Niendorf, T.; Bose, S. C.; Singh, K.; Swaminathan, J.; Sarma, D. S.; Birol, F.; Bauer, M.; Baral, J.; Ghosh, R. N.; Ashby, M. F.; Dyson, B. F.; Ahmadi, M. R.; Povoden-Karadeniz, E.; Öksüz, K. I.; Falahati, A.; Abson, D. J.; Rothwell, J. S.; Abd El-Azim, M. E.; Ibrahim, O. H.; El-Desoky, O. E. *IEA, World Energy Outlook 2010France IEA*, **2010**.
 - (10) *U.S. EIA, Annual Energy Outlook 2015*, www.eia.gov/forecasts/aeo.
 - (11) *National Energy Board of Canada, Energy Brief*, 2009.
 - (12) Carroll, J. *Natural Gas Hydrates: A Guide for Engineers*; 2, Ed.; Gulf Professional Pub, 2009, 2009.
 - (13) Rochelle, G. T. *Science* **2009**, 325 (5948), 1652.
 - (14) Arthur J. Kidnay, William R. Parrish, D. G. M. *Fundamentals of Natural Gas Processing, Second Edition*; CRC Press, 2011.
 - (15) Toreja, J., Chan, N., VanNostrand, B., Dickinson, J.P., 2011. Rotary-valve, fast-cycle pressure swing adsorption technology allows West Coast platofrm to meet tight California specifications and recover stranded gas. In: Laurance Reid Gas Conditioning Con.
 - (16) CO2 Sponge, 2011. 19 July 2011, /<http://www.iacx.com/co2sponge/S>.
 - (17) Wills, J., Mitariten, M. 2009. Production of Pipeline Quality Natural Gas With the Molecular

- GateTM CO₂ Removal Process. Tidelands Oil Production Company & Guild Associates.
- (18) UOP Overview of Gas Processing Technologies and Applications, 2010. 22 June 2011, <<http://www.uop.com/wp-content/uploads/2011/02/UOP-Overview-of-Gas-Processing-Technologies-and-Applications-tech-presentation.pdf>>.
- (19) Cavenati, S.; Grande, C. a; Rodrigues, a E. *J. Chem. Eng. Data* **2004**, *49*, 1095.
- (20) Simmons, J. M.; Wu, H.; Zhou, W.; Yildirim, T. *Energy Environ. Sci.* **2011**, *4*, 2177.
- (21) Ho, M. T.; Allinson, G. W.; Wiley, D. E. *Ind. Eng. Chem. Res.* **2008**, *47*, 4883.
- (22) Wang, J.; Huang, L.; Yang, R.; Zhang, Z.; Wu, J.; Gao, Y.; Wang, Q.; O'Hare, D.; Zhong, Z. *Energy Environ. Sci.* **2014**, *7* (11), 3478.
- (23) Li, J.; Sculley, J.; Zhou, H. **2012**, 869.
- (24) Li, H.; Eddaoudi, M.; O'Keeffe, M.; Yaghi, O. M. *Nature* **1999**, *402* (November), 276.
- (25) Eddaoudi, M.; Kim, J.; Rosi, N.; Vodak, D.; Wachter, J.; O'Keeffe, M.; Yaghi, O. M. *Science* **2002**, *295* (5554), 469.
- (26) Bruinsma, R. F.; Gennes, P. G. De; Freund, J. B.; Levine, D. *Nature* **2004**, *427* (February), 523.
- (27) 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* (November), 944.
- (28) Li, J.-R.; Kuppler, R. J.; Zhou, H.-C. *Chem. Soc. Rev.* **2009**, *38* (5), 1477.
- (29) Horcajada, P.; Gref, R.; Baati, T.; Allan, P. K.; Maurin, G.; Couvreur, P. *Chem. Rev.* **2012**, *112*, 1232.
- (30) Lee, J.; Farha, O. K.; Roberts, J.; Scheidt, K. A.; Nguyen, S. T.; Hupp, J. T. *Chem Soc Rev* **2009**, *38* (5), 1450.
- (31) Corma, A.; Garcíá, H.; Llabrés i Xamena, F. X. *Chem. Rev.* **2010**, *110* (8), 4606.
- (32) Ma, L.; Abney, C.; Lin, W. *Chem. Soc. Rev.* **2009**, *38* (5), 1248.
- (33) Ma, S.; Zhou, H.-C. *J. Am. Chem. Soc.* **2006**, *128* (36), 11734.
- (34) Kitagawa, S.; Kitaura, R.; Noro, S. *Angew. Chem. Int. Ed. Engl.* **2004**, *43* (18), 2334.
- (35) Park, K. S.; Ni, Z.; Cote, A. P.; Choi, J. Y.; Huang, R.; Uribe-Romo, F. J.; Chae, H. K.; O'Keeffe, M.; Yaghi, O. M. *Proc. Natl. Acad. Sci.* **2006**, *103* (27), 10186.
- (36) Cychosz, K. a.; Wong-Foy, A. G.; Matzger, A. J. *J. Am. Chem. Soc.* **2008**, *130* (22), 6938.
- (37) Manton, A.; Massüger, L.; Rabu, P.; Palivan, C.; McCusker, L. B.; Taubert, A. *J. Am. Chem. Soc.* **2008**, *130* (8), 2517.
- (38) Shimizu, G. K. H.; Taylor, J. M.; Kim, S. *Science (80-.)*. **2013**, *341* (6144), 354.
- (39) Furukawa, H.; Cordova, K. E.; O'Keeffe, M.; Yaghi, O. M. *Science* **2013**, *341* (6149), 1230444.
- (40) Peng, Y.; Li, Y.; Ban, Y.; Jin, H.; Jiao, W.; Liu, X.; Yang, W. *Science* **2014**, *346* (6215), 1356.
- (41) Rowsell, J. L. C.; Spencer, E. C.; Eckert, J.; Howard, J. a K.; Yaghi, O. M. *Science (80-.)*. **2005**, *309* (5739), 1350.
- (42) Vaidhyanathan, R.; Iremonger, S. S.; Shimizu, G. K. H.; Boyd, P. G.; Alavi, S.; Woo, T. K. **2010**, *330*, 650.
- (43) Deng, H.; Doonan, C. J.; Furukawa, H.; Ferreira, R. B.; Towne, J.; Knobler, C. B.; Wang, B.; Yaghi, O. M. *Science* **2006**, 327.
- (44) Wilmer, C. E.; Leaf, M.; Lee, C. Y.; Farha, O. K.; Hauser, B. G.; Hupp, J. T.; Snurr, R. Q. *Nat.*

- Chem.* **2012**, 4 (2), 83.
- (45) Xiang, S.; He, Y.; Zhang, Z.; Wu, H.; Zhou, W.; Krishna, R.; Chen, B. *Nat. Commun.* **2012**, 3, 954.
 - (46) 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 (7), 633.
 - (47) Yang, Q.; Liu, D.; Zhong, C.; Li, J. R. *Chem. Rev.* **2013**, 113, 8261.
 - (48) Li, J.; Sculley, J.; Zhou, H. *Chem. Rev.* **2012**, 112, 869.
 - (49) Burtch, N. C.; Jasuja, H.; Walton, K. S. *Chem. Rev.* **2014**, 114, 10575.
 - (50) Sumida, K.; Rogow, D. L.; Mason, J. a; McDonald, T. M.; Bloch, E. D.; Herm, Z. R.; Bae, T.-H.; Long, J. R. *Chem. Rev.* **2012**, 112 (2), 724.
 - (51) Cui, Y.; Yue, Y.; Qian, G.; Chen, B. *Chem. Rev.* **2012**, 112 (2), 1126.
 - (52) Suh, M. P.; Park, H. J.; Prasad, T. K.; Lim, D. **2012**, 782.
 - (53) Bétard, A.; Fischer, R. A. *Chem. Rev.* **2012**, 112 (2), 1055.
 - (54) Kreno, L. E.; Leong, K.; Farha, O. K.; Allendorf, M.; Van Richard P., D.; Hupp, J. T. *Chem. Rev. (Washington, DC, United States)* **2012**, 112, 1105.
 - (55) Zhong, R. Q.; Zou, R. Q.; Pandey, D. S.; Kiyobayashi, T.; Xu, Q. *Inorg. Chem. Commun.* **2008**, 11 (9), 951.
 - (56) Fang, Q. R.; Zhu, G. S.; Xue, M.; Zhang, Q. L.; Sun, J. Y.; Guo, X. D.; Qiu, S. L.; Xu, S. T.; Wang, P.; Wang, D. J.; Wei, Y. *Chem. - A Eur. J.* **2006**, 12 (14), 3754.
 - (57) Ortiz, A. U.; Boutin, A.; Fuchs, A. H.; Coudert, F. X. *J. Phys. Chem. Lett.* **2013**, 4 (11), 1861.
 - (58) Ling, Y.; Chen, Z.-X.; Zhai, F.-P.; Zhou, Y.-M.; Weng, L.-H.; Zhao, D.-Y. *Chem. Commun. (Camb).* **2011**, 47 (25), 7197.
 - (59) Rees, L. V. C. *Introduction to Zeolite Science and Practice.*; 1992.
 - (60) Baerlocher, C.; McCusker, L. B. *Database of Zeolite Structures*: <http://www.izastructure.org/databases/> (accessed Jul 10, 2014).
 - (61) James, S. L. *Chem. Soc. Rev.* **2003**, 32 (5), 276.
 - (62) Yaghi, O. M.; O'Keeffe, M.; Ockwig, N. W.; Chae, H. K.; Eddaoudi, M.; Kim, J. *Nature* **2003**, 423 (6941), 705.
 - (63) Wang, F.; Zhu, J.; Zhu, H.; Zhang, Q.; Lin, Z.; Hu, H. *Science (80-.).* **2009**, 336 (May).
 - (64) Mofs, Z.; Jasuja, H.; Zang, J.; Sholl, D. S.; Walton, K. S. **2012**, No. Iv, 2.
 - (65) Cmarik, G. E.; Kim, M.; Cohen, S. M.; Walton, K. S. *Langmuir* **2012**, 28, 15606.
 - (66) Lu, W.; Wei, Z.; Gu, Z.-Y.; Liu, T.-F.; Park, J.; Park, J.; Tian, J.; Zhang, M.; Zhang, Q.; Gentle III, T.; Bosch, M.; Zhou, H.-C. *Chem. Soc. Rev.* **2014**, 43 (16), 5561.
 - (67) Meyer, L. V; Schönfeld, F.; Müller-Buschbaum, K. *Chem. Commun. (Camb).* **2014**, 50 (60), 8093.
 - (68) Wang, Q.; Bai, J.; Lu, Z.; Pan, Y.; You, X. *Chem. Commun.* **2016**, 52 (3), 443.
 - (69) Jiao, Y.; Morelock, C. R.; Burtch, N. C.; Mounfield, W. P.; Hungerford, J. T.; Walton, K. S. *Ind. Eng. Chem. Res.* **2015**, 54 (49), 12408.
 - (70) Ho, M. T.; Allinson, G. W.; Wiley, D. E. **2008**, 1562.
 - (71) Britt, D.; Furukawa, H.; Wang, B.; Glover, T. G.; Yaghi, O. M. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, 106 (49), 20637.

- (72) Leus, K.; Couck, S.; Vandichel, M.; Vanhaelewyn, G.; Liu, Y.-Y.; Marin, G. B.; Van Driessche, I.; Depla, D.; Waroquier, M.; Van Speybroeck, V.; Denayer, J. F. M.; Van der Voort, P. *Phys. Chem. Chem. Phys.* **2012**, *14*, 15562.
- (73) Bae, Y.-S.; Farha, O. K.; Spokoyny, A. M.; Mirkin, C. a; Hupp, J. T.; Snurr, R. Q. *Chem. Commun. (Camb)*. **2008**, *0*, 4135.
- (74) Bae, Y.; Mulfort, K. L.; Frost, H.; Ryan, P.; Punnathanam, S.; Broadbelt, L. J.; Hupp, J. T.; Snurr, R. Q. *Langmuir* **2008**, *24* (18), 8592.
- (75) Hamon, L.; Jolimai, E.; Pirngruber, G. D. *Ind. Eng. Chem. Res.* **2010**, *49*, 7497.
- (76) Llewellyn, P. L.; Bourrelly, S.; Serre, C.; Vimont, A.; Daturi, M.; Hamon, L.; De Weireld, G.; Chang, J.-S.; Hong, D.-Y.; Kyu Hwang, Y.; Hwa Jung, S.; Férey, G. *Langmuir* **2008**, *24* (14), 7245.
- (77) Kuhn, J.; Castillo Sanchez, J. M.; Gascon, J.; Calero, S.; Dubbeldam, D.; Vlugt, T. J. H.; Kapteijn, F.; Gross, J. J. *Phys. Chem. C (Energy Convers. Storage, Opt. Electron. Devices, Interfaces, Nanomater. Hard Matter) IF - 4.524* **2010**, *114* (Figure 1), 6877.
- (78) Zornoza, B.; Martinez-Joaristi, A.; Serra-Crespo, P.; Tellez, C.; Coronas, J.; Gascon, J.; Kapteijn, F. *Chem. Commun. (Camb)*. **2011**, *47* (33), 9522.
- (79) Xiang, Z.; Peng, X.; Cheng, X.; Li, X.; Cao, D. *J. Phys. Chem. C* **2011**, *115* (40), 19864.
- (80) Zhao, Z.; Li, Z.; Lin, Y. S. *Ind. Eng. Chem. Res.* **2009**, *48* (22), 10015.
- (81) Allen, F. H. *Acta Crystallogr. Sect. B Struct. Sci.* **2002**, *58* (3 PART 1), 380.
- (82) Chung, Y. G.; Camp, J.; Haranczyk, M.; Sikora, B. J.; Bury, W.; Krungleviciute, V.; Yildirim, T.; Farha, O. K.; Sholl, D. S.; Snurr, R. Q. *Chem. Mater.* **2014**, *26*(21), 6185.
- (83) Boyd, P. G.; Woo, T. K. *CrystEngComm* **2016**.
- (84) Düren, T.; Bae, Y.-S.; Snurr, R. Q. *Chem. Soc. Rev.* **2009**, *38* (5), 1237.
- (85) Yang, Q.; Wiersum, A. D.; Llewellyn, P. L.; Guillerm, V.; Serre, C.; Maurin, G. *Chem. Commun. (Camb)*. **2011**, *47*, 9603.
- (86) Biswas, S.; Zhang, J.; Li, Z.; Liu, Y.-Y.; Grzywa, M.; Sun, L.; Volkmer, D.; Van Der Voort, P. *Dalton Trans.* **2013**, *42*, 4730.
- (87) Torrisi, A.; Bell, R. G.; Mellot-Draznieks, C. *Cryst. Growth Des.* **2010**, *10*, 2839.
- (88) Xiang, Z.; Leng, S.; Cao, D. *J. Phys. Chem. C* **2012**, *116*, 10573.
- (89) Dawson, R.; Adams, D. J.; Cooper, A. I. *Chem. Sci.* **2011**, *2* (6), 1173.
- (90) Watanabe, T.; Sholl, D. S. *Langmuir* **2012**, *28*, 14114.
- (91) Haldoupis, E.; Nair, S.; Sholl, D. S. *J. Am. Chem. Soc.* **2010**, *132* (21), 7528.
- (92) Goldsmith, J.; Wong-Foy, A. G.; Cafarella, M. J.; Siegel, D. J. *Chem. Mater.* **2013**, *25*, 3373.
- (93) Martin, R. L.; Simon, C. M.; Smit, B.; Haranczyk, M. *J. Am. Chem. Soc.* **2014**, *136*, 5006.
- (94) Wilmer, C. E.; Farha, O. K.; Bae, Y.-S.; Hupp, J. T.; Snurr, R. Q. *Energy Environ. Sci.* **2012**, *5* (12), 9849.
- (95) Campañá, C.; Mussard, B.; Woo, T. K. *J. Chem. Theory Comput.* **2009**, *5*, 2866.
- (96) Fernandez, M.; Boyd, P. G.; Da, T. D.; Aghaji, M. Z.; Woo, T. K. *J. Phys. Chem. Lett.* **2014**.
- (97) Iremonger, S. S.; Liang, J.; Vaidhyanathan, R.; Martens, I.; Shimizu, G. K. H.; Daff Thomas, D.; Aghaji, M. Z.; Yeganegi, S.; Woo, T. K. *J. Am. Chem. Soc.* **2011**, *133*, 20048.

- (98) Aghaji, M. Z.; Fernandez, M.; Boyd, P. G.; Daff, T. D.; Woo, T. K. *Eur. J. Inorg. Chem.* **2016**, 27, 4505-4511

2 Background Information

2.1 Criteria for MOF performance

Considering the PSA process for the separation of CO₂ from natural gas, the performance of the MOF is evaluated on the basis of three primary factors that affect the energetic cost of the process: (1) the amount of CO₂ uptake at adsorption and desorption conditions, (2) the selectivity of CO₂ over CH₄, and (3) CO₂ isosteric heat of adsorption. These three factors are discussed below along with some other factors, including MOF stability and kinetics, which are essential for MOFs when being implemented in a PSA unit.

2.1.1 CO₂ working capacity

The gas adsorption properties are typically examined by measuring a gas adsorption isotherm, as given in Figure 2.1. In this experiment, a porous material is exposed to a given partial pressure of gas at a fixed temperature and the equilibrium amount of the gas that is adsorbed is measured. Figure 2.1 represents an adsorption isotherm where the amount of CO₂ adsorbed in the material is plotted on the y-axis as a function of the partial pressure of CO₂ that is plotted on the x-axis. The closed circle corresponds to the partial pressure of CO₂ at which the CO₂ is adsorbed in the material, which is typically 1 bar. In contrast, the open circle represents the desorption pressure at which CO₂ is desorbed from the CO₂-saturated material. The difference in the amount of CO₂ adsorbed on the material at the adsorption pressure and the amount of CO₂ remaining at the desorption pressure is defined as working capacity, which is often quantified in the unit of CO₂ mmol per gram of the material.

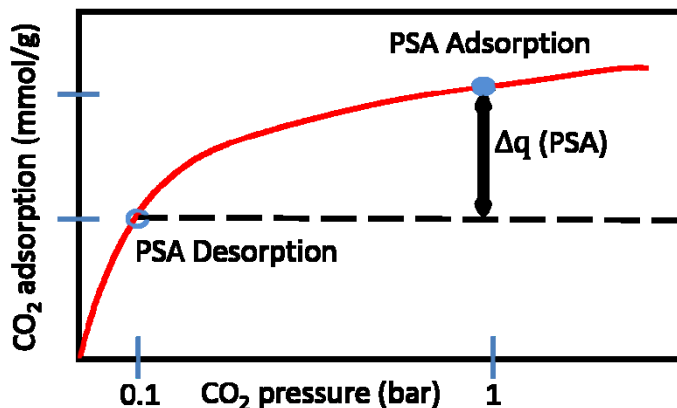


Figure 2.1: Adsorption isotherm of CO₂ in a MOF. The close circle represents the uptake of CO₂ at the CO₂ partial pressure relevant to natural gas processing, while the open one represents the uptake under desorption condition. Δq represent the CO₂ working capacity.

Working capacity dictates the amount of MOF required in the adsorbent bed in a PSA process. The higher the working capacity the less amount of material required for the PSA process, thus resulting in a decrease of capital and operation cost of CO₂ separation from natural gas.

2.1.2 CO₂/CH₄ selectivity

The CO₂/CH₄ selectivity is typically obtained using the single-component adsorption isotherms of CO₂ and CH₄, as it is challenging experimentally to measure the uptakes of individual gases from a mixture. Here, the selectivity is defined as a molar ratio of adsorbed amount of CO₂ (X_{CO_2}) over adsorbed amount of CH₄ (X_{CH_4}) which is normalized by the ratio of their partial pressures (P_{CO_2}/P_{CH_4}).

$$S = \frac{X_{CO_2} \times P_{CH_4}}{X_{CH_4} \times P_{CO_2}} \quad 2.1$$

However, the calculation that originates from two single pure isotherms instead of a binary mixture does not take into account the competition of CO₂ and CH₄ for specific adsorption sites that typically occurs in a binary mixture. Therefore, calculation in this manner

leads to a conservative estimation of selectivity, particularly in MOFs with polar functional groups or metal cation sites where there is a higher affinity for CO₂ over CH₄ because of the strong quadrupole moment of CO₂ ($13.4 \times 10^{-40} \text{ C}\cdot\text{m}^2$). In order to compensate, Ideal Adsorption Solution Theory (IAST)¹ has been used to predict the expected selectivity in a binary mixture using single gas isotherms. In this method, using a mathematical fitting of single-component isotherms, the mole fraction of gases in adsorbed phase are calculated by solving the following expression.

$$\int_0^{P \cdot X_i/Y_i} \frac{\text{Isotherm fit for component } i(p)}{P} dp = \int_0^{P \cdot X_j/Y_j} \frac{\text{Isotherm fit for component } j(p)}{P} dp$$

where X and Y represent the mole fraction of component i in adsorbed and bulk gas phases, respectively, and P is the total pressure. The assumptions behind the model are that the gases behave as ideal gases in the mixture and the surface is homogeneous. Despite these assumptions, it has been show that IAST predicts the binary selectivity very well in a wide range of MOFs.^{2,3}

As shown by equation 1.1, the amount of CO₂ adsorbed in MOFs increases by increasing the selectivity, which improves the purity of CH₄ in recovered gas when separating these two gases. Considering the sorbents in the current industrial PSA process that satisfiy the North American-defined pipeline specification (CO₂ <%2)⁴, an eligible MOF must have a selectivity of at least 15.⁵

2.1.3 Heat of adsorption

The regeneration cost of a CO₂-capture material ultimately depends on the binding strength of CO₂ with the framework, which is commonly expressed as the isosteric heat of

adsorption, Q_{st} . The Q_{st} value gives an estimation of the quantity of energy required to remove CO_2 during generation. More specifically, a material with too high of a Q_{st} for CO_2 adsorption, despite having strong affinity for CO_2 , suffers from the large quantity of energy required to break a strong bond between CO_2 and the framework, which increases regeneration costs. Meanwhile, despite being more readily regeneratable, a material with too low of a Q_{st} is likely to have a lower selectivity, which decreases the purity of CH_4 in pipeline gas. Thus, an ideal material must have a Q_{st} that provides a balance between the affinity for CO_2 and the required energy for removing CO_2 at the regeneration stage. The calculation of Q_{st} is often performed using two or more isotherms at different temperatures, usually within 10 K of each other. These isotherms are first fitted to a virial equation (i.e. as an expansion in terms of CO_2 loading (N)) to obtain an expression for pressure (p) in terms of CO_2 loading. Then, the Q_{st} value at each level of CO_2 loading is computed using the Clausius-Clapeyron equation,⁶ which is expressed as

$$(\ln p)_N = -(Q_{st}/R)(1/T) + C \quad 2.2$$

where T is the temperature, R is gases constant, and C is a constant. The Q_{st} at the limit of zero loading, so-called the zero-coverage isosteric heat of adsorption Q_{st0} , is an interesting parameter that evaluates the strength of CO_2 binding at the lowest CO_2 pressures. The value of zero-coverage isosteric heat of adsorption indicating the highest interaction energy sites within a material can be attributed to the certain chemical features of the material, such as functional groups and open metal sites.

2.1.4 MOF stability

As discussed in the previous chapter, MOFs are still limited in research laboratories where most studies focus on the synthesis of new MOFs and the optimization of the general

adsorption properties including selectivity, working capacity, and heat of adsorption. However, prior to implementing these materials in industrial applications, other factors such as stability, process design, and environmental impacts associated with these materials should also be addressed. Here, we focus on the stability of MOFs which directly indicates the feasibility of their practical applications in industrial settings. Specifically, it can address how long a MOF can be used in adsorption/desorption cycles of PSA processes before it would need to be replaced. In terms of a definition, a MOF's stability can be defined into structural stability and chemical stability.

Structural stability refers to the capability of a MOF to maintain its structure after the evacuation of its pores. This stability is critical when considering a material for frequent adsorption/desorption cycles in which gas is being adsorbed and evacuated from its pores. Generally speaking, a MOF with large pores exhibiting void fraction of 0.8 or higher usually suffers from low stability.⁷⁻⁹ This is due to fact that most of these MOFs contain solvent molecules in their pores, which support their crystallinity, and they lose their porosity after the solvent has been evacuated. This makes them ill-suited sorbents for the PSA process.

One further subclass of MOFs that also suffers from low framework stability is MOFs that show “breathing phenomena” during the adsorption process.¹⁰ Specifically, this involves enlarging and shrinking of the pore diameter of the MOF (i.e., change in the volume of the MOF) upon insertion or removal of guest molecules while bonds between building components are retained. Although it has been shown that this effect makes these MOFs selectively adsorb CO₂ over other gases,¹¹ they are not ideal to be used in PSA systems, as these MOFs are packed into a bed column with a fixed volume while their flexibility will cause these materials to

fracture.

The chemical stability is another important factor that must be addressed when considering MOFs for the PSA process of removing CO₂ from natural gas. Specifically, it shows the degree of vulnerability of the MOF when exposed to reactive chemicals which are often as simple as water. Although some MOFs show good stability toward acid gases such as H₂S,¹² the majority of them suffer from being degradable under humidity,¹³ which limits their application in industrial settings. The degradation of a MOF when exposed to humid conditions can be attributed to the fragile organic ligand – metal coordination bond that is easily broken by water. Providing strong bonds between organic ligands and metal nodes can increase the stability of MOFs when exposed to water. One strategy widely used by chemists to provide strong metal-ligand bonds is the use of an organic ligand on which the coordinating atom has high P_{Ka} . This was first noted by Long et al. who systemized a series of highly stable MOFs containing Co-, Zn, Ni, and Cu as metal nodes and the pyrazolate (P_{Ka} 19.8) as an organic ligand.^{14,15}

2.1.5 Kinetics of gas adsorption in MOF

Until now, many of the factors that have been addressed, such as working capacity and selectivity, are primarily based on equilibrium between adsorbed gas and bulk gas. However, in practice the gas flows through a fixed bed absorber where adsorption kinetics can influence the performance of the material. In this context, certain gases can enter a MOF's pores and become adsorbed faster than other gases based on differences in diffusion coefficients. The diffusion coefficient of the gas in a certain MOF, is typically quantified in cm²/s, and it indicates how easily the gas can travel through the MOF. On the other hand, it shows how quickly the MOF is able to adsorb and desorb the gas. Indeed, if the MOF exhibit a relatively high gas diffusion

coefficient, it can be rapidly loaded with gas and, subsequently, unloaded. There are several factors that can control the diffusion properties of MOFs, among which the pore size of the MOFs plays a significant role. It has been shown that the diffusion properties of a MOF are directly proportional to the ratio of the kinetic diameter of the gas to the pore size of the MOF.^{16–18} MOFs with pore sizes comparable to the kinetic diameter of gases show poor diffusion. In the case of PSA technologies, one should use MOFs with relatively high gas diffusion coefficients to meet desired adsorption-desorption cycle times in the PSA platform. In comparing with zeolites and activated carbons, some MOFs have shown larger diffusion coefficients for CO₂. For example, MOF-5 has a diffusion coefficient of $4.5 \times 10^{-2} \text{ cm}^2/\text{s}$ for CO₂,¹⁶ which is two orders of magnitude higher than the value observed in commercialized Zeolite 5A ($10^{-4} \text{ cm}^2/\text{s}$).¹⁹ This can be attributed to the larger pore volume in MOF-5, which facilitates diffusion of gases into MOFs by reducing congestion effect on gas molecules during the adsorption process. However, one should take care when choosing a large-pore MOF for PSA-based separation application as it has been reported when the pore size of MOF becomes too large, the MOF exhibits poor separation properties despite having promising diffusion properties.²⁰

2.2 Computational tools in MOF study

With the tremendous growth of computational resources, the use of computational tools that predict adsorption properties of MOFs have been widespread in the MOF community. These tools can be categorised into two groups. One group includes methods that predict the adsorption properties based on energy interactions between guest particles (gas) and the host (MOF). Of these methods, the most extensively used are grand canonical Monte Carlo (GCMC)

simulations, which is a molecular simulation based method. Another group includes methods that predict the adsorption properties using only structural information of MOFs, which refers to quantitative structure-property relationship (QSPR) modeling. What follows are the review of the underlying theory of GCMC simulations and structure-property relationship (QSPR) modeling, and the details of the major components that are involved in the GCMC simulations and QSPR modeling in this thesis. The discussion starts with the concept of ensemble in statistical mechanics, followed by details of the GCMC simulations and force field parameters used in the GCMC simulations, and the discussion ends with structure-property relationship (QSPR) modeling.

2.2.1 Grand Canonical Ensemble

The basis of Grand Canonical Monte Carlo simulation relies on sampling a statistical mechanical ensemble. Therefore, before going into details of the GCMC simulation, we briefly discuss on the ensemble concept from the point of view of statistical mechanics.

In the context of statistical mechanics, macroscopic properties of a system, such as pressure, temperature, and volume, are representations of averages over all possible microstates. To calculate such properties, it is convenient to define an ensemble. An ensemble is an imaginary collection of a large number of replicas of a system, which are all in contact with each other. These replicas are in different microstates but they have the same macroscopic attributes (i.e., thermodynamic variables) as the real system. For example, if a real system is defined with a fixed volume (V), a constant temperature (T), and fixed number of particles (N), all replicas will have the same N , V , and T . The ensemble average of any macroscopic property (M) can be determined by,

$$\langle M \rangle = \sum_i M_i P_i \quad 2.3$$

where M_i is the value of property M in a microstate i , P_i represents the probability of finding the system in a microstate in the ensemble, and the angled brackets denote an ensemble average. The form of P_i is determined by the type of the ensemble being sampled. For example, the P_i for Canonical ensemble (N, V, T) is given by:

$$P_i = \frac{e^{-\frac{E_i(N,V)}{KT}}}{\sum_i e^{-\frac{E_i(N,V)}{KT}}} \quad 2.4$$

where E_i is the energy of a system in a microstate i , T is the temperature, and K is the Boltzmann constant. As an example of Canonical ensemble, consider the diffusion of a CO_2 molecule through the pores of a MOF at constant temperature and volume, where the move of the CO_2 molecule through the pores is governed upon its velocity and its interaction with the MOF atoms. In this case, each snapshot of the CO_2 movement inside the pores represents one classical mechanical microstate (configuration) of the Canonical ensemble, and subsequently the movement of the CO_2 molecule inside the pore can be considered as sampling a number of microstates. If one observes the CO_2 molecule for sufficiently long times (i.e., samples a very large number of possible microsite), the average of all velocities that are collected at each snapshot would yield the ensemble average kinetic energy of the CO_2 molecule within the MOF.

Ensembles can be categorized into the two groups. There are different types of ensembles, including the microcanonical, canonical, isothermal-isobaric, and grand canonical ensembles. The first three ensembles describe closed systems where the numbers of particles are kept constant, while the grand canonical ensemble describes an open system where the number of particles in the system can change. The grand canonical ensemble, in which system

is defined with a fixed volume (V), a constant temperature (T), and a fixed chemical potential (μ), is often chosen for the study of gas adsorption in MOF materials, which is typically treated with Grand Canonical Monte Carlo Simulations (GCMC).

2.2.2 Grand Canonical Monte Carlo Simulation

Grand Canonical Monte Carlo (GCMC) Simulations have been widely used for study of gas adsorptions in porous materials. Studied in this manner, GCMC simulations enable us to determine the quantity of gas that will adsorb within MOFs at a given external pressure and temperature of the system. As the name of GCMC implies, this quantity is determined by Monte Carlo sampling of a Grand Canonical ensemble, where the gas chemical potential, the temperature, and the volume of a system are kept fixed, while the number of particles can change. Here, the chemical potential shows how the potential energy of the system changes as a function of number of particles in a given system. The choice of GC ensemble arises from the fact that in experimental setup adsorbed gas is in equilibrium with gas in the reservoir at a given pressure, and at equilibrium the chemical potential of adsorbed gas is equal to that of the gas in the reservoir, which further determines the equilibrium concentration of the adsorbed gas inside the porous material. For gas in the reservoir, the chemical potential is expressed as the following equation²¹:

$$\mu_{gas}(T,P) = \mu_{gas}^0 + RT \cdot \ln \frac{f}{f^0} \quad 2.5$$

where, P is the pressure of the gas in reservoir, μ_{gas}^0 is the chemical potential in a given standard state, T is temperature of a system, R is the gas constant, f^0 is the fugacity of the gas in the standard state, and f is the fugacity of the gas in the reservoir. The fugacity is obtained from the pressure using an equation of state such as Peng-Robinson equation, however, for low

pressures in the vicinity of 1 bar, one can assume the gas as an ideal gas and replace fugacity with pressure.

Monte Carlo (MC), in general, refers to any algorithm that uses random numbers. This method has been widely employed in many fields to solve diverse kinds of problems, such as telecommunication network optimization, finance, and computer chip design. In molecular simulations, 'Monte Carlo' refers almost exclusively to Metropolis Monte Carlo, where it preferentially samples low energy configurations of a system to prevent the sampling of configurations which likely have no significant contribution to the average properties. Indeed, Metropolis method generates configurations based on the Boltzmann distribution and then assigns them an equal weight in an ensemble average:

$$\langle A \rangle = \frac{1}{N_{sample}} \sum_i^{N_{sample}} A \quad 2.6$$

Such a sampling is remarkably more efficient than the pure MC random sampling where the configurations are generated equally and then are weighted by Boltzmann factor. The outline of Metropolis MC Algorithm for a system of interacting CO₂ molecules contained in a MOF is provided below.

1. One CO₂ molecule already placed in the framework is randomly picked.
2. The CO₂ molecule coordinates are randomly perturbed via a MC trial move. The typical MC trial moves are translation and rotation.
3. The potential energy of the new configuration, U_{new} , is evaluated.
4. If the energy, U_{new} , is lower than the energy of the old configuration, the new configuration is accepted. If it is higher than the energy of the old configuration,

then the new configuration is accepted based on an acceptance probability (ACC) which is given by the Boltzmann factor:

$$Acc = \min[1, e^{-\Delta U/kT}] \quad 2.7$$

where ΔU (i.e., $U_{\text{new}} - U_{\text{old}}$) is the change in energy between the new and old configuration and kT is the thermal energy of the system. To accept/reject this configuration, the value of this factor is compared to a random number generated between 0 and 1. If the random number is greater than this factor then new configuration is accepted.

5. If the new configuration is accepted, then it is added to the ensemble average.
6. If a new configuration is rejected, then the old configuration is re-added to the ensemble average.
7. This process is repeated by randomly perturbing the added configuration.

When Metropolis Monte Carlo sampling is coupled with the grand canonical ensemble, two additional trial moves insertion and deletion are added to the sampling to mimic the entrance or removal of a gas molecule from the MOF framework. Figure 2.2 represents some of the possible trial moves. The insertion and deletions moves are accepted or rejected according to the following acceptance probabilities:

Deletion trial move;

$$Acc_{N-1} = \min \left[1, \frac{\Lambda^3 N}{V} \exp^{-\mu/kT} \exp \left(\frac{-(U_{N-1} - U_N)}{k_B T} \right) \right] \quad 2.8$$

Insertion trial move;

$$Acc_{N+1} = \min \left[1, \frac{V}{\Lambda^3 (N+1)} \exp^{\mu/kT} \exp \left(\frac{-(U_{N+1} - U_N)}{k_B T} \right) \right] \quad 2.9$$

where Λ is the thermal de Broglie wave length, V is the volume of the system, P is the partial pressure of the gas in the reservoir, T is the temperature of the system, K is the Boltzmann constant, N is the number of molecules in preceding configuration, U_{N+1} is the potential energy of a configuration after the insertion trial move, U_{N-1} is the potential energy of a configuration after the deletion trail move, and U_N is the potential energy of the preceding configuration.

By substituting chemical potential (μ) with equation 2.5, following probabilities based on the pressure of the gas are obtained:

Deletion trial move;

$$Acc_{N-1} = \min \left[1, \frac{kTN}{VP} \exp \left(\frac{-(U_{N-1} - U_N)}{k_B T} \right) \right] \quad 2.10$$

Insertion trial move;

$$Acc_{N+1} = \min \left[1, \frac{VP}{kT(N+1)} \exp \left(\frac{-(U_{N+1} - U_N)}{k_B T} \right) \right] \quad 2.11$$

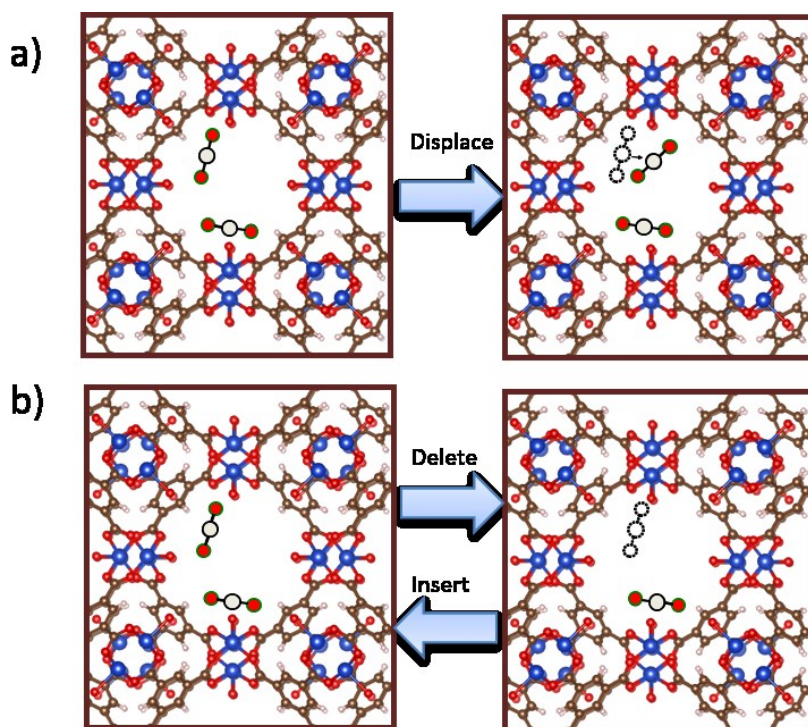


Figure 2.2: The basic moves in a GCMC simulation: a) A CO_2 guest molecule is displaced by a random translation and rotation. b) A molecule of CO_2 is deleted (to the right) and inserted (to the left).

As these two probabilities show, the acceptance of a new configuration for the deletion/insertion move is based on both chemical potential and potential energy in the GCMC scheme which is different from translation/rotation probability, where acceptance is based only on potential energy (equation 2.7). Indeed, these two acceptance probabilities allow the number of molecules to fluctuate in a system based on the Boltzmann distribution, where the more probable states (i.e., configurations) have more significant contributions to the average properties.

To determine the number of adsorbed gas molecules in a MOF at a given T and P , a GCMC simulation is performed at two stages: (1) equilibration and (2) production. In the equilibration stage, the MOF undergoes usually 10^7 GCMC trial moves (GCMC steps) to reach equilibrium. Specifically, at this stage the gas molecules are favourably inserted into the pores

of the MOF rather than deleted from the pores. This is attributed to the fact that the chemical potential of the molecules of the adsorbed phase must be equal to that of the gas phase at equilibrium. Once equilibrium is achieved, the system undergoes the production stage, where 10^7 additional GCMC steps are taken to calculate the average number of molecules that will adsorb in the MOF. Such simulation requires an efficient method to calculate the energy, as the energy of a system must be calculated at each GCMC step. For example, if it takes only a half second to calculate the energy at one step, then 10^7 energy calculations would require almost 4 months. For this reason, the energy is calculated using molecular mechanics force field rather than expensive electronic structure based methods such as density functional theory (DFT).

For the sake of computational efficiency and simplicity, the majority of GCMC simulations of gas adsorption in MOFs use three approximations: (1) the atom positions of the MOF framework are kept fixed in their crystallographic positions, (2) gas molecules behave as rigid molecules, and (3) the energy of a system, which refers to adsorbate-adsorbate interaction and adsorbate-adsorbent interaction, is described by short-range steric and dispersion interactions, and long-range Coulomb interactions. These interactions are assumed to be pairwise and are calculated using a set of potentials, also known as force fields, which are discussed in section 2.2.2.2.

2.2.2.1 Periodic boundary conditions

Because MOF is crystalline solid, the cluster model approach commonly used in computer simulations of discrete molecular systems cannot be used in the GCMC simulation of MOFs. For this reason, the GCMC simulation of a MOF is performed on a unit cell (or a set of multiple adjacent unit cells, called a 'supercell') which is a translationally invariant set of

representative atoms of its infinite crystal. However, such representation increases surface effects, which would strongly affect the overall properties of the system. Having large surface effects in the unit cell-based representation is attributed to the fact that the surface-to-bulk ratio of molecules in a finite system is larger than that of an infinite system. To remove the surface effects and simultaneously mimic the infinite nature of a MOF, periodic boundary conditions are imposed on the system such that an infinite number of image cells are placed around the main simulation cell. Here, in each image cell, molecules have relative positions and velocities identical to those in the main simulation cell. Treated in this manner, it seems that the numbers of interactions to be considered becomes infinite. However, the numbers of interactions that need to be calculated for the short-range interactions will be reduced by imposing a *cut-off* radius on a van der Waals potential (see details in section 2.2.2.3). Indeed, short-range interactions are calculated only for pairs of atoms whose internuclear distance falls within the *cut-off radius*. In addition to a *cut-off* radius, the *minimum image convention* is also implemented to treat the short-range interactions. According to the *minimum image convention*, molecules only interact with the closest image of other molecules, and consequently only these interactions need to be calculated. To follow the minimum image convention, the cut-off radius must be $< L_{\min}/2$, where $L_{\min}$ is the smallest dimension of the simulation cell. Therefore, a simulation cell must be constructed such that the length of its smallest dimension is larger than $2 \times r_{\text{cutoff}}$. A typical value for $L_{\min}/2$ is between 12 and 14 Å.

2.2.2.2 Force Field

The core of GCMC simulation is force field that is used to calculate the potential energy of a system at each GCMC step. The term Force Field arises from molecular mechanic methods

in which the potential energy of the system is expressed as a collection of simple mathematical functions and pre-defined parameters. A general force field can be broken down into bonded and non-bonded terms. The non-bonded terms, also known as intermolecular terms, pertain to steric and dispersion interactions as well as electrostatic interactions. The bonded terms, also known as intramolecular terms, describe how the energy changes due to bond stretching, angle bending, and torsional twisting. In modeling of gas adsorption in MOFs, only non-bonded terms are considered for energy calculations, as the MOFs' frameworks are kept fixed and small gas molecules behave as rigid molecules. Therefore, here we focus on non-bonded terms, including van der Waals and electrostatic interactions.

2.2.2.3 Short-range van der Waals Interaction Potential

Most MM force fields combine the dispersion interactions with steric interactions in a single 'van der Waals' potential function. The van der Waals potential most commonly used is the Lennard-Jones or 12-6 potential, which is expressed by:

$$E_{L-J} = \sum_A^N \sum_B^N 4\varepsilon_{AB} \left[\left(\frac{\sigma_{AB}}{r_{AB}} \right)^{12} - \left(\frac{\sigma_{AB}}{r_{AB}} \right)^6 \right] \quad 2.12$$

where ε_{AB} and σ_{AB} are specific parameters for interactions between atom A and B , r_{AB} is the distance between atom A and B , and N is the number of atoms in the simulation cell (restricting A and B to be on different molecules). The $1/r_{AB}^{12}$ branch of the potential represents steric repulsions at short distance and $1/r_{AB}^6$ branch arises from the dispersion attraction between atoms. A van der Waals Interaction, as shown in the equation 2.12, decays rapidly with distance and can be neglected after a certain distance. Such a behavior enables one to treat short-range van der Waals interactions with minimum image convention in systems that utilize periodic

boundary conditions, and to introduce a cut-off radius in van der Waals potential. As mentioned in section 2.2.2.1, only interactions within a cut-off radius need to be calculated. To calculate these interactions, one needs to deal with Lennard-Jones parameters. The L-J parameters that define the well depth, ϵ_{AB} , and the distance of the potential well, σ_{AB} , are usually based on the atom type of the two atoms involved. The atom type is a representation of an atom which contains information such as hybridization state and local environment in addition to atomic number. For example, “CA” atom type that is provided by the AMBER force field relates to sp^2 aromatic carbon in a 6-membered ring with 1 substituent.

The force field parameters are developed for specific classes of materials (such as AMBER for proteins and nucleic acid) or for general-purpose use. Because MOFs are a new generation of porous materials, the number of force field parameters that are specifically designed for them is not great. Therefore, L-J parameters for the framework atoms of MOFs have been taken from several generic force fields that are developed to cover a large portion of the periodic table. Among the generic force fields used, the DREIDING force field²² and the Universal Force Field (UFF)²³ are the most widely used and have reproduced experimental data with reasonable accuracy.^{3,24–26} In contrast to MOFs, the L-J parameters (and atomic charges) for gas molecules are often taken from well-established force fields such as the TraPPE force field²⁷ which are specifically designed for atoms in gas phase. These force fields are typically fit to experimental data such as vapour-liquid equilibrium properties to reproduce some observed properties of gases.

While aforementioned force fields provide L-J parameters for an interaction between two same types of atoms, the MOF-adsorbate interactions invariably involve vdW interactions

between different types of atoms. To calculate these interactions, the Lorentz-Berthelot mixing rules have been applied to determine the parameters for cross interaction:

$$\sigma_{AB} = \frac{1}{2}(\sigma_{AA} + \sigma_{BB}) \quad 2.13$$

$$\varepsilon_{AB} = \sqrt{\varepsilon_{AA}\varepsilon_{BB}} \quad 2.14$$

2.2.2.4 Long-range Electrostatic Interaction Potential

In contrast to the van der Waals Interactions, Coulombic interactions between atomic point charges decay proportional to $1/r$. Such a slow decay prevents a direct implementation of *cut-offs* or minimum image convention as are used in short-range van der Waals interactions. Electrostatic forces are long-ranged, and consequently the electrostatic forces from many image cells contribute to the force on a molecule. Coulombic potential energy of all molecules in a system with periodic boundary conditions is given by:

$$E_{\text{coul}} = \frac{1}{2} \frac{1}{4\pi\varepsilon_0} \sum_{\mathbf{n}} \sum_A^N \sum_B^N \frac{q_A q_B}{r_{AB} + \mathbf{nL}} \quad 2.15$$

where the interaction between partial charges q_A and q_B summed over all $\mathbf{n}$ replicas of a cell with side length L . Owing to the fact that the long-range convergence of the electrostatic potential is very slow, many $\mathbf{n}$ values in the summation are required, which dramatically increases the computation time. Moreover, equation 2.15 is infinite at $r \rightarrow 0$, which makes a Fourier transform technique, where a periodic function is expressed as a series of simple sine and cosine functions, inappropriate for this case. For these reasons, long-range electrostatic interactions in periodic systems are treated with the Ewald summation technique, where a long-range electrostatic interaction is divided into two parts: a short-range contribution, and a long-range contribution which is not infinite at $r \rightarrow 0$. Here, the error function, $\text{erf}(\alpha r)$, and

the complementary error function, $\text{erfc}(\alpha r)$, are chosen for this decomposition.

$$\text{erf}(r) = \frac{2}{\sqrt{\pi}} \int_0^r e^{-t^2} dt \quad 2.16$$

$$\text{erfc}(r) = \frac{2}{\sqrt{\pi}} \int_r^\infty e^{-t^2} dt \quad 2.17$$

replacing $\frac{1}{r}$ by $\frac{\text{erf}(r)}{r} + \frac{\text{erfc}(r)}{r}$ gives,

$$E_{\text{coul}} = \frac{1}{2} \frac{1}{4\pi\epsilon_0} \sum_n \sum_A \sum_B^N q_A q_B \left\{ \frac{\text{erf}(\alpha |r_{AB} + nL|)}{|r_{AB} + nL|} + \frac{\text{erfc}(\alpha |r_{AB} + nL|)}{|r_{AB} + nL|} \right\} \quad 2.18$$

where the first term, which is proportional to $\text{erf}(r)/r$, represents the long-range interactions which unlike the $1/r$ term is finite at the origin (blue curve in Figure 2.3), and the second term, which is proportional to $\text{erfc}(r)/r$, represents the short-range electrostatic interactions which converge faster than $1/r$ (black curve in Figure 2.3). The short-range $\text{erfc}(r)/r$ contributions are treated with the short-range cut-off distance approach as used in the van der Waals potential. The long-range contribution (first term) is calculated with Fourier analysis methods as there is no infinity at $r \rightarrow 0$. The range of the two terms in the expansion can be optimized with the choice of values for the α parameter which represents the rate of decay of the error function. Finally, to calculate the electrostatic interactions using equation 2.18, one needs to determine q_A and q_B according to methods discussed below.

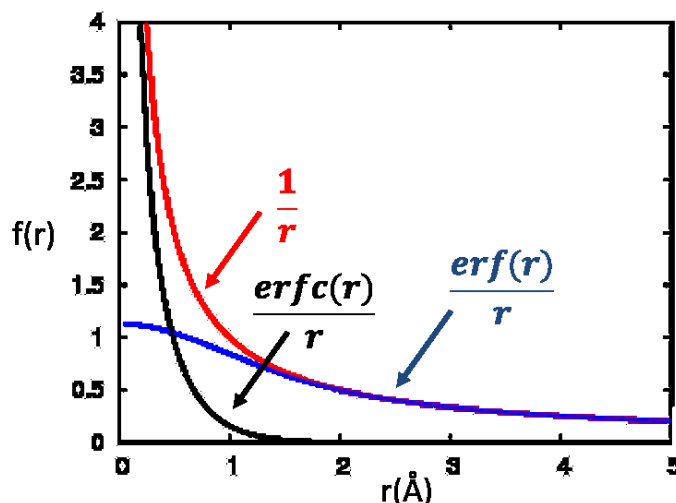


Figure 2.3: A behaviour of functions $\frac{1}{r}$ (red curve), $\frac{\text{erf}(r)}{r}$ (blue curve), and $\frac{\text{erfc}(r)}{r}$ (black curve) against a distance between two atoms, r .

In principle, the electrostatic interaction for adsorbate-MOF and adsorbate-adsorbate interactions would ideally be calculated from the Coulomb interactions between the charge densities (Figure 2.4a). However, this would require a lengthy QM electronic structure calculation of the potential energy surface. Thus, the electrostatic interactions are commonly calculated by replacing the complicated charge density on the interacting species with a fixed partial atomic charge model (Figure 2.4b). The most common charges used are electrostatic potential (ESP) fitted charges, which is extracted from a quantum mechanics calculation. Here a quantum chemical calculation of a molecule is performed and the electrostatic potential around the molecule evaluated as shown schematically in Figure 2.4a. Then, partial atomic charges on the molecule are adjusted such that the electrostatic potential due to these point charges best matches the electrostatic potential resulting from the QC calculation. Of the methods that use this approach, the REPEAT method, which was developed by Woo and coworkers,²⁸ has been widely used in a wide variety of MOFs. In the REPEAT method, the partial atomic charges on the MOFs are adjusted such that the electrostatic potential (ESP) due to

these point charges matches the ESP from the periodic QM calculation. More specifically, a least-squares fit is performed to minimize the relative difference between the QM ESP and the ESP from point charges.

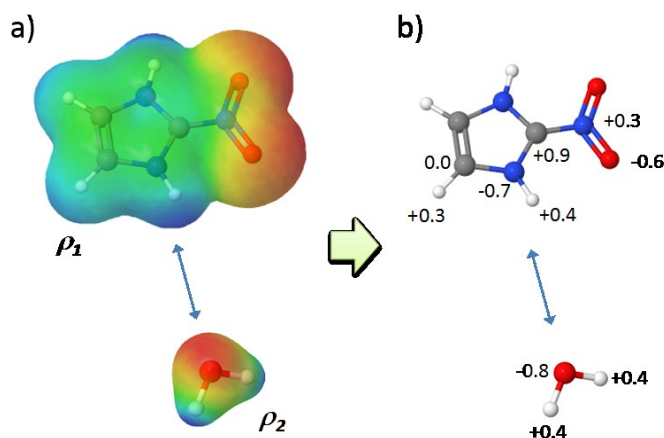


Figure 2.4: a) A representation of the electron density, ρ_1 and ρ_2 , of two interacting molecules. The surface around each molecule shown is an isosurface of the total electron density. The colors reflect the electrostatic potential due to the nuclear charges and electron density of the molecule at that isosurface point. Red regions represent regions of negative electrostatic potential, while blue regions represent regions of positive electrostatic potential. b) Partial atomic charges (in units of electron charge) commonly used to approximate the electrostatic interaction between molecules.

To use the REPEAT method, periodic DFT calculations need to be performed to obtain the ESP necessary for the fitting procedure. Here, the unit cell of a MOF is used in the periodic DFT calculations, where the wave function (or the traditional atom centered Gaussian functions) of the unit cell is optimized in the presence of its periodic images. This approach provides a more realistic model of crystalline systems compared to the molecular DFT methods that use cluster models where a subsection of a MOF is modeled as an isolated system. In the REPEAT method, periodic effects and long range electrostatic interactions are more properly accounted for. In this thesis, the periodic DFT calculations were performed using the Vienna Ab Initio Software Package (VASP)^{29–32} for two purposes: (1) to calculate the ESP needed for the REPEAT atomic

charge calculation, and (2) to optimize the positions of H atoms of MOFs as well as to relax other framework atoms if full optimization is required.

Although the REPEAT method is a robust method to derive ESP charges and obviates the need to construct a cluster model of a MOF as required for cluster-based charge derivation methods, the required DFT calculation usually takes an hour for one material. Such a time-demanding calculation makes this method infeasible in the screening of hundreds of thousands of MOF materials for gas-related applications. Alternatively, charge equilibration (QEq) methods, which compute atomic charges in seconds, have been used in several studies of MOFs. An example of this is the work was done by Snurr et al.³³ who used a QEq method to calculate the atomic partial charges in the screening of 134,000 hypothetical MOFs for CO₂ capture. Although they were able to recognize high-performing MOFs, it was later shown by Woo and coworkers that the QEq charges produced were particularly high for the Cl- and F-functionalized MOFs, which significantly over-estimated the CO₂ uptakes in those cases.³⁴ A brief discussion of the QEq method is provided in following section.

2.2.2.5 Charge Equilibrium methods

The QEq methods are based on an expansion of the total energy of a system in terms of partial atomic charges on the atoms as shown in equation 2.19.

$$\begin{aligned}
 E(q_1, q_2, \dots, q_N) &= \sum_{A=1}^N \left(E_{A0} + q_A \chi_A + \frac{1}{2} q_A^2 J_{AA} \right) \\
 &+ \frac{1}{2} \sum_{A=1}^N \sum_{B=1, B \neq A}^N q_A q_B J_{AB} + \frac{1}{2} \sum_{L \neq 0} \sum_{A=1}^N \sum_{B=1}^N q_A q_B J_{L,AB} \\
 &- \lambda \left(\sum_{k=1}^N q_k - q_T \right)
 \end{aligned}
 \tag{2.19}$$

Specifically, it states that the atomic charges for a system with N atoms can be obtained by minimizing equation 2.19 where the first term accounts for a sum of the charging energy of each atom, the second term represents a sum of interatomic Coulomb interactions ($J_{L,AB}$) between pairs of atoms in the unit cell, while the sum in the third term accounts for all L periodic images, and the last term which includes the Lagrange multiplier (λ) accounts for the constraint where the sum of individual atomic charges is equal to the total system charge q_T . E_{A0} , χ_A , and J_{AA} are the ground state energy, atomic electronegativity, and atomic hardness of atom A , respectively. The minimization of equation 2.19 is performed by taking the first partial derivative of the total energy with respect to all of the charges. Taking this procedure and considering the derivatives to be zero (stationary point), N linear equations are obtained, which can be written as:

$$\mathbf{G} \begin{pmatrix} \mathbf{Q} \\ \lambda \end{pmatrix} = \begin{pmatrix} -\mathbf{\chi} \\ \mathbf{0} \end{pmatrix}, \quad \mathbf{G} = \begin{pmatrix} \mathbf{J} & \mathbf{1} \\ \mathbf{1}^\dagger & \mathbf{0} \end{pmatrix} \quad 2.20$$

where $\mathbf{Q} = (q_1, q_2, \dots)^\dagger$ is the charge column vector, $\mathbf{\chi} = (\chi_1, \chi_2, \dots)^\dagger$ is the electronegativity column vector, and $\mathbf{J}$ is the Coulomb interaction matrix. Owing to the fact that the $\mathbf{J}$ matrix elements between two atoms decay very slowly ($\propto 1/r$), the Ewald summation approach is used to calculate these elements. The above matrix is solved iteratively until the charges have converged.

In this thesis, a “MOF electrostatic potential optimized” QEq (MEPO-QEq) method recently proposed by Kadantsev et al. was used to calculate the atomic charges for 340,000 hypothetical MOFs. In the MEPO-QEq method, the parameters of the QEq methods (i.e., electronegativity and hardness) have been trained to reproduce the DFT-derived electrostatic potential in a large and diverse set of MOFs.³⁴

2.2.3 Quantitative Structure–Property Relationship (QSPR)

Quantitative structure–Property relationship (QSPR) models have been widely considered as a cost-reduced and accelerated tool in the development of new materials in the fields of chemical/biological science and engineering. Nowadays, most material discovery processes are accompanied by QSPR approaches. Their unparalleled success in the recognition of high-performing chemical/biological materials^{35–39} makes them an efficient tool in the rational design of materials for a variety of applications. Basically, QSPR attempts to explore reliable relationships between the chemical structure and target property of studied materials, which can lead to define “rules”. These “rules” can be then used to evaluate new materials’ performance or to define priority features necessary for the design of high-performing materials.

Like other data mining techniques, QSPR requires a training set, upon which the QSPR models are developed, and a test set, upon which the predictive ability of QSPR models is validated. The training set should be structurally diverse enough to cover the range of physical and chemical space relevant to the target property of studied materials. Also, the training set must be excluded from any materials of the test set.

QSPR modeling is carried out in three major steps: (1) selecting appropriate descriptors, (2) applying QSPR modeling methods to create the model between descriptors values and target values, and (3) validating the QSAR model. The details of these steps are discussed below.

2.2.3.1 Descriptors

Descriptors are the “numerical values associated with chemical constitution for

correlation of chemical structure with various physical properties, chemical reactivity or biological activity.”⁴⁰ This includes a wide variety of physical and chemical features which can be broadly classified as constitutional, topological, geometric, and quantum-chemical. Among them, the constitutional descriptors are the simplest, most common descriptors that do not contain information on the electronic structure or the geometry of a material. Examples of constitutional descriptors can be molecular weight, atom count, and the number of aromatic rings.⁴¹ Topological descriptors describe the bonding information and connectivity between atoms in a material.⁴² Geometric descriptors describe the three-dimensional geometry of a material. Examples of structural descriptors include material surface area, material maximum pore diameter, void fraction, and material void volume.⁴³ Quantum-chemical descriptors are calculated using semiempirical, DFT, or ab initio methods. Examples of these descriptors are orbital energy, partial atomic charges, ESP, polarizability, and multipole moments.⁴⁴

The predictive ability and reliability of QSPR models strongly depend on the correct choice of descriptors as well as the number of calculated descriptors used in QSAR model development procedure. The number of calculated descriptors should be large enough to distinguish information between structures. However, if the number of descriptors is too large, particularly much larger than the size of data set (i.e., the number of materials), over-fitting can occur in QSAR model. Over-fitting can result in a number of different equivalent QSAR models which are built from different combinations of descriptors, none of which can reliably make predictions for a new data point. Also, when a very large number of descriptors are used, QSAR models are often complicated and are not easy to interpreted. Furthermore, owing to the fact that usually only a subset of calculated descriptors contains information relevant to a target

property, removing redundant and irrelevant descriptors can improve the performance of QSAR models. Therefore, QSAR modeling is often accompanied by descriptor selection.

Typically, descriptor selection is carried out through two approaches: wrapper methods and filter methods.⁴⁵ In wrapper methods, the space of all possible descriptor subsets is searched using a method which is wrapped around a regression or classification model. More specifically, a subset of various combinations of descriptors is generated by the search method, then is evaluated through training the regression or classification model (see below for definitions) on a subset of the training set, and is then assigned a “score” based on how well the model performs. This is performed on multiple different subsets of descriptors and the set with best “score” is selected. In some wrapper methods, information about subsets and “scores” are passed on other subsets. Examples of the wrapper method include the forward and backward method,⁴⁶ the genetic algorithm method,⁴⁷ and the artificial neural network method.⁴⁸ Unlike the wrapper method, filter methods do not interact with the regressing or classification model. Instead, it only deals with the statistical properties of descriptors. An example of a filter-type descriptor selection is removing correlated and low-variance descriptors from all calculated descriptors.^{49,50} In this thesis both approaches were used.

2.2.3.2 QSAR modeling Methods

Generally, QSPR modeling is carried out in two fashions: (1) regression, where a set of descriptors (predictor variables) is related to a set of targets (response variables), and (2) classification, where predictor variables are related to class labels (e.g., “high-performing” or “low performing”) rather than continuous response variables. More specifically, in classification QSAR modelling, classifier is trained on training data to return a score often between 0 and 1

for each material which is then transformed to a class label assigned to a material by choosing a classifier cut-off value. The common methods used in these two approaches are discussed in the following two sections.

2.2.3.3 Regression QSAR modelling methods

The simplest and most common method used in regression-based QSAR modelling is the multilinear regression (MLR), where a target property of a material is correlated to its descriptors with the use of a linear function:

$$y_i = c_0 + c_1x_{i1} + c_2x_{i2} + c_3x_{i3} + c_4x_{i4} + \dots + c_nx_{in} \quad 2.21$$

where n is the number of descriptors, x_i and y_i are descriptors and target property for the i th material, respectively, and c are coefficients that are found using the least-squares criterion. This method can also be extended to any order of polynomial regression. While this method is favoured due to its simplicity and ease of interpretation, it suffers from overfitting in cases where the number of descriptors is larger than that of samples. It is said that the number of descriptors used in this model should be at least one fifth the number of samples.⁵¹ In addition, this method is not appropriate for cases where the descriptors are correlated to each other as it does not take into account dependency of descriptors. As alternatives, principle component regression (PCR) and partial least-squares (PLS), where linear regression models are built on dimensionally reduced data, are used. Basically, PCR is performed in two stages: (1) descriptor values are first transformed onto a new coordinate system such that the first coordinate (principle component) has the largest possible covariance of descriptors and each succeeding component in turn has the highest variance, and (2) a subset of components, including the highest variances of descriptors, is selected to be used in the regression procedure. Similarly,

PLS tries to find a linear regression model by projecting both the descriptors and the target properties to a new space. More specifically, it searches for so-called latent variables (LVs), which explain as much as possible of the covariance between descriptors and target properties (i.e., response variables) while modelling the responses of the target property well. LVs themselves are not observable, but they are the underlying factors which accounts for most of the variation in the target property. Indeed, PLS avoids the problem of collinear descriptors by extracting these LVs while simultaneously models the response of the target property. This method is particularly suited when the number of descriptors exceed the number of samples, and when multiple target properties need to be predicted.

2.2.3.4 Classification QSAR modelling methods

As mentioned before, in classification problems, where materials are categorized into predefined classes, the predictor variables are related to the classes. The simplest method used to find this relationship is the decision tree (DT). A DT model is obtained (typically from simple binary rules) in a recursive fashion by partitioning a training set into successively purer subsets of materials from any of the predefined classes through choosing appropriate descriptors with certain thresholds. Figure 2.5 shows an example of the DT graph for a set of samples with a binary class, “high-performing” and “low-performing”. Each branch of the tree represents a threshold level given for the value of the descriptor, where the samples with values higher than the level threshold are branched to the right, while samples with lower values are branched to the left. Each non-terminal node of the tree represents the descriptor used to separate samples. Each terminal node represents an averaged DT score over all the samples in its corresponding branch, which can be related to a class label by choosing a classifier cut-off

value. For example, with the cut-off of 0.5, the grey terminal node represents the “high-performing” class of samples which follow the following rule: the values for descriptor A greater than X and the values for descriptor B greater than Y. In this thesis, the decision tree (DT) models were built using the Scikit-learn module for Python.⁵²

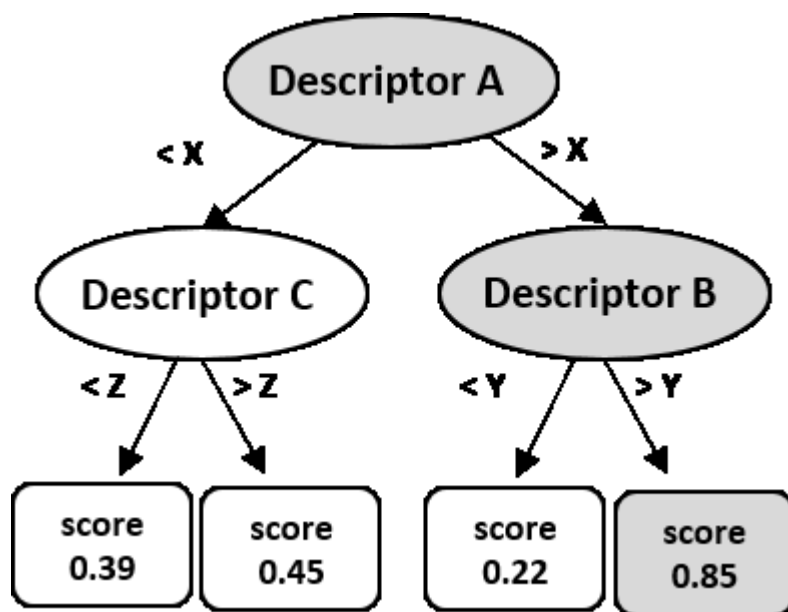


Figure 2.5: An example of a Decision Tree (DT) model with three descriptors A, B, and C. Each branch of the tree represents a threshold level given for the value of a descriptor (X for descriptor A, Z for descriptor C, and Y for descriptor B), where the samples with values higher than the level threshold are branched to the right, while samples with lower values are branched to the left. The high-performing node is highlighted in grey. The scores in the lowest branches are averaged QSPR scores over all samples in that branch.

While DT is favored for its ability to provide simple general rules which can be easily followed in rational design of materials, the greatest predictive power in classification problems lies in nonlinear machine learning techniques such as support vector machines (SVMs) capable of handling complex correlation patterns.⁵³ SVMs have been designed to minimize structural risk that prevents over-fitting by incorporating a regularization penalty into the model development procedure. The regularization penalty can be viewed as implementing a form

of Occam's razor that prefers simpler functions over more complex ones, controlling the bias/variance trade-off. Other machine learning methods, such as Artificial Neural Networks, are based on empirical risk minimization, which minimizes the difference between the predicted target property and the actual target property. Therefore, SVMs are less vulnerable to the over-fitting problem, and so they can typically deal with a larger number of features. Here, we briefly discuss the details of SVMs. For the sake of continuity and clarity, we first introduce linearly separable classification problem, and then discuss non-linearly separable classification problems.

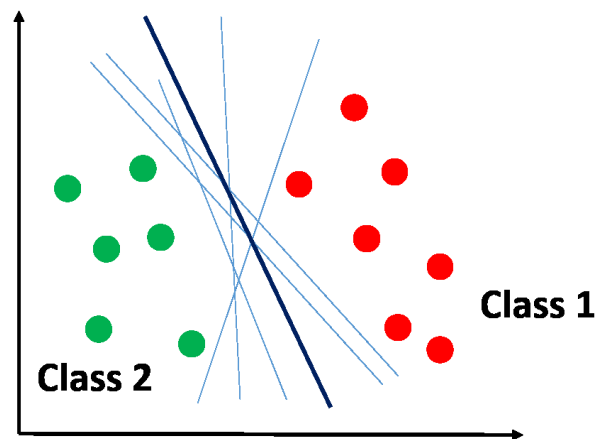


Figure 2.6: An example of a linearly separable two-class data with possible separating lines. Lines in light-blue color are assigned as “bad” lines for characterizing the two data classes within SVM’s concept, while the line in dark-blue color is an optimal line which SVM attempts to find.

As an example of linearly separable classification problem, consider a two-class data set in Figure 2.6, where red color represents one class and green color represents the other class. The goal of SVM is to find a line which optimally separates the space so that all existing points are classified into different regions of the two dimensional space, but more importantly, additional points that are added later to the two sets will have the greatest likelihood of being correctly classified as well. Clearly, there are many lines (see Figure 2.6) that can linearly

separate these two classes from each other, but all of them are not equally good from the SVM's theory point of view. A separating line too close to the data points in either class is bad because it will be noise sensitive and will not generalize correctly. In other words, a small change in the position of a point will move it from one class to the other. In fact, the line should be as far as possible from the data points in both classes. This is exactly what the operation of SVM relies on: finding the line that gives the largest minimum distance to the training examples. This distance receives the important name of **margin** within SVM's theory, and the 'line' is replaced by '**hyperplane**' for the sake of generalization to spaces of more than two dimensions (*in geometry a hyperplane is a subspace of one dimension less than its ambient space. If a space is 3-dimensional then its hyperplanes are the 2-dimensional planes, while if the space is 2-dimensional, its hyperplanes are the 1-dimensional lines*).⁵⁴ With this notation, SVM attempts to find the optimal separating hyperplane that *maximizes* the margin of the training data.

In most cases, training data are too noisy, and finding completely separating hyperplane (i.e. zero training error) will lead to poor generalisation. This is because the model is over-fitting the idiosyncrasies of the noise in the training data. To avoid this, SVM allows some points to be misclassified or even within the margin (see gray area (margin) in Figure 2.7), while still *maximizing* the margin of the training data. This refers to the balance between margin maximization and training error minimization.

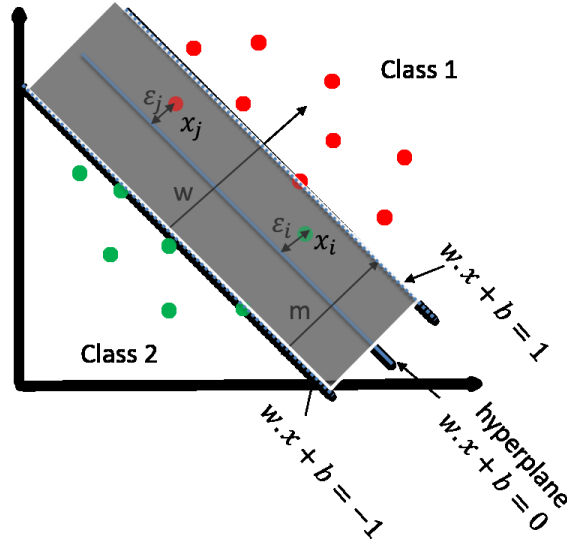


Figure 2.7: An example of noisy data, where SVM finds an optimal hyperplane (blue line) by maximizing margin (m) and allowing some points to be misclassified or in the margin (e.g., x_i and x_j). The gray area represents the margin.

In terms of formalism, an optimally separating hyperplane for two-class materials with n number of descriptors can be written as:

$$\mathbf{w} \cdot \mathbf{x} + b = 0 \quad 2.22$$

where the parameters w and b are found by maximizing the margin, $m = 1/\|\mathbf{W}\|$, through solving the following constrained optimization problem:

$$\begin{aligned} & \text{minimize } \frac{\|\mathbf{w}\|^2}{2} + C \sum_{i=1}^n \varepsilon_i \\ & \text{subject to } y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1 - \varepsilon_i, \quad \text{for } y_i \in \{1, -1\} \text{ and } i = 1, \dots, n \end{aligned} \quad 2.23$$

where the minimization of $\frac{\|\mathbf{w}\|^2}{2} + C \sum_{i=1}^n \varepsilon_i$ represents the maximization of the margin , $C \sum_{i=1}^n \varepsilon_i$ accounts for the possible training error (points x_i and x_j in Figure 2.7), y_i is the class label and $\mathbf{x}_i$ is the descriptor vector (i.e., $x_1, x_2, x_3, x_4, \dots, x_n$; n is the number of descriptors) for material i , ε_i is “slack variable” in an optimization which can be thought of as the distance from the separating hyperplane if an instance is misclassified and 0 if otherwise, $\mathbf{w}$ is the vector perpendicular to the hyperplane, and C is the regularization parameter which controls the trade-off between margin maximization and training error minimization. The reasons for

introducing the regularization parameter, C , is to control possible overfitting occurring in noisy data and to generalize the SMV model.⁵⁵

With the use of the Lagrange multiplier, the optimization problem becomes:

$$L(\mathbf{w}, \mathbf{b}, \alpha) = \frac{\|\mathbf{w}\|^2}{2} + C \sum_{i=1}^n \varepsilon_i + \sum_{i=1}^n \alpha_i [1 - \varepsilon_i - \mathbf{y}_i(\mathbf{w} \cdot \mathbf{x}_i + \mathbf{b})] - \sum_{i=1}^n r_i \varepsilon_i \quad 2.24$$

where α_i and r_i are Langrange multipliers and:

$$\begin{aligned} \alpha_i &\geq 0 \\ r_i &\geq 0 \end{aligned} \quad 2.25$$

The minimization of L with respect to b , w , and ε gives:

$$\mathbf{w} = \sum_{i=1}^n \alpha_i \mathbf{x}_i \mathbf{y}_i, \sum_{i=1}^n \alpha_i \mathbf{y}_i = \mathbf{0}, \text{ and } C = \alpha_i + r_i \quad 2.26$$

Substituting equation 2.26 into equation 2.24 gives the dual version of the constrained optimization problem:

$$\begin{aligned} \max L &= \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j \mathbf{y}_i \mathbf{y}_j (\mathbf{x}_i \cdot \mathbf{x}_j) \\ \text{Subject to } 0 &\leq \alpha_i \leq C, \sum_{i=1}^n \alpha_i \mathbf{y}_i = \mathbf{0} \end{aligned} \quad 2.27$$

where $\mathbf{x}_i \cdot \mathbf{x}_j$ is the dot product of descriptor vectors for materials i and j , and α_i is obtained by maximising the function L . Once α_i are found, w is obtained by $w = \sum_{i=1}^n \alpha_i \mathbf{x}_i \mathbf{y}_i$. Considering any $\mathbf{x}_s$ that satisfies $\mathbf{y}_j(\mathbf{w} \cdot \mathbf{x}_j + \mathbf{b}) = 1$ and substituting $w = \sum_{i=1}^n \alpha_i \mathbf{x}_i \mathbf{y}_i$ into $\mathbf{y}_s(\mathbf{w} \cdot \mathbf{x}_s + \mathbf{b}) = 1$ gives:

$$\mathbf{b} = \mathbf{y}_s - \mathbf{x}_s \cdot \sum_{i=1}^n \alpha_i \mathbf{x}_i \mathbf{y}_i \quad 2.28$$

Once b is obtained, substituting $w = \sum_{i=1}^n \alpha_i \mathbf{x}_i \mathbf{y}_i$ into $f(\mathbf{x}) = \text{sign}(\mathbf{w} \cdot \mathbf{x} + \mathbf{b})$, gives the hyperplane with the maximum margin:

$$f(\mathbf{x}) = \text{sign}\left(\sum_{i=1}^n \alpha_i \mathbf{y}_i (\mathbf{x}_i \cdot \mathbf{x}) + \mathbf{b}\right) \quad 2.29$$

where x is the descriptor vector for an unknown material.

For non-linearly separable data in input space (the space where point x_i are located), descriptor vectors are mapped onto one feature space (possibly with a higher dimension) where the separating hyperplane is found as illustrated in Figure 2.8.

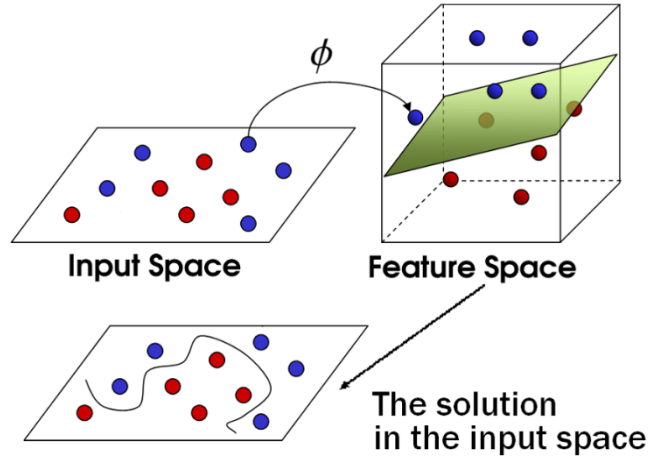


Figure 2.8: A representation of mapping descriptor vectors from an input space onto a feature space where data is linearly separable

By replacing the descriptor vectors with transformed descriptor vectors $\phi(x)$, the dual version of the constrained optimization problem becomes:

$$\begin{aligned} \max L &= \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j \phi(x_i) \cdot \phi(x_j) \\ \text{Subject to } 0 &\leq \alpha_i \leq C, \sum_{i=1}^n \alpha_i y_i = 0 \end{aligned} \quad 2.30$$

which requires very expensive computation of $\phi(x_i)$ (due to its extremely high dimensionality) for the calculation of $\phi(x_i) \cdot \phi(x_j)$. To avoid such an expensive calculation, since only inner product of $\phi(x_i)$ involves in the constrained optimization problem, SVM uses a kernel trick which replaces the inner product with a kernel function k where

$$K(x_i, x_j) = \phi(x_i) \cdot \phi(x_j) \quad 2.31$$

Considering equation 2.31, it is not necessary to know $\phi(x_i)$ explicitly, as kernel function is a

function of descriptor vectors in an input space, and returns the inner products between transformed descriptor vectors. Examples of kernel functions used in SMV methods are provided in Table 2.1.

Table 2.1: Examples of kernel functions used in support vector machine

Function	Name
$k(x_i, x_j) = (p + x_i \cdot x_j)^d$	Polynomial kernel
$k(x_i, x_j) = \exp\left(-\frac{\ x_i - x_j\ ^2}{2\sigma^2}\right)$	Gaussian radial basis function(rbf)
$k(x_i, x_j) = \exp\left(-\frac{\ x_i - x_j\ }{2\sigma^2}\right)$	Exponential kernel
$k(x_i, x_j) = (p + x_i \cdot x_j)^d \exp\left(-\frac{\ x_i - x_j\ ^2}{2\sigma^2}\right)$	Hybrid kernel
$k(x_i, x_j) = \tanh(kx_i \cdot x_j - \delta)$	Sigmoidal

By replacing $\phi(x_i) \cdot \phi(x_j)$ with $K(x_i, x_j)$, the dual version of the constrained optimization problem becomes:

$$\begin{aligned} \max L &= \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j K(x_i, x_j) \\ \text{Subject to } 0 &\leq \alpha_i \leq C, \sum_{i=1}^n \alpha_i y_i = 0 \end{aligned} \quad 2.32$$

where α_i is obtained by maximising the function L and the hyperplane is described as

$$f(x) = \text{sign}\left(\sum_{i=1}^n \alpha_i y_i k(x_i, x) + b\right) \quad 2.33$$

As given in Table 2.1, the kernel functions include parameters such as d and p in the case of a polynomial kernel. These parameters and even the regularization parameter C cannot be defined from the optimization problem, and must be manually tuned until the optimum values are achieved. This can be done by generating a set of models with different parameters, evaluating each of them on a set of unseen materials that were not used in the model development procedure, then selecting the model with parameters that show the highest

evolution metric value (e.g., R^2 metric for regression). Using a test set for validation increases the risk of overfitting in the model as the knowledge of the test set can “leak” into the model (parameters), and as a result evolution metrics no longer report the generalization performance. To prevent this problem, a cross-validation method is used. This involves dividing the training set into subsets, removing one subset (called validation set), building QSPR model using the remaining subsets, then validating the resulting model on the removed subset. The last three steps of this process are continued until all subsets have been validated against each other. The quality of the model against each omitted subset is measured using evolution metrics, which are discussed in the next section, and the average value is taken as an estimate for the model performance.

In this thesis, the RDF-based SVMs were built using the Python programming language and the scikit-learn toolbox.⁵² The kernel parameter and regularization parameter were optimized in a grid search fashion through cross-validation.

2.2.3.5 Validation of QSAR models

As mentioned before, in QSAR modeling, data is split into a training set and a test set. The training set is used to develop the model, and the test set is used to validate the model. Validation involves comparing predicted properties with observed properties in the test set, indicating the predictive power of the QSPR model. This predictive power is measured by the R^2 metric according to equation 2.34 in the case of regression and by accuracy metric according to equation 2.35 in the case of classification. If n is the number of samples, $\hat{y}_i$ and y_i are the observed property and the predicted property, respectively, for material i , $\bar{y}_i$ is the average of the predicted property, and $I(\hat{y}_i = y_i)$ is the indicator function where $I(\hat{y}_i = y_i) = 1$ if $\hat{y}_i = y_i$

and $I(\hat{y}_i = y_i) = 0$ if $\hat{y}_i \neq y_i$.

$$R^2 = 1 - \frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{\sum_{i=1}^n (\hat{y}_i - \bar{y}_i)^2} \quad 2.34$$

$$accuracy = \frac{1}{n} \sum_{i=1}^n I(\hat{y}_i = y_i) \quad 2.35$$

Good QSAR models have R^2 values and accuracy values close to 1 for regression and classification problems respectively.

2.3 References

- (1) Myers, a L.; Prausnitz, J. M. *AIChE J.* **1965**, *11* (1), 121.
- (2) Bae, Y.-S.; Farha, O. K.; Spokoyny, A. M.; Mirkin, C. a; Hupp, J. T.; Snurr, R. Q. *Chem. Commun. (Camb)*. **2008**, *0*, 4135.
- (3) Bae, Y.; Mulfort, K. L.; Frost, H.; Ryan, P.; Punnathanam, S.; Broadbelt, L. J.; Hupp, J. T.; Snurr, R. Q. *Langmuir* **2008**, *24* (18), 8592.
- (4) Manning F. S. and Thompson; E., R. *Oil Field Processing of Petroleum, Vol. I*; PennWell Publishing Co.: Tulsa, Okla, 1991.
- (5) Cavenati, S.; Grande, C. a; Rodrigues, a E. *J. Chem. Eng. Data* **2004**, *49*, 1095.
- (6) Pan, H.; Ritter, J. A.; Balbuena, P. B. *Langmuir* **1998**, *14* (11), 6323.
- (7) Wang, X. Sen; Ma, S.; Sun, D.; Parkin, S.; Zhou, H. C. *J. Am. Chem. Soc.* **2006**, *128* (51), 16474.
- (8) 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 (80-.)*. **2010**, *329* (5990), 424.
- (9) Yuan, D.; Zhao, D.; Sun, D.; Zhou, H. C. *Angew. Chemie - Int. Ed.* **2010**, *49* (31), 5357.
- (10) Alhamami, M.; Doan, H.; Cheng, C. H. *Materials (Basel)*. **2014**, *7* (4), 3198.
- (11) Boutin, A.; Coudert, F.-X.; Huet, M.-A. S.; Neimark, A. V; Férey, G.; Fuchs, A. H. *J. Phys. Chem. C* **2010**, *114*, 22237.
- (12) Hamon, L.; Serre, C.; Devic, T.; Loiseau, T.; Millange, F.; Férey, G.; De Weireld, G. *J. Am. Chem. Soc.* **2009**, *131* (25), 8775.
- (13) Burtch, N. C.; Jasuja, H.; Walton, K. S. *Chem. Rev.* **2014**, *114*, 10575.
- (14) Colombo, V.; Galli, S.; Choi, H. J.; Han, G. D.; Maspero, A.; Palmisano, G.; Masciocchi, N.; Long, J. R. *Chem. Sci.* **2011**, *2*, 1311.
- (15) Choi, H. J.; Dincă, M.; Dailly, A.; Long, J. R. *Energy Environ. Sci.* **2010**, *3* (1), 117.
- (16) Zhao, Z.; Li, Z.; Lin, Y. S. *Ind. Eng. Chem. Res.* **2009**, *48* (22), 10015.
- (17) Millward, A. R.; Yaghi, O. M. *J. Am. Chem. Soc.* **2005**, *127* (51), 17998.
- (18) Cheon, Y. E.; Suh, M. P. *Chem. - A Eur. J.* **2008**, *14* (13), 3961.
- (19) Saha, D.; Bao, Z.; Jia, F.; Deng, S. *Environ. Sci. Technol.* **2010**, *44* (5), 1820.
- (20) Bae, Y. S.; Snurr, R. Q. *Angew. Chemie - Int. Ed.* **2011**, *50* (49), 11586.
- (21) Atkins, Peter; De Paula, Julio (2006). *Atkins' Physical Chemistry (8th ed.)*. W. H. Freeman. p.

112.

- (22) Mayo, S. L.; Olafson, B. D.; Iii, W. a G.; Eb, E.; El, E. a E. T. *J. Phys. Chem.* **1990**, *101* (Suite 540), 8897.
- (23) Rappe, a. K.; Casewit, C. J.; Colwell, K. S.; Goddard III, W. a.; Skiff, W. M. *J. Am. Chem. Soc.* **1992**, *114*, 10024.
- (24) Vaidhyanathan, R.; Iremonger, S. S.; Shimizu, G. K. H.; Boyd, P. G.; Alavi, S.; Woo, T. K. **2010**, *330*, 650.
- (25) Wang, L.; Wang, L.; Zhao, J.; Yan, T. *J. Appl. Phys.* **2012**, *111* (11).
- (26) Bae, Y.-S.; Farha, O. K.; Spokoyny, A. M.; Mirkin, C. a; Hupp, J. T.; Snurr, R. Q. *Chem. Commun. (Camb)*. **2008**, *0* (35), 4135.
- (27) Martin, M. G.; Siepmann, J. I. *J. Phys. Chem. B* **1999**, *103* (21), 4508.
- (28) Campañá, C.; Mussard, B.; Woo, T. K. *J. Chem. Theory Comput.* **2009**, *5*, 2866.
- (29) Kresse, G.; Hafner, J. *Phys. Rev. B* **1993**, *47* (1), 558.
- (30) Kresse, G.; Hafner, J. *Phys. Rev. B* **1994**, *49* (20), 14251.
- (31) Kresse, G.; Furthmüller, J. *Phys. Rev. B* **1996**, *54* (16), 11169.
- (32) Kresse, G.; Furthmüller, J. *Comput. Mater. Sci.* **1996**, *6*, 15.
- (33) Wilmer, C. E.; Farha, O. K.; Bae, Y.-S.; Hupp, J. T.; Snurr, R. Q. *Energy Environ. Sci.* **2012**, *5* (12), 9849.
- (34) Kadantsev, E. S.; Boyd, P. G.; Daff, T. D.; Woo, T. K. *J. Phys. Chem. Lett.* **2013**, *4* (18), 3056.
- (35) Briard, J. G.; Fernandez, M.; De Luna, P.; Ben, R. N.; Woo, T. K. *Sci. Rep.* **2016**, No. October 2015, 1.
- (36) Basant, N.; Gupta, S.; Singh, K. P. *SAR QSAR Environ. Res.* **2016**, *27* (1), 67.
- (37) Wang, N.-N.; Dong, J.; Deng, Y.-H.; Zhu, M.-F.; Wen, M.; Yao, Z.-J.; Lu, A.-P.; Wang, J.-B.; Cao, D.-S. *J. Chem. Inf. Model.* **2016**, *56*, 763.
- (38) Mafud, A. C.; Ferreira, L. G.; Mascarenhas, Y. P.; Andricopulo, A. D.; Moraes, J. De. *Trends Parasitol.* **2016**, *xx*, 1.
- (39) Brogi, S.; Papazafiri, P.; Roussis, V.; Tafi, A. *Eur. J. Med. Chem.* **2013**, *67*, 344.
- (40) Kunal Roy, Supratik Kar, R. N. Das. *A Primer on QSAR/QSPR Modeling: Fundamental Concepts*; Springer, 2015.
- (41) Todeschini, R.; Consonni, V. *Handbook of Molecular Descriptors*, 11th ed.; 2000.
- (42) Lipkus, A. H. *J. Chem. Inf. Comput. Sci.* **2001**, *41* (2), 430.
- (43) Fernandez, M.; Woo, T. K.; Wilmer, C. E.; Snurr, R. Q. *J. Phys. Chem. C* **2013**, *117* (15), 7681.
- (44) Karelson, M.; Lobanov, V. S.; Katritzky, A. R. *Chem. Rev.* **1996**, *96* (3), 1027.
- (45) Dutta, D.; Guha, R.; Wild, D.; Chen, T. *J. Chem. Inf. Model.* **2007**, *47* (3), 989.
- (46) Yang, S.-P.; Song, S.-T.; Tang, Z.-M.; Song, H.-F. *Acta Pharmacol. Sin.* **2003**, *24* (9), 897.
- (47) Rogers, D.; Hopfinger, A. J. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 854.
- (48) Wikel, J.; Dow, E. *Bioorg. Med. Chem. Lett.* **1993**, *3* (4), 645.
- (49) Guha, R.; Jurs, P. C. *J. Chem. Inf. Comput. Sci.* **2004**, *44* (6), 2179.
- (50) Goll, E. S.; Jurs, P. C. *J. Chem. Inf. Model.* **1999**, *39* (6), 974.

- (51) Topliss, J. G.; Edwards, R. P. *J. Med. Chem.* **1979**, 22 (10), 1238.
- (52) Pedregosa, F.; Grisel, O.; Weiss, R.; Passos, A.; Brucher, M. **2011**, 12, 2825.
- (53) Cortes, C.; Vapnik, V. *Mach. Learn.* **1995**, 20 (3), 273.
- (54) <https://en.wikipedia.org/wiki/Hyperplane>.
- (55) Nello Cristianini, J. S.-T. *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*; Cambridge University Press, 2000.

3 Computational modelling of CO₂ adsorption on Phosphonate Monoesters-base MOF

This chapter discusses calculation performed on CuBDPMe, which is the first Phosphonate Monoesters-base MOF is reported. This study was performed in collaboration with the Shimizu group at the University of Calgary, and was published in part in *Iremonger, S. S.; Liang, J.; Vaidhyanathan, R.; Martens, I.; Shimizu, G. K. H.; Daff Thomas, D.; Aghaji, M. Z.; Yeganegi, S.; Woo, T. K. J. Am. Chem. Soc.* **2011**, *133*, 20048–20051. My role in this study was to perform the GCMC simulation and DFT calculation to elucidate the behavior of CO₂ adsorption in the MOF.

Abstract

Bidentate phosphonate monoesters are analogues of popular dicarboxylate linkers in MOFs, but with an alkoxy tether close to the coordinating site. Herein, we report a computational study on Cu(1,4-benzenediphosphonate bis(monomethyl ester), CuBDPMe, which has a modest CO₂ uptake capacity, with a high CO₂ heat of adsorption of 45 kJ mol⁻¹. Computational modeling reveals that the CO₂ uptake is extremely sensitive both to the flexing of the structure and to the orientation of the alkyl tether.

3.1 Introduction

Metal Organic Frameworks (MOFs)/Porous Coordination Polymers (PCPs) offer systematic and tuneable routes to the generation of new porous solids.¹ The role of the organic linkers is foremost to delineate the primary pore structure but these groups also define the chemical potential of the pore surfaces² and can possibly kinetically gate sorption phenomena.³ The fact that these solids can also demonstrate flexible architectures further extends their scientific interest and offers potential for new applications tied to their dynamic behaviour.⁴

Many MOF/PCP materials reported incorporate carboxylate anions as, in appropriate rigid and polyvalent forms,⁵ these linkers provide sufficient thermodynamic stability to enable permanent porosity but also sufficient kinetic lability that crystalline products are typically

obtained. Phosphonate monoesters (RP(=O)(-OH)(-OR)) are a largely unexplored ligating functionality in coordination polymer science.^{6,7} It was reported the compound Zn 1,4-benzenediphosphonate bis(monoethyl ester), Zn(BDPet) , a van der Waals compound with permanent porosity despite having only two-dimensional connectivity. In this compound, the monoanionic and bidentate coordination mode of the phosphonate monoesters was reminiscent of the ligation displayed by dicarboxylate linkers in MOFs. A fundamental difference between a phosphonate monoester and a carboxylate is obviously the presence of the alkoxy tether on the P atom.

The following describes the computational modeling of CO_2 adsorption in a phosphonate-based MOF, CuBDPMe . The characteristics of CO_2 uptakes, including adsorption isotherm, heat of adsorption, and binding sites, in CuBDPMe are modelled via a combination of grand canonical Monte Carlo (GCMC) simulation and density functional theory (DFT). Our modeling reveals that the CO_2 uptake is sensitive both to the flexing of the structure and to the orientation of the methyl tether.

3.2 Computational details

DFT calculations were carried out in the Vienna ab initio simulation package (VASP).⁸⁻¹¹ The PBE exchange-correlation functional^{12,13} and PAW potentials^{14,15} were used with an energy cutoff of 520 eV. All DFT calculations used a single unit cell with a gamma centred $2 \times 2 \times 2$ k-point grid on the Brillouin zone. The semi-empirical dispersion potential correction of Grimme¹⁶ was included in the energy and force calculations with the default parameters for the PBE functional. Spin polarisation was enabled for all calculations to account for the d^9 -Cu(II) centers. Depicted in Figure 3.1 is an isosurface plot of the net spin density for the empty framework of

Cu. This reveals that the ground state is antiferromagnetic with localization of the spin on the copper centers. The localization of the spin on the different metal centers could be controlled by using the MAGMOM key in VASP which spin polarizes the initial density guess in the calculation. Negligible difference in energy was found for the permutations of spin up and spin down metal centers.

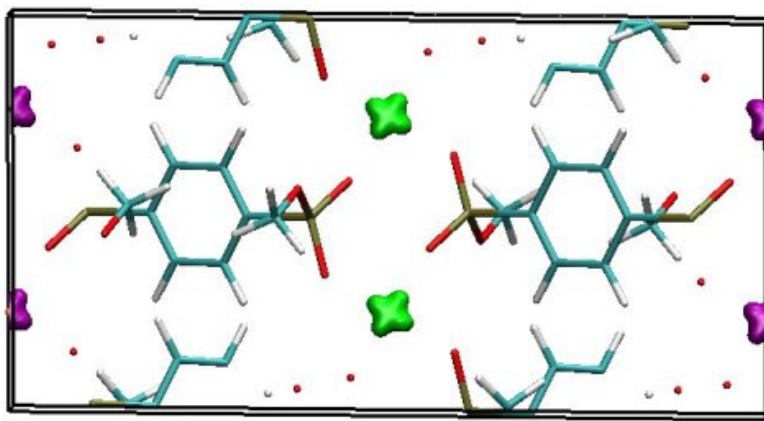


Figure 3.1: Ground state spin-density localization from DFT calculations of CuBDPMe at an isosurface value of $0.5e$. Up (purple) and down (green) unpaired spins are localized on the copper atoms.

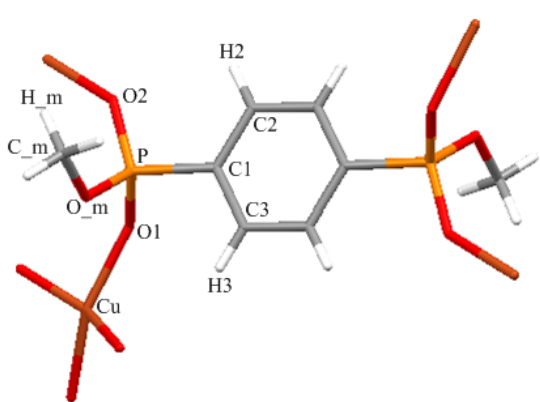
Grand canonical Monte Carlo (GCMC) simulations were based on classical force-field using Lennard-Jones parameters given in Table 3.1 and partial atomic charges given in Table 3.2. For the framework, the Lennard-Jones (LJ) potential were taken from the Universal Force Field (UFF)¹⁷ and combined with Lorentz-Berthelot mixing rules for cross terms. The non-bonded parameters of CO₂ were taken from García-Sánchez et al.¹⁸ 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 CO₂ molecule had a fixed bond length of 1.16 Å and partial charges of $+0.645e$ and $-0.3325e$ for carbon and oxygen. Partial atomic charges for Coulombic interactions involving the framework were calculated using the REPEAT method,¹⁹ where atomic charges are fitted to

reproduce the electrostatic potential derived from the DFT calculations. To derive these charges, the hydrogen atom positions of the X-ray structure of CuBDPMe were optimized at the DFT level and the then optimized structure was used for the derivation of the REPEAT charges.

Table 3.1: Lennard-Jones parameters from the UFF.

Atomic type	σ / Å	ϵ / kcalmol ⁻¹
Cu	3.1137	0.005
C	3.4309	0.105
P	3.6946	0.305
O	3.1181	0.060
H	2.5711	0.044
C (CO ₂)	2.7450	0.05948
O (CO ₂)	3.0170	0.17023

Table 3.2: REPEAT partial atomic charges for the X-ray structure of CuBDPMe.

	atom	q / e
	Cu	1.023
	P	1.322
	O1	-0.672
	O2	-0.782
	O_m	-0.467
	Cu	1.023
	C1	-0.167
	C2	-0.179
	C3	-0.164
	C_m	0.134
	H1	0.163
	H2	0.181
	H_m	0.005

The GCMC simulations were run using an in-house developed code based on the open-source DL_POLY classic package (formerly known as DL_POLY 2).²⁰ For the guest molecule addition and deletion moves, the ideal gas chemical potential was used for the acceptance probability. The geometry of the framework was frozen and taken from the X-ray structure, except for the hydrogen atom positions which were optimized at the DFT level. A 2x4x4 supercell was generated from the unit cell which enabled use of a 12.5 Å cutoff for the Lennard-

Jones interactions. Electrostatic interactions are calculated using the Ewald summation method. Systems were equilibrated for 10 million MC steps, and results calculated for 12.5 million steps and averaged over four separate MC chains.

3.3 Result & Discussion

The crystal structure of CuBDPMe is shown in Figure 3.2c. In CuBDPMe, one-dimensional Cu(R-PO₃Me) chains are formed down the c-axis (Fig. 3.2b), with the chains being linked by fully deprotonated 1,4-benzendiphosphonate bis (mono methyl ester), BDPMe, dianions forming a rhombohedral grid of 14.933(4) x 7.024(2) Å (P-P distance and metal-metal distances as depicted by the arrows in Fig. 3.2c). The methyl groups of the phosphonate esters protrude into the pores but leave an aperture of ~6 Å.

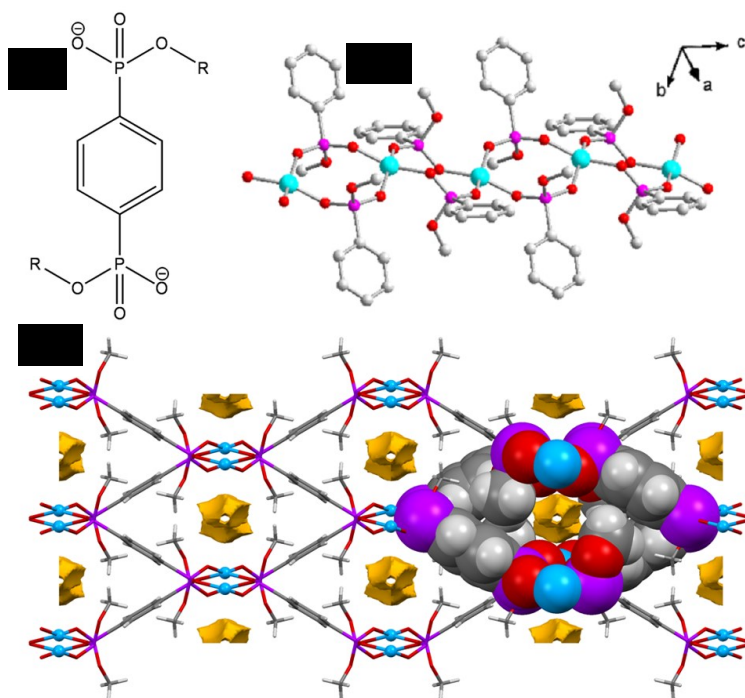


Figure 3.2: a) The dianionic ligand, BDPR, where R = Me. b) Single crystal structure of one column of CuBDPMe, showing the bidentate coordination of the phosphonate monomethylester and the 4-coordinate geometry of Cu. c) crystal structure with solvent accessible voids (1.2 Å probe in gold). In c) one channel is shown in a space filling representation.

A modest uptake of CO₂ at ambient temperature was observed, 1.38 mmol/g (6.06 wt%) at 273 K and 1200 mbar. The enthalpy of adsorption (ΔH_{ads}) for CO₂ was determined as a function of loading. At zero loading, indicative of the highest energy sites, the ΔH_{ads} was calculated as 45 kJ/mol, a very high value for a physisorptive process.²¹ Recent reports have shown a strong correlation between high ΔH_{ads} and high selectivities for CO₂ capture.²² Often, high ΔH_{ads} can be connected to the presence of coordinately unsaturated metal centers (e.g. HKUST-1, 35 kJ/mol;²³ MIL-101, 44 kJ/mol;²⁴ Mg-MOF-74/CPO-27-Mg, 47 kJ/mol;²⁵ MIL-100, 63 kJ/mol²⁴). Considering three pieces of data (high value for ΔH_{ads} ; stoichiometry of 0.45 CO₂/Cu; transannular Cu...Cu distance of 7.02 Å between pseudo-square planar Cu centers), it is tempting to hypothesize that the CO₂ molecules in this system are in a dual end-on bridging mode between Cu centers. ΔH_{ads} value can also be augmented by the presence of polar functional group,²⁶ such as amines lining the pores, although higher amination is not always necessarily beneficial for CO₂ uptakes.²⁷

To investigate the nature of the CO₂ binding and sorption properties of CuBDPMe, dispersion corrected periodic DFT and classical grand canonical Monte Carlo (GCMC) simulations were employed. The dual end-on bridging interaction of CO₂ with two Cu centers was first examined with DFT as such a binding mode might involve weak orbital interactions. All attempts to locate a minimum energy structure corresponding to this mode of CO₂ failed even when the framework and cell vectors were allowed to fully relax. Instead, all structures optimized to geometries in which the CO₂ molecules were oriented roughly perpendicular to the hypothesized end-on bridging mode in the direction of the channels.

To further investigate the CO₂ sorption in CuBDPMe, GCMC simulations were performed using the UFF parameters¹⁷ and REPEAT¹⁹ charges that were able to successfully reproduce the experimental CO₂ binding sites and isotherms of a Zn amino-triazole-oxalate MOF.²⁸ Using the X-ray structure of 2, Figure 3.3 shows the CO₂ oxygen atom probability distribution resulting from a GCMC simulation at 1 atm CO₂ and 273 K. The plots show that there are four symmetrically equivalent binding sites in the unit cell. The guest molecules align along the channels with either the carbon or oxygen of the CO₂ at the midpoint between the methyl groups. Consistent with the DFT calculations, no dual end-on bridging was seen in with the GCMC simulations. One key experimental observation supporting the lack of direct interaction of CO₂ with the Cu sites is that in the crystal structure, although the Cu centers are 4-coordinate, there is no ligated water. Typically, an open metal site requires activation²³⁻²⁵ which is not the case with CuBDPMe.

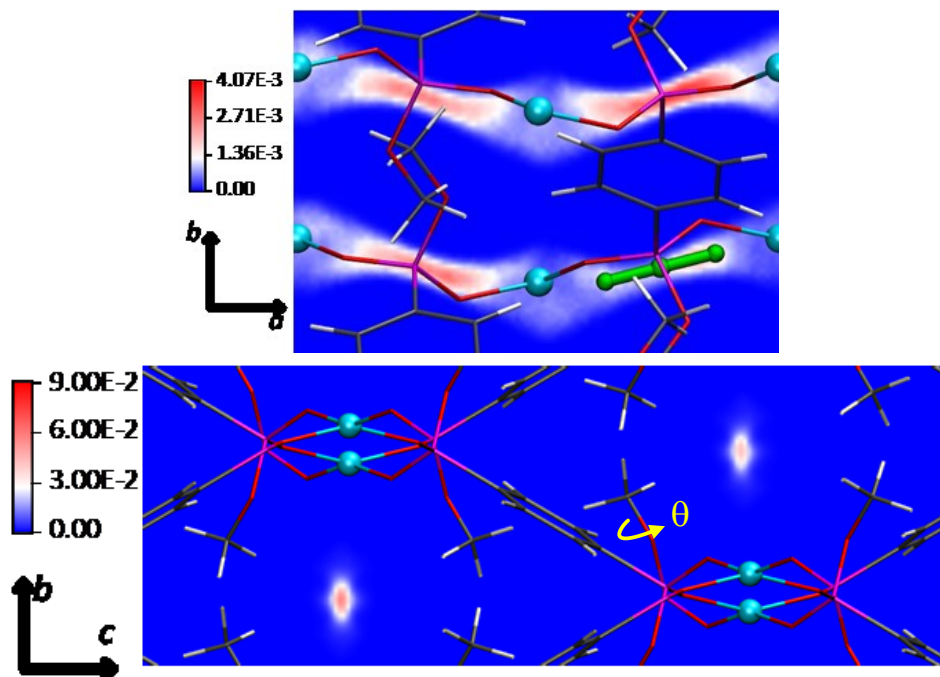


Figure 3.3: Probability density plots of the CO₂ oxygen atom in a unit cell of CuBDPMe at 1 bar, 273 K from simulation. In a) probabilities along the c-axis are projected onto the a-b plane while in b) probabilities along the b-axis are projected onto the b-c plane. A green CO₂ molecule is shown in a) for reference.

CO₂ adsorption isotherms were also determined from the GCMC simulations. The CO₂ adsorption behaviour of CuBDPMe was found to be highly sensitive to the structure of the framework. For example, when the solvent included X-ray structure of CuBDPMe is used in the simulation, the CO₂ adsorption is found to be 15-20% lower than the experimental result throughout the pressure range (Figure 3.4a). However, if this structure is adjusted to the lattice parameters determined in situ with 1.2 bar CO₂, the simulated isotherm reproduces the experimental uptake well throughout the pressure range examined. In fact, although the cell parameters changed by a less than 2% between the X-ray structure and adjusted one, the CO₂ loading changed by more than 20% at 1 atm. We also found a strong CO₂ uptake dependence on the methoxy group orientation, which can be defined by the C_{me}-O-P-O dihedral angle, φ ,

as highlighted in Figure 3.3. To elucidate this, a systematic variation of the pore width was achieved by altering the O-P-O-C_{Me} dihedral angles, ϕ . All 8 of these dihedrals in the unit cell are changed simultaneously. The energy, determined from single point DFT calculations, as a function of the change in the torsion is shown in Figure 3.4, indicating that the $\Delta\phi = -2$ and $+2$ degrees structures are viable. The REPEAT charges were recalculated for each structure and 2x4x4 supercells used in the GCMC isotherm calculations. Shown in Figure 3.4b are the simulated adsorption isotherms with the ϕ angle of all methoxy groups in the framework changed by $\Delta\phi = -2$ and $+2$ degrees from that in the original structure. Negative $\Delta\phi$'s correspond to the channel closing, while positive values open the channel. A 2° change in ϕ , which corresponds to a ~ 0.08 Å change in pore width, results in a dramatic increase or decrease in the simulated gas adsorption. Therefore, the slight inflection in the experimental isotherm of CuBDPMe at ~ 0.6 bar can be attributed to a subtle change in the framework geometry upon CO₂ uptake. While not explicitly demonstrated in this work, in general, the high sensitivity of the uptake to the methoxy group orientation in CuBDPMe suggests the potential of the methyl tethers in the phosphonate monoester based MOFs to modulate the adsorption behavior. The dynamic nature of the methyl tether may also permit optimal arrangement between CO₂ molecules to maximize cooperative interactions.²⁹

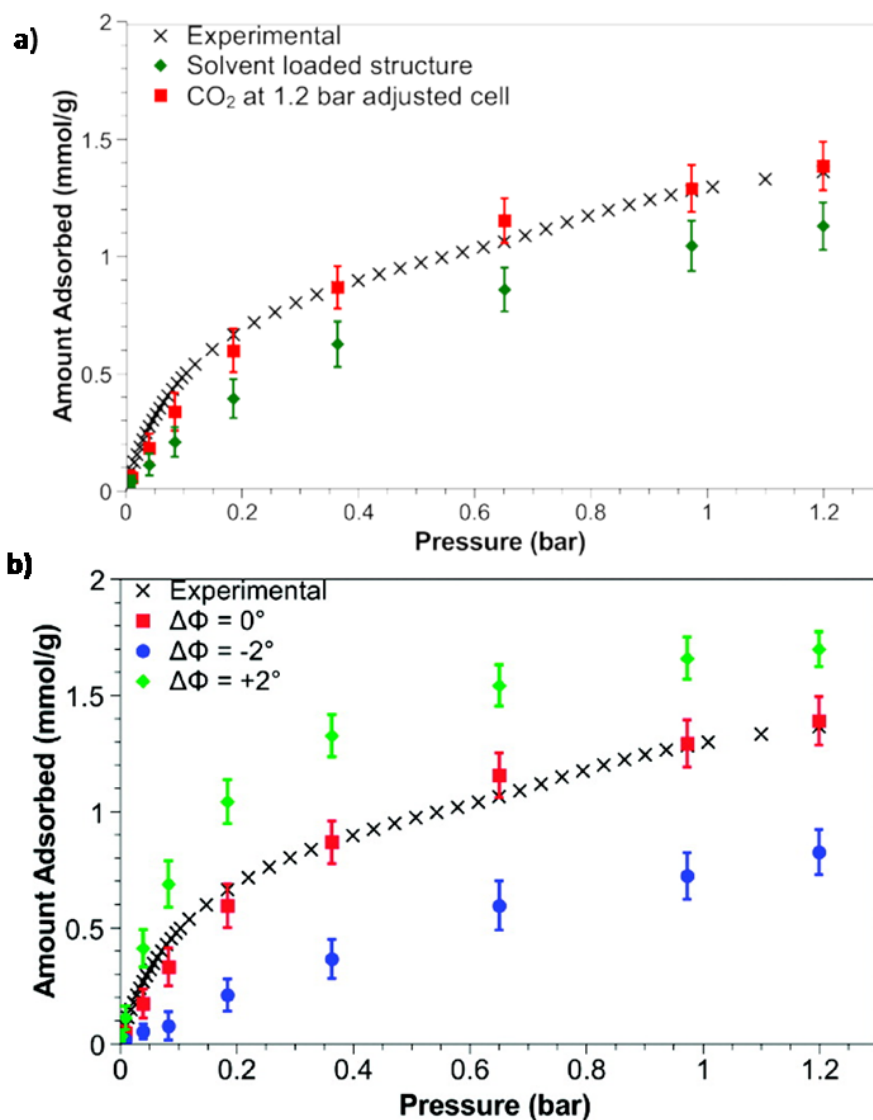


Figure 3.4: Comparison of experimental and simulated CO₂ adsorption isotherms at 273 K. a) The simulated isotherms were calculated with the X-ray structure, which was solvent loaded, and that where this structure's cell parameters were adjusted to that determined upon solvent removal and loaded with CO₂ at 1.2 bar. b) The simulated isotherms were calculated with the X-ray structure, where the geometry of the methoxy groups have been altered from the X-ray structure in silico

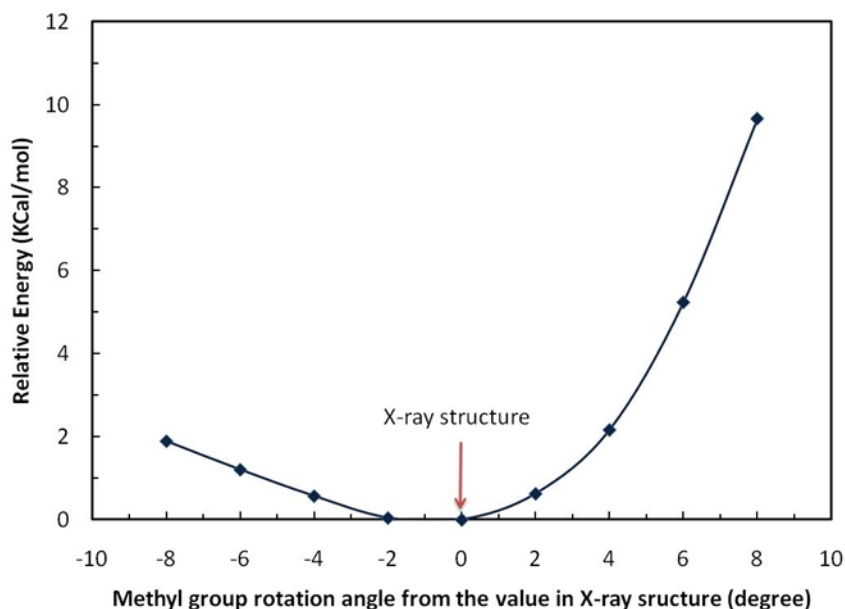


Figure 3.5: Change in the DFT calculated energy for variation of the torsion angle to methyl tethers

Isosteric heats of adsorption (HOA) were also determined from the GCMC simulations, calculated via the Green–Kubo fluctuation theory expression,

$$q_{isosteric} = \frac{\langle UN \rangle - \langle U \rangle \langle N \rangle}{\langle N^2 \rangle - \langle N \rangle \langle N \rangle} + RT \quad 3.1$$

Where N is the number of CO₂ molecules in the GCMC simulation and U is configurational energy for the CO₂ molecules.³⁰ Calculated in this manner, the binding enthalpies as a function of CO₂ loading are plotted in Figure 3.6. for CuBDPMe using the crystal structures with the methoxy group rotated by $\phi = 0^\circ, 2^\circ$ and 4° . The calculated heats of adsorption are significantly less than that determined experimentally (~ 36 kJ/mol versus 46 kJ/mol), however it is notable that the simulated HOA is very sensitive to the framework geometry as is found with the CO₂ uptake. Therefore, it would be difficult to exactly reproduce the experimental HOA without

considering a fully flexible framework due to the energetics of small distortions and relaxations of the framework in the presence and absence of the guest molecule.

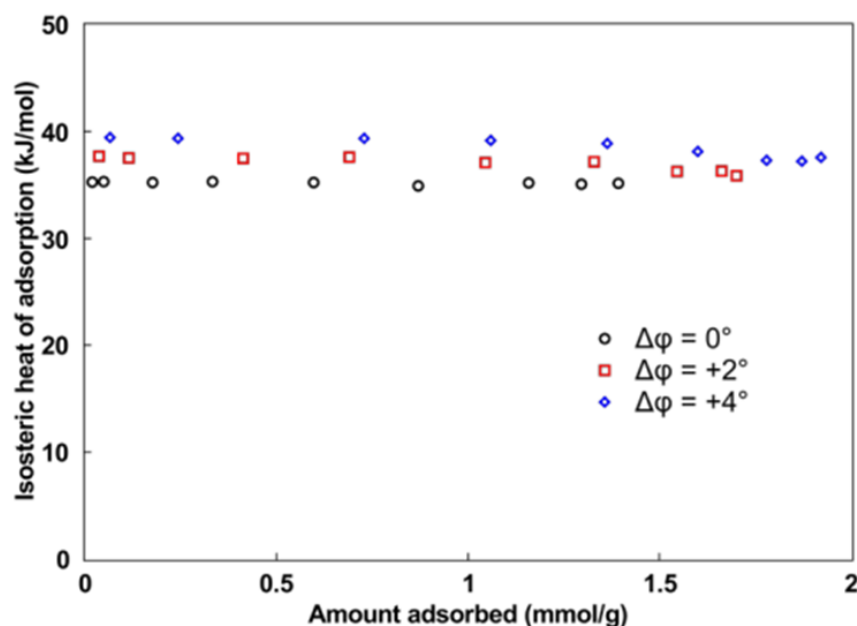


Figure 3.6: Isosteric heat of adsorption of CuBDPMe with lattice parameters scaled for CO₂ loading at 1.2 bar, as a function of guest loading determined from GCMC simulations at 273 K with the methyl group rotated by $\phi = 0^\circ$, 2° and 4°

3.4 Conclusion

Results from DFT calculation, followed by GCMC simulations, reveal that the CO₂ molecules in CuBDPMe align itself along the channel and they don't have a direct interaction with Cu sites. The simulated isotherms represent that the CO₂ adsorption behavior on CuBDPMe is highly sensitive to the methoxy orientations and the flexing of the structure. Less than 2% change in the lattice parameters results in 20 % change in the CO₂ uptake on CuBDPMe. The extreme dependency of uptake to the methoxy orientations suggests a gate opening mechanism for CO₂ diffusion in the channel.

3.5 References

- (1) (a) Kitagawa, S.; Matsuda, R. *Coord. Chem. Rev.* **2007**, *251*, 2490. (b) Zhao, D.; Meek, S. T.; Greathouse, J. A.; Allendorf, M. D. *Adv. Mater.* **2011**, *23*, 249. (c) Farha, O. K.; Hupp, J. T. *Acc. Chem. Res.* **2010**, *43*, 1166. (d) Tanabe, K. K.; Cohen, S. M. *Chem. Soc. Rev.* **2011**, *40*, 498. (e) Kuppler, R. J.; Timmons, D. J.; Fang, Q. R.; Li, J. R.; Makal, T. A.; Young, M. D.; Yuan, D. Q.; Zhao, D.; Zhuang, W. J.; Zhou, H. C. *Coord. Chem. Rev.* **2009**, *253*, 3042. (f) Czaja, A. U.; Trukhan, N.; Mueller, U. *Chem. Soc. Rev.* **2009**, *38*, 1284
- (2) (a) Phan, A.; Doonan, C. J.; Uribe-Romo, F. J.; Knobler, C. B.; O’Keeffe, M.; Yaghi, O. M. *Acc. Chem. Res.* **2010**, *43*, 58. (b) Li, J. R.; Kuppler, R. J.; Zhou, H. C. *Chem. Soc. Rev.* **2009**, *38*, 1477. (c) Wang, B.; Cote, A. P.; Furukawa, H.; O’Keeffe, M.; Yaghi, O. M. *Nature* **2008**, *453*, 207. (d) Herm, Z. R.; Swisher, J. A.; Smit, B.; Krishna, R.; Long, J. R. *J. Am. Chem. Soc.* **2011**, *133*, 5664. (e) Mason, J. A.; Sumida, K.; Herm, Z. R.; Krishna, R.; Long, J. R. *Energy Environ. Sci.* **2011**, *4*, 3030.
- (3) (a) Noro, S.; Tanaka, D.; Sakamoto, H.; Shimomura, S.; Kitagawa, S.; Takeda, S.; Uemura, K.; Kita, H.; Akutagawa, T.; Nakamura, T. *Chem. Mater.* **2009**, *21*, 3346. (b) Devic, T.; Horcajada, P.; Serre, C.; Salles, F.; Maurin, G.; Moulin, B.; Heurtaux, D.; Clet, G.; Vimont, A.; Greneche, J. M.; Le Ouay, B.; Moreau, F.; Magnier, E.; Filinchuk, Y.; Marrot, J.; Lavalley, J. C.; Daturi, M.; Ferey, G. *J. Am. Chem. Soc.* **2010**, *132*, 1127. (c) Burrows, A. D.; Frost, C. G.; Mahon, M. F.; Richardson, C. *Angew. Chem., Int. Ed.* **2008**, *47*, 8482. (d) Kaye, S. S.; Long, J. R. *J. Am. Chem. Soc.* **2008**, *130*, 806. (e) Morris, W.; Doonan, C. J.; Furukawa, H.; Banerjee, R.; Yaghi, O. M. *J. Am. Chem. Soc.* **2008**, *130*, 12626. (f) Matsuda, R.; Tsujino, T.; Sato, H.; Kubota, Y.; Morishige, K.; Takata, M.; Kitagawa, S. *Chem. Sci.* **2010**, *1*, 315. (g) Demessence, A.; D’Alessandro, D. M.; Foo, M. L.; Long, J. R. *J. Am. Chem. Soc.* **2009**, *131*, 8784.
- (4) (a) Tanaka, D.; Nakagawa, K.; Higuchi, M.; Horike, S.; Kubota, Y.; Kobayashi, L. C.; Takata, M.; Kitagawa, S. *Angew. Chem., Int. Ed.* **2008**, *47*, 3914. (b) Xiao, B.; Byrne, P. J.; Wheatley, P. S.; Wragg, D. S.; Zhao, X.; Fletcher, A.; Thomas, K. M.; Peters, L.; Evans, J. S. O.; Warren, J. E.; Zhou, W.; Morris, R. E. *Nat. Chem.* **2009**, *1*, 289. (c) Wang, Z.; Cohen, S. M. *J. Am. Chem. Soc.* **2009**, *131*, 16675. (d) Dubbeldam, D.; Krishna, R.; Snurr, R. Q. *J. Phys. Chem. C* **2009**, *113*, 19317. (e) Seo, J.; Matsuda, R.; Sakamoto, H.; Bonneau, C.; Kitagawa, S. *J. Am. Chem. Soc.* **2009**, *131*, 12792.
- (5) (a) Horike, S.; Shimomura, S.; Kitagawa, S. *Nat. Chem.* **2009**, *1*, 695. (b) Ferey, G.; Serre, C. *Chem. Soc. Rev.* **2009**, *38*, 1380. (c) Fletcher, A. J.; Thomas, K. M.; Rosseinsky, M. J. *J. Solid State Chem.* **2005**, *178*, 2491. (d) Llewellyn, P. L.; Maurin, G.; Devic, T.; LoeraSerna, S.; Rosenbach, N.; Serre, C.; Bourrelly, S.; Horcajada, P.; Filinchuk, Y.; Ferey, G. *J. Am. Chem. Soc.* **2008**, *130*, 12808. (e) Trung, T. K.; Trens, P.; Tanchoux, N.; Bourrelly, S.; Llewellyn, P. L.; LoeraSerna, S.; Serre, C.; Loiseau, T.; Fajula, F.; Ferey, G. *J. Am. Chem. Soc.* **2008**, *130*, 16926. (f) Maji, T. K.; Matsuda, R.; Kitagawa, S. *Nat. Mater.* **2007**, *6*, 142
- (6) (a) Braun, M. E.; Steffek, C. D.; Kim, J.; Rasmussen, P. G.; Yaghi, O. M. *Chem. Commun.* **2001**, 2532. (b) Rao, C. N. R.; Natarajan, S.; Vaidhyanathan, R. *Angew. Chem., Int. Ed.* **2004**, *43*, 1466.
- (7) (a) Chang, Y. D.; Zubieta, J. *Inorg. Chim. Acta* **1996**, *245*, 177. (b) Evans, O. R.; Manke, D. R.; Lin, W. *Chem. Mater.* **2002**, *14*, 3866. (c) Ayyappan, P.; Evans, O. R.; Cui, Y.; Wheeler, K. A.; Lin, W. *Inorg. Chem.* **2002**, *41*, 4978. (d) Chandrasekhar, V.; Sasikumar, P. *Dalton Trans.* **2008**, 6475. (e) Zurowska, B.; Brzuszkiewicz, A.; Ochocki, J. *Polyhedron* **2008**, *27*, 1721.
- (8) G. Kresse and J. Furthmüller, *Comput. Mater. Sci.*, **1996**, *6*, 15.

- (9) G. Kresse and J. Furthmüller, *Phys. Rev. B*, 1996, **54**, 11169.
- (10) G. Kresse and J. Hafner, *Phys. Rev. B*, 1993, **47**, 558.
- (11) G. Kresse and J. Hafner, *Phys. Rev. B*, 1994, **49**, 14251.
- (12) J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, 1996, **77**, 3865.
- (13) J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, 1997, **78**, 1396
- (14) P. E. Blöchl, *Phys. Rev. B*, 1994, **50**, 17953.
- (15) G. Kresse and D. Joubert, *Phys. Rev. B*, 1999, **59**, 1758.
- (16) S. Grimme, *J. Comput. Chem.*, 2006, **27**, 1787.
- (17) A. K. Rappe, C. J. Casewit, K. S. Colwell, W. A. Goddard and W. M. Skiff, *J. Am. Chem. Soc.*, 1992, **114**, 10024.
- (18) A. Garcia-Sanchez, C. O. Ania, J. B. Parra, D. Dubbeldam, T. J. H. Vlugt, R. Krishna, S. Calero, *Journal of Physical Chemistry C* 2009, **113**, 8814.
- (19) C. Campañá, B. Mussard and T. K. Woo, *J. Chem. Theory Comput.*, 2009, **5**, 2866.
- (20) T. R. Forrester, W. Smith, DL_POLY ver. 2 (C.L.R.C. Daresbury Laboratory, Daresbury, UK, 1995)
- (21) (a) Keskin, S.; van Heest, T. M.; Sholl, D. S. *ChemSusChem* **2010**, *3*, 879. (b) Li, J. R.; Ma, Y.; McCarthy, M. C.; Sculley, J.; Yu, J. M.; Jeong, H. K.; Balbuena, P. B.; Zhou, H. C. *Coord. Chem. Rev.* **2011**, *255*, 1791.
- (22) Yazaydin, A. O.; Snurr, R. Q.; Park, T. H.; Koh, K.; Liu, J.; LeVan, M. D.; Benin, A. I.; Jakubczak, P.; Lanuza, M.; Galloway, D. B.; Low, J. J.; Willis, R. R. *J. Am. Chem. Soc.* **2009**, *131*, 18198.
- (23) Plant, D. F.; Maurin, G.; Deroche, I.; Llewellyn, P. L. *Microporous Mesoporous Mater.* **2007**, *99*, 70.
- (24) Llewellyn, P. L.; Bourrelly, S.; Serre, C.; Vimont, A.; Daturi, M.; Hamon, L.; De Weireld, G.; Chang, J. S.; Hong, D. Y.; Hwang, Y. K.; Jung, S. H.; Ferey, G. *Langmuir* **2008**, *24*, 7245.
- (25) Caskey, S. R.; Wong-Foy, A. G.; Matzger, A. J. *J. Am. Chem. Soc.* **2008**, *130*, 10870
- (26) (a) Lin, J. B.; Zhang, J. P.; Chen, X. M. *J. Am. Chem. Soc.* **2010**, *132*, 6654. (b) Zheng, B. S.; Bai, J. F. F.; Duan, J. G.; Wojtas, L.; Zaworotko, M. J. *J. Am. Chem. Soc.* **2011**, *133*, 748. (c) Lin, J. B.; Lin, R. B.; Cheng, X. N.; Zhang, J. P.; Chen, X. M. *Chem. Commun.* **2011**, *47*, 9185. (d) An, J.; Geib, S. J.; Rosi, N. L. *J. Am. Chem. Soc.* **2010**, *132*, 38. (e) Panda, T.; Pachfule, P.; Chen, Y. F.; Jiang, J. W.; Banerjee, R. *Chem. Commun.* **2011**, *47*, 2011.
- (27) Vaidhyanathan, R.; Iremonger, S. S.; Shimizu, G. K. H.; Boyd, P. G.; Alavi, S.; Woo, T. K. *Angew. Chem.*, in press
- (28) Vaidhyanathan, R.; Iremonger, S. S.; Shimizu, G. K. H.; Boyd, P. G.; Alavi, S.; Woo, T. K. *Science* **2010**, *330*, 650

-
- (29) 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.
- (30) T. Vuong, P. A. Monson, *Langmuir* **1996**, *12*, 5425.

4 Investigation of MIL-47 and its functionalized derivatives on Natural gas purification: A Computational Study

Abstract

In this work, the performance of MIL-47 and its seven functionalized derivatives on natural gas purification with pressure swing adsorption (PSA) technology was investigated using grand canonical Monte Carlo simulations. The computational results show that introducing polar non-bulky functional groups on the solid MOF type MIL-47 leads to an enhancement in its CO₂ working capacity and a significant improvement in the CO₂/CH₄ selectivity. Furthermore, their performance was compared with that of other well-known MOFs, Zeolite 13x, and titanosilicate for use in PSA gas separation systems. The comparison suggests that MIL-47-NO₂ could be a possible candidate as a solid sorbent for natural gas purification.

4.1 Background

As discussed in Chapter 1, the use of solid adsorbents within pressure swing adsorption (PSA) systems is a more energy efficient, and affordable alternative to the traditional amine-based technology for CO₂ scrubbing of natural gas. Molecular Gate™ units based on the titanosilicate adsorbents,¹ UOP MOLSIV systems based on the Zeolite-13x,² and Xebec Adsorption BGX™ systems³ are examples of commercial solid sorbents used in the industrial scale purification of natural gas.

The purity and quantity of CO₂ recovered by PSA-based processes depends on adsorbent used, and conditions in which a gas is trapped and released by the adsorbent, which will ultimately affect the PSA performance. It has been estimated that if a material whose CO₂

selectivity is three times higher than that of Zeolite 13x and CO₂ recovery is two times higher than that of Zeolite 13x is used, the cost of PSA performance could be decreased to half of the current value.⁴ In the past decade, many studies have been performed with the intent of finding materials with high adsorption capacity and high separation ability.^{5–11} Traditional adsorbents, such as zeolites, activated carbons and mesoporous silica, and novel porous materials, such as MOFs, have been widely evaluated for their performance with PSA gas separations. However, the traditional materials have low CO₂ working capacity or they suffer from poor CO₂/CH₄ selectivity, leading to a low efficiency and high energetic costs.¹² In contrast, MOFs have shown great promise for CO₂-related gas applications due to their remarkable properties such as ultra-high surface areas.^{13–15} Figure 4.1 compares the CO₂ working capacity for a set of MOFs with commercial porous materials, including Zeolite 13x and activated carbon Norit RB2. Most of the MOFs, as seen in Figure 4.1, have a CO₂ working capacity higher than that of traditional porous materials. Interestingly, PCN-11 has CO₂ working capacity of 9.1 mmol/g, which is four times higher than that of Zeolite 13x. In addition, it has been demonstrated that introducing functional groups to MOFs can tune the pore characteristics without altering the underlying topology, offering a systematic approach for developing more efficient CO₂ capture materials. Biswas et al.¹⁶ enhanced selectivity of CO₂ over CH₄ 50% in UiO-66 through carboxylate functionalization. Couck et al.¹⁷ increased the CO₂/CH₄ selectivity in MIL-53 by orders of magnitude by integrating amines into the pores. Recently, Lee et al. functionalized Mg-MOF-74 with ethylenediamine, which presented an exceptionally high CO₂ adsorption capacity, 2.83 mmol/g, at 0.39 mbar. Many simulation studies have also predicted a

similar enhancement in the CO₂ adsorption properties of a divers set of MOFs through functionalization.^{18–20}

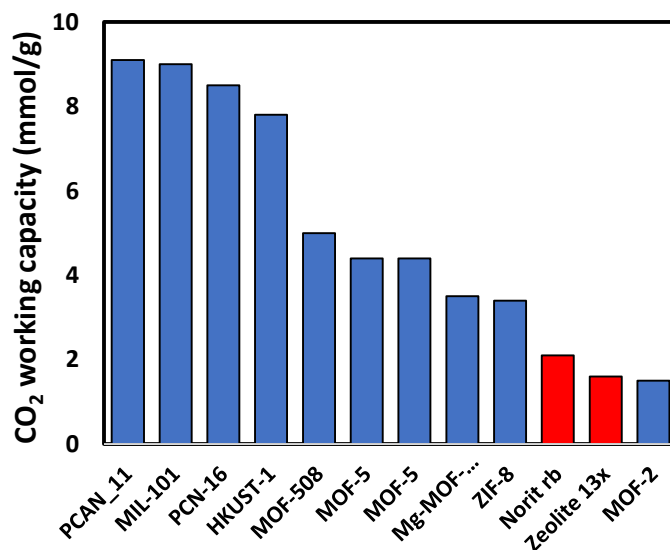


Figure 4.1: Working capacity for CO₂ in MOFs (blue) which is the difference between the amount of CO₂ adsorbed at 6 bar and the amount of CO₂ adsorbed at 1 bar. Non-MOF materials (red) are included for comparison.

On the down side, most of the MOFs suffer from low physicochemical stability when exposed to humidity or subsequent thermal regeneration, which decreases their capacity to adsorb CO₂. Thus, Matzger et al.²¹ studied the performance of Mg-MOF-47 for CO₂ adsorption in the presence of humidity, and they found that its adsorption capacity for CO₂ dropped to 16% of its original capacity after exposure to 70% relative humidity and subsequent thermal regeneration. Walton et al.²² studied the stability of a series of the pillared MOFs under humid conditions, and they found that DMOF and its functionalized derivatives containing NO₂, Br, Cl₂, and OH lose more than 95% of their surface area after exposure to 80% relative humidity, which will significantly impact CO₂ adsorption capacity. Such a reduction in the CO₂ adsorption capacity limits MOFs' applications in industrial settings, as trace amounts of water are present

in many industrial streams. Thus, having sufficient thermal and chemical stability is a necessary criterion in order for MOFs to be implemented in industrial settings.

Of the MOFs that are available in the Cambridge Crystallographic Data Centre (CCDC),²³ the MIL-n family, originally developed by Feyer's group, has attracted much attention for their thermal and chemical stability. Among this family, MIL-47 is thermally stable up to 673.15 K in air, and does not show any significant alteration of its porosity upon the inclusion of water vapour.^{24,25} It has been also demonstrated that this material is easily regenerateable and retains its structure after exposure to H_2S , which contrasts the majority of porous solids reported thus far.²⁶ Furthermore, this material does not exhibit 'breathing' phenomena (i.e., opening/closing pores) during gas sorption process, which makes it ideal for packing in a bed column of fixed volume in a PSA unit. Such unique properties have made MIL-47 a widely studied case for a variety of applications including catalysis, and gas separation and storage.^{27,28}

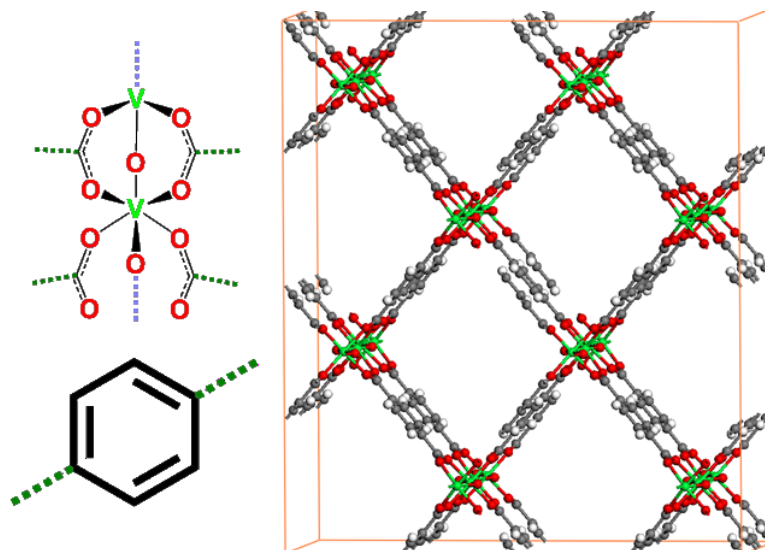


Figure 4.2: Crystal structure of MIL-47 (right) its inorganic structural building unit (top-left) and organic structural building unit (bottom-left)

Llewellyn et al.²⁷ evaluated the MIL-47 performance for CO₂-related applications, and found that among the MOFs chosen in their work, it has the highest CO₂ working capacity for a pressure swing between 1 and 6 bar at 303 K. Voort et al.²⁹ synthesized a series of MIL-47-Xn functionalized solids (Xn = -Cl, -Br, -CH₃, -CF₃, -OH, -OCH) and evaluated their performance for CO₂ capture from flue gas. Among these six functionalized MIL-47 materials, MIL-47-OCH₃ was found to have the highest CO₂ uptake. Many simulation studies have been also conducted to investigate the sorption properties of MIL-47 for a variety of adsorbates including H₂,³⁰ CO₂,³¹ CH₄, N₂,³² alkanes,³³ benzene, and xylenes.³⁴ In most of these studies, pure component isotherms and low coverage heats of adsorption were computed using GCMC simulations, and the adsorption selectivity, in the case of the mixture, was calculated from the computed single isotherms data using the ideal adsorbed solution theory (IAST). For example, in the study of the separation of xylene isomers in MIL-47 by Calero et al.,³⁴ the Henry coefficients and low coverage heats of adsorption were computed at 543 K, which presented excellent agreement with experimental results. A few GCMC simulations have also been used to investigate the sorption properties of functionalized MIL-47 materials. An example of this is the study by Snurr et al.³⁵ who investigated the ammonia adsorption in MIL-47-Xn (Xn = -OH, -C=O, -Cl, -COOH), and found that ammonia adsorption increases significantly with increasing density of functional group. In that study, the crystal structures of the functionalized materials were built by computationally substituting hydrogen atoms within the MIL-47 unit cell with the aforementioned functional groups.

As far as we were able to ascertain, no simulation or experimental studies on natural gas purification with substituted MIL-47 materials, except to the MIL-47-NH₂, have been reported in the literature so far. Herein, we present a computational study on natural gas purification in the MIL-47 and its functionalised derivatives (i.e., MIL-47-X, X = -Cl, -Br, -CH₃, -CHO, -OH, -OCH₃, -COOH, -NH₂, NO₂). First, we computed the adsorption properties, including CO₂/CH₄ selectivity and CO₂ working capacity, using GCMC simulations of the CO₂/CH₄ mixture at PSA conditions for natural gas purification. Using simulation data, we then investigated the effect of functional groups on this application in MIL-47-Xn. Finally, we compared the performance of these materials with that of other materials that are already reported in literature.

4.2 Computational Methods

4.2.1 Optimization and Generation of functionalized MIL- 47 materials

Since the structures of substituted MIL-47 materials we are interested in have not been synthesized before, the structures were built from the known crystal structure of MIL-47 using an in-house program, FAPSWITCH written by former members of the Woo lab. Here, one hydrogen atom of each ligand within the MIL-47 unit-cell was randomly chosen to be replaced with a desired functional group that had been pre-oriented to an energy minimized configuration with respect to a benzene ring using the universal force field (UFF).^{36,37} After the insertion of the pre-oriented functional group, FAPSWITCH checks for unphysical overlaps of the inserted functional group with the base MOF structure. The functional group is rotated about its linking bond by 30° until all the functional group atoms fall outside the van der Waals radius of MIL-47 framework atoms. Once the functional group is successfully added,

the generated structure is geometry optimized at the molecular mechanics level with the UFF force field and the GROMACS package.³⁸ Here, the optimization was performed in three steps. First, all atomic positions within the generated structure were allowed to relax to their energy minima with fixed cell parameters and an L-BFGS optimizer with a cutoff of $10 \text{ kJ}\cdot\text{mol}^{-1}\cdot\text{nm}^{-1}$ for the forces. Then, cell parameters were relaxed for 2000 steps using quenched molecular dynamics (MD). Finally, the L-BFGS optimization was repeated to reduce the forces to less than $10 \text{ kJ}\cdot\text{mol}^{-1}\cdot\text{nm}^{-1}$, which would increase during the MD simulation. Following the classical force field optimization, periodic DFT calculations were performed using the Vienna ab initio simulation package (VASP)^{39,40} to relax the GROMACS-optimized structures. In these calculation, the PBE exchange-correlation functional⁴¹ and PAW potentials^{42,43} were used with a plane wave cut-off of 520 eV. Gamma-point was used to sample the Brillouin zone. A conjugate gradient algorithm with a cutoff of $2\times 10^{-5} \text{ eV}/\text{\AA}$ for forces was used to relax the atom positons within structures. The semi-empirical dispersion potential correction of Grimme⁴⁴ was included in the energy and force calculations with the default parameters specified for the PBE functional.

4.2.2 Force Fields

To describe an interaction between framework atoms and guest molecules, the Lennard-Jones (LJ) and Coulomb potentials were used. LJ parameters between atoms of different types were calculated using the Lorentz-Berthelot mixing rules (i.e., geometric average of well depths and arithmetic average of diameters). LJ parameters for framework atoms, given in Table 4.1, were taken from the Universal Force Field (UFF).³⁶ For Coulombic interactions, partial atomic charges for each structure were calculated using the REPEAT method⁴⁵ with the electrostatic potential derived from the periodic DFT calculations. All partial atomic charges calculated are

given in Table 4.2. The non-bonded parameters of CO₂ were taken from García-Sánchez et al.⁴⁶ 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 CO₂ molecule had a fixed bond length of 1.16 Å and partial charges of +0.645e and -0.3325e for carbon and oxygen. The non-bonded parameters of CH₄ were taken from TraPPE force field,⁴⁷ where CH₄ molecule was modeled as a single sphere. This force field has been fit to reproduce the experimental vapor-liquid phase equilibrium data. The force field parameters for the MOF atom types, CO₂ and CH₄ are given in Table 4.1.

Table 4.1: Lennard-Jones parameters from the UFF.

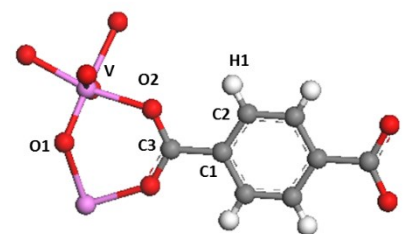
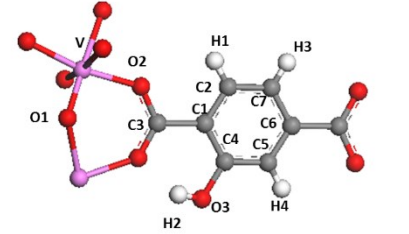
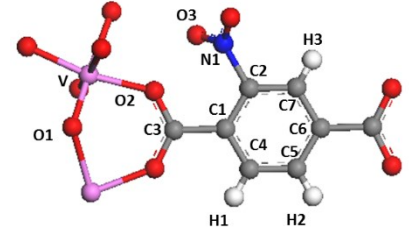
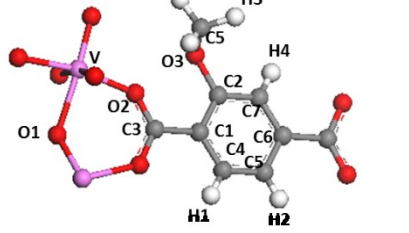
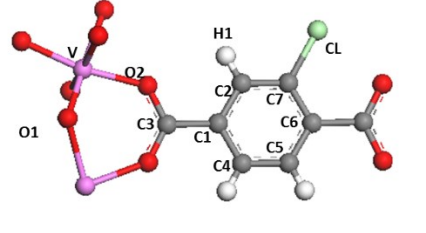
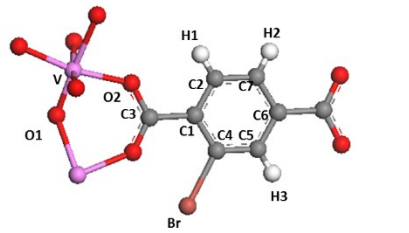
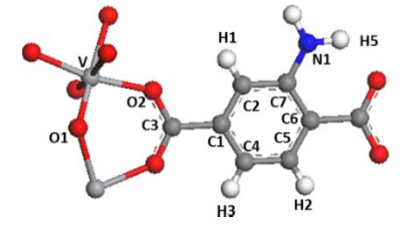
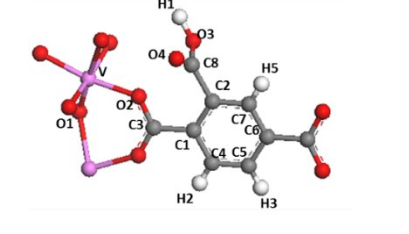
Atomic type	σ / Å	ϵ / kcalmol ⁻¹
V	2.8010	0.0160
C	3.4309	0.105
N	3.2607	0.0690
O	3.1181	0.060
H	2.5711	0.044
Cl	3.5164	0.2270
Br	3.7320	0.2510
CH ₄	3.7300	0.2941
C (CO ₂)	2.7450	0.05948
O (CO ₂)	3.0170	0.17023

4.2.3 GCMC simulations

To calculate the adsorption of pure components and their mixture at natural gas conditionings, Grand canonical Monte Carlo (GCMC) simulations were run using an in-house developed code based on the open-source DL_POLY classic package formerly known as DL_POLY 2. The addition, deletion and displacement with equal probabilities were included in GCMC moves. In the case of mixture isotherms, there was also a swapping move that allowed a molecule of one type to be substituted with one molecule of another type. Each system was

equilibrated for 10 million MC steps, followed for further 12.5 million steps, over which the desired properties were averaged. The constant chemical potential was calculated using the ideal-gas equation of state for pressures up to 1 bar, while for higher pressures the Peng-Robinson equation of state was used. The geometry of the framework was frozen during simulations. A 2x4x4 supercell was generated from the unit cell which enabled use of a 12.5 Å cutoff for the Lennard-Jones interactions. Electrostatic interactions were calculated using the Ewald summation method.

Table 4.2: REPEAT partial atomic charges (q/e) for the X-ray structure of (1) MIL-47, (2) MIL-47-OH, (3) MIL-47-NH₂, (4) MIL_47-OME, (5) MIL-47-CL, (6) MIL-47-Br, (7) MIL-NH₂, and (8) MIL_47-COOH

 V: 1.34, C1: 0.064, C2: -0.157, C3: 0.546, O1: -0.646, O2: -0.564, H1: 0.124,	 V: 1.43, C1: -0.26, C2: 0.44, C3: 0.646, C4: 0.44, C5: -0.29, C6: -0.09, C7: -0.29, O1: -0.54, O2: -0.60, O3: -0.54, H4: 0.19, H3: 0.14, H1: 0.12, H2: 0.21
 V: 1.38, C1: 0.09, C2: -0.13, C3: 0.27, C4: 0.44, C5: -0.17, C6: -0.006, C7: -0.20, O1: -0.57, O2: -0.40, O3: -0.38, N1: 0.72, H3: 0.12, H1: 0.14, H2: 0.21	 V: 1.32, C1: -0.20, C2: 0.33, C3: 0.54, C4: -0.009, C5: -0.29, C6: -0.09, C7: -0.14, O1: -0.52, O2: -0.47, O3: -0.26, H4: 0.19, H3: 0.057, H1: 0.12, H2: 0.21, C5: -0.088
 V: 1.20, C1: 0.07, C2: -0.15, C3: 0.56, C4: -0.009, C5: -0.29, C6: -0.09, C7: 0.26, O1: -0.47, O2: -0.52, O3: -0.26, H4: 0.19, H3: 0.057, H1: 0.18, CL: -0.21	 V: 1.37, C1: 0.11, C2: -0.06, C3: 0.44, C4: -0.08, C5: -0.06, C6: -0.03, C7: -0.22, O1: -0.12, O2: -0.30, H3: 0.098, H1: 0.08, Br: -0.21
 V: 1.36, C1: 0.09, C2: -0.35, C3: 0.53, C4: -0.29, C5: -0.13, C6: -0.23, C7: -0.44, O1: -0.57, O2: -0.51, O3: -0.53, N1: -0.69, H3: 0.12, H1: 0.14, H2: 0.21, H5: 0.30	 V: 1.39, C1: -0.09, C2: 0.07, C3: 0.63, C4: -0.07, C5: -0.11, C6: -0.01, C7: -0.17, O1: -0.54, O2: -0.52, O3: -0.48, O4: -0.45, H1: 0.38, H2: 0.12, H3: 0.12, H5: 0.13

4.3 Result and Discussion

4.3.1 Adsorption of pure components

To validate the forced fields used in our simulations, we simulated adsorption isotherms of pure CO_2 and CH_4 in MIL-47 as well as the adsorption isotherm of CO_2 in MIL-47- OCH_3 , and compared them with reported experimental data. These two MOFs were chosen, as no experimental data of CO_2 and CH_4 adsorption at high pressure are reported for other MOFs studied here. Figure 4.3 compares the simulated isotherms of pure CH_4 and CO_2 with experimentally measured adsorption for MIL-47 and MIL-47- OCH_3 at 303 K up to 20 bar. The simulation results show fairly good agreement with the experimental isotherms, albeit there is an overestimation in simulated isotherms at low pressures. Overestimation in simulated isotherm occurs for two reasons. First, the experimental sample usually contains impurities, which block the channels and reduce the adsorption capacity. In other words, the real samples can never be perfectly purged before the uptake experiments are performed. Second, the calculated adsorption isotherm in GCMC simulations are obtained with the consideration as a perfect crystal for the adsorbent while in experiments the sample may contain crystal defects which reduce the adsorption capacity. Although we could adjust the Lennard-Jones parameters of the guest molecules to match the experimental isotherms, since we are interested in selectivity (i.e., ratios of the uptake), and working capacity (i.e., differences of the uptakes), we have decided not to adjust the parameters.

Figure 4.3a also shows that indeed MIL-47 adsorbs more CO_2 than CH_4 over the entire pressure range tested. There are two reasons that can be used to explain this. First, CO_2 has

stronger dispersion and electrostatic interactions with the framework than CH_4 . In terms of dispersion, CO_2 has more electrons than CH_4 , resulting in stronger dispersion interactions. With respect to the electrostatic interactions, CO_2 has a relatively large molecular quadrupole moment dipole, whereas CH_4 has neither a quadrupole nor a dipole moment (only a weak octapole moment). Second, the considered temperature of 303 K is supercritical for CH_4 ($T_c = 190.5$), but subcritical for CO_2 ($T_c = 304.4$ K); that is, CO_2 is more condensable than CH_4 at 303 K.

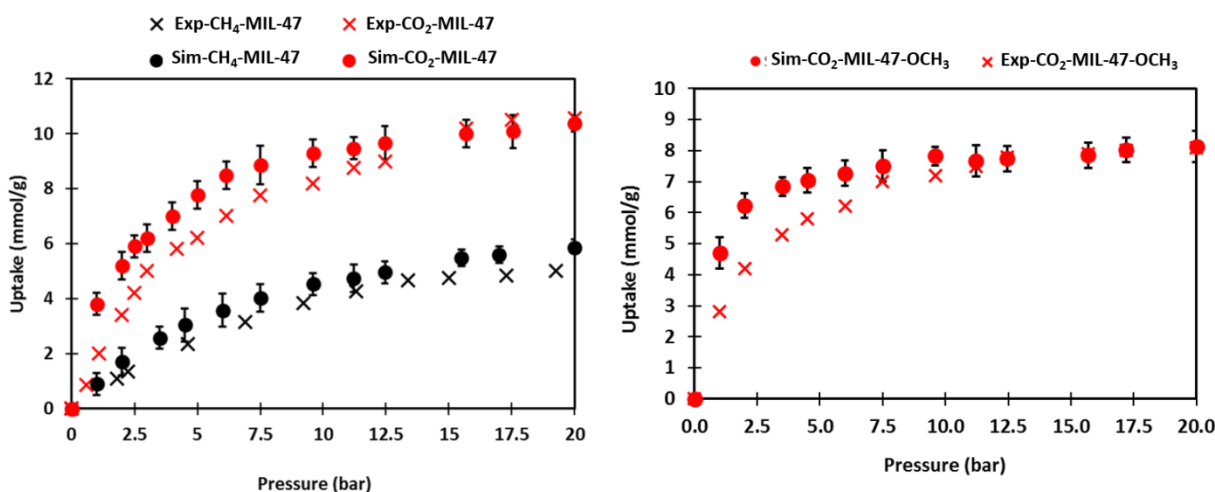


Figure 4.3: a) Adsorption isotherms of CO_2 (red circle: simulation, red cross: experiment) and CH_4 (black circle; simulation, black cross: experiment) in MIL-47. b) Adsorption isotherm of CO_2 (red circle: simulation, red cross: experiment) in MIL-47- OCH_3 .

4.3.2 Adsorption of Binary Mixtures

To investigate the performance of MIL-47 and its functionalized derivatives for natural gas purification under PSA conditions, we simulated the adsorption in a binary mixture of CO_2 and CH_4 with a composition of 1:9 at 303 K and total pressures up to 30 bar. Figure 4.4 shows the CO_2 adsorptions obtained from GCMC simulations of a 1:9 CO_2 - CH_4 mixture on MIL-47 and its functionalized derivatives as a function of partial pressure of CO_2 . The CO_2 uptakes in the -OH, - NO_2 , - NH_2 , -COH, functionalized derivatives are higher than those in MIL-47. Particularly, at

pressures below 1 bar, the adsorptions for MIL-47-CHO are three times as high as those in MIL-47. Such enhancement in uptake can be attributed to the highly polar nature of these functional groups. However, although the polarity of the -COOH is strong, the CO₂ uptakes decreases in MIL-47 COOH due to its very small accessible pore volume compared to the other polar functionalized derivatives. The CO₂ uptake also increases in MIL-47-Cl while for the other three MOF derivatives, the CO₂ uptakes are close to those in MIL-47.

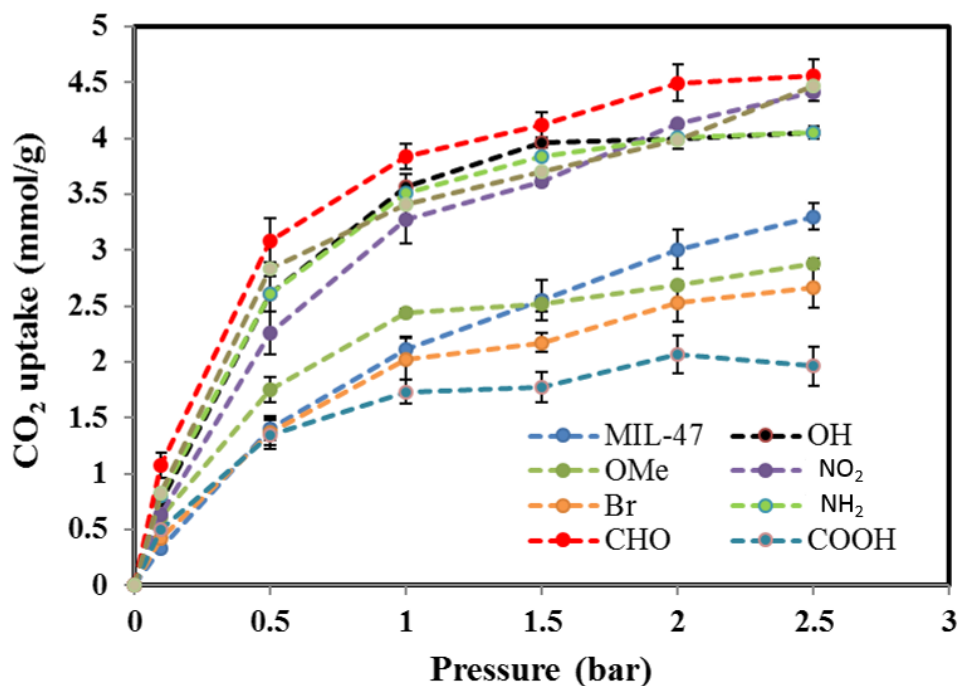


Figure 4.4: CO₂ adsorption isotherms from GCMC simulations of a 1:9 CO₂-CH₄ mixture on MIL-47 (light blue), MIL-47-OMe (light gray), MIL-47-Br (dark green), MIL-47-CHO (red), MIL-47-Cl (dark gray), MIL-47-COOH (dark blue), MIL-47-NO₂ (brown), MIL-47-OH (black), and MIL-47-NH₂ (light green) at 303 K.

Figure 4.5 compares the CO₂ working capacity in MIL-47 and its functionalized derivatives obtained from GCMC simulations of CO₂-CH₄ mixture with a pressure swing of 10 bar and 1 bar (total pressure). The working capacities follow a trend similar to that seen for absolute CO₂ uptake capacities given in Figure 4.4. In the cases of MIL-47-OH, MIL-47-NO₂, MIL-47-NO₂, MIL-

47-CHO, and MIL-47-Cl, the working capacity is higher than that of MIL-47 by nearly 1 mmol/g. For MIL-47-OMe and MIL_47-Br the working capacity is close to that in the parent MIL-47 MOF while MIL-47-COOH has the lowest working capacity.

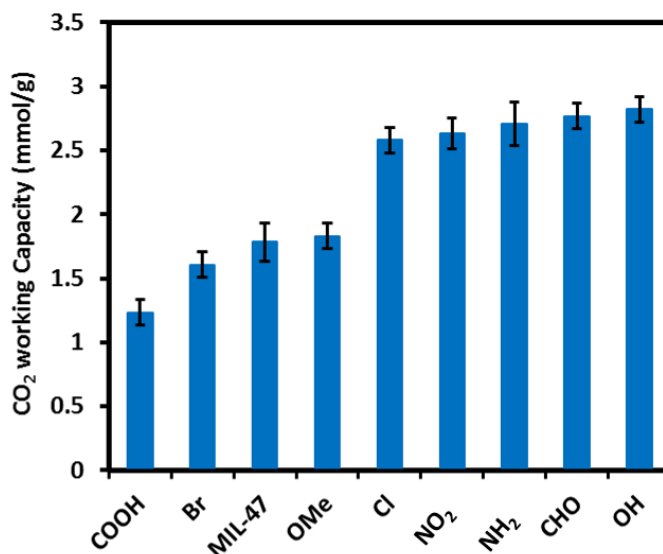


Figure 4.5: CO₂ working capacity obtained from GCMC simulations of CO₂-CH₄ mixture on MIL-47 and its functionalized derivatives using a pressure swing of 10 to 1 bar

Figure 4.6 shows the computed CO₂/CH₄ selectivities in MIL-47 and its functionalized derivatives. Here, the selectivity is defined as $S = (x_A/x_B)(y_B/y_A)$, where x_A and x_B are the amounts of components A and B in the adsorbed phase, and y_A and y_B are the partial pressures of components A and B in the bulk phase, respectively. All functionalized MIL-47 derivatives, have selectivities higher than those of the parent MOF, MIL-47, across the entire pressure range studied. MIL-47-NO₂-and MIL-47-OH have the highest selectivities. For example, at 10 bar MIL-47-NO₂ has selectivity that is three times as high as that of MIL-47. Such enhancement in selectivity can be attributed to the permanent dipole moments of these functional groups, which are more attractive to CO₂ (large quadrupole moment, 13.4 C·m²) than CH₄ (nonpolar),

and consequently resulting in stronger interactions of these functional groups with CO₂ compared to CH₄.

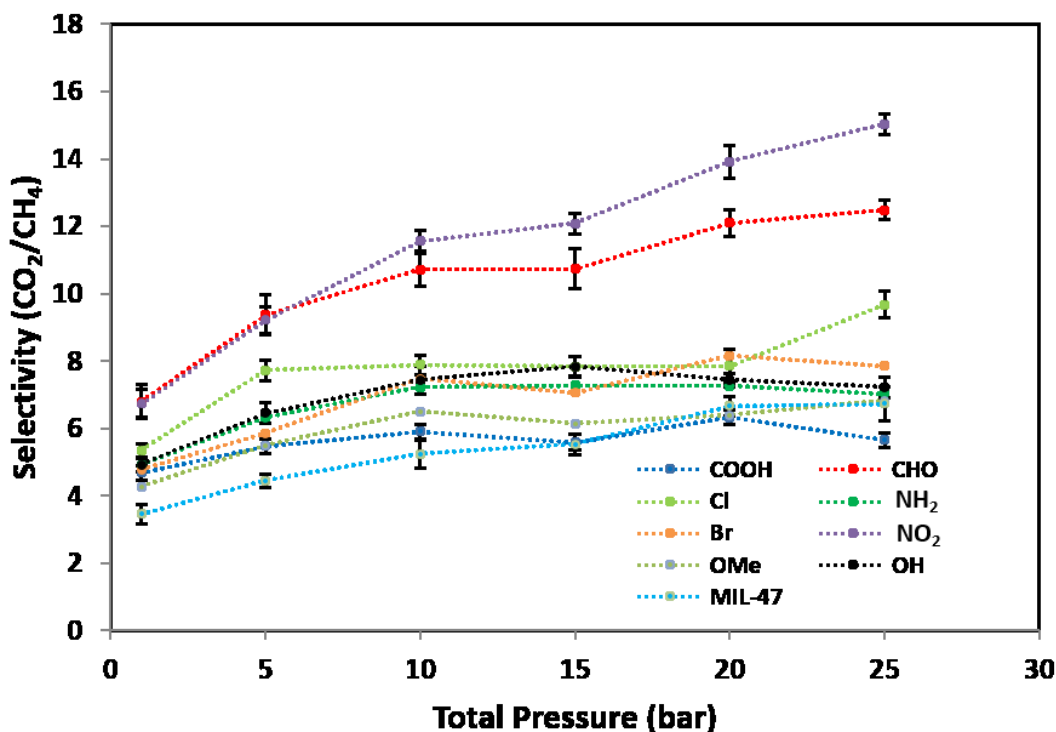


Figure 4.6: Simulated selectivities for CO₂ over CH₄ from their 1:9 gas mixture at 303 K in MIL-47 and its functionalized derivatives, as a function of the total pressure.

To provide further investigation, we computed the isosteric heat of adsorption for both CO₂ and CH₄ in MIL-47 and its derivatives from the GCMC simulations via the Green–Kubo fluctuation theory expression,

$$q_{isotheric} = \frac{\langle UN \rangle - \langle U \rangle \langle N \rangle}{\langle N^2 \rangle - \langle N \rangle \langle N \rangle} + RT \quad 4.1$$

where U is the potential energy of the system, N is the number of adsorbed molecules, T is temperature and R is the ideal gas constant. The calculated isosteric heat of adsorptions at zero loading, which indicates the highest energy sites, are given in Figure 4.7. As expected, the

difference in heat of adsorption between CO₂ and CH₄ for MIL-47-NO₂ and MIL-47-COH are larger than that for other functionalized MOFs. Furthermore, it is found from Figure 4.7 that the CH₄ heats of adsorption are moderately affected by the functionalization.

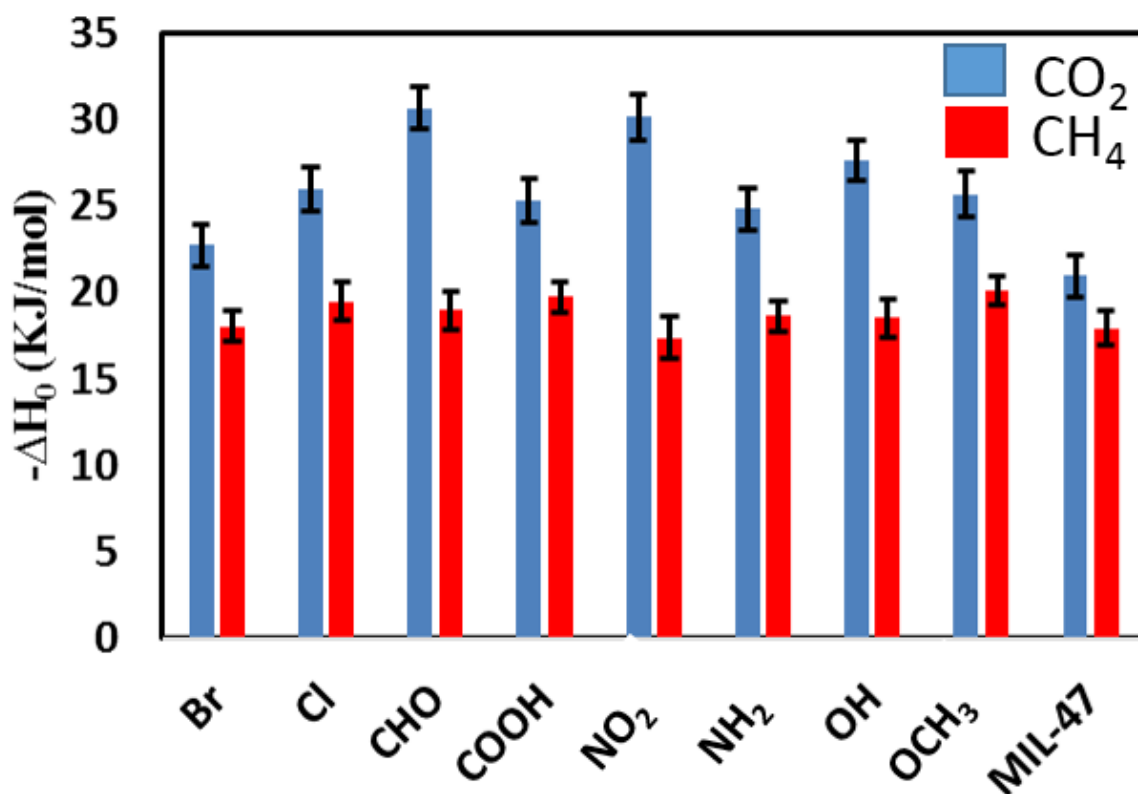


Figure 4.7: Calculated isosteric heats of adsorption at zero loading of pure CO₂ (blue bar) and CH₄ (red bar) gases in MIL-47 and its functionalized derivatives at 303 K.

To provide molecular insights on the adsorption of the CO₂-CH₄ mixture in MIL-47 and the top-performer, we characterized CO₂ binding sites within MIL-47 and MIL-47-NO₂ using probability distributions derived from the GCMC simulations of the CO₂-CH₄ mixture at total pressure of 10 bar in these two MOFs. More specifically, the probability distributions of CO₂ were obtained by storing CO₂ positions from every Monte Carlo production steps on the 3D grid with a resolution of 0.1 Å in each direction³. This histogram consisted of a 3D grid of millions of

points and was obtained separately for each atom of CO_2 . From these noisy probability distributions, CO_2 binding sites in each MOF were found using an in-house program, ABSL (Automatic Binding Site Locator), which smooths the noisy 3D histograms via a Gaussian filter, and isolates the maxima of probability.

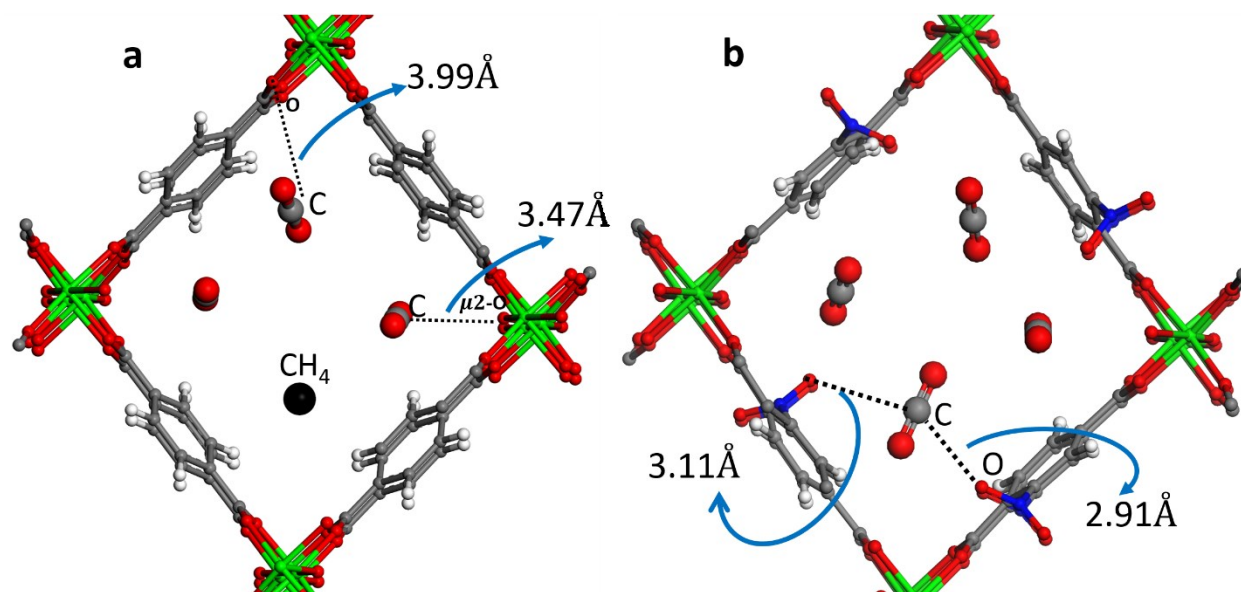


Figure 4.8: a) three CO_2 binding sites and one CH_4 binding site (black circle) are obtained from the GCMC simulation of a CO_2 - CH_4 mixture at a total pressure of 10 bar in a single pore of the MIL-47 where the carboxyl O and the $\mu 2$ -O atoms interact with CO_2 molecules. b) four CO_2 binding sites are obtained from the GCMC simulation of a CO_2 - CH_4 mixture at a total pressure of 10 bar in a single pore of the MIL-47- NO_2 where O atoms from NO_2 groups interact with CO_2 molecules. Green: V, Red: O, Gray: C, and Blue: N

Figure 4.8 shows the CO_2 binding sites, obtained from the GCMC simulations, in a single pore of MIL-47 (a) and MIL-47- NO_2 (b). While CO_2 binding sites within a MIL-47 appear nearest to either the carboxyl O or the $\mu 2$ -O atoms, for MIL-47- NO_2 CO_2 binding sites are located rather close to the NO_2 functional group. The binding energy of CO_2 in MIL-47- NO_2 was calculated to be -32.5 KJ/mol on average, while for MIL-47 the CO_2 binding energy was calculated to be -25.6 KJ/mol on average. The difference in the binding energy is attributed to the interaction of NO_2

functional group with CO₂, which is stronger than the interaction of the carboxyl O or the μ 2-O atoms with CO₂ within MIL-47.

4.3.3 Comparison of performance of functionalized MIL-47 materials with other materials

To compare the CO₂/CH₄ selectivity and CO₂ working capacity of the MOFs studied here with other materials, we collected the data from literature that reported the selectivity and working capacity of other materials at conditions relevant to natural gas purification. Figure 4.9 shows CO₂ working capacity and CO₂/CH₄ selectivity of three functionalized MIL-47 materials, UIO-66,¹⁶ UIO-66-COOH,¹⁶ UIO-66-SO₃H,¹⁶ IR-MOF-74,⁴⁸ mixed-ligand MOF,⁴⁹ MIL-53,⁵⁰ Carborane-based MOF,⁵¹ MOF-508b, Cu-BTC,⁵² PCN-16,⁵³ PCN-11,⁵³ MOF-177,⁵⁴ MOF-5,⁵³ UMCM-1,⁵⁵ ZIF-8,⁵³ and two non-MOF materials including Zeolite 13X and titanosilicate.^{8,11}

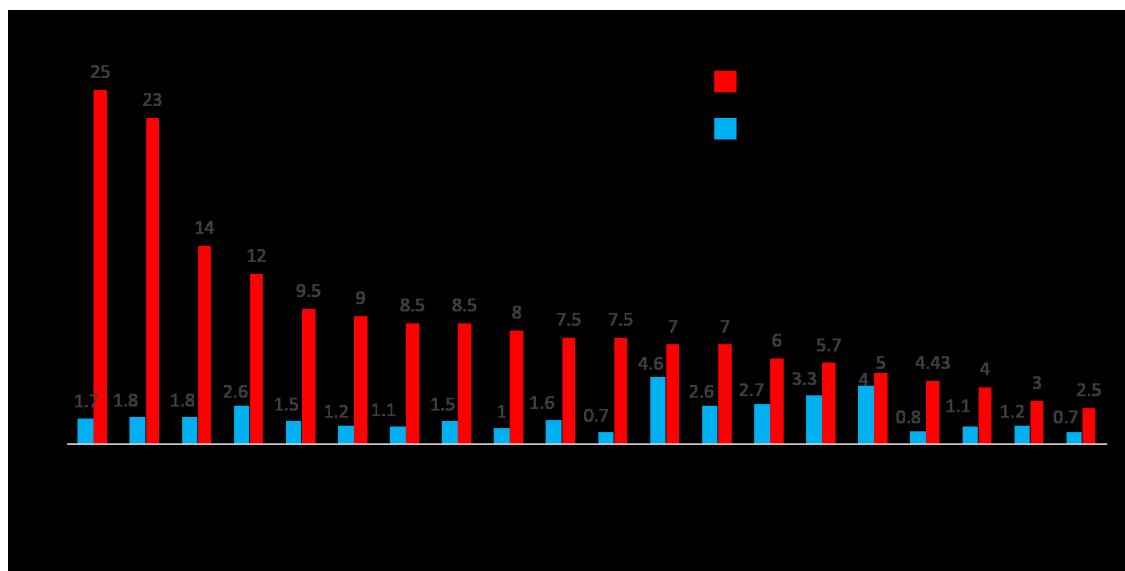


Figure 4.9: CO₂/CH₄ selectivities and CO₂ working capacity for a 1:9 CO₂-CH₄ mixture at conditions relevant to natural gas purification in a variety of MOFs and two non-MOF materials. The chosen conditions correspond to an adsorption pressure of 10 bar, a desorption pressure of 1 bar, and temperature of 300 K.

In regards to working capacity, functionalized MIL-47 materials show relatively high working capacity in comparison to other MOFs and non-MOF materials except for Cu-BTC (world record), PCN-16 and PCN-11. Among the functionalized MIL-47 materials, MIL-47-NO₂ has a selectivity higher than Cu-BTC by nearly 40%, which is very close to that of Zeolite 13X. Considering both selectivity and working capacity, MIL-47-NO₂ may be considered a candidate for use in a PSA unit, as it offers a good combination of selectivity and working capacity.

4.4 Conclusion

In this work, a computational study was performed to evaluate the performance of MIL-47 and its functionalization derivatives on methane purification under PSA conditions. CO₂ working capacity and CO₂/CH₄ selectivity were computed using GCMC simulations for each MOF studied here. The results from the GCMC simulations show that introducing polar non-bulky functional groups within MIL-47 leads to an enhancement in its CO₂ working capacity and a significant improvement in the CO₂/CH₄ selectivity. Particularly, MIL-47 functionalized with NO₂ offers the highest selectivity, which is two times as high as that of unfunctionalized MIL-47. In comparison with other materials whose CO₂ working capacity and CO₂/CH₄ selectivity are reported in literature, MIL-47-NO₂ shows a relatively good combination of CO₂ working capacity and CO₂/CH₄ selectivity, which suggests that this material could be a possible candidate as a solid adsorbent in natural gas purification with PSA technologies.

4.5 References

- (1) Wills, J., Mitariten, M. 2009. Production of Pipeline Quality Natural Gas With the Molecular Gate™ CO₂ Removal Process. Tidelands Oil Production Company & Guild Associates.

- (2) UOP Overview of Gas Processing Technologies and Applications, 2010. 22 June 2011, <[http://www.uop.com/wp-content/uploads/2011/02/UOP-Overview of-Gas-Processing-Technologies-and-Applications-tech-presentation.pdf](http://www.uop.com/wp-content/uploads/2011/02/UOP-Overview-of-Gas-Processing-Technologies-and-Applications-tech-presentation.pdf)>.
- (3) <http://www.xebecinc.com/about-profile.php>
- (4) Ho, M. T.; Allinson, G. W.; Wiley, D. E. *Ind. Eng. Chem. Res.* **2008**, *47*, 4883.
- (5) Mason, J. a.; Sumida, K.; Herm, Z. R.; Krishna, R.; Long, J. R. *Energy Environ. Sci.* **2011**, *4*, 3030.
- (6) Bae, Y. S.; Snurr, R. Q. *Angew. Chemie - Int. Ed.* **2011**, *50* (49), 11586.
- (7) Wang, J.; Huang, L.; Yang, R.; Zhang, Z.; Wu, J.; Gao, Y.; Wang, Q.; O'Hare, D.; Zhong, Z. *Energy Environ. Sci.* **2014**, *7* (11), 3478.
- (8) Primera-Pedrozo, J. N.; Torres-Cosme, B. D.; Clardy, M. E.; Rivera-Ramos, M. E.; Hernández-Maldonado, A. J. *Ind. Eng. Chem. Res.* **2010**, *49* (5), 7515.
- (9) Bellussi, G.; Broccia, P.; Carati, A.; Millini, R.; Pollesel, P.; Rizzo, C.; Tagliabue, M. *Microporous Mesoporous Mater.* **2011**, *146* (1-3), 134.
- (10) Bao, Z.; Yu, L.; Dou, T.; Gong, Y.; Zhang, Q.; Ren, Q.; Lu, X.; Deng, S. **2011**, 4017. Journal missing
- (11) Cavenati, S.; Grande, C. a; Rodrigues, a E. *J. Chem. Eng. Data* **2004**, *49*, 1095.
- (12) Tagliabue, M.; Farrusseng, D.; Valencia, S.; Aguado, S.; Ravon, U.; Rizzo, C.; Corma, A.; Mirodatos, C. *Chem. Eng. J.* **2009**, *155*, 553.
- (13) Millward, A. R.; Yaghi, O. M. *J. Am. Chem. Soc.* **2005**, *127* (51), 17998.
- (14) Deng, H.; Doonan, C. J.; Furukawa, H.; Ferreira, R. B.; Towne, J.; Knobler, C. B.; Wang, B.; Yaghi, O. M. *Science* **2006**, 327.
- (15) Britt, D.; Furukawa, H.; Wang, B.; Glover, T. G.; Yaghi, O. M. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (49), 20637.
- (16) Biswas, S.; Zhang, J.; Li, Z.; Liu, Y.-Y.; Grzywa, M.; Sun, L.; Volkmer, D.; Van Der Voort, P. *Dalton Trans.* **2013**, *42*, 4730.
- (17) Leus, K.; Couck, S.; Vandichel, M.; Vanhaelewyn, G.; Liu, Y.-Y.; Marin, G. B.; Driessche, I. Van; Depla, D.; Waroquier, M.; Speybroeck, V. Van; Denayer, J. F. M.; Voort, P. Van Der. *Phys. Chem. Chem. Phys.* **2012**, *14*, 15562.
- (18) Liu, Y.; Ye, Q.; Shen, M.; Shi, J.; Chen, J.; Pan, H.; Shi, Y. *Carbon dioxide capture by functionalized solid amine sorbents with simulated flue gas conditions*; 2011; Vol. 45.
- (19) Yang, Q.; Wiersum, A. D.; Llewellyn, P. L.; Guillerm, V.; Serre, C.; Maurin, G. *Chem. Commun. (Camb)*. **2011**, *47*, 9603.

- (20) Mu, W.; Liu, D.; Yang, Q.; Zhong, C. *Microporous Mesoporous Mater.* **2010**, *130* (1-3), 76.
- (21) Cychosz, K. a.; Wong-Foy, A. G.; Matzger, A. J. *J. Am. Chem. Soc.* **2008**, *130* (22), 6938.
- (22) Jasuja, H.; Huang, Y. G.; Walton, K. S. *Langmuir* **2012**, *28*, 16874.
- (23) Allen, F. H. *Acta Crystallogr. Sect. B Struct. Sci.* **2002**, *58* (3 PART 1), 380.
- (24) Riou, D.; Barthelet, K. *J. Am. Chem. Soc.* **2002**, *2* (2), 281.
- (25) Salles, F.; Bourrelly, S.; Jobic, H.; Devic, T.; Guillermin, V.; Llewellyn, P.; Serre, C.; Ferey, G.; Maurin, G. *J. Phys. Chem. C* **2011**, *115*, 10764.
- (26) Hamon, L.; Serre, C.; Devic, T.; Loiseau, T.; Millange, F.; Frey, G.; De Weireld, G. *J. Am. Chem. Soc.* **2009**, *131* (25), 8775.
- (27) Llewellyn, P. L.; Bourrelly, S.; Vagner, C.; Heymans, N.; Leclerc, H.; Ghoufi, A.; Bazin, P.; Vimont, A.; Daturi, M.; Devic, T.; Serre, C.; De Weireld, G.; Maurin, G. *J. Phys. Chem. C* **2013**, *117* (V), 962.
- (28) Zornoza, B.; Martinez-Joaristi, A.; Serra-Crespo, P.; Tellez, C.; Coronas, J.; Gascon, J.; Kapteijn, F. *Chem. Commun. (Camb)*. **2011**, *47* (33), 9522.
- (29) Biswas, S.; Vanpoucke, D. E. P.; Verstraelen, T.; Vandichel, M.; Couck, S.; Leus, K.; Liu, Y.-Y.; Waroquier, M.; Speybroeck, V. Van; Denayer, J. F. M.; Voort, P. Van Der. *J. Phys. Chem. C* **2013**, *117*, 22784.
- (30) Salles, F.; Jobic, H.; Maurin, G.; Koza, M. M.; Llewellyn, P. L.; Devic, T.; Serre, C.; Ferey, G. *Phys. Rev. Lett.* **2008**, *100* (24), 1.
- (31) Ramsahye, N. a.; Maurin, G.; Bourrelly, S.; Llewellyn, P. L.; Devic, T.; Serre, C.; Loiseau, T.; Ferey, G. *Adsorption* **2007**, *13*, 461.
- (32) Liu, B.; Smit, B. *Langmuir* **2009**, *25* (19), 5918.
- (33) D. Roche, I.; Rives, S.; Trung, T.; Yang, Q.; Ghoufi, A.; Ramsahye, N. A.; Trens, P.; Fajula, F.; Devic, T.; Serre, C.; Frey, G.; Jobic, H.; Maurin, G. *J. Phys. Chem. C* **2011**, *115* (28), 13868.
- (34) Castillo, J. M.; Vlugt, T. J. H.; Calero, S. *J. Phys. Chem. C* **2009**, *113* (49), 20869.
- (35) Yu, D.; Ghosh, P.; Snurr, R. Q. *Dalt. Trans.* **2012**, *41* (14), 3962.
- (36) Rappe, a. K.; Casewit, C. J.; Colwell, K. S.; Goddard III, W. a.; Skiff, W. M. *J. Am. Chem. Soc.* **1992**, *114*, 10024.
- (37) Boyle, N. M. O.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. **2011**, 1.
- (38) Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; Van Der Spoel, D.; Hess, B.; Lindahl, E. *Bioinformatics* **2013**, *29* (7), 845.

- (39) Kresse, G.; Furthmüller, J. *Comput. Mater. Sci.* **1996**, 6, 15.
- (40) Kresse, G.; Furthmüller, J. *Phys. Rev. B* **1996**, 54 (16), 11169.
- (41) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, 77 (3), 3865.
- (42) Kresse, G. *Phys. Rev. B* **1999**, 59 (3), 1758.
- (43) Blöchl, P. E. *Phys. Rev. B* **1994**, 50 (24), 17953.
- (44) Steffen, C.; Thomas, K.; Huniar, U.; Hellweg, A.; Rubner, O.; Schroer, A. *J. Comput. Chem.* **2010**, 31, 2967.
- (45) Campañá, C.; Mussard, B.; Woo, T. K. *J. Chem. Theory Comput.* **2009**, 5, 2866.
- (46) 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 (3), 8814.
- (47) Martin, M. G.; Siepmann, J. I. *J. Phys. Chem. B* **1998**, 102 (97), 2569.
- (48) Babarao, R.; Jiang, J.; Sandler, S. I. *Langmuir* **2009**, 25 (9), 5239.
- (49) Bae, Y.; Mulfort, K. L.; Frost, H.; Ryan, P.; Punnathanam, S.; Broadbelt, L. J.; Hupp, J. T.; Snurr, R. Q. *Langmuir* **2008**, 24 (18), 8592.
- (50) Boutin, A.; Coudert, F.-X.; Huet, M.-A. S.; Neimark, A. V; Férey, G.; Fuchs, A. H. *J. Phys. Chem. C* **2010**, 114, 22237.
- (51) Bae, Y.-S.; Farha, O. K.; Spokoyny, A. M.; Mirkin, C. a; Hupp, J. T.; Snurr, R. Q. *Chem. Commun. (Camb)*. **2008**, 0 (35), 4135.
- (52) Hamon, I Elsa Jolimaî ^tre, and G. D. P. *Ind. Eng. Chem. Res.* **2010**, 49, 7497.
- (53) Simmons, J. M.; Wu, H.; Zhou, W.; Yildirim, T. *Energy Environ. Sci.* **2011**, 4, 2177.
- (54) Saha, D.; Bao, Z.; Jia, F.; Deng, S. *Environ. Sci. Technol.* **2010**, 44 (5), 1820.
- (55) Mu, B.; Schoenecker, P. M.; Walton, K. S. *J. Phys. Chem. C* **2010**, 114, 6464.

5 Water Adsorption in Metal Organic Frameworks and the Effect of Water on Natural Gas Purification in MOFs: A Computational Study

Abstract

In this work, we examined the accuracy of six “standard” water force fields (SPC/E, TIP3P, TIP4P, TIP4P/Ew, TIP5P, and TIP5P/Ew) for reproducing the experimental water adsorption isotherms on both hydrophobic and hydrophilic MOFs. For hydrophobic MOFs, GCMC simulations reveal that the predicted saturation pressure is quite different for all six models, and the TIP4P model performs better than other models in reproducing experimental saturation pressure by 50% on average. For hydrophilic MOFs, simulations show that TIP4P-Ew water model gives the best agreement with experiment compared to all other water models. The simulations also show how water adsorption in hydrophobic and hydrophilic MOFs is different. Hydrophilic MOFs gradually uptake water, whereas hydrophobic MOFs show little uptake until the saturation pressure where there is rapid uptake. Simulations were also performed on CO₂/CH₄ mixtures in both dry and humid conditions to examine the effect of water on natural gas purification in hydrophobic and hydrophilic MOFs. It is found that water has a minimal effect on the adsorption properties of hydrophobic MOFs for natural gas purification where the humidity levels meet pipeline transport specifications.

5.1 Background

Metal-Organic Frameworks (MOFs) are a novel class of nanoporous materials that have gained intense interest because of their exceptionally high porosities and chemically tunable structures.¹⁻⁵ As a result, they have come under intense investigations for gas storage and

separation applications.^{6–10} For example, MOF-905 has a volumetric working capacity (desorption at 5 bar) of $203 \text{ cm}^3 \text{ cm}^{-3}$ at 80 bar and 298 K, which is the highest value reported among all methane storage materials.¹¹ Mg-MOF-74 which has a CO_2 uptake capacity of $8.81 \text{ mmol} \cdot \text{g}^{-1}$ at 1 bar and 298 K, is three times larger than that of zeolite 13X, a material used industrially for CO_2 scrubbing of natural gas.¹² It is important to realize that these record adsorption properties were measured in dry conditions, whereas water vapour is present in virtually all industrial gas streams. For example, natural gas extracted from the shale gas is typically saturated with water vapor.¹³ The water content in industrial combustion flue gases can reach 10% by volume. Thus, water cannot be ignored when considering real world applications of nanoporous materials. For example, although Mg-MOF-74 possesses record CO_2 uptake capacities under dry conditions, the material irreversibly degrades even with trace amounts of water. Obviously, prior to considering any new material for use in an industrial scale process, it is critical to examine the material's behavior under the applicable humid conditions.

Molecular simulation has developed into a powerful tool to study the gas adsorption in MOFs.^{14–16} Molecular simulation can be used to screen the properties of MOFs, to provide molecular level insight into the adsorption processes, and to identify the nature of the guest-host interactions. Although many simulation studies have been done on MOFs for the prediction of gas adsorption properties, few of them have taken into account the effect of water in MOFs. Greathouse and Allendorf,¹⁷ who performed the first simulation study of water in MOFs, investigated the interaction of water with MOF-5 using molecular dynamics simulations. Although they described the mechanism of degradation of MOF-5 in the presence of water in an accurate fashion, they did not reproduce water adsorption isotherms for MOF-5,

as the experimental water adsorption isotherms were unavailable. Calero et al.¹⁸ employed Grand Canonical Monte Carlo (GCMC) simulations to understand the water adsorption in a Cu-BTC MOF. They choose the TIP5P-Ew model for water and scaled-up the partial atomic charges of framework atoms to reproduce experimental adsorption water isotherm. In another work, Paranthaman et al.¹⁹ studied the water adsorption in the hydrophobic Al(OH)(1,4-ndc) MOF using GCMC simulations, but they choose TIP4P/Ew model for water and scaled-down the partial atomic charges of the framework atoms. More specifically, they scaled the partial atomic charges by 80, 60, 40, 30 and 20%, and found that the best agreement between the simulated and experimental water adsorption isotherm occurred when the charges were scaled down by 30%. In a similar work, Salles et al.²⁰ studied the water adsorption in the MIL-53 using GCMC simulation with TIP4P/Ew model for water. They used the scaled-down approach for the partial charges on the framework atoms to reproduce experimental isotherms. Kapteijn and coworkers.²¹ used three distinct water models, including TIP5P/Ew, TIP4P/Ew and SPC/E, to investigate the water-water interactions in mesoporous MIL-101 without scaling the charges. They found that the simulated isotherm using the SPC/E model was in good agreement with experimental isotherm. Recently, Snurr and co-workers studied the adsorption of water-ammonia mixture in three hydrophobic MOFs using GCMS simulations with TIP4P and TIP4P/2005.²² They found that the TIP4P model reproduced the experimental adsorption isotherms the best. These studies were performed in a somewhat *ad hoc* manner to try and replicate adsorption data for specific porous materials. However, no comprehensive study has been performed to investigate the performance of “standard” water models for water adsorption in a wide range of MOFs.

There are a large number of simple, ‘standard’ water models that can be used in standard molecular mechanics packages. they are typically parameterized to reproduce bulk water properties,^{23–26} such as the density, or heat of vaporization. However, water molecules confined in nanoporous materials have different properties compared to the bulk. For example, the dipole moment of water is much smaller in the confined liquid compared to in bulk water.^{27,28} Therefore, it is important to examine how the ‘standard’ water models describe water interactions within MOFs.

In this work, we perform a comprehensive simulation study of water adsorption in a variety of MOFs and examine the effect of water on their performance for natural gas purification. The MOFs examined in this work are: MIL-47, ZIF-8, Al(OH)(1,4-naphthalenedicarboxylate), UiO-66, Zn-DMOF-A, Zn-DMOF-TM, CAU-OH-10, MOF-806, and Cu-BTC. The MOFs were selected based on the availability of experimental water adsorption data, and to provide structural diversity and a range of expected hydrophobicities/hydrophilicities. First, we evaluate the accuracy of seven “standard” non-polarizable water models, including SPC/E, TIP3P, TIP3P/Ew, TIP4P, TIP4P/Ew, TIP5P, and TIP5P/Ew, by calculating pure water adsorption isotherms and comparing them to experimental data for each of these MOFs. Following an evaluation of the water force field, we investigate the effect that water has on the MOFs’ adsorption properties of a mixture of CO₂-CH₄ under natural gas purification conditions.

5.2 Systems and Computational Details

Figure 5.1 shows the crystal structures of nine MOFs that were evaluated in this study. ZIF-8 (Figure 5.1a) is a widely studied material from a class of MOFs known as Zeolitic

Imidazolate Frameworks (ZIFs) which have zeolite topologies. It is composed of tetrahedrally-coordinated Zn ions connected by imidazolate SBUs. ZIF-8 has a hexagonal channel with a small aperture size of 3.4 Å. MIL-47 (Figure 5.1b) is a hydrophobic MOF from the Matériaux de l'Institut Lavoisier (MIL) series, which does not experience any change in porosity upon water adsorption. MIL-47 is made of vanadium atoms connected by 1,4-benzenedicarboxylate linkers and has a one-dimensional diamond-shaped pore system with a maximum pore diameter of 12.0 Å. Al(OH)(1,4naphthalenedicarboxylate) (Figure 5.1c) is composed of parallel, one-dimensional Al(OH) chains connected together in two directions by 1,4-naphthalenedicarboxylate SBUs. This provides two different one dimensional channels (with 7×7 Å and 3×3 Å cross sections) that are not connected together. UiO-66 (Figure 5.4d) is one of the most studied MOFs and is built from $Zr_6O_4(OH)_4$ clusters bridged by 1,4-benzenedicarboxylate SBUs. It is composed of octahedral cages with a free diameter of 11 Å, and tetrahedral cages with a free diameter of 8 Å. These cages are connected through narrow windows with the diameter of 6 Å. Zn-DMOF-A (Figure 5.1e) is a non-catenated pillared MOF in which the coordination between 9,10-antheracenacarboxylates (ADC) and zinc ions form 2D layers, and 1,4 diazabicyclo[2,2,2]-octane(DABCO) ligands connect the 2D layers to each other. Cu-BTC (Figure 5.1g) is an open metal site MOF which is composed of copper paddle-wheel units that are connected to each other by 1,3,5-benzenetricarboxylate (BTC) organic SBUs, giving rise to square 9.0×9.0 Å pores. CAU-10-OH (Figure 5.1h) is a hydrophilic MOF which is composed of AlO_6 polyhedra connected together in two directions by OH functionalized 1,3-benzenedicarboxylate linkers. This provides one dimensional channels with an aperture size of 5.8 Å. MOF-806 (Figure 5.1i) is built from $Zr_6O_4(OH)_4$ clusters bridged by dihydroxyl

functionalized biphenyldicarboxylate. It is composed of tetrahedral and octahedral cages with free diameters of 10.1 Å and 12.6 Å, respectively.

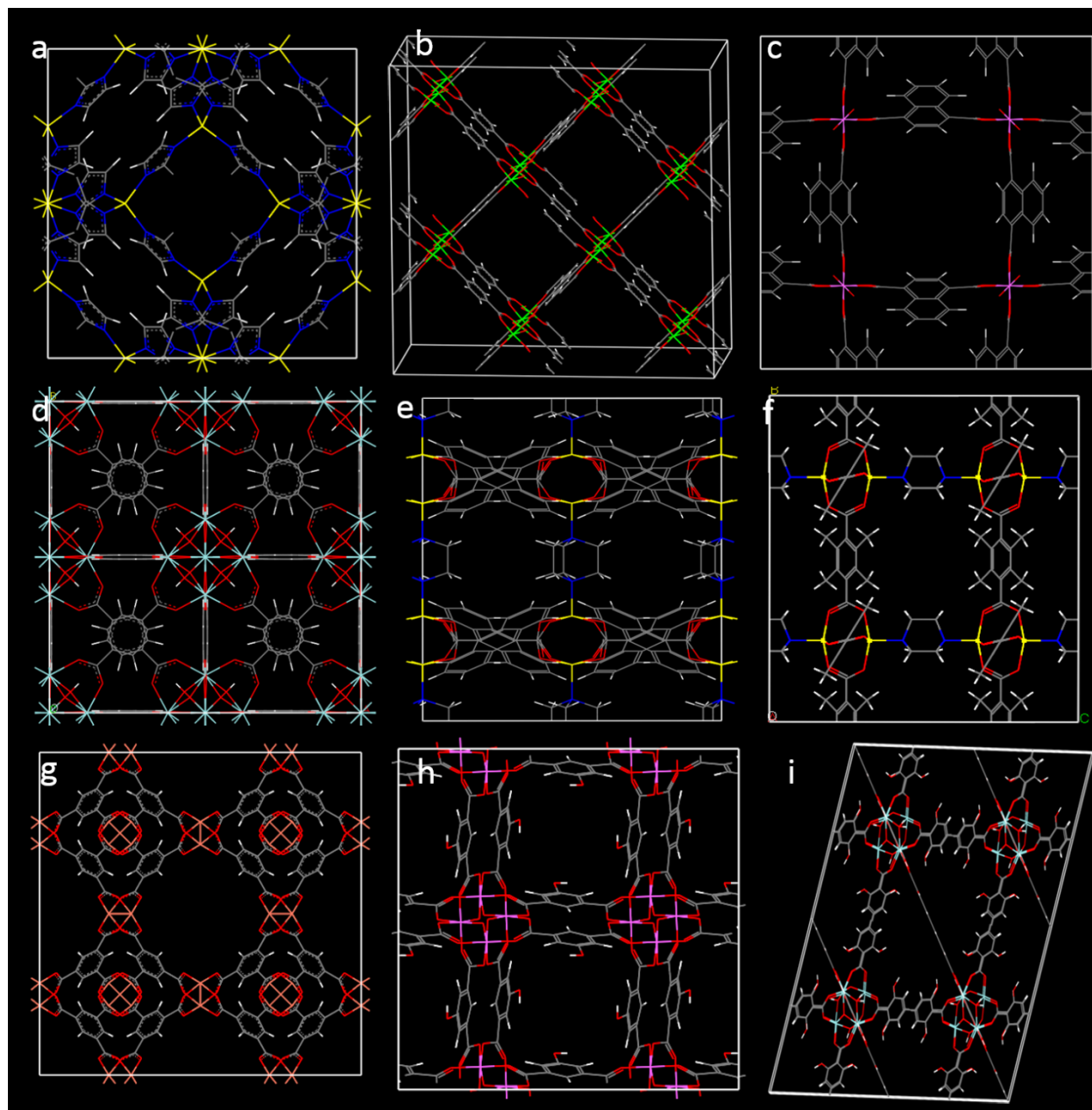


Figure 5.1: Atomic structure of (a) ZIF-8, (b) MIL-47, (c) Al(OH)(1,4-naphthalenedicarboxylate), (d) UiO-66, (e) Zn-DMOF-A, (f) Zn-DMOF-TM, (g) Cu-BTC, (h) CAU-OH-10, and (i) MOF-806. N, blue; C, gray; Cu, orange; O, red; V, green; Al, purple; Zn, yellow; and Zr, cyan.

To describe the guest-host interactions in these systems, standard Lennard-Jones and Coulomb potentials were used. The Lennard-Jones parameters for the framework atoms of each system were taken from the Universal Force Field (UFF)²⁹ and are provided in Table 5.1. REPEAT³⁰ electrostatic potential fitted atomic charges were used for the MOF framework atoms where the electrostatic potential was derived from periodic DFT calculations using the Vienna ab initio software package VASP.^{31,32} For each system, the DFT calculation used a single unit cell of the system with a 2x2x2 k-point sampling of the Brillouin zone centered at the gamma point. The PBE exchange-correlation functional³³ and PAW potentials^{34,35} were used with a plane wave cut-off of 520 eV with a geometry optimization force convergence set to 2×10^{-5} eV/Å. The REPEAT charges obtained from these calculations and used for the simulations are given in Table 5.2 with the atom numberings shown in Figure 5.2.

Table 5.1: The UFF Lennard-Jones parameters for atoms in the MOFs studied in this work

Element	$\sigma / \text{\AA}$	$\epsilon / \text{kcal mol}^{-1}$
V	2.8010	0.016
C	3.4309	0.105
N	3.2607	0.069
O	3.1181	0.060
H	2.5711	0.044
Al	4.0082	0.505
Zr	2.7832	0.069
Cu	3.1137	0.005
Zn	2.4616	0.124

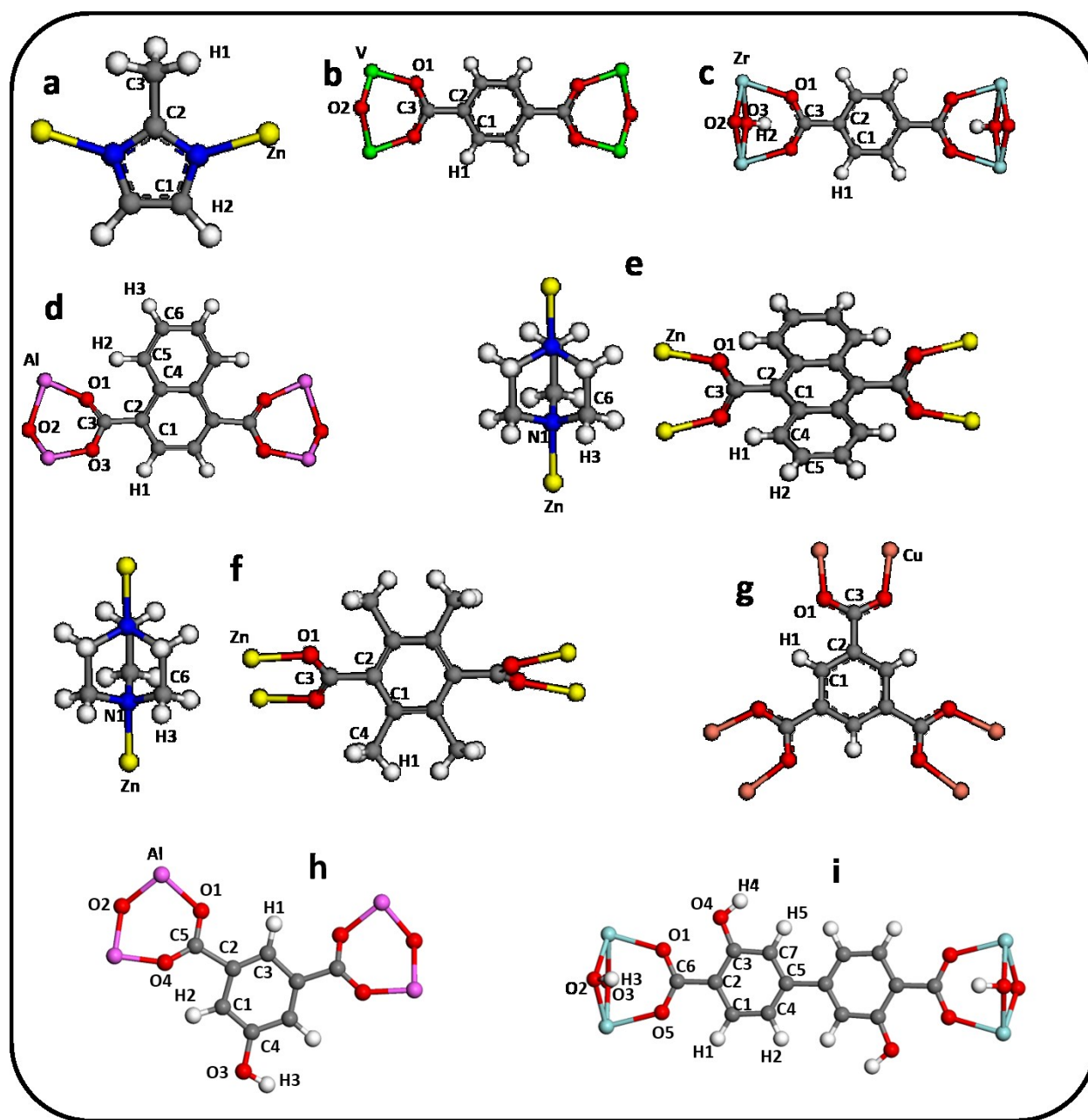


Figure 5.2: Atom labelling of (a) ZIF-8, (b) MIL-47, (c) UiO-66, (d) Al(OH)(1,4 naphthalenedicarboxylate), (e,) Zn-DMOF-A, (F) Zn-DMOF-TM, (g) Cu-BTC, (h) CAU-OH-10, and (i) MOF-806 for the partial atomic charges given in Table 5.2.

Table 5.2: Partial atomic charges (e) of ZIF-8, MIL-47, UiO-66, Al(OH)-MOF, Zn-DMO-TM, Zn-DMOF-A, and Cu-BTC

Atomic types	ZIF-8 (a)	MIL-47 (b)	UiO-66 (c)	Al(OH) (d)	Zn-DMO-A (e)	Zn-DMOF-TM (f)	Cu-BTC (g)	CAU-10-OH (h)	MOF-806 (i)
C1	-0.192	-0.157	-0.147	-0.062	0.012	0.21	-0.121	-0.407	-0.130
C2	0.441	0.064	0.023	-0.199	-0.426	-0.470	-0.016	0.114	-0.073
C3	-0.682	0.542	0.742	0.658	0.870	0.956	0.669	-0.337	0.360
C4	-	-	-	0.167	-0.11	-0.222	-	0.613	-0.280
C5	-	-	-	-0.218	-	-0.116	-	0.584	0.230
C6	-	-	-	-0.103	-0.366	-0.120	-	-	0.592
C7	-	-	-	-	-	-	-	-	-0.402
H1	0.101	0.124	0.137	0.107	0.123	0.150	0.131	0.183	0.163
H2	0.141	-	0.632	0.167	-	0.110	-	0.259	0.128
H3	-	-	-	0.122	0.040	0.070	-	0.407	0.089
H4	-	-	-	-	-	-	-	-	0.403
H5	-	-	-	-	-	-	-	-	0.113
O1	-	-0.564	-0.682	-0.481	-0.633	-0.655	-0.55	-0.522	-0.499
O2	-	-0.646	-0.881	-0.534	-	-	-	-0.625	-0.621
O3	-	-	-1.673	-	-	-	-	-0.590	-1.128
O4	-	-	-	-	-	-	-	-	-0.642
O5	-	-	-	-	-	-	-	-0.599	-0.524
N1	-0.281	-	-	-	-0.219	-	-	-	-
N2	-	-	-	-	-	-	-	-	-
Zn	0.711	-	-	-	0.835	0.850	-	-	-
V	-	1.340	-	-	-	-	-	-	-
Al	-	-	-	1.145	-	-	-	1.477	-
Zr	-	-	2.522	-	-	-	-	-	2.231
Cu	-	-	-	-	-	-	0.908	-	-

The force field parameters of water models considered in this work are provided in Table 5.3, with the definitions of geometric parameters given in Figure 5.3. All six models use a single Lennard-Jones interaction site (i.e., oxygen), while the models use different sites to describe the electrostatic interactions. SPC/E and TIP3P place a negative charge on the oxygen atom while TIP4P and TIP5P type models use one and two dummy atoms, respectively, to distribute the negative charge.

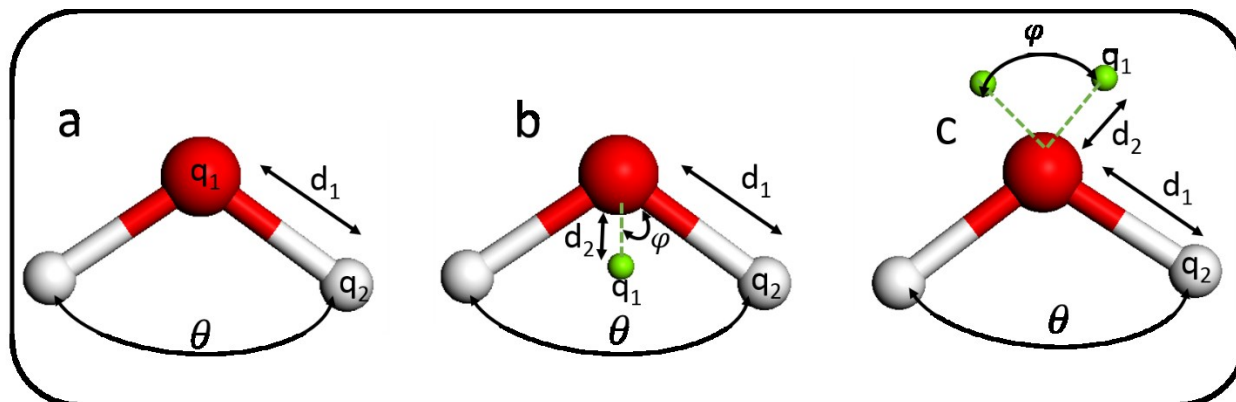


Figure 5.3: Definition of the geometric parameters used in the water models: (a) SPC/E and TIP3P, (b) TIP4P, TIP4P/Ew, and (c) TIP5P, TIP5P/Ew

Table 5.3: Force field parameters and geometries of SPC/E, TIP3P, TIP4P, TIP4P/Ew, TIP5P, and TIP5P/Ew

Models	σ (Å)	ϵ (kcal mol ⁻¹)	q_1 (e)	q_2 (e)	d_1 (Å)	d_2 (Å)	θ (°)	ϕ (°)
SPC/E	3.1656	0.15535	-0.8480	0.4240	1.000	-	109.47	-
TIP3P	3.1501	0.15210	-0.8340	0.4170	0.957	-	104.52	-
TIP4P	3.1540	0.15481	-1.0400	0.5200	0.957	0.150	104.52	52.26
TIP4P/Ew	3.1643	0.16275	-1.0484	0.5242	0.957	0.125	104.52	52.26
TIP5P	3.1200	0.15999	-0.2410	0.2410	0.957	0.700	104.52	109.47
TIP5P/Ew	3.0970	0.17801	-0.2410	0.2410	0.957	0.700	104.52	109.46

The force field parameters of CO₂ were taken from García-Sánchez et al.³⁶ who optimized them from a self-consistent fitting of the experimental adsorption isotherm data of a small set of zeolites. The CO₂ molecule has a fixed bond length of 1.16 Å and partial charges of +0.645 e and -0.3325 e for carbon and oxygen, respectively. The force field parameters for CH₄ were taken from the TraPPE force field,³⁷ where the CH₄ molecule was modeled as a single sphere. This force field has been fit to reproduce the experimental vapor-liquid phase equilibrium data.

The GCMC simulations were run using an in-house developed code based on the open-source DL_POLY Classic molecular dynamics package. The GCMC moves include insertion, deletion and displacement with equal probabilities. The geometry of the framework atoms was taken from X-ray crystallographic data and was held fixed during the GCMC simulations. A

Lennard-Jones cut-off distance of 12.5 Å was used for all simulations, and electrostatic interactions were calculated using the Ewald summation method. In the case of the pure water isotherm, each system took between 300 million and 1 billion steps to reach equilibrium. After being equilibrated, each system underwent further 500 million production steps for data collection. To ensure that each isotherm point was truly equilibrated, GCMC simulations were initiated from two different initial states. One in which the framework was initially empty and one where it was completely loaded with water. This test was performed to confirm that at a given pressure the simulation initiated from the fully loaded state equilibrated to the same point as the simulation that was initiated from the empty state. All GCMC simulations for CH₄-CO₂ adsorption in the absence and presence of water included 20 million equilibrium steps, followed by 20 million production steps. The constant chemical potential used for the GCMC simulations was calculated using the ideal-gas equation of state for pressures up to 1 bar, while for higher pressures the Peng-Robinson equation of state was used to relate the fugacity and pressure.

5.3 Results and discussion

5.3.1 Water adsorption

For the sake of simplicity, this discussion starts with the study of water adsorption in hydrophobic MOFs, followed by hydrophilic MOFs, and ends with an open metal site MOF, Cu-BTC.

Figure 5.4 presents simulated water isotherms for six distinct water models used in hydrophobic MOFs including MIL-47, ZIF-8, Al(OH) (1,4 NDC), UiO-66, Zn-DMOF-A, and Zn-

DMOF-TM at 298 K as a function of relative pressure (p/p_0), supplemented with experimental water adsorption data. p_0 is the experimental saturation pressure of water at room temperature (0.032 bar). Both the experimental and simulated isotherms show the same general behaviour and follow the typical type-V isotherm shape. At low pressure, there is little to no observable adsorption until there is a very abrupt increase to the saturation limit. This behaviour is attributed to the hydrophobic nature of these MOFs in which the water-MOF interactions are much weaker than water-water interactions, which makes the adsorption of water unfavorable at very low pressures. At saturation points, the filling of the MOFs' pores occurs such that the water molecules cluster together through hydrogen bond network, which is detailed later.

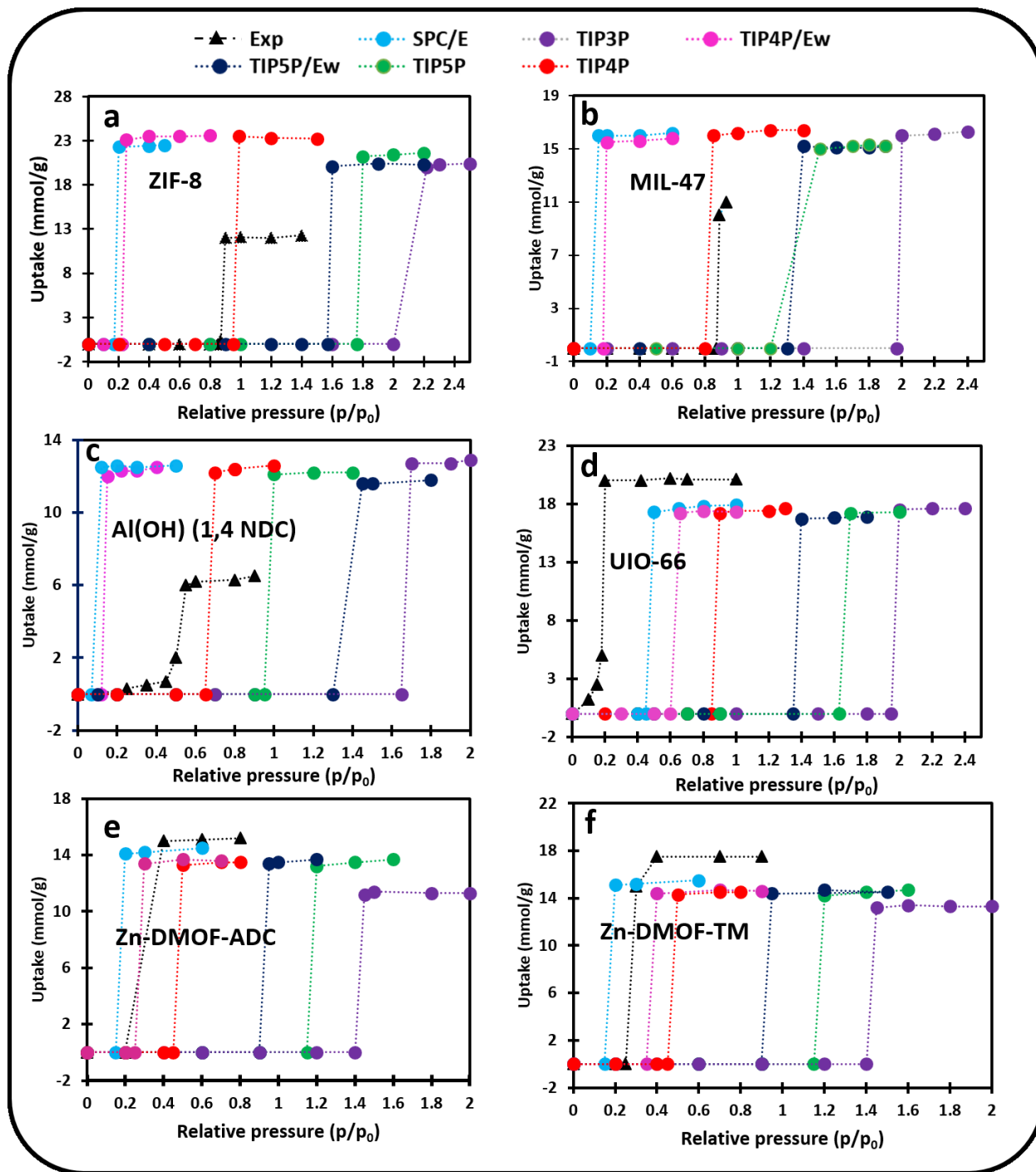


Figure 5.4: Simulated and experimental adsorption isotherms for water in (a) ZIF-8, (b) MIL-47, (c) Al(OH) (1,4 NDC), (d) UiO-66, (e) Zn-DABCO-ADC, (f) Zn-DABCO-TM. (Experimental Data taken from: ZIF-8,³⁸ MIL-47,²⁰ Al(OH) (1,4 NDC),³⁹ UiO-66,⁴⁰ Zn-DABCO-ADC,⁴¹ Zn-DABCO-TM⁴¹)

For the hydrophobic MOFs (Figure 5.4a-f), the saturation pressure observed for each of the water models is quite distinct. Since the Lennard-Jones interaction parameters are fairly similar for the water models, the observed differences in the saturation pressures can be mostly attributed to the electrostatic component of the models. The TIP5P, TIP5P/Ew and TIP3P water models significantly overestimate the saturation pressure for all MOFs tested. On the other hand, the SPC/E model tends to underestimate the saturation pressure for the hydrophobic MOFs (Figure 5.4a-f), with the exception of UiO-66 for which all models overestimate the saturation pressure. The TIP4P and TIP4P/Ew models both under and overestimate the saturation pressure. Overall, the TIP4P water model gives the best agreement with experiment compared to all other water models.

Of all of the MOFs, the experimental water adsorption isotherm for UiO-66 stands out in that it has the lowest saturation pressure ($<0.2 p/p_0$) and that all water models significantly overestimate the saturation pressure. There may be several reasons, why all of the models overestimate the saturation pressure. Our force fields only account for non-covalent bonding, and it may be that defects within the UiO-66 framework that have been reported^{42,43} allow for water-framework interactions that ‘seed’ the saturation. Since our simulations use defect free models of the MOFs, we are not accounting for these interactions resulting in an overestimating of the saturation pressure.

Figure 5.5 provides the number of guest molecules, along with snapshots, during the GCMC simulation of TIP4P water in a simulation cell of MIL-47 at 0.025 bar and 300 K. The plateaus, along with the snapshots from the GCMC simulations, indicate that one pore is filled to saturation before water adsorption occurs in another pore. The adsorption of a water

molecule in a pore is a stochastic process, but once it occurs, the pore with the water molecule will favor the adsorption of further water molecules. In other words, no intermediately filled states of the pores have been observed, and pores are filled independently of each other. At the saturation point, on average 48 water molecules occupies each pore, which form water clusters inside the MIL-47 pores.

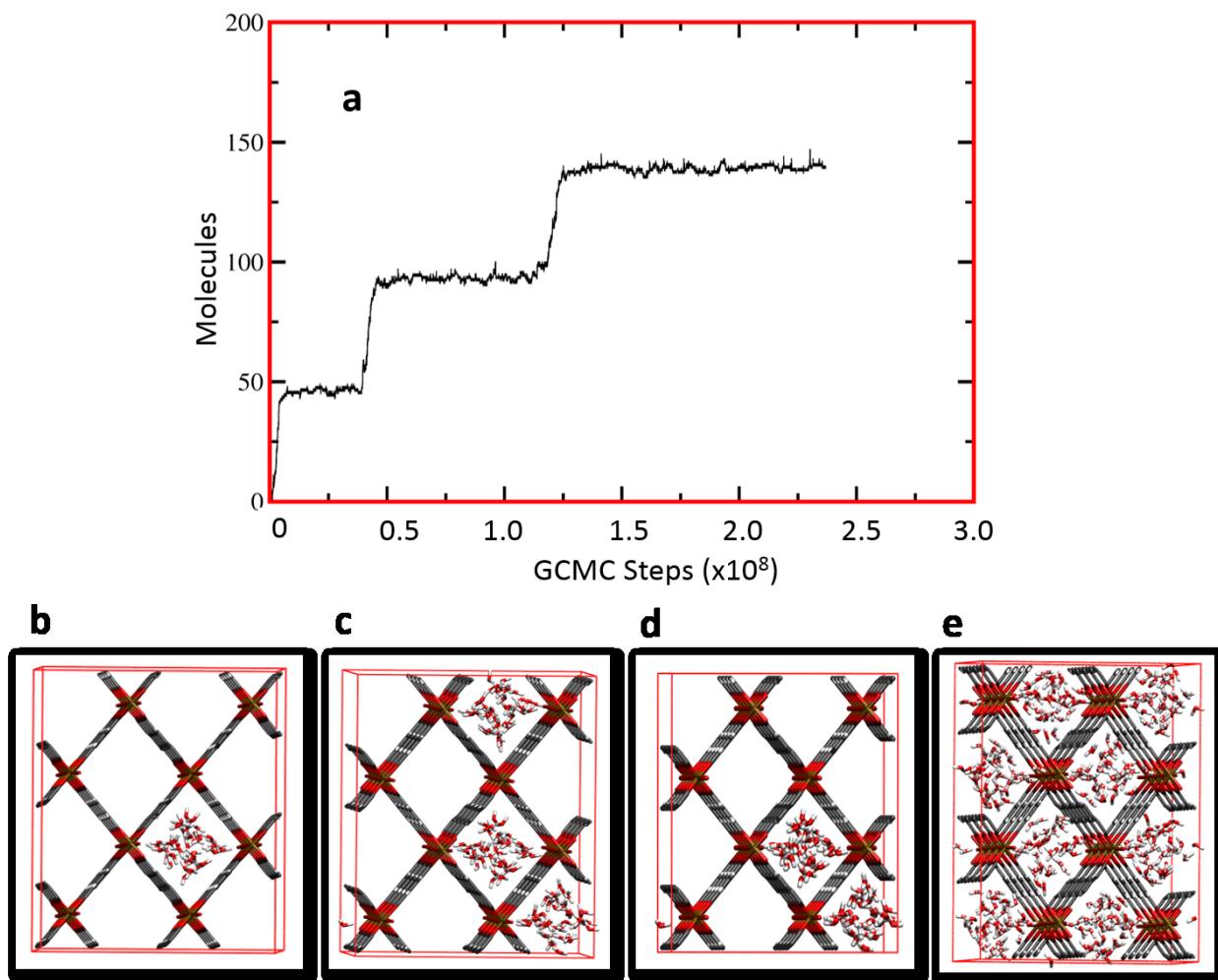


Figure 5.5: (a) the number of guest molecules (TIP4P-Ew water molecules) in MIL-47 as a function of GCMC steps. (b-e) Snapshots of water adsorption during GCMC simulation: (b) 40 millions steps, (c) 80 millions steps, (d) 200 millions steps, and (e) saturation point.

Water clusters formed in MIL-47-type MOFs are different from those of the bulk phase. For example, Figure 5.6 shows that most water molecules prefer to form four hydrogen bonds in the bulk phase while in MIL-47 most water molecules have three hydrogen bonds. This reduction in the number of hydrogen bonds can be attributed to the fact that most water molecules in a confined system interact with the framework atoms of the pore and consequently have less opportunities to hydrogen bond with other water molecules compared to the bulk water. The distribution shown in Figure 5.6 was obtained using reported criteria⁴⁴ where a hydrogen bond for a pair of water molecules is considered if the oxygen-oxygen distance of two water molecules is less than 3.5 Å and the OHO angle is greater than 150°.

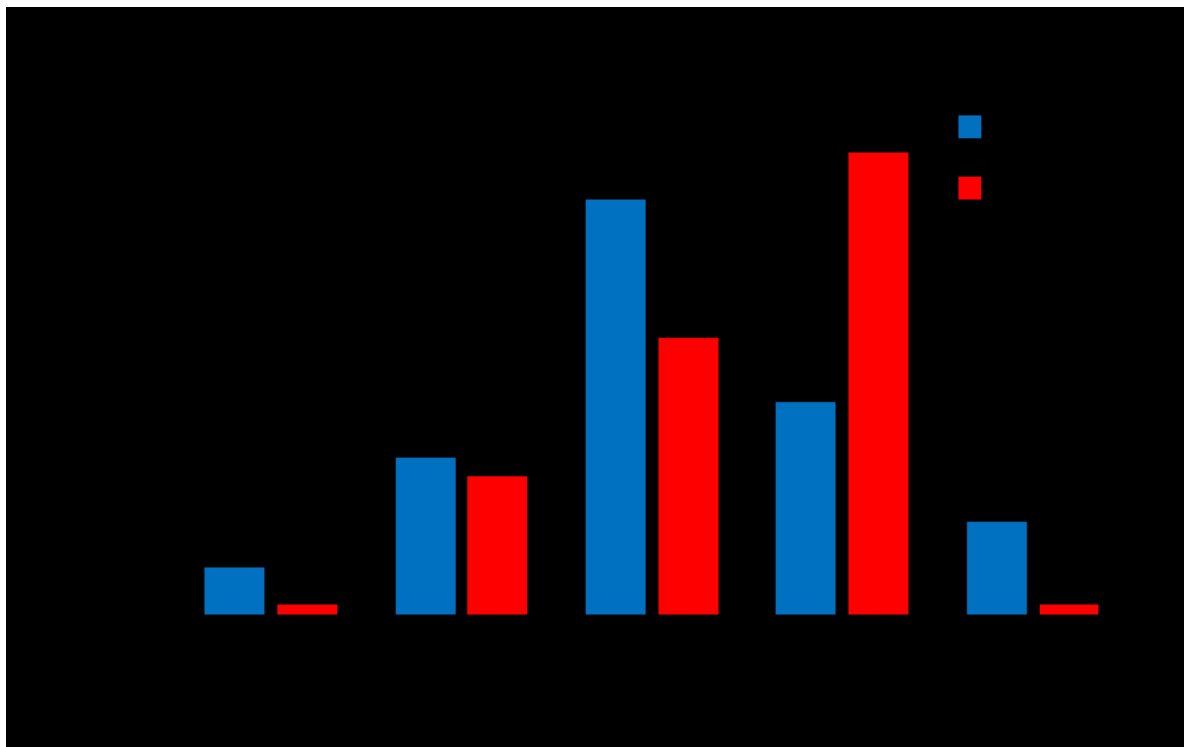


Figure 5.6: Histogram of the number of hydrogen bonds per water molecule formed in MIL-47 (blue bar) and bulk (red bar). The distribution for bulk water is obtained through the simulation of TIP4P water molecules in a box with the same size as the MIL-47.

Figure 5.7 presents simulated and experimental water isotherms for hydrophilic MOFs including CAU-OH-10 and MOF-806. Of the models used, for both MOFs, the TIP4P-Ew water model gives the best agreement with experiment, although it overpredicts the uptakes at the plateau of the isotherms. The TIP5P, TIP5P/Ew and TIP3P water models perform poorly in the simulation of experimental isotherms as they significantly underpredict uptake, and the SPC/E model tends to underestimate the pressure at which the isotherm reaches a plateau. With regards to the shape of the isotherms, both the experimental and TIP4P-Ew simulated isotherms exhibit a different general behaviour than the hydrophobic MOFs discussed above. The adsorption of water in hydrophilic MOFs is more gradual. This behavior can be attributed to the hydrophilic hydroxyl functional groups that line the pores of CAU-OH-10 and MOF-806. To provide more details, Figure 5.8 shows a series of GCMC-derived snapshots showing water molecules adsorbed in CAU-OH-10 at different relative pressures. At low relative pressure, as shown in Figure 5.8a the water molecules are preferentially located next to OH groups. At higher relative pressure (Figure 5.8b), additional water molecules are adsorbed around the initially adsorbed water molecules (green region in Figure 5.5b) located next to the OH groups. This suggests that once a water molecule is located next to an OH group, it promotes the uptake of water at higher pressures.

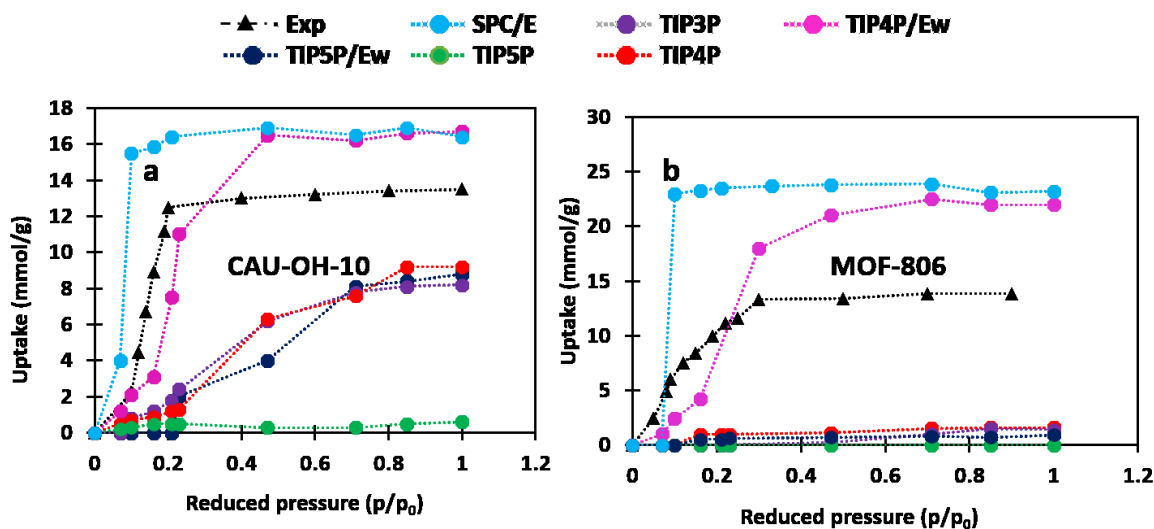


Figure 5.7: Simulated and experimental adsorption isotherms for water in (a) CAU-OH—10 and (b) MOF-806. (Experimental Data taken from: CAU-OH-10⁴⁵ and MOF-806⁴⁶)

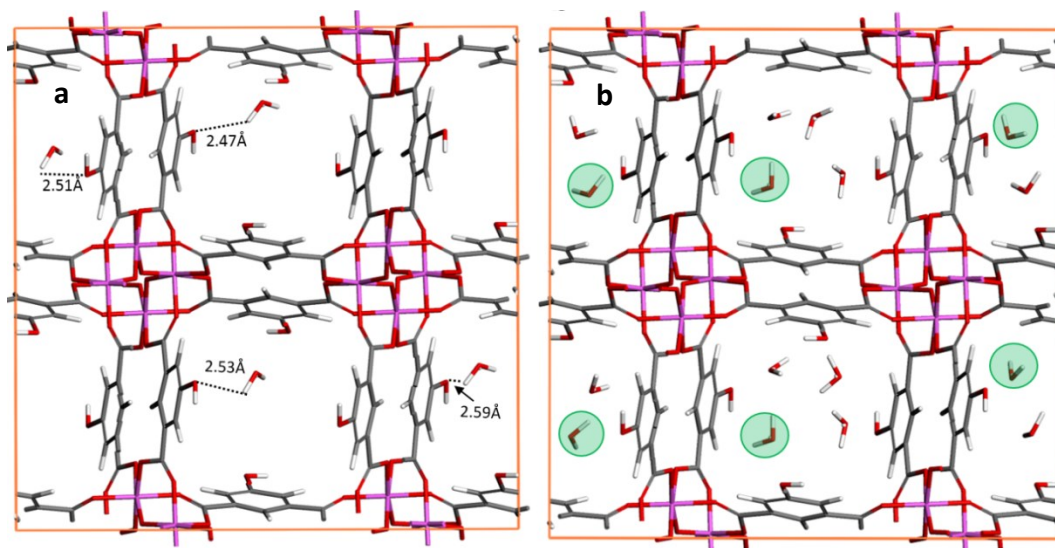


Figure 5.8: Snapshot showing TIP4P-Ew water molecules adsorbed in CAU-OH-10 at (a) $p/p_0=0.17$ and (b) $p/p_0=0.45$. Green region in (b) represents water molecules located next to OH groups.

Figure 5.9a shows that none of the six models correctly reproduce the experimental adsorption isotherm in Cu-BTC. This is attributed to the exposed Cu sites in Cu-BTC which covalently interact with water molecules. These interactions are not captured by the traditional force fields used in this study. For this reason, we also simulate water adsorption on a framework of Cu-BTC in which water molecules are attached to each Cu site through the

oxygen of the water. Here, we consider two cases, one where the partial atomic charges of the coordinated water were obtained using a REPEAT calculation on the MOF framework containing the water molecules, and one where force field-type water molecules are added and partial atomic charges were taken from the force field. The results for REPEAT-derived charges and force field-based charges are given in Figure 5.9 b and c, respectively. In both cases, TIP4P-Ew and SPC-E perform better than other models in the prediction of saturation pressure. However, they both under-predict the amount of water adsorbed at low pressures.

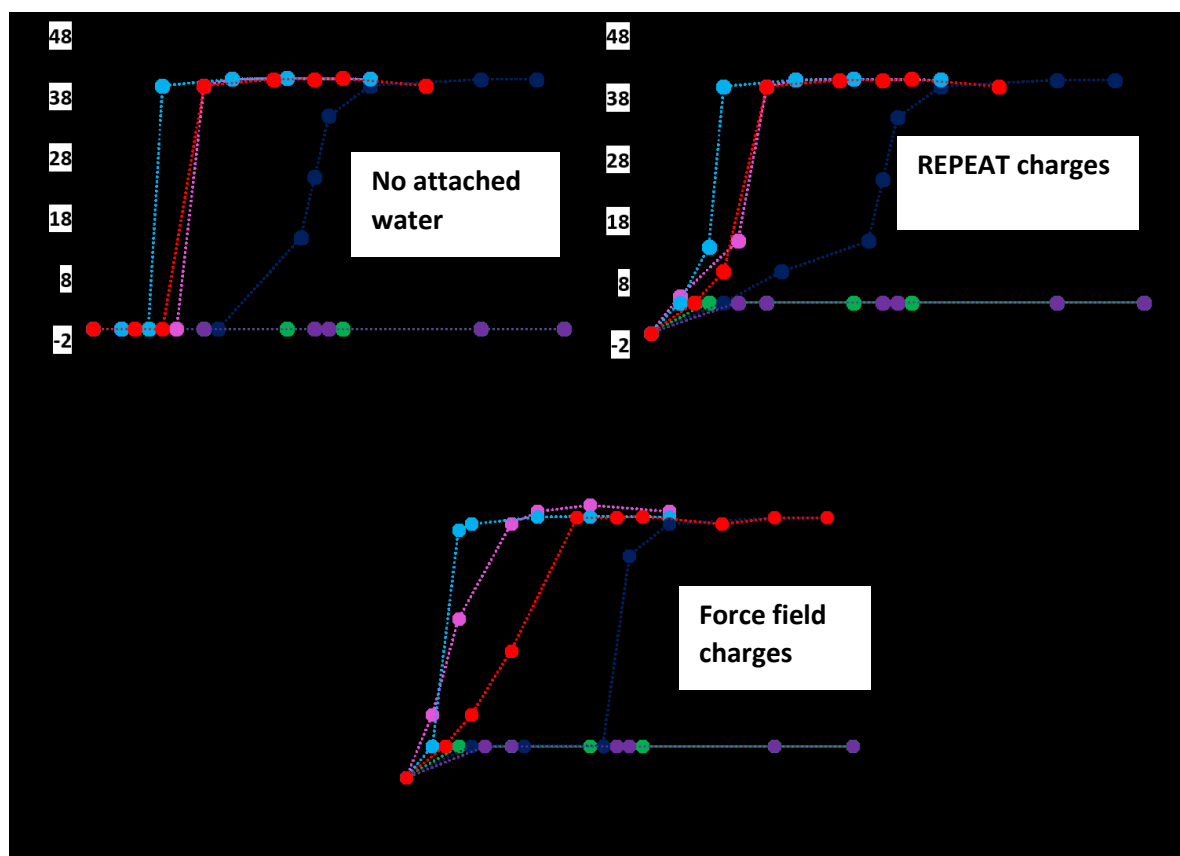


Figure 5.9: Simulated and experimental adsorption isotherms for (a) Cu-BTC without attached water, (b) Cu-BTC with attached water molecules, and (c) Cu-BTC with attached force field-type water molecules. (Experimental Data taken from: ref 47).

5.3.2 The effect of water on natural gas purification in MOFs

To understand the effect of water on natural gas purification in MOFs, the adsorptions of dry and wet CO_2/CH_4 mixtures with a composition of 1:9 were simulated at 303 K and pressures up to 30 bar for the MOFs considered in this work. For the wet mixture, the partial pressure of water was chosen to be 0.03% of the total pressure, as the water specification in natural gas pipelines is 160 mg/m^3 .⁴⁸ In the case of the wet mixture with the open metal site MOF Cu-BTC, one water molecule was coordinated to each Cu atom in the framework, such that the coordinated water molecules were considered as part of the framework. The adsorption selectivities for both mixtures were calculated using the $(x_i/x_j)/(y_i/y_j)$, where the x_i and x_j are the loadings of component i and j in the adsorbed phase while the y_i and y_j are the partial pressures of component i and j in the gas phase.

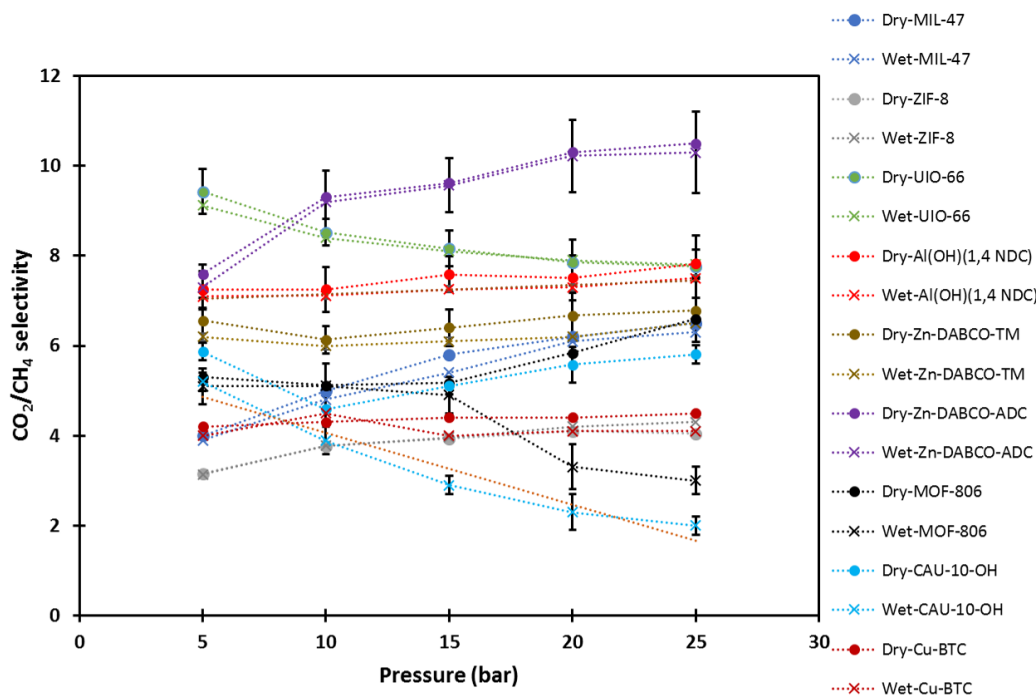


Figure 5.10: Selectivities of CO_2 in 1:9 mixtures of CO_2/CH_4 with (cross) and without (circle) water as a function of pressure at 303 K.

As seen in Figure 5.10, water has a minimal effect on the selectivities for CO₂ in MIL-47, ZIF-8, Al(OH)(1,4 NDC), UiO-66, Zn-DMOF-A, and Zn-DMOF-TM. The reason, as discussed in section 5.3.1, is that the hydrophobic MOFs have a weak affinity for water at very low pressures, and consequently the CO₂ adsorption is not affected by water. Figure 5.10 also shows that the selectivity of CO₂ remains constant through the entire pressure range in Cu-BTC for the wet mixture. On the other hand, the selectivity of CO₂ gradually decreases with increasing pressure in CAU-OH-10 and MOF-806 for the wet mixture, which can be attributed to strong affinity of OH functional groups for water over CO₂, subsequently resulting in decreased adsorption of CO₂.

5.4 Conclusion

We investigated the accuracy of seven popular non-polarizable water models (SPC/E, TIP3P, TIP4P-Ew, TIP4P, TIP5P and TIP5P-Ew) for examining water adsorption in hydrophobic and hydrophilic MOFs with GCMC simulations. The water adsorption isotherms of the MOFs MIL-47, ZIF-8, Al(OH)(1,4-NDC), UiO-66, Zn-DMOF-A, Zn-DMOF-TM, CAU-OH-10, MOF-806, and Cu-BTC were calculated and compared to experimental isotherms. The results revealed that TIP4P water model performs better than other models in the prediction of saturation pressure in hydrophobic MOFs. For hydrophilic MOFs, the results show that TIP4P-Ew gives the best agreement with experiment compared to all other water models. Our GCMC simulations also revealed that water adsorption in hydrophobic and hydrophilic MOFs are considerably different from each other. Hydrophilic MOFs gradually uptake water, whereas hydrophobic MOFs show rapid water uptake at the saturation pressure. Snapshots from GCMC simulations in a

hydrophobic MOF indicate that one pore is filled to saturation before water adsorption is seen in another pore. We also investigated the effect of water on natural gas purification in these MOFs, and found that water has a negligible effect on natural gas purification in hydrophobic MOFs under maximum humidity specifications for pipeline natural gas transport.

5.5 References

- (1) 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* (80-.). **2010**, 329 (5990), 424.
- (2) Peng, Y.; Li, Y.; Ban, Y.; Jin, H.; Jiao, W.; Liu, X.; Yang, W. *Science* **2014**, 346 (6215), 1356.
- (3) Lu, W.; Wei, Z.; Gu, Z.-Y.; Liu, T.-F.; Park, J.; Park, J.; Tian, J.; Zhang, M.; Zhang, Q.; Gentle III, T.; Bosch, M.; Zhou, H.-C. *Chem. Soc. Rev.* **2014**, 43 (16), 5561.
- (4) Choi, E.-Y.; Wray, C. a.; Hu, C.; Choe, W. *CrystEngComm* **2009**, 11 (4), 553.
- (5) Lin, C. K.; Zhao, D.; Gao, W. Y.; Yang, Z.; Ye, J.; Xu, T.; Ge, Q.; Ma, S.; Liu, D. J. *Inorg. Chem.* **2012**, 51 (16), 9039.
- (6) Li, J.; Sculley, J.; Zhou, H. *Chem. Rev.* **2012**, 112, 869.
- (7) Li, J.-R.; Kuppler, R. J.; Zhou, H.-C. *Chem. Soc. Rev.* **2009**, 38 (5), 1477.
- (8) Keskin, S.; Sholl, D. S. *J. Phys. Chem. C* **2007**, 111, 14055.
- (9) Britt, D.; Furukawa, H.; Wang, B.; Glover, T. G.; Yaghi, O. M. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, 106 (49), 20637.
- (10) Watanabe, T.; Sholl, D. S. *Langmuir* **2012**, 28, 14114.
- (11) Jiang, J.; Furukawa, H.; Zhang, Y.; Yaghi, O. M. **2016**.
- (12) Mason, J. a.; Sumida, K.; Herm, Z. R.; Krishna, R.; Long, J. R. *Energy Environ. Sci.* **2011**, 4, 3030.
- (13) Ribeiro, A. M.; Sauer, T. P.; Grande, C. A.; Moreira, R. F. P. M.; Loureiro, J. M.; Rodrigues, A. E. *Ind. Eng. Chem. Res.* **2008**, 47 (18), 7019.
- (14) Yang, Q.; Liu, D.; Zhong, C.; Li, J. R. *Chem. Rev.* **2013**, 113, 8261.
- (15) Düren, T.; Bae, Y.-S.; Snurr, R. Q. *Chem. Soc. Rev.* **2009**, 38 (5), 1237.
- (16) Krishna, R.; van Baten, J. M. *Phys. Chem. Chem. Phys.* **2011**, 13, 10593.
- (17) Greathouse, J. A.; Allendorf, M. D. *J. Am. Chem. Soc.* **2006**, 128 (33), 10678.
- (18) Castillo, J. M.; Vlugt, T. J. H.; Calero, S. *J. Phys. Chem. C* **2008**, 112, 15934.
- (19) Paranthaman, S.; Coudert, F.-X.; Fuchs, A. H. *Phys. Chem. Chem. Phys.* **2010**, 12, 8123.

- (20) Salles, F.; Bourrelly, S.; Jobic, H.; Devic, T.; Guillerm, V.; Llewellyn, P.; Serre, C.; Ferey, G.; Maurin, G. *J. Phys. Chem. C* **2011**, *115*, 10764.
- (21) De Lange, M. F.; Gutierrez-Sevillano, J.-J.; Hamad, S.; Vlugt, T. J. H.; Calero, S.; Gascon, J.; Kapteijn, F. *J. Phys. Chem. C* **2013**, *117* (lii), 7613.
- (22) Ghosh, P.; Kim, K. C.; Snurr, R. Q. *J. Phys. Chem. C* **2014**, *118*, 1102.
- (23) Rick, S. W. *J. Chem. Phys.* **2004**, *120* (13), 6085.
- (24) Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T. *J. Chem. Phys.* **2004**, *120* (20), 9665.
- (25) Harrach, M. F.; Drossel, B. *J. Chem. Phys.* **2014**, *140* (17).
- (26) Mark, P.; Nilsson, L. *J. Phys. Chem. A* **2001**, *105* (43), 9954.
- (27) Jungwirth, P. *Faraday Discuss.* **2009**, *141*, 9.
- (28) Coudert, F. X.; Vuilleumier, R.; Boutin, A. *ChemPhysChem* **2006**, *7* (12), 2464.
- (29) Rappe, a. K.; Casewit, C. J.; Colwell, K. S.; Goddard III, W. a.; Skiff, W. M. *J. Am. Chem. Soc.* **1992**, *114*, 10024.
- (30) Campañá, C.; Mussard, B.; Woo, T. K. *J. Chem. Theory Comput.* **2009**, *5*, 2866.
- (31) Kresse, G.; Furthmüller, J. *Phys. Rev. B* **1996**, *54* (16), 11169.
- (32) Kresse, G.; Hafner, J. *Phys. Rev. B* **1994**, *49* (20), 14251.
- (33) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77* (1992), 3865.
- (34) Blöchl, P. E. *Phys. Rev. B* **1994**, *50* (24), 17953.
- (35) Kresse, G. *Phys. Rev. B* **1999**, *59* (3), 1758.
- (36) 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* (3), 8814.
- (37) Martin, M. G.; Siepmann, J. I. *J. Phys. Chem. B* **1998**, *102* (97), 2569.
- (38) Küsgens, P.; Rose, M.; Senkovska, I.; Fröde, H.; Henschel, A.; Siegle, S.; Kaskel, S. *Microporous Mesoporous Mater.* **2009**, *120* (3), 325.
- (39) Comotti, A.; Bracco, S.; Sozzani, P.; Horike, S.; Matsuda, R.; Chen, J.; Takata, M.; Kubota, Y.; Kitagawa, S. *J. Am. Chem. Soc.* **2008**, *130*, 13664.
- (40) Jasuja, H.; Walton, K. S. **2013**.
- (41) Jasuja, H.; Huang, Y. G.; Walton, K. S. *Langmuir* **2012**, *28*, 16874.
- (42) Wu, H.; Chua, Y. S.; Krungleviciute, V.; Tyagi, M.; Chen, P.; Yildirim, T.; Zhou, W. *J. Am. Chem. Soc.* **2013**, *135* (28), 10525.

-
- (43) Katz, M. J.; Brown, Z. J.; Colón, Y. J.; Siu, P. W.; Scheidt, K. a; Snurr, R. Q.; Hupp, J. T.; Farha, O. K. *Chem. Commun. (Camb)*. **2013**, 49 (82), 9449.
- (44) Luzar, A.; Chandler, D. *Phys. Rev. Lett.* **1996**, 76 (6), 928.
- (45) Reinsch, H.; Veen, M. A. Van Der; Gil, B.; Marszalek, B.; Verbiest, T.; Vos, D. De; Stock, N. **2013**.
- (46) Furukawa, H.; Ga, F.; Hudson, M. R.; Yaghi, O. M. **2014**.
- (47) Liang, Z.; Marshall, M.; Chaffee, A. L. *Energy and Fuels* **2009**, 23 (5), 2785.
- (48) Carroll, J. *Natural Gas Hydrates: A Guide for Engineers*; 2, Ed.; Gulf Professional Pub, 2009, 2009.

6 Large-Scale Computational Screening of MOFs for Natural gas purification

Abstract

Metal organic frameworks (MOFs) have attracted attention in many gas separation applications, including natural gas purification where CO₂ is separated from CH₄, due to the highly tunable nature of the material. A seemingly unlimited number of combinations of inorganic and organic structural building units can be used to construct a highly diverse set of MOFs. To identify the key structural and chemical features that provide high performance for natural gas purification, we have screened a database of 324,000 hypothetical MOFs for their CO₂ working capacity and CO₂/CH₄ selectivity using GCMC simulations. We find that MOFs with a pore diameter less than 6.6 Å but possess surface areas higher than 2299 m²g⁻¹ generally perform well for natural gas purification. We have applied a scheme to locate the CO₂ binding sites in the best performing MOFs from the probability distributions resulting from the GCMC simulations. A similarity analysis of the binding sites was formed to identify the key chemical and geometric traits of these MOFs. The most common CO₂ binding environment identified in the high performance MOFs are layered sheets of aromatic groups separated by roughly 6.5 Å.

6.1 Introduction

Due to both the global increase in energy consumption and rising greenhouse gas emissions,^{1,2} there is a demand to shift towards methane-based energy sources since the combustion of natural gas releases less CO₂ per unit of energy compared to coal and oil.³ More and more, unconventional natural gas reservoirs are being tapped, which can have CO₂ concentrations of 15 vol% or more. Prior to sale and transport, the removal of CO₂ is required

because it not only reduces the energy density of the fuel but CO_2 causes pipeline and equipment corrosion through the formation of carbonic acid. Moreover, in cold regions CO_2 can form solid clathrate hydrates in the pipelines that can cause serious clogging problems.

One common method used in industry to scrub CO_2 from natural gas involves using solid sorbents and pressure swing adsorption gas separation technology.⁴ In this process, CO_2 from natural gas is selectively adsorbed onto a solid porous material which later undergoes a regeneration step where the CO_2 is desorbed by applying a vacuum to decrease the pressure. The most commonly used adsorbents in these PSA systems are zeolite 13x, 4A, and 5A, which selectively adsorb CO_2 from natural gas.⁵ Despite having a strong affinity for CO_2 over CH_4 , these materials have low CO_2 working capacity. The working capacity is the amount of CO_2 recovered per unit mass (or volume) of the material over a single adsorption/desorption cycle. Put in another way, it is the difference in the CO_2 uptake capacity of the material at the adsorption conditions and at the desorption conditions. Zeolite 13x has CO_2 working capacity of 1.86 mmol/g,⁶ which is low compared to the other porous materials: Cu-BTC (4.0 mmol/gr), PCN-11 (5.5 mmol/gr), and Mg-MOF (7.7 mmol/gr).⁷ The working capacity of the material is important since it affects the size and number of adsorbent beds that are required, and ultimately in the cost of the PSA system. In addition to CO_2 working capacity, the selectivity of the material for CO_2 over CH_4 , which is measured as the molar ratio of CO_2 adsorbed over CH_4 normalized by their partial pressures, is important. The higher the CO_2/CH_4 selectivity, the purer the gas is recovered. Thus there is a considerable interest in finding promising materials with working capacities and selectivities above zeolite 13x for use in PSA systems for scrubbing CO_2 from natural gas.

Among the porous materials, metal organic frameworks (MOFs) have attracted considerable attention as promising candidates to replace zeolites in PSA units for removing CO₂ from natural gas due to the low energy demands of their regeneration process.^{8–11} With the wide availability of unique MOFs, researchers have conducted screening studies on these materials for their application to separate CO₂ from natural gas.^{12,13} The intent of these studies is to identify high performing targets to synthesize or to identify any structural features that can lead to higher performance. However, these studies have been limited to screening of a small set of 10-15 different MOFs, which is insufficient to provide robust structure-property relationships. Moreover, the best candidates identified from these studies were not practical to use on an industrial scale. For example, Krishna et al.,¹³ who screened 20 different MOFs for methane purification under PSA conditions, showed that Mg-MOF-74 outperforms the other candidate MOFs in their study. However, this MOF irreversibly degrades in the presence of trace amounts water because it possesses unsaturated metal sites. This limits their real world use, as there are always traces of water in any feed gas.^{14–16}

An alternative approach to find high performing materials is to screen hypothetical MOFs, which covers an extremely large number of possible MOFs structures. The utility of this approach was first demonstrated by Wilmer et al.¹⁷ whose pioneering large-scale screening study examined 138,000 hypothetical MOFs for methane storage. In that study, they identified MOFs with methane-storage capacities above the world-record holders (at the time). They also identified a MOF that had been previously reported as the NOTT-107 MOF,¹⁸ and found their predicted uptakes were in good agreement with the experimental ones, thus validating that high-throughput screening approach. Recently, they computationally screened the same

database with the intent of finding structure-property relationships of MOFs for CO₂ separation from natural gas.¹⁹ Although the ‘bounds’ of MOFs physical properties were well described in their work, the data indicated that physical and geometric properties of the framework, such as pore size and pore volume alone are not sufficient design factors for creating high-performing MOFs. They further claimed that decorating MOFs’ pores with halogens was an important factor for high performance materials. Unfortunately, Woo and co-workers²⁰ discovered that the high performance of the halogen substituted materials was an artifact of the method in which they used to calculate the charges used in their simulations. Indeed, in the study of CO₂/CH₄ separation performed by Wilmer et al.²¹, and even in similar following studies, the issue of what pore environments contribute to strong and selective CO₂ adsorption has been not fully addressed. In other words, these studies suffer from a lack of chemical insight relevant for CO₂- adsorption, which is a critical for rational design. Thus, there is a need to identify common chemical and geometric features of binding sites, which we will call binding motifs, that allow for strong and selective CO₂ adsorption. Gaining such information requires one to identify the CO₂ binding sites within a MOF’s pore, which is quite difficult to determine experimentally (for physisorbed gases) but easily determined from simulation. For example, Woo, Shimizu and co-workers²² who studied CO₂ binding within the CALF-15 using GCMC simulations were able to identify the CO₂ binding sites both experimentally and by simulation. The agreement between the predicted and experimental binding sites was excellent. From their analysis, it was found that the electrostatic interaction of amine functional groups did not dominate the CO₂ binding as initially thought – rather the CO₂-MOF binding was mostly a

dispersion interaction. A large scale CO₂ binding site analysis has been not performed in the screening of MOFs for natural gas purification.

In this work, we present a computational high throughput screening study of 324,500 hypothetical MOFs for their performance for natural gas purification at typical conditions used for natural gas purification with PSA systems. For each MOF, the performance parameters (selectivity and working capacity) were calculated using data obtained from grand canonical Monte Carlo (GCMC) simulations of CO₂-CH₄ binary mixture, and the geometry of MOFs was characterized by calculating the surface area, void fraction, and maximum pore diameter. We investigate the structure property relationships. Further, we explore the two interesting strong CO₂ binding motifs in the high-performing MOFs, whose selectivity and working capacity are above zeolite 13x, a commercially used sorbent in PSA systems. These adsorption pockets were obtained using a substructure search among frameworks atoms located within 5 Å of the CO₂ binding sites. We show that *layered sheets* consisting of poly-aromatic molecules separated by a perpendicular distance of roughly 7 Å are an important structural-chemical feature that leads to strong adsorption of CO₂. Finally, we identify “hit” materials for targeted synthesis.

6.2 Methods

6.2.1 Hypothetical MOF Database

A hypothetical MOF database was generated from a selection of secondary building units (SBUs), typically organic linkers and metal clusters, derived from existing MOF crystal structures. A program for the assembly of hypothetical MOF structures was developed by the Woo group at the University of Ottawa which uses a method similar to the recursive method established by Wilmer et al.¹⁷ A brief summary of the database generation procedure follows.

*Any building block from the database can combine with any other building block provided that the local geometry and chemical composition at the point of connection are similar to an existing MOF crystallographic structure. Building blocks are combined in a stepwise manner, and when an unphysical atomic overlap occurs at a particular step, a different building block or a different connection site is chosen until all possibilities for crystallographic linkages in the database are exhausted. When no further building blocks can be added, the crystal generation procedure is considered complete.*¹⁷

All generated hypothetical MOFs underwent functionalization whereby hydrogen atoms were replaced by a given functional group in order to enhance the chemical diversity of the chemical substances in the database. The symmetry of the generated structure, determined based on the Grosse-Kunstleve method,²³ was implemented to recognize the symmetrically equivalent hydrogen atoms so that the functionalized structure has symmetry as similar as possible to the parent structure's symmetry. This enhances the subsequent synthetic viability of the functionalized hypothetical MOFs. In total, based on the procedure described above, 345,332 hypothetical MOFs were generated from a library containing 7 inorganic SBUs, 59 organic SBUs and 19 functional groups. The entire list is provided in the appendix to this chapter.

6.2.2 Screening Protocol

In PSA-based natural gas (NG) purification the gas is fed into a fixed volume vessel containing the solid sorbent material. CO₂ is selectively adsorbed from the gas feed by the adsorbent material, resulting in an outgoing gas that is methane-rich. After the adsorbent bed is saturated with CO₂, the feed gas is redirected to another bed and the CO₂ in the existing bed

is desorbed by reducing pressure. Once the CO₂ is desorbed the adsorbent bed is ready to be used again. The optimal adsorption and desorption conditions (T and P) are different for each PSA system, feed gas composition and adsorbent material. In this work, we use a fixed set of adsorption and desorption conditions that are typical of natural gas PSA systems. Specifically, we screened each MOF for adsorption of a binary mixture of CO₂:CH₄ with a composition of 1:9 at a total pressure of 10 bar and a temperature of 298 K, and desorption at a pressure of 1 bar and a temperature of 298 K with the same composition. These conditions have been used in various studies which evaluated the capability of sorbent materials for NG purification.^{24–27} In order to compare the performance of MOFs, we used two traditional critical factors that control the economics of the PSA unit, namely the CO₂:CH₄ selectivity, S , and the CO₂ working capacity. The selectivity is defined in equation 6.1,

$$S = \frac{q_i p_j}{q_j p_i} \quad 6.1$$

where the q_i and p_i are the loading of the gas in the sorbent material and partial pressure of the gas i , respectively. The CO₂ working capacity (Δq) of sorbent materials was obtained from the difference in CO₂ uptake capacity at the adsorption and desorption conditions.

$$\Delta q = q_{CO_2_{ads}} - q_{CO_2_{des}} \quad 6.2$$

Since rigid SBUs were used to construct the hypothetical MOFs, each hypothetical MOF in the database underwent a classical force-field optimization using the UFF force field and the GULP software package.²⁸ The force field optimization of the structures removes unphysical structural features such as overlapping atoms. During the optimizations, the fractional coordinates of the metal ions and atoms directly coordinated to them were fixed, and the positions of other

atoms, along with lattice parameters, were relaxed using a conjugate gradient algorithm until the forces were all below 0.001 eV/Å. The reason to treat the geometry around the metals as rigid goes back to the fact that our inorganic SBUs include metal atoms in geometries that have not been parameterized in the UFF. Indeed, optimization without the rigid treatment would adjust the metal atoms to the geometric configurations parameterized by the UFF, resulting in highly distorted structures.

Following the classical force field optimizations, each MOF was examined for irregularities using a code written by the author. These defects include missing hydrogen atoms and non-sensible lengths between both bonded and non-bonded atoms. To determine whether a given pair of bonded atoms i and j are within a chemically reasonable distance, we compare the distance between atom i and j (d_{ij}) to their covalent radii (r) with scaling factors using the below criteria.

$$0.75 \times (r_i + r_j) < d_{ij} < 1.25 \times (r_i + r_j) \quad 6.3$$

For non-bonded atoms, i and j , the following criteria with scaling factors $\alpha=1.25$ was used.

$$1.25 \times (r_i + r_j) < d_{ij}. \quad 6.4$$

With these criteria, approximately 20,000 MOFs were excluded from screening because they were deemed to possess unphysical structural features.

GCMC simulations were then performed on all of the structure where each structure was subjected to a short GCMC simulation of a binary mixture of CO₂ and CH₄ with 5000 equilibrium cycles and 5000 production cycles. A single cycle contains N Monte Carlo steps, where N is the fluctuating number of guests in the GCMC simulation.

The structures were labelled “high-performing” if the CO₂ working capacity was greater than 1.65 mmol/g and whose CO₂/CH₄ selectivity was >10 and were subjected to more detailed screening. (These cut-off values were chosen as they are comparable or better than that of zeolite 13x) The more advanced screening involved deriving the charges for simulation from DFT calculations using the REPEAT method²⁹ whereas in the initial screening charges were derived from the charge equilibration (QEq) method. Following the calculation of the REPEAT charges, each high-performing MOF was subjected to a long GCMC simulation using 5×10⁶ equilibration and 5×10⁷ production steps. Then, using the data extracted from the long GCMC simulation, binding site analysis was performed on each high-performing MOF. The details of binding site analysis are provided in the following section. The screening procedure is outlined in the flowchart given in the Figure 6.1.

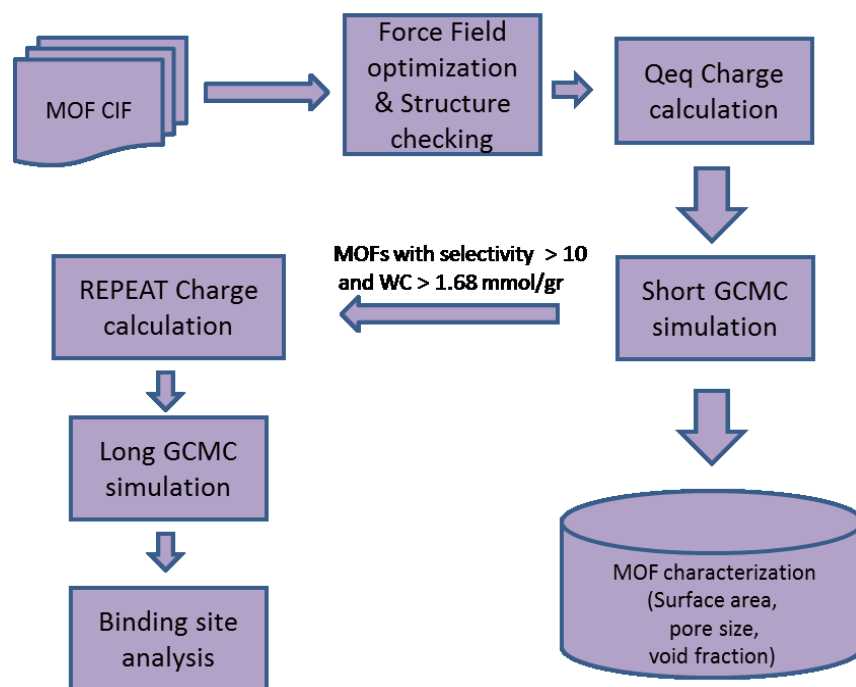


Figure 6.1: Flowchart of screening and analysing of 324,000 MOFs.

6.2.3 Binding Site Analysis

In addition to the prediction of the CO₂ adsorption isotherm, GCMC simulations generate probability distributions of the guest molecules inside the MOFs by adding CO₂ positions from every Monte Carlo step on to a 3D grid with a resolution of $0.1 \times 0.1 \times 0.1 \text{ \AA}^3$ (Figure 6.2a). This histogram consists of a 3D grid of millions of points and was obtained separately for each the C and O atoms of CO₂. From these noisy probability distributions, CO₂ binding sites in each MOF were found using an in-house program developed by the group called ABSL (Automatic Binding Site Locator), which smooths the noisy 3D histograms via a Gaussian filter, and isolates maxima of probability. Each atom of the CO₂ molecule is then placed in the respective maxima in the probability distributions and the guest molecule with a fixed geometry is then best fit to the maxima to give the location of a binding site. (Figure 6.2b). ABSL uses an occupancy cut-off such that only the binding sites with the highest probability are identified and ranked according to their occupancies. Once the binding sites are located, the interaction energy of the binding site (in the absence of any other CO₂ molecules) is calculated with the same force field for which the GCMC simulations were run with. With the procedure described above, the binding sites of the ~3900 high-performing MOFs were identified. In total, approximately 30,000 binding sites were identified. With this data, we then looked for common chemical features in the binding sites of the high-performing MOFs.

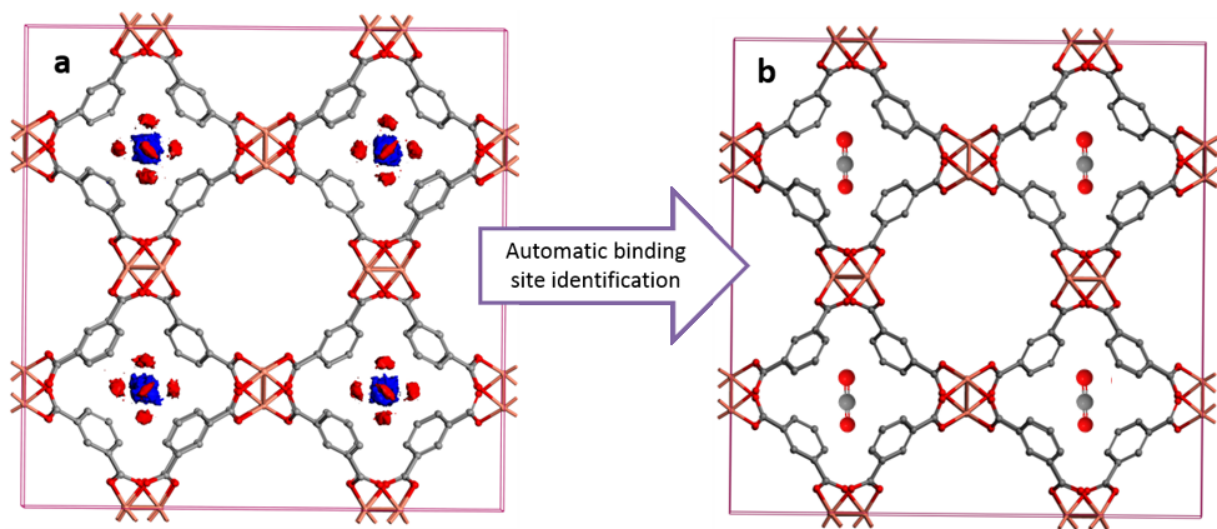


Figure 6.2: a) GCMC-derived 3-D probability distribution of CO₂ carbon (blue cloud) and oxygen (red cloud) in the Cu-BTC MOF and b) representative CO₂ binding sites are identified in the Cu-BTC using the ABSL program.

6.2.4 Discovery of Common Binding Site Motifs

Another program developed within the Woo lab can identify the most common chemical and structural features among a large dataset of MOF binding sites. In the past, the group has called these common binding site motifs, *adsorbophores*, in analogy to the term pharmacophore used in the drug discovery field. The term *pharmacophore* was officially defined in 1998 by IUPAC as “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 adsorption interactions with CO₂ can be considered conceptually similar to this definition, and thus the lab coined the term *adsorbophore* to represent these features.

Briefly, the adsorbophore discovery algorithm works in the following way. Given a set of MOF coordinates with an associated set of guest molecule coordinates, the framework atoms associated with the ‘binding site’ are extracted. Framework atoms located within 5 Å of each

atom of the CO₂ molecule are selected. The most common binding site features are identified by Maximum Clique Detection (MCD) method (detailed later) in a recursive, pairing fashion. Here, each binding site is randomly paired together to produce a set of common substructures using the MCD method detailed later. The common substructures are subsequently paired to produce a new set of common structures. This is repeated until no new substructures could be discovered. The resulting frequent substructures are then ranked by the number of binding sites they contain. In the MCD approach, a graph theory representation of each binding site is used, where labels for vertices are chemical elements, and for edges are the pairwise distances between the atoms (covalently bonded or not). Graphs from two binding sites are then compared. As shown in Figure 6.3, a correspondence graph (G_C) of two binding sites is constructed (Figure 6.3b), where each vertex in G_C represents an atom from first and second binding site, so long as they possess the same labels (i.e., same chemical elements). Edges are assigned between two vertices in G_C if the corresponding atoms from both the binding sites possess the same distance (within a given tolerance). Once the correspondence graph is constructed, maximum clique detection (red circle in Figure 6.3b) is performed to find the common set of atoms between two binding sites.

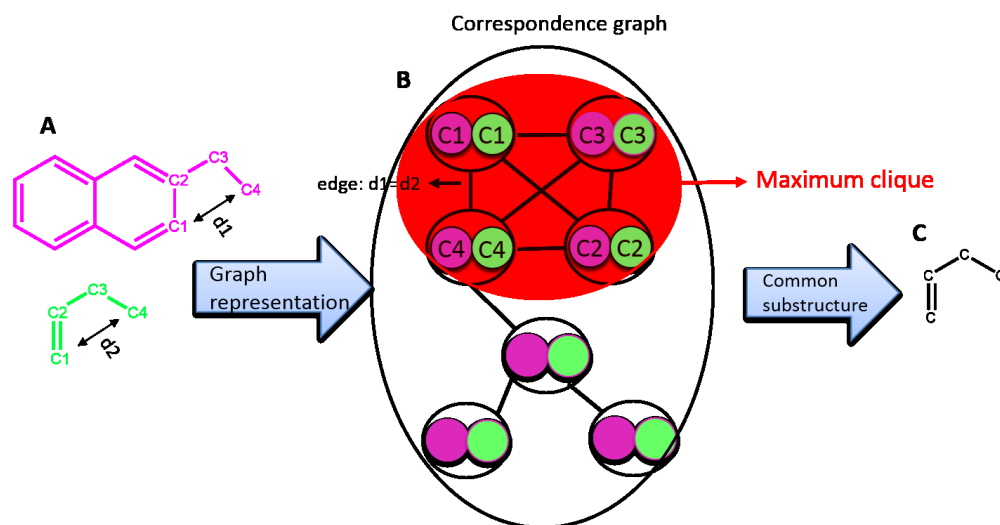


Figure 6.3: A) Examples of two binding sites and B) the correspondence graph of the two binding sites. C) The largest common substructure of two binding sites is obtained from the maximum cliques shown within the red circle in B.

The minimum number of atoms allowed when finding a common substructure (maximum clique) between two binding sites was set to 14 in this study. If a too small minimum is used, then too many common fragments will be identified. The end-point of the search of adsorbophores is reached when no new single substructure is found after 200 consecutive pairings and the program ceases to randomly pair binding sites. As the random pairing method does not scan every possible substructure between binding sites, the most common *substructures* discovered from this method were then examined in an exhaustive manner using the same clique detection method detailed above. This way, the total number of binding sites that possessed each *common substructure* was recognized.

6.2.5 GCMC Simulation Details

Gas uptake properties in all of the hypothetical MOFs were predicted using an in-house GCMC simulation code at fixed chemical potential, volume and temperature. The constant chemical potential was calculated using the ideal-gas equation of state for pressures up to 1

bar, while for higher pressures the Peng-Robinson equation of state was used with binary interaction parameters.³⁰ All simulations included random insertion, deletion, translation, and swap moves of molecules with equal probabilities. During the simulation, framework atoms were held fixed at their crystallographic positions and the guests were treated as rigid molecules. The simulation box for each hypothetical MOF was created in which half the length of the smallest orthogonal distance in the box was larger than the long range interaction energy cut-off, 12.5 Å. All interaction energies between non-bonded atoms were computed through the Lennard-Jones potential for dispersion and repulsive interactions and a pairwise point charge Coulombic potential for the electrostatic interactions. LJ parameters for framework atoms were taken from the Universal Force Field (UFF).³¹ LJ parameters between atoms of different types were calculated using the Lorentz-Berthelot mixing rules (i.e., the geometric average of well depths and the arithmetic average of diameters). Charges on the framework atoms were computed using an in-house code in which the Rappe-Goddard charge equilibration method parameters are trained to reproduce the electrostatic potential derived from the DFT calculation in a large diverse set of 543 hypothetical MOFs.²⁰ Regarding the guest molecules, The LJ parameters of CO₂ were taken from García-Sánchez et al. who optimized them from a self-consistent fitting of the experimental adsorption isotherm data in large class of zeolite frameworks with the predictions of Gibbs ensemble Monte Carlo calculations on these systems.³² The CO₂ molecule had a fixed bond length of 1.16 Å and partial atomic charges of +0.645e and -0.3325e for carbon and oxygen, respectively. LJ parameters and partial charges for CH₄ were taken from the TraPPE force field,³³ where the CH₄ molecule was modeled as a

single sphere. This force field has been fit to reproduce the experimental vapor-liquid phase equilibrium data for methane.

Binding site energies were calculated using the binding site coordinates identified from the ABSL algorithm and equation 6.5:

$$E_{binding\ site} = E_{CO_2+MOF} - E_{MOF} - E_{CO_2} \quad 6.5$$

where E_{CO_2} corresponds to a configuration energy of only CO_2 , E_{MOF} is a potential energy of a MOF which does not contain a guest molecule, and E_{CO_2+MOF} is a energy of MOF and a bound CO_2 with no other guest molecules present. If the binding site energy is calculated with a periodic DFT method, then E_{CO_2} is determined from a calculation of an 'isolated' CO_2 molecule in the same unit cell as that of the MOF.

As mentioned in the section outlining the screening protocol, several high performing MOFs were subject to high-level DFT calculations. All DFT calculations were carried out in the Vienna *ab initio* simulation package (VASP).^{33,34,35} The PBE exchange-correlation functional³⁶ and PAW potentials^{37,38,38} were used with a plane-wave energy cut-off of 520 eV. All DFT calculations used a single unit cell with a γ -centered $2 \times 2 \times 2$ k-point grid on the Brillouin zone. The semi-empirical dispersion potential correction of Grimme³⁹ was included. A conjugate gradient algorithm was used for relaxing the atoms and lattice parameters. The lattice configuration was fully optimized when all forces were smaller than 2×10^{-4} eV/Å.

6.3 Results & Discussion

Our high-throughput screening protocol, summarized in Figure 6.1, enables us to compute CO_2 working capacity and CO_2/CH_4 selectivity of 324,000 MOFs, and to identify key

structural and chemical features of high performing MOFs from screening data for methane purification, which can be further useful for future experimental synthesis efforts. In the following sections, we investigate the structure-property relationship of MOFs for CO₂/CH₄ separation, explore common binding motifs among the high-performing MOFs, and finally identify 'hits' or potential target materials for synthesis.

6.3.1 CO₂-CH₄ Screening

Figure 6.4 shows a scatter plot of the GCMC computed working capacity vs selectivity for entire the database. There is no simple linear relationship between working capacity and selectivity, except that there seems to be a maximum working capacity of about 8 mmol/g with a selectivity of about 10. Ideally we are looking for a material with both large working capacity and high selectivity. The shaded portion of the scatter plot in Figure 6.4 include MOFs whose performance is comparable or better than that of zeolite 13x and were categorized as high performing MOFs.

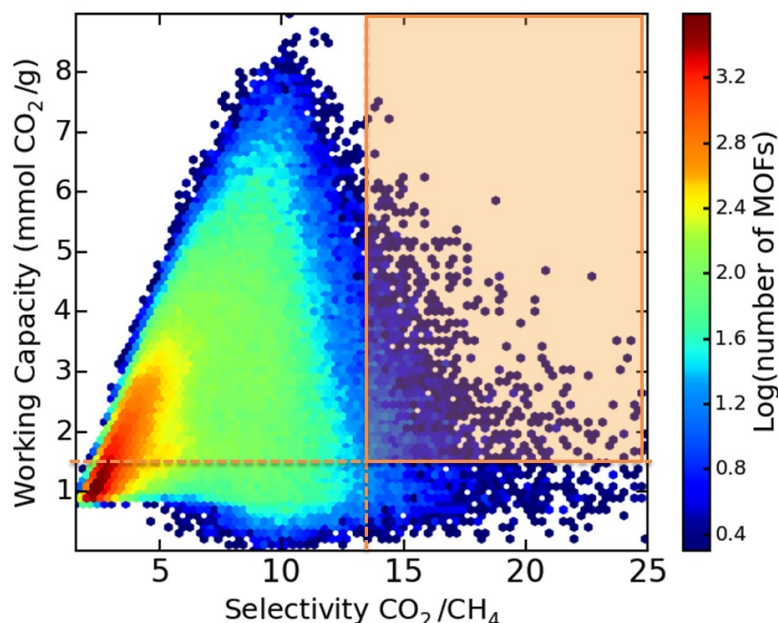


Figure 6.4: A color map of the 324,474 MOFs screened for working capacity and CO_2/CH_4 selectivity with GCMC simulations. A pink shaded zone in the upper right quadrant represents MOFs whose selectivity and working capacity are both comparable to those of zeolite 13x.

6.3.2 Structure-property relationships

An interesting aspect of large-scale screening of MOFs for gas-related applications is a possibility to obtain structure-property relationships (i.e., correlations between the performance and structural properties), which can be used for further guiding the design of high-performing MOFs. For example, Wilmer and Snurr¹⁷ screened $\sim 130,000$ hypothetical MOFs for methane storage and from the screening data found that the optimal pore size for the high performing MOFs is $\sim 8 \text{ \AA}$. In the follow-up study, the same database was screened for CO_2/N_2 separation, and it was found that the high performing MOFs would have surface area in the $1000\text{-}2000 \text{ m}^2/\text{g}$ range.²¹ Here, we expand the scope of these previous studies to include over 300,000 hypothetical MOFs discussed in Section 6.2.1, and present structure-property relationship of these MOFs for natural gas purification. We characterized the geometric

features of the MOFs using the zeo++ code⁴⁰ by calculating void fraction (i.e., the fraction of empty space within a MOF) with the use of Helium probe atom, maximum pore diameter (i.e., the diameter of the largest sphere that can fit within the pores), and surface area, all of which have received significant attention in the MOF literature.^{41,42}

The working capacity vs selectivity plots with color coding based on different geometric properties of the MOFs including (a) void fraction, (b) surface area, and (c) max-pore diameter are provided in Figure 6.5. The rectangular orange zone in all plots represents a region of MOFs which possess working capacity and selectivity above zeolite 13x, and illustrates the increase in performance offered by MOF-based PSA technology for natural gas purification. We consider zeolite 13x as a reference because it is used commercially in PSA-based technology for methane purification.⁴³

Figure 6.5a shows that for MOFs with selectivity around 10, working capacity increases with increasing void fraction up to 0.36. Meanwhile, selectivity decreases below 5 with increasing void fraction above 0.4, which is attributed to the fact that MOFs with higher void fractions have additional room to be occupied by methane molecules at the adsorption pressure of 9 bar. In contrast, lower void fractions (in the 0.1- 0.2 range) correspond to higher selectivity. Figure 6.5a also shows that working capacity drops off whenever the void fraction is too large or too small. Similar to the void fraction, whenever a MOF's surface area is too large or too small, it has a negative effect on working capacity (Figure 6.5b). Selectivity increases above 10 when surface area is smaller than $\sim 3000 \text{ m}^2/\text{g}$. Finally, Figure 6.5c depicts a similar pattern where a higher value of selectivity corresponds to the lower max-pore diameter (in the 4-6 Å range) and working capacity increases beyond 2 mmol/g for pores that are no larger than

8 Å. However, while such good discrimination shows limiting (i.e., highest and lowest) geometrical features essential for constructing high-performing MOFs, one can readily observe from Figure 6.5 that structural/geometrical features alone are not sufficient factors for discriminating high performing and non-high performing MOFs. For example, as shown in Figure 6.6, while high selectivity (>10) is observed for MOFs with pores in the 4 - 6 Å range, a negligible number of MOFs with the same pore size are also found to exhibit selectivities between 5 and 10.

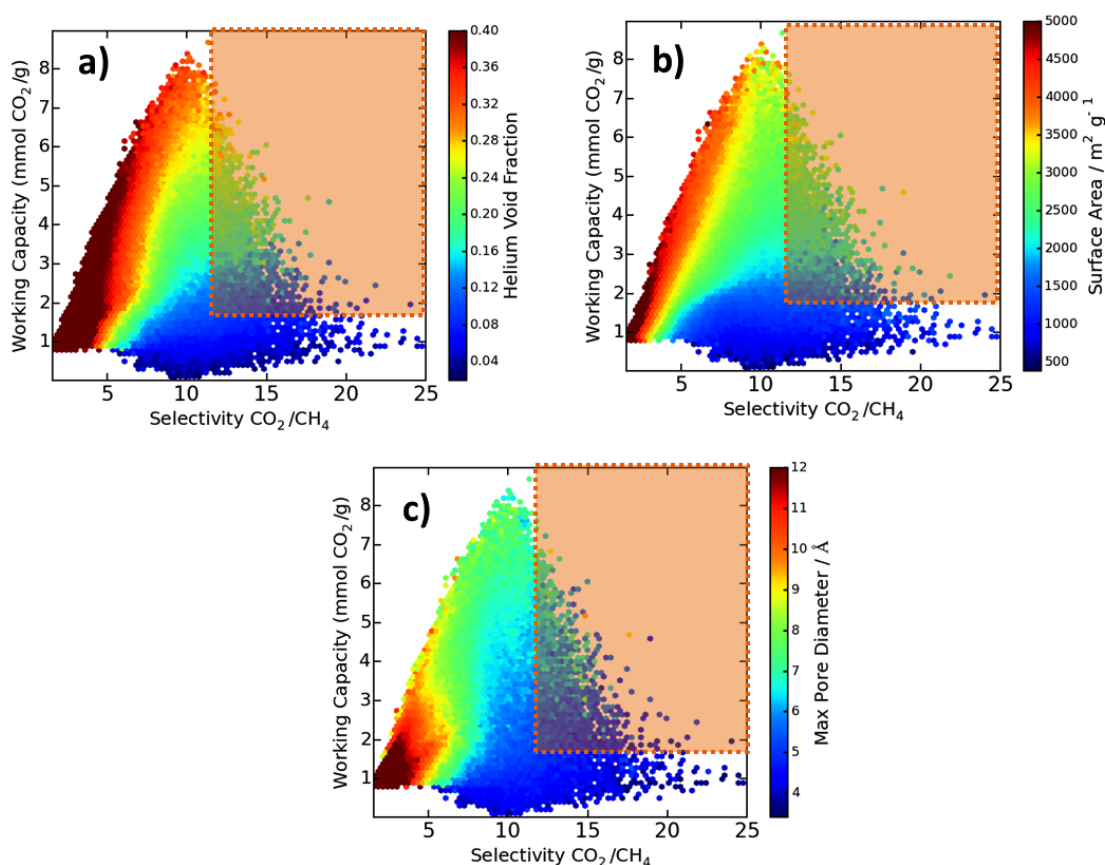


Figure 6.5: Scatter plots of CO₂ working capacity vs CO₂/CH₄ selectivity for 324,000 hypothetical MOFs color-mapped according to (a) void fraction determined by helium probe atom, (b) surface area and (c) max-pore diameter. The rectangular orange zone in the upper right hand corner of all three figures represents the region of MOFs which possess working capacities and selectivities above zeolite 13X.

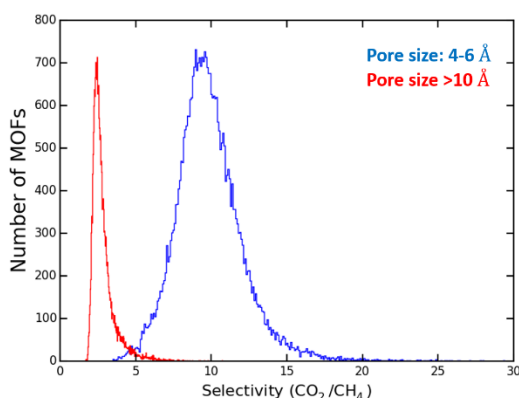


Figure 6.6: Distribution of CO_2/CH_4 selectivity of MOFs with the pore size above 10 (bred) and the pores size in in the range 4-6 Å (blue).

In addition to the geometrical features, we have also observed that the heat of adsorption of CO_2 (Q_{st}) strongly influences the CO_2 separation ability. As shown in Figure 6.7, CO_2/CH_4 selectivities are positively correlated with Q_{st} - although notably a few data points at very high Q_{st} values have selectivities around 8. As with the geometrical features, working capacity significantly decreases whenever Q_{st} is too large or too small. Overall, our data suggest that high-performing MOFs would have Q_{st} between 8-9 kcal/mol. However, it is difficult to design MOFs based on a chosen Q_{st} value. In fact, it is easier to design MOFs based on a chosen geometrical features, or based on a particular chemical feature.

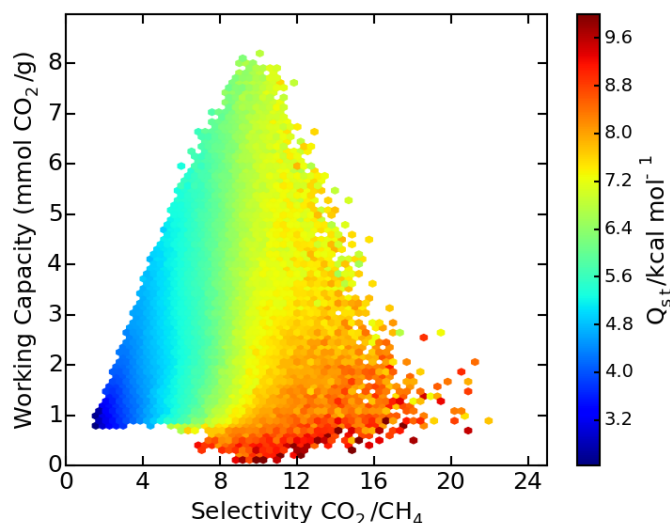


Figure 6.7: Scatter plots of CO₂ working capacity vs selectivity CO₂/CH₄ of 324,000 hypothetical MOFs with a colour map of the CO₂ heat adsorption (Q_{st})

Another important feature that can affect a MOF's performance is the presence of functional groups. To study the effect of functional groups in MOFs on the performance with regards to CO₂ separation, we calculated the frequency of finding functional groups in MOFs with working capacity higher than 2 mmol/g as a function of selectivity of CO₂/CH₄. Figure 6.8 shows that MOFs with NO₂ and HCO functional groups dominate high performers; nearly 50 % of MOFs which have selectivity greater than 12 contain NO₂ and HCO functional groups. This is likely due to the permanent dipole moments in the vicinity of NO₂ and HCO groups, which is favorable for interactions with the quadrupole of CO₂.

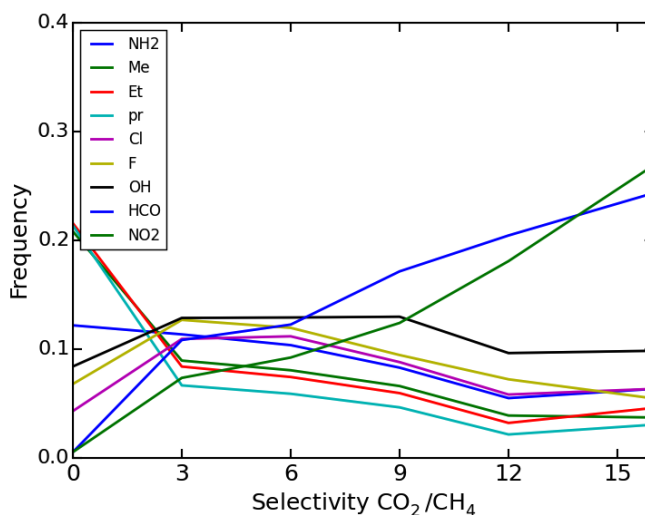


Figure 6.8: Frequency of selected functional groups present in the database as a function of selectivity for MOFs with working capacity above 2 mmol/g.

As a complement to above analysis, the weighted radial distribution function (RDF) was calculated for CO₂ binding sites vs. functional groups in the high-performing MOFs. Here, CO₂ binding sites were identified through ABSL analysis of the CO₂ probability distribution obtained from GCMC calculations using REPEAT charges. The occupancy of the CO₂ binding sites was used to determine the number of CO₂ molecules within a specific radius of a functional group, and the following equation was used to calculate the RDF of ABSL-derived CO₂ binding sites around center mass of each type of functional group;

$$g_{ij}(r) = \frac{\Delta N_{ij}(r, r + \Delta r)V}{4\pi r^2 \Delta r N_i N_j} \quad 6.6$$

where r is the distance between i (CO₂ binding site) and j (functional group), $\Delta N_{ij}(r, r + \Delta r)$ is the number of species j around i within a shell from r to $r + \Delta r$, V is the system volume, and N_i and N_j are the numbers of species i and j .

Figure 6.9 shows the RDF of CO₂ binding sites around a selection of functional groups averaged over 3990 high-performing MOFs. The peak height in the $g(r)$ indicates that the NO₂ and HCO functional groups are preferred functional groups in our high-performing MOFs, which is consistent with our findings from Figure 6.8.

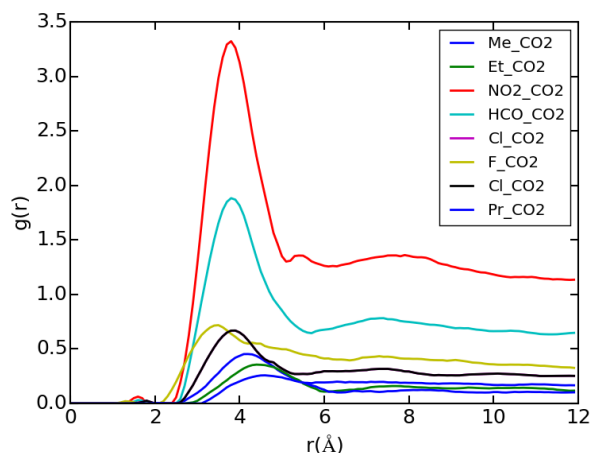


Figure 6.9: Averaged weighted radial distribution functions $g(r)$ of CO₂ binding sites for selected functional groups found in over 3990 high-performing MOFs.

With respect to the preferred functional groups, our findings are inconsistent with the study of Wilmer et al. who found that fluorine and chlorine are the best functional groups for scrubbing CO₂. A reason for this discrepancy is related to the way we calculated the non-bonded interactions. For CO₂, a significant portion of binding energy comes from the electrostatic interaction between CO₂ and the MOF's framework atoms. This interaction depends on the MOF's partial atomic charges, which in this work was computed by the MEPO-QEq method as opposed to the EQEq method used in their studies. Woo and coworkers recently reported that partial atomic charges assigned to halogen atoms in the study of Wilmer et al.²¹ using the EQEq method were typically much larger than those derived from DFT calculations which subsequently results in an over-estimation of the CO₂ uptake.²⁰ To make

sure that our high-performing MOFs were precisely modeled, a comparison of CO₂ uptake determined from MEPO-QEq and ab initio REPEAT is given in Figure 6.10. The slope ($m=0.87$) of the best fit line illustrates that MEPO-QE method tends to underestimate CO₂ uptake, by an average of 1.2 mmol/g, which further validates our modeling.

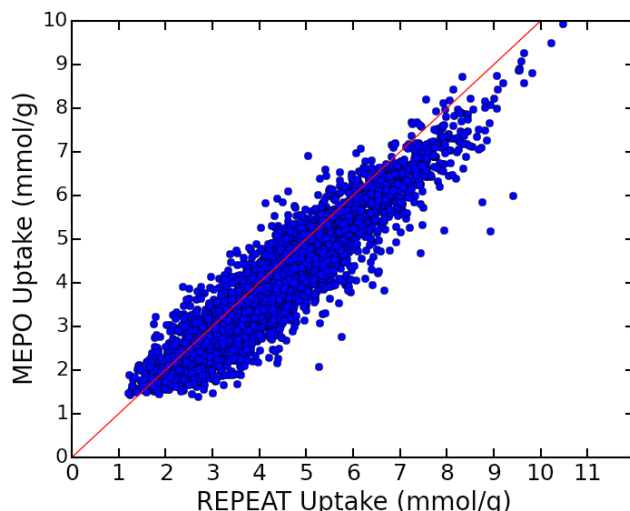


Figure 6.10: CO₂ gas uptake and at 1 bar, 298 K for high-performing MOFs obtained from the GCMC simulations using MEPO-QEq charges as compared with those obtained using REPEAT charges. The line of perfect correlation is shown in red.

6.3.3 Adsorbophore Analysis

As elucidated in Section 6.3.2, a pore size lower than 6.6 Å and a surface area higher than 2299 m²g⁻¹ are essential, but not sufficient, factors for MOFs to be high performing materials for natural gas purification. It is also found that MOFs with NO₂ and HCO functional groups dominate high performers, however, these functional groups are seen in low-performing MOFs as well (Figure 6.8). As a complement to the previous analyses, this section discusses the discovery of two binding site motifs (adsorbophore) of CO₂, S1 and S2, shown in Figure 6.11. These adsorbophores are common substructures among the binding sites of all high performing

MOFs, which were discovered using maximum clique detection methods developed in the Woo lab (see details in the Methods Section).

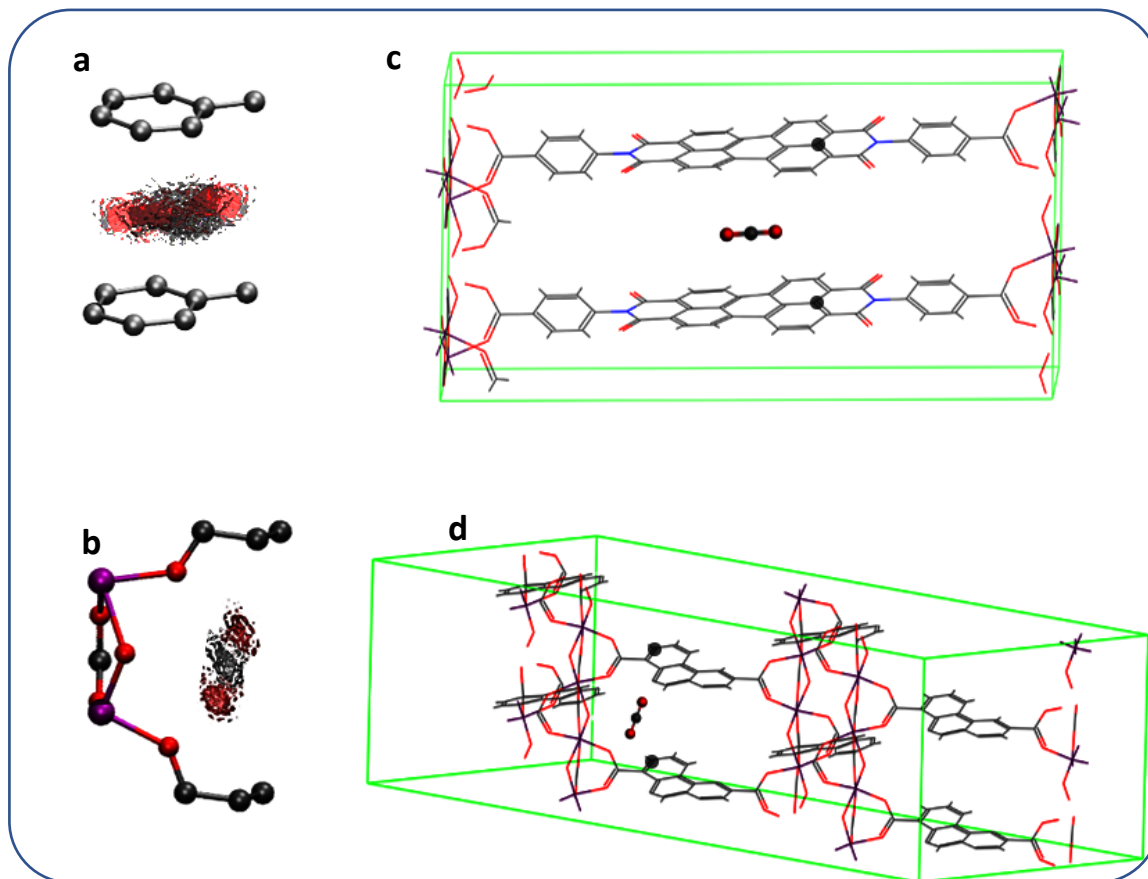


Figure 6.11: a,c) Two most common binding site motifs identified with a maximum clique analysis of the high performing MOFs. The distribution of the CO₂ molecules within the binding sites are shown as the isosurfaces (red for oxygen and gray for carbon). b) and d) show examples of each binding site motif within a representative hypothetical MOF. Gray: carbon, purple: vanadium, red: oxygen, blue: nitrogen, white; hydrogen.

Binding site motif S1 can be described as two parallel aromatic rings, separated roughly by 7.2 Å, with two carbons from the extended ligands (shown in figure 6.11c with black color). The average CO₂ binding energy with binding site S1 was determined to be -8.5 kcal/mol, where ~95% of the energy is governed by dispersion interaction. A question that immediately arises from this observation is whether methane molecules can diffuse into the S1 binding motif? In

other words, is the selectivity driven only by high binding affinity for CO₂ or is it also driven by some sort of size selectivity (i.e., 7.2 Å)? To examine this, we looked at the probability distributions of pure CO₂ and CH₄ in a set of high performing MOFs containing the S1 binding motif, and in Figure 6.12 we provide isosurface plots of the probability distributions of CO₂ and CH₄ in one of them. As shown in Figure 6.12a, CH₄ can diffuse into the S1 binding motif, but its binding energy with the S1 is -4.2 kcal/mol, which is significantly less than the CO₂ binding energy with the S1 (-8.2 kcal/mol). This indicates that the selectivity is governed by high binding affinity for CO₂. Figure 6.12 also shows that, unlike CO₂, the methane molecules do not show definite adsorption sites in the MOFs framework, and multiple diffuse adsorption sites are observed. This is due to the fact that methane has weaker dispersion interactions and does not have electrostatic interactions.

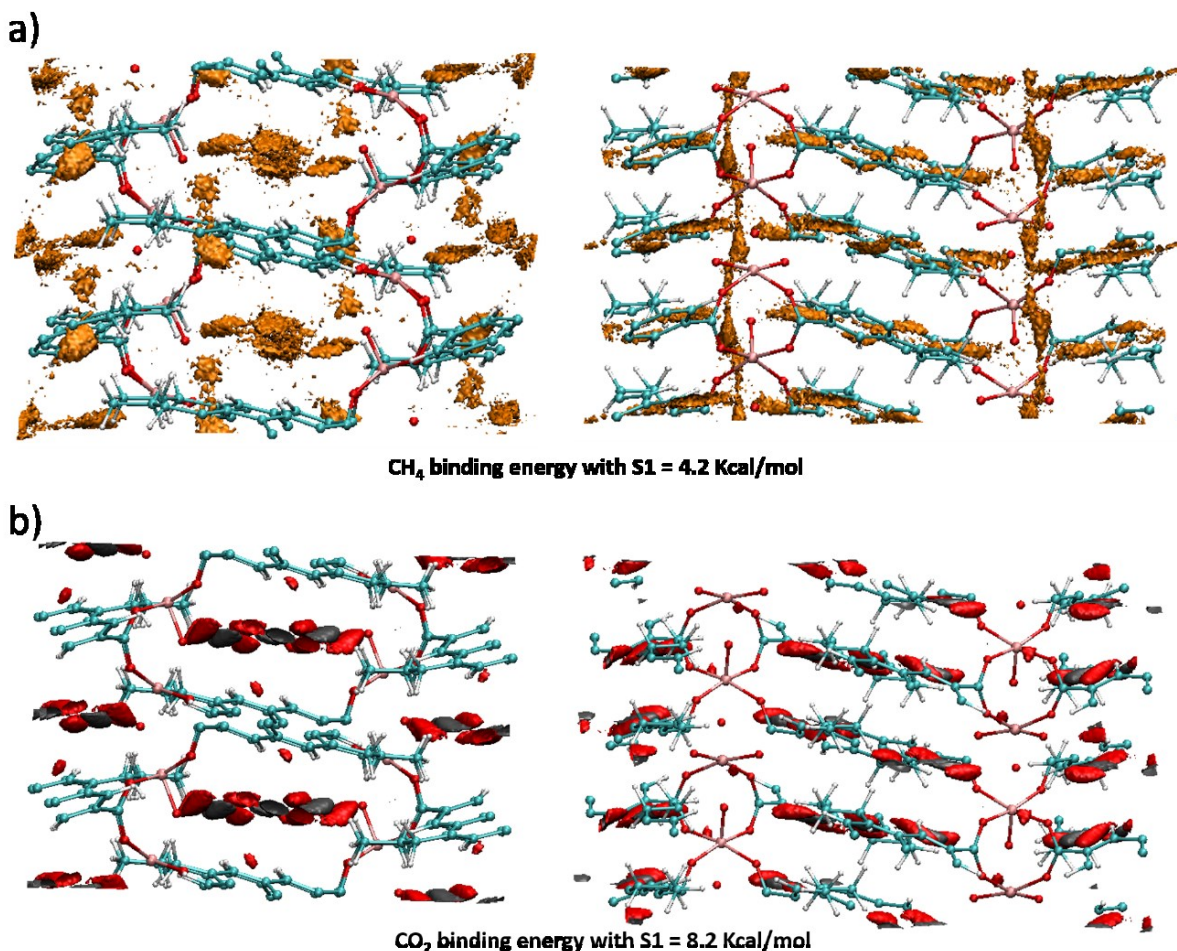


Figure 6.12: a) Isosurface plots of the probability distribution of CH₄ in a high performing MOF containing S1 binding motif at isovalue 0.01; front view (right) and side view (left). b) Isosurface plots of the probability distribution of CO₂ in the same MOF at isovalue 0.03; front view (right) and side view (left).

Binding site motif S2, shown in Figure 6.11b, involves the pillaring V₂O₂ SBU of MIL47 and is unique to hypothetical MOFs constructed with this SBU. In comparison to S1, the distribution of CO₂ molecules in the binding sites are much more localized. The binding energy was computed to be on average, -9.5 kcal/mol where 20% of this energy is electrostatic. This electrostatic portion is governed by the positively charged carbon atom of CO₂ interacting with the negatively charged bridging oxygen between two vanadium atoms. The two end carbons in this binding motif, separated by roughly 7 Å, correspond to the linear 2-connected organic

SBU, which extend from the vanadium SBUs. Notably, although this motif compared to the S1 motif provides more binding energy for CO₂ (-9.5 kcal/mol vs -8.5 kcal/mol), it is less common. This is due to the fact the S1 motif is unique to hypothetical MOFs constructed with V₂O₂ SBU, which constitutes 10% of total high performing MOFs, while the S1 motif is seen across different topologies and inorganic SBUs within the database (see Section 6.3.4).

The average binding energies reported were calculated with the classical force field which was used to perform the GCMC simulations. To ensure that the calculated binding energies using this force field are reasonable, we computed the binding energies of the two binding site motifs within the representative MOFs shown in Figure 6.12 with dispersion corrected periodic DFT calculations. The DFT-derived binding energies were -8.2 and -9.1 kcal/mol for S1 and S2, respectively, indicating a good agreement with the force-field derived binding energies of -8.3 for S1 and -9.3 kcal/mol for S2.

Binding motifs S1 and S2 were defined by a substructure matching algorithm where the size of the substructures was limited to 14 atoms. One interesting question is, how much of the total binding energy do the 14 atoms contributed to the total CO₂ binding energy? With a force field, the total binding energy is approximated with pairwise Lennard-Jones terms and electrostatic terms. As a result, we can examine how much of the total binding energy results from the 14 atoms of the S1 and S2 substructures. Averaging over all S1 binding sites identified, we found that 90% of the total binding site energy comes from the 14 atoms of the S1 substructure. For the S2 binding site, the same analysis shows that 75% of the total binding energy can be contributed to the 14 atoms of the S2 substructure. This indicates that the

binding of CO₂ is mostly localised within the S1 and S2 binding motifs, and the remainder of the framework does not contribute significantly to the CO₂ binding.

6.3.4 Conjugated Aromatic-Based High Performing MOFs

Motivated by the S1 binding motif which would be interesting from the perspective of MOF design, we looked at the high performing MOFs containing the S1 sorption pocket for further investigation. We found that the common feature among these MOFs is the one dimensional channels of *layered sheets* consisting of poly-aromatic molecules separated by a perpendicular distance of roughly 7 Å. The three examples of these MOFs provided in Figure 6.13 show that this feature is found in different classes of MOFs (i.e., different inorganic SBUs and different topology) within the database. This observation proposes that the important factor relevant to these high-performing MOFs is the optimal distance of ~7 Å between two poly-aromatic organic SBUs.

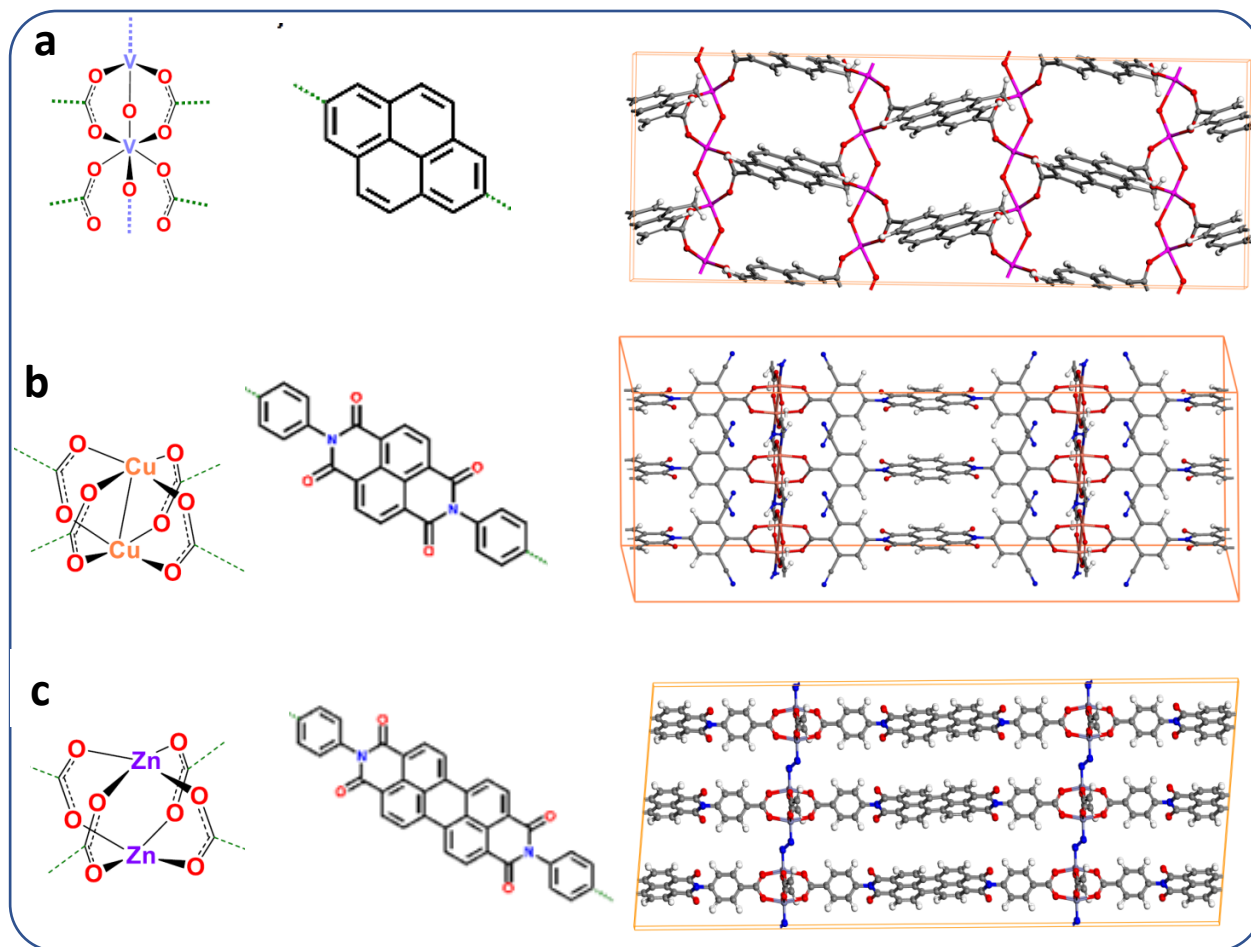


Figure 6.13: Three examples of high-performing MOFs containing 1-D sheet-like channels spaced by 7 Å with planar poly-aromatic organic SBU: a) A MOF with sra topology which is assembled of the conjugated organic SBUs coordinated to 1-dimensional chains of vanadium oxide, and b & c) MOFs with pcu topology in which the large poly-aromatic organic SBUs are connected to hexacoordinate copper and zinc paddlewheels.

From the above observation, some questions that arise are –how many of these structures are not actually high-performing? Can we identify these structures only through physical properties? To answer these questions, we looked at MOFs in the database which were built from metal and organic SBUs which had a potential to form layers of poly aromatic molecules with the ideal spacing (7-7.4 Å). This gave a total of 31,000 MOFs. Of these, 987 MOFs are recognized as “high-performers”, which is 30% of all high performing MOFs. The

remaining 30,193 MOFs are flagged as low-performing materials, which constitutes 9% of the low-performing MOFs in our database. This distribution indicates that having a poly aromatic ligand does not guarantee a MOF to be high-performing. Note that these structures are generated through the functionalization of a set of base structures, where symmetrical equivalent hydrogen atoms are replaced with functional groups. Following the functionalization, these structures undergo geometry optimization which yields pores with different shapes and sizes, thus resulting in significant variability in performance. Figure 6.14 shows the CO_2/CH_4 selectivity of this set of MOFs as a function of the maximum pore diameter with color mapped S1 density. Here, the S1 density across the MOFs was computed by finding the instances of S1 binding motif in each MOF using the maximum clique detection algorithm discussed in the Methods section, and by converting the number of S1 binding motifs found in each MOF to a molar value and then by normalizing by the MOF's unit cell volume. As shown in Figure 6.14, MOFs with a maximum pore diameter larger than 8 Å have CO_2/CH_4 selectivity below 10, and their S1 density are 0. In contrast, MOFs with a maximum pore diameter in the range of 4-4.5 Å show a sizeable distribution of the selectivity from 2 to 40. The selectivity increases above 12 for MOFs with S1 density higher than 0.8 mmol/cm³, and the selectivity drops below 10 when the S1 density is close to 0. Such behaviour in the performance indicates that having maximum pore diameter between 4-4.5 Å is necessary but not a sufficient condition for high performance. The maximum pore diameter is the largest sphere that could be grown inside the framework. Since the van der Waals radii of aromatic carbons is 1.7 Å, the peak at 4Å in Figure 5.16 correspond to a spacing of 7.4 Å in the S1 binding motif.

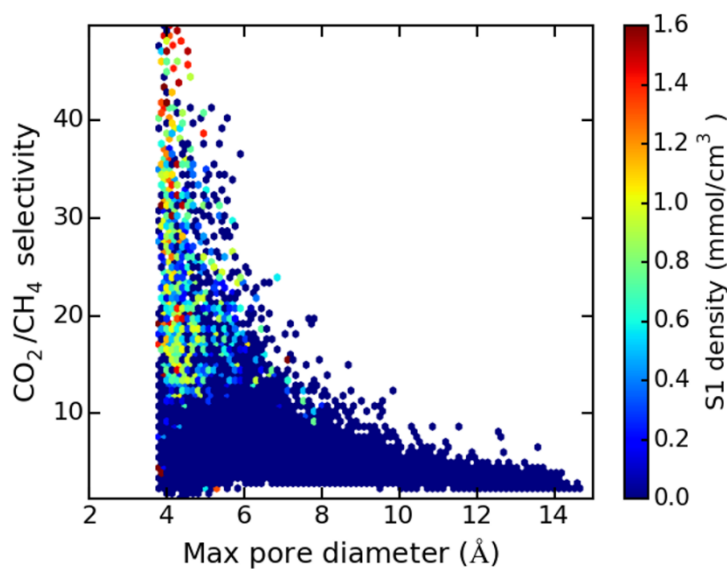


Figure 6.14: CO_2/CH_4 selectivity of 31,180 MOFs identified as possible layered sheet-like MOFs vs. the maximum free pore diameter with a color mapped S1 density.

To simultaneously see the correlation of the S1 binding motif with the selectivity and working capacity, we plot the selectivity of 31,180 MOFs identified as possible layered sheet-like MOFs as a function of working capacity with a color mapped S1 density in Figure 6.15. The MOFs with selectivity above 14 and working capacity above 1.68 mmol/g that constitute the green zone in figure are the high performing MOFs. From the color map in Figure 6.15, it can be seen that the presence of S1 is a good indicator for MOFs to be high-performing since there are no low performing MOFs with S1 density greater 0.4 mmol/cm³. This further demonstrates the importance of the S1 binding motif in the rational design of MOFs for natural gas purification.

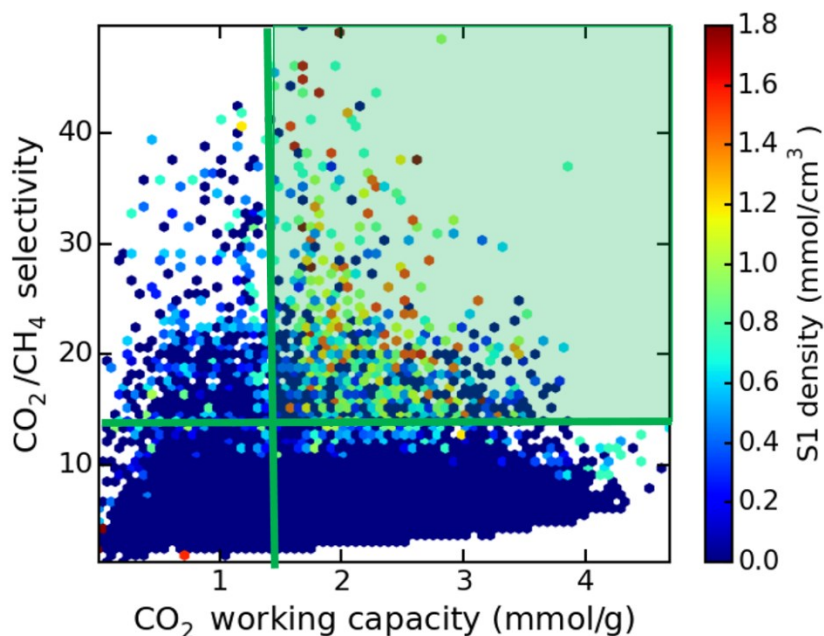


Figure 6.15: CO₂/CH₄ selectivity of ~31,000 MOFs identified as possible layered sheet-like MOFs vs. the working capacity with a color mapped S1 density.

6.3.5 *Ab initio* geometry optimization of the top performers

The adsorption properties of materials are highly dependent on the geometry of material. The geometry of the MOFs in our hypothetical database were refined via a molecular mechanics force field optimization. Although DFT optimized geometries are more accurate, it was not feasible to perform the optimizations at this level on the whole database. In this section, we explore the sensitivity of the adsorption properties on the geometry by optimizing the 50 top performers from our screening at the DFT level and comparing the adsorption properties before and after DFT optimization. We note that both the nuclear coordinates and the cell geometries have been optimized.

Table 6.1: CO₂ working capacities (WC) and CO₂/CH₄ selectivities (S) of the top high-performing MOFs optimized at the MM and DFT level. Each MOF in the table is described by the functional group and SBUs which were used to build them.

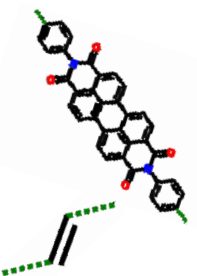
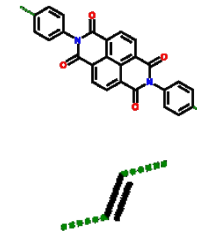
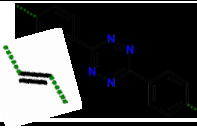
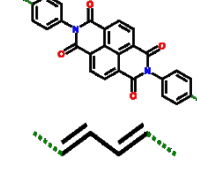
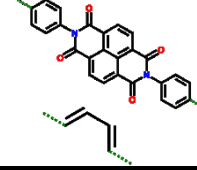
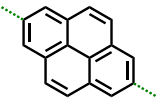
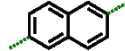
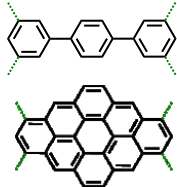
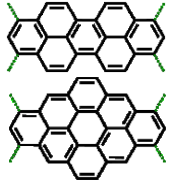
Organic SBUs	Metal SBU	Functional Group	Class	MM Optimized Structure		DFT Optimized Structure		Δ WC (mmol/g)	Δ S
				WC (mmol/g)	S	WC (mmol/g)	S		
	V ₂ O ₂	Cl	A	2.6	24.8	3.48	23.2	0.88	-1.6
	V ₂ O ₂	NO ₂	A	3.54	25.0	2.18	18.3	-1.36	-6.7
	V ₂ O ₂	OMe	A	3.19	20.9	2.7	30.9	-0.49	10
	V ₂ O ₂	F	A	3.36	39.4	4.22	24.4	-0.86	-15
	V ₂ O ₂	HCO	A	4.11	28.7	2.22	23.2	-1.89	-5.5
	V ₂ O ₂	HCO	A	3.35	22.5	3.72	20.2	0.37	-2.3
	V ₂ O ₂	NH ₂	A	3.85	23.8	2.91	15.6	-0.94	-8.2
	V ₂ O ₂	SO ₃ H	A	3.02	26.2	2.77	10.5	-0.25	-10.7
	V ₂ O ₂	OH	A	3.67	33.1	1.72	19.2	-1.95	-13.9
	V ₂ O ₂	COOH	A	3.9	31.2	3.57	22.1	-0.33	-9.1
	V ₂ O ₂	Cl	A	2.61	24.7	2.1	18.3	-0.51	-6.4
	Cu ₂	I	A	2.7	23.2	2.51	23.4	-0.19	0.2
	Cu ₂	I	A	2.4	33.1	1.93	29.3	-0.47	-3.8
	Zn ₂	Et	A	4.2	36.4	3.23	20.1	-0.97	-16.3
	Cu ₂	CN	A	3.87	40.9	2.24	15.3	-0.16	-25.6
	V ₂ O ₂	CN	A	4.3	27.7	1.9	13.5	-2.4	-14.2
	Cu ₂	OPr	A	3.00	33	2.00	20.1	-1	-12.9
	Cu ₂	I	A	3.4	32	3.17	28.2	-0.23	-3.8
	Cu ₂	CHO	A	3.78	88	2.07	19	-1.69	-69
	Cu ₂	NHMe	A	3.01	33	2.0	19.9	-1.01	-13.1
	V ₂ O ₂	Et	A	2.9	33.7	2.3	24.4	-0.6	-9.3
	Cu ₂	OH	B	3.8	27.6	3.9	35.3	0.1	7.7
	Zn ₂	CHO	B	4.0	28.8	2.7	17.9	-1.3	-10.9
	Zn ₂	SO ₃ H	A	2.84	37.8	2.6	11.3	-0.24	-26.5
	Cu ₂	OEt	A	3.71	24.2	3.38	18.4	-0.33	-5.8
	Cu ₂	NHMe	A	4.52	24.9	3.37	16.6	-1.15	-8.3
	Zn ₂	SO ₃ H	A	3.5	26.2	2.22	14.5	-1.28	-11.7
	Cu ₂	SO ₃ H	A	3.3	32.4	3.0	14.2	-0.3	-18.2
	V ₂ O ₂	Br	A	3.3	27.9	3.1	26.7	-0.2	-1.1
	V ₂ O ₂	NO ₂	A	3.7	21.2	2.4	22.0	-1.3	1.8
	V ₂ O ₂	OMe	A	2.8	22.8	1.9	7.7	-0.9	-15.1
	V ₂ O ₂	OH	A	3.0	21.4	1.8	9.0	-1.2	-12.4
	V ₂ O ₂	Me	A	3.9	25.2	1.4	4.9	-2.5	-20.3
	V ₂ O ₂	CHO	A	4.01	24.5	1.43	3.7	-2.58	-20.6

Table 6.1. Continued.

Organic SBUs	Metal SBU	Functional Group	Class	MM Optimized Structure		DFT Optimized Structure		ΔWC (mmol/g)	ΔS
				WC (mmol/g)	S	WC (mmol/g)	S		
	V ₂ O ₂	OH	B	2.7	31.5	3.0	32.5	0.3	1.0
	Cu ₂	Br	B	3.9	22.4	2.1	30.2	-1.8	7.8
	V ₂ O ₂	OH	B	2.9	37.9	3.0	21.0	0.1	-16.9
	V ₂ O ₂	OH	A	3.2	40.4	1.15	5.3	-2.05	-35.1
	V ₂ O ₂	Pr	A	2.23	42.6	2.36	41.1	0.13	-1.5
	Zn ₂	H	A	3.6	20.9	2.63	16.2	-0.07	-4.7
	V ₂ O ₂	OH	A	3.2	40.4	1.15	5.3	-2.05	-35.1
	V ₂ O ₂	F	B	2.8	32.8	3.1	23.8	0.3	-9.0
	V ₂ O ₂	COOH	B	4.02	29.3	3.8	24.1	0.22	-5.2
	Cu ₂	OH	B	2.5	28.8	3.7	24	1.2	-4.8
	Cu ₂	SO ₃ H	B	3.97	29.7	3.23	23.3	-0.74	-6.4
	Zn ₂	NO ₂	B	3.44	28.2	3.80	19	0.36	-9.2
	Cu ₂	NO ₂	B	2.72	23.2	3.09	17.7	0.37	-5.5
	Zn ₂	NH ₂	B	3.01	23.2	3.41	21.8	0.4	-1.4
	Zn ₂	CHO	B	2.30	28.1	2.29	20.1	-0.01	-8.0
	Zn ₂	NO ₂	B	4.20	25.2	3.90	20.1	-0.3	-5.1
Average CO₂ working capacity for MM-optimized structures: 3.4 mmol/g Average CO₂ working capacity for DFT-optimized structures: 3.1 mmol/g Average CO₂/CH₄ selectivity for MM-optimized structures: 29.2 Average CO₂/CH₄ selectivity for DFT-optimized structures: 19.7									

Tables 6.1 compares the calculated adsorption properties determined with an MM optimized structure to those determined with a DFT optimized structure of the top 50 performing MOFs, where Class A refers to 1-D sheet-like MOFs and Class B refers to MOFs that

are not composed of layered sheets. The DFT-optimized structures of the top 50 performing MOFs have an average CO₂ working capacity of 3.1 mmol/g, which is a slight drop from that of the MM-optimized structure (3.4 mmol/g). Unlike the CO₂ working capacity, the average CO₂/CH₄ selectivity drops more significantly from 29.2 with the MM-optimized structures to 19.7 for the DFT optimized structures. As given by ΔS in Table 6.1, the decrease in the selectivity is mostly contributed by Class A materials (i.e., 1-D sheet-like MOFs), which suggests that the S1 binding site is sensitive to small changes in the geometry. Examination of the geometries of the layered sheet MOFs listed in Table 6.1 before and after DFT optimization reveals that parallel alignment of the aromatic rings or the optimal separation distance of 7 Å between the aromatic rings is sometimes lost after DFT optimization. An example of this is shown in Figure 6.16, where upon DFT optimization, the aromatic sheet structure becomes distorted from the idealized planar structure. It is notable however, that the layered-sheet MOFs that retained their performance following DFT optimization were decorated by bulky functional groups such as iodine and propyl. The bulky functional groups may help to preserve the pore shape of MOFs after DFT optimization, and they may make good synthetic targets as the optimal pore geometries appear to be more robust with these functional groups.

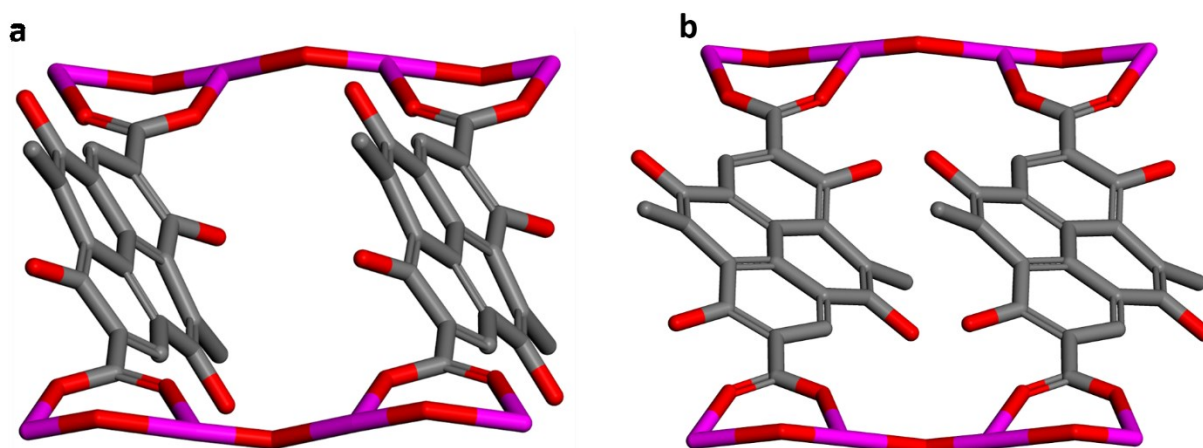


Figure 6.16: a) The optimal alignment of poly aromatic organic SBUs separated by $\sim 7\text{\AA}$ in MM-optimized structure of a high-performing layered sheet MOF b) The same MOF following optimization at the DFT level showing distortions in the aromatic sheet.

Following DFT optimization, 19 MOFs were found to have both CO_2/CH_4 selectivity higher than 20 and CO_2 working capacity higher than 2.0 mmol/g, which outperform zeolite 13x which has a CO_2/CH_4 selectivity of 14 and the CO_2 working capacity of 1.68 mmol/g under dry conditions. In the next section, performance of these 19 MOFs were further examined in the presence of water.

6.3.6 The Effect of Water on the Performance of Top High-performing MOFs

As discussed in Chapter 5, the amount of water in natural gas is reduced to a concentration no greater than of 112 mg of water per standard m^3 of gas to meet typical natural gas pipeline specification across Canada. Even though this level of humidity is much lower than with other gas separations such as post-combustion CO_2 capture, this amount of water vapor could still affect a MOF's performance. For example, Mg-MOF-74, often considered the benchmark MOF for CO_2 gas separations irreversibly decomposes with any trace amounts of water. Even if a MOF is hydrolytically stable, water may more favourably occupy the guest

binding sites, thereby reducing the CO₂ adsorption performance. As such, we have investigated the performance of our top-performing MOFs under trace amounts of water at levels relevant to natural gas transport. Here, the DFT optimized structures of the top-performing MOFs which possessed both CO₂/CH₄ selectivity higher than 20 and CO₂ working capacity higher than 2.0 mmol/g were subjected to long GCMC simulations (i.e., 50 millions steps for equilibrium stage and 50 million steps for production stage) under PSA conditions with 10% relative humidity ([also add the partial pressure]). In those simulations, The LJ parameters for framework atoms and the non-bonded parameters of CO₂ and CH₄ were taken from the force fields mentioned in the Computational Details section. Electrostatic interactions were calculated using the Ewald summation method where the partial atomic charges were calculated with REPEAT method. The water molecules were modeled using TIP4P-Ew force fields.

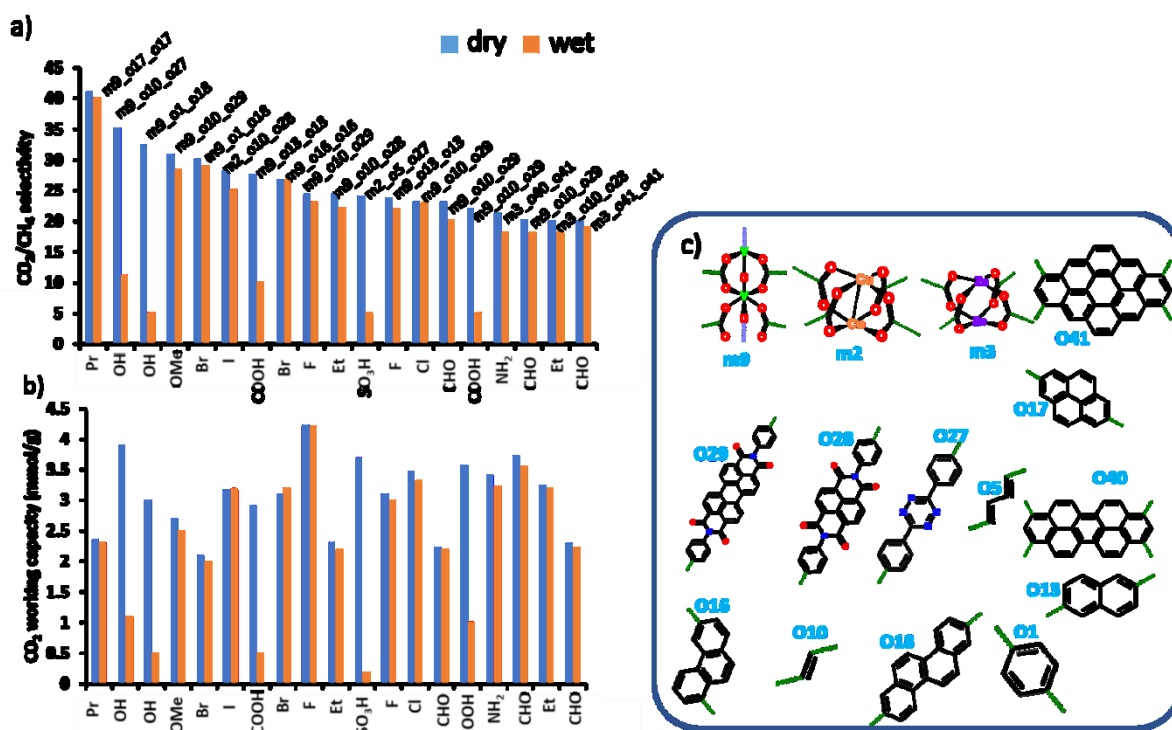


Figure 6.17: a) CO₂/CH₄ selectivity and b) CO₂ working capacity of the 14 top-performing MOFs in the presence and absence of water. c) shows the SBUs codes in the labels given in A).

Figure 6.17 shows the CO₂/CH₄ selectivity and CO₂ working capacity of 19 top-performing MOFs in dry and humid conditions, where each MOF was labeled with the constituent SBUs given in C) and functional groups. Of the 19 top-performing MOFs, 14 MOFs still outperform zeolite 13x in the presence of water. Five MOFs saw a drop of ~70% in the selectivity and ~80% drop in their working capacity in the presence of 10% humidity. These MOFs contained COOH, OH, and SO₃H functional groups. On the other hand, MOFs with halogen functional groups were least affected by humid conditions. 15 MOFs that retain their high-performance adsorption properties following DFT optimization in the presence of water, represent potential synthetic targets for high performance CO₂/CH₄ for natural gas purification. The MOF with the highest selectivity (40) under humid conditions is shown in Figure 6.18, and can be considered a top target for synthesis.

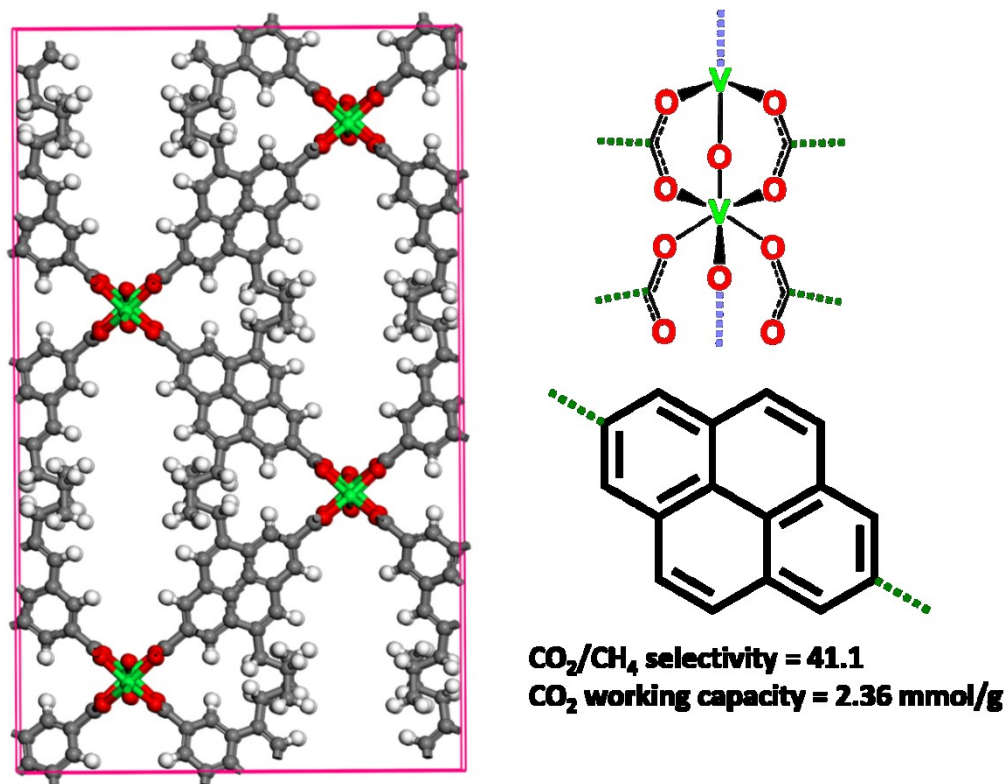


Figure 6.18: The structure of one of the top candidates for CO₂/CH₄ gas separations for natural gas purification that has the highest computed selectivity in humid conditions. It is composed of pillared V SBU bonded to a Pyrene organic SBU functionalized with propyl groups.

6.4 Conclusion

In this work, we defined the general protocol for the screening of large numbers of MOFs for the purpose of natural gas purification. 324,500 hypothetical MOFs were screened for CO₂ working capacity and CO₂/CH₄ selectivity at the PSA conditions relevant for NG purification. From the screened data, some structure-property relationships were identified, providing the limiting (i.e., highest and lowest) gravimetric surface area, void fraction, and max pore diameter essential for the design of high-performing MOFs. From this analysis, MOFs for high performance methane purification should have a pore size lower than 6.6 Å but surface area higher than 2299 m²g⁻¹.

Of the 324,500 MOFs examined, 3990 MOFs were identified as out performing zeolite 13X, a material that is currently used for industrial scale natural gas purification. These structures underwent further analysis to identify common chemical and geometric features of the binding sites. Using maximum clique detection methods, we identified two common selective binding site motifs, S1 and S2 both of which exhibit a high (non-covalent) binding energy with CO₂ (8.5-9.5 kcal/mol). The most common binding site, S1, was composed of layered sheets of poly-aromatic rings spaced roughly 7 Å apart. This structural motif can be targeted by synthetic chemists for high performance CO₂/CH₄ separations.

In order to identify “hit” materials, and not just binding sites, the 50-top high performing structures were further examined. Since the 324,500 MOFs in the database were optimized at the molecular mechanics level, these 50 top performing MOFs were all optimized at the DFT level. Following DFT optimization, only 19 MOFs of the top 50 were found to retain CO₂/CH₄ selectivities above 20 and CO₂ working capacities greater than 2 mmol/g. The adsorption performance of these selected structures were then examined in humid conditions relevant to NG purification. With 10% relative humidity, 14 of the 19 MOFs retained their high performance adsorption levels with selectivities greater than 20 and working capacities greater than 2.0 mmol/g. The best performing material was found to have a CO₂/CH₄ selectivity of 41 and CO₂ working capacity of 2.36 mmol/g, under humid conditions, which is significantly higher than the performance of zeolite 13x (1.65 mmol/g for CO₂ working capacity and 14 for selectivity).

6.5 References

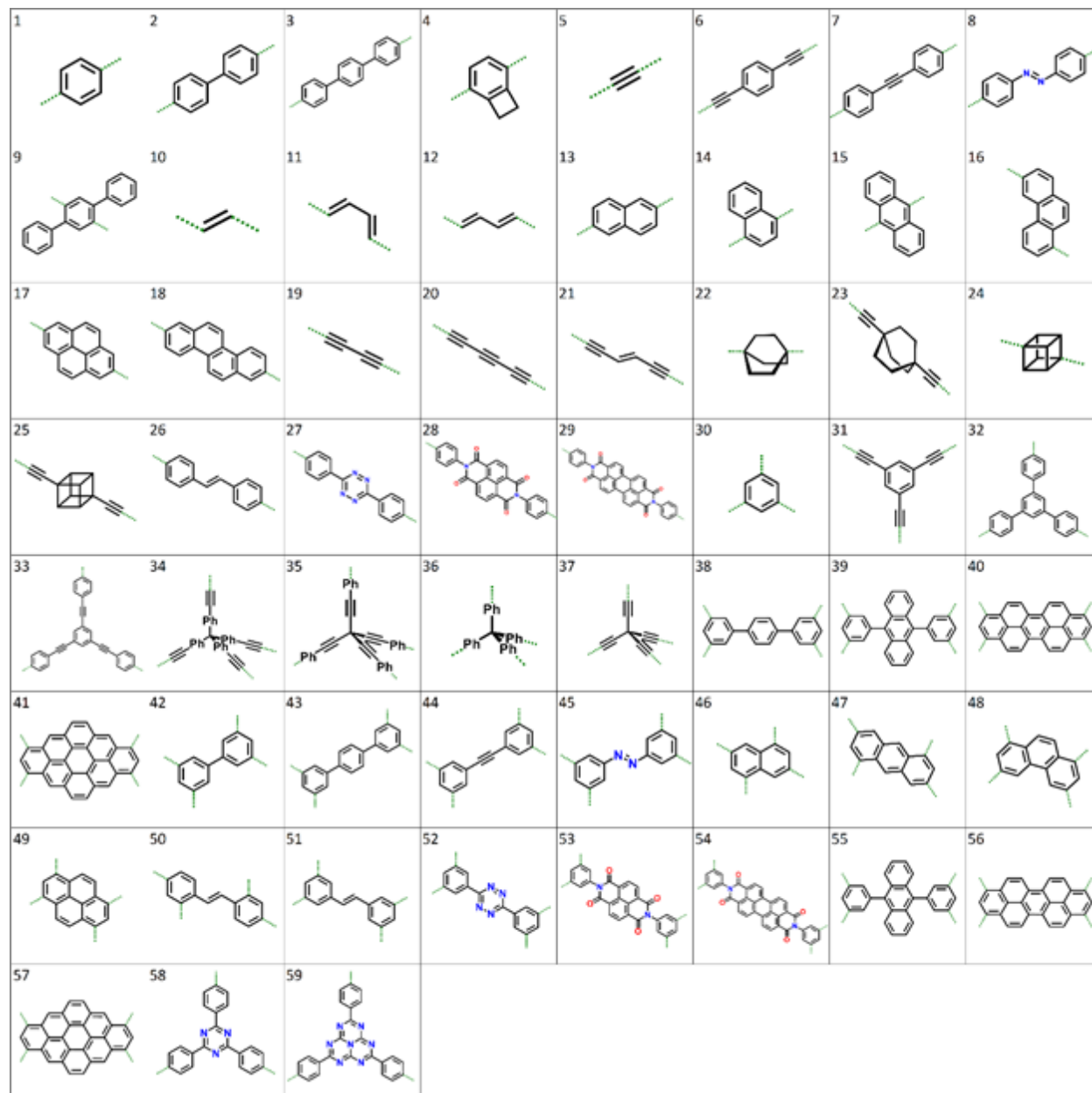
- (1) Solomon, S.; Plattner, G.-K.; Knutti, R.; Friedlingstein, P. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (6), 1704.
- (2) Liu, Y.; Ye, Q.; Shen, M.; Shi, J.; Chen, J.; Pan, H.; Shi, Y. *Carbon dioxide capture by functionalized solid amine sorbents with simulated flue gas conditions*; 2011; Vol. 45.
- (3) U.S. Energy Information Agency. *International Energy Outlook 2013*; 2013; p 312.
- (4) Grande, C. a. *ISRN Chem. Eng.* **2012**, *2012*, 1.
- (5) Rufford, T. E.; Smart, S.; Watson, G. C. Y.; Graham, B. F.; Boxall, J.; Diniz da Costa, J. C.; May, E. F. *J. Pet. Sci. Eng.* **2012**, *94-95*, 123.
- (6) Cavenati, S.; Grande, C. a; Rodrigues, a E. *J. Chem. Eng. Data* **2004**, *49*, 1095.
- (7) Simmons, J. M.; Wu, H.; Zhou, W.; Yildirim, T. *Energy Environ. Sci.* **2011**, *4*, 2177.
- (8) Li, J.-R.; Kuppler, R. J.; Zhou, H.-C. *Chem. Soc. Rev.* **2009**, *38* (5), 1477.
- (9) Li, J.; Sculley, J.; Zhou, H. *Chem. Rev.* **2012**, *112*, 869.
- (10) Sumida, K.; Rogow, D. L.; Mason, J. a; McDonald, T. M.; Bloch, E. D.; Herm, Z. R.; Bae, T.-H.; Long, J. R. *Chem. Rev.* **2012**, *112* (2), 724.
- (11) Zhang, Z.; Yao, Z.; Xiang, S.; Chen, B. *Energy Environ. Sci.* **2014**, *7* (9), 2868.
- (12) Krishna, R.; van Baten, J. M. *Phys. Chem. Chem. Phys.* **2011**, *13*, 10593.
- (13) Krishna, R.; Long, J. R. *Phys. Chem. C.* **2011**, *2*, 12941.
- (14) Burtch, N. C.; Jasuja, H.; Walton, K. S. *Chem. Rev.* **2014**, *114*, 10575.
- (15) DeCoste, J. B.; Peterson, G. W.; Jasuja, H.; Glover, T. G.; Huang, Y.; Walton, K. S. *J. Mater. Chem. A* **2013**, *1* (18), 5642.
- (16) Low, J. J.; Benin, A. I.; Jakubczak, P.; Abrahamian, J. F.; Faheem, S. A.; Willis, R. R. *J. Am. Chem. Soc.* **2009**, *131* (43), 15834.
- (17) Wilmer, C. E.; Leaf, M.; Lee, C. Y.; Farha, O. K.; Hauser, B. G.; Hupp, J. T.; Snurr, R. Q. *Nat. Chem.* **2012**, *4* (2), 83.
- (18) 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* (6), 2159.
- (19) Wilmer, C. E.; Farha, O. K.; Bae, Y.-S.; Hupp, J. T.; Snurr, R. Q. *Energy Environ. Sci.* **2012**, *5* (12), 9849.
- (20) Kadantsev, E. S.; Boyd, P. G.; Daff, T. D.; Woo, T. K. *J. Phys. Chem. Lett.* **2013**, *4* (18), 3056.
- (21) Wilmer, C. E.; Farha, O. K.; Bae, Y.-S.; Hupp, J. T.; Snurr, R. Q. *Energy Environ. Sci.* **2012**, *5* (12), 9849.
- (22) Vaidhyanathan, R.; Iremonger, S. S.; Shimizu, G. K. H.; Boyd, P. G.; Alavi, S.; Woo, T. K. *Science*, **2010**, *330*, 650.
- (23) Togo, A. spglib (<http://spglib.sourceforge.net>), 2012.
- (24) Bellussi, G.; Broccia, P.; Carati, A.; Millini, R.; Pollesel, P.; Rizzo, C.; Tagliabue, M. *Microporous*

- Mesoporous Mater.* **2011**, 146 (1-3), 134.
- (25) Mu, W.; Liu, D.; Yang, Q.; Zhong, C. *Microporous Mesoporous Mater.* **2010**, 130 (1-3), 76.
- (26) Martín-Calvo, A.; García-Pérez, E.; Manuel Castillo, J.; Calero, S. *Phys. Chem. Chem. Phys.* **2008**, 10, 7085.
- (27) Demir, B.; Ahunbay, M. G. *J. Phys. Chem. C* **2013**, 117 (30), 15647.
- (28) Gale, J. J. *Chem. Soc., Faraday Trans.* **1997**, 93 (4), 629.
- (29) Campaña, C.; Mussard, B.; Woo, T. K. *J. Chem. Theory Comput.* **2009**, 5, 2866.
- (30) Fateen, S. E. K.; Khalil, M. M.; Elnabawy, A. O. *J. Adv. Res.* **2013**, 4 (2), 137.
- (31) Rappe, a. K.; Casewit, C. J.; Colwell, K. S.; Goddard III, W. a.; Skiff, W. M. *J. Am. Chem. Soc.* **1992**, 114, 10024.
- (32) 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 (3), 8814.
- (33) Martin, M. G.; Siepmann, J. I. *J. Phys. Chem. B* **1998**, 102 (97), 2569.
- (34) Kresse, G.; Hafner, J. *Phys. Rev. B* **1993**, 47 (1), 558.
- (35) Kresse, G.; Furthmüller, J. *Phys. Rev. B* **1996**, 54 (16), 11169.
- (36) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, 77 (1992), 3865.
- (37) Kresse, G. *Phys. Rev. B* **1999**, 59 (3), 1758.
- (38) Blöchl, P. E. *Phys. Rev. B* **1994**, 50 (24), 17953.
- (39) Steffen, C.; Thomas, K.; Huniar, U.; Hellweg, A.; Rubner, O.; Schroer, A. *J. Comput. Chem.* **2010**, 31, 2967.
- (40) Willems, T. F.; Rycroft, C. H.; Kazi, M.; Meza, J. C.; Haranczyk, M. *Microporous Mesoporous Mater.* **2012**, 149 (1), 134.
- (41) Duren, T.; Millange, F.; Ferey, G.; Walton, K. S.; Snurr, R. Q. *J. Phys. Chem. C* **2007**, 111 (42), 15350.
- (42) Gómez-Gualdrón, D. a; Wilmer, C. E.; Farha, O. K.; Hupp, J. T.; Snurr, R. Q. *J. Phys. Chem. C* **2014**, 118, 6941.
- (43) Wills, J., Mitariten, M. 2009. Production of Pipeline Quality Natural Gas With the Molecular Gate™ CO2 Removal Process. Tidelands Oil Production Company & Guild Associates.

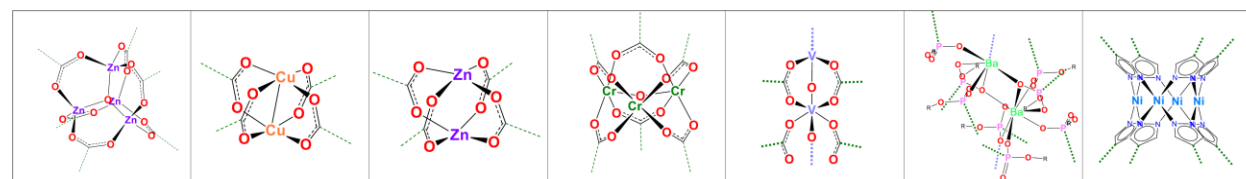
6.6 Appendix - Hypothetical MOF Database

The entire list of SBUs used to construct the hypothetical MOFs is shown in Figure 6.18

Organic SBUs



Inorganic BUs



Functional groups

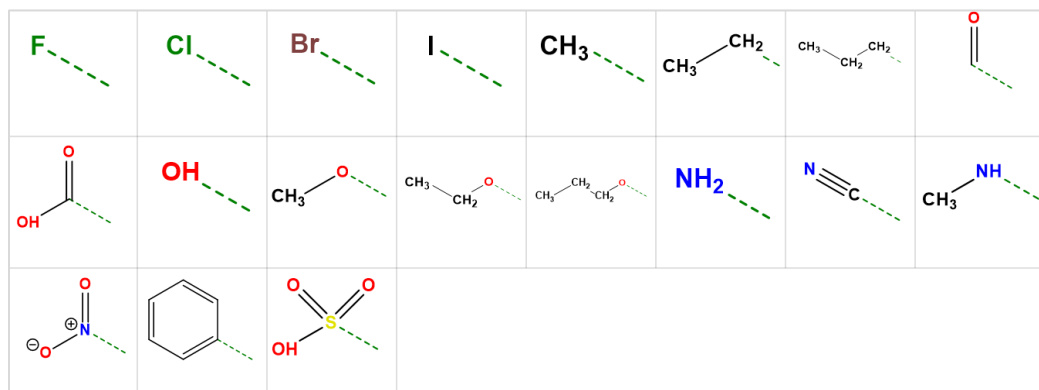


Figure 6.19: Structural Building Units (SBUs) and functional groups used to generate the ~350,000 hypothetical MOFs from the Northwestern University database.

7 Quantitative Structure-Property Relationship Models for Recognizing Metal Organic Frameworks (MOFs) with High CO₂ working Capacity and CO₂/CH₄ Selectivity for Methane Purification

The contents of this chapter were published in part in Aghaji, M. Z.; Fernandez, M.; Boyd, P. G.; Daff, T. D.; Woo, T. K. *J. European Journal of Inorganic Chemistry*. **2016**. The GCMC simulations, MOF's characterizations, and model developments were performed by the author. The manuscript was prepared by the author, M. Fernandez, with assistance by T. Woo.

Abstract

Nearly infinite number of Metal Organic Frameworks (MOFs) can be theoretically constructed, which represents a combinatorial design challenge that demands the use of computational tools to explore the chemical space. Here we report Quantitative Structure-Property Relationship (QSPR) models to help identify high-performing MOFs for methane purification solely using geometrical features. The CO₂ working capacity and CO₂/CH₄ selectivity of ~320,000 hypothetical MOF structures was computed at conditions relevant to natural gas purification using grand canonical Monte-Carlo (GCMC) simulations. From this large database a training set of 32,500 MOF structures was used to calibrate binary decision tree (DT) and support vector machines (SVMs) models that can accurately identify high-performing MOFs based on their pore size, void fraction and surface area. DT models yielded guidelines of pore size, void fraction and surface area for designing high-performing materials. The SVM machine learning classifiers can be used to quickly pre-screen MOFs, such that the compute intensive GCMC simulations are not performed on all structures. The SVM classifiers were tested on ~290,000 MOFs that were not part of the training set and could correctly identify up to 90% of high-performing MOFs while only flagging a fraction of the MOFs for more rigorous screening. QSPR models constitute efficient computational tools for the virtual screening of large

structural libraries and provide rational design rules for the discovery of sorbents for methane purification.

7.1. Introduction

Metal organic frameworks (MOFs) are nanoporous crystalline materials that are promising candidates to replace zeolites as solid sorbents in gas separation applications such as natural gas purification.^{1,2,3,4} MOFs can be constructed from a virtually limitless combination of inorganic and organic secondary building units (SBUs), yielding a diverse range of pore volume, surface area, void fraction and functionality suitable for different gas separation applications.⁵ Designing MOFs with optimal combinations of these properties for a specific application presents a remarkable challenge due to the high structural diversity.

Recently, several computational high-throughput screening studies on MOFs with the intent of finding their optimal adsorption properties have been performed. Wilmer *et al.*⁶ screened over 130,000 hypothetical MOFs for CO₂/CH₄ and CO₂/N₂ separations applications, Lin *et al.*⁷ screened a database of hundreds of hypothetical zeolitic MOFs for their parasitic energies in CO₂/N₂ separation, and Watanabe *et al.*⁸ and Qiao *et al.*⁹ screened large sets of MOFs taken from the Cambridge Structural Database (CSD)¹⁰ for their adsorption properties in CO₂ gas separation applications. In these studies, computationally expensive GCMC simulations are utilized that typically require the evaluation of millions of guest-host interactions.¹¹ To evaluate these interactions, molecular mechanics force fields are used where the non-bonded Lennard-Jones and atomic partial charge parameters are important. While the van der Waals parameters are in general transferable between one MOF to another, the partial atomic charges need to be estimated for each individual MOF. Even with the use of fast parametric

approaches to calculate the atomic partial charges^{12,13} the GCMC simulations used to determine the adsorption properties takes on the order of tens of minutes to perform. This prohibits the brute-force high-throughput screening of databases of millions of structures that can now be easily constructed.^{14,15,13,16}

Machine learning and molecular similarity can be applied to extract relevant knowledge from the screening data and predictive models can be trained to rapidly identify high-performing candidate MOFs to obviate the need for performing GCMC simulations on all structures of a database.¹⁷ For example, the large-scale similarity analysis of millions of hypothetical zeolites using descriptors derived from the Voronoi network representation of the frameworks allowed the rapid identification of promising candidates for CO₂ capture.¹⁸ Recently, the Woo lab has reported that geometrical properties and specialized radial distribution function descriptors (RDF) that account for the chemistry of the pores can be combined with machine learning techniques to yield quantitative structure-property relationship (QSPR) models that can predict the potential of MOFs for low pressure CO₂ uptake capacities.^{17,19} Although the uptake capacity of a material is often examined, it is the working capacity of a material that is a more relevant metric for gas separation processes, where the working capacity is the difference in uptake capacity at the adsorption and desorption conditions. The working capacity corresponds to the amount of gas a given material can capture and release in a single adsorption/desorption cycle. Another critical adsorption property for gas separation applications is the selectivity of the sorbent material, which is defined for a binary mixture of species *i* and *j* by

$$S = \frac{q_i/q_j}{p_i/p_j} \quad 7.1$$

where the q_i is the uptake capacity and p_i is the partial pressure of the gas i .

In this work, we develop machine learning QSPR models of the CO₂ working capacity and CO₂/CH₄ selectivity at conditions typical of natural gas purification with PSA systems using the purely geometric properties of pore size, void fraction and surface area. The models are trained on the GCMC calculated adsorption properties of 32,450 MOFs using binary decision trees (DTs) and nonlinear SVMs. These machine learning models were then tested on a set of 292,050 diverse MOF structures and compared to the results of GCMC simulations. These QSPR models allowed for the accurate identification of high-performing MOFs, thereby allowing for rapid pre-screening of materials in order to significantly reduce the number of computational intense GCMC simulations. In brief, the remainder of this chapter is as follows: first a description of the data set, computational methods and QSPR descriptors used to calibrate the model is provided. Then an explanation of how the classifiers were trained is given, followed by a discussion of the results and conclusions.

7.2 Data set and computational methods

7.2.1. Hypothetical MOF dataset and GCMC simulations

The 324,500 hypothetical MOF structures in our database were generated from 70 SBUs and 20 functional groups using an similar algorithm to that from Wilmer *et al.*²⁰. Care was taken to ensure reasonable non-bonded atom distances during structure generation by ensuring that atoms did not enter within half the sum of their van der Waal radii. Functional group placement was accomplished by replacing available hydrogens on organic SBUs within each material.

Where available, symmetry was imposed on functional group placement so that symmetric hydrogens would be replaced with the same functional group. To relax the atomic positions, a molecular mechanics force field was used to optimize the atomic positions in the GULP 4.0 software package.²¹ Bonding information was included in each hypothetical MOF to ensure the correct connectivity information was provided to the optimizer. A small sample of the SBUs used to construct the database are depicted in Figure 7.1. The list of all SBUs and functional groups used are given in the appendix of Chapter 5.

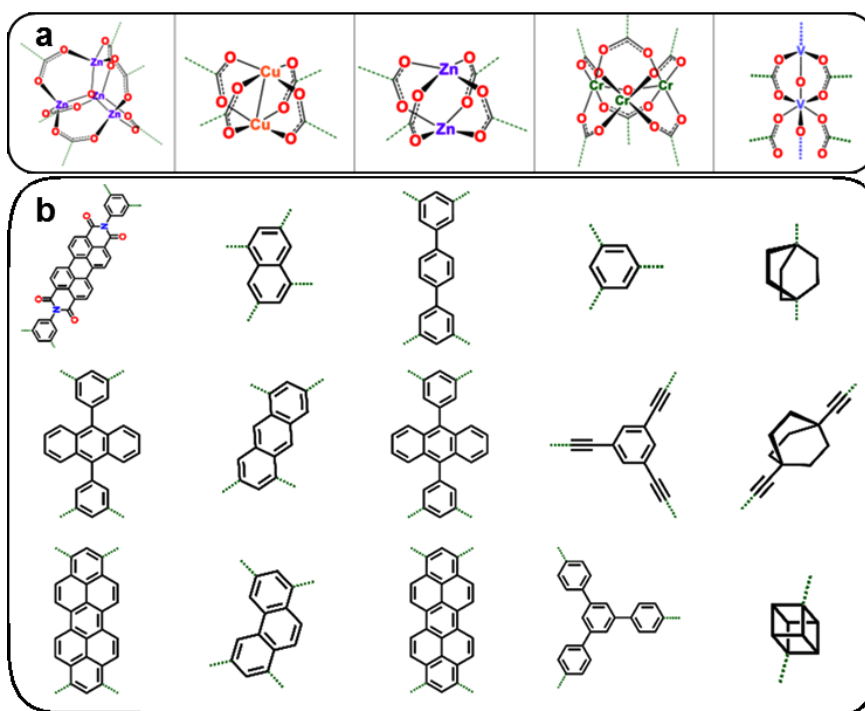


Figure 7.1: Examples of the inorganic (a), organic (b) SBUs used to generate the 324,500 hypothetical MOFs in our database.

GCMC simulations were performed on the entire 324,500 member database to estimate the adsorption of a binary mixture of $\text{CO}_2:\text{CH}_4$ with a composition of 1:9 at a total pressure of 10 bar and 1 bar at 298 K. In these simulations, the constant chemical potential was calculated using the ideal-gas equation of state for pressures up to 1 bar, while for higher pressures the

Peng-Robinson equation of state was used with binary interaction parameters²². The GCMC simulations were performed with a 3x3x3 supercell. Interactions between guest molecules and the fixed framework atoms were evaluated with Lennard-Jones (LJ) and Columbic potentials. The LJ parameters for framework atoms were taken from the UFF and combined with Lorentz-Berthelot mixing rules for cross terms. All LJ parameters are shown in Table 6.1. Electrostatic interactions were calculated using the Ewald summation method where the partial atomic charges were calculated with MEPO-QEq method¹² which was trained to reproduce DFT derived electrostatic potential fitted charges using the REPEAT method.²³ The non-bonded parameters of CO₂ were taken from García-Sánchez et al. 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 CO₂ molecule had a fixed bond length of 1.16 Å and partial charges of +0.645e and -0.3325e for carbon and oxygen. The CH₄ molecules were modeled using TraPPE force field parameters,²⁴ which were developed to reproduce the vapor-liquid phase equilibrium of methane. In this force, methane is modeled as a single site with LJ parameters shown in Table 7.1.

Table 7.1: LJ parameters for methane, CO₂ and framework atoms in all hypothetical MOFs.

Atom type	σ (Å)	ϵ/k_B (K)	Atom type	σ (Å)	ϵ/k_B (K)
C	3.43	52.83	Cu	3.114	2.516
zO	3.12	30.19	V	2.80	8.05
H	2.57	22.14	Zr	2.783	34.72
N	3.26	34.72	C(CO ₂)	2.745	29.933
F	2.997	25.16	O(CO ₂)	3.017	85.671
Cl	3.517	114.23	CH ₄ (united)	3.75	148.0
Br	3.73	126.30	Zn	2.46	62.40

7.2.2. Machine learning classification models

To classify the methane purification performance of MOFs, we used two machine learning techniques: decision tree (DT) and support vector machine (SVM). While DT is a linear regressor,²⁵ the SVMs²⁶ are a machine learning method of broad applicability to many types of pattern recognition problems. The decision tree (DT) regressions were built using the C4.5 algorithm²⁵ implemented in the *Scikit-learn* module²⁷ for Python. The nonlinear SVMs discussed in Chapter 1 were built using the Python programming language and the SHOGUN toolbox²⁸. Our implementation of SVMs is briefly introduced as follows. The input vectors (i.e. MOF structural features) are first mapped onto one feature space (possibly with a higher dimension) by means of a kernel function. Then, a hyperplane is built to separate the positive and negative inputs in the data within this feature space. Only relatively low-dimensional vectors in the input space and dot products in the feature space will evolve by a mapping function. A radial-basis-function (RBF) kernel was implemented in the SVM framework to handle the nonlinearity in the data to yield higher accuracies. The optimum width of the RBF function and the regularization parameter of the SVM regressions were tuned by a grid search that maximized the cross-validation **AUC** of the ROC plot of the training set. For the classification, all data elements (e.g., gravimetric surface area) were normalized to be in the range [-1, 1], where -1 and 1 correspond to the minimum and maximum values of each category of data (e.g., the lowest and highest gravimetric surface area values in the database). The MOFs in the database were labeled as high-performing or low-performing at given threshold values according to their GCMC simulated CO₂ working capacity and CO₂/CH₄ selectivity. A set of 32,450 MOFs were randomly

chosen from the database to form the training set used to construct the QSPR models, while the remaining 292,050 structures formed the test set to validate the QSPR model.

7.2. Result and Discussion

The complete 324,500 member hypothetical MOF database was screened for carbon dioxide and methane adsorption capacity by GCMC simulations of a binary mixture of CO_2 : CH_4 with a composition of 1:9 at a total pressure of 10 bar (adsorption) and 1 bar (desorption) at 298 K. The distributions of CO_2 working capacity and CO_2/CH_4 selectivity of the entire database are given in Figure 7.2. We found that approximately 60% of the database exhibited CO_2 working capacities higher than 2 mmol/g and ~10% had working capacities greater than 4 mmol/g. Similarly, ~40% possessed CO_2/CH_4 selectivity higher than 5 and ~10% had selectivities greater than 10. To put these thresholds in perspective, the solid sorbent zeolite 13X currently used in commercial PSA systems to scrub CO_2 from natural gas has a working capacity and selectivity of 1.8 mmol/g and 14,²⁹ respectively, under the conditions studied here.

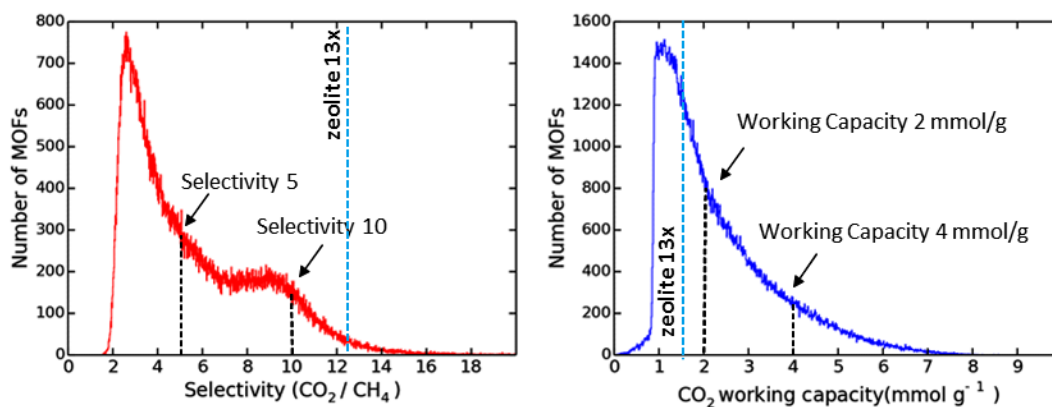


Figure 7.2: Distribution of CO_2/CH_4 selectivity and CO_2 working capacity of entire database. Black dashed lines represent the ‘high-performing’ thresholds used in the QSPR models, while the blue dashed line shows the selectivity and working capacity of zeolite 13x that is currently in use.

We characterized the geometry of the MOFs in our database by calculating the maximum pore diameter, void fraction and surface area. These properties were calculated with the Zeo++ computer software³⁰ using helium as a probe molecule. Initially, we performed a univariate analysis where the separation parameters (working capacity and selectivity) are scatter plotted against the geometrical features as given in Figure 7.3. The scatter plots of their interactions depict well-defined patterns of CO₂ working capacity and CO₂/CH₄ adsorption selectivity. In general, the CO₂/CH₄ selectivity (Figure 7.3a) decreases for surface area, pore diameter and void fraction higher than 4000 m²/g, 10 Å and 0.4, respectively. In contrast, MOFs with lower surface area, pore diameter and void fraction exhibited selectivity higher than 5. Figure 7.3b shows a different behaviour for the working capacity, which increases beyond 2 mmol/g for MOFs with surface areas in the range of 2000 to 5000 m²/g, pore diameter values in the range of 5 Å to 10 Å and void fraction values in the range of 0.1 to 0.5. Additionally, the CO₂ working capacity decreases below 1 mmol/g for surface area and void fraction lower than 1000 m²/g and 0.1, respectively. In general, the optimum ranges of geometrical features found here are in agreement with previous virtual HT screening for methane purification made by Wilmer et al.⁶

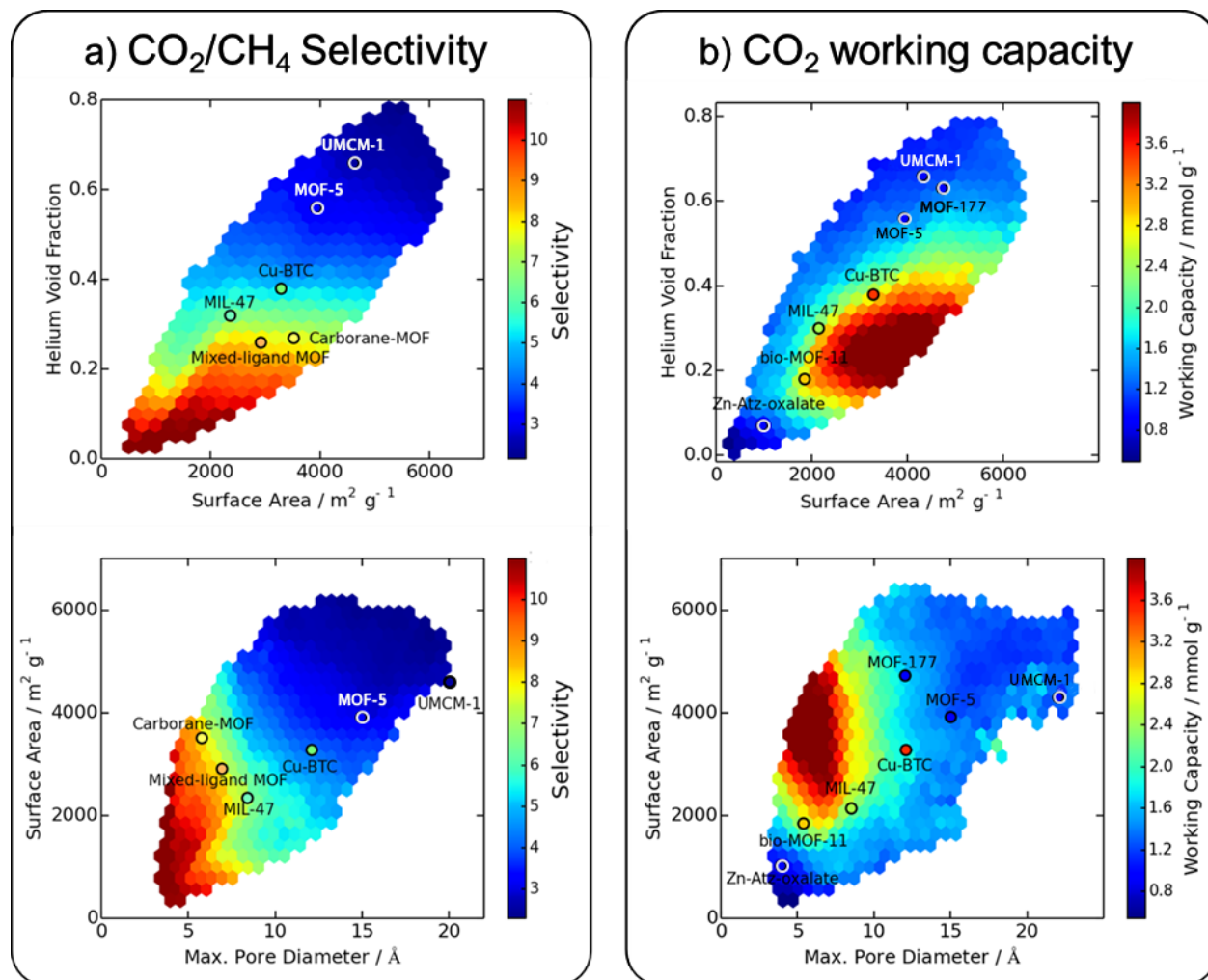


Figure 7.3: Interaction scatter plots of the max. pore diameter, helium void fraction and surface area of 324,500 MOFs in the calibration data set with color mapped (a) CO_2/CH_4 selectivity and (b) CO_2 working capacity. Each hexbin contains more than one MOFs and the color represents the average (a) CO_2/CH_4 selectivity or and (b) CO_2 working capacity of MOFs. The colored dots correspond to the experimental working capacity and selectivity of published MOFs. (Cu-BTC³¹, Zn-Atz-oxalate³², MOF-5³³, Carborane-MOF³⁴, Mixed-ligand MOF³⁵, UMCM-1³⁶, bio-MOF-11³⁷, MOF-177³⁸, MIL-47³⁹)

We also computed the geometrical features of nine MOFs with experimentally reported CO_2 working capacity and CO_2/CH_4 selectivity. These are shown in Figure 7.3 as the labelled open dots where the colour map of the dot corresponds to the MOF's experimental working capacity and selectivity. As can be observed, these experimental adsorption parameters matched, quite well, the simulated values of neighbouring hypothetical MOFs from our

database. The exception was the MOF Cu-BTC (bottom plots of Figure 7.3a and 7.3b), which possesses experimental working capacity and selectivity of ~ 4 mmol/g and ~ 8 , respectively. With a pore diameter of ~ 12 Å, the experimental working capacity and selectivity of Cu-BTC are notably higher than that predicted by our virtual screening. Cu-BTC is a special case of an open-metal site MOF where the CO₂ uptake is not physisorptive, but rather chemisorptive, which our model potentials do not account for.

The good graphical discrimination in Figure 7.3 suggests that correlation models can be derived from this data to rapidly filter through MOF databases for methane purification. We focused on QSPR models that use pore size, surface area and void fraction to classify MOFs as high- or low-performing defined by threshold values of working capacity and selectivity. In addition to enabling a rapid way to screen through candidate materials, this type of binary classification can yield valuable structure-property relationship ‘rules-of-thumb’ that can be used for the rational design process. For this purpose, we defined a high-performance threshold for each property: 2 and 4 mmol/g for the CO₂ working capacity; and 5 and 10 for the CO₂/CH₄ selectivity. For each of the four thresholds, a different model is built. To derive the models, we randomly selected 10% of the structures in the database so that 32,450 MOFs formed a representative calibration set. For each of the thresholds, we labelled every MOF in the calibration set as 1 or 0 depending on whether its property value was above or below the threshold. Then, a QSPR model was created for each classification scheme where we trained the model with the geometrical features of the calibration set to return a score between 0 and 1 for each MOF. This score is then transformed to a binary label indicating a high-performing or low-performing structure by choosing a sensitivity *cut-off* value, a parameter that is adjusted

when applying the classifier in order to adjust the number of ‘high performing’ structures that are returned, depending on the user’s needs. A comparison of the predicted labels, from the SVM model, and the labels from the GCMC data can be depicted using a receiver operating characteristic (ROC) plot, as shown in Figure 7.4 and 7.5. The ROC curve illustrates the ability of the SVM model to correctly label MOFs. The fraction of true positives out of the positives (TPR = true positive rate) is plotted versus the fraction of false positives out of the negatives (FPR = false positive rate). TPR is also referred to as the sensitivity, in contrast to the true negative rate (TNR) which is referred to as the specificity. The FPR is then one minus the specificity. The ROC depicts relative trade-offs between true positives (benefits) and false positives (costs). The best possible prediction method would yield a point in the upper left corner of the ROC plot, representing 100% sensitivity (no false negatives) and 100% specificity (no false positives) and an area under the curve (AUC) of one. Completely random guesses would yield points along the diagonal line from the bottom left to the top right corners with an AUC value of 0.5. Hence, the AUC measures the probability that the model will label a randomly chosen MOF correctly. The AUC for all of the models built are considered excellent and are reported in Table 7.2.

Table 7.2: Area under the curve (AUC) values of the receiver operator curve (ROC) plots of the recognition of high-performing MOFs for CO₂/CH₄ separations using DT and SVM models of pore size, void fraction and surface area.

parameter	threshold	model type	AUC
working capacity	2 mmol/g	DT	0.889
		SVM	0.916
	4 mmol/g	DT	0.908
		SVM	0.919
selectivity	5	DT	0.948
		SVM	0.953
	10	DT	0.928
		SVM	0.934

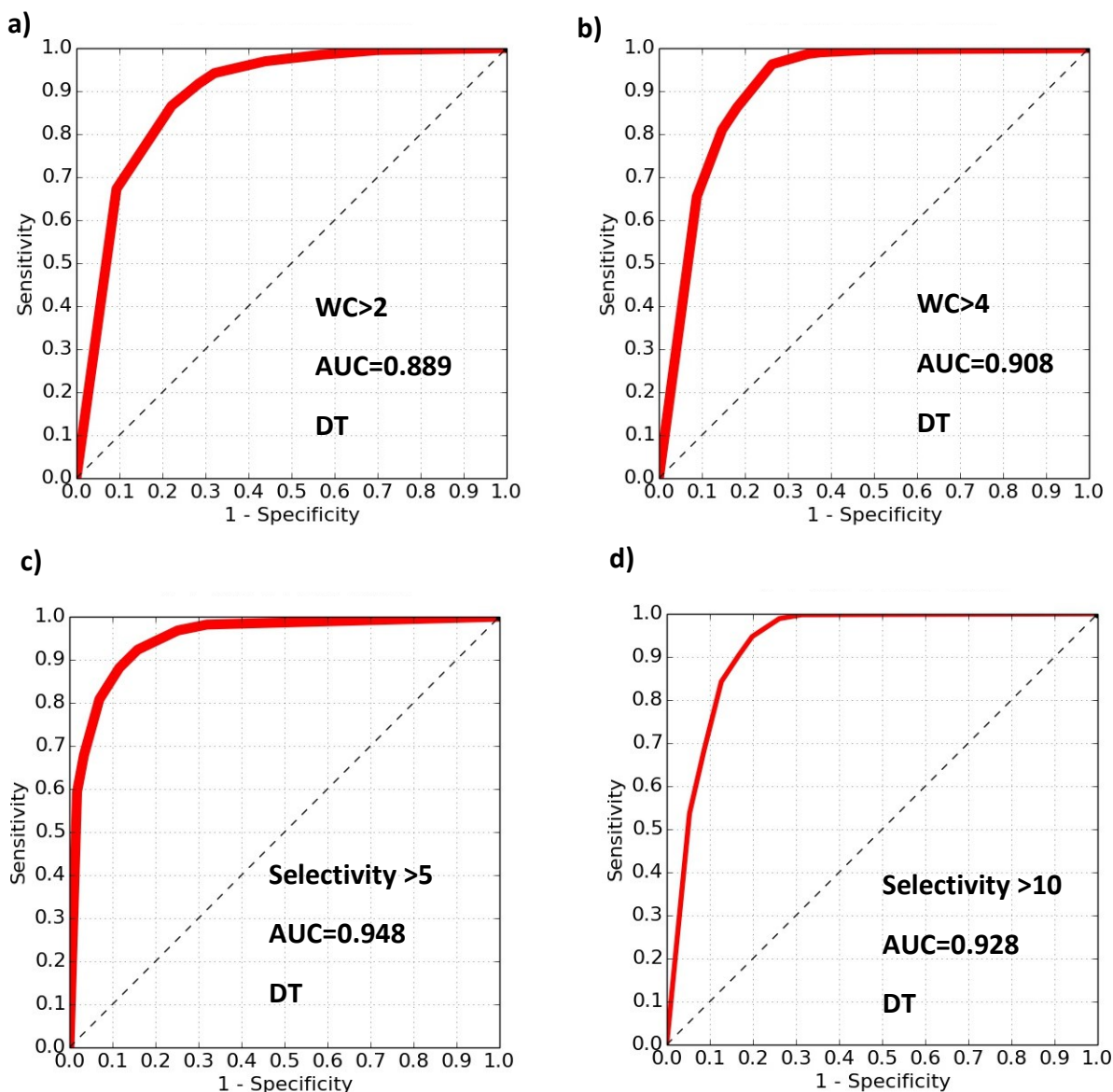


Figure 7.4: Calibration receiver operating curve (ROC) plots that show how well the decision tree (DT) models label MOFs as "positive" or "negative" depending on whether the CO₂ working capacity is greater or less than 2 mmol/g (a) and greater or less than 4 mmol/g (b); and CO₂/CH₄ selectivity greater or less than 5 (c) and greater or less than 10 (d). The MOFs are labelled correctly most of the time, as indicated by the high AUC value. The dashed line represents the curve that would be obtained by random labelling of the MOFs as positive or negative.

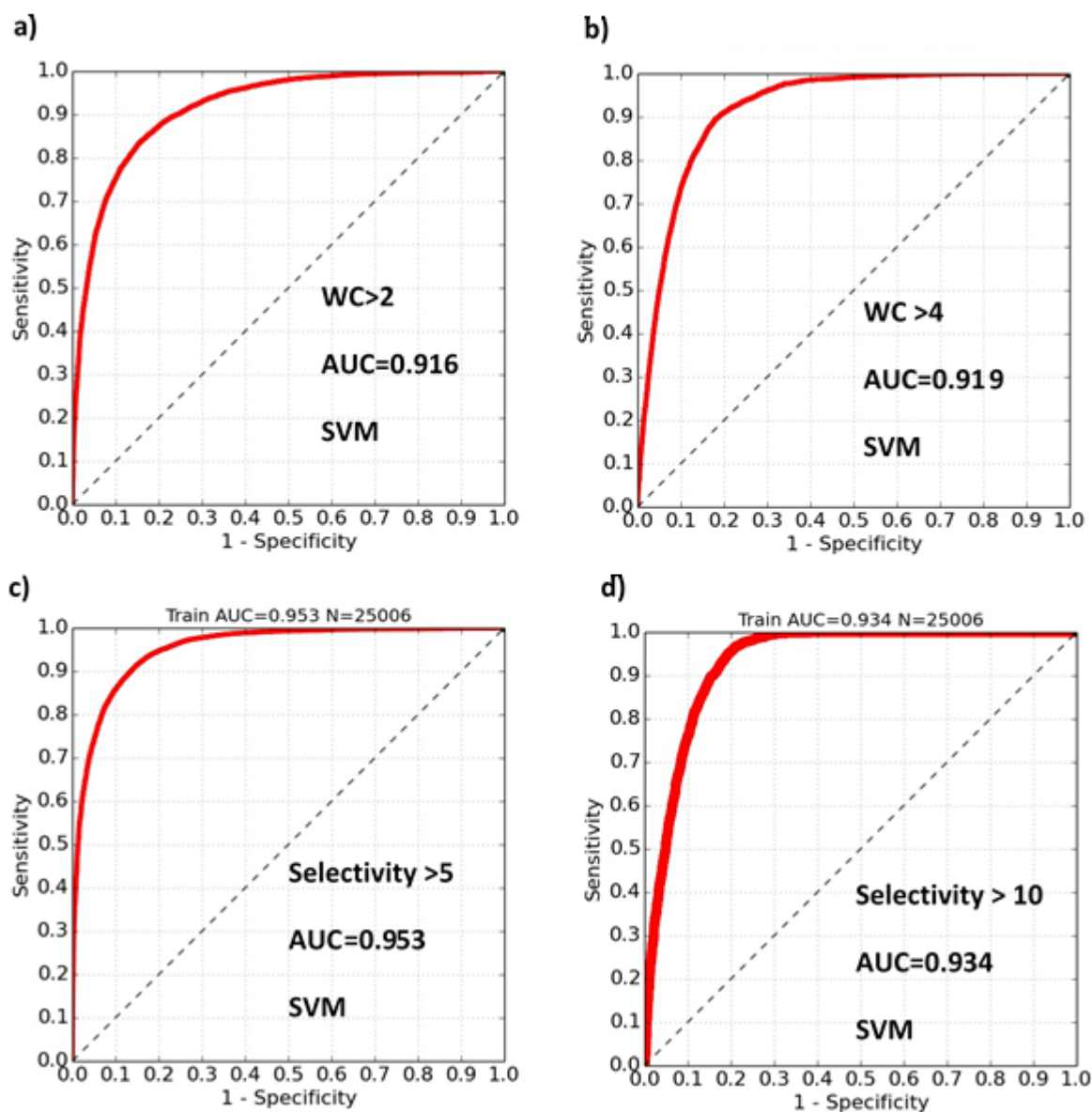


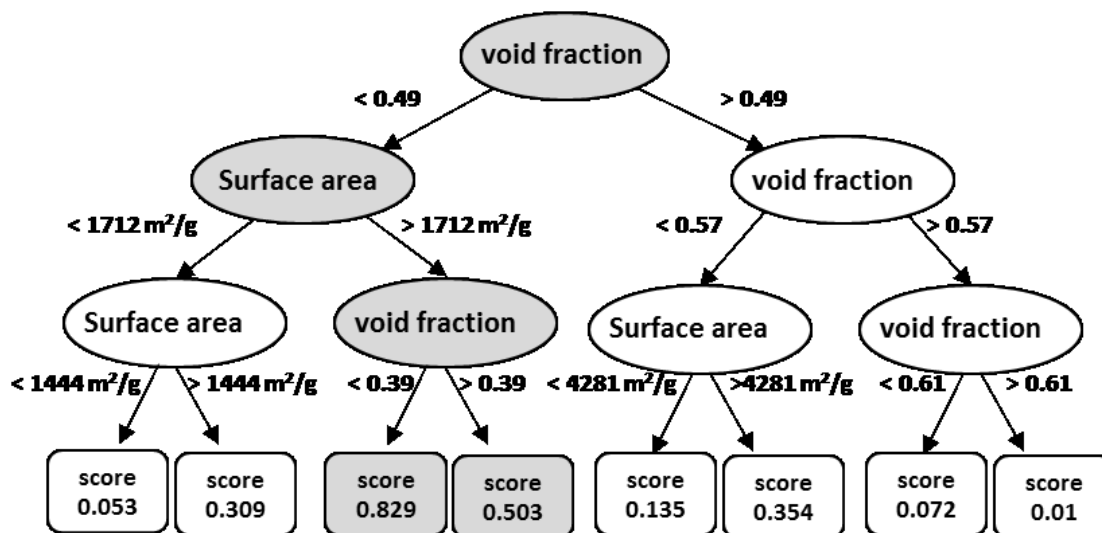
Figure 7.5: Calibration receiver operating curve (ROC) plots that show how well the support vector machine (SVM) models label MOFs as “positive” or “negative” depending on whether the CO₂ working capacity is greater or less than 2 mmol/g (a) and greater or less than 4 mmol/g (b); and CO₂/CH₄ selectivity greater or less than 5 (c) and 10 (d). The MOFs are labelled correctly most of the time, as indicated by the high AUC value. The dashed line represents the curve that would be obtained by random labelling of the MOFs as positive or negative.

Decision tree (DT) predictions were made using simple binary rules (e.g. void fraction must be greater than 0.5) that can be easily followed in a MOF design process to achieve a

desired property target.^[30] Only three rules were considered because if too many levels are used in the tree the practical utility of the DT model diminishes. Despite the low complexity of the DT models, MOFs with CO₂ working capacities higher than 2 mmol/g and 4 mmol/g were identified with high AUC values of 0.889 and 0.908, respectively. Similarly, MOFs with CO₂/CH₄ selectivity higher than 5 and 10 were recognized with AUC values of 0.948 and 0.928, respectively. Binary DT graphs shown in Figure 7.6a and 7.6b illustrate general design rules to obtain CO₂ working capacity higher than 2 mmol/g and 4 mmol/g, respectively. The DT modes encompass three decision levels, all of them relevant from the AUC point of view, but from a model interpretation perspective, we found that two rules discriminate well between high and low performance as it is indicated by the similar high scores in the lowest branches of the high performance nodes in Figures 7.6 and 7.7. Therefore, the ‘rule-of-thumb’ for CO₂ working capacity >2 mmol/g (T=298 K, P_{ad}=10 atm, P_{de}=1 atm) is to have void fraction lower than 0.49 and gravimetric surface area higher than 1712 m²/g. Meanwhile, the “rule-of-thumb” for CO₂ working capacity >4 mmol/g is to have pore size lower than 8.5 Å and gravimetric surface area greater than 2300 m²/g. Binary DT graphs for CO₂/CH₄ selectivity appear in Figure 7.7. From Figure 7.7, the ‘rule-of-thumb’ for CO₂/CH₄ selectivity >5 is to have void fraction lower than 0.32 and pore size lower than 8.30 Å. Meanwhile, the ‘rule-of-thumb’ for CO₂/CH₄ selectivity >10 is to have void fraction lower than 0.27 and pore size lower than 6.6 Å. It is not surprising that only combinations of void fraction and pore size “rules” appear in the DT models of selectivity in Figure 7.7. The amount of available space in the pore can be realized as more important than surface area for the capture of CO₂ molecules while methane diffuses through the material.

Despite the fact that the rules for high working capacity and selectivity are not necessary mutually exclusive, we only found 6,256 MOFs with working capacity greater than 4 and selectivity greater than 10 in our database, which represents less than 2% of the structures. For example, selectivity >10 requires a pore size lower than 6.6 Å while working capacity >4 mmol/g allows for pore size lower than 8.5 Å but with surface area higher than 2299 m²g⁻¹. Therefore, if we combine both two rules the “best” MOFs for methane purification would have a pore size lower than 6.6 but surface area higher than 2299 m²g⁻¹ lower.

a) Working Capacity > 2 mmol/g



b) Working Capacity > 4 mmol/g

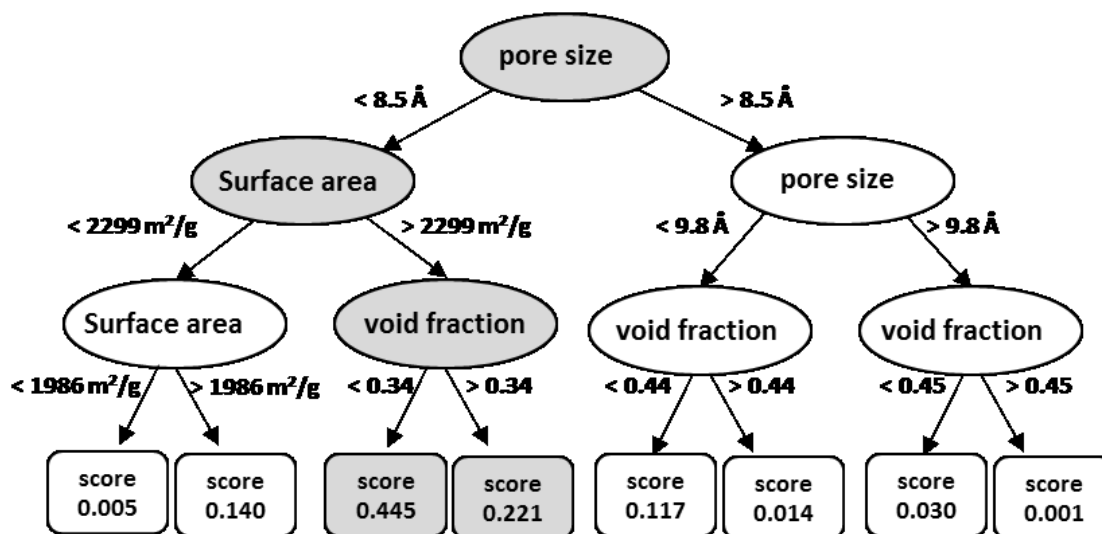
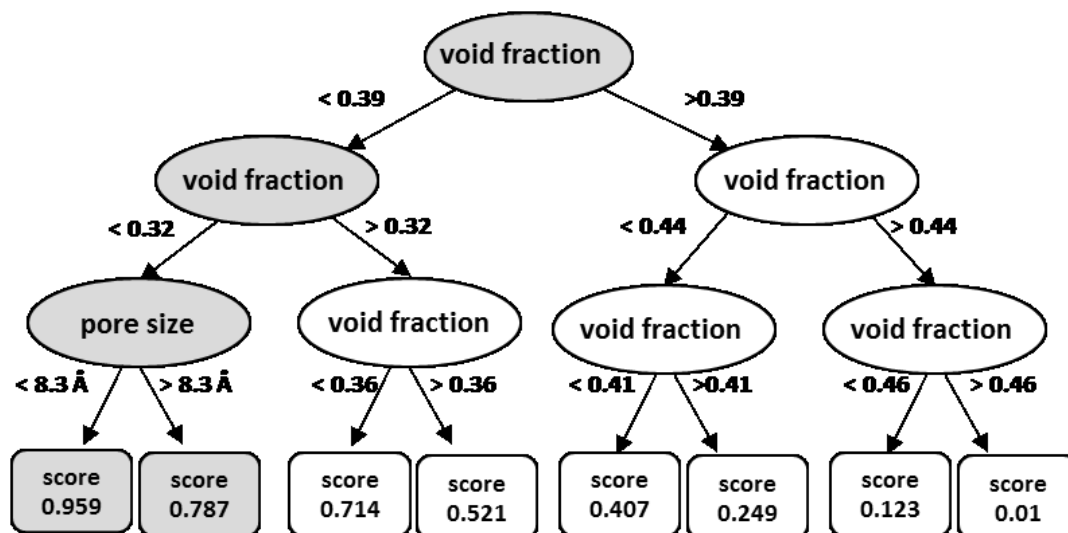


Figure 7.6: Decision tree (DT) models of the CO₂ working capacity higher than 2 mmol/g (a) and 4 mmol/g (b) of the 32,450 MOFs in the calibration set. Each branch of the tree represents a threshold level given for the value of a property, where the MOFs with values higher than the level threshold branched to the right, while MOFs with lower values branched to the left. High-performing nodes are highlighted in gray. The scores in the lowest branches are averaged QSPR scores over all the MOFs in that branch.

a) Selectivity > 5



b) Selectivity > 10

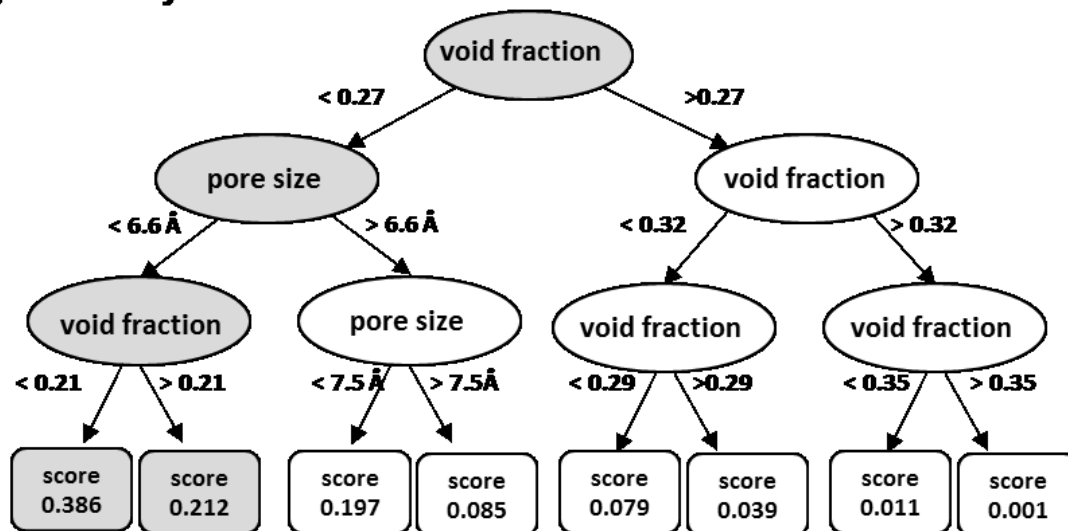


Figure 7.7: Decision tree (DT) models of the CO₂/CH₄ selectivity higher than 5 (a) and 10 (b) of the 32,450 MOFs in the calibration set. Each branch of the tree represents a threshold level given for the value of a property, where the MOFs with values higher than the level threshold branched to the right, while MOFs with lower values branched to the left. High-performing nodes are highlighted in gray. The scores in the lowest branches are averaged QSPR scores over all the MOFs in that branch.

Simple DT models yielded reasonably accurate and easy to use QSPR models but the greatest predictive power lies in nonlinear machine learning techniques such as SVMs capable of handling complex correlation patterns. Therefore, we calibrated nonlinear SVMs to identify high-performing MOFs using the same threshold values of the working capacity and selectivity as the DT models. As expected, we found that the SVM models outperformed the DT models with higher AUC values as shown in Table 7.2. Specifically, MOFs with CO₂ working capacities higher than 2 mmol/g and 4 mmol/g were recognized with AUC values of 0.916 and 0.919, respectively. For CO₂/CH₄ selectivity, the SVM classifiers identified high-performing MOFs at threshold values of 5 and 10 with the highest AUC values of 0.953 and 0.948, respectively.

Although the QSPR models successfully recognized MOFs in the training set with very high accuracy, the important measure of the predictive capability of the QSPR models is the performance on MOFs not used to calibrate the models. For this purpose, we applied the QSPR models on the test set of the remaining 292,050 MOFs in our database that were not used to train the models. GCMC simulations have been performed on the whole test set allowing us to compare the QSPR predicted results to what we will refer to as the ‘actual’ results. Whereas the DT models allow for ‘rule-of-thumb’ structure-property relationships to be elucidated for rational design, we envision the SVM models to be used for rapid pre-screening of large structural databases or large search spaces. The goal is to drastically reduce the number compute intensive GCMC simulations that are performed during the virtual screening process. The performance of the SVM models for pre-screening purposes are quantified in the plots in Figure 7.8. Plotted are the number of MOFs (y-axis) that are classified as high-performing by the SVM model with different sensitivity cut-off values that have been adjusted to recover a certain

percentage of the true-positives (30%, 50% and 90% in Figure 7.8a-c and 30%, 50% and 70% in Figure 7.8d) from the test set. In these plots, the true positives are denoted in blue and the false positives are shown in red, while the dashed line denotes the actual number of ‘high-performing’ MOFs or the total number of true-positives.

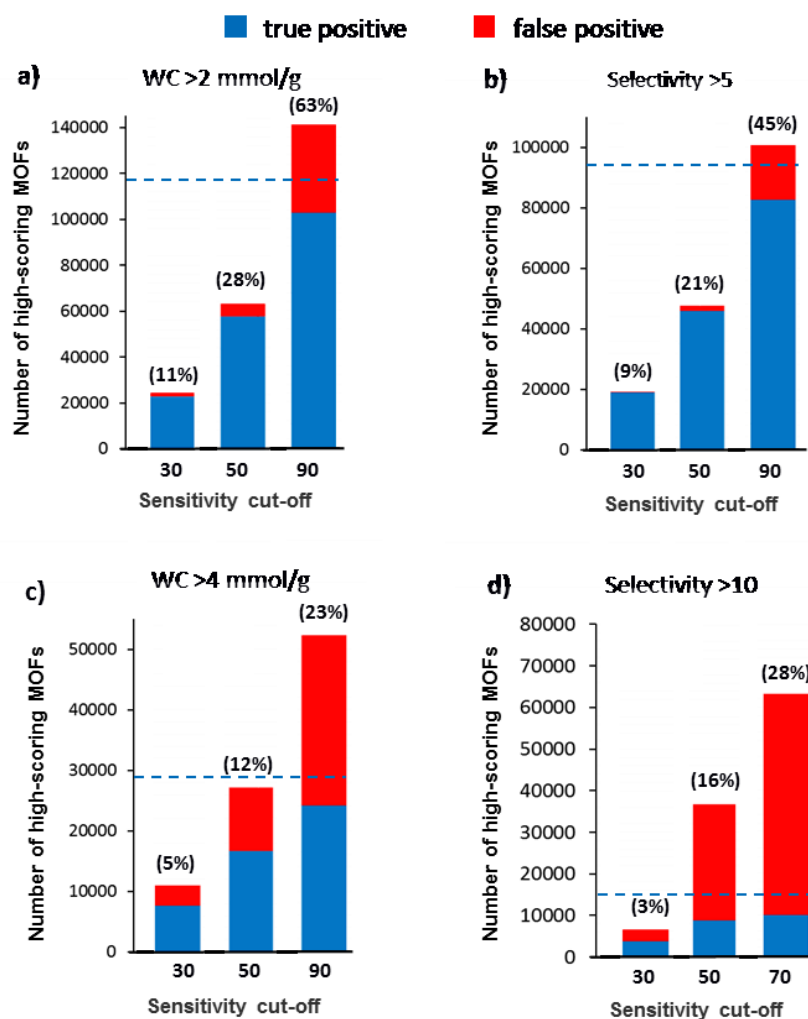


Figure 7.8: Number of MOFs identified by the SVM classifiers as having a working capacity greater than 2 mmol/g (a) or greater than 4 mmol/g (b) and CO₂/CH₄ selectivity greater than 5 (c) or greater than 10 (d) from the test set of 292,050 MOFs. Three different sensitivity cut-offs are reported for each. The dashed line indicates the total number of actual high-performing MOFs (all of the true positives), whereas the number in parenthesis gives the percentage of the total test set that the QSPR classifier identified as high-performing.

For the lower thresholds of 2 mmol/g working capacity and selectivity of 5 the SVM classifier returns mostly true-positives. This can be seen in Figures 7.8a and 7.8b where the red fraction is small compared to the blue fraction. At a sensitivity cut-off of 30% (Figures 7a and 7b) virtually all of the MOFs returned as high-performing by the classifier are actually true-positives. Thus, if these MOFs are sent for more computationally expensive screening, very little compute time will be ‘wasted’ on low performing MOFs. On the other hand, with this sensitivity cut-off only 30% of all true positives are recovered by the classifier. Depending on the virtual screening scenario, one may prefer to maximize the “benefits” of the screening by using the 90% sensitivity cut-off, which returns more high-performing MOFs, albeit at the expense of including a large number of false positives.

For the higher thresholds of working capacity (>4 mmol/g) and selectivity (>10) the SVM classifiers (Figures 7.8c and 7.8d) return a greater fraction of false positives. At the same time, there are 4-6 times fewer true positives than that yielded by the lower thresholds. It is notable that the classifier for working capacity >4 mmol/g in Figure 7.8c can recover 90% of all the true positives, while classifying only 23% of the test set as high-performing. In practical terms, one would only need to perform GCMC simulations on 23% of the database in order to recover 90% of the high-performing MOFs. The model for scoring MOFs with selectivity >10 is the least accurate of the models, that returns the greatest fraction of false-positives. However, even with this model, one can recover 70% of all the true-positives while screening only 28% of the total database.

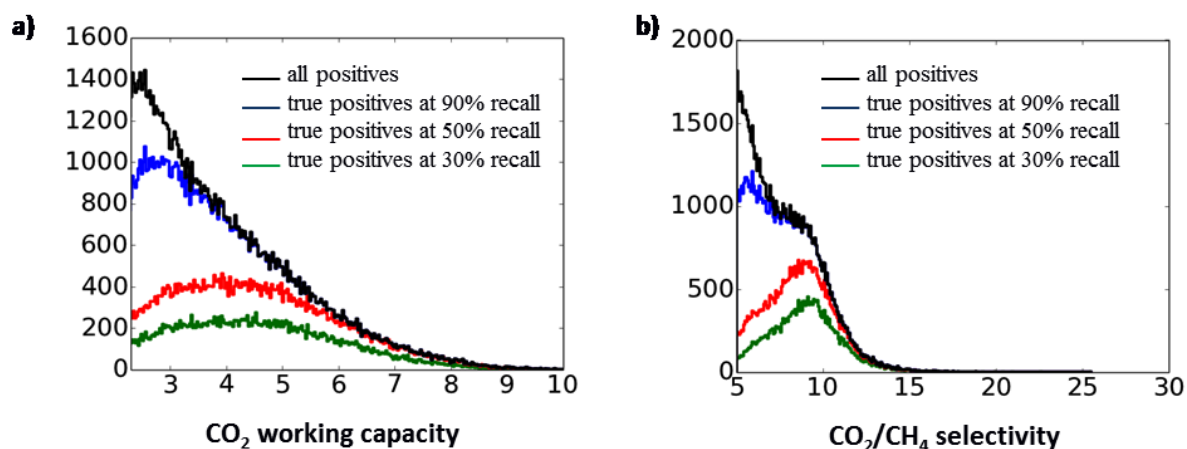


Figure 7.9: Distribution of the calculated CO₂ working capacity and (b) CO₂/CH₄ selectivity of all true positives in the test set using the classifiers of working capacity >2 mmol/g and selectivity >5, compared to the distribution of true positives recovered by the QSPR classifiers at different cut-offs.

The sensitivity cut-off is used to adjust the number of structures the classifier returns as high-performing. With a low sensitivity cut-off (i.e. 30%), 70% of the true positives are missed. Thus, the question of importance is – what is the quality of the true positives that are missed? To examine this, in Figure 7.9 we plot the distributions of CO₂ working capacity and CO₂/CH₄ selectivity for the true positives captured by the classifiers at various cut-offs (coloured) and compare these to the distribution of all the actual true positives (black). These distributions are shown for the CO₂ working capacity classifier with a high/low performance threshold of 2 mmol/g in Figure 7.9a and for the CO₂/CH₄ selectivity classifier with a threshold of 5 in Figure 7.9b. The majority of the ‘actual’ high-performing MOFs (black) have CO₂ working capacity less than 4 mmol/g, while there are still hundreds of MOFs that have uptake capacities greater than 6 mmol/g, which is considered very high under these conditions. Deviation of the QSPR predicted true positives (coloured lines) from the total true positive distribution (black line) reveals the working capacity of the high-performing MOFs that the QSPR classifiers miss (the false-negatives). For all cut-offs, we see that the false-negatives are strongly biased towards the

low end of CO₂ working capacity and CO₂/CH₄ selectivity. For example, in Figure 7.9a, the distribution of working capacity for the 30% sensitivity cut-off (green) is well below the black distribution in the range between 2-4 mmol/g, meaning that the classifier is missing most of the MOFs in that uptake range. However, in the uptake range of 6 mmol/g or higher, the green and black distributions nearly match, revealing that the classifier is capturing most of the MOFs with the best working capacities. Similar behaviour was observed for the selectivity in Figure 7.9b. The majority of the high-performing MOFs have CO₂ selectivity less than 7, while there are still hundreds of MOFs that have selectivity greater than 10. More specifically in Figure 7.9b, the distribution of selectivity for 30% sensitivity cut-off (green) is well below the black distribution in the selectivity range between 5-8, meaning that the classifier is ‘missing’ most of the MOFs in that selectivity range. However, in the selectivity range of 10 or higher, the green and black distributions nearly match, revealing that the classifier is capturing most of the MOFs with the best selectivities.

As a different evaluation of the model prediction performance, we have examined how many of the top-1000 high-performing MOFs are captured by the classifiers. These are given in Table 7.3 for different sensitivity cut-offs for the working capacity >2 mmol/g and selectivity > 5 classifiers. With the very selective 30% sensitivity cut-off, the classifier for working capacity > 2 mmol/g is still able to recover 396 of the top 1000 MOFs from the 292,050 member test set. If the sensitivity cut-off is adjusted to be less selective, then as expected, the number of the top 1000 MOFs increases up to 995 for the 90% sensitivity cut-off. Similar results are obtained for the selectivity, albeit with less favourable results. Nonetheless, Table 7.3 reveals that using the QSPR classifier with the selective 50% cut-off to pre-screen materials for methane purification,

one would recover ~90% of the top-1000 high-performing materials while only carrying out expensive simulations on ~20-30% of the total search space.

Table 7.3: Number of the top 1000 MOFs from the test set that the classifiers of working capacity >2 mmol/g and selectivity>5 correctly identifies with different cut-offs. The percentage in parenthesis is the fraction of the test set that would require GCMC simulations.

Parameter	Sensitivity cut-off		
	30%	50%	90%
working capacity	396 (11%)	872 (28%)	995 (63%)
selectivity	711 (9%)	898 (21%)	1000 (45%)

7.3. Conclusions

QSPR models of CO₂/CH₄ separation parameters under methane purification conditions have been developed using only simple geometric features. The models were constructed using a diverse set of over 32,000 hypothetical MOFs for which the separation parameters were computed with GCMC simulations. The binary decision tree analysis provides ‘rules-of-thumb’ to follow in rational design of MOFs, based on optimal range of pore size, void fraction and surface area. The ‘rule-of-thumb’ for a MOF to possess a CO₂/CH₄ selectivity >10 is to have void fraction lower than 0.27 and pore size lower than 6.6 Å. The “rule-of-thumb” for CO₂ working capacity >4 mmol/g is to have pore size lower than 8.5 Å and gravimetric surface area greater than 2300 m²/g. Meanwhile, the SVM models were developed that allowed the rapid identification of high-performing MOF candidates. This could be used as a filtering tool when screening large databases of materials by rapidly discarding the majority of the poor candidate materials. Significant computation time would then be saved as only a fraction of the database would need to be screened with compute intensive GCMC simulations. It was demonstrated on a test set of 292,050 MOFs that the classifiers could recover ~900 of the top 1000 MOFs in the

database, while only flagging ~25% of the total database for more compute intensive screening. This would result in an approximately 4 times speed up in screening compared to a full brute force approach where GCMC simulations are performed on all MOFs. The general approach can also be applied to other classes of materials as long as relevant data is made available to calibrate QSPR models.

7.4. References

- (1) Dong, F.; Kodama, A.; Goto, M.; Hirose, T. In *Fundamentals of Adsorption*; Le Van, D., Ed.; Kluwer Academic Publisher: Boston, 1998.
- (2) Molecular Gate™ Systems: Nitrogen Rejection and Carbon Dioxide Removal Made Easy. Product Brochure, Guild Associates Inc., Dublin, OH. <[http:// www.moleculargate.com](http://www.moleculargate.com)>.
- (3) Zhang, Z.; Yao, Z.; Xiang, S.; Chen, B. *Energy Environ. Sci.* **2014**, 7 (9), 2868.
- (4) Xiang, S.; He, Y.; Zhang, Z.; Wu, H.; Zhou, W.; Krishna, R.; Chen, B. *Nat. Commun.* **2012**, 3, 954.
- (5) Li, J.-R.; Kuppler, R. J.; Zhou, H.-C. *Chem. Soc. Rev.* **2009**, 38 (5), 1477.
- (6) Wilmer, C. E.; Farha, O. K.; Bae, Y.-S.; Hupp, J. T.; Snurr, R. Q. *Energy Environ. Sci.* **2012**, 5 (12), 9849.
- (7) 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 (7), 633.
- (8) Keskin, S.; van Heest, T. M.; Sholl, D. S. *ChemSusChem* **2010**, 3, 879.
- (9) Qiao, Z.; Zhang, K.; Jiang, J. *J. Mater. Chem. A* **2016**, 4 (6), 2105.
- (10) Allen, F. H. *Acta Crystallogr. Sect. B Struct. Sci.* **2002**, 58 (3 PART 1), 380.
- (11) Smit, B.; Maesen, T. L. M. *Chem. Rev.* **2008**, 108 (10), 4125.
- (12) Kadantsev, E. S.; Boyd, P. G.; Daff, T. D.; Woo, T. K. *J. Phys. Chem. Lett.* **2013**, 4 (18), 3056.
- (13) Wilmer, C. E.; Kim, K. C.; Snurr, R. Q. *J. Phys. Chem. Lett.* **2012**, 3, 2506.
- (14) Kim, J.; Abouelnasr, M.; Lin, L. C.; Smit, B. *J. Am. Chem. Soc.* **2013**, 135, 7545.
- (15) Martin, R. L.; Simon, C. M.; Smit, B.; Haranczyk, M. *J. Am. Chem. Soc.* **2014**, 136, 5006.
- (16) Boyd, P. G.; Woo, T. K. *CrystEngComm* **2016**.
- (17) Fernandez, M.; Tre, N. R.; Woo, T. K. *J. Phys. Chem. c* **2013**, 117, 14095.
- (18) Martin, R. L.; Willems, T. F.; Lin, L.-C.; Kim, J.; Swisher, J. a; Smit, B.; Haranczyk, M. *Chemphyschem* **2012**, 13 (16), 3595.

- (19) Fernandez, M.; Boyd, P. G.; Da, T. D.; Aghaji, M. Z.; Woo, T. K. *J. Phys. Chem. Lett.* **2014**.
- (20) Wilmer, C. E.; Leaf, M.; Lee, C. Y.; Farha, O. K.; Hauser, B. G.; Hupp, J. T.; Snurr, R. Q. *Nat. Chem.* **2012**, 4 (February), 83.
- (21) Gale, J. J. *Chem. Soc., Faraday Trans.* **1997**, 93 (4), 629.
- (22) Fateen, S. E. K.; Khalil, M. M.; Elnabawy, A. O. *J. Adv. Res.* **2013**, 4 (2), 137.
- (23) Campañá, C.; Mussard, B.; Woo, T. K. *J. Chem. Theory Comput.* **2009**, 5, 2866.
- (24) Martin, M. G.; Siepmann, J. I. *J. Phys. Chem. B* **1998**, 102 (97), 2569.
- (25) Quinlan, J. R. *C4.5: Programs for Machine Learning*; California: Morgan Kaufmann Publishers: San Mateo, 1993.
- (26) Cortes, C.; Vapnik, V. *Mach. Learn.* **1995**, 20 (3), 273.
- (27) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, E. *J. Mach. Learn. Res.* **2011**, 12, 2825.
- (28) Sonnenburg, S.; Raetsch, G.; Henschel, S.; Widmer, C.; Behr, J.; Zien, A.; De Bona, F.; Binder, A.; Gehl, C.; Franc, V. *J. Mach. Learn. Res.* **2010**, 11, 1799.
- (29) Cavenati, S.; Grande, C. a; Rodrigues, a E. *J. Chem. Eng. Data* **2004**, 49, 1095.
- (30) Willems, T. F.; Rycroft, C. H.; Kazi, M.; Meza, J. C.; Haranczyk, M. *Microporous Mesoporous Mater.* **2012**, 149 (1), 134.
- (31) Hamon, L.; Jolimai, E.; Pirngruber, G. D. *Ind. Eng. Chem. Res.* **2010**, 49, 7497.
- (32) Vaidhyanathan, R.; Iremonger, S. S.; Dawson, K. W.; Shimizu, G. K. H. *Chem. Commun. (Camb)*. **2009**, No. 35, 5230.
- (33) Yang, Q.; Zhong, C. *J. Phys. Chem. B* **2006**, 110 (36), 17776.
- (34) Bae, Y.-S. S.; Farha, O. K.; Spokoyny, A. M.; Mirkin, C. a; Hupp, J. T.; Snurr, R. Q.; Mulfort, K. L.; Frost, H.; Ryan, P.; Punnathanam, S.; Broadbelt, L. J.; Hupp, J. T.; Snurr, R. Q. *Chem. Commun. (Camb)*. **2008**, 0 (35), 4135.
- (35) Bae, Y.; Mulfort, K.; Frost, H.; Ryan, P. *Langmuir* **2008**, No. 18, 8592.
- (36) Mu, B.; Schoenecker, P.; Walton, K. *J. Phys. ...* **2010**, 6464.
- (37) An, J.; Geib, S.; Rosi, N. *J. Am. Chem. ...* **2009**, 38.
- (38) Yazaydin, a O.; Snurr, R. Q.; Park, T.-H.; Koh, K.; Liu, J.; Levan, M. D.; Benin, A. I.; Jakubczak, P.; Lanuza, M.; Galloway, D. B.; Low, J. J.; Willis, R. R. *J. Am. Chem. Soc.* **2012**, 131 (51), 18198.
- (39) Leus, K.; Couck, S.; Vandichel, M.; Vanhaelewyn, G.; Liu, Y.-Y.; Marin, G. B.; Van Driessche, I.; Depla, D.; Waroquier, M.; Van Speybroeck, V.; Denayer, J. F. M.; Van der Voort, P. *Phys. Chem. Chem. Phys.* **2012**, 14 (44), 15562.

8 Investigation of Atomic Property-Weighted Radial Distribution Function and Voronoi Hologram Descriptors in Machine Learning Recognition of High Performing Metal Organic Frameworks for Natural Gas Purification

Abstract

Metal-organic frameworks have attracted interest due to high structural diversity and gas separation applications such as natural gas purification. In this direction, recent *In-silico* generation of libraries with thousands to millions of hypothetical materials and subsequently high-throughput (HT) computational screening as a part of material discovery process, illustrate the demand of more efficient computational tools compared to traditional grand canonical Monte Carlo (GCMC) simulations which can be resource-demanding. Here, we report machine learning classifier models for CO₂/CH₄ separation parameters (i.e., CO₂ working capacity and CO₂/CH₄ selectivity) that utilize separately the Voronoi hologram and Atomic Property-Weighted Radial Distribution Function (AP-RDF) descriptors. We compare their performance with previously reported classifiers models. For this, the classifiers are trained on 32,450 MOFs and then are tested on the 292,050 MOFs that are not part of the training set. From the comparison, it is found that including AP-RDF and Voronoi hologram descriptors into the classifiers improves the performance of classifiers by 20% in capturing high-performing MOFs.

8.1 Introduction

Growing concern over increasing atmospheric CO₂ concentrations¹ and its effects on climate change² have prompted immediate action to reduce CO₂ emissions. The two main strategies currently being proposed to reduce emissions from fossil fuel usage when generating power are the following: (1) carbon capture and storage (CCS)³ and (2) the replacement of coal-

or oil-based energy sources with natural gas⁴. Both strategies involve the separation of CO₂ from a gas mixture (i.e., CO₂/N₂ separation for post-combustion CCS, and CO₂/CH₄ separation in the case of NG purification), which is traditionally performed using aqueous amines (primarily ethanolamine) that react with CO₂ to form carbamates.⁵ The regeneration of aqueous amines requires the breaking of the N-C bond in carbamate using heat. Due to the high heat capacity of aqueous amines, the process is very energy intensive. An alternative method is to perform the CO₂ separation using a solid adsorbent that selectively binds CO₂ via physical adsorption, and can be regenerated using less energy once saturated.⁶

The success of the adsorption-based CO₂ separation process relies on the development of novel adsorbent materials that possess a high CO₂ capacity, and has a strong affinity for CO₂ over other gases (N₂ or CH₄). Owing to the unique properties of metal organic framework materials (MOFs), such as exceptional internal surface areas and chemical tunability, this class of porous materials has emerged as potential adsorbents for the adsorption-based CO₂ separation process.^{7,8} However, finding the optimal MOF for this purpose is remarkably challenging, as millions of MOFs can potentially be constructed from virtually limitless combinations of inorganic and organic secondary building units (SBUs).⁹ In order to overcome this combinatorial problem, a large-scale computational screening of hypothetical materials, which combines a large-scale combinatorial design and advanced computer simulations, has been undertaken in the MOF community. Using this approach, several high-throughput (HT) computational studies have been reported on MOFs for CO₂-related separation applications.^{10–13} While informative, these studies are limited to hundreds of thousands of MOFs, which represents only a small fraction of the millions of MOFs which make up the current databases

of experimental and theoretical MOFs.⁹ Moreover, most of these studies employed Grand Canonical Monte Carlo (GCMC) simulations for the prediction of adsorption properties of MOFs, which become practically impossible when screening millions of MOFs due to the fact that the calculation of partial atomic charges required for a GCMC simulations takes several minutes per MOF.^{14,15} Thus, new computational methods that do not rely on atomistic molecular simulations are essential.

Machine learning (ML) methods have been integrated into the study of gas adsorption of MOFs as an attractive alternative to high-throughput screening using computationally expensive GCMC simulations, and they have shown capability towards the screening of MOFs for a variety of gas separation applications, including CO₂ capture and NG purification.^{16,17} More specifically, it was demonstrated for CO₂ capture in MOFs that ML classifiers which incorporate atomic property-weighted radial distribution function (AP-RDF) as “descriptors” are able to rapidly identify 90% of high-performing MOFs (CO₂ uptake > 1 mmol/g at 0.15 bar) from a test set of 292,050 MOFs, indicating that AP-RDF-based ML classifiers could be used as a filtering tool prior to GCMC simulations during the screening of large databases of materials for CO₂ capture.¹⁶ With regards to NG purification (i.e., CO₂/CH₄ separation), it was demonstrated that using ML classifiers that incorporate simple geometrical features, namely void fraction, max-pore diameter, and surface area, one could recover 70% of high-performing MOFs (CO₂/CH₄ selectivity > 10 at 10 bar) while only performing GCMC simulations on 28% of the total database.¹⁷ Although this is a reasonable result given the significantly reduced amount of time-intensive GCMC simulations performed, using the simple geometrical descriptors-based classifiers missed 30% of the high-performing MOFs for CO₂/CH₄ separation. To improve these

models for CO₂/CH₄ separation, it was reported that one could incorporate descriptors that describe the pore shape or chemistry of the MOF.¹⁷ Recently, Martin et al.^{18,19} introduced a new geometry-based descriptor, the Voronoi hologram, which contains information on shape and size of void space of a material. Unlike simple one-dimensional geometrical descriptors such as surface area, the Voronoi hologram consists of higher-dimensional descriptors which represent the pore topology of a material. More specifically, each hologram descriptor describes the shape of some localized part of pore, whereas the surface area (or other simple one-dimensional descriptors) describes the entire pore network with a single number. Indeed, Voronoi hologram reflects the arrangement of atoms within a pore of a material, which is not addressed by one-dimensional geometrical descriptors. A detailed definition of the Voronoi hologram is given in section 8.2.3. Using Voronoi hologram descriptors, Martin et al.¹⁸ were able to rapidly identify high-performing zeolites for CO₂ capture. Motivated by the performance of the Voronoi hologram and AP-RDF descriptors in the recognition of promising candidates for CO₂ capture, and with the aim of improving previously reported classifiers models for CO₂/CH₄ separation, we have investigated the performance of ML classifiers using the Voronoi hologram descriptors and AP-RDF descriptors for CO₂/CH₄ separation.

In this work, we present ML classifier models for CO₂/CH₄ separation parameters (i.e., CO₂ working capacity and CO₂/CH₄ selectivity) that incorporate separately the Voronoi hologram and AP-RDF descriptors, and we compare their performance with previously reported classifiers models. All classifiers were trained on the GCMC calculated adsorption properties of a diverse set of 32,450 MOFs using nonlinear support vector machines (SVMs), accompanied by descriptor selection where genetic algorithm (GA) and cross validation techniques were

applied. These ML classifiers were then tested on a set of 292,050 diverse MOF structures and compared to the results of GCMC simulations. In brief, the remainder of this chapter is as follows: first, an overview of our approach, including a description of the data set, computational methods, descriptors, computation methods, and selection procedure, is given. Then an account of how the classifiers were trained is provided, followed a discussion of the results and conclusions.

8.2 Methods

Our approach to develop machine learning classifiers based on Voronoi holograms and AP-RDF descriptors of MOFs for the prediction of methane purification consists of the following basic steps: (1) A training set of MOFs is selected from a large database of hypothetical MOFs. (2) Using GCMC simulations, the CO₂ working capacity and CO₂/CH₄ selectivity are calculated under typical methane purification conditions. (3) AP-RDF, geometric and hologram descriptors for all MOFs are evaluated. (4) A genetic algorithm is used to prune the hologram descriptors using cross-validation data from the training set. (5) Machine learning SVM models are trained using each of these descriptors using the training set data. 6) The models are tested on the test set of MOFs.

8.2.1 Hypothetical MOF dataset

The training and test set of MOFs were selected from a large database of 324,500 hypothetical MOFs developed in the Woo lab. This database was generated from 66 structural building units (SBUs) and 19 metal SBU's using a simple geometric algorithm¹¹ that was developed by a former Ph.D. student in the lab. Care was taken to ensure reasonable non-bonded atom distances during structure generation by ensuring that atoms did not enter within

half the sum of their Van der Waal radii. Functional group placement was accomplished by replacing available hydrogens on organic SBUs within each material. Where available, symmetry was imposed on functional group placement so that symmetric hydrogens would be replaced with the same functional group. To relax the atomic positions, the UFF force field was used to optimize all atomic positions in the GULP 4.0 software package²⁰. Bonding information was included in each hypothetical MOF to ensure the correct connectivity information was provided to the optimizer. 10% of the database was randomly chosen to form a training set, while the remaining 90% of the database was considered as a test set.

8.2.2 Grand Canonical Monte-Carlo (GCMC) simulations

Atomistic grand canonical Monte Carlo (GCMC) simulations were performed to estimate the adsorption of a binary mixture of CO₂:CH₄ with a composition of 1:9 in all of the hypothetical MOFs. Intermolecular interaction energies were computed through a sum of Lennard-Jones (LJ) and point charge Coulomb potentials. LJ parameters between atoms of different types were calculated using the Lorentz-Berthelot mixing rules (i.e., geometric average of well depths and arithmetic average of diameters). LJ parameters for framework atoms were taken from the Universal Force Field²¹. The MEPO-QEq partial atomic charges on the MOF framework were used to evaluate electrostatic interactions.¹⁴ The force field parameters for CO₂ were taken from the force field developed by Garcia-Sanchez et al.,²² while for CH₄ were taken from TRAPPE force field parameters.²³ All GCMC simulations of CH₄-CO₂ adsorption included an M-cycle equilibration period followed by an M-cycle production run, where M was 40,000. A cycle consists of n Monte Carlo steps, where n is equal to the number of guest molecules (which fluctuates during a GCMC simulation). All simulations included random

insertion, deletion, and translation moves of the gas molecules with equal probabilities. Atoms in the MOF were held fixed at their crystallographic positions. A LJ cutoff distance of 12.5 Å was used for all simulations. A 2x2x2 unit cell of every MOF crystal was used for the simulations.

8.2.3 Descriptors

Geometrical descriptors: Six structural features, namely dominant pore diameter, maximum pore diameter, void fraction, gravimetric surface area, volumetric surface area, and density, in all possible combinations, were used to build multi regression models for predicting the CO₂/CH₄ selectivity. From all of the combinations, the set including maximum pore diameter, void fraction, and gravimetric surface area yielded the best model and was chosen as the geometrical descriptors set. The maximum pore diameter is the largest sphere that can be grown inside the framework when calculating the pore size distribution. The void fraction (i.e., the fraction of empty space within a MOF) is calculated by the helium insertion method and is described elsewhere.²⁴ The gravimetric surface areas were determined by geometric calculations described previously.²⁵ All the geometrical features of the MOFs in our database were calculated using the zeo++ code.²⁶

Atomic Property Weighted Radial distribution Function (AP-RDF) descriptors: The AP-RDF descriptors²⁷ (i.e., scores) for each MOF in our database were calculated by:

$$RDF^P(R) = f \sum_{i,j}^{all\ atom\ pairs} P_i P_j e^{-B(r_{ij}-R)^2} \quad 8.1$$

where the summation is over all atom pairs in the unit cell, r_{ij} is the minimum image convention distance of these pairs, B is a smoothing parameter, f is simply a scaling or normalization factor, and P is an element-based atomic property for incorporating chemical information into the descriptors. For each MOF, the radial probabilities were weighted by three tabulated atomic properties, specifically the: electronegativity, polarizability, and van der Waals volume, whose values are given in Table 8.1. Using a distance range of 2.0 to 30 Å in steps of 0.25 Å, 113 scores were computed per atom property, yielding a total of 339 AP-RDF descriptors for each MOF. Examples of AP-RDF profiles of two MOFs are given in Figure 8.1.

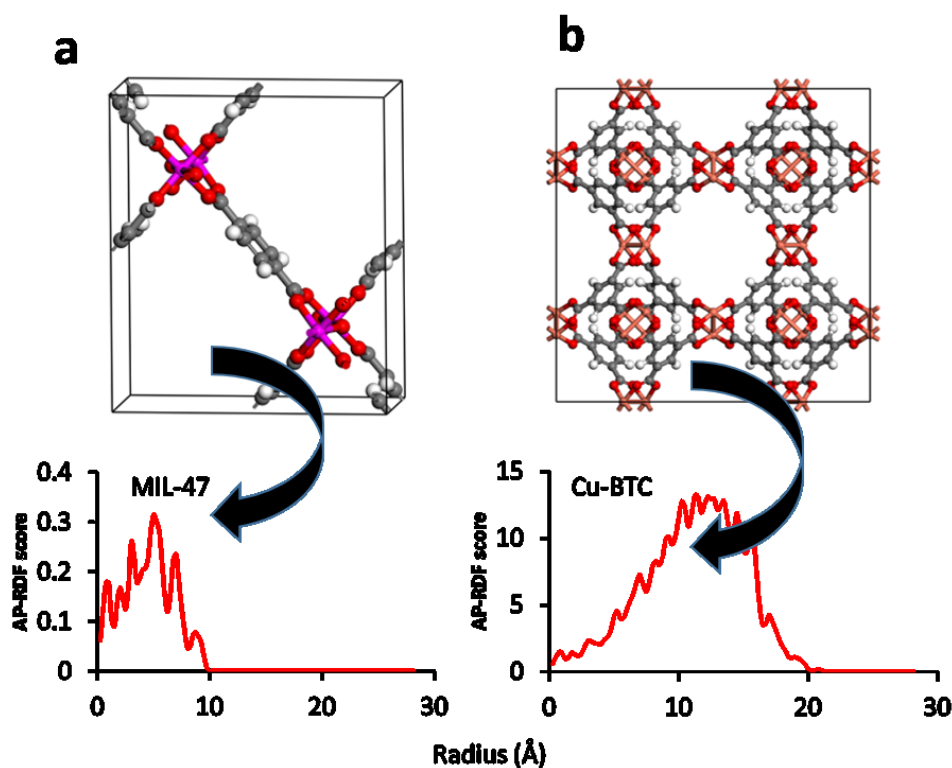


Figure 8.1: Averaged electronegativity weighted AP-RDF plot of two MOFs, a) MIL-47 and b) Cu-BTC

Table 8.1: Tabulated atomic properties used to weight the radial distribution function (RDF) descriptors. Each property value is normalized by the values for the carbon atom.

Element	Electronegativity	Polarizability	VdW Volume	Element	Electronegativity	Polarizability	VdW Volume
Al	0.6242	3.8636	1.6264	I	1.0117	3.0398	1.7280
B	0.8285	1.7216	0.7962	In	0.7786	5.3889	1.4633
Ba	0.2487	1.1252	1.7941	Mg	0.4800	6.3600	1.0538
Br	1.1723	1.7330	1.3835	Mn	0.6045	5.2220	1.8287
C	1.0000	1.0000	1.0000	N	1.1631	0.6250	0.6949
Cd	0.7210	4.6600	1.3941	Ni	0.7283	3.8636	0.7643
Cl	1.2655	1.2386	1.0347	Ol	1.3307	0.4557	0.5120
Co	0.7283	4.2614	1.5609	P	0.9159	2.0625	1.1814
Cr	0.7200	6.4444	1.6283	S	1.0768	1.6477	1.0882
Cu	0.7403	3.4659	0.5120	Si	0.7786	3.0568	1.4244
F	1.4567	0.3165	0.4100	Sn	0.8369	4.3750	2.0415
Fe	0.7283	4.7727	1.8287	V	0.6392	7.0455	1.6283
Gd	0.7283	13.3523	3.2418	Zn	0.8095	4.0341	1.7084
H	0.9439	0.3790	0.2989	Zr	0.5216	10.1705	1.6283

Voronoi Hologram descriptors: The Voronoi hologram descriptors of a MOF in our database were obtained from the Voronoi decomposition of a MOF's accessible void space. In mathematics, a Voronoi decomposition is a partitioning of a space containing n sites into n cells such that each cell contains only one site and all points within any given cell are closer to its contained site than to any other site.²⁸ Such decomposition forms a Voronoi network, which is shown in Figure 8.2 for a two-dimensional space.

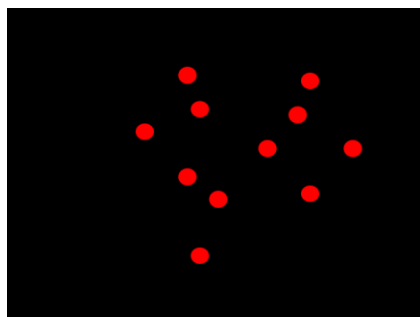


Figure 8.2: Example of Voronoi decomposition of two-dimensional space containing n sites, (red circle) into n cells such that each cell consists of all points closer to its site than to any other.

Similarly, the Voronoi decomposition of a space surrounding a MOF's atoms, as shown in Figure 8.3a, yields a periodic Voronoi network represented by edges and nodes, where each edge is defined by its length (L) and the radii of the two nodes it connects (r_a and r_b) as shown by the red oval in Figure 8.3a. Here, the radius of the node corresponds to the radius of the largest spherical probe that can be placed at the node without colliding with any atom. In other words, the radius of the node is the distance from the centre of the node to the surface of the nearest atom. Using CO_2 probe radius ($1.7 \text{ \AA}$), the accessible nodes and edges are obtained which represent the CO_2 -accessible region of the void space (purple region in Figure 8.3b).

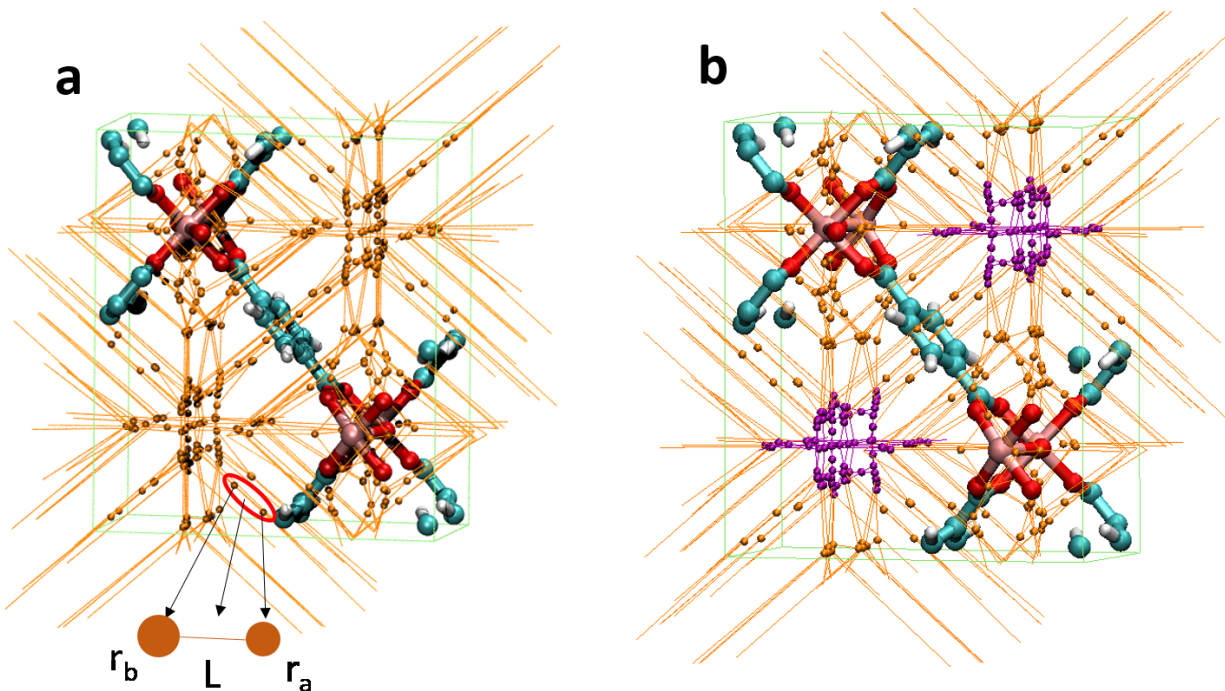


Figure 8.3: a) MIL-47 (purple: V, red: O, cyan: C, and white: H) and its 3D Voronoi network represented by edges (orange lines) and nodes (orange points). b) The CO₂ accessible Voronoi networks are highlighted in purple within the Voronoi network of MIL-47. For clarity of presentation, nodes are represented by spheres of equal, small radii.

Considering continuous variables L , r_a , and r_b , we constructed a binning system for edge lengths and node radii. For this, we computed CO₂-accessible Voronoi networks of 30,000 MOFs randomly chosen from the database, and we then determined the bounds of bins such that computed nodes and edges were equally populated in the bins. This approach (i.e., non-equal width bins) was taken from Smit and co-workers¹⁹ who encoded the pore shapes of hundreds of zeolites as edges and nodes. Similar to their work, 16 non-equal width bins for each L , r_a , and r_b were chosen, which yielded a total of 2176 distinct bins arranged in a three-dimensional, cubic grid. Here, each bin represents one Voronoi hologram descriptor, and its population (i.e., descriptor's value) represents the frequency of occurrence of edges within a bin. Indeed, these 2176 descriptors, along with their values, can be interpreted as a fingerprint for a MOF, which

includes information about the pore shape and void space of the MOF. For each MOF in our database, the Voronoi hologram descriptors were computed using the zeo++ code with the help of several Python scripts. Examples of Voronoi hologram of two MOFs are given in Figure 8.4.

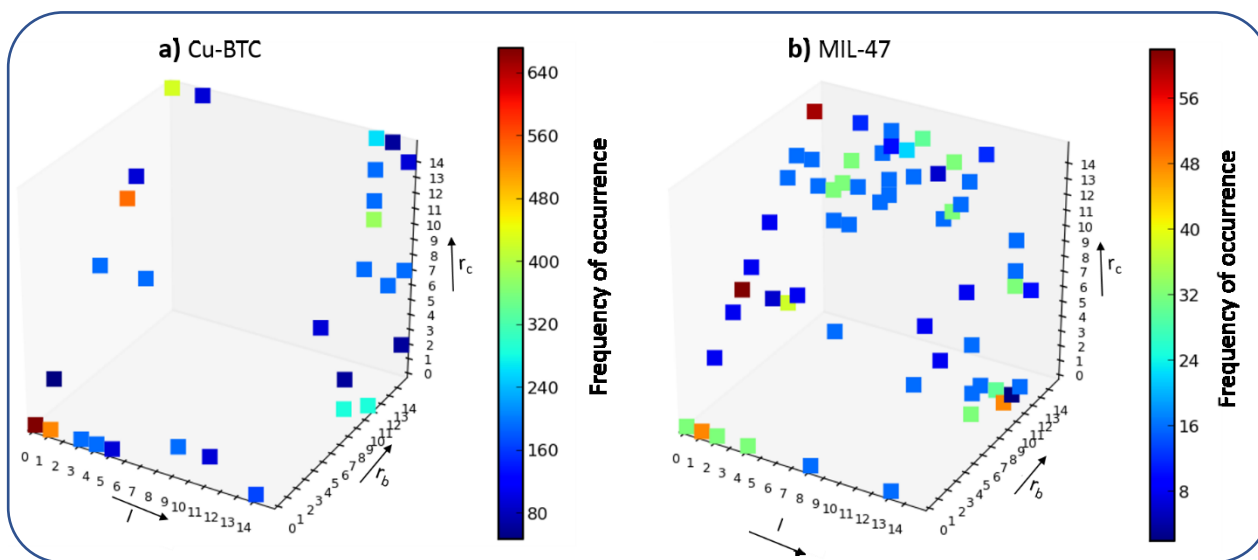


Figure 8.4: Voronoi hologram of the MOFs Cu-BTC (a) and MIL-47 (b). Features are assigned a color on the basis of their frequency of occurrence.

8.2.4 Genetic algorithm-based descriptor selection

The Genetic Algorithm (GA) has been widely used to optimize large and complex search spaces in a timely and effective manner.^{29,30} Its basis relies on evolving a population of candidate solutions, also called individuals, toward a better solution as defined by some sort of ‘fitness’ parameter. The genetic representation of individuals is referred to as a “chromosome”, where each chromosome makes up of pre-determined number of “genes” (five in Figure 8.5) which are typically represented in binary as strings of 0s and 1s. A schematic representation of the steps involved in a genetic algorithm optimization is provided in Figure 8.5. The first step involves generating an initial population of individuals, typically with randomly generated

chromosomes. The population size is usually of a pre-determined size (four in Figure 8.5). In the next step, the ‘fitness’ of each individual is computed with a predefined fitness function. A new generation of individuals is created by mating two parents in the current generation to create a child. Here, the chromosomes of each individual are mixed in some what to create the child chromosome. As in natural selection, the most ‘fit’ individuals are allowed to have a higher probability of mating and to pass on their traits to the future generations.³¹ Mutations are also introduced into the population, where some genes of a chromosome are randomly altered. In addition to mating and mutation operations, promotion operation, also known as elitism, is performed. This involves promoting a small number of top performing chromosomes without any alteration from the current population to next generation. Elitism ensures that high performing genetic material to be present throughout the run of the GA.³² The GA cycle, including evaluation, mating, mutation and elitism, is repeated until the most fit individual remains the same for a sufficient number of generations. Once the best chromosomes stop changing, then the GA is converged.³³

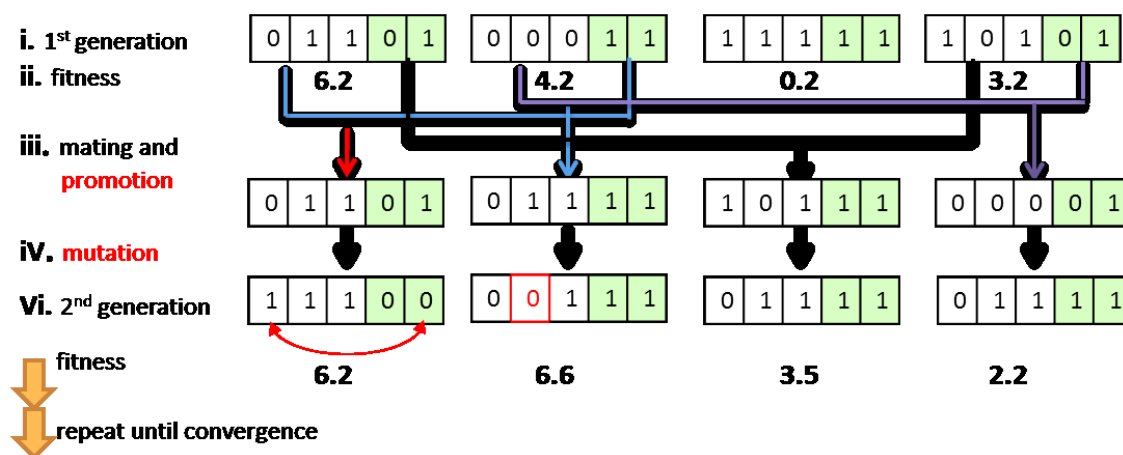


Figure 8.5: A schematic representation of the steps involved in a generalized genetic algorithm optimization.

In this work, a specialized GA code was written in Python by the author. The GA was specifically used to prune the number of Voronoi hologram descriptors which totaled 1260. Here, chromosomes were made up of binary strings of length 1260, where the 1's corresponded to including the descriptors from the total descriptors set, while 0s correspond to leaving out descriptors. In this work, population sizes of 50-75 were used, with the initial population generated by using a random binary string for each individual. To evaluate the performance of chromosomes (i.e., assign fitness values to chromosomes), we used Equation 8.2 as a fitness function in the course of cross validation. The validation involved dividing a training set into four groups, removing one group (called validation set), building an ML model using the remaining three groups for each chromosome, then validating the resulting model on the group removed using Equation 8.2. The removing, building and validating cycle is repeated until all four groups have been validated against each other.

$$\begin{aligned} \text{fitness} = & \sqrt{\text{sensitivity} \times \text{specificity}} \\ & - \left(\left(1 - \sqrt{\text{sensitivity} \times \text{specificity}} \right) * \left| \frac{N}{a - N - 1} \right| \right) \end{aligned} \quad 8.2$$

In Equation 8.2, the first term accounts for the geometric mean of sensitivity (i.e., recovery rate of high performing MOFs) and specificity (i.e., recovery rate of low performing MOFs) obtained from the validation of a ML model incorporating a chromosome as a descriptor set. The second term represents the number of descriptors, N , (i.e., the number of 1's in a chromosome) used to build the ML model where a is an arbitrary number which was selected to be 100, 200, and 300 for the GAs. This fitness function is used to find an optimum subset of descriptors by removing irrelevant or redundant descriptors. An ML model with the optimum subset of descriptors would have a high recovery value of high-performing MOFs while discarding as

many of the low-performing MOFs as possible. After each chromosome was assigned a fitness values based on Equation 8.1, a number of parent chromosomes (as much as 50% of the entire population) was selected to pass on their traits to the next generation. Here, we used the roulette selection method, where the higher “fit” chromosomes the more likely they will be chosen. Once a sufficient number of chromosomes were selected, the mating operation was performed using either a 1 or 2-cut mating scheme (described later). The selection between the 1 and 2-cut mating scheme was chosen by Single Cut Rate parameter which is a probability of performing a 1-cut vs 2-cut mating during mating process. Different values for this parameter were tested, and the optimum value was found to be 0.7, meaning that 70% of the time the 1-cut mating scheme was selected while the other 30% of the time the 2-cut mating scheme was selected.

In the 1-cut mating scheme, the chromosomes of two parents are cut at a single, random crossover point and complementary segments from the two parents are combined. This process is shown in Figure 8.6.

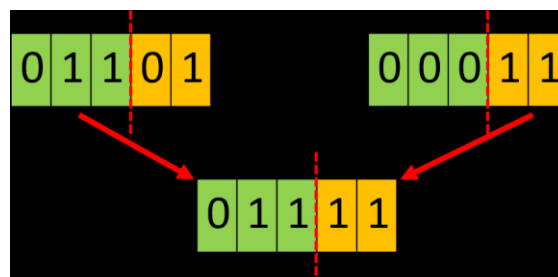


Figure 8.6: Schematic of the 1-cut mating process

The 2-cut mating operation is similar to that of the 1-cut mating scheme except that two cut-points are randomly generated instead one. Subsequently, the middle segment of one

parent is combined with the outer segments of the other parent to form the new child, as shown in Figure 8.7.

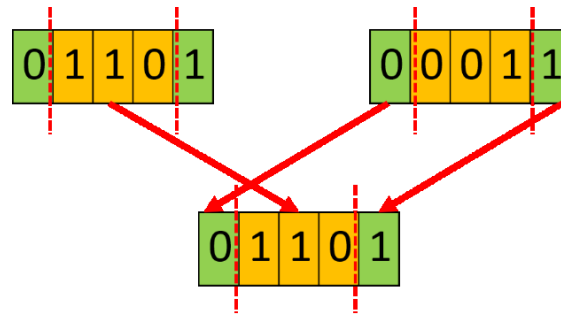


Figure 8.7: Schematic of the 2-cut mating process

Following the mating process, each new chromosome underwent two distinct mutations, “flip” and “swap” Mutations (described later), where the probability of performing each of them was given by Mutation Rate parameter. As with the Single Cut Rate parameter, different values for the Mutation Rate parameter were tested and the optimum value was found to be 0.2.

In the flip mutation, as seen in Figure 8.8a, the binary value of a randomly selected gene was flipped, while in the swap mutation, as given in Figure 8.8b, the binary values of two randomly chosen genes were swapped.

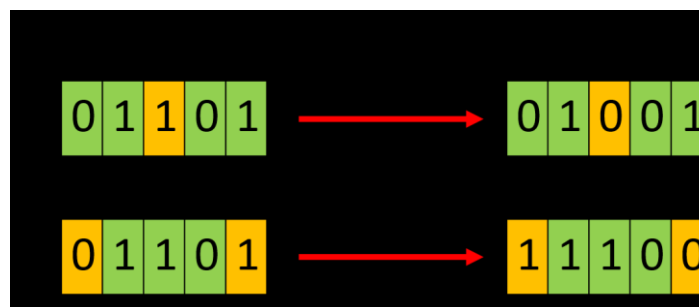


Figure 8.8: Schematic of (a) flip mutation and (b) swap mutation

In addition to the mating and mutation, the elitism operation was also performed where a number of elite chromosomes as much as 10% of the population size were selected without

any changes for the next population. The fitness evaluation, mating, mutation, and elitism cycle which represents one generation of GA was repeated until the elite genes have remained the same for a sufficient number of generations.

8.2.5 Support Vector Machine (SVM) models

The scikit-learn toolbox³⁴ was used to implement SVM models using custom Python code written by the author. A radial-basis-function (RBF) kernel was implemented in the SVM framework to handle the nonlinearity in the data to yield higher accuracies. The optimum width of the RBF function and the regularization parameter of the SVM classifications were tuned by a grid search approach. The descriptor values required for SMVs were normalized to be in the range $[-1, 1]$, where -1 and 1 correspond to the minimum and maximum values, respectively, of each category of data (e.g., the lowest and highest gravimetric surface area values in the database).

8.3 Results and Discussion

8.3.1 Evaluating MOFs for NG Purification

The performance of all 324,500 hypothetical MOFs in the database on NG purification was evaluated using the protocol defined in Chapter 6. Evaluated using that protocol, distributions of the CO₂ working capacity and CO₂/CH₄ selectivity are plotted in Figure 8.9. We found that approximately 10% of the database possessed CO₂ working capacity higher than 4 mmol/g, which is twice the CO₂ working capacity for zeolite 13X currently used in commercial PSA systems to scrub CO₂ from natural gas.³⁵ Similarly, 10% of the database had CO₂/CH₄ selectivities greater than 10, which is comparative to the CO₂/CH₄ selectivity of 13 for zeolite 13X under the conditions studied here. Encouraged by these observations, we defined a

threshold of high-performance for each metric: 4 mmol/g for the CO₂ working capacity and 10 for the CO₂/CH₄ selectivity. These thresholds were used in the training procedure of SVMs classifiers, which are discussed in the next section.

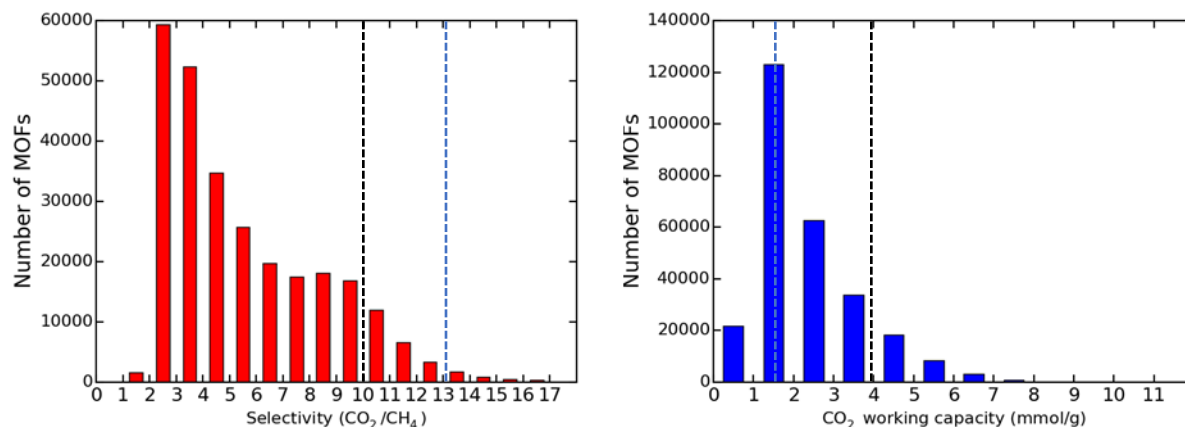


Figure 8.9: Distribution of CO₂/CH₄ selectivity and CO₂ working capacity of entire database. Black dashed lines represent the ‘high-performing’ thresholds used in the current QSPR models, while the blue dashed line shows the selectivity and working capacity of zeolite 13X that is currently in use

8.3.2 Training of Support Vector Machine Classifiers

For each type of descriptors (i.e., simple geometrical, AP-RDF, and Voronoi hologram descriptors), two sets of SVM classifiers were trained, one to recognize high-performing MOFs with CO₂ working capacity above 4 mmol/g and the other to recognize MOFs with CO₂/CH₄ selectivity above 10. In the cases of the simple geometrical and AP-RDF descriptors, a training procedure accompanied by the optimization of SVMs parameters (i.e., the width of the radial basis kernel and the regularization parameter) was performed similar to our previously reported work^{16,17} for each property. Every MOF in the training set was labeled 1 or 0 depending on whether its working capacity or selectivity was above or below the corresponding threshold. Then, SVM classifiers were trained to return a score between 1 and 0 for each MOF,

which is then transformed to a binary label indicating high-performing or low-performing MOF by choosing a cutoff between 0 and 1. It is worthy to note that once the classifier is created, the cut-off value can be adjusted when applying the classifier in order to adjust the number of ‘high performing’ structures that are returned, depending on the user’s needs. For the Voronoi hologram descriptors, as the number of non-zero descriptors was too large (1296), the descriptors had to be ‘trimmed’ using a GA descriptor selection algorithm we developed, which is discussed in the methods section. Using a fitness function given in Equation 8.2 the GA descriptor selection trimmed the number of Voronoi hologram descriptors from 1296 down to 52 for the selectivity classifier and down to 68 for the working capacity classifier.

To compare the performance of the SMV classifiers for each property, the cut-off value for each classifier was tuned to its optimum value such that the classifier captured as many of the actual high performing structures as possible in the course of a cross-validation, while discarding as many of the actual low performing structures as possible. Then, the performance of each SMV classifier at its corresponding optimum cut-off value was quantified using the following G-mean score function which assesses the model’s ability to perfectly classify high-performing and low-performing structures.

$$G - mean = \sqrt{sensitivity \times specificity} \quad 8.3$$

In this equation, sensitivity is the recovery rate of actual high-performing MOFs while specificity is the rate of discarding actual low performing MOFs. Obviously, G-mean becomes close to 1 when the sensitivity is close to 1 (i.e., all the actual high-performing MOFs are recovered by the classifier) and the specificity becomes close to 1, (i.e., all the actual low-performing MOFs are discarded by the classifier). The reason to define the above score function

goes back to the goal of this work, which is the development of classifiers as filtering tools for the screening of large databases of structures. As given by the G-mean score in Table 8.2, both GA-derived VH and AP-RDF classifiers perform better than the simple geometric classifier. The GA-derived VH and AP-RDF classifiers recovered more than 90% of the actual high-performing MOFs and also identified approximately 80% of the actual low-performing MOFs. On the other hand, the simple geometric classifier captured less than 80% of actual high-performing MOFs. Such difference in performance could be attributed to two facts: (1) there are far more descriptors used with AP-RDF and Voronoi hologram classifiers than the simple geometric classifiers, and more importantly, (2) the VH and AP-RDF descriptors capture more detailed information on the MOF structure which affects their performance. Indeed, the pore shape of the MOF would indicate how much of the accessible void space of MOFs could be useable for gas adsorption. Thus, VH and AP-RDF descriptors encoding the pore shape of the MOF, albeit in different way, show better performance than the simple geometric classifiers. However, the inclusion of “atomic properties” in RDF does not increase the performance as much as seen in previous work.²⁷ This is likely due to the high-pressure adsorption condition used for methane purification in this study. At high pressure, the adsorption may depend more on the geometry of the pores, whereas at very low pressure the adsorption depends more on the chemistry of the pores. Regarding CO₂ working capacity, we note that all three descriptor types perform about the same.

Table 8.2: G-mean, sensitivity, and specificity values for all of the SVM classifiers for CO₂/CH₄ selectivity at 10 bars, and CO₂ working capacity. Adsorption and desorption pressures are 10 and 1 bars, respectively, and the temperature is 298 K.

property	descriptor	G-mean	sensitivity	Specificity
CO ₂ /CH ₄ selectivity	simple geometric	0.78	0.78	0.78

	GA-derived VH	0.87	0.92	0.82
	AP-RDF	0.85	0.91	0.80
CO ₂ working capacity	simple geometric	0.83	0.90	0.77
	GA-derived VH	0.85	0.92	0.78
	AP-RDF	0.84	0.91	0.79

In addition to the G-mean score evaluation, we used the area under the curve (AUC) of the receiver operating characteristic (ROC) plot as a metric to evaluate the performance of the classifiers on the training set. The AUC measures the probability that the classifier will correctly rank one MOF higher than another, with the ideal AUC being unity. The ROC plots for all of the classifiers were constructed and their corresponding AUCs are reported in Table 8.3. As seen with the G-mean score, the GA-derived VH and AP-RDF classifiers perform better than the simple geometric classifier on the basis of AUC values. For CO₂/CH₄ selectivity, the GA-derived VH and AP-RDF classifiers exhibited AUC values of ~0.94, while the simple geometrical classifier yielded a lower AUC of 0.89. Similarly, in the case of CO₂ working capacity, the GA-derived VH and AP-RDF classifiers have AUC values higher than that of the simple geometric classifiers.

Table 8.3: Area under the curve (AUC) values of the receiver operator curve (ROC) plots of the recognition of high-performing MOFs for CO₂/CH₄ separations.

property	Descriptor	AUC
CO ₂ /CH ₄ selectivity	simple geometric	0.889
	GA-derived VH	0.948
	AP-RDF	0.936
CO ₂ working capacity	simple geometric	0.919
	GA-derived VH	0.934
	AP-RDF	0.931

8.3.3 Predictive Capability of the Classifiers

Although Tables 8.2 & 8.3 show that including AP-RDF and GA-derived VH into classifiers improves the classification of MOFs in the training set, their performance still needs to be validated on MOFs which were not used in calibrating the classifiers. For this purpose, we

applied the classifiers on the test set of 292,050 MOFs in our database that were not used to train the classifiers, and we compared the sensitivity and specificity values of the classifiers' prediction at the cut-off values that yield the highest G-mean score values. In other words, we evaluated how well the classifiers recovered as many of the actual high-performing MOFs as possible (sensitivity) while discarding as many of actual low-performing MOFs as possible. Figure 8.10 shows the sensitivity and specificity values of all classifiers at their corresponding maximum G-mean score values. The grey bars in Figure 8.10 correspond to the ratio of MOFs flagged as high-performing, which include actual high-performing MOFs recovered by the classifiers and actual low-performing MOFs incorrectly recovered by the classifiers, to the total number of MOFs in the database.

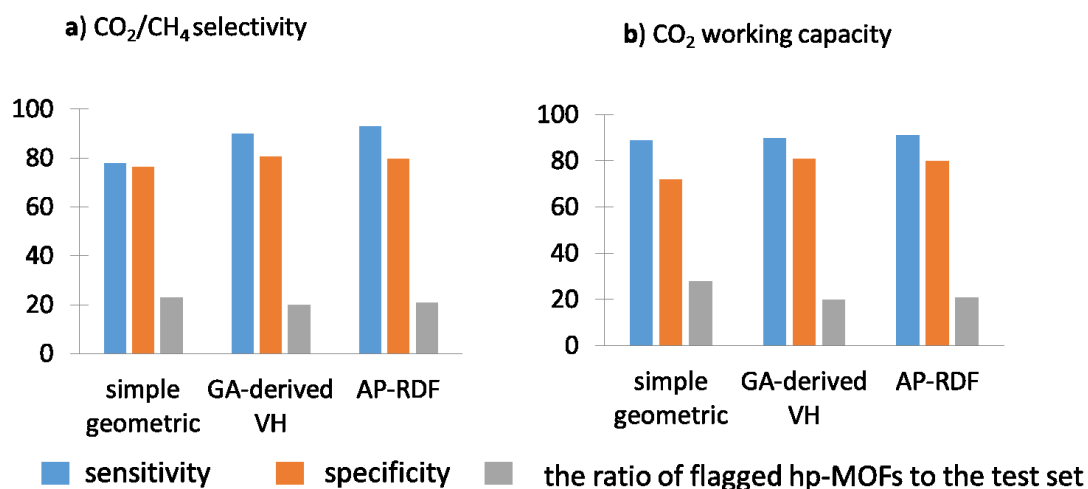


Figure 8.10: Sensitivity (blue bar) and specificity (red bar) values at maximum G-mean score of the all classifiers for a) CO₂/CH₄ selectivity and b) CO₂ working capacity. The gray bars indicate the ratio of MOFs flagged as high-performing by the classifier to the total number of MOFs in the database.

As it can be observed by Figure 8.10, in general, the specificity and sensitivity of the test set predictions are in good agreement with the cross-validation accuracies in Table 8.2. This indicates that the classifiers learned the structural information from the calibration set, and

more importantly, it could be applied to the test set with comparable predictive capability, suggesting the models were not over fit. Consequently, both the AP-RDF and GA-derived VH classifiers are better than the simple geometric classifiers. For CO_2/CH_4 selectivity, the AP-RDF classifier recovered 93% of all the high-performing MOFs by discarding $\sim 80\%$ of all the low-performing MOFs, while the simple geometric classifier captured only 75% of all the high performing MOFs and discarded $\sim 76\%$ of the low-performing MOFs. Similar to the AP-RDF classifier, the GA derived VH classifier recovered 90% of the actual high-performing MOFs while discarding 80% of the low-performing MOFs. These results, along with the gray bars in Figure 8.10 show that using AP-RDF classifiers, one would only need to perform GCMC simulations on 21% of the database in order to recover 93% of MOFs with the CO_2/CH_4 selectivity above 10. Regarding CO_2 working capacity, although all the classifiers possess almost the same value of the sensitivity (90%), their specificity values are different. The GA-derived VH and AP-RDF classifiers have a specificity value of 82% (i.e., discarded 82% of the low performing MOFs), while the simple geometric classifier prediction yielded a specificity value of 72% (i.e., discarded only 72% of the actual low-performing MOFs). This, along with the gray bars in Figure 8.6, indicates that using the AP-RDF and GA-derived classifiers, less computation time will be ‘wasted’ on incorrectly flagged high-performing MOFs compared to the simple geometric classifier. As seen with the CO_2/CH_4 selectivity, one can recover 90% of all actual high-performing MOFs while screening only 20% of the total database.

In the case of CO_2/CH_4 selectivity, while VH and AP-RDF descriptors retrieve $\sim 90\%$ of the high-performing MOFs, simple geometric classifier can still recover $\sim 75\%$ of the high-performing MOFs. This brings up the question of what specific SBUs or functional groups are

captured more effectively by the VH or AP-RDF classifiers than the simple geometric classifiers. To investigate this, we examined the frequency of the SBU's and functional groups in the high performing MOFs retrieved by the classifiers. These statistics were measured for each category at the cut-off values that yield the highest G-mean score values. Figure 8.11 shows the sensitivity values divided into MOFs with different inorganic SBUs. As illustrated by this Figure, for MOFs with V_2O_2 and Cu_2 metal nodes, the different classifiers predict fairly different performances with regards to sensitivity values. More specifically, AP-RDF and VH classifiers have sensitivity values of 0.85 and 0.95 for V_2O_2 while the sensitivity values of simple geometric classifier for V_2O_2 is 0.66. This indicates that the AP-RDF and VH classifiers recover more high-performing MOFs with V_2O_2 compared to the simple geometric MOF. In other words, the improved performance is attributed to the MOFs with **sra** topology (i.e., vanadium-based MOF) in the database.

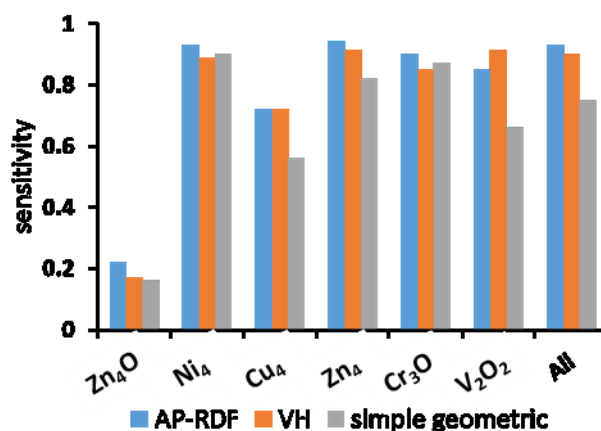


Figure 8.11: Sensitivity of the classifiers for recognizing MOFs with high performance for CO_2/CH_4 selectivity categorized according to inorganic SBUs.

Figure 8.12 depicts the effect of the different classifiers on the predicted sensitivity values of MOFs functionalized with different functional groups. As can be observed in this Figure, $-NO_2$, $-Cl$, and $-OH$ functional groups exhibited significantly higher sensitivities values for

the AP-RDF classifier compared to the simple geometric classifier. This is due to the fact that “atomic properties” in AP-RDF could take into account the interaction of these functional groups with CO₂ which has a high contribution to MOF selectivity.

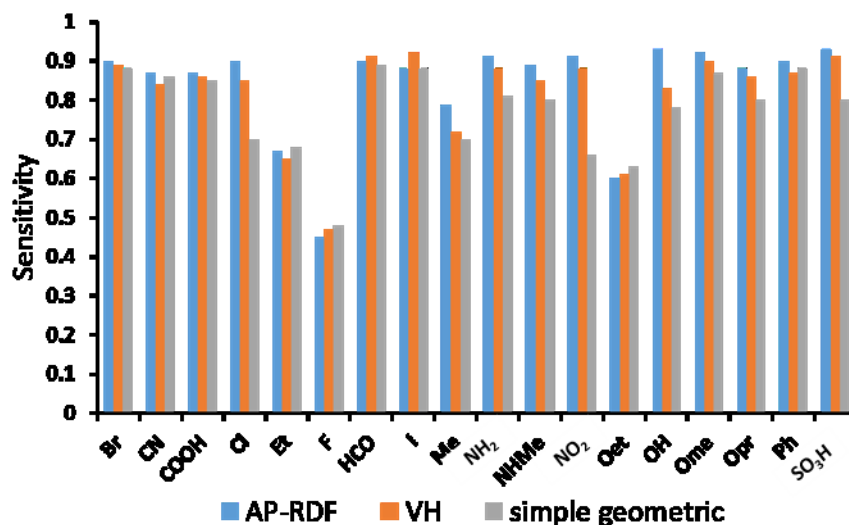


Figure 8.12: Sensitivity of the different classifiers for recognizing high-performing MOFs for CO₂/CH₄ selectivity categorized according to functional groups.

As previously mentioned, the evaluation is performed at the cut-off values that yielded the highest G-mean score for each classifier. However, other cut-offs can be chosen for our optimized classifiers that could return higher sensitivity or higher specificity depending on the screening scenario and the needs of the user. When the validation of the top candidates is inexpensive, one could choose a lower cut-off value (i.e., high sensitivity value) to capture nearly all the potential high-performing MOFs even though the number of MOFs incorrectly flagged as high-performing would increase. On the other hand, for situations where the higher level calculations are very compute intensive, one may prefer to choose a higher cut-off value (i.e., low sensitivity value) such that only a small number of high performers are recovered even though a considerable fraction of good performing MOFs would be categorized as low . As such,

it is important to investigate how many of top high-performing MOFs a classifier is able to capture under these two scenarios. For this, we have examined the number of top 1000 MOFs recovered by the AP-RDF and GA-derived VH classifiers where the cut-off was adjusted such that the sensitivity values were 30%, 50% and 90%. These results are given in Table 8.4, where the number in parentheses gives the percentage of the total test set that the classifier identified as high-performing. With the very selective 30% sensitivity cut-off for CO₂/CH₄ selectivity, VH classifier is still able to recover 720 of the top 1000 MOFs from the 292,050 member test set. Meanwhile, AP-RDF classifier could retrieve 780 of the top 1000 MOFs. If the sensitivity cut-off is adjusted to be less selective, then as expected, the number of the top 1000 MOFs increases up to 990 in case of VH classifier and to 993 in the case of AP-RDF for the 90% sensitivity cut-off. Interestingly, with the selective 50% cut-off, one would recover 889 of the highest performing materials for CO₂/CH₄ selectivity while performing expensive simulations only on 12% of the total database. Following a similar trend, with 50% sensitivity cut-off for CO₂ working capacity, classifiers retrieved 880 of the top-1000 high-performing MOFs while only flagging 15% of the database as high-performing set.

Table 8.4: Number of the top 1000 MOFs from the test set that the classifiers of working capacity and selectivity correctly identifies with different cut-offs. The percentage in parenthesis is the fraction of the test set that would require GCMC simulations

Parameter	Descriptor	Sensitivity cut-off		
		30%	50%	90%
Selectivity	Voronoi hologram	720 (9%)	889 (12%)	990 (21%)
	AP-RDF	780 (10%)	898 (13%)	993 (20%)
Working capacity	Voronoi hologram	405 (11%)	880 (15%)	992(20%)
	AP-RDF	422 (11%)	886 (17%)	994(21%)

8.4 Conclusion

As a prescreening tool, ML classifiers for CO₂/CH₄ separation parameters under methane purification conditions have been developed using Voronoi hologram and AP-RDF descriptors. Using simulated gas adsorption data from GCMC simulations on ~32 000 MOFs, we have constructed classifiers based on simple geometric descriptors, AP-RDF descriptors and Voronoi hologram descriptors. To validate these models, they were applied to a test set of ~300,000 MOFs for which were not part of the training set. It was found that both AP-RDF and Voronoi hologram classifiers correctly retrieved about 90% of high-performing MOFs and correctly discarded about 80% of low-performing MOFs, while simple geometric classifier only retrieved 75% of all the high performing MOFs and discarded ~76% of the low-performing MOFs. It was also demonstrated on the test set that the classifiers could recover ~900 of the top 1000 MOFs in the database, while only flagging about 13-17 % of the total database for more compute intensive screening. Since the classifiers can be applied to a material on the orders of milliseconds rather than tens of minutes to calculate the adsorption properties with GCMC simulations, applying the classifiers would result in roughly a 6-7 times faster high throughput screening of materials.

8.5 References

- (1) McLaughlin, J. F.; Hellmann, J. J.; Boggs, C. L.; Ehrlich, P. R. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, 99 (9), 6070.
- (2) Solomon, S.; Plattner, G.-K.; Knutti, R.; Friedlingstein, P. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, 106 (6), 1704.
- (3) Liu, Y.; Ye, Q.; Shen, M.; Shi, J.; Chen, J.; Pan, H.; Shi, Y. *Carbon dioxide capture by functionalized solid amine sorbents with simulated flue gas conditions*; 2011; Vol. 45.
- (4) Outlook, A. E.; Protocol, K.; Markets, E.; Activity, E.; States, U. *Energy Inf. Adm. Nat. Gas* 1998

Issues Trends **2012**, 49.

- (5) Rochelle, G. T. *Science* **2009**, 325 (5948), 1652.
- (6) Kumar, R. *Ind. Eng. Chem. Res.* **1994**, 33, 1600.
- (7) Sumida, K.; Rogow, D. L.; Mason, J. a; McDonald, T. M.; Bloch, E. D.; Herm, Z. R.; Bae, T.-H.; Long, J. R. *Chem. Rev.* **2012**, 112 (2), 724.
- (8) Li, J.; Sculley, J.; Zhou, H. **2012**, 869.
- (9) Boyd, P. G.; Woo, T. K. *CrystEngComm* **2016**.
- (10) Keskin, S.; Sholl, D. S. *J. Phys. Chem. C* **2007**, 111, 14055.
- (11) Wilmer, C. E.; Leaf, M.; Lee, C. Y.; Farha, O. K.; Hauser, B. G.; Hupp, J. T.; Snurr, R. Q. *Nat. Chem.* **2012**, 4 (2), 83.
- (12) Wilmer, C. E.; Farha, O. K.; Bae, Y.-S.; Hupp, J. T.; Snurr, R. Q. *Energy Environ. Sci.* **2012**, 5 (12), 9849.
- (13) 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 (7), 633.
- (14) Kadantsev, E. S.; Boyd, P. G.; Daff, T. D.; Woo, T. K. *J. Phys. Chem. Lett.* **2013**, 4 (18), 3056.
- (15) Wilmer, C. E.; Kim, K. C.; Snurr, R. Q. *J. Phys. Chem. Lett.* **2012**, 3, 2506.
- (16) Fernandez, M.; Boyd, P. G.; Da, T. D.; Aghaji, M. Z.; Woo, T. K. *J. Phys. Chem. Lett.* **2014**.
- (17) Aghaji, M. Z.; Fernandez, M.; Boyd, P. G.; Daff, T. D.; Woo, T. K. *Eur. J. Inorg. Chem.* **2016**, 1.
- (18) Martin, R. L.; Willems, T. F.; Lin, L.-C.; Kim, J.; Swisher, J. a; Smit, B.; Haranczyk, M. *Chemphyschem* **2012**, 13 (16), 3595.
- (19) Martin, R. L.; Smit, B.; Haranczyk, M. *J. Chem. Inf. Model.* **2012**, 52 (2), 308.
- (20) Gale, J. J. *J. Chem. Soc., Faraday Trans.* **1997**, 93 (4), 629.
- (21) Rappe, a. K.; Casewit, C. J.; Colwell, K. S.; Goddard III, W. a.; Skiff, W. M. *J. Am. Chem. Soc.* **1992**, 114, 10024.
- (22) 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 (3), 8814.
- (23) Martin, M. G.; Siepmann, J. I. *J. Phys. Chem. B* **1998**, 102 (97), 2569.
- (24) Frost, H.; Düren, T.; Snurr, R. Q. *J. Phys. Chem. B* **2006**, 110 (19), 9565.
- (25) Duren, T.; Millange, F.; Ferey, G.; Walton, K. S.; Snurr, R. Q. *J. Phys. Chem. C* **2007**, 111 (42),

15350.

- (26) Willems, T. F.; Rycroft, C. H.; Kazi, M.; Meza, J. C.; Haranczyk, M. *Microporous Mesoporous Mater.* **2012**, *149* (1), 134.
- (27) Fernandez, M.; Trefiak, N.; Woo, T. *J. Phys. ...* **2013**, 130617161909001.
- (28) franz aurenhamer. *ACM Comput. Surv.* **1991**, *23* (3).
- (29) Hooper, J.; Hu, A.; Zhang, F.; Woo, T. K. *Phys. Rev. B* **2009**, *80*, 20.
- (30) Collins, S. P.; Daff, T. D.; Piotrkowski, S. S.; Woo, T. K. *Sci. Adv.* **2016**, *2*, 1.
- (31) Magalhães-mendes, J. *Wseas Trans. Comput.* **2013**, *12* (4), 164.
- (32) Chakraborty, B.; Chaudhuri, P. *Austrian J. Stat.* **2003**, *32* (1), 13.
- (33) S.N. Sivanandam and S. N. Deepa. *Introduction to Genetic Algorithms*; Springer: New York, 2008.
- (34) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, E. *J. Mach. Learn. Res.* **2011**, *12*, 2825.
- (35) Cavenati, S.; Grande, C. a; Rodrigues, a E. *J. Chem. Eng. Data* **2004**, *49*, 1095.

9 MOFFinder: A Novel Web-based Tool for Aiding in the Discovery of New Metal Organic Framework Materials

MOFFinder is publicly available at <http://titan.chem.uottawa.ca/moffinger>. The data collection, the development and implementation of MOFFinder were performed by the author.

Abstract

MOFFinder is a searchable on-line database of hypothetical metal organic framework (MOF) materials that currently contains each material's simulated natural gas purification and CO₂ adsorption properties. It has a user-friendly graphical interface to promote easy exploration of over 300,000 hypothetical MOFs. It enables synthetic chemists to find MOFs of interest by searching the database for Secondary Building Units (SBUs), geometric features, functional groups and adsorption properties. It also provides substructure, similarity searches of the organic SBUs to allow chemists to search for potential synthetic targets based on their desired application.

9.1 Introduction

Due to tremendous growth in the amount of data and availability of open-source software components, many biological and chemical databases have been developed or refined, such as the Protein Data Bank (PDB),¹ the NMRShiftDB,² and the ChemSpider.³ A number of integrated databases, such as UniChem,⁴ ZINC,⁵ eMolecules,⁶ and PubChem,⁷ which provide access to an enormous number of structures from different databases have also emerged. However, most of these databases contain information on biological macromolecular structures and small molecules, and up to now, far less attention has been paid to specifically designing a database for porous materials, particularly MOFs. Of the databases containing

information on porous materials, the most widely used is the Cambridge Crystallographic Data Centre (CCDC)⁸ which contains the crystallographic structures of 22,700 synthesized MOFs. However, physical properties associated with these structures, including the adsorption properties and the geometrical features, have not been deposited so far. Therefore, one cannot query CCDC based on the adsorption properties of MOFs. Also, from the design point of view, its substructure and similarity searches are unusable for extended structures. Moreover, 22,700 structures deposited in CCDC are only a tiny fraction of possible MOFs structures that can be assembled.⁹

Recent efforts in MOF research have resulted in the computational construction of hundreds of thousands of hypothetical MOF structures which have undergone high-throughput screening studies within gas-related applications. Snurr et al.^{10,11} computationally screened 135,000 hypothetical MOFs for a variety of applications, including methane and hydrogen storage, CO₂ capture, and Xe/Kr separation. Similarly, Woo et al.¹² computationally screened a massive library of over 300,000 hypothetical MOFs for CO₂ capture. One aim of these studies was to identify materials that perform better than existing materials for the given application. Another aim was to investigate the structure-property relationships-in other words to relate the MOF's structure to its performance for a desired application, such as CO₂ separation. Regardless of whether these aims were achieved or not, a large amount of data is generated in these studies (e.g., one million data points in the study of screening of 135,000 hypothetical MOFs for methane storage by Wilmer et al.¹¹), from which many researchers can undoubtedly benefit. With such a large amount of data, Snurr et al.¹³ specifically designed a searchable database which contains information on 135,000 hypothetical MOFs. More specifically, it

includes the methane uptake at 35 bar, the crystal structure, and the geometric features including surface area, pore diameter, void volume, and density. Although its search engine enables users to query the database based on user-defined CH₄ uptake or geometric features, it does not provide substructure and similarity searches which are available in many biological and molecular databases, and could give benefits to MOF synthesizers. Moreover, it is populated only with the methane uptake data and does not include data relevant to CO₂ capture.

Here, we have developed a search engine database of 300,000 hypothetical MOFs, called MOFFinder, which includes the data relevant to CO₂ capture and natural gas purification, and provides the capability to perform substructure/similarity searches, in addition to standard property search for users. In what follows are the details of the MOFFinder database, how it works and how it was constructed. For the sake of simplicity, the discussion starts with the functionality of MOFFinder, followed by an introduction to its user interface, and ends with an explanation of how MOFFinder was developed and implemented.

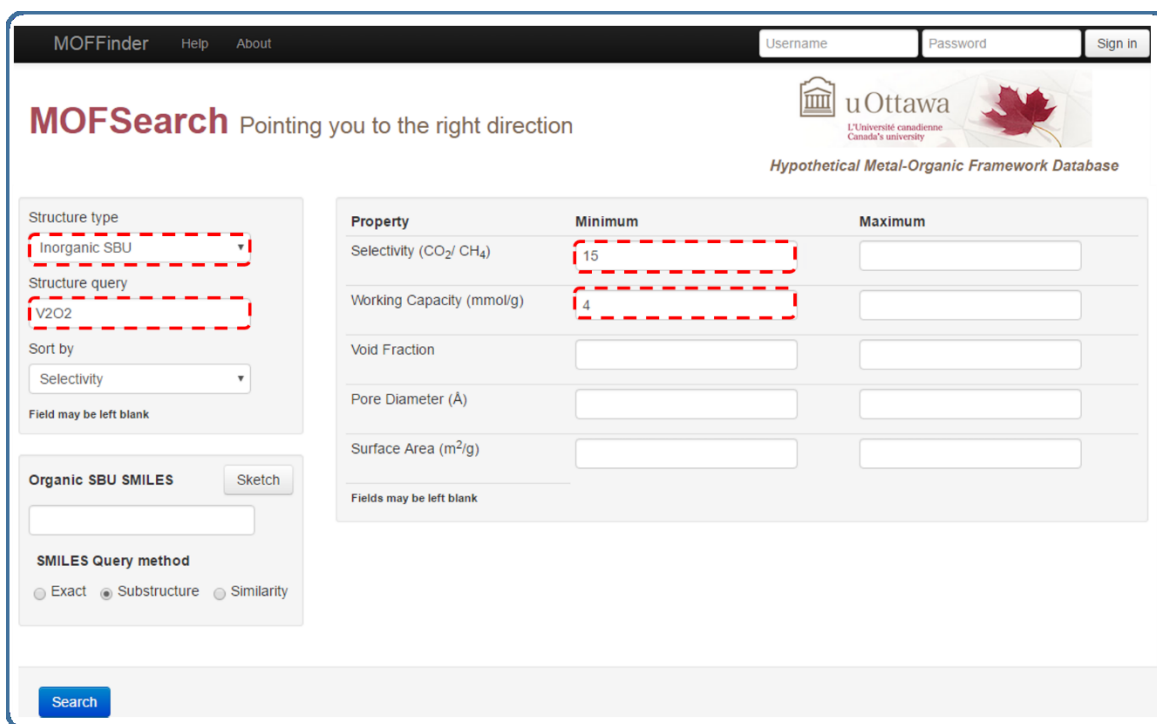
9.2 Functionality

MOFFinder provides access to over ~300,000 hypothetical metal organic frameworks, and their predicted adsorption properties relevant to both natural gas purification and post-combustion CO₂ capture. It is specifically designed to allow users to search the database of hypothetical metal organic frameworks database, and currently provides search results in under a second. Indeed, MOFFinder can help synthetic chemists to find more synthesis targets, such as MOFs with high surface area, hydrophobic MOFs including non-polar functional groups,

etc. The following two examples illustrate how MOFFinder can help synthetic chemists to find synthesis targets for their desired application.

Example 1: The query for vanadium-based MOFs to use in the NG application

The solid sorbent Zeolite 13X is commercially used in natural gas purification and has a CO_2/CH_4 selectivity of 13 and a CO_2 working capacity of 1.8 mmol/g.¹⁴ In order to find the promising MOFs to potentially replace Zeolite 13X, one use of MOFFinder would be to perform the following search for a V containing MOF. In the MOFFinder web-interface (Figure 9.1), first the 'Inorganic SBU' is selected for the 'Structure type' search and 'V2O2' is then written in the 'Structure query' field. Afterwards the 'minimum' option for selectivity (CO_2/CH_4) and working capacity (mmol/g) are set 15 and 4, respectively, in order to outperform of Zeolite 13X.



The screenshot shows the MOFFinder web interface. At the top, there is a navigation bar with 'MOFFinder', 'Help', and 'About' links, along with fields for 'Username', 'Password', and a 'Sign in' button. Below this, the main header features the 'MOFSearch' logo with the tagline 'Pointing you to the right direction', the uOttawa logo, and the text 'Hypothetical Metal-Organic Framework Database'.

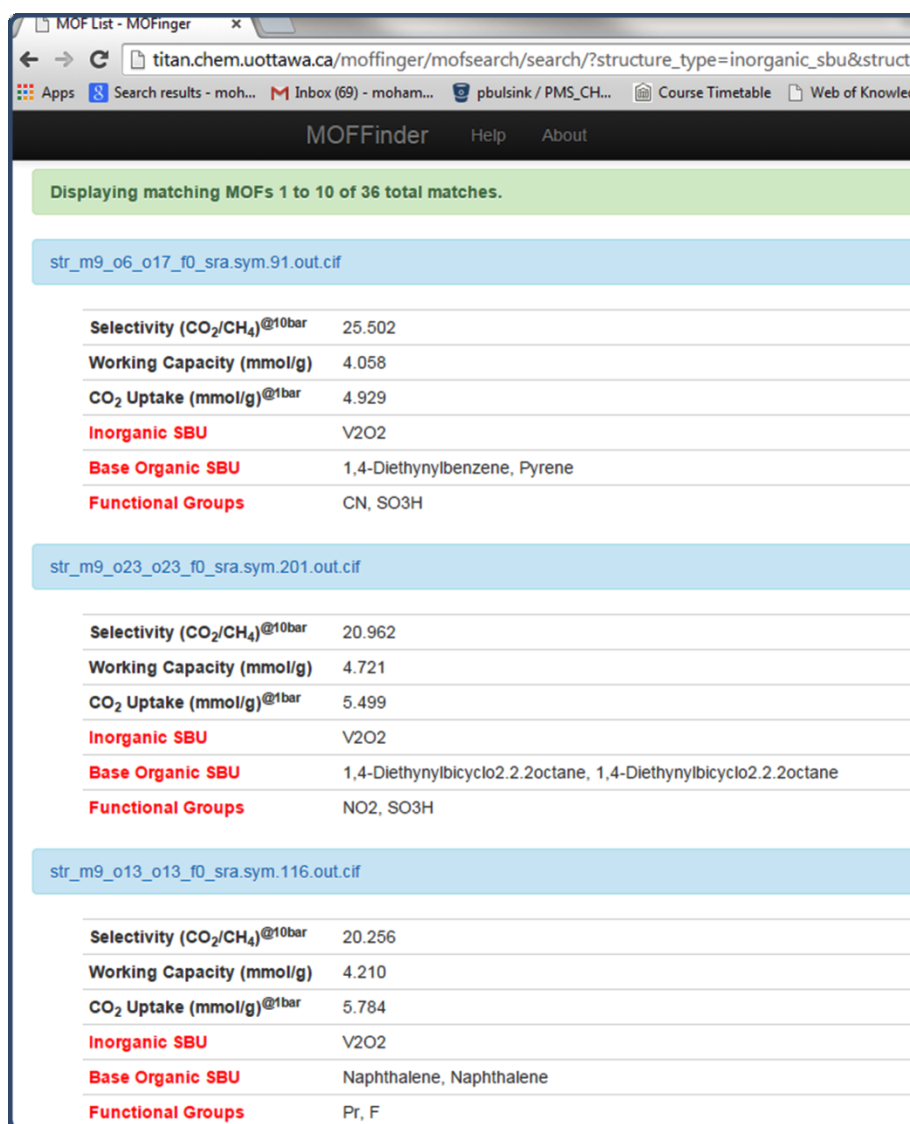
The search interface is divided into several sections:

- Structure type:** A dropdown menu with 'Inorganic SBU' selected.
- Structure query:** A text input field containing 'V2O2'.
- Sort by:** A dropdown menu with 'Selectivity' selected.
- Field may be left blank:** A note below the sort by dropdown.
- Organic SBU SMILES:** A section with a 'Sketch' button and a text input field.
- SMILES Query method:** Radio buttons for 'Exact', 'Substructure' (selected), and 'Similarity'.
- Property table:** A table with columns 'Property', 'Minimum', and 'Maximum'.

Property	Minimum	Maximum
Selectivity (CO_2/CH_4)	15	
Working Capacity (mmol/g)	4	
Void Fraction		
Pore Diameter (Å)		
Surface Area (m^2/g)		
- Fields may be left blank:** A note below the property table.
- Search:** A blue button at the bottom left.

Figure 9.1: The query for vanadium-based MOFs with the CO_2/CH_4 selectivity and CO_2 working capacity above 15 and 4 mmol/g, respectively.

With this query, MOFFinder filters MOFs with the V_2O_2 SBU. Then, selected MOFs are filtered for selectivity higher than 15 and working capacity higher than 4 mmol/g. At the end, the search hits will be displayed in a paged and tabular manner. The results page for above query is shown in Figure 9.2. The numbers of hits (36) are highlighted on the green panel of the results page.



MOF List - MOFFinder

titan.chem.uottawa.ca/moffinder/mofsearch/search/?structure_type=inorganic_sbu&structu

MOFFinder Help About

Displaying matching MOFs 1 to 10 of 36 total matches.

str_m9_o6_o17_f0_sra.sym.91.out.cif

Selectivity (CO_2/CH_4)@10bar	25.502
Working Capacity (mmol/g)	4.058
CO_2 Uptake (mmol/g)@1bar	4.929
Inorganic SBU	V_2O_2
Base Organic SBU	1,4-Diethynylbenzene, Pyrene
Functional Groups	CN, SO_3H

str_m9_o23_o23_f0_sra.sym.201.out.cif

Selectivity (CO_2/CH_4)@10bar	20.962
Working Capacity (mmol/g)	4.721
CO_2 Uptake (mmol/g)@1bar	5.499
Inorganic SBU	V_2O_2
Base Organic SBU	1,4-Diethynylbicyclo2.2.2octane, 1,4-Diethynylbicyclo2.2.2octane
Functional Groups	NO_2 , SO_3H

str_m9_o13_o13_f0_sra.sym.116.out.cif

Selectivity (CO_2/CH_4)@10bar	20.256
Working Capacity (mmol/g)	4.210
CO_2 Uptake (mmol/g)@1bar	5.784
Inorganic SBU	V_2O_2
Base Organic SBU	Naphthalene, Naphthalene
Functional Groups	Pr, F

Figure 9.2: A resulting first page for the query of vanadium based-MOFs with selectivity and working capacity higher than those of Zeolite 13X.

For each hit, the selectivity, the working capacity and the CO₂ uptake, the building blocks name and the chemical formula of functional groups are given below the hit name. Each hit name on the blue panel is a clickable link, which takes the user to a single structure view page that interactively displays the 3D structure of the hit using JSmol. As an example, the view page of the third hit (str_m9_013_013_f0_sra.sym.116.cif) of above query is given in Figure 9.3. The 2x2x2 unit cell of this hit is shown as a ball and stick model on the top of the view page with different colors representing the different atom types. The user can change the view style by using the JSmol popup menu.

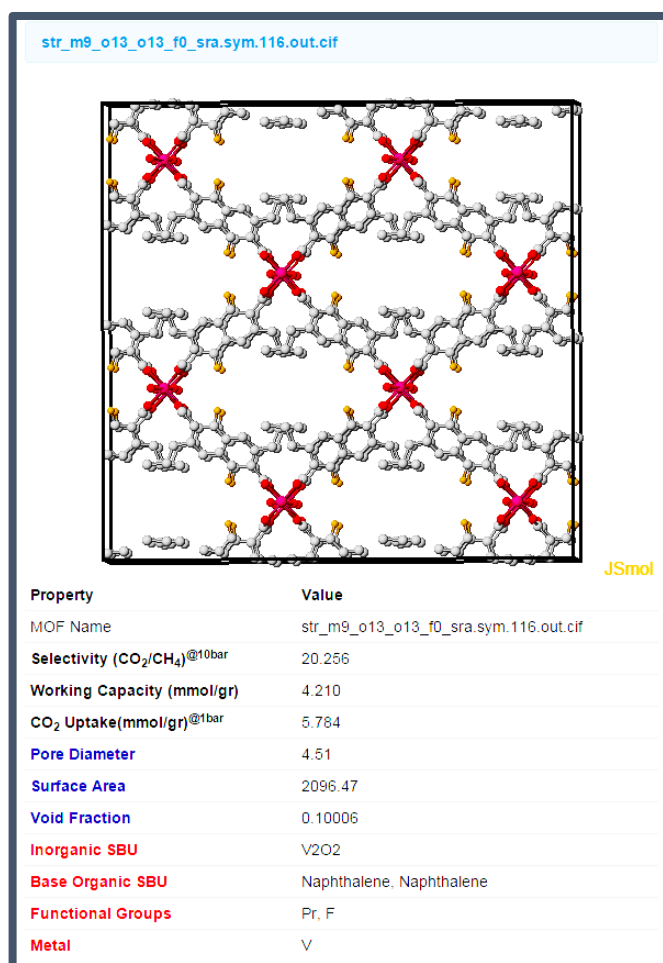
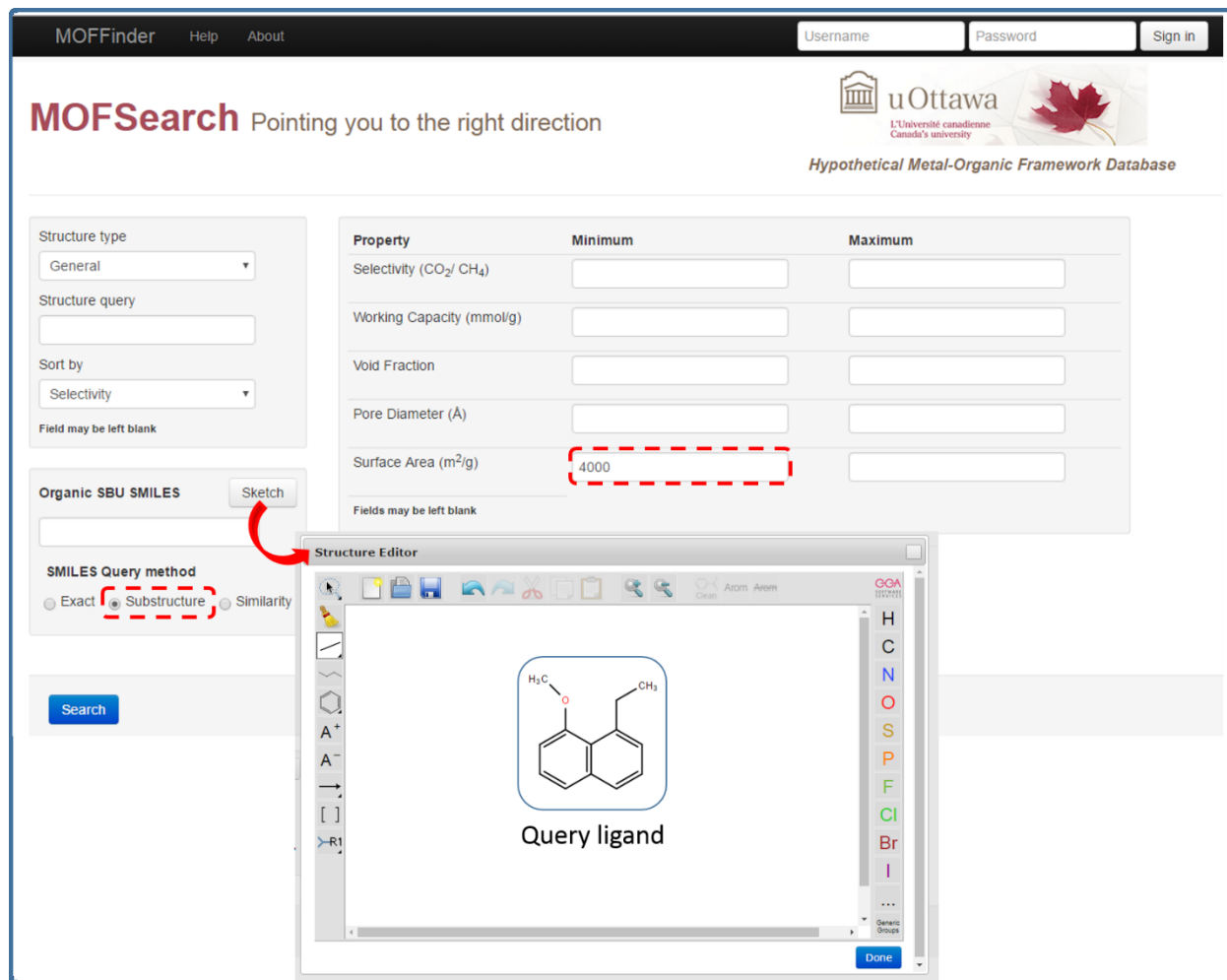


Figure 9.3: An example of a view page of a hit resulting from structure type search

Example 2: The query for hydrophobic MOFs along with medium surface area

In various industrial streams, gas mixtures usually contain a trace of water vapor, limiting the application of MOFs in gas storage and separation processes due to ligand-exchange or pore filling.^{15,16} Introducing hydrophobicity into MOFs has been proposed as a short-term strategy to diminish the water vapor effect. One possible way of addressing this challenge is to graft hydrophobic functional groups such as alkyl chains to the MOFs.¹⁷ In this direction, to demonstrate the capability of MOFFinder to search for MOFs with hydrophobic character and desired geometric properties, we provide one example of the query for MOFs which have surface area of at least $4000 \text{ m}^2\text{g}^{-1}$ and contain naphthalene with CH_3CH_2 and OCH_3 functional groups as part of their ligands. On the MOFFinder web-interface, the query ligand (i.e., Naphthalene with CH_3CH_2 and OCH_3 functional groups) was drawn and the minimum value of surface area was set to 4000 (Figure 9.4). By selecting 'Substructure' on the 'SMILES Query Method' field and clicking the 'Search' button, the MOFFinder first performs a substructure search for the query ligand over the $\sim 200,000$ ligands. Then, the filtered ligands are linked to their related MOFs, and subsequently the resulting MOFs are filtered for surface area higher than $4000 \text{ m}^2\text{g}^{-1}$. At the end, the hits will be displayed in the same manner as in the first example on the result page. The 3D view of two examples of hits is shown in Figure 9.5.



MOFFinder Help About Username Password Sign in

MOFSearch Pointing you to the right direction

Structure type
General

Structure query
Field may be left blank

Sort by
Selectivity

Organic SBU SMILES Sketch

SMILES Query method
Exact **Substructure** Similarity

Search

Property	Minimum	Maximum
Selectivity (CO ₂ / CH ₄)		
Working Capacity (mmol/g)		
Void Fraction		
Pore Diameter (Å)		
Surface Area (m ² /g)	4000	

Fields may be left blank

Structure Editor

Query ligand

Chemical structure of a naphthalene derivative with CH₃CH₂ and OCH₃ functional groups.

Figure 9.4: The query for MOFs which have surface area of at least 4000 m²g⁻¹ and contain naphthalene with CH₃CH₂ and OCH₃ functional groups as part of their ligands

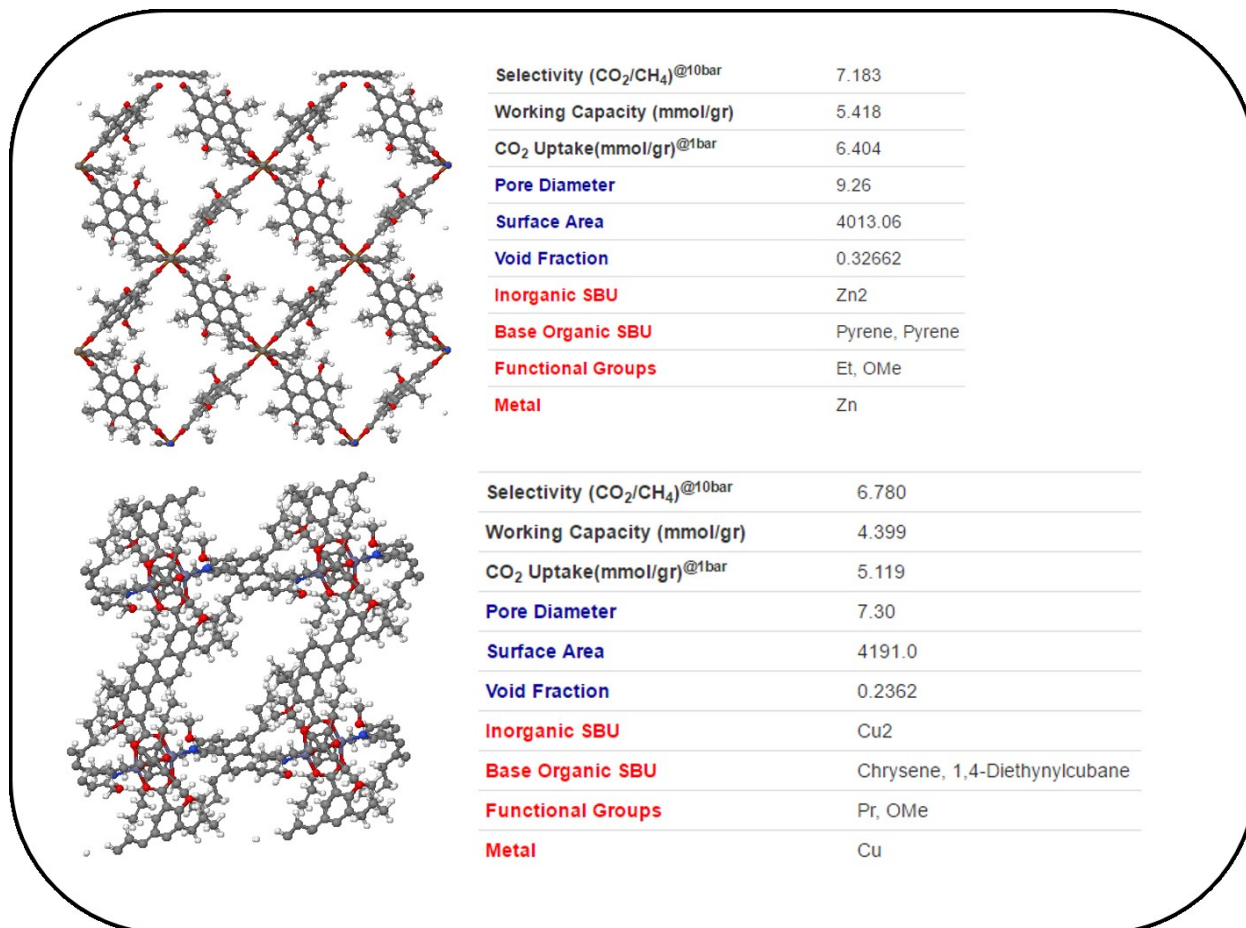


Figure 9.5: Examples of two hits resulting from the query for MOFs with a hydrophobic nature and a surface area higher than 4000 m²/g. Both hits have CO₂ uptakes higher than Zeolite 13X (6.4 and 5.1 vs 4.2 mmol/g)

9.3 Graphical User Interface

In the MOFFinder homepage, a session can be launched by users via either clicking the 'Try it' button or 'MOFSearch' option. The interface shown in Figure 9.6 is designed to allow the user to query the database with a variety of options. It includes the 'Structure type' search (Figure 9.7a), the 'Organic SBU SMILES' search (Figure 9.7b), and a search by properties where one can set the min/max of the property. Additionally, MOFFinder provides a chemical structure editor via the 'Sketch' button which allows users to draw structures in order to

perform substructure searches. The main parts of these features are described below. All information needed to use the interface has been provided in 'MOFFinder Help' option.

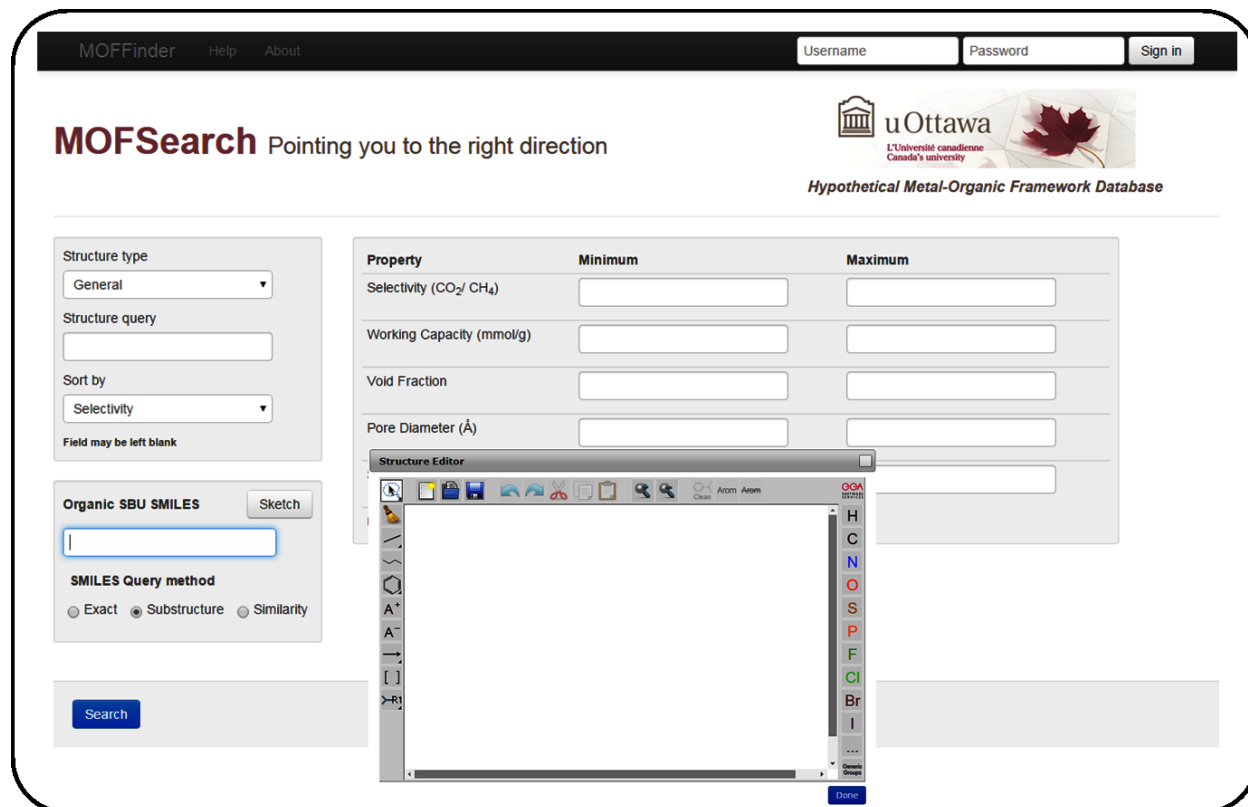


Figure 9.6: The interface in MOFFinder. The 'Structure type' search is in the upper left including five options namely Functional Groups, Organic SBU, Inorganic SBU, Metal and General. The 'Structure query' field is below the 'Structure type' to write a query as text. The 'Sort by' option can be used to order the search results by sorting the numerical properties based on the user's request. The minimum and maximum options of adsorption and geometric properties are on the upper right. The 'Organic SBU SMILES' search which is on the lower left is responsible for similarity, substructure and exact search. The 'Sketch' button invokes a pop-up the chemical structure editor.

Two types of searches are available on the web interface, namely 'Structure type' and 'Organic SBU SMILES'. The 'Structure type' search includes five different options: Functional Groups, Organic-SBUs, Inorganic-SBUs, Metal Center, and General, which provides text-based searches. Here, a user has to select one option from the 'Structure type' options and write related text to the selected option in the 'Structure query' field and click the 'Search' button on

the interface. The ‘Organic SBU SMILES’ search allows the user to perform substructure searches by directly inputting the structure in SMILES format, or by drawing the structure using the pop-up structure editor. Three search methods are allowed - ‘Exact’, ‘Substructure’, and ‘Similarity’. All these searches are achieved by comparing ligands’ fingerprints which include a series of binary digits (bits) that represent the presence or absence of particular fragment in ligand. In the case of similarity, this similarity is quantified using the following metrics supported by the Bingo cartridge:

$$\text{Tanimoto:} \quad \frac{c}{a + b} \quad 9.1$$

$$\text{Tversky:} \quad \frac{c}{(0.9 \times (a - c) + 0.1 \times (b - c) + c)} \quad 9.2$$

$$\text{Euclidean:} \quad \frac{c}{a} \quad 9.3$$

where a is the count of bits in the fingerprint of the query, b is the count of bits in fingerprint of the target, and c is the count of coincident bits in the fingerprints. A simple filtered search is also possible where the user can set ‘Minimum’ and ‘Maximum’ limits for adsorption properties, including CO₂/CH₄ selectivity and CO₂ working capacity, and for geometric features, such as the void fraction, pore diameter and surface area. With these various searches, users can conduct different searches, such as two examples provided in section 9.2, to retrieve candidates which are most relevant to their queries from the MOFFinder database. A graphical representation of the ‘Structure type’ search and Organic SBU SMILES’ search is given in Figure 9.7.

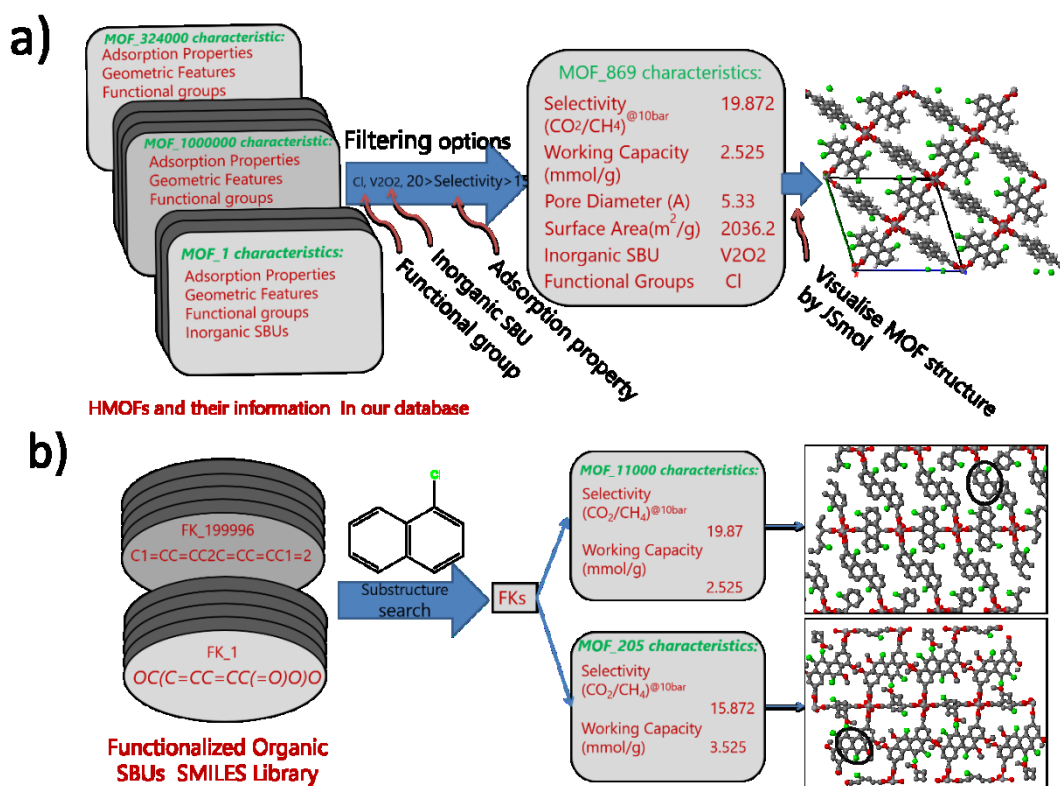


Figure 9.7: a) A graphical representation of the ‘Structure type’ search where MOFs can be filtered by adsorption properties, functional groups, and inorganic and organic SBUs. b) ‘Organic SBU SMILES’ search where functionalized organic SBUs SMILES can be filtered by ‘Exact’, ‘Substructure’, and ‘Similarity’ search, and then the resulting SMILES are related to their parent MOFs via FKs.

9.4 Data Collection

The 324,500 hypothetical MOF structures in our database were generated from 66 SBUs and 19 functional groups using a similar algorithm developed by a former Ph.D. student in the Woo lab, Peter Boyd. The Crystallographic Information Files (CIFs) of all hypothetical MOF structures were deposited into the MOFFinder database. GCMC simulations were performed for the adsorption of 1:9 CO₂-CH₄ mixture at 10 bar and 1 bar at 298 K, and consequently the CO₂/CH₄ selectivity and the CO₂ working capacity were computed and deposited into MOFFinder. Also, the simulation data of the pure CO₂ adsorption in all MOFs at 1 bar was

provided in the database. The geometric features of all MOFs including void fraction, surface area, and max pore diameter were calculated using the zeo++ code¹⁸ and provided in the database.

9.5 Implementation

MOFFinder was developed by integrating various open-source software components that are given in Table 9.1. These components are either supported by commercial organizations which offer services, or are well-known and have a large base of users like PostgreSQL. At the moment, these components are well maintained and it is expected that they can be continued to use as MOFFinder is further developed.

Table 9.1: MOFFinder components with their functionality

component	source	functionality
Django	https://www.djangoproject.com/	High-level python web framework
PostgreSQL	http://www.postgresql.org/	Provide storing data, and retrieving data at request of clients
Bingo PostgreSQL Cartridge	http://lifescience.opensource.epam.com/	Provide chemical structure searches within the PostgreSQL.
Ketcher	http://lifescience.opensource.epam.com/	Provide a platform to Draw structures of ligands
Jsmol	http://www.jmol.org/	Provide the visualization of the structures interactively in 3D format

The key part of searchable engines is the type of database used. The most commonly used databases are relational databases whose organization is based on the relational model of data defined by E. F. Codd in 1970.¹⁹ In this scheme, the data is organized into one or more tables (or "relations") of columns and rows, where each row contains a unique instance of

data for the categories defined by the columns, and each row has its own unique key for being linked to rows in other tables. In the context of database definition, rows are called records, and categories defined by the columns are referred to as fields. In terms of querying and maintaining database, all relational databases use SQL²⁰ (Structured Query Language) which is the standard language for relational database management systems. Examples of commercial relational databases include Oracle²¹, Sybase²², Informix²³, MySQL²⁴, and PostgreSQL²⁵. In order to provide efficient storage and searching solution for chemical information within these databases, several chemistry cartridges are developed as an extension for them. Examples of commercial cartridges include Bingo,²⁶ RDkit,²⁷ and Indigo cartridges²⁸. These cartridges integrate chemistry into relational databases, and enable scientists to store and search chemical moieties within one underlying relational database server. A relational database and its extension (e.g., RDkit), which make up the “backend” of cheminformatics framework, are tied together using sophisticated programming languages such as C++ and Python.

The underlying database of MOFFinder is PostgreSQL, which is available for different operating systems (Linux, Windows, OS X). To allow for chemical structure searches within PostgreSQL, the Bingo²⁶ PostgreSQL extension was implemented. The ‘backend’ of MOFFinder was tied together with custom Python code. The data structures of the underlying database within MOFFinder is summarized in Figure 9.8, and is organized into three main tables. One reason to divide the data into three tables goes back to the fact that some MOFs in our database are constructed using two different organic ligands. Thus, if the ligands of these MOFs are included in a single table of ~300,000 rows (where each row corresponds to one MOF in the database), the column corresponding to ligand information will have two items for

these MOFs. This will make substructure and similarity searches slow within the Bingo PostgreSQL cartridge, as its functionality is optimized to search over columns, each of which contains one item for each row of the table.²⁶ Another reason, which is more important, is the difference between the number of structures (i.e., ~300,000) and the number of unique organic ligands used to construct the structures (~200,000). This means that some MOFs in the database are composed of the same ligands. Thus, providing ligand information in a separate table (i.e., a table of ~200,000 rows where each row corresponds to one unique ligand) ensures that substructure searches is performed only once for a ligand used in the construction of different MOFs, reducing the number of substructure search compared to when the ligand information for each MOF (row) is provided in a table of ~300,000 rows. Because of this, structure properties and ligand information are provided in two separate tables, where these two tables are connected together by another table. In Figure 9.8, the tables' names and fields' names are shown in upper-case to describe the database architecture. The MOFSEARCH_MOF table holds all of the structures' names with their properties, where each structure in the database is assigned an identifier (s-id). The s-id is an integer which is stored in the ID field of MOFSEARCH_MOF table and acts as a primary key (PK) for this table. The simulated geometric properties of structures, including maximum pore diameter (Å), BET surface area (m²/g) and void fraction, are stored in the PORE_DIAMETER, SURFACE_AREA and VOID_FRACTION fields of MOFSEARCH_MOF table respectively. The data taken from GCMC simulations for each structure, including the amount of adsorbed CO₂ and CH₄, the CO₂ working capacity and CO₂/CH₄ selectivity are stored in the CO₂_UPTAKE, CH₄_UPTAKE, WORKING_CAPACITY and SELECTIVITY fields of the MOFSEARCH_MOF table, respectively. To represent the inorganic

secondary building unit (SBU) and its metal center of each structure, the chemical formula of the inorganic-SBU with the chemical symbol of the metal center are stored in the INORGANIC_SBU and METAL fields of the MOFSEARCH_MOF table, respectively. The base-organic SBUs with their related functional groups for each structure are stored in the ORGANIC_SBU and FUNTIONAL_GROUPS fields of the MOFSEARCH_MOF table using a systematic name for base-organic SBUs and chemical formula for functional groups. The unique functionalized organic SBUs used to generate the hypothetical structures are expressed by SMILES representations²⁹ and provided in the SMILES field of the MOFSEARCH_UNI_LIGAND_BINGO table. Each unique functionalized organic SBU is given an identifier which stores in the ID field of the MOFSEARCH_UNI_LIGAND_BINGO table and acts as a PK for this table. The assignment of the functionalized organic SBUs to their parent structures is provided by the MOFSEARHC_MOF_LIGAND table, where each row in MOFSEARCH_MOF_LIGAND table provides a link between a ligand and its related MOF via the LIGAND_FK and MOF_FK fields. The LIGAND_FK field is populated with integers which are foreign keys (FK) to the MOFSEARCH_UNI_LIGAND_BINGO table while integers in the MOF_FK field are foreign keys to the MOFSEARCH_MOF table.

The 'front-end' of MOFFinder, is constructed from a combination of HTML, Bootstrap, and JavaScript utilizing the jQuery library. The open-source JavaScript-based molecular viewer from Jmol (JSmol, <http://www.jmol.org>) is used to display the structures interactively in 3D format. To draw structures, for substructure searches, the JavaScript-based chemical structure editor, Ketcher,³⁰ was integrated. Note that users of MOFFinder do not need Java or Flash plugins in their browser to use both JSmol and Ketcher.

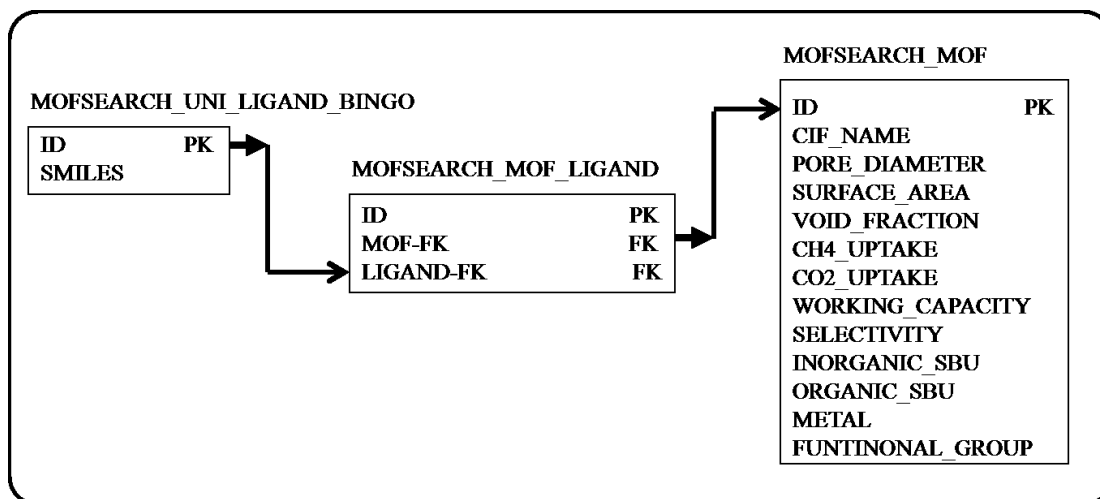


Figure 9.8: The database layer of MOFFinder consists of three main tables. All of the structure names with their geometric features, adsorption properties and structural representations of SBUs are stored in the MOFSEARCH_MOF table using their related fields. The SMILES representations of unique ligands are stored in the MOFSEARCH_UNI_LIGAND_BINGO table. The MOFSEARCH_MOF_LIGAND table provides the relation between MOFSEARCH_MOF and MOFSEARCH_UNI_LIGAND_BINGO tables. Primary/foreign key constraints are shown with solid arrows. PK=Primary Key, FK= Foreign Key.

It is noteworthy that any searches in MOFFinder are done in seconds. This is because any type of search in MOFFinder is done in its database (PostgreSQL environment) and not within the application code (python-environment). Consequently, any query only takes one trip to the database. If any part of the search is done in the application code, the query will take at least two trips to the database.

9.6 Conclusion

MOFFinder is able to successfully abstract data for ~300,000 hypothetical MOFs and perform structure searching. It provides fast access to over ~300,000 hypothetical MOFs for synthetic chemists, chemical and material engineers to find their MOFs of interest. The web interface is easy to use and enables adsorption and geometry searches with text-based structure type search via organic/inorganic SBUs, metal types and functional groups. In

addition, it provides a combination of substructure/similarity searches and property searches, and displays the 3D view of resulting MOFs for searches without the need for any Java plugins. These capabilities were achieved by intelligent design of the database and the integration of widely used open-source software that are offered by commercial organisations. Both examples provided in section 9.2 demonstrate the MOFFinder's capabilities for different applications. However, to use MOFFinder for any other application such as methane storage, MOFFinder needs to be updated by adding new FIELDS containing the data of desired applications for each MOF to the MOFSEACH_MOF table in the MOFFinder database. Indeed, MOFFinder is a flexible software which can be used for different purposes only by updating the database schema. From both examples in section 9.2, one can see that MOFFinder would also provide a fairly large number of hits for a given application. However, in order to get a subset of hits for which there are more known synthetic methods compared to others, the user needs to strategically create a query based on their chemical knowledge and experience. In general, one could start with a query including a substructure or similarity search of the existing ligands or ligands which are under consideration followed by a search for other properties, which would narrow down the set of promising structures for which the experimental synthesis could be attempted.

9.7 References

- (1) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, 28 (1), 235.
- (2) Steinbeck, C.; Krause, S.; Kuhn, S. *J. Chem. Inf. Comput. Sci.* **2003**, 43 (6), 1733.
- (3) Society, R.; Circle, T.; Forest, W. **2010**.
- (4) Chambers, J.; Davies, M.; Gaulton, A.; Hersey, A.; Velankar, S.; Petryszak, R.; Hastings, J.;

- Bellis, L.; McGlinchey, S.; Overington, J. P. *J. Cheminform.* **2013**, 5 (1), 1.
- (5) Irwin, J. J.; Sterling, T.; Mysinger, M. M.; Bolstad, E. S.; Coleman, R. G. *J. Chem. Inf. Model.* **2012**, 52 (7), 1757.
- (6) Hersey, A.; Chambers, J.; Bellis, L.; Patrícia Bento, A.; Gaulton, A.; Overington, J. P. *Drug Discov. Today Technol.* **2015**, 14, 17.
- (7) Kim, S.; Thiessen, P. A.; Bolton, E. E.; Chen, J.; Fu, G.; Gindulyte, A.; Han, L.; He, J.; He, S.; Shoemaker, B. A.; Wang, J.; Yu, B.; Zhang, J.; Bryant, S. H. *Nucleic Acids Res.* **2015**, 44 (D1), D1202.
- (8) Allen, F. H. *Acta Crystallogr. Sect. B Struct. Sci.* **2002**, 58 (3 PART 1), 380.
- (9) Boyd, P. G.; Woo, T. K. *CrystEngComm* **2016**.
- (10) Fernandez, M.; Woo, T. K.; Wilmer, C. E.; Snurr, R. Q. *J. Phys. Chem. C* **2013**, 117 (15), 7681.
- (11) Wilmer, C. E.; Leaf, M.; Lee, C. Y.; Farha, O. K.; Hauser, B. G.; Hupp, J. T.; Snurr, R. Q. *Nat. Chem.* **2012**, 4 (2), 83.
- (12) Fernandez, M.; Boyd, P. G.; Da, T. D.; Aghaji, M. Z.; Woo, T. K. *J. Phys. Chem. Lett.* **2014**.
- (13) <http://hmofo.northwestern.edu/hc/crystals.php>.
- (14) Cavenati, S.; Grande, C. a; Rodrigues, a E. *J. Chem. Eng. Data* **2004**, 49, 1095.
- (15) Han, S.; Huang, Y.; Watanabe, T.; Nair, S.; Walton, K. S.; Sholl, D. S.; Carson Meredith, J. *Microporous Mesoporous Mater.* **2013**, 173, 86.
- (16) Burtch, N. C.; Jasuja, H.; Walton, K. S. *Chem. Rev.* **2014**, 114, 10575.
- (17) Canivet, J.; Fateeva, A.; Guo, Y.; Coasne, B.; Farrusseng, D. *Chem. Soc. Rev.* **2014**, 5594.
- (18) Willems, T. F.; Rycroft, C. H.; Kazi, M.; Meza, J. C.; Haranczyk, M. *Microporous Mesoporous Mater.* **2012**, 149 (1), 134.
- (19) E.F. codd. *Commun. ACM* **1970**, 13, 377.
- (20) Don Chamberlin. 2009.
- (21) <https://www.oracle.com/>.
- (22) www.sybase.com/.
- (23) <http://www.ibm.com/analytics/us/en/technology/informix/>.
- (24) <https://www.mysql.com/>.
- (25) <https://www.postgresql.org/>.
- (26) Pavlov, D.; Rybalkin, M.; Karulin, B. *J. Cheminform.* **2010**, 2 (Suppl 1), F1.

-
- (27) <http://www.rdkit.org/>.
- (28) <http://lifescience.opensource.epam.com/indigo/>.
- (29) Weininger, D. *J. Chem. Inf. Comput. Sci.* **1988**, 28 (1), 31.
- (30) Karulin, B.; Kozhevnikov, M. *J. Cheminform.* **2011**, 3 (SUPPL. 1), P3.

10 Conclusion and Future Direction

This thesis demonstrated the potential of MOF materials for use in PSA systems for natural gas purification. It also showed the potency of computational tools in the discovery of optimal MOFs for natural gas purification. Indeed, presented in this thesis is a systemic, computational approach for aiding the design of optimal MOFs for natural gas purification. This involved validating the accuracy of our models for the prediction of gas adsorption on MOFs, employing efficient computational tools to screen the large database of hypothetical MOFs for NG purification, revealing important physical and chemical features relevant for high-performers, and finally developing new computational tools to accelerate the MOF discovery process for NG purification.

In Chapter 3, we investigated the adsorption of CO₂ on the first reported phosphonate monoesters-based MOF, CuBDPMe. The CO₂ adsorption properties, including the heats of adsorption and binding sites, were modelled via a combination of grand canonical Monte Carlo (GCMC) simulation and density functional theory (DFT) calculations. The GCMC-derived CO₂ adsorption isotherms revealed that the CO₂ uptake is sensitive both to the flexing of the structure and to the orientation of the methyl tether. Less than a 2% change in the lattice parameters resulted in a 20 % change in the CO₂ uptake on CuBDPMe. Similarly, a 2° change in the methoxy orientations which corresponds to a ~0.08 Å change in pore width resulted in a 25 % change in the CO₂ uptake on CuBDPMe. The GCMC-derived CO₂ binding sites suggested that the CO₂ molecules align along the channels, and do not align perpendicular to the channels as to interact with the copper atoms. It is conceivable that the perpendicular CO₂ binding is not

being captured by the GCMC simulations because these simulations do not account for orbital interactions and covalent bonding. To more thoroughly investigate the CO₂ binding, we performed a series of DFT calculations of the MOF with CO₂ in various configurations. All attempts to optimize the structure with CO₂ interacting with two copper atoms led to an optimized structure where the CO₂ molecules align itself along the channel. Thus, we are confident that the CO₂ binding in this material is along the channels, and does not involve any covalent bonding or coordination. This is in contrast to the dual end-on bridging interaction of CO₂ with two Cu centers that was hypothesized based on high heat of adsorption observed and the transannular Cu...Cu distance of 7.02 Å between pseudo-square planar Cu centers. The DFT-derived binding mode, along with the extreme dependency of uptake to the methoxy orientations, suggests a gate opening mechanism for CO₂ diffusion in the CuBDPMe. The agreement between experimental and simulated isotherms demonstrates the accuracy of parameters and methods used in this thesis.

In Chapter 4, we studied MIL-47 (a strong candidate MOF for NG purification) and its functionalised derivatives (i.e., MIL-47-X, X = -Cl, -Br, -CH₃, -CHO, -OH, -OCH₃, -COOH, -NH₂, NO₂) for their performance in natural gas purification in a systematic manner. All functionalised derivatives were hypothetically constructed from the pristine MIL-47, where one hydrogen atom of each ligand within the MIL-47 unit-cell was randomly replaced with a desired functional group. All functionalised structures were subjected to the classical force field optimizations, and subsequent geometry optimization at the DFT level. To investigate the performance of MIL-47 and its functionalised derivatives in NG purification, we calculated the CO₂ working capacity and

CO₂/CH₄ selectivity for each MOF from GCMC simulations with a 1:9 CO₂-CH₄ mixture at typical conditions used in PSA based natural gas purification. The results from GCMC simulation showed that introducing polar non-bulky functional groups in the MOF MIL-47 solid leads to an enhancement of its CO₂ working capacity and to a significant improvement of the CO₂/CH₄ selectivity. Particularly, MIL-47 functionalized with NO₂ offers the highest selectivity, which is two times as high as MIL-47. We also compared the performance of MIL-47 and its functionalised derivatives with other materials whose CO₂ working capacity and CO₂/CH₄ selectivity are reported in literature. From the comparison, MIL-47-NO₂ showed a relatively good combination of CO₂ working capacity and CO₂/CH₄ selectivity, suggesting it as a possible candidate for use in PSA unit for NG purification.

Chapter 5 presented a comprehensive simulation study of water adsorption in a variety of MOFs including MIL-47, ZIF-8, Al(OH)(1,4 NDC), UIO-66, Zn-DMOF-A, Zn-DMOF-TM, CAU-OH-10, MOF-806, and Cu-BTC. We evaluated the performance of six “standard” water models (i.e., SPC/E, TIP3P, TIP4P, TIP4P/Ew, TIP5P, and TIP5P/Ew) by calculating pure water isotherms and comparing them to experimental data for each of these MOFs. The simulation results showed that the saturation pressures are different for different water models, which is most likely attributed to the electrostatic parameters of these models. By comparing the predicted saturation pressures with experimental values, it was found that four site models (i.e., TIP4P and TIP4P-Ew) performed better than other models for all MOFs considered. Our simulations also revealed that water adsorption in hydrophobic and hydrophilic MOFs are considerably different. Hydrophilic MOFs gradually uptake water, whereas hydrophobic MOFs show little

uptake until the saturation pressure where there is rapid uptake. Snapshots from GCMC simulations in a hydrophobic MOF indicates that one pore is filled to saturation before water adsorption is seen in another pore. Interestingly, it was found that water clusters formed in MIL-47-type MOFs are somehow different from those in the bulk phase. Most water molecules prefer to form four hydrogen bonds in the bulk phase while in MIL-47 most water molecules have three hydrogen bonds. We also investigated the effect of water on natural gas purification in these MOFs using the GCMC simulations of a 1:9 CO₂-CH₄ mixture in the presence and absence of water. We found that water has negligible effects on natural gas purification in hydrophobic MOFs under the humidity thresholds allowable for natural gas transport in pipelines.

In Chapter 6, we computationally screened 324,500 hypothetical MOFs for their performance for natural gas purification in a high-throughput fashion based on the protocol we defined, including screening conditions (e.g., adsorption pressure and temperature), simulation methods, and simulation parameters. We calculated the performance parameters (i.e., CO₂ working capacity and CO₂/CH₄ selectivity) for each MOF, and we found 3900 high-performing MOFs whose selectivity and working capacity are above zeolite 13x commercially used in PSA systems. From the screened data, we identified key structural and chemical features relevant to high-performers. A pore size smaller than 6.6 Å and surface area higher than 2299 m²g⁻¹ were the most frequent geometric features among the high-performers. MOFs with NO₂ and CHO functional groups were found to dominate the high-performers; 50% of MOFs which have selectivity greater than 12 contain NO₂ and HCO functional groups. A similarity search of the

binding sites in the high performing MOFs identified two common CO₂ binding motifs: 1) in between parallel aromatic rings separated roughly 7 Å and 2) adjacent to a bridging oxygen in an oxo-vanadium chains. These binding motifs were discovered by extracting the binding sites within all high-performers and using a clique detection method to identify the most common substructures among them. Finally, we proposed several MOFs, which exhibit CO₂ working capacity higher than 2 mmol/g and CO₂/CH₄ selectivity around 20, as “hit” materials for targeted synthesis.

In Chapter 7, we developed machine learning (ML) *QSPR* models for CO₂ working capacity and CO₂/CH₄ selectivity at conditions typical of natural gas purification with PSA systems using the purely geometric properties of pore size, void fraction and surface area. The models were trained on the GCMC calculated adsorption properties of 32,450 MOFs using binary decision trees (DTs) and nonlinear support vector machines (SMVs). DT analysis provided clues and principles to follow in the rational design of MOFs, based on optimal range of pore size, void fraction and surface area. The “rule-of-thumb” for a MOF to possess a CO₂/CH₄ selectivity >10 is to have void fraction lower than 0.27 and pore size less than 6.6 Å, while in order to have CO₂ working capacity >4 mmol/g is to have pore size less than 8.5 Å and gravimetric surface area greater than 2299 m²g⁻¹ MOF. Meanwhile, the SVM models allowed rapid identification of high-performing MOF candidates by discarding majority of the low-performing materials in the database. This saves computational resources by allowing one to run GCMC simulations on a reduced fraction of the database. It was demonstrated on a test set of 292,050 MOFs that the classifiers could recover ~900 of the top 1000 MOFs in the database, while only flagging ~25%

of the total database for more computationally expensive screening. This approach could be applied to more efficiently screen of large structural libraries by avoiding the brute force application of GCMC simulations on all materials. The approach could also be applied to other classes of materials as long as the relevant data is made available to calibrate the QSPR models.

In **Chapter 8**, we presented machine learning QSPR models that use a subset of hologram descriptors generated from the Voronoi network of MOFs and atomic property weighted radial distribution functions (AP-RDF) descriptors for the recognition of high-performing MOFs for natural gas purification. The CO₂-accessible Voronoi network of MOFs was calculated and resulted in 2176 hologram descriptors for each MOF. To reduce the number of descriptors, a descriptor selection was performed in two stages: first, variance-based filtering was used to remove irrelevant descriptors, which was followed by the application of a custom genetic algorithm to efficiently select the “best” subset of hologram descriptors. The selected descriptors were used in the model development procedure where the models were trained on the GCMC calculated adsorption properties of 32,450 MOFs using nonlinear support vector machines (SVMs). To validate the models, they were applied to a test set of ~300,000 MOFs which were not part of the training procedure. Similar to the work presented in **chapter 7**, the SVM models allowed the rapid identification of high-performing MOF candidates by discarding majority of the low-performing materials in the database. Specifically, they recognised 90% of high-performing MOFs in the test set by flagging only 20% of the test set as a high-performing set. We also compared the performance of the hologram and AP-RDF SVM models with the performance of simple geometrical descriptors-based SVM model using a cross-validation

method on 32,450 MOFs. From the comparison, it was found that both AP-RDF and Voronoi hologram classifiers retrieved about 90% of the high-performing MOFs while discarding about 80% of all MOFs. On the other hand, the simple geometric classifiers only retrieved 75% of all the high performing MOFs while only discarding ~76% of all MOFs.

Finally, in **Chapter 9**, we developed a novel cheminformatics tool, *MOFFinder*, to aid in the discovery of new MOFs for CO₂ scrubbing from natural gas. It provides fast access (in a second) to over ~300,000 hypothetical MOFs for researchers to find MOFs of interest. It provides easy-to-use search tools where the user can perform searches based on the adsorption properties and/or geometric characteristics. It also provides substructure/similarity searches, which have not been previously implemented in any other database of porous materials, allowing for graphical input of the substructures. Moreover, it displays the 3D view of resulting MOFs for searches without the need for any Java plugins. All these capabilities were achieved by intelligent design of the database and the integration of widely used open-source software that are offered by commercial organisations.

10.1 Future directions

In Chapter 5, we systematically evaluated the performance of conventional non-polarizable water models for predicting the uptake of water in MOFs, and investigated how water affects the adsorption of CO₂/CH₄ mixture. It was found that there is no water model that could predict the water adsorption in all of the MOFs considered in chapter 4. One future direction of study would be to specifically develop a water model for water adsorption in MOFs and to explore the use of polarizable water models. However, one significant barrier to this is

that the published experimental water adsorption isotherms are quite limited. Thus, future work in this direction would likely require collaborations with experimental labs. If a new water model could be developed, it would be an important contribution to the field as water present in essentially all industrial gas streams and must be considered when considering materials for practical applications.

Perhaps the biggest criticism of the work presented in chapter 6 is whether a broad range of MOF structures have been sampled to provide the design principles. The 324,500 hypothetical MOF structures in our database screened for NG purification were generated from 66 SBUs and 19 functional groups. These SBUs were chosen from existing MOF crystal structures, and varied significantly in their geometries, number of connection sites and chemical composition. As a comparison, the Northwestern University's MOFs database that was published in *Nature Chemistry* and screened for methane storage and CO₂ capture applications includes five kinds of inorganic SBUs and 13 kinds of functional group, while our database contains seven kinds of inorganic SUBs and 19 kinds of functional groups. However, it would be difficult to rigorously prove that our database is sufficiently diverse. Therefore, the design principles proposed from this work remains conjecture at this point.

Recently, Boyd et al.¹ developed a MOF generating algorithm, which used topological nets of existing MOFs as a basis for building MOFs by searching for compatible organic and inorganic SBUs and placing them in the vertices and edges of the nets, thereby generating MOFs with the targeted topology. Using this algorithm, they generated over one million hypothetical MOFs which are currently undergoing a screening process for methane storage

applications. Thus, one future direction of study would be to screen this new database using the screening protocol defined in this chapter.

Another future direction, which is very important from our point of view, will be the successful synthesis and validation of the materials we suggested as ideal candidates for use in PSA unit for CO₂ separation from natural gas. The validation would include analysing their thermal and chemical stability, and investigating the kinetics aspect of the separation of CO₂ from CH₄ using these materials.

In chapters 7 & 8, we developed machine learning classifier models of CO₂/CH₄ separation parameters using a variety of descriptors (i.e. pore size, void fraction, surface area, Voronoi network, and AP-RDF) of the materials. We showed that these QSPR models could be used to pre-screen large libraries of materials for CO₂/CH₄ separation application, which could save computational resources by running GCMC simulations on a reduced fraction of available data. While this is informative, a future direction of study would be to develop QSPR models that predict the separation performance (e.g. CO₂ working capacity) of MOFs instead of only classifying them as high or low performing. Currently, one of the Woo lab's members is working on this using the approach presented in these two chapters.

In Chapter 9, MOFFinder was presented which provides fast access to over 300,000 hypothetical MOFs, properties and associated information relevant for NG purification. Owing to its structure design, it potentially provides access to over millions of MOF structures. One possibility of future work would involve the commercialization of MOFFinder. This could be achieved by updating its database for any gas separation related application, adding millions of

MOF structures that the lab is currently screening, and integrating new *utilities* to its framework. Utilities of interest would be QSPR tools that predict performance for MOFs provided by users. Another utility of interest would be a MOF generation program, which predicts a MOF's structure based on the SBUs and net topology provided by users.

10.2 References

- (1) Boyd, P. G.; Woo, T. K. *CrystEngComm* **2016**.